THE DESIGN OF GENE REGULATORY NETWORKS WITH FEEDBACK AND SMALL NON-CODING RNA

ANDREAS WILLIAM KISLING HARRIS

Thesis submitted for the

Degree of Doctor of Philosophy

Exeter College

University of Oxford

Michaelmas 2016

ABSTRACT

The objective of the field of Synthetic Biology is to implement novel functionalities in a biological context or redesign existing biological systems. To achieve this, it employs tried and tested engineering principles, such as standardisation and the design-build-test cycle. A crucial part of this process is the convergence of modelling and experiment. The aim of this thesis is to improve the design principles employed by Synthetic Biology in the context of Gene Regulatory Networks (GRNs). Small Ribonucleic Acids (sRNAs), in particular, are focussed on as a mechanism for post-transcriptional expression regulation, as they present great potential as a tool to be harnessed in GRNs. Modelling sRNA regulation and its interaction with its associated chaperone Host-Factor of Bacteriophage Qβ (Hfq) is investigated. Inclusion of Hfq is found to be necessary in stochastic models, but not in deterministic models. Secondly, feedback is core to the thesis, as it presents a means to scale-up designed systems. A linear design framework for GRNs is then presented, focussing on Transcription Factor (TF) interactions. Such frameworks are powerful as they facilitate the design of feedback. The framework supplies a block diagram methodology for visualisation and analysis of the designed circuit. In this context, phase lead and lag controllers, well-known in the context of Control Engineering, are presented as genetic motifs. A design example, employing the genetic phase lag controller, is then presented, demonstrating how the developed framework can be used to design a genetic circuit. The framework is then extended to include sRNA regulation. Four GRNs, demonstrating the simplest forms of genetic feedback, are then modelled and implemented. The feedback occurs at three different levels: autoregulation, through an sRNA and through another TF. The models of these GRNs are inspired by the implemented biological topologies, focussing on steady state behaviour and various set-ups. Both deterministic and stochastic models are studied. Dynamic responses of the circuits are also briefly compared. Data is presented, showing good qualitative agreement between models and experiment. Both culture level data and cell population data is presented. The latter of these is particularly useful as the moments of the distributions can be calculated and compared to results from stochastic simulation. The fit of a deterministic model to data is attempted, which results in a suggested extension of the model. The conclusion summarises the thesis, stating that modelling and experiment are in good qualitative agreement. The required next step is to be able to predict behaviour quantitatively.

In memory of Stuart Alexander Hugh Harris,

whose life was a lesson in how to be happy and spread love

in the face of severe trials.

24^{th} of April, 1978 – 1^{st} of July, 2016

ACKNOWLEDGMENTS

I begin by thanking my supervisor, Antonis Papachristodoulou, without whom it is needless to say, that this document would never have been written. Thank you for your endless guidance, patience, understanding, and trust in me backed up by the continual belief, not always shared by me, that I would be able to pull this off at some point. You have taught me an uncountable number of lessons over the last three and half years, for which I will always be grateful.

Now to Ciarán L. Kelly: thank you for all your tips, tricks, banter and guidance in the lab, as well as your unfailing good humour, energy and laughs. I find myself shouting 'bosh it in!' in all sorts of contexts these days.

Thanks in equal measure go to Edward J. Hancock, who guided me through the initial phase of my project. You helped shape the thinking I brought forward, and I think you might recognise parts of your approach in some of my thesis.

On a more general note, can I thank the SYSoS group members at Oxford, both past and present, these are James Anderson, Giorgio Valmorbida, Axel Nyström, Thomas Prescott, Richard Mason, Xuan Zhang, Dhruva Raman, James Scott-Brown, Mohamadreza Ahmadi, Yang Zheng, Harrison Steel, Aivar Sootla and Nick Delalez. Many of you have had to put up with me talking at you at regular intervals. Please accept my apologies for this. I hope you all have enjoyed the interactions as much as I have.

A special thanks to the chaps in ETB30.23: Ross Drummond, Mehdi Imani Masouleh, Manuel Rainer Schaich, Chen Xiaoqing and Ming Lee. Thanks for the AoEII games and the crazy conversations on topics both relevant, but mainly irrelevant to our work.

Thank you also to the Lower Ground Floor crew in Biochemistry, who were always supportive and helpful when a theorist asked stupid questions in the lab. I would particularly like to mention Felicity Alcock and George Wadhams.

From a personal dimension, my experience at Oxford was made particularly colourful by the people I met there. Most important of all is Florence A. Benn and, by extension, the Benn clan at large. Thank you Flo, for the time you spent with me and a million other things. You have an energy and love of life that is unique. This must have been, at least in part, bestowed upon you by your wonderful extended family, the most kindhearted, interesting, warm and welcoming collection of people I have ever met.

An non-Oxford friend, that deserves to be mentioned, is Christopher Namih. He has helped me through all my ups and downs for almost ten years now and is as dear to me as any family member could be. Thank you for your time and patience. I hope you benefit from the relationship as much as I do.

Now to thank my family, without whose support and encouragement throughout my life I would not be who I am today. Mum, thanks for your bottomless love and generosity. Your heart is gold and you are the best person I know. Dad, thank you for all the guidance and support, and the memorable experiences. They have supplied me with endless stories. Katrine and Matthias, I love you. Let us always make sure we remain such a strong team. Leticia Panzo, second mother to all three of us. Thank you for loving us, as if we were your own. Thanks also goes to all the family in Denmark: Camilla, Niels, Andrea, Jens, Carl-Johan, Frans-Ulrik, Philip, Caroline, Victoria and August. More dinners at Brobyhus, please !

Finally, can I thank Exeter College, in particular the MCR, ECBC and the staff, for the social life and the distractions that followed. It is in large part due to these, that I have managed to retain my sanity over the last four and a half years.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

2-D	Two Dimensional
aTc	Anhydrotetracycline
CAP	Catabolite Activator Protein
DNA	Deoxyribonucleic Acid
DNAP	DNA Polymerase
EZ	EZ Rich Media
GRN	Gene Regulatory Network
GFP	Green Fluorescent Protein
Hfq	Host-Factor of Bacteriophage Qβ
IPTG	Isopropyl-β-D-1-Thiogalactopyranoside
LB	Luria Broth
mRNA	Messenger Ribonucleic Acid
M9	M9 Minimal Media
ODE	Ordinary Differential Equation
P_{rhaBAD}	Rhamnose Promoter
P_{tet}	*tet*-promoter
PID	Proportional Integral Derivative
RhaS	Rhamnose Activator
RBS	Ribosome Binding Site
RNA	Ribonucleic Acid
RNAP	RNA Polymerase
rRNA	Ribosomal RNA

SDE Stochastic Differential Equation

sRNA Small Ribonucleic Acid

sfGFP super-folder Green Fluorescent Protein

SBML Systems Biology Markup Language

TetR Tetracycline Repressor

tRNA Transfer RNA

TF Transcription Factor

UTR Un-Translated Region

1 INTRODUCTION

Synthetic Biology is the Engineering of Biology. The fields objective is to design and construct novel functionalities in a biological context through the re-design of naturally occurring systems or the assembly of completely novel ones. It is a rapidly growing field that has received a lot of attention. Scientists have known the potential of the field for a while and set very ambitious goals [3, 14, 47, 115, 141, 152, 174, 181, 192], but more recently the public and notably politicians have come to expect the field of Synthetic Biology to yield great leaps in technology and industry, which has lead to considerable public investment [48, 187].

Synthetic Biology spans a wide range of fields. At the extreme synthetic end, DNA nanotechnology hijacks DNA and uses it as a building material with which to construct nanoscale structures which function mechanically [166]; a task that DNA did not evolve to perform. At the other extreme, pathway engineering takes metabolic and signalling pathways in cells and redesigns them, making changes to optimise for a certain yield or other design objectives [131, 152].

A vital element of the approach taken to tackle both the above challenges is the combination of theoretical and experimental tools. In the case of DNA nanotechnology, an intimate understanding of Watson-Crick base-pairing has facilitated the development of tools such as NUPACK [59] that accurately predict DNA annealing patterns. At the other extreme, mathematical tools such as Flux Balance Analysis [160] assist in the comprehension, study and design of metabolic pathways.

1.1 AN OVERVIEW OF SYNTHETIC BIOLOGY

This synergy between mathematics and experiment is the keystone of the field of Synthetic Biology and its sister field, Systems Biology. Early examples of such synergies are found in two papers, that are often identified as seminal papers in the field of Synthetic Biology. These were published back to back in the same issue of Nature in 2000. The synthetic mechanisms studied in these papers are the Toggle Switch [76] and the Repressilator [66]. These are two and three gene circuits respectively, that are studied using the combination of mathematics and experiment. The first consists of two repressors repressing each other's expression. This system is bi-stable with

either one gene or the other being expressed, hence the name 'Toggle Switch'. This behaviour was predicted mathematically and then confirmed experimentally. The switch can be toggled from one state to the other using either chemical or thermal induction. The second consists of three repressors connected in a loop: first repressing second, second repressing third, third repressing first. In this case, the modelling was found to poorly predict the observed behaviour of the synthetic circuit. The models exhibited sustained oscillations with a constant period and what was observed were noisy oscillations that were often not sustained and had fluctuating periods. The above systems were inspired by natural systems, such as the λ-phage switch [173] in the case of the Toggle Switch and the circadian clock found in cyanobacteria [107] in the case of the Repressilator, but do not consist of any parts that are present in these natural circuits.

Another equally important paper [23], published a matter of weeks after the Toggle Switch and Repressilator, performs a similar study of a synthetic circuit studying the effects of feedback. The simple circuit studied consists of a single gene that inhibits its own expression. The benefits of the genetic motif are laid out in detail both mathematically and experimentally, showing a reduction in noise of the circuit and an increase in stability. It was shown shortly after that this genetic motif, called negative autoregulation or autorepression, is widely used in nature to regulate a large number of Transcription Factors (TFs) [179].

These examples fit neatly into the area of focus of this thesis; the design of *in vivo* dynamical systems exploiting the cellular process of genetic expression. Such systems are broadly referred to as Gene Regulatory Networks (GRNs). The approach taken is circuit building from the bottom up, by connecting biological parts. This is analogous to early electrical engineering, where basic parts, such as resistors, capacitors, and inductors, were connected in a huge variety of configurations to form larger complex circuits with increasingly intricate behaviours.

The vast majority of parts used to build GRNs have been mined from nature directly or are based on a natural system which has then been engineered. A good example of this is the widely used fluorescent reporter protein super-folder Green Fluorescent Protein (sfGFP) [164]. The original fluorescent protein was taken from the jellyfish *Aqueorea victoria*, which was then engineered for stability. The parts in general consist of genes encoding various functional elements. The range of tools available is broad and includes elements such as the TFs mentioned above, but also many others. To mention a few: light sensitive elements [144], orthogonal transcription and translational systems [12] that do not interfere with the endogenous system in the cell, degradation tags that speed up degradation of protein in the cell [37], and libraries of terminators [45]. A recent review presenting a broad overview of the tools available to a synthetic biologist can be found in [31].

Through the combination of these parts, circuits are engineered and metabolic pathways are manipulated. A great example of this is a TF based GRN published by Stricker *et al.* in 2008 describing a circuit capable of sustained oscillation with a tuneable period [200]. This circuit managed to do what the original Repressilator could not. Even the Repressilator was recently shown to exhibit robust and sustained oscillation after a considerable amount of tuning and modification [168]. Another impressive example of what can be achieved is found in [226]. Here, engineering techniques were used to improve the yield of a metabolic pathway inserted into the bacteria *E. coli* producing biodiesel. A transcription factor capable of sensing a key intermediate is used to ensure stability of the production by reducing the toxicity to the host, increasing the final yield three fold.

Despite this success, the combination of well characterised parts into larger modules is not completely straightforward. Well defined parts used to construct larger circuits are often found to behave differently in the new context of the circuit [41]. The interconnection of parts also places a load on the parts upstream [55]. In addition to this, transforming a host with foreign DNA often burdens the host, sapping resources that would have been used elsewhere and limiting growth [29]. This is in stark contrast to the implementation of electronic circuits, where the behaviour of individual parts, such as resistors and inductors, is independent of the larger context. Such issues can be dealt with in a number of ways. One is the inclusion of higher level regulatory mechanisms in models [84], another is the careful selection and tuning of parts [15].

Due to these challenges and the lack of a formalised design methodology, the implementation of GRNs has in essence remained an artisanal craft, where every system is laboriously implemented and tuned in its specific context [36]. The quick design iterations guided by predictable modelling, which is a key feature of other fields of engineering, still eludes Synthetic Biology as a whole.

Inspired by other fields of engineering, tools are being developed to facilitate the design process in the context of GRNs. Systems Biology Markup Language (SBML) is a widely used standardised language for sharing models [101], which has driven the development of a large number of simulation and analysis tools employing deterministic and stochastic approaches; one such tool, CudaSim [227], is used in this project. Repositories and libraries of parts are being created, most notably the iGEM BioBricks library, which span out of Massachusetts Institute of Technology in 2012[1]. Once the necessary parts for a GRN have been selected, software such as j5 [98], automatically provides an assembly protocol to objectives such as cost efficiency using a wide range of cloning methods such as Gibson [80] and Golden Gate [67].

1 http://parts.igem.org

The final aim of this progression is complete design automation in a biological context, in a manner similar to current capabilities for electronic systems [56]. Having selected the input-to-output behaviour of a system to be built, a user can employ Electronic Design Automation tools to design, analyse and test the system prior to implementation. An ambitious project, called Cello [156], provides a such a platform in the context of biological computation. For a specified binary input-to-ouput relationship, a genetic design is provided that behaves accordingly without the need of further tuning. The platform has been tested and shown to design GRNs with up to 55 parts with an input-output accuracy of 92%. This is an impressive feat, but this platform is limited as it, so far, only functions in a binary paradigm. The extension of such a platform to the continuous paradigm, in which these biological systems exist, would revolutionise the field.

1.2 HARNESSING FUNCTIONAL RNAS

One major class of parts not mentioned above is functional Ribonucleic Acid (RNA)s [105]. These extend the regulatory toolbox to post-transcriptional control. RNA devices were successfully engineered early on in the development of the field [106], but the main focus was on protein devices such as TFs [36] as there was a large natural repertoire to draw from and their behaviour spans a large dynamic range. Recently there has been a resurgence in interest in RNAs as a means of regulation [43, 165] for a number of reasons. In large part this is due to a vastly increased understanding of the role of natural regulatory RNAs, but is also due to the predictability of the folding and interactions of RNA structures [165]. The natural RNA regulators perform a broad range of functions [86, 217], and harnessing these in synthetic circuits offers a large number of benefits. Functional RNAs are short and therefore use little energy and are quick for a host cell to produce. This means they are both genetically compact and constitute a lower burden to the host. They also have the potential to propagate signals rapidly as is demonstrated in [202], as their dynamics are tied to the high degradation rate of RNA. However, due to the high degradation rate, functional RNAs are more noisy than their protein-based counterparts [142]. In addition to this, off-target effects are common and can interfere with endogenous systems in the cell [134]. Both these issues must be taken into account when designing. Despite this, with correct implementation, sRNAs can be powerful tools in the optimisation of metabolic pathways [151, 197].

There is a subclass of these functional RNAs that are native to prokaryotes called Small Ribonucleic Acid (sRNA)s. These classically regulate translation of a target Messenger Ribonucleic Acid (mRNA). The mechanism is well studied from a biochemical [11, 85, 133, 199, 207] and

regulatory perspective [1, 103, 127, 129, 142, 182, 190], and provides a method of thresholding expression in a number of ways [128].

Given the large potential of these mechanisms, a focus of this work is to take the next step and include sRNA regulation in implemented GRNs. Initially, Chapter 3 focusses on the modelling of sRNA regulation in the context of a GRN, studying its interaction with the catalyst chaperone protein Host-Factor of Bacteriophage Qβ (Hfq) and looking at different common implementation set-ups: a one plasmid [102] and a two plasmid set-up [129]. In Chapter 5 two circuits including sRNA regulation are modelled in detail and a biological implementation is presented. Finally in Chapter 6, data from the circuits studied in Chapter 5 is presented.

1.3 HARNESSING FEEDBACK

Feedback is a crucial tool in engineering and is implemented ubiquitously in man-made systems. Natural systems are also highly dependent on feedback to accurately self-regulate. Three studied examples of this are the regulation of chemotaxis [223], the calcium homeostasis in cows [65] and the heat-shock response system [120]. These represent feedback across all biological scales from intracellular to organism level. Feedback is also present in a number of the above mentioned synthetic systems: the Toggle Switch [76], the Repressilator [66], the Autorepressor [23] and the circuit exhibiting sustained oscillations [200] all harness feedback to achieve their objectives.

Feedback is an essential tool as it is a means of overcoming the robustness problem, arguably the central challenge faced by Synthetic Biology [13]. Robust systems perform reliably when exposed to a wide range of perturbations. In the context of a cell, these can either be internal or environmental. The correct implementation of feedback will facilitate the predictable design and implementation of robust GRNs, and therefore the scale-up of systems constructed from smaller GRN modules [3, 174].

Methods for analysis and design of feedback systems are readily available in the field of Control Engineering [61]. Some of these ideas have had extensive application in Systems Biology [8, 104] and more recently in Synthetic Biology [3, 157, 170, 191]. The example of Oishi and Klavin's work from 2011 [157] is particularly interesting as this presents a linear design framework for the implementation of synthetic circuits harnessing feedback. This framework employs Deoxyribonucleic Acid (DNA) as the active species, relying on the highly developed understanding of Watson-Crick base pairing to accurately predict interactions between individual strands resulting in the desired behaviour. As all parts of these circuits are made of DNA, these

circuits function *in vitro*, outside a natural biological host in highly controlled conditions. But there are strong predictive modelling techniques for such circuits [122], and in combination with the feedback design methodology presented in [157], this offers an excellent basis for the correct implementation of feedback in this context.

The second focus of this project is the implementation of feedback in GRNs. To this end Chapter 4 develops a linear design framework focussing on regulation by TFs. This includes a block diagram method to visualise systems and extract transfer functions and an extension to include sRNA regulation. A biological design example is presented with a proposed biological implementation to show how the framework can be used to design a real system. In addition to this, Chapter 5 studies four simple GRNs employing negative feedback. These are then implemented and experimental data for the four circuits is presented in Chapter 6 and compared to the models.

1.4 OVERVIEW OF THE THESIS

In summary, the overall aim and objective is to improve engineering techniques used to design and implement GRNs. In Chapter 2, the thesis begins with an in depth look at the biological mechanisms involved in GRNs and the mathematical modelling techniques used. The post-transcriptional sRNA regulation mechanism is focussed on, as it promises benefits over classical regulatory units such as TFs when harnessed in GRNs. This regulation mechanism is studied in detail in Chapter 3. In addition to this, a linear design methodology to harness and engineer feedback in the context of GRNs is developed in Chapter 4. Chapter 5 studies four GRNs in depth, two of which employ sRNA regulation. These are then implemented and experimental data is presented and compared to modelling in Chapter 6. Finally, Chapter 7 summarises the results of the thesis and gives a perspective on the future of the project.

2 BACKGROUND

The aim of this chapter is to provide the reader with an introduction to the biology and mathematics, on which the thesis is based. As the work performed here sits neatly between Engineering and Biology, elements of both subjects are covered to make the text accessible to readers approaching from either discipline.

Cellular and molecular biology is the focus of the first half of the chapter. This describes the context, environment and specific mechanisms relevant to the work presented in this thesis. As the focus of the thesis is Gene Regulatory Networks (GRNs), the biological section presents the hereditary mechanism, the process of gene expression and mechanisms of gene expression regulation at a transcriptional and translational level relevant to this project. This is done first in a general context and then more specifically in the context of *Escherichia coli*, the host chosen for this work. Key aims are presenting the biology from an engineering perspective, presenting a number of tools useful to a synthetic biologist with the aim of designing and implementing GRNs, and making it clear what can be designed when implementing systems.

The information outlined below, if not cited, can be found in standard texts such as Molecular Biology of the Cell [6] and Molecular Genetics of Bacteria [194]. These are recommended for introductory reading; [194] in particular due to its specific treatment of bacteria. A great recent review article covering the specifics of GRN design and implementation is [31]. This covers a much wider range of building blocks and tools than is presented here.

In the later sections of this chapter, it is described how the biological mechanisms presented can be translated into models. Finally, a series of tools and methods for the study, analysis and design of these systems are introduced. An understanding of calculus and differential equations is assumed for the mathematical section.

Let us begin with a tour of the world from the wet and squishy perspective of a cell...

2.1 BIOLOGICAL BACKGROUND

'Life: The condition that distinguishes animals, plants, and other organisms from inorganic or inanimate matter, characterised by continuous metabolic activity and the capacity for functions such as growth, development, reproduction, adaptation to the environment, and response to stimulation.'

— Oxford English Dictionary, Def. I1.d.

Life is found in virtually every environment on Earth [208]. It surrounds us visibly in the shape of plants, animals, insects and of course the ubiquitous *Homo sapiens*, but it also manages to flourish in most extreme environments on the planet. Organisms have evolved to survive at extreme temperatures, pressures and levels of pH, salinity, desiccation, oxygen and radiation [180]. Specific examples of these aptly named extremophiles exist in Lake Whillans [171], a body of water found 800 metres below the surface of the Antarctic ice sheet in the pitch black, and in the Sahara [21], one of the driest, hottest and most exposed places on earth.[1]

All life can be broken down into the unit of living matter: the cell. Cells consume nutrients found in their environments, which they process to allow them to grow and replicate. They employ a vast number of different strategies to survive; they can exist as single cells, in communities, or in multicellular lifeforms. Taking the human as an example, it was recently estimated that body contains approximately 3.72×10^{13} human cells [27]. This vast number of cells are organised into a large number of different organs and tissues and are present in body fluids such as lymph and blood. These cells perform hugely different tasks depending on their context. Liver cells, or hepatocytes, are, among other purposes, the bodies main source of lipoproteins that allow the transportation of lipids through the body. Muscle cells, or myocytes, work together contracting and relaxing when stimulated by the nervous system. Despite the various roles played by human cells in different organs or contexts, they all[2] carry identical genetic codes that are the precise biological recipe book unique to each individual human. These clusters of specialised human cells all cooperate and live in symbiosis with a huge range of non-human cells to form a fully functional human, communicating via an intricate network of chemical balances. The best known example of this symbiosis is the gut microbiome, a complex community of microorganisms, mainly bacteria, found in the digestive tracts of animals. The composition of the gut microbiome is highly dependent on the host species and the diet of the specific host.

1 The name of the desert extremophile is *Deinococcus radiodurans*. It can handle so many different extremes that is has been referred to as 'Conan the Bacteria'.
2 With the notable exceptions of red blood cells and platelets.

The immense diversity of nature can be contrasted with the high degree of consistency observed in the fundamental building blocks and mechanisms involved. As mentioned already, all cells can replicate and grow and therefore they exist far from chemical equilibrium. They do this by consumption of energy taken from their environment. If this process stops, they decay towards chemical equilibrium and death. Cellular boundaries are all defined by the plasma membrane, a lipid bilayer, that acts as a selective barrier, letting only specific species in and out. By using the plasma membrane the cell can intricately control the internal chemical environment of the cytoplasm, functioning as a molecular factory processing the absorbed nutrients and synthesising chemical species it needs to survive, grow and replicate.

The next two sections explain what two of these fundamental building blocks are, namely the nucleic acids and proteins, and then explain how these are linked through the Central Dogma of Molecular Biology.

2.1.1 *Deoxyribonucleic Acid*

The nucleic acids, Deoxyribonucleic Acid (DNA) and Ribonucleic Acid (RNA), are essential to all life. Each and every cell carries its genome encoded in DNA. This is a linear chain of chemicals that stores all the hereditary information needed for the cell to live and replicate. A simple Two Dimensional (2-D) representation of a section of DNA can be seen in Figure 2.1a.

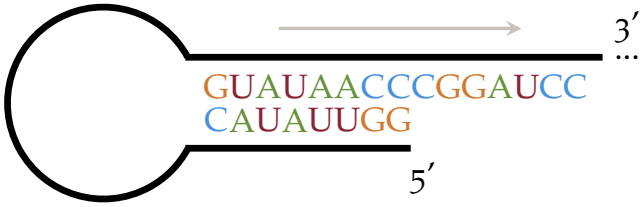
DNA is a double helix[3] consisting of two polysaccharide chains linked together by the four nucleotide bases: guanine (G), adenine (A), thymine (T) and cytosine (C). It is the specific sequence of these bases that defines the genetic sequence of a living organism. This is called the primary structure of DNA.

A DNA strand is read in a direction defined by the orientation of the polysaccharide backbone, from the 5′ end to the 3′.[4] These names are taken from the chemical labels of the carbon atoms in the saccharide subunits of the backbone. The standard double stranded conformation of DNA consists of two strands lying antiparallel to each other: one 5′ to 3′ the other 3′ to 5′, as shown in Figure 2.1a.

The double helix, defined as DNA's secondary structure, is formed by the pairing through hydrogen bonds of the bases opposite each other on the strands. The bases pair in a specific manner: guanine to cytosine (G to C) with three hydrogen bonds and adenine to thymine (A to T) with two hydrogen bonds. Strands with higher numbers of guanine and cytosine base pairs

3 When asked what Watson and Crick discovered, I usually reply: 'Rosalind Franklin's notes, of course.'
4 These are pronounced 'five-prime' and 'three-prime' respectively.

(a) DNA



(b) RNA

Figure 2.1 – 2-D representations the nucleic acids in common conformations. The nucleotides guanine (G), adenine (A), cytosine (C), thymine (T) and uracil (U) are present. The polysaccharide backbones are represented by the black lines and the dots (...) represent the continuation of the strands. The directionality of the strands is given by the grey arrows and runs from 5′ to 3′. (a) A section of double stranded DNA. The sections are perfectly complementary and the two strands are antiparallel. Letters representing nucleotides opposite each other on the strands are base-paired. This is the conventional way of representing DNA in 2-D. (b) This is a section of an RNA strand demonstrating RNA's ability to base-pair with itself forming a hairpin structure. The looped section of backbone without labelled nucleotides indicates a section of RNA with no complementarity.

bind more tightly than those with large numbers of adenine and thymine base pairs. [5] The two DNA strands, that make up the double stranded helix, base-pair at every nucleotide and are therefore said to be perfectly complementary.

DNA is sturdy ensuring the long term storage of crucial genetic information and therefore costs the cell a large amount of energy to produce. DNA essentially acts as a passive information storage device for the cell.[6]

2.1.2 *Ribonucleic Acid*

RNA is very similar to DNA, but there are certain important differences. It is made up of a linear polysaccharide chain, but with the sequence encoded by a modified set of the four nucleotide bases: guanine (G), adenine (A), cytosine (C) and uracil (U). Uracil (U) is used instead of thymine (T) of DNA, but the bases pair in a similar manner: G to C and A to U. RNA contains ribose, as opposed to deoxyribose of DNA, in its saccharide chain, which makes RNA less stable than DNA

5 It is a common misconception that this is due to the number of hydrogen bonds between the base pairs. It is actually due to stacking of π-orbitals around the aromatic rings of the bases [222].

6 Assuming the human genome has about 3×10^9 base pairs, these can each be represented by two bits (00, 10, 01, 11) and there are 2^{33} bits per gigabyte. This means the complete human genome represents about $(3 \times 10^9 \times 2)/2^{33}\,\mathrm{GB} = 0.70\,\mathrm{GB}$ of data.

and therefore also less energetically costly to produce. As it is less stable, RNA can be damaged easily. For this reason RNA is regularly degraded and broken down into individual nucleotides, typically having a half-life of a few minutes[87]. The nucleotides can then be used to make new RNA.

Another important difference is that genomic DNA is always present as a perfectly complementary double helix while, RNA on the other hand is found in all sorts of different conformations. Strands of RNA can be a large variety of lengths, from just a few bases to several thousand. RNAs base-pair in any manner that is energetically favourable, usually interacting with short regions of complementarity. They can base-pair with other strands of RNA or even, as the polysaccharide backbone is flexible, fold onto themselves forming what is known as a hairpin. A section of an RNA strand with a hairpin can be seen in Figure 2.1b. The topologies formed by RNA interacting in this manner are called secondary structure, analogous to the double helix of DNA. In a similar way to DNA, sections with a high guanine and cytosine content base-pair more tightly than sections with a high adenine and uracil content.

Finally, RNA is not just passive like DNA. In various contexts RNA plays an active role in the cell. Functional RNAs are crucial in some of the large biomolecular machinery of the cell through the interaction of several different functional RNA strands and proteins. An example of this is the ribosome, described in detail in a later section.

2.1.3 *Protein*

Proteins are the work-horses of the cell. They execute almost all processes of the cell and represent most of the dry mass of a cell. All proteins interact with other chemical species in the cell with some purpose: some are links in signalling pathways, some process metabolites in the cell, some transport chemical species around the cell or even untangle DNA. Like DNA and RNA, proteins are linear, unbranched chains of linked subunits defined by the subunits' specific order. The subunits of which proteins consist, are the 20 naturally occurring amino acids. They form a polypeptide chain with a covalent bond linking nitrogen and carbon to join the subunits together. These are called peptide bonds. As the order of amino acids in a protein is important, proteins have directionality in a similar way to DNA and RNA. They are said to run from N-terminus to C-terminus, that is from the end of the protein with exposed nitrogen to the end with exposed carbon.

Though the specific order of the amino acids in a protein define it, it is the shape and characteristics of the protein's surface that imbue it with functional properties. Each protein folds into a distinct and intricate shape allowing it to perform its task. A protein's conformation is generally energetically favourable to it, but the conformation may change slightly on interaction with other chemical species, enabling the protein to reduce its free energy further.

Non-covalent interactions are the main drivers of the protein folding process. Examples of these are hydrogen bonds, electrostatic and van der Waals interactions. These interactions exist internally between subunits of a protein, but also with the surrounding water and other chemical species. Protein folding within a cell is a reliable process, though the exact details of it remain a mystery.

When urea, which blocks all non-covalent interactions, is added to a highly pure solution of a specific protein, the protein is unfolded, or 'denatured'. If the urea is then washed away, many of the proteins spontaneously re-fold. The remaining proteins that stay unfolded are not functional. This indicates that the proteins themselves contain enough information to form their three dimensional conformations independently. Despite this, in a cell the folding process is facilitated by proteins belonging to a class of proteins known as chaperones.

Analogous to the nucleic acids, the specific sequence of amino acids defines a protein's primary structure. Folded linear strands of protein form a range of common topologies. Two of the most common are α-helices and β-sheets. α-helices are long helical structures that are right handed. β-sheets occur when two or more strands of a protein lie parallel or antiparallel to each other forming a sheet like structure connected by hydrogen bonds. Such complex structures are linked by 'turns' and 'loops'. Turns are small, consisting of two to five amino acids where the protein chain doubles back on itself. Loops are a catch-all for longer more disorganised sections of the protein chain. These elements can broadly be denoted as secondary structure.[7]

Elements of secondary structure come together to form tertiary structure, which is the full three dimensional conformation of a protein. A protein's tertiary structure can be split into what are called 'domains'. These are tightly folded sections or subunits of the protein that often fold and are stable independently of the rest of the protein. As the protein is a single linear strand, each domain is connected to the next by the protein's single strand forming a chain of domains. A number of proteins of the same type, or different types can then come together to form yet larger cellular machinery. This type of interaction is called quaternary structure. Two proteins that bind to form a functional unit are called a dimer, three a trimer, four a tetramer, and so on...

7 The specific definition of secondary structure requires patterned hydrogen bonds in the structure.

Accurate *in silico* prediction of protein structure, also known as The Protein-Folding Problem, is not yet feasible, but a way to gain some intuition of protein folding is to study the amino acids and the protein sequence. The amino acids can be split into two types: hydrophilic and hydrophobic. The interior of the cell is aqueous and for this reason, hydrophobic acids shield themselves, as best they can, from the surrounding water.

One of the more accurate methods of protein structure prediction is through studying the homology of proteins with similar sequences. Imagine the structure of a certain protein is known and this protein can be split into two domains. A second protein is then found. It has a sequence very similar to the section representing the first domain of the protein with known structure. It is then a fair guess that it will fold in a similar to the first domain in the known protein. In such a manner an educated guess at a protein's conformation can be achieved and the domains used will hint at its possible function. Ultimately if the structure of a protein is to be known, experimental techniques such at X-ray Crystallography of Nuclear Magnetic Resonance must be employed.

2.1.4 *The Central Dogma of Molecular Biology*

The Central Dogma of Molecular Biology, presented in Figure 2.2, is what connects the nucleic acids of DNA and RNA, and proteins. It was first stated in 1956 by Francis Crick and was formally described in 1970 and published in Nature [50]. The DNA itself is passive and is in essence a device for storing information. It does not take part in the functions of the cell directly. The information in DNA is stored in the form of genes.

The definition of a gene is a region of DNA that encodes either a protein or a functional RNA. Gene expression is the process of transferring the passive information stored in the DNA to the active form of the gene. This can either be in the form of functional RNA or, as in the vast majority of cases, protein.

As DNA is double stranded, both strands can contain genes. An example of the possible orientation of genes on genomic DNA is presented in Figure 2.3. By convention the top strand runs from 5′ to 3′ and is referred to at the forward strand. Genes on this strand run from left to right. Genes on the bottom, or reverse, strand run in the opposite direction, from right to left.

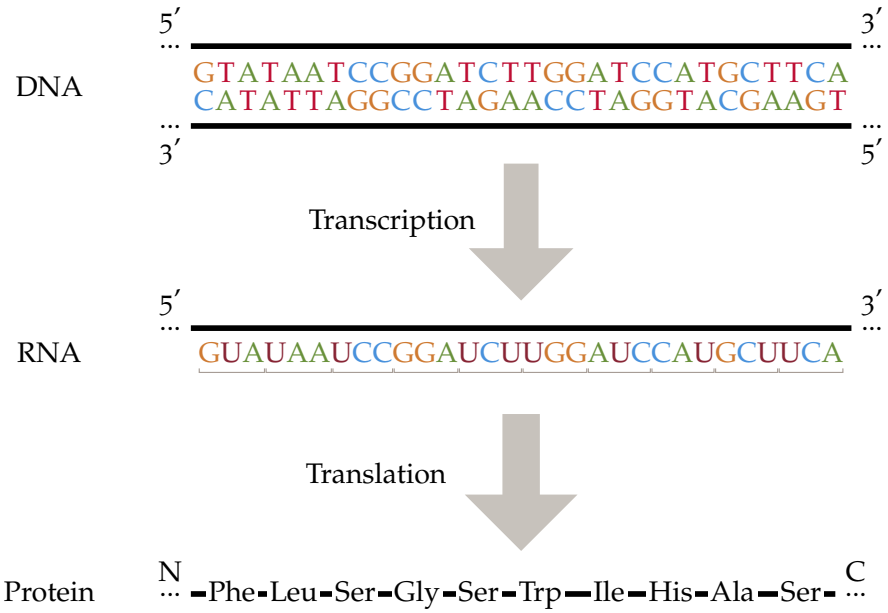


Figure 2.2 – The Central Dogma of Molecular Biology is represented above. It shows the transfer of information from DNA through RNA to protein. The top strand of DNA is transcribed to RNA. Note that the RNA has the same sequence as the top strand of DNA, except that thymine (T) has been switched with uracil (U). The RNA strand can then be split into codons, indicated by the brackets. The RNA is then translated to protein, converting the codons to the relevant amino acids. As the DNA and RNA are read and synthesised from 5′ to 3′, protein is polymerised from N-terminus to C-terminus. The degeneracy in the genetic code can be seen as three different codons in the RNA code for serine, Ser (see Table 2.1).

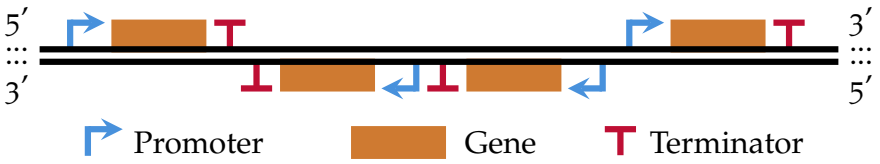


Figure 2.3 – Possible gene orientation on genomic DNA. An organism's DNA consists of two antiparallel strands that both can carry genetic information. The blue arrows indicate the position of the promoters from which transcription is initiated. The genes are indicated by the orange blocks and the red T-shapes indicate the positions of terminators that end transcription.

2.1.4.1 *Transcription*

The first step in the process of gene expression is transcription. Transcription transfers the information stored in DNA to RNA. The reason it is called transcription is that the 'languages' of DNA and RNA are very similar, so this process can be thought of as transcribing certain paragraphs or pages of a book to a clean page. At a more abstract level, the languages of DNA and RNA are related through a one-to-one mapping; each DNA nucleotide is directly translated to a single RNA nucleotide.

Transcription of a gene is performed by a large enzyme called RNA Polymerase (RNAP). One of these must be recruited to the correct position on the DNA to transcribe a specific gene. The regions that recruit RNAP and from which transcription begins are called promoters. A promoter is positioned upstream of a gene such that RNAP can move downstream from the promoter to transcribe.

To transcribe a gene on the forward strand, such as one of those seen in Figure 2.3, RNAP is recruited by the relevant promoter. Once recruited, the DNA is read by RNAP moving downstream, from left to right, from the promoter. The transcribing RNAP synthesises a strand of RNA while moving from 5′ to 3′ on the forward strand, copying the section of DNA representing the gene. This produces a strand of RNA with an identical sequence to the gene on the forward strand of DNA, except that thymine (T) is replaced by uracil (U). This can clearly be seen in the transcription section Figure 2.2. Transcribing the forward strand, RNAP uses the reverse strand as a template, taking advantage of complementarity to guide the correct nucleotides into place before fusing them to the RNA transcript.

Transcription continues until the RNAP hits a sequence of DNA that terminates transcription by forcing the RNAP off the DNA. These regions are referred to as terminators. The transcription of a gene is now complete; the DNA and the strand of RNA is released from the machinery. Depending on the specific situation, some RNA transcripts need processing prior to being fully functional RNAs or being ready for the next step in expression. If the gene encodes a protein, the ready RNA strand is called an Messenger Ribonucleic Acid (mRNA).

2.1.4.2 *Translation*

The transferal of information from mRNA to protein is called translation. This label was chosen as the 'languages' of the nucleic acids and protein differ hugely. The analogy of translating from one language to another has therefore been adopted.

Table 2.1 – This table shows the relationship between the twenty naturally occurring amino acids that are combined to produce proteins and the 64 possible codons. There is substantial redundancy in the number of codons; up to six codons can code for the same amino acid. It is worth noting that in the cases of such redundancy, it is common for there to be variation in the last nucleotide of the associated codons. Alanine, glycine, proline, threonine and valine are prime examples of this, as their first two nucleotides remain unchanged and they take all four nucleotides in the third position. AUG is the Start Codon that initiates translation. The amino acid corresponding to the Start Codon in eukaryotes is methionine. Prokaryotes use a modified version: *N*-formylmethionine, often denoted fMet. The three Stop Codons terminate translation.

AMINOACID	LABEL	NO.	ASSOCIATED CODONS
Alanine	Ala	4	GCA, GCC, GCG, GCU
Cysteine	Cys	2	UGC, UGU
Aspartic acid	Asp	2	GAC, GAU
Glutamic acid	Glu	2	GAA, GAG
Phenylalanine	Phe	2	UUC, UUU
Glycine	Gly	4	GGA, GGC, GGG, GGU
Histidine	His	2	CAC, CAU
Isoleucine	Ile	3	AUA, AUC, AUU
Lysine	Lys	2	AAA, AAG
Leucine	Leu	6	UUA, UUG, CUA, CUC, CUG, CUU
Methionine	Met	1	AUG (Start Codon)
Asparagine	Asn	2	AAC, AAU
Proline	Pro	4	CCA, CCC, CCG, CCU
Glutamine	Gln	2	CAA, CAG
Arginine	Arg	6	AGA, AGG, CGA, CGC, CGG, CGU
Serine	Ser	6	AGC, AGU, UCA, UCC, UCG, UCU
Threonine	Thr	4	ACA, ACC, ACG, ACU
Valine	Val	4	GUA, GUC, GUG, GUU
Tryptophan	Trp	1	UGG
Tyrosine	Tyr	2	UAC, UAU
-	-	3	UAA, UAG, UGA (Stop Codons)
TOTAL:		64	

The protein encoded on the mRNA is read off in sequential blocks of three nucleotides, called codons. There are four nucleotides and therefore there are $4^3 = 64$ possible distinct codons. These are translated into one of the twenty naturally occurring amino acids that make up protein or the Stop Codon. Table 2.1 shows the relationship between the codons made up of the RNA nucleotides (A, G, C, U) and the amino acids. It is clear to see the substantial degeneracy between codons and amino acids. As a result of this, a protein with a specific sequence of amino acids can be encoded in a large number of different ways. Applying the abstraction employed to describe transcription, translation uses a many-to-one mapping.

A group of functional RNAs called Transfer RNA (tRNA) recognises the codons on the mRNA being translated. These are short RNAs of around 80 nucleotides in length [188]. They form very distinct secondary structure in the shape of a clover leaf. On the middle leaf of the structure there is an exposed section with an anticodon. The anticodon facilitates the recognition of codons through complimentary base-pairing. Attached to the stem of the clover leaf is the amino acid corresponding to the specific anticodon. For example, the anticodon for the first of the alanine codons (see Table 2.1) is UGC, which is complimentary and so read in the opposite direction. A tRNA with this anticodon carries the amino acid alanine ready to be fused to the C-terminus of a protein in production. Interestingly enough, the redundancy found in amongst the codons is mirrored in the number of tRNAs, but it is reduced. There is not a different tRNA for each of the 61 amino acid codons. Some amino acids have several tRNAs, but some tRNA can recognise several codons. To do this, tRNAs take advantage of what is called 'wobble'. Wobble takes advantage of the fact that many amino acids, such as alanine, have codons only differing in the third position. In these cases, tRNAs are often designed such that they only require tight base-pairing at the first two positions to be able to recognise the codon.

Using tRNAs the cell can relate the different codons to their amino acids. The fusion of these amino acids to form the protein is performed by a complex molecular machine called the ribosome, completing the process of translation. The ribosome, in large part, consists of pieces of functional RNA called Ribosomal RNA (rRNA). The ribosome is huge and has two subunits called the small subunit and the large subunit. When not translating, these are separate. Translation is initiated when these subunits combine around an mRNA. Within the ribosome there are pockets where the tRNAs fit and can access the mRNA being translated.

Translation initiates at the AUG Start Codon. From there the ribosome moves downstream along the mRNA fusing the amino acids brought in by tRNAs by forming covalent peptide bonds as it goes. This, in effect, prints out the protein. The ribosome adds amino acids to the C-terminus of the protein, pushing the N-terminus end into the cell as it progresses. Once the tRNAs have

had their amino acids fused to the protein, they are ejected back into the interior of the cell, ready to be loaded up with another amino acid. This process continues until the ribosome hits one of the three Stop Codons (UAA, UGAor UAG). These force the ribosome to disassemble, releasing the mRNA and the newly translated protein.

Protein folding occurs as the protein is pumped out of the ribosome or shortly afterwards facilitated by chaperone proteins. Despite the high accuracy of translation, one mistake in 10000 amino acids, the cell has a number of mechanisms to ensure that proteins are translated correctly. These classically degrade faulty proteins. This completes the description of the Central Dogma of Molecular Biology and the process of gene expression.

In both the case of transcription and translation, as the relevant cellular machinery moves down the strand, the initiation sites are freed up, allowing access by more machinery. Many RNAPs can be transcribing the same gene at the same time and many ribosomes can be translating the same mRNAs at the same time. Combining this with rates of transcription of between 20 and 50 nucleotides per second and translation of two to 20 amino acids per second [6, Chap. 6], these effects can lead to rapid expression rates.

2.1.5 *Harnessing Biology*

It is incredible that all known life shares all of the above properties, but that is where the similarities end. All living organisms can be split into three domains: bacteria, archaea and eukaryotes. Bacteria and archaea are both classed as prokaryotes. These generally exist as single cell organisms or in loose organisations such as colonies or cultures. Eukaryotes are much larger and more complex. The volume of a eukaryote is usually around one thousand times that of a bacteria. Eukaryotes include all multicellular organisms from plants to humans.

When modelling and designing organisms, the mechanisms that are being employed and manipulated, whether specific to an organism or a set of organisms such as bacteria, need to be well understood to ensure that the assumptions and context of a certain model are clear. For this reason, from here on in, the focus is on bacteria and bacterial mechanisms. That is not to say that some of these processes do not also take place or are possible in archaea and eukaryotes. It is also hoped that with a little adaptation, the modelling and design techniques presented in this work can be generalised to higher organisms.

That being said, the focus of this document is generally on bacterial mechanisms, and more specifically on *Escherichia coli*, written *E. coli* in short, the chosen bacterial host for this work.

Similarly to the implementation of any engineered system, from a computer to commercial aircraft, designed Gene Regulatory Networks (GRNs) are created through the careful combination of many different biological parts. To a large extent, Nature has provided a huge library of such parts that can be used, but synthetic biologists have been slowly adding to this library[31, 45]. The chosen parts in an engineered system need to interact and function in unison in a predictable manner, and therefore they need to be chosen carefully. Throughout the next sections, a number of different parts and mechanisms are introduced.

2.1.6 *Bacterial Structure and Function*

The aim of this section is to give an impression of the type of environment found within a bacterial cell. Understanding this is important as this contextualises the mathematical models that are constructed later in this thesis.

The interior of a bacterial cell is a busy aqueous environment. As opposed to eukaryotes, that are split into many compartments and organelles, bacterial structure is relatively simple.

The innermost space in a bacteria is referred to as the cytoplasm. The cytoplasm is where the large majority of cellular processes occur including metabolic, biosynthetic and genetic processes. Bacterial DNA and all the transcriptional and translational machinery exist in the cytoplasm. This is very different to eukaryotes, where the genetic material is kept within the nucleus and where access to the DNA is very carefully regulated. Any species within the bacterial cytoplasm has direct access to the DNA.

Around the cytoplasm is a lipid cell membrane often referred to as the cytoplasmic membrane or just the cell membrane. Around the cell membrane is a rigid cell wall that gives the cell its shape. In the case of Gram positive[8] bacteria, this cell wall is thick and robust. In the case of Gram negative bacteria, such as *E. coli*, this membrane is relatively thin and somewhat flexible. In addition to the cell wall, Gram negative bacteria have an additional lipid bilayer referred to as the outer membrane.

The space between the cytoplasmic membrane and outer membrane containing the cell wall is referred to as the periplasm. The periplasm contains proteins that degrade larger nutrients from outside the cell. The broken down nutrients can then be brought into the cytoplasm and metabolised. In the periplasm are also sensor proteins that detect membrane damage, chaperone

8 The Gram staining is a historical method used to differentiate between types of bacteria. It basically looks at the ability of a certain type of bacteria to retain a dye. Gram positive species retain the dye well, Gram negative do not. It was discovered that this was due to important structural and ancestral differences.

proteins that help proteins fold and move between compartments, and proteases that degrade proteins, among others.

The cytoplasmic membrane contains membrane proteins used to pass information into the cytoplasm by signalling and conformational change. There are also proteins that transport a large range of chemical species between the cytoplasm and periplasm. The outer membrane contains channels that allow the passage of certain proteins and small molecules, receptor proteins used for signalling, and also provides anchorage for bacterial structures such as flagella and pili that bacteria use to move, sense and communicate. In Gram positive bacteria, the cell wall performs many of roles that the outer membrane does in the case of Gram negative bacteria.

2.1.7 *Bacterial Transcription*

To perform transcription, bacteria have a single type of RNAP, whereas eukaryotes have three different types that perform different tasks. Transcription initiation is an incredibly important process to all cells as this is the point at which they decide which genes to express. In bacteria this process is initiated by what is called the RNAP holoenzyme. This is the RNAP in complex with a detachable subunit called a sigma (σ) factor. This complex only weakly binds DNA and slides along rapidly scanning for promoters.

The σ factor recognises promoter regions and binds tightly to them pinning the RNAP to the specific region of DNA. The DNA strands are then separated and the process of transcription begins inefficiently as the RNAP begins to transcribe using the 3′ to 5′ strand as template while slowly releasing itself from the σ factor. After transcribing about ten nucleotides, the σ factor completely dissociates and transcription progresses at around 50 nucleotides per second [215], depending on temperature and media.

2.1.7.1 *Promoter Architecture in E. coli*

Individual promoters for naturally occurring genes vary greatly, but the comparison of many promoter regions gives a sequence of nucleotides most commonly found in a promoter. This is called the consensus sequence. An early study of this type found the consensus sequence of a major class of promoter in *E. coli* [92]. The sequence consists of two hexamers, sections of six nucleotides, separated by between 15 and 19 nucleotides. The two hexamers are named -35 and -10 from their approximate position relative to the position of transcription initiation denoted +1. The -35 consensus sequence is TTGACA and the -10 consensus sequence is TATAAT. Distance

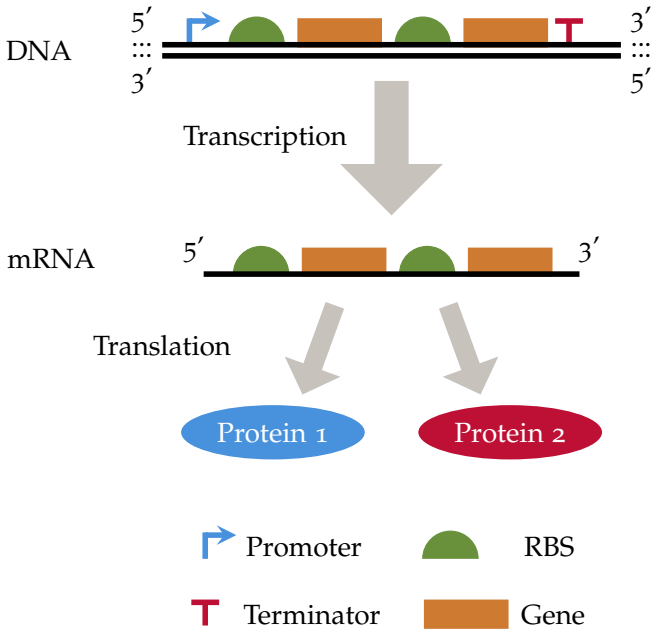


Figure 2.4 – Bacterial gene expression from an operon coding two genes. The diagram is annotated, clearly indicating the promoter, terminator, RBSs and coding regions of the operon and polycistronic mRNA. The mRNA is then translated into two different proteins. As bacterial translation initiation occurs at a region of the mRNA referred to as the RBS just upstream of the gene to be transcribed, bacterial mRNAs can code for several proteins which are then translated independently of each other. This is discussed further in Section 2.1.8.

between these two sequences varies, but there is a strong preference for a distance of 17 nucleotides. This consensus sequence is what is recognised by the major σ factor, also denoted σ^{70}, which initiates transcription of most genes. The other σ factors generally specialise in initiating the transcription of genes with specific purposes. For example σ^{32} targets genes induced by heat shock and σ^{34} initiates genes that deal with mis-folded proteins in the periplasm.

In general, as the promoter sequence diverges from the consensus, it binds the relevant σ factor more weakly, reducing the probability that a transcription event occurs. Many other factors, from the DNA architecture around a promoter to the specific situation of the bacteria, influence the rate of transcription initiation at a promoter. In this way different promoters can be thought of as having different strengths.

The specific choice of promoter in the context of the implementation of GRNs is a crucial one. It is discussed in Section 2.1.14.

2.1.7.2 *Transcription Termination*

Transcription initiation and elongation have been presented in detail, but the mechanism by which transcription terminates is yet to be presented. In bacteria, there are two different types of transcription termination: factor-independent and factor-dependent.

Factor-independent termination is simply recognised in the genetic code. It consists of a section of DNA with an inverted repeat, separated by three or four nucleotides, followed by a number of Ts. The inverted repeat produces a hairpin structure in the RNA, such as that seen in Figure 2.1b. This structure knocks the RNAP off the DNA as it comes out of the machinery, terminating transcription somewhere along the run of Ts. The tail of an RNA strand produced using such a terminator has a hairpin followed by a section of Us that continue till the end of the strand.

Factor-dependent termination is a complicated process that is not well understood. It is best understood in *E. coli* where there are three factors that play a role. Only one of these factors, called the rho (ρ) factor, has a postulated functional mechanism. In the simplest terms, it attaches to the RNA strand at a recognition site, as the strand is being pushed out of the RNAP. The ρ factor forms a homohexameric[9] ring around the RNA strand and then moves downstream towards the RNAP. When this ring encounters the RNAP, it knocks the RNAP off the DNA strand forcing the termination of transcription.

Large libraries of transcriptional terminators for *E. coli* exist and are well documented [35, 45]. These are factor-independent terminators and consist of a mixture of natural and synthetic terminators. The library presented in [45] should satisfy the initial needs of a synthetic biologist using *E. coli* as a host.

2.1.8 *The Ribosome Binding Site*

Bacterial mRNAs are different to eukaryotic mRNAs, as they are not capped at the 5′ end. Translation initiation occurs by a different mechanism. Usually, bacterial mRNAs contain regions that are not translated close the the 5′ and 3′ ends of the mRNA. Such regions are called Un-Translated Regions (UTRs).

In the 5′ UTR of an mRNA, the section just upstream of the Start Codon is referred to as the Ribosome Binding Site (RBS). The RBS has no clear boundaries and can be up to 50 nucleotides long. An RBS contains a sequence of six to eight nucleotides called the Shine-Dalgarno sequence which is broadly conserved across bacteria and archaea with the sequence AGGAGG. This sequence varies from species to species; in *E. coli* it is AGGAGGU. The Shine-Dalgarno sequence is usually five to ten nucleotides upstream of the Start Codon. In addition to this, secondary structure of the RBS plays a large role in the initiation of translation, making prediction of initiation rates difficult. As in the case of promoters, RBSs can be thought of as strong or weak.

9 A quaternary structure consisting of six identical subunits of ρ factor.

As translation initiation in bacteria occurs at an RBS, and not at the 5′ cap of an mRNA as in eukaryotes, initiation occurs anywhere there is an RBS on an RNA strand. In fact, many genes in bacteria share promoter and terminator sequences, resulting in mRNAs with two or more genes on them. Such a structure is called an operon. Expression from an operon coding for two genes is presented in Figure 2.4. An mRNA produced through the transcription of an operon has several genes separated by UTRs. Each of these genes has its own RBSs regulating translation. Such mRNAs are polycistronic.

The design of RBSs to a strength over at five fold range can be performed using Salis *et al.*'s Ribosome Binding Site Calculator [184], which is detailed further in a book chapter [183]. Though this offers a measure of relative strength of an RBS, this is a rough score, accurate to within a factor 2.3, and so the actual performance of the RBS must be tested *in vivo*.

2.1.9 *Bacterial Translation*

Another factor that plays a role in translation initiation is the use of Start Codons. The classic AUG Start Codon is not the only one used. Another relatively common Start Codon is GUG. The codons UUG and AUA are also used but are much rarer and in a single case in *E. coli* AUU is used.

The machinery crucial to translation, the small and large subunits of the ribosome, is found separately in the cytoplasm when not translating. The small subunit associates with three translation initiation factors, labelled IF1 to IF3, and the tRNA carrying fMet that recognises the Start Codon. There is a section of this complex that recognises and associates with the Shine-Dalgarno in the RBS and the tRNA carrying fMet associates with the specific Start Codon. As the Start Codon used diverges from the classic AUG, it becomes harder and harder for the fMet tRNA to recognise it, reducing the probability of translation initiation. Choice of Start Codon therefore affects translation rates.

In prokaryotes, the second a gene encoding a protein is transcribed, the mRNA produced is ready to be translated. As the translational machinery is in the cytoplasm with the DNA, translation can begin the second the RNA strand exits the RNAP. In this way transcription and translation of a gene can occur at the same time. In eukaryotes, this is not the case. The mRNA has to be processed and exit the nucleus prior to translation.

Once the initiation complex is recruited to an RBS and the Start Codon is found, the large ribosomal subunit rapidly joins and translation commences, only terminating when the ribosome

comes across a Stop Codon. Translation progresses at around 20 amino acids per second [154, p.1559]. The Stop Codons, in effect, jam the ribosome as they do not code for an amino acid and therefore have no associated tRNAs. Release factors recognise this and free the polypeptide from the translational machinery.

The freed protein is modified at the N-terminus after translation through removal of the fMet, the first amino acid in the chain. Folding of the protein is then facilitated by numerous chaperone proteins preparing the protein for its functional role in the cell.

Interestingly enough, there are examples of bacterial mRNAs that do not have 5' UTRs [146]. This implies that translation initiation can occur without an RBS, relying solely on the Start Codon. In these cases, it is believed that initiation occurs through recognition of the Start Codon by a free fMet tRNA, that then recruits the small and large subunits of the ribosome to commence translation.

Another fact worth noting is that many of the most common bacterial antibiotics function by translation inhibition by interfering with the ribosome [5, p.384]. Many of these are expressed in fungi that compete with bacteria in many environments. Two among these that were used in the course of this project are tetracycline and chloramphenicol.

2.1.10 *Degradation*

The degradation of chemical species in the cell is an important process that allows the cell to recycle and reuse components. In many ways, it is not nearly as well understood or studied as biosynthetic processes such as gene expression and metabolic pathways. As degradation plays an important role in the turnover of species, the mechanisms by which it occurs are introduced here.

2.1.10.1 *Degradation of RNA*

The degradation of RNA is a continual and rapidly occurring process. Rapid turnover is important as it ensures that mRNA is not corrupted prior to translation and that the cell can allocate its ribonucleotides to the expression of genes that are needed. The basics of RNA degradation are covered in [57, 58, 99].

Generally, there is a large redundancy in the degradation machinery and it overlaps extensively with the RNA processing machinery that prepares functional RNAs such as rRNA. RNA can

either be degraded from either end by exonuclease proteins or from sites in the middle by endonuclease proteins. Endonucleases are associated with the RNA degradosome [38].

The degradation of RNA is in large part regulated by access to the targets. In this way, highly translated mRNAs are bound by ribosomes and other cellular machinery that can block access slowing down degradation. Degradation of a mono or polycistronic mRNAs usually occurs through cleavage by the RNA degradosome within its translation initiation regions.

2.1.10.2 *Degradation of Protein*

Protein degradation is another powerful tool harnessed by the bacterial cell. It plays a crucial role, not just in recycling components, but also protecting the cell from a pathogenesis [83]. Bacteria can selectively degrade proteins which are found to be abnormal or damaged due to either having mis-folded or containing toxic mutations [82, 172].

A mechanism by which this occurs is through a ribosome rescue mechanism that uses a special type of functional RNA called transfer-messenger RNA (or just tmRNA) [108, 111]. This system rescues stalled ribosomes, by pushing out the faulty mRNA and replacing it with a sequence of about thirty nucleotides that are translated and attached to the end of the abnormal protein. This tags the proteins for immediate degradation by endogenous degradation machinery. The nucleotides are encoded in the *ssra* gene, and the protein is said to have been *ssra* tagged.

This mechanism has then been highjacked for use in synthetic biology to control the degradation rates of synthetically expressed proteins [221]. This can be done by fusing the sequence to the end of an encoded protein. This immensely increases the turnover rate of the tagged protein. A library of mutants of natural *ssra* tags have been made using an analogous but orthogonal mechanism to that found in *E. coli* which produces graded or tuneable degradation of the tagged proteins [37].

2.1.11 *The Bacterial Genome and Fission*

In most bacteria and in *E. coli* in particular, genetic information is stored on a single ring of DNA called the chromosome. Most bacteria have circular chromosomes, and in a few cases, they have more than one. The *E. coli* chromosome contains about 4.6×10^6 base pairs and 4288 genes split into 2584 operons [28]. For comparison, the smallest know natural genome is that of *Mycoplasma genitalium* which has 5.8×10^5 base pairs and just 535 genes [72]. It is estimated that the human genome contains a little over 20000 genes.

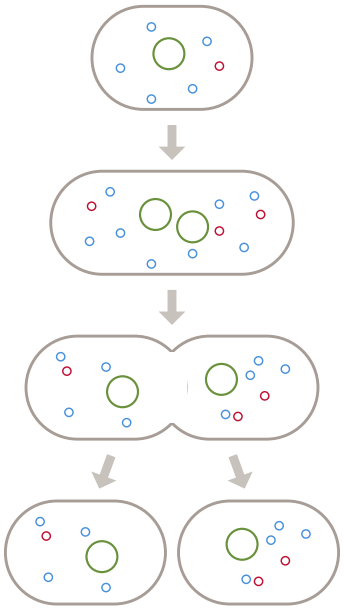


Figure 2.5 – Bacterial fission can be separated into four stages. Initially the whole cell begins with one whole copy of the chromosome and a number of plasmids. In this case the mother cell has two types: five blue plasmids and one red plasmid. The cell then copies the chromosome and the plasmids and grows by increasing the size of the cell wall and the overall volume of the cell. Note that the plasmids often do not exactly double in number. The cell then separates the two chromosomes and the plasmids into the two halves and begins division. The final result is two daughter cells with identical chromosomes, but often with different numbers of plasmids to the mother cell.

Bacteria proliferate by replication: the production of two copies of the original. This is also called binary fission and is a method of asexual reproduction. To be able to divide, the mother cell needs to equip each daughter cell with everything needed to survive, crucial to this is, of course, the genetic information stored in the chromosome.

The replication of a chromosome begins at a region called the Origin of Replication. This usually has a region that is very A-T heavy as this facilitates separation of the two DNA strands allowing access by DNA Polymerase (DNAP), the cell's DNA replication machinery. As the chromosome contains two completely complimentary DNA strands, DNAPs bind to each of these two strands using them both as templates. The DNAPs work their way around the chromosomal ring, the end result being two complete chromosomes, both consisting of one DNA ring from the old chromosome and one newly synthesised ring.

As this chromosomal replication occurs the bacteria grows its membranes and cell walls and divides the cellular machinery in two, pulling each half towards either end of the cell. Soon after the chromosome has been copied, each copy is then separated and the cell divides producing two complete daughter cells. A diagram of this process is presented in Figure 2.5.

2.1.12 *Plasmids*

Apart from the chromosomes, there is another carrier of genetic information that can be found in nearly all bacteria. These are called plasmids. They are usually found in the same form as bacterial chromosomes: double stranded rings of DNA. They can be anywhere from a few thousand base pairs to several hundred thousand base pairs long, rivalling the size of the chromosome itself.

In nature, plasmids carry genes of use to the host cell, but they usually do not carry genes vital to survival. The number of genes carried can vary from a few genes to hundreds. The genes carried on plasmids usually confer the ability to survive[10] in a particular type of environment or deal with certain toxins or nutrients, the classic example of this being antibiotic resistance genes. A cell usually contains more than one copy of the same plasmid and can carry several different plasmids.

To be useful, replication of the plasmids is required. Plasmids and the bacterial chromosomes are DNA elements called replicons that have the ability to replicate autonomously. Plasmids contain Origins of Replication, much like the bacterial chromosome, usually just referred to as the *ori* region. The *ori* region contains the replication initiation site, but also genes encoding a single or several factors that allow DNAP to initiate replication of the plasmid. In this way, the plasmids ensure their continued existence, playing a role in bacterial fission presented in Figure 2.5.

Using plasmids, bacteria can increase their chances of survival through a heterogeneous population. Having a plasmid adds burden to the host cell, slowing growth and giving a competitive disadvantage. However, if the environment then suddenly changes allowing the bacteria with a specific plasmid to benefit from the genes on the plasmid, they immediately have a competitive advantage and the future of the whole population is ensured. Not only that, but the rest of the population can benefit through horizontal gene transfer. This is when single celled organisms transfer genetic information to other single celled organisms in a population.

2.1.12.1 *Conjugation*

One method of horizontal gene transfer is conjugation, a process by which plasmids can be directly transferred. The donor emits a signal attracting bacteria that do not have the plasmid. The donor then makes a physical connection with the recipient bacteria using a structure extending

10 Being a nerd, I like the analogy with magical items in Dungeons & Dragons or other role playing games. You pick an item up and suddenly you can resist fire!

from the outer membrane called a pilus. The cells are brought together and a pore is formed between them. One of the two DNA strands of the plasmid is then transferred through this pore. The pore closes, the bacteria separate and both bacteria synthesise the missing DNA strands, resulting in two complete plasmids, one in each cell. Conjugation can occur, not only between individuals of the same species, but also separate and even quite distantly related species.

2.1.12.2 *Roles of the* Ori *Region*

The *ori* region defines a number of attributes of a plasmid. The first of these is the host range of the plasmid. Some *ori* regions can only be replicated in a very small number of hosts. One such example are ColE1 type *ori* regions. These can only replicate in *E. coli* and close relatives such as *Salmonella*. The RK2 *ori* region from *Klebsiella aerogenes* on the other hand can replicate in a large number of different hosts. In general for natural plasmids, the more flexible the *ori* region, the larger the plasmid.

A second important attribute regulated by the *ori* region is the plasmid copy number. This is formally defined to be the average number of plasmids in a new cell immediately post a fission event. Plasmid copy numbers vary from one to several hundred. Methods of replication control used by the *ori* regions vary greatly depending on number. High copy number plasmids generally regulate their copy numbers through negative feedback, whereas low copy number plasmids regulate replication initiation very carefully to ensure that this only occurs once or twice during a cell cycle. There are a number of different copy number control mechanisms, some of which are touched on in [194, Chap. 4]. The key to remember is that the copy number is not exact. In reality the copy number of a plasmid is the average around which the actual number of a certain plasmid in a cell varies. An example of how this works is presented in Figure 2.5.

When division occurs, the two complete genomes in the bacteria are physically separated, ensuring that one complete genome is found in each daughter cell. The chromosomes are paired and then separated. Plasmids perform a similar dance called partitioning. Similarly to chromosomes, plasmids are paired in the centre of the cell and then are separated by some means prior to division. The precise method by which this process is governed is defined by the plasmid's *ori* region. Different *ori* regions can use the same method. Methods of partitioning can be read about in more detail in [194, Chap. 4], but are not discussed in more detail in this document.

2.1.12.3 *Incompatibility*

As *ori* regions define how plasmids are replicated in the cell and propagated through the generations, they are crucial to a plasmid's survival. Many different *ori* regions share similar mech-

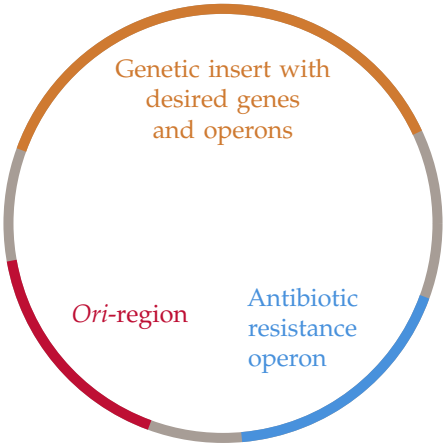


Figure 2.6 – The example structure of a plasmid used for molecular biological study. The double stranded DNA is of the plasmid is represented by the single grey line. The plasmid will have three main parts: The *ori*-region, which determines plasmid copy number, replication mechanism and how the plasmids are distributed in fission events. The antibiotic resistance operon is used to select for cells with a certain plasmid. Finally, the genetic insert that contains the genes and operons of to be studied.

anisms for copy number regulation and partitioning and this causes incompatibility. Sharing various mechanisms across different plasmids produces a number of problems for the survival and propagation of individual plasmids, hence the term incompatibility. When a plasmid is removed from a line of cells, the cells are said to have been cured of the plasmid.

If two different plasmids share the same copy number regulation mechanism, the machinery responsible for regulation sees the two plasmids as the same and regulates them both. The machinery in effect counts the two different plasmids as one plasmid. Due to the stochasticity of replication of plasmids, heterogeneity of the population is unavoidable. The culture will end up with (at least) three subsets of bacteria; two subsets each cured of one of the two plasmids and a third still carrying both.

The a similar process occurs, if two plasmids in the same bacteria share the partitioning mechanism. Different plasmids with the same partitioning mechanism will get paired at the centre of the cell just prior to cellular fission. The paired, but different, plasmids are then physically separated, distributing the two plasmids between the daughter cells unevenly. As this continues during growth of a culture, curing of one of the two plasmids in a cell line becomes inevitable, resulting in a heterogeneous culture.

Some of the cured cells may receive plasmid again through processes such as conjugation, but this is not thorough enough to ensure a homogenous culture and cannot be relied on.

Table 2.2 – Details of the plasmid backbones used in this project. AMP is ampicillin, CARB is carbenicillin and CML is chloramphenicol. The copy numbers can be found in [194, p.230].

LABEL	*ori-region*	ANTIBIOTIC RES.	COPY NUMBER
pJ404	pBR322	AMP/CARB	15-20
pSUtat	p15A	CML	10-12

2.1.12.4 *Plasmids in Molecular Biology*

Plasmids are an incredibly versatile and powerful tool for molecular biologists. They are small, easy to purify, easy to transform into cells and easily edited. Plasmids are therefore the standard tool used in many experiments as a means to transfer genetic information into a cell. Modern techniques such as Gibson Cloning [80] have greatly increased the speed and reliability of such processes, making them now routine tasks in the laboratory.

A plasmid classically has three regions chosen by the designer: an *ori*-region, an antibiotic resistance operon, and the genetic insert. An example plasmid structure can be seen in Figure 2.6. The *ori*-region is selected based on copy number and compatibility requirements as discussed above.

Antibiotic resistance is used in all synthetic plasmids to ensure their presence in the cells. The cells are selected by constantly being fed the antibiotic. This ensures that any cells that are accidentally cured of the plasmid are killed.

The genetic insert on the plasmid contains the genes and operons crucial to the study being performed. This could consist of a whole GRN or a select few of the genes in the complete network. In this work, to facilitate composition of different parts to construct different circuits, operons were often split across two plasmids, the details of which are in Table 2.2. Two plasmids can be used together simply by transforming both plasmids into the chosen host simultaneously and selecting cells containing both plasmids with a combination of both corresponding antibiotics.

2.1.13 *Gene Expression Control*

Cells perform a huge number of different tasks and actions. Depending what nutrients are available they metabolise differently, requiring different proteins and metabolic pathways. During growth they follow a cell cycle, in which there are a huge number of different mechanisms and steps that are taken and triggered by various specialised proteins. In hard times, they employ specific machinery to reuse and recycle waste in the cell.

All the information needed to do all of this is encoded in genes, which the cell needs to express to be able to use. Instead of all genes being expressed at once in equal measure, cells intricately regulate the expression of all genes. Genes are only expressed when required and the level of expression is also controlled carefully so that the cell has enough of the protein in question to execute the desired function. In this way, a cell can be thought of as an intricate computer constantly analysing and reacting on its environment and its own situation.

As an example, at a high level, bacteria, such as *E. coli*, regulate expression through the use of the aforementioned σ factors, with expression of each of the different σ factors activating the transcription of a group of genes. For example, one of these factors, σ^{32}, is expressed in response to heat shock and triggers the expression of proteins to help *E. coli* survive.

Another means of high level regulation used by bacteria is, confusingly called, catabolite repression [194, p.548-558]. This system ensures that carbon sources that are more easily metabolised are consumed first, by hierarchically repressing the expression of pathways metabolising other carbon sources. The mechanism is effected by the Catabolite Activator Protein (CAP), which binds and activates transcription of its target operons when the metabolic conditions are right. For example, *E. coli* has a preference for glucose. Grown on agar plates of glucose and lactose, the glucose is metabolised first, leaving the lactose for later consumption. This is due to catabolite repression.

Mechanisms of gene expression control are crucial to the construction of GRNs. The majority of parts used to implement designed GRNs are borrowed and adapted from natural systems of gene expression regulation [15, 31, 33, 43, 192]. Bradley *et al.* published a broad overview of the available tools and methods that can be employed in their 2016 article [31].

Regulation can occur at every stage in the expression process. The next sections describe a selection of these mechanisms relevant to this project.

2.1.14 *Transcriptional Control*

The most common method of expression regulation is regulation of the transcription rate of a gene or operon. This is efficient for the cell as it controls expression at the earliest possible point in the process.

In the natural world, genes and operons can roughly be divided into two groups: constitutive and facultative. A constitutive gene or operon is one that is expressed at a relatively constant rate, no matter the cell's state or environment. These genes are governed by constitutive promoters.

A facultative gene or operon, on the other hand, varies its level of expression depending on the needs of the host cell. These are governed by promoters that can vary transcription initiation rates through processing information present in the cell.

The information in the cell is transferred to the DNA by a special class of protein called a Transcription Factor (TF). These proteins are able to bind DNA and often share structural similarity in their DNA binding regions, employing a famous secondary structure unit consisting of two alpha helices connected by a turn. This is referred to as the Helix-Turn-Helix motif. The regions of DNA with which a TF interacts are called operators.

The way TFs transfer information is through interaction with small molecules in the cytoplasm. These small molecules are generally referred to as effectors, but have more specific names depending on the type of TF they interact with. When effectors associate with a TF, they form a complex that changes the conformation of the TF, thereby changing the behaviour of the TF. This is referred to as allosteric regulation.

TFs have the ability to either increase transcription rates, in which case they are called an activator, or decrease transcription rates, in which case they are called a repressor. A TF can regulate the expression of one or several genes, either through an operon structure or several different promoters. In many cases TFs regulate their own expression; these are referred to as autoregulatory TFs. In some particular cases, the same TF can act as an activator in some instances and a repressor in others [149].

As a single TF can regulate several genes, so can expression of a gene be regulated by several TFs. This is common in nature and results from a complicated operator structure in the region of the promoter. Specific examples of this are presented shortly.

2.1.14.1 *Transcriptional Repression*

There are two common mechanisms which regulate the activity of repressing TFs.

In the first case, the TF is functionally active on its own and represses expression of its target. On the binding of an effector, in this case called an inducer, the conformation of the TF changes, inhibiting it from binding DNA and repressing transcription of its target. Therefore, when an inducer is present, transcription occurs. This process is referred to as induction of expression, and hence the name inducer.

This case describes how the famous lactose (*lac*) repressor works. This is the system that controls the expression of machinery needed to process lactose in *E. coli* and was one of the seminal examples of gene expression regulation [178]. Another example is the tetracycline repressor, which is discussed in detail in the next section.

The second case is where binding of the effector activates the TF allowing it to repress. In this case the effector is called a corepressor and the TF is called an aporepressor. The combined and active TF-effector complex is a repressor. When the active complex is formed, it can bind DNA and inhibit transcription initiation.

This form of mechanism is often used for biosynthesis genes, such as the tryptophan (*trp*) repressor. In this case, tryptophan is the corepressor molecule. When it is plentiful, the cell does not need to synthesise any more of it, so expression of the synthesis machinery is shut down by the repressor.

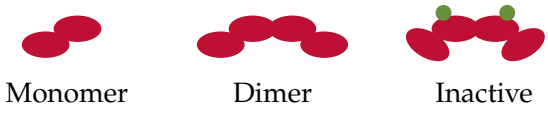
Active repressor proteins function in numerous different ways. Some act as monomer proteins, but in many cases proteins combine to form more complex quaternary structure such as dimers, tetramers or even larger structures. The models of repression for different repressors varies greatly too, though steric interactions are common, blocking access by the RNAP holoenzyme to the promoter. The *lac* repressor is thought to inhibit transcription by bending the DNA in such a way that the RNAP holoenzyme transcription initiation complex cannot bind, as well as sterically interacting with it [194, p.506]. In extreme cases, stabilisation of the interaction between the RNAP holoenzyme and the promoter region can occur to such an extent, that the transcriptional machinery cannot leave the promoter. An example of this is used by a bacteriophage, which uses the same TF and mechanism to both activate and repress [149]. In the case of transcription activation, the stabilisation occurs to a slightly lesser extent.

2.1.14.2 *The Tetracycline Repressor*

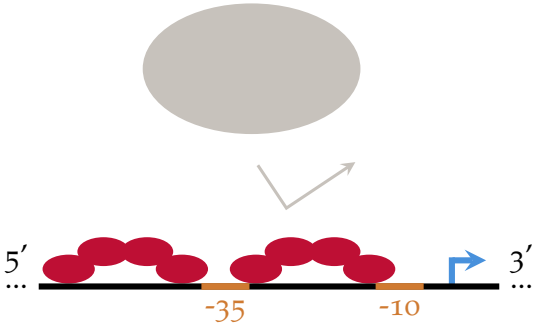
For the implementation of gene circuits designed in this project the Tetracycline Repressor (TetR) was chosen as it is commonly used [16, 19, 23, 62, 155, 176, 179] and is well documented [54, 94, 95, 97, 176, 177]. A diagram of the states of TetR and the mechanism of repression are presented in Figure 2.7.

Tetracycline is an antibiotic that acts by translation inhibition through interfering with the ribosome. Resistance to tetracyclines is common amongst bacteria and is distributed by horizontal gene transfer (see Section 2.1.12.1) [94, 176]. It is the regulatory protein of this resistance mechanism that has been highjacked for implementation of synthetic GRNs. As it stems from antibiotic resistance, TetR mechanism provides highly responsive and tightly regulated system. Despite being a bacterial mechanism, it can be used in some eukaryotes [94, 176].

The repressor, TetR, is functional as a dimer and has a very high dimerisation rate [18, 97]; it is hardly present in the cell as a monomer. Each monomer unit of the TetR consists of two domains: a DNA binding domain and a dimerisation and inducer binding domain. TetR dimers



(a) States of TetR



(b) TetR Repression

Figure 2.7 – The repression mechanism of the tetracycline repressor is a classic example of a repressing TF. (a) The different states of TetR. It is translated as a monomer with two domains: one DNA binding domain and the other responsible for dimerisation and inducer binding. It rapidly dimerises with a high affinity. The dimer form is the active repressor. Binding the inducer, aTc or another tetracycline, inactivates the dimer through a conformational change. (b) Repression mechanism of the tetracycline promoter, P_{tet}, employed in this project (BioBrick label BBa_R0040). Transcriptional repression occurs sterically through blocking access to the transcription initiation region of the promoter by a dimer repressor binding between the σ^{70} hexamers is shown above. The transcriptional machinery, RNAP holoenzyme, cannot bind and initiate the expression process.

bind operator regions with a very high affinity for the inverted repeat GATAG with three spacer base pairs in between [94, 176].

The standard inducer used is aTc and when added to a bacterial culture, aTc readily diffuses into the cytoplasm and has a low antibiotic activity [54]. Each monomer TetR binds a single aTc molecule with a high affinity [97]; the dimer can therefore bind two aTc molecules. This changes the conformation of the DNA binding domain of TetR, greatly reducing its affinity for the operator regions and, in effect, inactivating the protein (see Figure 2.7a) [125].

The wild type tetracycline promoter region is rather complicated [94], and due to that a simpler engineered promoter was used to implement the GRNs designed in this project. The promoter is from the iGEM BioBricks repository and is labelled BBa_R0040[11]. The action of this specific P_{tet} is presented in Figure 2.7b. The promoter is a σ^{70} promoter with one TetR operator nestled between the -35 and -10 hexamers that σ^{70} binds to. A second operator is just upstream of the -35 hexamer. Binding of TetR dimers to these operators physically blocks access to the DNA, stopping transcription initiation from occurring.

11 http://parts.igem.org/Part:BBa_R0040

2.1.14.3 *Transcriptional Activation*

Activation of transcription is commonly used to express enzymes used to metabolise nutrients and synthesise required biomolecules. Most commonly, activating TFs are inactive when not bound to an effector. The effector is again called an inducer, as binding functionalises the TF, enabling it to activate transcription. This inducer is often the molecule to be metabolised or a close chemical relative. On binding of the inducer the TF changes conformation and becomes active. The activator then binds DNA and increases the transcription initiation rate, and therefore the expression.

Whereas many repressors bind the DNA in the middle of the RNAP binding region, most common activators bind operator regions slightly upstream so that they do not physically obstruct the machinery. With no activator bound, binding between the transcription initiation complex, consisting of the RNAP bound to a σ factor, is very weak, making transcription initiation almost impossible. When bound to their operator regions, activators connect with the transcription initiation complex stabilising its interaction with the DNA and ensuring that the initiation complex is bound tightly enough for transcription to occur.

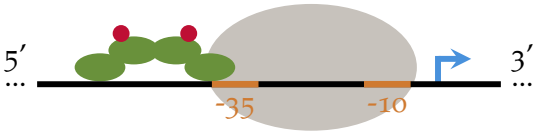
The above described mechanism is used by the arabinose activator, also called AraC. This activator (and others) can perform various roles in gene expression regulation and use a range of mechanisms [194, p. 524]. In addition to activating transcription in the presence of arabinose, when no arabinose is present AraC binds the DNA making it fold in a way that inhibits transcription. Also, if the concentration of AraC is very high, it autorepresses, binding its own promoter and inhibiting its own transcription. MalT, the regulator of the maltose pathway uses yet another mechanism. Activation by MalT occurs when many copies of MalT bind the DNA unravelling it and facilitating access to the promoter by the RNAP holoenzyme [194, p. 527].

2.1.14.4 *The RhaS Activator*

The chosen activator for implementation of designed systems in this project is Rhamnose Activator (RhaS) in conjunction with one of the two natural promoters in the rhamnose operon, referred to as the P_{rhaBAD}.[12] There are a number of reasons for this: P_{rhaBAD} is capable of very high levels of expression without being leaky [89, 203] and also responds linearly over a wide range of inducer concentrations [79, 203]. The rhamnose system is native to *E. coli*. The inducer, L-rhamnose, is also naturally occurring and non-toxic to cells, but is metabolised in *E. coli*. RhaS can also be induced by non-metabolised analogs of L-rhamnose, such as L-mannose [113, 213], which are

12 BioBrick number BBa_K914003: http://parts.igem.org/Part:BBa_K914003

(a) States of RhaS



(b) RhaS Activation

Figure 2.8 – The mechanism of the rhamnose activator. This is a classic example of an activating TF. (a) The various states of RhaS. It is expressed in monomer form, consisting of two domains one DNA binding domain and a second dimerisation and inducer binding domain. It dimerises, but is still inactive as the DNA binding region is disrupted. When bound to inducer, the conformation changes. DNA binding and transcription activation can then occur. (b) Transcriptional activation by RhaS. The active dimer RhaS binds to its target promoter, P_{rhaBAD} just upstream of and slightly overlapping the -35 hexamer. This allows it to interact with RNAP holoenzyme, stabilising the complex and ensuring transcription initiation.

more reliable experimentally as the response with non-metabolised inducers is sustained [113]. The various states of RhaS and its mechanism are presented in Figure 2.8.

The activator protein RhaS is part of the AraC/XylS family of transcription activators [75]. It is found in *E. coli* as part of the L-rhamnose metabolism. The position of the genes and the action of an activator in the pathway was first discovered in 1967 [169], but the regulatory mechanism in this pathway was first presented in detail by Tobin and Schleif in 1987 [204]. They described the *rhaSR* promoter lying upstream of a region encoding a polycistronic mRNA with both *rhaR* and *rhaS* genes. These genes encoded the transcriptional activators RhaR and RhaS which are constitutively expressed at low levels [89]. In the presence of L-rhamnose, RhaR transcriptionally activates *rhaRS* [204] thereby autoactivating. A cascade is then triggered as the accumulation of RhaS in the cell leads to the activation of the *rhaBAD* operon governed by P_{rhaBAD} [63] and the *rhaT* gene [213]. The *rhaT* gene encodes a rhamnose-proton symporter [213] that then transports L-rhamnose into the cell, further increasing the concentration. The *rhaBAD* operon encodes enzymes required for the processing of L-rhamnose. The promoter P_{rhaBAD} is subject to catabolite repression, but it seems that at high enough concentrations of RhaS, the promoter can be reliably induced without activation by CAP which binds just upstream of RhaS [113].

RhaR is the better characterised of the two activators [205], as it is water soluble at high concentrations. This turns out to be due to the specific mechanism of regulation. In the absence of the inducer L-rhamnose, RhaR binds its target promoter, the *rhaSR* promoter, but does not recruit RNAP [206] as the region of the protein responsible for recruitment of RNAP is disrupted [118]. RhaR is a dimer [205] and RhaS is assumed to be as well due to its homology with AraC [63]. In support of this is that RhaR and RhaS both bind inverted repeats [26, 63, 64, 213]. The reason

why RhaS has not been well characterised is that the DNA binding sequence is disrupted when no inducer is present [118]. This exposes hydrophobic amino acids which lead to aggregation of RhaS when it is expressed in high concentrations, making the molecule difficult to deal with experimentally [64]. Both RhaR and RhaS have two domains, a DNA binding domain and a dimerisation and inducer binding domain. It was shown that the specific sequence of linkers between them did not affect the behaviour of the proteins [117] and that in the case of RhaS, the DNA binding domain expressed without the dimerisation and inducer binding domain actually activates expression efficiently without the need of inducer, though to a much lesser extent that the full protein [219]. The inducer activated RhaS dimer binds just upstream of the -35 hexamer and overlaps it slightly [26]. RhaS stabilises the interaction between the RNAP holoenzyme using σ^{70} and the DNA ensuring transcriptional initiation, as shown in Figure 2.8b.

2.1.15 *Small RNAs*

A prominent member of the functional RNA family is the sRNA. These are short, non-coding strands of RNA that play a number of different regulatory roles in bacterial cells including, but not limited to, translation initiation regulation, translation quality control, mRNA cleavage, transcription attenuation, protein inhibition and secretion regulation [85, 217]. They are usually between about forty and a couple of hundred nucleotides long and provide a powerful tool to synthetic biologists.

sRNAs are a prokaryotic device and were discovered in the early 80'ies [199, 207] long before their more complex analogs in eukaryotes [70]. Translation regulation by sRNA, which is the focus of this section, can occur through one of four mechanisms: inhibition or activation of translation, stabilisation or destabilisation of the mRNA target [85, 159, 198, 217].

sRNAs are the predominant method of translation regulation in prokaryotes. In excess of 80 sRNAs have been characterised in *E. coli*, with a wide range of functions [10, 85, 86, 198, 217]. Prokaryotes mainly use sRNAs in their stress response pathways [88, 138, 159] as they allow for rapid propagation of signals through a network [2].

In natural systems, sRNA genes are often found in *trans*: encoded in a location on the genome away from the target gene. In many cases sRNAs are also found in what is called *cis*-antisense, where the sRNA is encoded in the DNA directly across from and overlapping with the target gene [78]. Such a set-up can affect transcription of a gene in several ways [34], in addition to the translational regulation provided by the sRNA.

The interaction between sRNAs and their target mRNAs occurs through complementarity. The affinity and binding strength between an sRNA and its target depends on the length of the strand, the level of its complementarity and the secondary structure of the RNAs involved. *Cis*-antisense sRNAs have a very high affinity due to containing a section of perfect complementarity with their targets [78]. The annealing of an sRNA to its target mRNA creates an RNA heteroduplex, a double stranded section of RNA consisting of two strands of different primary structure and length.

The mechanisms of translation regulation focussed on here are inhibition and activation of translation initiation. In spite of this destabilisation and stabilisation are worth mentioning. The destabilisation of target mRNAs by sRNAs functions by recruiting the RNA degradation machinery of the cell, called the degradosome, prematurely such as to reduce the half-life of the target mRNA, thereby minimising the translational product and expression of the gene [136]. Recently, the exact mechanisms by which these interactions occur have begun to come to light [123, 185]. These often work in conjunction with the translation inhibition mechanism discussed below.

In addition to the above, Storz *et al.* postulated [198] and then confirmed [159] that sRNAs are also used to stabilise mRNA. This increases mRNA levels in the cell by increasing the target mRNA's half-life, thereby increasing expression of the gene. It was shown in [159] that the sRNA binds the 3′ end of the target mRNA and that this was crucial to stabilisation. Though the mechanism is not fully understood, it is thought that the sRNAs block ribonucleases from accessing the mRNA, prolonging the lifetime of the target.

2.1.15.1 *Translation Inhibition by sRNA*

This is the most commonly observed sRNA regulation mechanism. The predominant mechanism by which this occurs is through an sRNA blocking physical access to the translation initiation region of the target by binding across the RBS, covering the Shine-Dalgarno region and/or Start Codon of the target mRNA (see Figure 2.9). This blocks the ribosome from binding and therefore inhibits translation. Alternative mechanisms might be possible as it has for example been found that there is a window of up to five codons from the Start Codon in which sRNA binding can lead to translation inhibition [30].

2.1.15.2 *Translation Activation by sRNA*

Activation by sRNAs occurs as shown in Figure 2.10. A target mRNA folds into a hairpin structure rapidly after transcription sequestering the RBS and/or Start Codon of a gene, blocking access by the ribosome and thereby inhibiting translation. The hairpin opened up by an incoming activ-

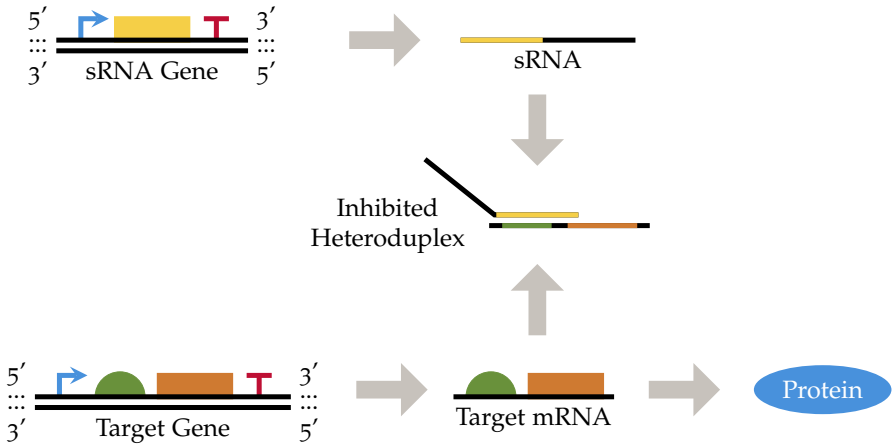


Figure 2.9 – The sRNA inhibition mechanism. Two genes encoded in DNA are presented on the left of the figure. The top gene is the inhibitory sRNA gene which is transcribed into its active form. Its target is the bottom operon, encoding a protein, which is first transcribed and then translated into protein. The active sRNA blocks transcription by binding the target mRNA covering the RBS and Start Codon. This forms an RNA heteroduplex that is inactive.

ating sRNA. The ribosome can then access the RBS of the target mRNA and the gene is translated and expressed.

2.1.15.3 *Complex Roles of sRNAs*

The relationship between naturally occurring sRNAs and their targets is not always one-to-one; sRNAs play complex multifaceted roles in the cellular environment, often having several targets or combining the above mentioned mechanisms [24, 73, 190]. To give examples of the rich behaviour demonstrated by sRNAs, the following includes some natural examples.

Oxidative stress in *E. coli* induces the transcription of *oxyS*, an sRNA gene that has at least eight targets [11]. Not only that, but the sRNA OxyS inhibits translation of some targets and activates translation of others. The resulting heteroduplexes are not degraded rapidly.

Another example: The sRNA RhyB, part of the iron metabolism, has 18 targets, some of them polycistronic, negatively regulating a total of 56 proteins [137, 138]. Some of these heteroduplexes are degraded rapidly, and it is also thought that the sRNA can cleave certain operons, degrading some sections of the target and leaving others to be translated [137].

The *rpoS* gene encodes a sigma factor in *E. coli* and other bacteria called σ^s or σ^{38} that is not present in the cell during exponential growth, which occurs when nutrition is abundant, but is expressed when the cell is under stress or in stationary phase. This is an example of translation regulation of one target by several sRNAs. The *rpoS* mRNA has at least three associated sRNAs that all activate translation via the mechanism described in Figure 2.10, and studies point to there being more [20]. The *rpoS* mRNA has a hairpin at the 5′ end that occludes the RBS inhibiting

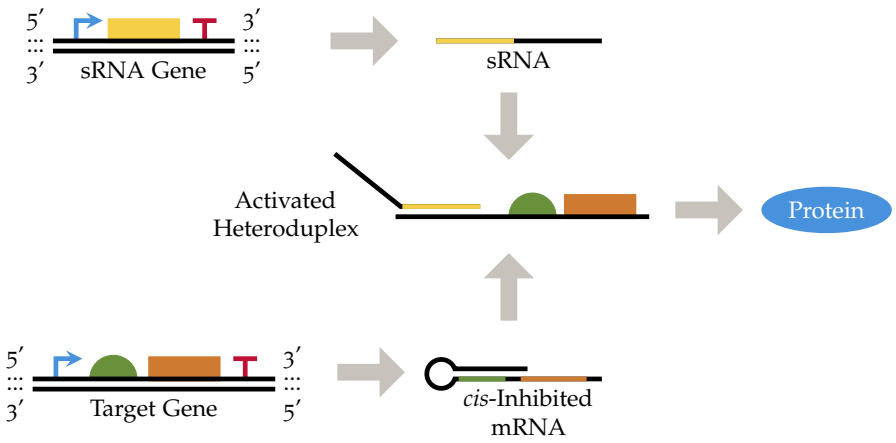
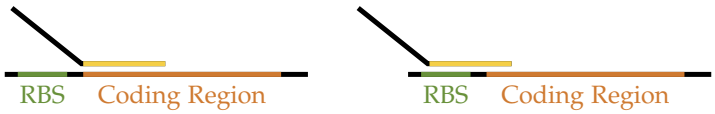


Figure 2.10 – The sRNA activation mechanism. Here the target gene produces a *cis*-repressed mRNA that folds up on itself forming a hairpin that blocks access to its RBS and/or Start Codon, thereby inhibiting translation. When expressed, an activating sRNA then opens up the hairpin in the target mRNA allowing access by the ribosome and translation of the gene encoded in the mRNA.



(a) Na *et al.* (b) Man *et al.*

Figure 2.11 – The sRNA target binding strategies suggested by (a) Na *et al.* in [151, 224] and (b) Man *et al.* in [134]. Na *et al.* suggest that the tightest regulation is produced when the sRNA binding region is complementary to the target starting with the start codon at the beginning of the coding region of the target mRNA. Man *et al.* on the other hand suggest that the sRNA binding region be complementary to the target starting with the Shine-Dalgarno sequence in the RBS and finishing 20-30 nucleotides downstream in the coding region.

translation when the cells are growing. Under stress, the cells express the sRNAs and the hairpin is opened up. One of the *rpoS* associated sRNAs called DsrA acts as both a transcriptional inhibitor and a temperature sensitive activating sRNA [133, 198]. DrsA turns out to be modular, with two different sections, one performing each role.

There are even examples of sRNAs regulating each other. This mechanism is used in *Salmonella* where an inhibiting sRNA with an mRNA target is destabilised by a second sRNA, facilitating the expression of the message [69].

2.1.15.4 *Implementation of sRNAs*

These rich networks and dynamics point to a high level of flexibility offered by sRNA regulation which can be transferred to synthetic systems, potentially allowing for more complex and designed regulation patterns in larger systems.

Experimental implementations of sRNAs are becoming more common. Implementations using simple synthetic sRNAs to inhibit translation of targets in bacteria have been reported by several

groups [41, 134, 151]. These groups use sRNA inhibition to knockdown target genes. The sRNAs designed are all encoded in *trans*, away from their target genes, as this facilitates design. An sRNA can be thought of a having three main regions: a target binding region, a Host-Factor of Bacteriophage Qβ (Hfq) protein binding region and a terminator region, though the two last regions heavily overlap [210].

Na *et al.* performed metabolic engineering assays to optimise certain yields [151] using a protocol that was published [224]. This protocol suggests using the scaffold of MicC, an sRNA native to *E. coli* [44]. The scaffold is the native MicC sRNA with the 24 nucleotide target binding region removed. It includes the Hfq binding region and the factor-independent terminator. The 24 nucleotide target binding region can then be designed and added synthetically. Na *et al.* and the associated protocol suggest that the binding region of the sRNA with the highest efficiency of knockdown is from the start codon and downstream for a total of 24 nucleotides. This results in an sRNA that only binds to and occludes a section of the gene coding region of the target mRNA.

Man *et al.* in [134] also supply a method for synthetic design and implementation of inhibitory sRNAs. Man *et al.* do not suggest the use of any specific scaffold. They experimented with several natural scaffolds and confirm the need for careful design of the Hfq binding and terminator regions of an sRNA. Their study suggests that an AU rich region is key to Hfq binding. Man *et al.* contradict Na *et al.* as they suggest that the target binding region should start at the Shine-Dalgarno sequence in the middle of the RBS of the target and then continue down stream for 20-30 nucleotides, thereby covering the Start Codon as well. The differences between these binding strategies can be seen in Figure 2.11. Man *et al.* also showed that in the case of some of their synthetic sRNAs, the target mRNA was destabilised by the interaction with the sRNA and rapidly degraded.

Both of these methods are relevant for design of synthetic sRNAs, and a key factor mentioned by both is the effect of off target binding. Checking off target binding of a designed sRNA can be effected using the 'Target RNA' tool [114] that suggests possible off target interactions of a submitted RNA sequence given a specific host and its genome.

Many synthetic sRNA-like regulation mechanisms have been developed due to our good understanding of RNA secondary structure and these exhibit a wide variety of behaviours [43, 51, 106, 130, 220]. An example of this are sRNA riboswitches, such as that presented in Figure 2.12. An sRNA is transcribed with a structure that is inactive as the sRNA's target binding region is occluded. The addition of a small molecule effector that interacts with the sRNA then changes its secondary structure, revealing the target binding region, which can then bind and regulate its target. Methods to allow the reliable design of such systems have been developed [22].

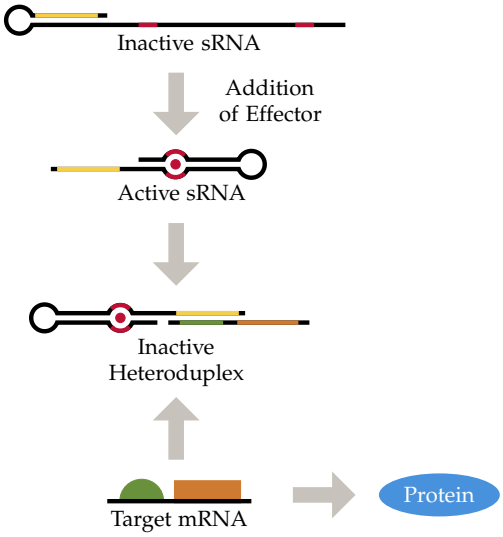


Figure 2.12 – Effector activated sRNAs. An sRNA (top) is inactive when transcribed due to occlusion of the sRNA binding region. Its target mRNA (bottom) is translationally active post transcription. Addition of an effector molecule changes the secondary structure of the sRNA revealing the the sRNA binding region. The active sRNA can now bind its target. In this case, the sRNA covers parts of the RBS and/or gene coding region thereby inhibiting translation, but the sRNA can also be designed to activate translation, or regulate in other ways.

One of the foci of this project is the use of the basic sRNA inhibition and regulation mechanisms. The extension to incorporate these more elaborate sRNAs is the natural next step.

2.1.16 *Hfq Chaperone Proteins*

In prokaryotic cells, the binding of a chaperone protein is necessary for effective regulation by a large number of sRNAs [4, 32, 53, 88, 198]. Originally discovered almost fifty years ago, this chaperone was found to be required for the replication of bacteriophage Qβ [71]. This gave it the name Host-Factor of Bacteriophage Qβ (Hfq).

Hfq is a homohexamer[13] that forms a toroid, similar to the aforementioned ρ factor. It has been studied in detail due to its pleiotropic role and is well conserved across bacteria [32, 198, 210]. It is primarily an RNA binding protein but performs a number of other roles in the cell, including interacting with DNA and other proteins such as ρ-factors [53, 195, 214].

In the context of sRNA regulation, Hfq catalyses pairing of the target mRNA and regulating sRNA. In the cell, this is a complex process as there are many target-sRNA pairs competing for Hfq to catalyse their finely balanced interactions. Hfq facilitates these interactions through its three RNA binding surfaces named the proximal and distal faces, and the rim [210]. The proximal face has been shown to bind poly-U-tails and a hairpin associated factor-independent terminator

13 It is made up of six identical subunits.

regions (see Section 2.1.7.2), as are found on most natural sRNAs. The distal face binds more flexibly. In *E. coli*, the distal face binds RNA sections with a preferred AAN repeat motif, where N can be any ribonucleotide. The rim binds AU rich sequences. In addition to this, each of the six hexamers interacts with RNAs independently of the others, which gives the protein advantageous properties [68]. sRNAs can be broadly split into two classes: class I has the sRNA binding the proximal face and the rim and the target mRNA binding the distal face and class II vice-versa.

Hfq proteins act as chaperones, protecting the sRNAs from degradation and modifying secondary structure [57, 77, 99, 102, 136, 211]. By having multiple binding surfaces, Hfq is able to actively recruit other RNAs, in effect increasing local concentrations and binding rates [4, 147]. In addition, due to its electrostatic properties, the protein aids in overcoming the negative charge barrier to RNA binding and stimulates nucleation of RNA heteroduplexes [214]. It is estimated that there are between 400 and 10000 Hfq hexamers per cell, which are heavily associated with the ribosomes [32, 210]. Their number is a definitive limiting factor. Overloading the cell with sRNAs to chaperone has a disruptive effect on regulation [102, 150].

Experimental results studying the dissociation constants of complexes between various RNAs and Hfq find the half-lives to be on the scale of several cellular generations [68, 158, 210], with binding between Hfq and RNAs being extremely tight. This does not agree with the observed response times of cellular stress responses that are on the scale of minutes [163], as given slow low rates of dissociation, sRNAs or targets expressed in response to stress would not be able to interact on this time scale.

The observed mechanistic solution to this is termed 'cycling' [68, 216]. In the presence of a single RNA, Hfq binds it tightly, but when other RNAs are introduced, the dissociation rate increases dramatically reducing half-lives of the complexes to minutes as is required for cellular stress responses. Though not shown explicitly, it has been suggested from observed behaviour that the pool of RNAs capable of binding Hfq in a cell is always large enough to ensure that cycling occurs, reducing the half-lives of Hfq-RNA complexes to the required range [210]. This mechanism occurs through reversible interactions between the individual monomer subunits in the Hfq hexamer. An RNA bound to Hfq is often in contact with many, but not all, of the six subunits. The free subunits can then interact with other free RNAs. Newly bound RNAs can then reversibly replace the resident RNA replacing the resident's interaction with the other subunits, as shown in Figure 2.13. Despite this cycling mechanism being a generally observed effect, an individual RNA's ability to compete for Hfq depends on its specific structure [158]. As a result, some RNAs compete more efficiently than others.

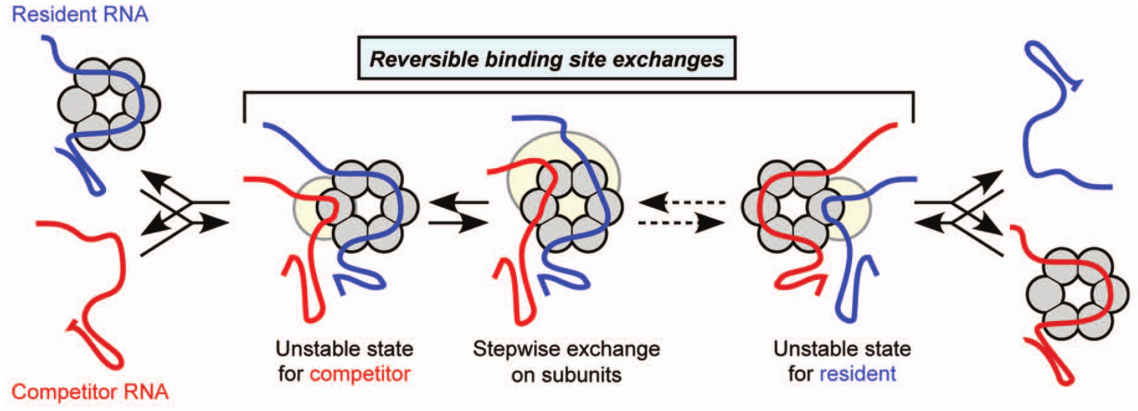


Figure 2.13 – The mechanism of cycling of RNAs bound to Hfq. The resident RNA is replaced in a reversible and stepwise manner by the competitor by being ejected one subunit at a time. Taken from [216].

2.1.17 *Reporter Proteins*

There are a last few pieces that need to fall into place to complete the picture. The explanations above discuss how genes in bacteria are expressed and present tools for controlling expression at various levels. This section presents one method by which expression of a protein can be measure in real time, through what is known as a reporter gene.

The most common type of reporter gene currently used is probably the fluorophore. Fluorophores are chemical species that absorb light of a certain wavelength and then emit the absorbed energy as light of a slightly longer wavelength. Expression of a protein is measured using the fluorescence of such reporters as a proxy. The cells are illuminated with light at the excitation wavelength, and then the intensity of the emitted light is measured. This can either be done at culture level, where a large number of cells in solution are illuminated, or at single cell level, using techniques such as flow cytometry.

In flow cytometry, cells are passed through a narrow canal one after the other. The individual cells are illuminated and their fluorescence intensity measured. Modern machines analyse cell cultures in this manner at a rate of thousands of cells a minute, yielding data that accurately describes variations in levels of protein expression across a population.

The most common choice of reporter protein are a group of fluorophores derived from Green Fluorescent Protein (GFP). Originally discovered in the early 60'ies and named Aequorin [189], this protein stems from the jellyfish *Aequorea victoria*. It has since been engineered to produce more robust versions [49, 164] and many different colours including blue, yellow and cyan.

The version of GFP used in this project is super-folder Green Fluorescent Protein (sfGFP) [164]. This mutant has been shown to fold incredibly efficiently, even when tethered. The protein is

reported to have peak absorption at 485nm and have peak emission at 510nm. Photobleaching, which is the destruction of the protein by absorption and emission of light was reported to occur relatively slowly, with a 20-25% reduction in intensity after 40 minutes of constant illumination.

2.1.18 *Putting the Pieces Together*

The above sections describe a subset of the toolbox available to a synthetic biologist and many specifics of all the biological mechanisms relevant to this project. The pieces now need to be designed and put together to achieve the desired objectives. To perform analysis and rational design of these biological systems, the world of mathematics is now called upon.

The next sections describe how to model these systems and present some approaches that can be used to design.

2.2 MODELLING OF GENE REGULATORY NETWORKS

This section presents standard methods of mathematical modelling in the context of biochemical networks. It is worth remembering that there are many types of model and that all models are an abstraction of reality intended to show specific properties and facilitate study. To that end, the descriptions of biological mechanisms in the above section now need to be distilled into a form that can be studied, analysed and even designed using mathematical tools.

Rather than more generally treating modelling of chemical networks, this text refers specifically to the modelling of gene regulatory networks, as this is the context of the project. The cellular environment in which the systems modelled exist is a busy one. Modelling of an entire cell, including all possible interactions, has been performed with some success [110], but comes at huge computational cost with little benefit to understanding. The approach taken here is to model a subsystem of a cell, only focussing on directly relevant species. This provides a much more tractable model, both analytically and computationally.

This section focusses on modelling gene expression and then presents the standard method of modelling expression control via activation and repressor of transcription. For the details of models of sRNA regulation, see Chapter 3. For further reading, [8, 104] are useful in the context of continuous modelling, while [209] focusses on stochastic techniques.

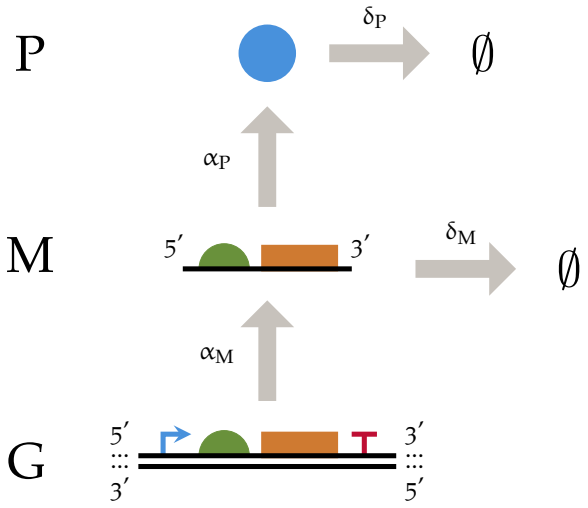


Figure 2.14 – A model of gene expression. This is a visual representation of a possible model of gene expression from a constitutive promoter. The chemical species are labelled G, M, P which refer to the gene, the mRNA and the protein respectively. Transcription and translation are represented by the grey arrows between the gene and mRNA and the mRNA and protein respectively. The degradation of the mRNA and protein are represented by the grey arrows pointing to the empty set symbols, $\emptyset$. The rates at which these reactions occur are defined by the rate constants: the rate of transcription of the gene is α_M, the rate of translation of the mRNA is α_P and the rates of degradation of the protein are δ_M and δ_P respectively.

2.2.1 *Construction of Model*

The fist step in producing a mathematical model is to provide a biochemical model to convert into the language of mathematics. The biochemical model model consists of all the biological interactions to be included in the model. The clearest way to list these interactions is using chemical reactions.

The clearest means of introducing this method is through example, and to this end the expression of a single gene is used. A diagram of the model is presented in Figure 2.14 and the chemical reactions are written as follows using standard notation:

$$G \longrightarrow G + M, \tag{2.1a}$$

$$M \longrightarrow M + P, \tag{2.1b}$$

$$M \longrightarrow \emptyset, \tag{2.1c}$$

$$P \longrightarrow \emptyset, \tag{2.1d}$$

where the gene being expressed is labelled G. This is transcribed into mRNA, labelled M, in Equation 2.1a and then translated into protein, labelled P, in Equation 2.1b. Degradation of mRNA and protein are included through the reactions (2.1c) and (2.1d) respectively. The species

on the left hand side of these equations are referred to as the reactants, and the species on the right as the products. The reactants are all required for the reaction to occur and the products are produced by the reaction. Note how, in transcription and translation, the gene and mRNA respectively function as a catalyst, appearing on both sides of a chemical equation, being neither consumed nor produced.

The reactions listed in Equation 2.1 are one possible representation of gene expression. Though transcription and translation are complex processes that require the execution of a number of different chemical interactions with a large number of species involved, it is standard practise to model these as single steps [8, 104].

As this model is considerably simplified, it is based on a number of assumptions. For example, the model assumes an endless supply of biosynthetic parts such a nucleotides and amino acids and it does not take the energy consumption of the expression process into account. Assuming that these factors do not play a role might be reasonable if the cells are surrounded by a large amount of nutrients which they can absorb and use freely. These simplifications of, and abstractions from the real system are core elements of modelling.

Now that the chemical reaction network has been outlined, attention is turned to the world of mathematics to provide tools to aid in the study, analysis and even design of such systems.

2.2.2 *The Law of Mass Action*

Reactions occur through the physical interaction of the chemical species involved. Given two assumptions:

A1: the container, i.e. the cell is well stirred and homogeneous,

A2: and the numbers of particular chemical species in the cell is high enough to be accurately described using continuous concentrations as opposed to discrete values,

the Law of Mass Action states that the rate at which a chemical reaction occurs is proportional to the concentrations of the reactants involved. These constants of proportionality are called

reaction rate constants and are always positive. Here they are added to the reactions describing gene expression:

$$\begin{aligned} G &\xrightarrow{\alpha_M} G + M \\ M &\xrightarrow{\alpha_P} M + P \\ M &\xrightarrow{\delta_M} \emptyset \\ P &\xrightarrow{\delta_P} \emptyset \end{aligned} \tag{2.2}$$

where the reaction rate constants for transcription and translation are labelled α_i, where, in this case, $i \in [M, P]$, where M refers to the mRNA and P refers to the protein. The degradation rates of mRNA and protein are δ_M and δ_P respectively. In general δ_i represents the degradation or removal rate of species i from a system, whereas α_i represents the synthesis of species i.

The reaction rate constants, α_i and δ_i in the above example, are the parameters of the mathematical model. These are mechanistic as the can be related to biochemical processes. For example, the transcription rate, α_M, can be related to the strength and efficiency of the promoter, defined by how efficiently it interacts with the RNAP holoenzyme and other species to initiate transcription. Likewise, the translation rate, α_P can be related to the strength of the RBS on the mRNA.

2.2.3 *Ordinary Differential Equations*

Using an Ordinary Differential Equation (ODE), the evolution over time of the concentration of a chemical species can be described. The concentration of the chemical species is referred to as the state. More generally a set of ODEs describing a system with a vector of states denoted x, can be written

$$\dot{x} = f(x), \tag{2.3}$$

where $\dot{(-)}$ refers to the first derivative with respect to time, the state $x \in \mathbb{R}^N_{\geqslant 0}$, where $\mathbb{R}_{\geqslant 0}$ is the set of all positive rational numbers and zero, and N is the number of states or order of the system and $f(-)$ is a nonlinear function. As the states are chemical concentrations, the elements of x cannot take negative values, but $f(x)$, and therefore $\dot{x}$, can in principle take any value in $\mathbb{R}^N$.

The reactions in Equation 2.2 describe the transcription or production of species M, and the translation or production of species P, and the degradation of both of these. There is no change to the species G, as the gene simply catalyses transcription. Two ODEs are therefore required

to describe this system, making it a second order system. The chemical reactions must now be converted into mathematical statements. Each chemical reaction contributes one term to the differential equation for each of the species it consumes or produces; consumption terms are negative, production terms are positive. In the context of this example, the transcription and translation reactions produce M and P respectively, and therefore contribute a single positive term to each of their differential equations.

Denoting the concentration of a species by its lower case, m for the mRNA M, p for the protein P and g for the gene G, the following differential equations describing the system can be written:

$$\begin{aligned} \dot{p} &= \alpha_P m - \delta_P p \\ \dot{m} &= \alpha_M g - \delta_M m, \end{aligned} \tag{2.4}$$

where the parameters are the same as in Equation 2.2. Such a system of ODEs can be simulated using standard solvers implemented in software such as MATLAB. ODEs can also be analysed and manipulated using a number of techniques.

In the context of ODEs, the rate of a reaction is measured in units of concentration over time. In biochemical networks in cells the units are typically $[\frac{nM}{min}]$ or 'nano-molar per minute'. As the rate of a reaction is dependent on the concentrations of all reactants involved, the units of the reaction rate constants vary. In reactions with one reactant, such as the degradation reactions, the unit of the reaction rate constant is just $[min^{-1}]$. In general, the unit of the reaction rate constant is $[min^{-1} nM^{(1-n)}]$, where n is the number of reactants in the reaction.

2.2.3.1 *Model Reduction by Time Scale Separation*

Model reduction techniques are a classic tool used to simplify a model, making it more tractable. In the context of biochemical reactions, time scale separation is a common method of doing this and it is therefore introduced in this section.

As discussed in Section 2.1.10, mRNAs have a short half-life, on the order of minutes [87]. Unless specifically tagged for degradation, protein in the cell is not degraded. There is however another significant effect that gives rise to the removal of protein, and other species from the system; that is dilution. When surrounded by lots of nutrients, cells grow and fission to multiply. During this process, cells grow their cell walls, increasing the volume within the cells. This process constantly reduces the concentration of protein and other species in the cell by diluting it. Assuming that the cell is twice its original volume just prior to division and that it grows at a relatively constant

rate during the cell cycle, the rate of dilution is then closely tied to the average time it takes for cells to divide. This is called the doubling time and for *E. coli* this can be anywhere from about 20 minutes to hours depending on the media the cells are in, the temperature, and how many and at what concentrations antibiotics are used (if any). In reality, proteins are also degraded due to damage and destabilisation, but the rate at which this occurs is much slower than the doubling time, on a scale of up to 20 hours [140], meaning that it can be neglected. The removal of protein from the system is generally referred to as degradation when modelling, analogous to mRNA degradation, but it is worth keeping in mind that unless *ssra* tags are involved (see Section 2.1.10.2), protein is only removed by dilution.

This large difference in time scales between degradation of mRNA and protein can be exploited to simplify the ODE model in Equation 2.4 from two equations to one using singular perturbation. As this method is used later in this thesis, it is introduced here.

The degradation rate of mRNA, δ_M, is much larger than the degradation rate of protein, δ_P assuming that protein is only diluted. This provides a time scale separation which is exploited to remove the mRNA state m in the following way. Starting from Equation 2.4, perform the following change of variables, rescaling the two states m and p and time t to nondimensionalise the system:

$$\begin{aligned} p &= \phi \bar{p} \\ m &= \psi \bar{m} \\ t &= \theta \tau, \end{aligned}$$

where $\phi = \frac{\alpha_P \psi}{\delta_P}$, $\psi = \frac{\alpha_M g}{\delta_P M}$ and $\theta = \frac{1}{\delta_P}$ and $(\bar{-})$ signifies the nondimensionalised states and τ nondimensionalised time. This yields the equations

$$\begin{aligned} \frac{d\bar{p}}{d\tau} &= \bar{m} - \bar{p} \\ \epsilon \frac{d\bar{m}}{d\tau} &= 1 - \bar{m} \end{aligned}$$

where $\epsilon = \frac{\delta_P}{\delta_M}$. As the rate of degradation of mRNA is much larger than that of protein, $\epsilon \ll 1$. In the limit $\epsilon \to 0$, the left side of the second equation goes to zero and yields $\bar{m} = 1$. Substituting into the first equation and reverting to the original variables yields:

$$\dot{p} = \alpha'_P - \delta_P p$$

where $\alpha'_P = \frac{\alpha_P \alpha_M g}{\delta_M}$ can be thought of as the expression rate of the protein. The validity of this approximation is reasonable as long as ϵ remains small. If degradation tags are used to bring δ_P to a size comparable to δ_M, then it no longer holds.

2.2.3.2 *ODE Modelling of Promoter Activation and Repression*

Transcriptional regulation, which is presented in Section 2.1.14, is the most common method of expression regulation. This section introduces the standard way that these mechanisms are modelled. Modelling these interactions introduces reaction rate functions into a chemical reaction network. To derive the functions, only the interactions between the Transcription Factor (TF), whether activating or repressing, and the promoter region it associates with are considered. Let us assume that n TFs need to bind the operator regions to regulate a gene. This interaction can be written

$$nX + G \underset{k_r}{\overset{k_f}{\rightleftharpoons}} C$$

where X is the active TF, G the operator of the gene it regulates, C is the complex between the n TFs and the gene. The parameters k_f and k_r are the forward and reverse reaction rates of the interaction respectively. Let us now write the ODE governing the change in concentration of the complex:

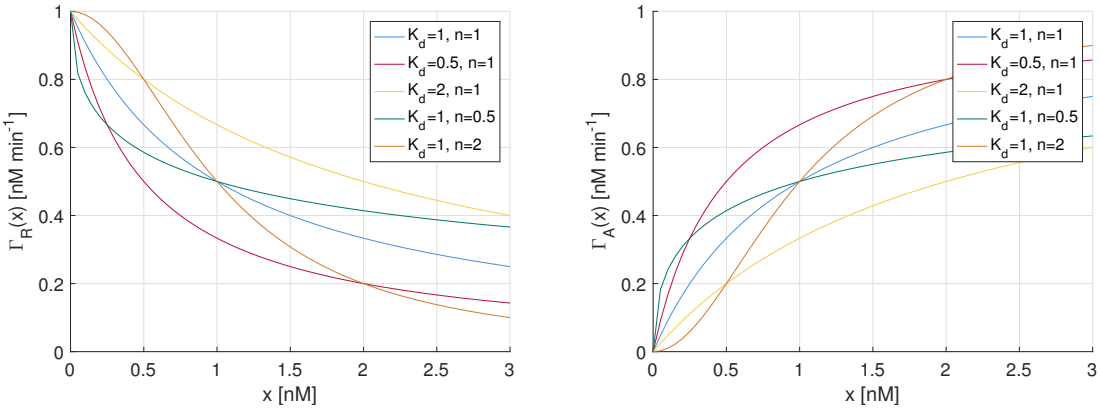
$$\dot{c} = k_f x^n g - k_r c$$

where the concentrations of the chemical species are denoted by their lower case. Note that the concentration of the TF is raised to the power n as n TF act as reactants in the interaction with the gene. The total concentration of gene is the sum of free unbound gene, g, and the gene bound in the complex c: $g_T = g + c$. Assuming equilibrium by setting the left hand side of the above equation equal to zero and substituting $c = g_T - g$ in yields:

$$\frac{g}{g_T} = \frac{K_d^n}{K_d^n + x^n} \tag{2.5}$$

where $K_d = \left(\frac{k_r}{k_f}\right)^{\frac{1}{n}}$ is often called the dissociation constant. This is called the Hill Function and is the fraction of free gene given a concentration of TF x. Substituting for g yields:

$$\frac{c}{g_T} = \frac{x^n}{K_d^n + x^n} \tag{2.6}$$



(a) Transcriptional Repression

(b) Transcriptional Activation

Figure 2.15 – The behaviour of the rate functions detailed in Equation 2.7. The parameter values used are presented in the legend, except the maximal rate, which is fixed to $\alpha = 1\text{nMmin}^{-1}$. The functions are related: $\Gamma_R(x) = 1 - \Gamma_A(x)$. Changing K_d shifts the point x for which $\Gamma_R(x) = \Gamma_A(x) = 0.5$ and changing the Hill Coefficient n changes whether the function is sensitive to low or high concentrations of TF.

where K_d is the same as above. This is the fraction of gene-TF complex given a concentration of TF x, and is often called the Michaelis-Menten Function. In both the above cases K_d gives the concentration of TF at which the fractional occupation of the gene is a half.

It is assumed that in the case of a repressor, the transcription initiation is blocked when the active TF is bound to its operator region, associating it with Equation 2.5. Similarly, in the case of an activator, the transcription initiation only occurs when the active TF is bound to its operator region, associating it with Equation 2.6. These are reasonable assumptions given the biological mechanisms.

Assuming that the expression from a regulated promoter is proportional to the fractional occupation, the rate functions of transcription or expression from repressed and activated promoters can be written

$$\Gamma_R(x) = \frac{\alpha K_d^n}{K_d^n + x^n} \qquad \text{and} \tag{2.7a}$$

$$\Gamma_A(x) = \frac{\alpha\, x^n}{K_d^n + x^n} \tag{2.7b}$$

respectively, where $\Gamma_R(x)$ is the rate function of a repressed promoter, $\Gamma_A(x)$ is the rate function of an activated promoter and α is the maximal rate of transcription or expression. In these functions n is referred to as the Hill Coefficient and describes the cooperativity of the TFs when binding their target gene. The behaviour of these functions for a number of different parameters is presented in Figure 2.15.

Some promoters have a tendency to 'leak'. This is when a regulated promoter initiates transcription despite being completely down regulated. This is also often referred to as the basal expression rate of a promoter. This effect is included by the addition of a constant term to the transcription rate functions above, for example in the case of a repressed promoter the total transcription rate taking leakiness into account is $\alpha_0 + \Gamma_R(x)$, where α_0 is the basal transcription rate and the maximal transcription rate possible from the promoter is $\alpha_0 + \alpha$. In the case of tight promoters, such as the tetracycline promoter or the rhamnose promoter, this basal rate can be ignored.

Incorporating regulation into models is then straight forward. Changing the system modelled in Equation 2.2 to include repression of the gene being expressed by a TF denoted X requires only the inclusion of the relevant rate function in the chemical reactions:

$$\begin{aligned} G &\xrightarrow{\Gamma_R(x)} G + M \\ M &\xrightarrow{\alpha_p} M + P \\ M &\xrightarrow{\delta_m} \emptyset \\ P &\xrightarrow{\delta_p} \emptyset \end{aligned} \tag{2.8}$$

and the equivalent change in the associated ODEs.

2.2.3.3 *Modelling the Transcription Factor Effector Interaction*

The most common way of modelling the interaction between a effector molecule and its target TF is using fractional functions on the same form as Equation 2.5 and Equation 2.6, but with $n = 1$ [8]. The chemical reaction used to model the interaction is

$$X + E_X \underset{k_r}{\overset{k_f}{\rightleftharpoons}} C$$

where the TF is denoted X and the effector associated with X is denoted E_X, C is now the complex of the effector and TF, and k_f and k_r are the forward and reverse reaction rates of the interaction. This one to one relationship between TF and effector is reasonable in many cases as with TetR and RhaS, as each subunit of the dimer TFs binds an effector molecule each, though setting $n = 1$ excludes any form of cooperativity from the model. The next step is to consider which state is the active form of the TF.

In the case of a repressor, binding of effectors, known as the inducers, inactivates the TF. Therefore the promoter can only interact with the free and active repressor. Similarly to Equation 2.5, the concentration of free and active repressor takes the form:

$$x = \frac{K_d x_T}{K_d + e_X}, \tag{2.9}$$

where x is the concentration of free and active repressor and x_T is the total concentration of repressor written $x_T = x + c$, where c is the concentration of TF bound to inducer, and $K_d = \frac{k_r}{k_f}$ is again referred to as the dissociation constant.

In the case of the activators and apo- and corepressors, the TF is only active when bound to effector, so the function of interest is the concentration of complex. Similarly to Equation 2.6, this takes the form:

$$c = \frac{x_T e_X}{K_d + e_X}, \tag{2.10}$$

where the symbols have the same significance as in Equation 2.9.

Now that we have expressions for the active forms of the TFs, it is now possible to model adding an effector to a system. Take the interaction between a repressor X, with total concentration x_T and free concentration x given by Equation 2.9, interacting with inducer E with concentration e_X. The rate function in this case is

$$\Gamma_R(e_X, x_T) = \frac{\alpha K_G^n}{K_G^n + \left(\frac{K_E x_T}{K_E + e_X}\right)^n} = \frac{\alpha K_G^n (K_E + e_X)^n}{K_G^n (K_E + e_X)^n + K_E^n x_T^n}, \tag{2.11}$$

where α is the maximal expression rate, K_G and n are the dissociation constant and Hill coefficient of the interaction between the active TF and the gene operator regions and K_E is the dissociation constant of the interaction between the inducer and the TF. As can be seen, the value of the Γ function grows with increasing inducer concentration, which is what is expected.

An analogous expression for the activator and apo-corepressor mechanisms can be derived in the same manner.

2.2.4 *Stochastic Simulation Using the Gillespie Algorithm*

The cellular environment is innately stochastic in nature. The cytoplasm of a cell is densely populated with lots of different chemical species constantly colliding. On the time scale of mo-

lecular collisions, reactions between chemical species are rare events. The correct molecules need to collide in the correct manner to trigger a reaction. Despite this chaotic sounding foundation, living organisms manage to behave in a reliable manner, reconciling the deterministic and the stochastic, harnessing the latter to their own benefit [175].

When modelling a biological system, there are two types of noise that come into play: extrinsic and intrinsic noise. Extrinsic noise stems from changes in cell or cellular environment, which might change the make up or focus of the cellular machinery. Intrinsic noise is noise the stems from the interactions of the system itself. In the case of modelling gene expression, the intrinsic noise is the noise that stems from the stochasticity of the transcription and translation processes, whereas extrinsic noise can be thought of as fluctuations in the parameters.

The Gillespie Algorithm [81] provides an exact method of simulating chemical reaction networks taking intrinsic noise into account. The state space in this case is discrete, working only with integer numbers of the individual chemical species involved. The state space of the stochastic system is denoted X_i and lives in $\mathbb{Z}^I_{\geqslant 0}$, where $i \in \{1, 2, \ldots, I\}$ and I is the number of states or equivalently, chemical species. The chemical species associated with the i^{th} state is denoted S_i.

In a modelled system, each of the J reactions is denoted Q_j with an associated reaction rate constant c_j, where $j \in \{1, 2, \ldots, J\}$. A subset of the chemical species take part in each reaction in relative quantities known as the stoichiometry. The chemical reactions can be written on the general form

$$\sum_{i=1}^{I} a_{ij} S_i \xrightarrow{c_j} \sum_{i=1}^{I} b_{ij} S_i,$$

where a_{ij} and b_{ij} are the reactant and product stoichiometry coefficients of the species S_i in the reaction Q_j respectively and belong to the set $\mathbb{Z}_{\geqslant 0}$. These numbers define which and how many chemical species are consumed as reactants and produced as products in each reaction. From these numbers it is possible to set up the stoichiometry matrix with the elements

$$\mathcal{S}_{ij} = b_{ij} - a_{ij}.$$

In the matrix $\mathcal{S}$, each row corresponds to a species, and each column to a reaction. This formulation is useful when describing the Gillespie Algorithm.

The stochastic reaction rate constants, c_j, are closely related to the rate constants employed when modelling with ODEs. Analogous to the Law of Mass action, it is assumed that the rate

a reaction occurs at is proportional to the number of each reactant in the 'container' of the reactions, in this case the cell. It is assumed that the cell is about one micro metre in each dimension, giving a total volume $V = 10^{-18}m^3$ or $10^{-15}l$. Essentially, the ODE rate constants must be rescaled to be used to model the equivalent stochastic system. Denoting the ODE rate constant for the j^{th} reaction k_j and assuming that it has units of $s^{-1}nM^{(1-n)}$, where n is the order of the reaction, calculating the stochastic reaction rate from the ODE reaction rate requires the following formula:

$$c_j = (10^{-9} \times N_A V)^{(1-n)} k_j = 0.6022^{(1-n)} k_j, \tag{2.12}$$

where the factor 10^{-9} comes from nano in nM unit of k_j, N_A is Avogadro's Number and the second equality holds given the above assumption that the cell volume $V = 10^{-15}l$.

In the ODE case, the rate at which a reaction occurs is assumed to be proportional to the concentrations of the reactants involved. In the discrete case, this is slightly different, as the quantity used the is probability of the reaction occurring in a certain small time interval. This is proportional to the total number of unique sets of reactants in the system. For example a chemical reaction with the left side $S_1 + S_2 \xrightarrow{c_j}$ has the propensity $\mathcal{P}_j = c_j X_1 X_2$. This is very similar to the corresponding ODE term. It is more complicated when a species reacts with itself, as for example in the case of dimerisation, a reaction with the reactants $2S_1 \xrightarrow{c_j}$. Here the number of unique pairs of S_1 is given by $\frac{X_1(X_1-1)}{2}$, but it is standard to absorb the factor $\frac{1}{2}$ in the reaction rate constant.

2.2.4.1 *The Gillespie Algorithm*

To model chemical reactions stochastically, two elements must be decided at each step: 1) when the next reaction occurs and 2) which reaction it is. Dealing with when first; reactions occur as Poisson processes, occurring with independent time intervals. Given only the j^{th} reaction, its propensity, $\mathcal{P}_j$, is the mean frequency of the reaction, given that the reactants do not change. When the reaction does occur, the state of the system does change, consuming reactants and producing products depending on the reaction. Summing the propensities of all J reactions to find the total propensity $\mathcal{P}_{Tot} = \sum_{j=1}^{J} \mathcal{P}_j$, the cumulative distribution function for the time of the next reaction can be set-up:

$$\mathcal{F}(\tau) = 1 - e^{-\mathcal{P}_{Tot}\tau}.$$

Now the reaction that occurs must be chosen. The probability that the j^{th} reaction occurs is simply the relative size of its propensity to the total, $\frac{\mathcal{P}_j}{\mathcal{P}_{Tot}}$. The associated cumulative distribution function is then just a sum of these up to the k^{th} reaction:

$$\mathcal{G}(k) = \sum_{j=1}^{k} \frac{\mathcal{P}_j}{\mathcal{P}_{Tot}}.$$

During simulation, the time of the next reaction and which reaction occurs are chosen through sampling two uniformly distributed random numbers per step of the algorithm on the interval $(0;1)$. This is done using a pseudorandom number generator, such as the Mersenne Twister [139]. The above cumulative distribution functions are simply inverted to find the time and reaction. Labelling the two random numbers r_1 and r_2, the time of the next reaction is $\tau = \frac{-\ln(r_1)}{\mathcal{P}_{Tot}}$ and the reaction is chosen by finding k such that $\mathcal{G}(k-1) < r_2 \leqslant \mathcal{G}(k)$.

The Gillespie algorithm can now be presented in full:

Step 1: Initialise the system with initial states X_{i0}, a starting time $t = t_0$ and an end time t_{end}.

Step 2: Choose two random numbers, r_1 and r_2, that are uniformly distributed on the interval $(0;1)$.

Step 3: Calculate the propensities for each reaction and then the total propensity. If these are all equal to zero, then terminate the simulation.

Step 4: Find the time of the next reaction: $\tau = \frac{-\ln(r_1)}{\mathcal{P}_{Tot}}$ and update the time to $t + \tau$.

Step 5: Check if $t \geqslant t_{end}$, if this is true, terminate the simulation.

Step 6: Find the occurring chemical reaction by finding k such that $\mathcal{G}(k-1) < r_2 \leqslant \mathcal{G}(k)$. Once k has been found, the states can be updated according to the corresponding column in the stoichiometry matrix: $X_i + \mathcal{S}_{ik}$.

Step 7: Go back to Step 2 and continue the simulation.

Instead of saving every single reaction and corresponding state change, the standard is to select an array of times from t_0 to t_{end} at which to save the state. The state is saved when the time of the next reaction exceeds the next time in the array, but before the next reaction is chosen and the state updated. As the states are discrete, this gives the exact state at a given time.

2.2.4.2 *The Execution of Simulations*

As simulations employing the Gillespie algorithms are stochastic, they must be performed in large ensembles to facilitate the reliable extraction of the moments of the system modelled. This

is a computationally intensive task that requires considerable CPU time for large systems. To this end, all Gillespie simulations executed for this thesis employ a description of the studied system in SBML [101], which is then simulated using CudaSim [227], a package which runs biochemical models on NVIDIA CUDA GPUs in a highly parallelised manner.

2.2.4.3 *Gene Expression: An Example*

To give an example of how to use the Gillespie algorithm, take the chemical reactions describing gene expression in Equation 2.2. There are 3 states, yielding the state vector $X = [X_G, X_M, X_P]'$, where the subscript refers to the chemical species and $'$ is the transpose, and four reactions. The reactant and product matrices are:

$$a = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

The stoichiometry matrix is then:

$$S = b - a = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix},$$

where the columns refer to the four reactions, the first being transcription, producing mRNA M. The second is translation, producing protein P, and the third and fourth are degradation of M and P respectively. As can be seen in a and b, the gene plays a role as a reactant and product, but its numbers do not change as can be seen in the first row of S. The initial state vector must include $X_G > 0$ to ensure that the states do not decay to zero, but the two other states just need to be in $\mathbb{Z}_{\geqslant 0}$. The choice of starting time and end time is also arbitrary but $t_0 = 0$ is standard.

2.3 ANALYSIS AND DESIGN TECHNIQUES

This section presents classical control theoretical tools and methods applied to the biological systems designed and studied in this project. The areas covered are the state space representation of a system, linearisation, Laplace transforms, and transfer functions.

2.3.1 *Non-Linear Input/Output System*

The inner workings of a system are described by its states. In addition to states, real systems also have inputs and outputs. The inputs are the means through which the system can be influenced externally, through the experimental addition of inducer for example. Outputs of a system can be measured. In a biological system the output is often a proxy for protein concentration through the use of a reporter gene such as sfGFP.

With inputs and outputs, Equation 2.3 becomes the standard state space representation

$$\begin{aligned} \dot{x} &= f(x,u), \\ y &= g(x,u), \end{aligned} \tag{2.13}$$

where as above, $(\dot{-})$ refers to the first derivative with respect to time, the state $x \in \mathbb{R}^N_{\geqslant 0}$, where N is the number of states, $u \in \mathbb{R}^M$ is the input to the system, where M is the dimension of the input, $y \in \mathbb{R}^L$ is the output of the system with dimension L, and $f(x,u)$ and $g(x,u)$ are nonlinear functions of both the state and input.

The above is a non-linear system, which is difficult to deal with analytically. To facilitate the analysis of such systems, it is standard to linearise.

2.3.2 *Linearisation*

This section describes the method of linearisation of a system around a point of interest. Usually, an equilibrium of the system is chosen as this simplifies the form of the linearisation. Taking a generic non-linear system, consider Equation 2.3. A Taylor series expansion of Equation 2.3 yields the following:

$$\begin{aligned} \dot{x} &= f(x^*) + \frac{f'(x^*)}{1!}(x-x^*) + \frac{f''(x^*)}{2!}(x-x^*)^2 + \dots \\ &= \sum_{i=0}^{\infty} \frac{f^{(i)}}{i!}(x-x^*)^i, \end{aligned}$$

where $x^* \in \mathbb{R}^N_{\geqslant 0}$, in the case of a biochemical system, is the point of interest chosen, $f^{(i)}(x^*)$ defines the i^{th} derivative w.r.t. x evaluated at x^*. Finally, $i!$ denotes the factorial of i. When linearising, this series is truncated at the linear term i.e.

$$\begin{aligned}\dot{x} &\approx f(x^*) + f'(x^*)(x - x^*) \\ &= f(x^*) + \left.\frac{df}{dx}\right|_{x=x^*} (x - x^*),\end{aligned}$$

where we have written the so called Jacobian, $\left.\frac{df}{dx}\right|_{x=x^*}$, more explicitly in the second line. Setting

$$\mathbf{A} = \left.\frac{df}{dx}\right|_{x=x^*}$$

and performing the transformation $\bar{x} = x - x^*$, noting that if x^* is chosen to be at the equilibrium $f(x^*) = 0$, yields

$$\dot{\bar{x}} = \mathbf{A}\bar{x},$$

which is the linearised system where $\bar{x} \in \mathbb{R}^N$ is a small perturbation around a point of interest and is a matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$.

2.3.3 *Linear State-Space Representation*

Taking the full non-linear input/output system described in Equation 2.13, this can now be linearised using the above method, both with respect to the states x and inputs u. The linearised full state-space system is

$$\dot{\bar{x}} = \mathbf{A}\bar{x} + \mathbf{B}\bar{u}, \tag{2.14a}$$

$$\bar{y} = \mathbf{C}\bar{x} + \mathbf{D}\bar{u}, \tag{2.14b}$$

where $\bar{x}$, $\bar{y}$ and $\bar{u}$ indicate perturbations to the equilibrium state, output and input respectively, $\mathbf{A} \in \mathbb{R}^{N \times N}$ is called the state or system matrix, $\mathbf{B} \in \mathbb{R}^{N \times M}$ is called the input matrix, $\mathbf{C} \in \mathbb{R}^{L \times N}$

is called the output matrix and $\mathbf{D} \in \mathbb{R}^{L\times M}$ is called the feedthrough matrix. The state-space representation and the matrices above have the following definitions:

$$\mathbf{A} = \left.\frac{df}{dx}\right|_{x=x^*,u=u^*}, \quad \mathbf{B} = \left.\frac{df}{du}\right|_{x=x^*,u=u^*}$$
$$\mathbf{C} = \left.\frac{dg}{dx}\right|_{x=x^*,u=u^*}, \quad \mathbf{D} = \left.\frac{dg}{du}\right|_{x=x^*,u=u^*}$$

where u^* is the input at the equilibrium x^*.

For all systems considered in this project $\mathbf{D} = 0$ as there is no direct link between the input and output. Another way to think of this is, that there is always a time delay between a change in input and its effect on the output. The output matrix $\mathbf{C}$ usually has a very simple form picking one of the states or summing a few of them as the output. The bar notation for small perturbations is commonly also dropped to simplify notation and the symbols x, y and u must therefore be considered in context to either mean the full variable, in the case of the full non-linear system, or to mean a perturbation on the equilibrium, in the case of the linearised system.

2.3.4 *Laplace Transforms*

Laplace Transforms are a standard mathematical tool that facilitate the analysis and design of systems of linear ODEs, by putting the equations on a form that can be manipulated algebraically as opposed to by means of calculus. This is done by transforming the system from the time domain into the frequency domain. This method is widely used in engineering and a formulation of a framework in which this method can be applied to analyse and design biological systems is presented in Chapter 4. For more information any standard engineering textbook, such as [119], is recommended. Here the transform is introduced and some standard results that are used in the thesis, are presented.

The definition of the Laplace Transform is as follows:

$$F(s) = \mathcal{L}(f(t)) = \int_0^\infty e^{-st} f(t)dt$$

where $s \in \mathbb{C}$ is the frequency domain variable, $\mathbb{C}$ is the set of complex numbers and $f(t)$ is the function to be transformed. Note that the transformed function is usually denoted by its capital, such as $F(s)$ in this case. The transform is linear, which is to say that

$$\mathcal{L}(af(t) + bg(t)) = aF(s) + bG(s).$$

Another result that will be used regularly is the Laplace transform of the derivative of a function:

$$\mathcal{L}\left(\frac{df(t)}{dt}\right) = sF(s) - f(0).$$

As we generally assume the system is relaxed at $t = 0$, i.e. at equilibrium, and therefore $f(0) = 0$ and the above equation can be simplified by dropping the second term.

2.3.5 *The Generalised Transfer Function*

The transfer function of a system describes the relationship between its inputs and outputs in the frequency domain. In this section, how to calculate the generalised transfer function of the state-space representation is demonstrated. Starting with Equation 2.14, but ignoring the bar notation for clarity and setting $\mathbf{D} = 0$, the Laplace transform is applied to obtain

$$\begin{aligned} sX(s) &= \mathbf{A}X(s) + \mathbf{B}U(s), \\ Y(s) &= \mathbf{C}X(s), \end{aligned}$$

where $X(s)$, $U(s)$ and $Y(s)$ are the Laplace transforms of $x(t)$, $u(t)$ and $y(t)$ respectively. Re-arranging the first equation in terms of $X(s)$ and substituting into the second yields

$$Y(s) = \mathbf{C}(s\mathbb{I} - \mathbf{A})^{-1}\mathbf{B}U(s)$$

where $\mathbb{I}$ is the $N \times N$ identity matrix. This can be written on the form $Y(s) = H(s)U(s)$, where $H(s)$ is called the transfer function and relates the input signal to the output signal at a specific frequency. In this case the transfer function is

$$H(s) = \mathbf{C}(s\mathbb{I} - \mathbf{A})^{-1}\mathbf{B}.$$

In Chapter 4, the transfer functions are calculated for a number of biological systems and is used to aid the design process as it facilitates characterisation of input output properties of a system.

2.4 CONCLUSION

This chapter provides an introduction to the main concepts required to understand this text. The focus is initially on the description of biological systems, beginning with the Central Dogma of Molecular Biology, which is key to the systems studied in this project. Bacteria and relevant bacterial mechanisms of expression and regulation are presented as these are the specific systems relevant to this work. Throughout there is a focus on how to design and construct the GRNs by combining biological parts. Finally, the chapter covers first modelling and simulation techniques, then analytical methods useful to the study of ODE models.

3 REGULATION BY SRNAS

Regulation by Small Ribonucleic Acids (sRNAs), presented in Section 2.1.15, fulfils a large number of complex and intricate roles in expression regulation of a cell. Due to their flexibility and low energetic burden, sRNAs present a potentially powerful tool to synthetic biologists. To be able to design systems employing sRNAs, their behaviour must be well understood. For a large subset of naturally occurring sRNAs, the Host-Factor of Bacteriophage Qβ (Hfq), a bacterial chaperone protein, plays a key role in their function [4, 32, 53, 88, 198]. In particular, the sRNA inhibition mechanism has been researched in depth. In the case of this mechanism, there is good agreement on a dynamic model of the system [103, 127–129, 142, 190] and it has been experimentally verified to some extent [127, 129]. The sRNA activation mechanism, though well documented from a biochemical perspective [73, 196], is somewhat less quantitatively studied [103].

The question posed in this chapter is whether the inclusion of Hfq in the modelling of sRNA regulation is important when designing a circuit. The system studied is that of a single sRNA-target mRNA pair.

Initially, basic models of both sRNA inhibition and activation are studied in depth using ODEs. This yields the classic threshold-linear behaviour expected in sRNA regulation. Novel results are presented through stochastic simulation of the regulation mechanism employing the Gillespie Algorithm [81]. The stochastic data provides the noise profile of the system through the extraction of moments. This information is of interest to designers and gives the possibility of experimental verification using methods such as those employed in [62]. In these basic stochastic models, noise is expected to peak just off the threshold. In addition to this, the effect of separating the genes of the target mRNA and the sRNA by encoding them on separate plasmids, such as is done for the experimental set-up in [129], is studied. This is shown to increase and broaden the noise peak at the threshold substantially.

A biochemical model is then presented including the interactions between the RNA species and Hfq. This model is referred to as 'the symmetric model' because sRNA and mRNA interact with Hfq in a symmetric manner, and is studied in the context of varying transcription rates of the RNA species, presenting steady state data over a large range of parameters. The study of the symmetric model predicts that the mRNA behaves in a non-linear manner above the threshold.

A novel extension to the model, including the assumption that mRNA is degraded rapidly while in complex with Hfq, is shown to correct this non-linear behaviour. As a result of adding two chemical reactions to the network, the 'extended model' exhibits the expected threshold-linear behaviour of the concentration of free mRNA, as is the case in the basic model.

A further novel assumption is included: for synthetic circuits which only employ a natural scaffold for the design of the sRNA, such as in [134, 151], there is a binding order preference biased towards Hfq binding the sRNA first. This is because the mRNA is not intentionally designed with a sequence known to bind to Hfq. The result of this greatly increases the range of transcription rates over which regulation is efficient.

Stochastic simulations of the models including Hfq, both with and without plasmid copy number variation, are performed. Results show that when target and sRNA are on the same plasmid stochastic modelling with Hfq is probably not necessary, when designing a larger system, though the peak in noise seen in the basic model is greatly reduced. If the genes are split across two plasmids, the inclusion of Hfq is important to the behaviour of the system, both increasing noise and greatly softening the transition of the threshold-linear response.

Finally, time courses comparing the basic model of sRNA regulation to the extended model including Hfq are presented. These show that the dynamic behaviour of the two models is very similar in a regime where regulation is efficient and Hfq is neither scarce nor too abundant.

It is concluded that, in the design of a circuit employing a single sRNA-target mRNA pair, it is not required to include Hfq and its interactions when ODE models are employed. It is not as clear cut in the case of stochastic models, as it depends on the experimental set-up of the system.

3.1 BASIC MODELS OF SRNA REGULATION

The mechanisms by which sRNAs regulate translation are described in detail in Section 2.1.15. The mechanisms focussed on here are those of translation initiation control. In the case of translation inhibition, the sRNA regulates by covering the Shine-Dalgarno and/or Start Codon of a target mRNA. The translation activation mechanism functions occurs through an sRNA providing access to the Shine-Dalgarno and Start Codon of a target mRNA.

3.1.1 *sRNA Inhibition*

The standard chemical reaction model of sRNA inhibition found in the literature [103, 127, 129, 142, 190] is a simple three state model. The states are the sRNA, target mRNA, and the protein expressed through translation of the target. The chemical reactions included in the model are as follows:

$$\begin{aligned} M &\xrightarrow{\alpha_P} M + P, & P &\xrightarrow{\delta_P} \emptyset, \\ G_M &\xrightarrow{\alpha_M} G_M + M, & M &\xrightarrow{\delta_M} \emptyset, \\ G_R &\xrightarrow{\alpha_R} G_R + R, & R &\xrightarrow{\delta_R} \emptyset, \\ M + R &\xrightarrow{K} \emptyset, \end{aligned} \tag{3.1}$$

where P is the expressed protein, M is the target mRNA, R is the sRNA, G_M is the gene that encodes the protein and therefore also the target mRNA, G_R is the gene that encodes the sRNA. The constants α_P, α_M and α_R are the rates of the translation of protein, and the transcription of target mRNA and sRNA respectively. The degradation rate of the protein, target mRNA and sRNA are given by δ_P, δ_M and δ_R respectively. The parameter K is the rate of the sRNA regulation interaction between target mRNA and sRNA, also called the rate of heteroduplex formation.

The model assumes that the association of sRNA to target mRNA is irreversible. This is reasonable as the complementarity of the sRNA and target is often very high, and in synthetic cases, usually complete, ensuring tight binding between the two species, with a half-life much longer than the usual turn-over rate of RNA. As the duplex formation reaction is irreversible, the quantity of the heteroduplex in the system is not of interest. This interaction can therefore be thought of as a tuneable degradation rate, dependent on both species involved. This provides a sort of comparison between the number of the two species present.

The above chemical reactions can be turned into a simple ODE model of the system:

$$\begin{aligned} \dot{p} &= \alpha_P m - \delta_P p, \\ \dot{m} &= \alpha_M g_M - \delta_M m - K m r, \\ \dot{r} &= \alpha_R g_R - \delta_R r - K m r, \end{aligned} \tag{3.2}$$

where $\dot{-}$ signifies the derivative with respect to time, the lower case letters p, m, r, g_M and g_M signify the concentrations of the species denoted by the upper case, in this case, the protein,

mRNA, sRNA, mRNA gene and sRNA gene respectively. The parameters signify the rates of the reactions, as in Equation 3.1.

From these ODEs, it is simple to extract the steady state concentrations by setting the derivatives on the left hand side of Equation 3.2 to zero and solving for the individual states p, m and r.

$$p^{ss} = \frac{\alpha_P}{\delta_P} m^{ss}, \tag{3.3a}$$

$$m^{ss} = \frac{B_M}{2K\delta_M}\left[\sqrt{1+\frac{4K\alpha_M g_M \delta_R \delta_M}{B_M^2}} - 1\right], \tag{3.3b}$$

$$r^{ss} = \frac{B_R}{2K\delta_R}\left[\sqrt{1+\frac{4K\alpha_R g_R \delta_R \delta_M}{B_R^2}} - 1\right], \tag{3.3c}$$

where $B_M = K(\alpha_M g_M - \alpha_R g_R) - \delta_R\delta_M$ and $B_R = K(\alpha_R g_R - \alpha_M g_M) - \delta_R\delta_M$ and the superscript ss denotes steady state. Note that Equation 3.3b and Equation 3.3c are symmetric, as are the related ODEs in Equation 3.2. The steady state protein concentration, p^{ss}, is simply proportional to m^{ss}.

3.1.1.1 *Notes on Parameter Values*

The set of biologically feasible parameters used to simulate and study the system are shown in Table 3.1. These have all been carefully selected with guidance from the literature. Employing the above mentioned notation, when simulating G_M and G_R are plasmid copy numbers associated with pBR322 *ori*-region, which has a copy number of 15-20 per cell. Therefore the associated concentrations are $g_M = g_R = 28.2$nM. This value assumes a gene copy number of 17 in a cellular volume of 10^{-15}l. The important point is that the genes encoding target mRNA and sRNA are in single copies on the same plasmid, or on the chromosome, ensuring that the copy number of each gene can be assumed to be the same.

Commenting on the remaining parameter values, the degradation rate of mRNA, δ_M, represents a half-life of about 2.8 minutes, in line with the literature [87]. Turn over of protein and sRNA are on the scale of the doubling time of the cell. For ease, this has been set to 28 minutes. In exponential growth, this is a reasonable estimate for the doubling time in rich media. The reason for this is that protein is assumed to only be diluted and the sRNAs are protected in the cell by Hfq.

Table 3.1 – Parameters used to model the sRNA inhibition mechanism detailed in Equation 3.1.

Param.	Value	Param.	Value
α_P	15min^{-1}	α_R	0 to 30min^{-1}
δ_P	0.025min^{-1}	δ_R	0.025min^{-1}
α_M	0 to 30 min^{-1}	g_M	28.2nM
δ_M	0.25min^{-1}	g_R	28.2nM
K	0.005 to $5\text{nM}^{-1}\text{min}^{-1}$		

3.1.1.2 *The Threshold-Linear Response*

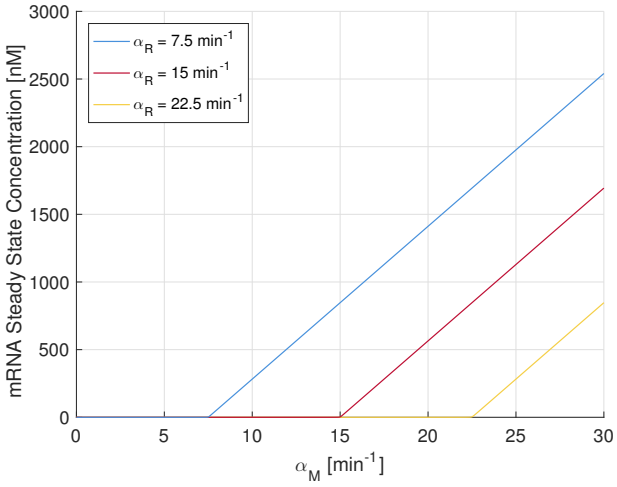
The steady state behaviour of the system can be studied using the functions in Equation 3.3. This yields the well documented threshold-linear response of the steady state exhibited by the sRNA inhibition mechanism [103, 127–129, 142, 190].

The steady state of the system is calculated having selected three values for α_R, sweeping α_M from 0min^{-1} to 30min^{-1} and otherwise using the parameters given in Table 3.1. The results in Figure 3.1a show the steady state behaviour of the target mRNA. The steady state of the protein behaves in an identical manner, but rescaled by α_P/δ_P as seen in Equation 3.3a. Figure 3.1b shows the steady state of the sRNA.

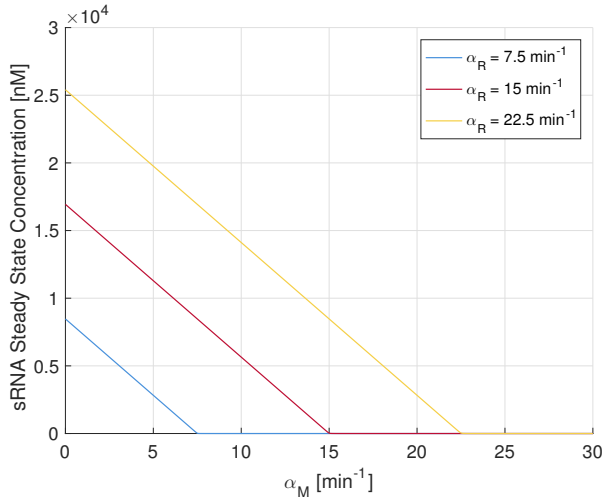
The system operates in two regimes dependent on the transcription rates of the sRNA and target mRNA. When $\alpha_R > \alpha_M$, there is an excess of sRNA in the system and virtually no free target mRNA, and hence hardly any expression of protein. In this state the gene is inhibited or silenced. In the other regime, when $\alpha_R < \alpha_M$, there is an excess of mRNA and translation occurs. A key feature of this mechanism, seen in both Figure 3.1a and Figure 3.1b, is that experimentally both the rates of transcription of the target mRNA, α_M and the sRNA, α_R can be controlled through transcription regulation mechanisms. This allows the threshold at which $\alpha_M = \alpha_R$ to be shifted dynamically as a subunit of a larger system.

Another important parameter of the system is the rate of heteroduplex formation, K. The effect of varying this is shown in Figure 3.1c. As can be seen, K influences how sharp the transition at the threshold is. As K becomes smaller, describing less and less efficient formation of the heteroduplex, the threshold becomes less smooth. This might be due to low complementarity between the sRNA and target, secondary structure of the RNAs, or the effect of interactions with the chaperone protein Hfq.

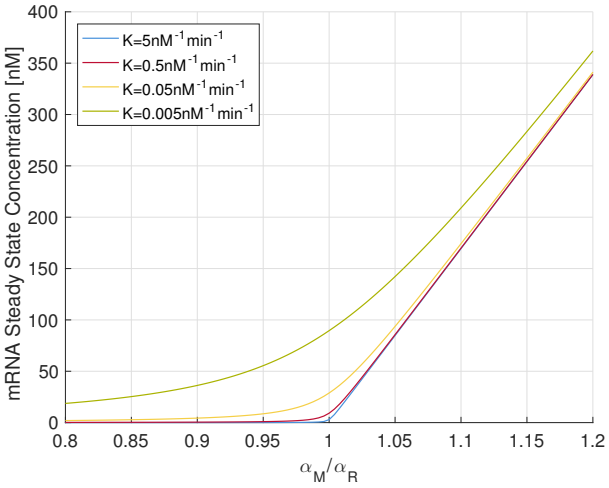
The threshold-linear response is confirmed experimentally in [129] and [103]. Levine *et al.* in [129] construct a synthetic system using RhyB, an sRNA endogenous to *E. coli*, that regulates the iron homeostasis. Its strongest interaction is with the target gene *sodB*. The 5′ UTR and the first 11 codons of *sodB* is encoded upstream of the reporter GFP on a plasmid. This



(a) Steady State of mRNA



(b) Steady state of sRNA



(c) Varying Rate of Heteroduplex Formation

Figure 3.1 – The threshold-linear behaviour of the steady state observed in sRNA inhibition. See the text for discussion.

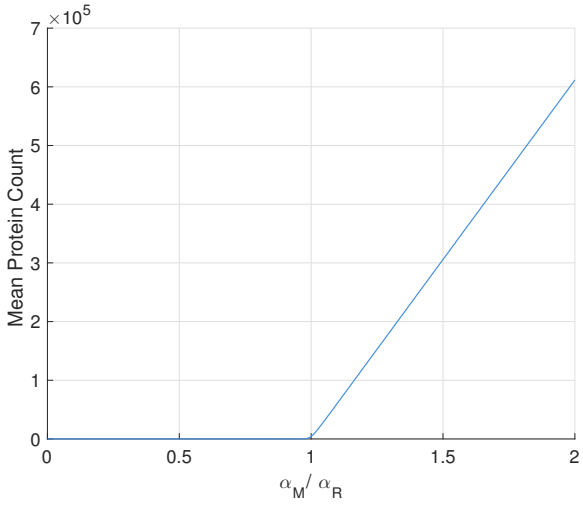
fused target gene is expressed under control of the *lac*-promoter, which allows the scientists to control the expression rate of the target gene though addition of the inducer Isopropyl-β-D-1-Thiogalactopyranoside (IPTG). Relating this to the model in Equation 3.1, α_M is varied experimentally through the concentration of IPTG. The sRNA RhyB is both expressed from the chromosome and via a separate plasmid vector to that of the fused target gene. On the plasmid vector, expression of the sRNA is under the control of the *tet*-promoter and induced with Anhydrotetracycline (aTc). The dose response data presented in this paper shows evidence of a smooth transition at the threshold in line with a low K value.

Hussein and Lim in [103] implement a similar system, placing expression of the sRNA MicC under the control of the *lac*-promoter. Expression is induced by IPTG. The synthetic target uses the natural scaffold of *ompC*, including the 5′ UTR and a few of the first codons of the gene. This is fused to GFP, which acts as a reporter, to create the target. The target is expressed under control of the *tet*-promoter with no TetR present in the system. In this case, the target and sRNA were encoded on the same plasmid. The data for this system shows a sharp transition from expressing to silenced, indicating a high K value.

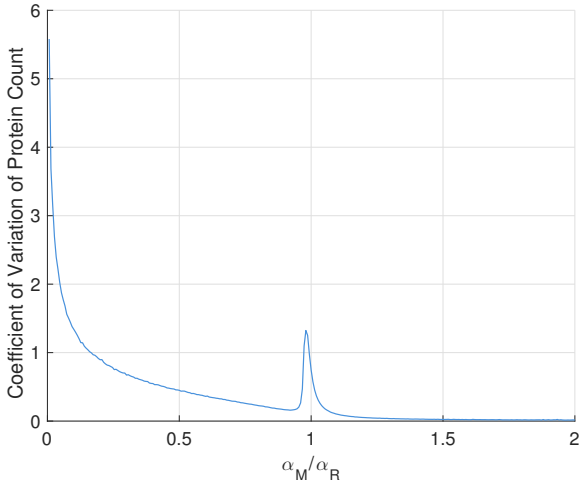
3.1.1.3 *Gillespie Simulation of sRNA Inhibition*

Using the reactions described in Equation 3.1, the system is simulated using the Gillespie Algorithm, detailed in Section 2.2.4. With data from such a simulation, the noise spectrum of the system at steady state is extracted. This is comparable to experimental data gathered through techniques such a flow cytometry. Through the use of a reporter protein, such as sfGFP, in conjunction with flow cytometry, the fluorescence of each individual cell in a population can be determined. The fluorescence data from each cell provides the distribution of fluorescence across a population from which the mean and variance can be extracted. Assuming the fluorescence is proportional to the number of proteins in a cell, this experimental data can then be compared with data extracted from stochastic simulations, as has been done for other systems in [62]. This interpretation treats each simulation in an ensemble as a representation of the chemical reactions that occur in a single cell.

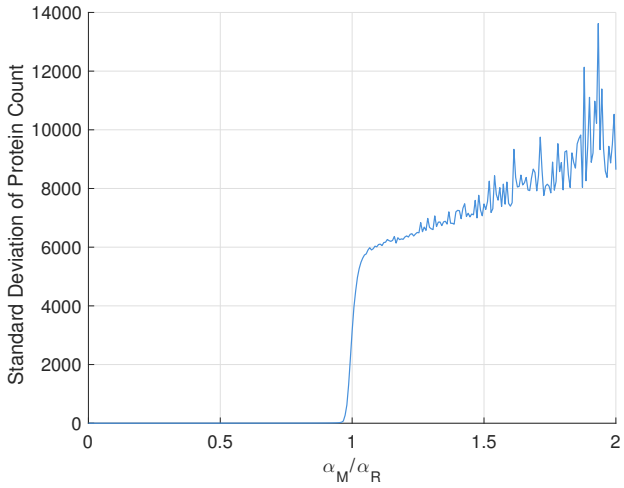
The results presented in Figure 3.2, Figure 3.3 and Figure 3.4a are from end point data of an ensemble of 10000 simulations. The simulations each run long enough, till $t = 300\text{min}$, to ensure that the end point behaviour represents the steady state behaviour of the system. The parameters used are the stochastic equivalent of the parameters presented in Table 3.1. The gene copy numbers $G_M = G_R = 17$ are used and the rate of heteroduplex formation is corrected to $K = 8.3\text{min}^{-1}$ to ensure that it has the correct value and units, assuming that the reactions



(a)



(b)



(c)

Figure 3.2 – Protein species number data for the sRNA inhibition model specified by Equation 3.1. See the text for discussion.

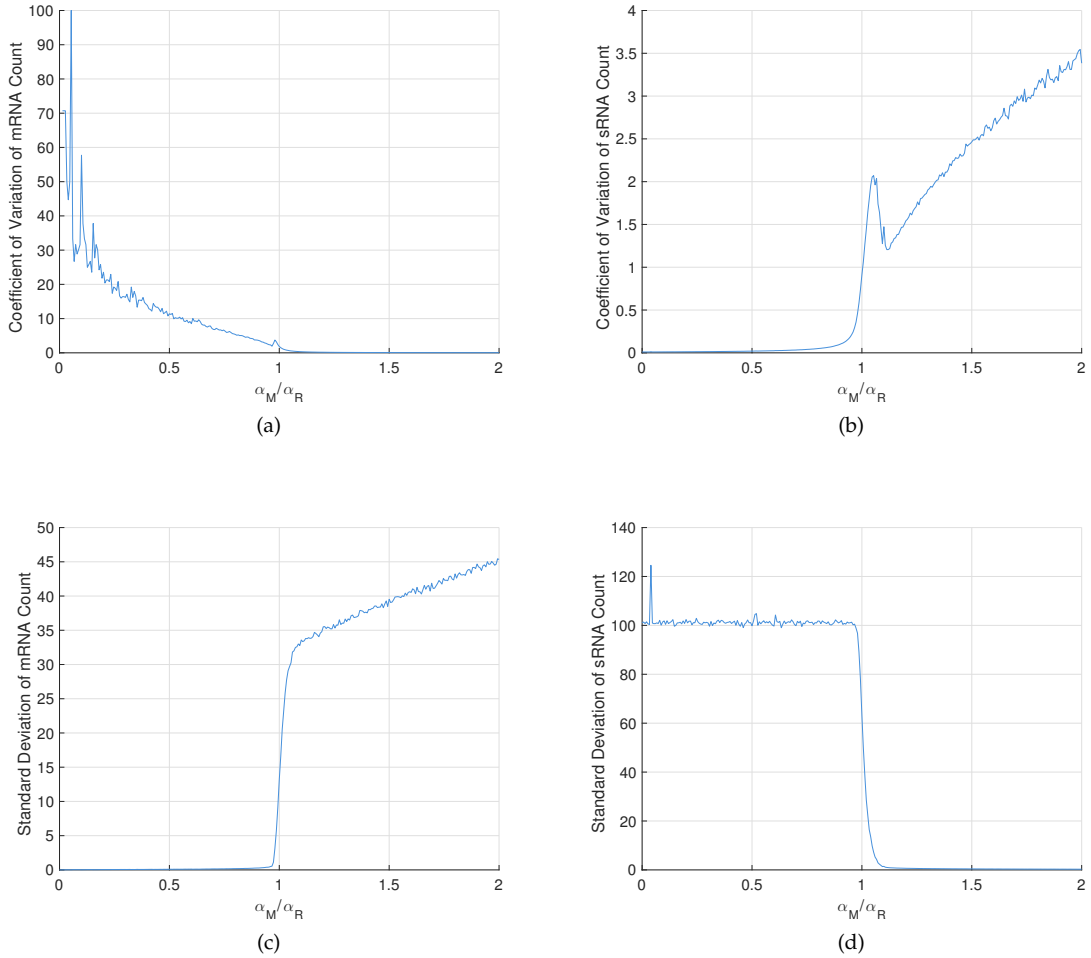
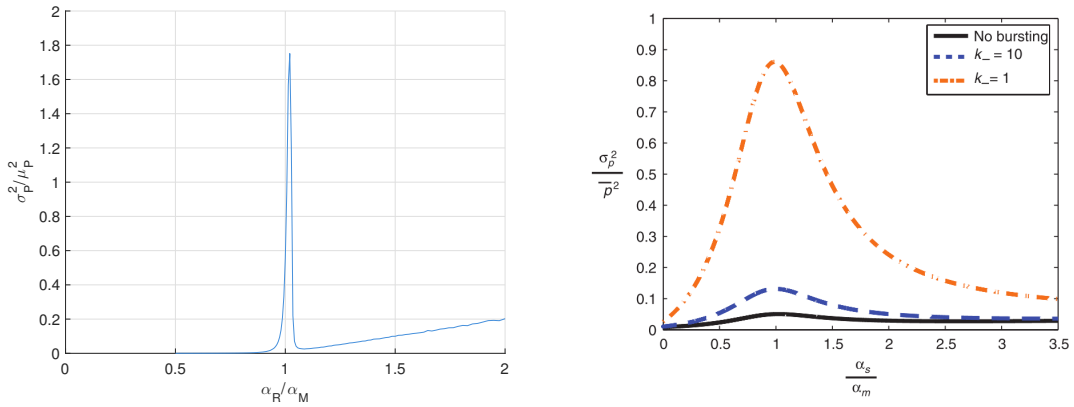


Figure 3.3 – mRNA and sRNA species number data from stochastic simulation of sRNA inhibition detailed in Equation 3.1. See the text for discussion.



(a) Gillespie Simulation

(b) From [142]

Figure 3.4 – This compares the result from the Gillespie simulation performed here with the result from [142]. The peak in noise predicted by the Gillespie simulation is narrow and just off the threshold, whereas the peak predicted in [142] using stochastic differential equations is broad and on the threshold. The parameter α_s in (b) is the same as α_R in (a), $\bar{p}$ in (b) is used to denote the mean protein concentration and k in (b) denotes the strength of the modelled bursting mechanism. Note the same metric of protein has been presented, namely the variance over the mean squared, and that the X-axis has been flipped in comparison to Figure 3.2 and Figure 3.3.

occur in a cell of volume 10^{-15}l. The transcription rate of sRNA is constant with the value $\alpha_R = 15\text{min}^{-1}$ and the transcription rate of target mRNA is swept from 0min^{-1} to 30min^{-1}. This is an identical set up to that used in the case of the ODE model above, albeit for the stochastic system.

The protein species number data presented in Figure 3.2a shows the expected threshold-linear response that is expected for this system as a function of the ratio between the transcription rates of target mRNA and sRNA, α_M/α_R. The equivalent mRNA and sRNA results are identical, apart for the change of units, to those in Figure 3.1. Defining the mean as μ and the standard deviation as σ, the coefficient of variation of the number of proteins, defined as σ/μ, is presented in Figure 3.2b. The standard deviation of the number of proteins is reported in Figure 3.2c. The coefficient of variation and standard deviations of the target mRNA and sRNA are presented in Figure 3.3.

The coefficient of variation of the protein 3.2b, mRNA 3.3a and sRNA 3.3b show an increase in noise just off the threshold on the side where the species quantity approaches zero. In the case of protein, when α_M/α_R is just under 1. This can be thought of as the system moving through the noisy boundary of a phase transition. This does not agree with the result in [142] as presented in Figure 3.4, where analysis is performed on an Stochastic Differential Equation (SDE) model of the sRNA inhibition mechanism including bursting. The bursting mechanism models the repeated translation by the cellular machinery of a single mRNA. The peak in noise predicted by Mehta *et al.* in [142] is exactly at the threshold and is broad, whereas the peak predicted by the stochastic

simulations performed here expect a much narrower peak just off the threshold, as shown in Figure 3.4a.

These two differing results could be checked experimentally following protocols checking the dose responses to induction of this system such as executed in [62]. The exact systems studied in [102, 129] could be used to do this. Flow cytometry data would be required to check how the coefficient of variation behaves around the threshold, and because the peak in noise predicted by the Gillespie simulations is so narrow, it might go undetected. In addition to this, due to low quantities of expression in the output protein around and below the threshold, the experiment might prove challenging. If the experimental system behaves as predicted by the model in [142] including the bursting mechanism on the other hand, a broad peak in noise should be visible in the data.

3.1.2 *sRNA Activation*

The sRNA activation mechanism, though well documented from a biochemical perspective [73, 196, 217], is the less studied of the two sRNA regulation mechanisms from a quantitative perspective. In [103], Hussein and Lim published both theoretical and experimental results documenting its behaviour at steady state.

A simple chemical reaction model of sRNA activation equivalent to that of sRNA inhibition in Equation 3.1 and also used in [103] is as follows:

$$\begin{aligned} \mathrm{MR} &\xrightarrow{\alpha_P} \mathrm{MR} + \mathrm{P}, & \mathrm{P} &\xrightarrow{\delta_P} \emptyset, \\ \mathrm{G_M} &\xrightarrow{\alpha_M} \mathrm{G_M} + \mathrm{M}, & \mathrm{M} &\xrightarrow{\delta_M} \emptyset, \\ \mathrm{G_R} &\xrightarrow{\alpha_R} \mathrm{G_R} + \mathrm{R}, & \mathrm{R} &\xrightarrow{\delta_R} \emptyset, \\ \mathrm{M} + \mathrm{R} &\xrightarrow{K} \mathrm{MR}, & \mathrm{MR} &\xrightarrow{\delta_{MR}} \emptyset, \end{aligned} \tag{3.4}$$

where the symbols have the same significance as in Equation 3.1 with the addition of MR which denotes the heteroduplex between the sRNA and target mRNA and the degradation rate of the heteroduplex, which is denoted by δ_{MR}.

The key difference between these two models is that the protein is now translated from the heteroduplex MR, rather than the free mRNA. This requires that the heteroduplex is included as a state. Again the heteroduplex formation is treated as irreversible; once the sRNA is bound to its target, it does not release it again prior to being degraded.

It is worth noting that this model specifically describes the translational activation by an sRNA, as is described in Section 2.1.15.2. The stabilisation mechanism, whereby expression of a protein is increased by sRNAs stabilising the target mRNA and thereby prolonging its lifetime, requires a model where protein is translated from both free mRNA and the heteroduplex. The choice of mechanism is due to the fact that translation activation is better understood and documented than stabilisation, and so the design of systems employing it is more straight forward. In addition to this, regulation is tighter, giving the possibility of a 'off' state, when transcription rates of both mRNA and sRNA are close to zero. This is expected to be preferable in synthetic contexts.

As in the case of sRNA inhibition, the chemical reactions in Equation 3.4 yield an analogous ODE model of the system:

$$\dot{p} = \alpha_P m_R - \delta_P p, \tag{3.5a}$$

$$\dot{m} = \alpha_M g_M - \delta_M m - K m r, \tag{3.5b}$$

$$\dot{r} = \alpha_R g_R - \delta_R r - K m r, \tag{3.5c}$$

$$\dot{m}_R = K m r - \delta_{MR} m_R, \tag{3.5d}$$

where the symbols signify the same as they do in Equation 3.2, with the addition of m_R which is the concentration of the heteroduplex, and δ_{MR}, the degradation rate of the heteroduplex.

The steady state concentrations of the species included in the sRNA activation model can be found explicitly, as in the case of sRNA inhibition, by setting the left hand side of Equation 3.5 to equal zero. The resulting solutions are:

$$p^{ss} = \frac{\alpha_P}{\delta_P} m_R^{ss}, \tag{3.6a}$$

$$m^{ss} = \frac{B_M}{2K\delta_M}\left[\sqrt{1 + \frac{4K\alpha_M g_M \delta_R \delta_M}{B_M^2}} - 1\right], \tag{3.6b}$$

$$r^{ss} = \frac{B_R}{2K\delta_R}\left[\sqrt{1 + \frac{4K\alpha_R g_R \delta_R \delta_M}{B_R^2}} - 1\right], \tag{3.6c}$$

$$m_R^{ss} = \frac{K}{\delta_{MR}} m^{ss} r^{ss}, \tag{3.6d}$$

where $B_M = K(\alpha_M g_M - \alpha_R g_R) - \delta_R \delta_M$ and $B_R = K(\alpha_R g_R - \alpha_M g_M) - \delta_R \delta_M$ and the superscript ss again denotes steady state.

The study of the model uses the parameters presented in Table 3.2. The parameters are identical to those in Table 3.1 except for the addition of $\delta_{MR} = 0.25 \text{min}^{-1}$. This is the same

Table 3.2 – Parameters used to model the sRNA activation mechanism detailed in Equation 3.4.

Param.	Value	Param.	Value
α_P	15min^{-1}	α_R	0 to 30min^{-1}
δ_P	0.025min^{-1}	δ_R	0.025min^{-1}
α_M	0 to 30min^{-1}	δ_{MR}	0.25min^{-1}
δ_M	0.25min^{-1}	g_M	28.3nM
K	0.005 to $5\text{nM}^{-1}\text{min}^{-1}$	g_R	28.2nM

as the degradation rate of free mRNA and is a reasonable value assuming the sRNA facilitates access to the RBS, activating translation, but does not stabilise the mRNA.

The similarity between the inhibition and activation models is high. As the ODEs describing the mRNA concentration and the sRNA concentration are identical, so are the steady state solutions. Therefore the steady state behaviours of the concentrations of free mRNA and sRNA are identical to those presented in Figure 3.1.

The steady state behaviour of the concentration of heteroduplex is of interest as this gives insight into the steady state behaviour of the protein concentration as well, as this is proportional to that of the heteroduplex. The steady state of heteroduplex is shown in Figure 3.5a. Similar behaviour is reported experimentally in [103].

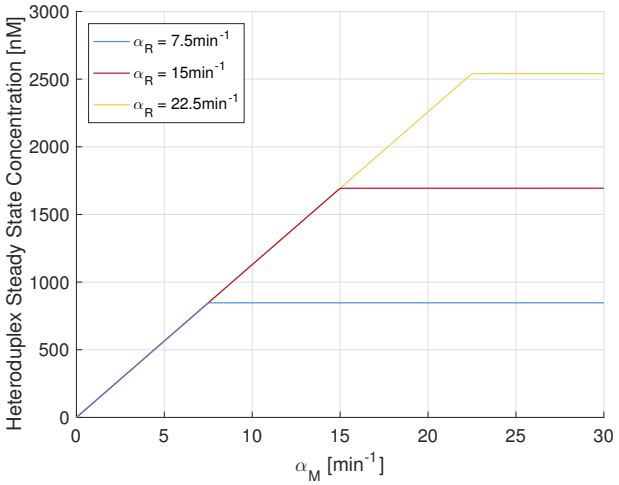
As can be seen in Figure 3.5a, there is a threshold at $\alpha_M = \alpha_R$. The threshold, in this case, decides the saturation point of the system, ensuring that protein cannot be expressed above a certain level. With a tight promoter controlling the rate of transcription of the sRNA ensuring that translation of protein does not occur when the sRNA is not induced, this would be a powerful tool in a synthetic circuit, by also providing an upper limit on the possible expression of protein.

Again, the effect of changing the rate of heteroduplex formation is presented in Figure 3.5b. As in the case of sRNA inhibition, decreasing K softens the transition at the threshold.

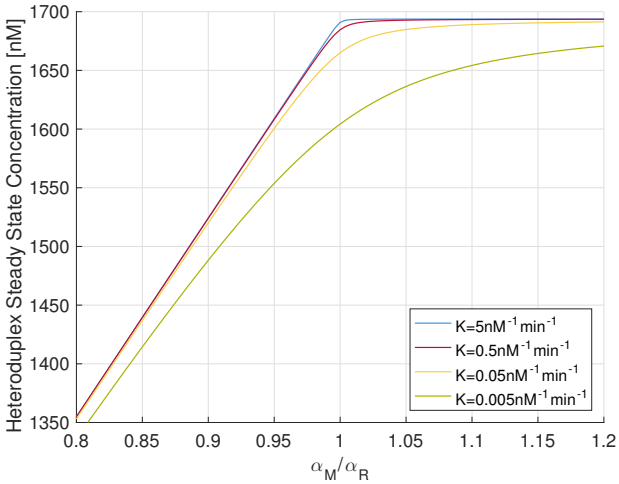
3.1.2.1 *Gillespie Simulation of sRNA Activation*

As in the case of sRNA inhibition, a stochastic ensemble simulation of the system is performed with 10000 individual runs using the parameters presented in Table 3.2. The parameters that need a change of units to be used for stochastic simulation are the gene copy numbers, that are set to $G_M = G_R = 17$ and the rate of heteroduplex formation which is corrected to $K = 8.3\text{min}^{-1}$. The merit of the stochastic simulation is that it yields a predicted noise profile of the system that can be compared to experimental data.

The data most easily comparable to that which could be extracted using flow cytometry to measure fluorescence of a reporter gene is presented in Figure 3.6. The mean protein number



(a) Steady State of the Heteroduplex



(b) Varying Rate of Heteroduplex Formation

Figure 3.5 – The threshold-linear behaviour of the steady state concentration of heteroduplex observed in sRNA activation. See the text for discussion.

has the expected threshold-linear profile found by solving the system's ODEs (see Figure 3.8a). A striking difference between these results and the sRNA inhibition case is that there is no spike in the coefficient of variation at the threshold $\alpha_M = \alpha_R$. The coefficient of variation of the protein count is also very low, staying below 2% for most of the data presented in Figure 3.8b.

The lack of noise peak at the threshold carries over to the coefficient of variation of the heteroduplex, as can be seen in Figure 3.7.

On the other hand, due to the dynamics being identical for the mRNA and sRNA in both the sRNA inhibition and activation models, the spike in the coefficient of variation is present as expected in the steady state data for both mRNA and sRNA (see Figure 3.8), and the qualitative behaviour of these plots is identical to that seen in Figure 3.3.

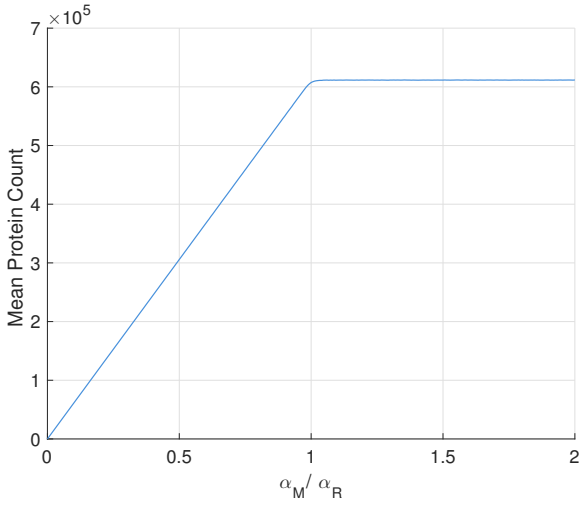
The fact that this system exhibits low noise in the expressed quantity of protein over such a wide operating range is attractive in the design of larger systems. The sRNA activation mechanism does not produce a noise peak around the threshold as the sRNA inhibition mechanism does. The peak in coefficient of variation produced by sRNA inhibition would need to be accounted for in a design process for a larger system and might potentially limit a system's operating range. A similar system, employing the sRNA activation mechanism, does not have the same issue.

3.1.3 *sRNA Regulation with Plasmid Copy Number Variation*

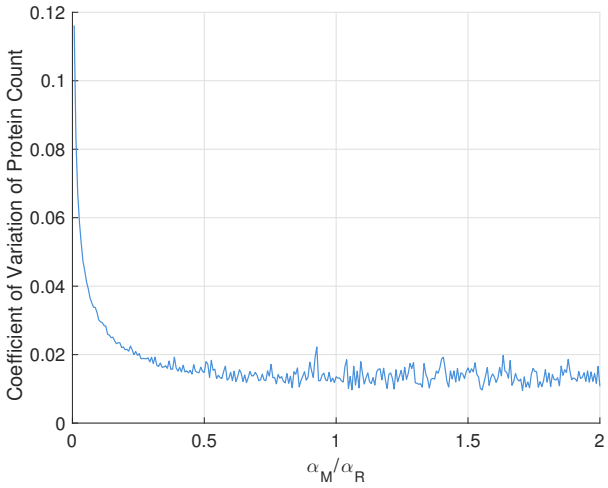
Plasmids are a standard tool used to implement GRNs in cells (see Section 2.1.12) and they play a role in the dynamics of the genetic circuits encoded on them. Plasmids are replicated in a stochastic manner throughout the cell cycle. The variation in copy number of a plasmid in a constant volume within a cell can be modelled as a simple birth-death process. The birth process models the replication within that volume and the death process models the dilution of the plasmids as cells grow and divide. This chemical reaction can be written as

$$\emptyset \xrightarrow{\alpha_G} G \xrightarrow{\delta_G} \emptyset, \tag{3.7}$$

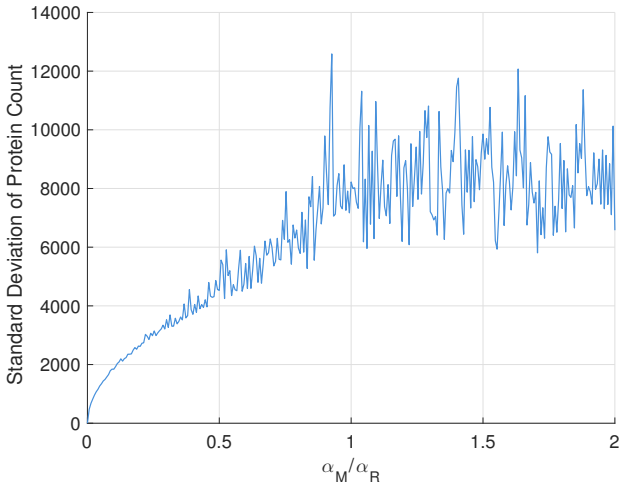
where G is the plasmid, α_G is the replication rate and δ_G is the degradation rate. When modelling this process, the degradation rate is set to equal that of a diluted protein, as active degradation does not occur and therefore the only contributing factor is dilution. Here this is chosen to be $\delta_G = 0.025\text{min}^{-1}$, which equates to a cellular doubling time of about 28min. To decide the replication rate, the average copy number of the plasmid is used. In the case of the commonly used *ori*-region pBR322 the copy number is between 15 and 20 [194, p.230]. Since the average copy



(a)

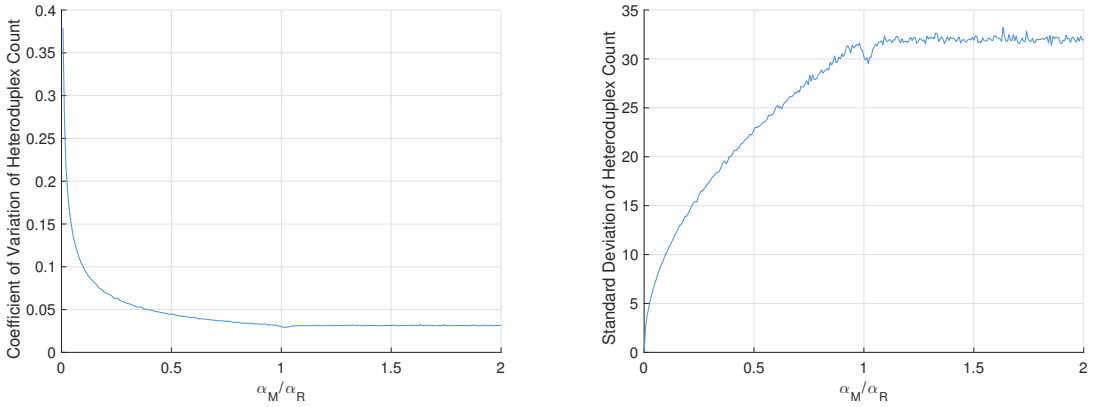


(b)



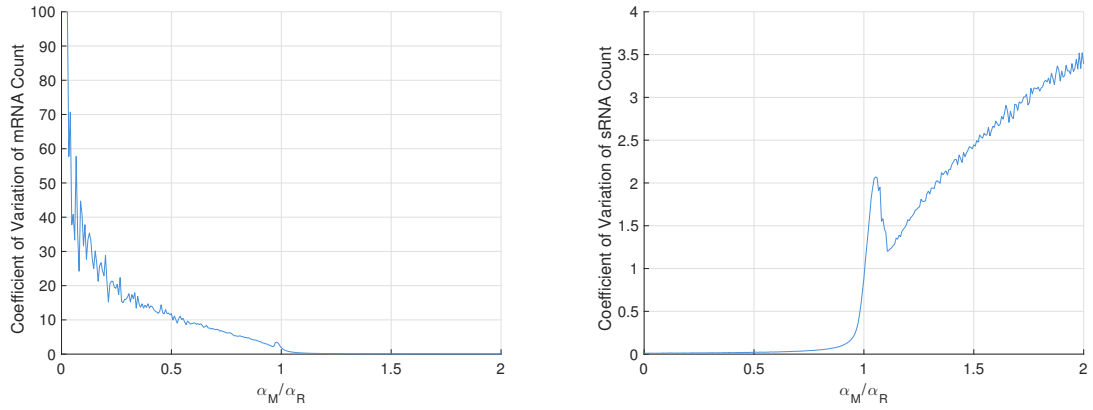
(c)

Figure 3.6 – Protein species number data for the sRNA activation model specified in (3.4). See the text for discussion.



(a) (b)

Figure 3.7 – The coefficient of variation and standard deviation of heteroduplex species number data for the sRNA activation model specified in (3.4). Note that there is no visible spike in the coefficient of variation, as in the case of the protein data, presented in Figure 3.6.



(a) (b)

Figure 3.8 – The coefficient of variation of the mRNA and sRNA species number data for the sRNA activation model specified in (3.4). Note that the behaviour is identical, within stochastic error, to that seen in Figure 3.3a and Figure 3.3b.

Table 3.3 – The parameters used to simulate both sRNA inhibition and activation including modelling the sRNA and target genes as being encoded on separate plasmids. Note that these parameters have the units necessary for a stochastic simulation using the Gillespie Algorithm.

Param.	Value	Param.	Value
α_P	5min^{-1}	α_{G_M}	0.4375min^{-1}
δ_P	0.025min^{-1}	δ_{G_M}	0.025min^{-1}
α_M	0 to 10min^{-1}	α_{G_R}	0.4375min^{-1}
δ_M	0.25min^{-1}	δ_{G_R}	0.025min^{-1}
α_R	5min^{-1}	K	8.3min^{-1}
δ_R	0.025min^{-1}	δ_{MR}	0.25min^{-1}

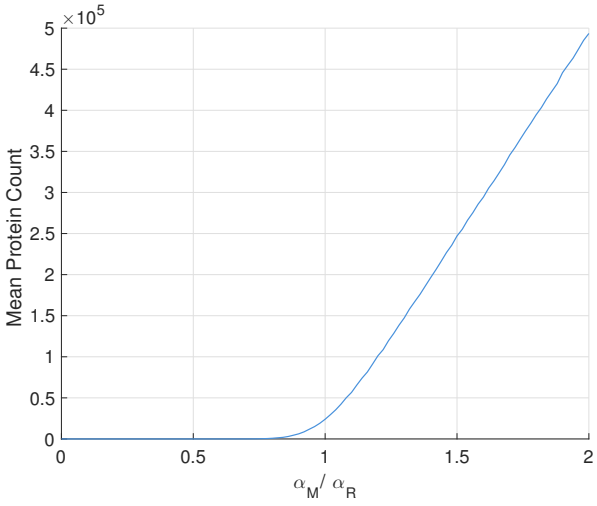
number of G in the above process is α_G/δ_G, and taking the average copy number of 17.5, the required replication rate can be calculated to be $\alpha_G = 0.4375\text{min}^{-1}$ for a stochastic simulation.

In the specific case of sRNA regulation, where both target and sRNA genes are encoded on the same plasmid, such as the systems used by Hussein and Lim in [103], the total number of plasmid vector varies according to Equation 3.7. Models of sRNA regulation, both inhibition and activation, including copy number variation of the plasmid displays qualitatively identical behaviour to the models without plasmid copy number variation. This is because, despite variation in the total plasmid copy number, the relative copy number of both the sRNA and target gene, and therefore the relative transcription rates of both genes, remains constant as they are physically on the same plasmid.

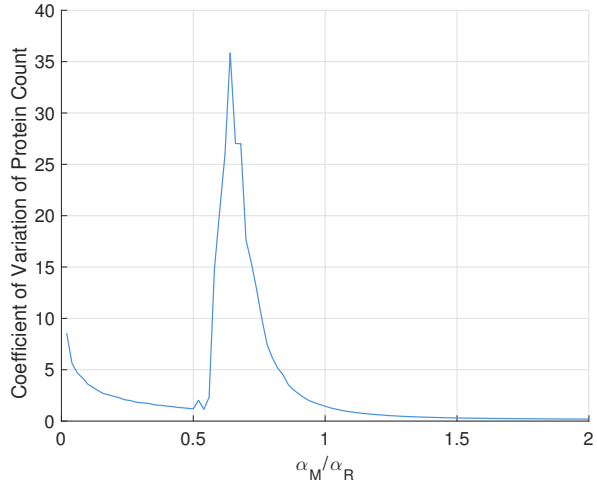
For purposes of modularity to facilitate the construction of a number of a different circuits, the target gene and the associated sRNA can be encoded on separate plasmids, such as is done by Levine *et al.* [129]. To include this in the models of sRNA regulation, the birth-death process of the two individual plasmids involved, and thereby the fluctuating gene copy numbers of target and sRNA, must be modelled. To expand the models of sRNA inhibition detailed in Equation 3.1 and sRNA activation detailed in Equation 3.4, all that is required is the inclusion of the following reactions:

$$\begin{aligned} \emptyset &\xrightarrow{\alpha_{G_M}} G_M \xrightarrow{\delta_{G_M}} \emptyset, \\ \emptyset &\xrightarrow{\alpha_{G_R}} G_R \xrightarrow{\delta_{G_R}} \emptyset, \end{aligned} \tag{3.8}$$

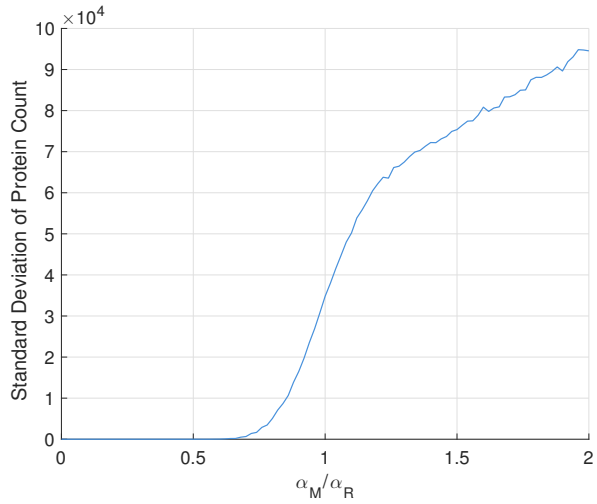
where G_M and G_M represent the plasmid carrying the target gene and the sRNA gene respectively, α_{G_M} and α_{G_R} are the replication rates of the target gene plasmid and the sRNA gene plasmid respectively, and δ_{G_M} and δ_{G_R} are the degradation rates of the target gene plasmid and the sRNA gene plasmid respectively.



(a)



(b)



(c)

Figure 3.9 – Protein count data from stochastic simulation of sRNA inhibition including modelling the target and sRNA genes on separate plasmids. The parameters used are presented in Table 3.3.

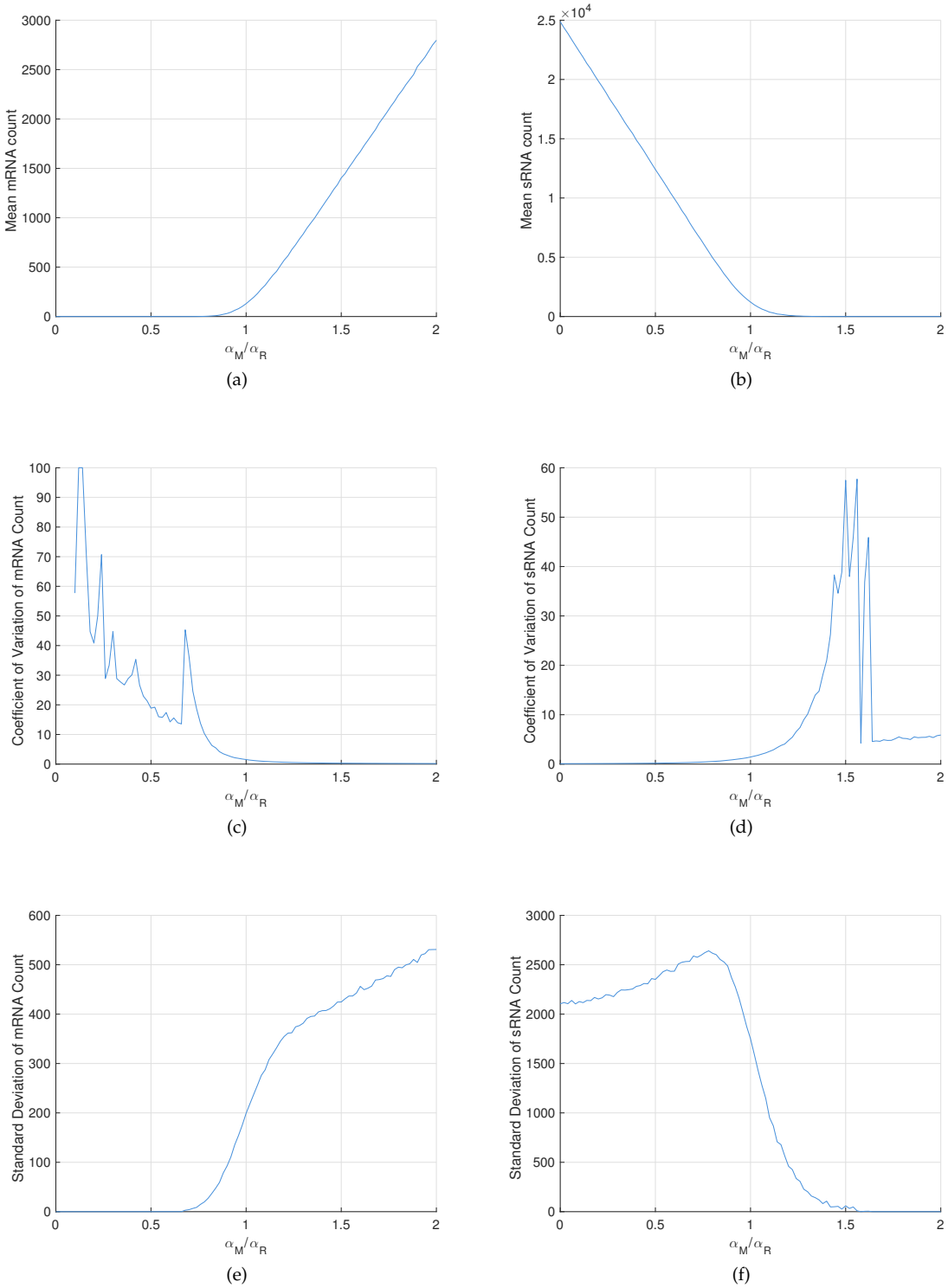
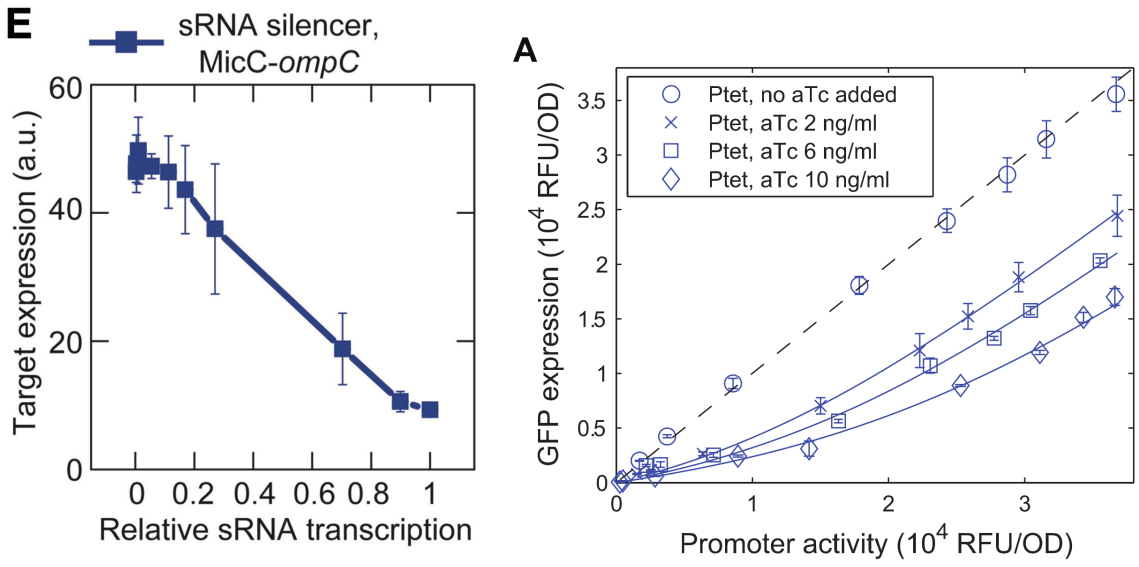


Figure 3.10 – mRNA and sRNA data from stochastic simulation of sRNA Inhibition with the sRNA and target gene encoded on two separate plasmids. This data is also representative of the behaviour of the RNA species in the case of sRNA activation with two plasmids. The parameters used are presented in Table 3.3.



(a) Fig. 2E from [103].

(b) Fig. 3A from [129].

Figure 3.11 – sRNA Inhibition data from the literature. (a) Data by Hussein and Lim shows relatively tight regulation, with both target and sRNA genes on the same plasmid. (b) Data by Levine *et al.* with a soft threshold studying a system where target and sRNA genes are on separate plasmids. What is varied experimentally is the transcription rate of mRNA. The solid lines are fits to Equation 3.2.

During simulation of both sRNA inhibition and activation including plasmid copy number fluctuation ensembles of 10000 runs are executed using the parameters presented in Table 3.3. The transcription rate of the target gene, α_M, is swept from 0 to 10min^{-1} and the transcription rate of sRNA is kept constant at $\alpha_R = 5\text{min}^{-1}$. For simplicity, values for the replication and degradation rates of the two plasmids are the same, giving them the same mean copy number of 17.5. The transcription rates selected provide data around the threshold at $\alpha_M = \alpha_R$.

Starting with the model of sRNA inhibition, the data on protein, mRNA and sRNA number are presented in Figure 3.9 and Figure 3.10. There are two major differences between this case and the basic model detailed in Equation 3.1. The first is that the threshold behaviour of the steady state responses are softened. Secondly, the coefficients of variation peak in a more marked manner away from the threshold. Comparing Figure 3.2b to Figure 3.9b, the peak in the coefficient of variation of the mean protein count when the target genes are on separate plasmids is well over ten times higher than when the gene copy numbers are equal. In addition to this, the peak is at its highest point at $\alpha_M/\alpha_R \approx 0.7$ as opposed to peaking at $\alpha_M/\alpha_R \approx 0.95$ when gene copy numbers are equal. This peak is mirrored in the mRNA and sRNA data presented in Figure 3.10. Such a noise spike would be important to take into account in a design process for a larger circuit, particularly if it operates below the threshold.

For model verification purposes, a feature such an increase in noise might be measurable experimentally using flow cytometry to measure the distribution of expression across cells in a

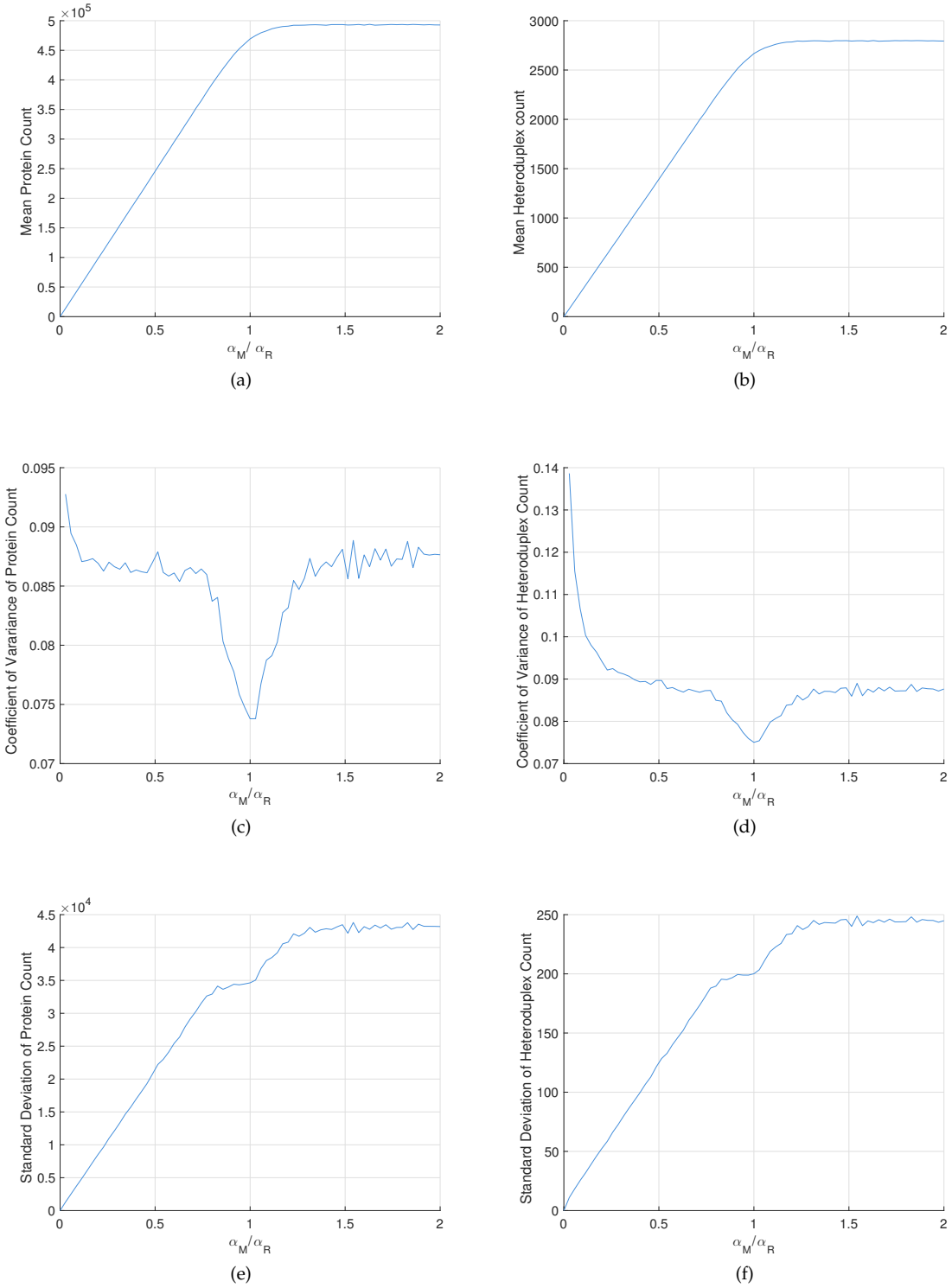


Figure 3.12 – Protein and Heteroduplex data from stochastic simulation of sRNA activation with the sRNA and target gene encoded on two separate plasmids. The parameters used are presented in Table 3.3.

culture. As the noise is seen below the threshold where the level of protein expression is very low, data collection could be challenging. If possible, this could be done using the constructs implemented by Levine *et al.* in [129], comparing the results to the same circuit implemented on a single plasmid, such as done by Hussein and Lim in [102]. As work by Hussein and Lim shows a sharp threshold with their target and sRNA genes are on the same plasmid (see Figure 3.11a), and a soft threshold is observed by Levine *et al.* (see Figure 3.11b), this mechanism provides a possible explanation.

The equivalent data for the model of sRNA activation, including target and sRNA genes on separate plasmids, is less dramatic. The data for the protein and heteroduplex numbers is presented in Figure 3.12, and as the dynamics of the sRNA and target mRNA are identical for both sRNA regulation mechanisms, the steady state behaviour of these species, to within stochastic error, is represented by Figure 3.10. As in the case of sRNA inhibition, the transition at the threshold is softened. The noise seen in the protein numbers is substantially higher: a coefficient of variation of 9% when the dynamics of the plasmids are included, as opposed to 2% without. Interestingly enough, the stochastic data predicts a drop in coefficient of variation at the threshold for both protein and heteroduplex. This might be due to the spikes in mRNA and sRNA noise on either side of the threshold.

3.2 THE ROLE OF HFQ IN SRNA REGULATION

The Host-Factor of Bacteriophage Qβ (Hfq) plays a key role in the sRNA regulation, and is outlined in Section 2.1.16. The aim of this section is to study the interaction of Hfq with RNAs in the context of the sRNA regulatory mechanisms of inhibition and activation. Understanding the effect of Hfq on sRNA regulation improves the modelling and design of systems incorporating these mechanisms. Specifically, the models and results in this section are compared to the basic models of sRNA regulation presented in the previous section.

All detailed qualitative analysis of Hfq's role in sRNA regulation, to the best of my knowledge, has been produced by the Lim Group at The Department of Integrative Biology, University of California, Berkeley. The group has published three papers on the topic, the relevant results of which are presented here.

The first of the three papers was published in 2011 by Hussein and Lim [102]. This is an experimental paper that explores the effect of varying conditions on the sRNA regulation mechanisms of inhibition and activation. The experimental set-up uses four naturally occurring sRNA-target

pairs: RhyB-*sodB*, MicC-*ompC*, OxyS-*fhlA*, which all are inhibited by their sRNAs, and DsrA-*rpoS*, which activates. The target genes are fused to GFP, which is used as a reporter.

In experiments testing the role of Hfq, expression of both the sRNA and the target genes are controlled by the *tet*-promoter without the associated repressor, Tetracycline Repressor (TetR), present in the system. This ensures continuous strong expression. The expression of Hfq is controlled using the *lac*-promoter, which can be induced using IPTG. Cells with and without the gene encoding Hfq on the chromosome are used.

The results of the paper begin by showing that the chosen systems depend on Hfq to function and that the amount of Hfq in a cell is indeed a limiting factor on the inhibition and activation mechanisms of the chosen systems. This is done by showing that efficiency of inhibition and activation increases when the amount of Hfq in the cell is increased. A detail in their results of interest to this section, is that it is shown that sRNA lifetimes are much longer with Hfq present than without. But this is not the case for all RNAs; it was shown that mRNA lifetimes do not change substantially when Hfq is present.

The second paper [1], by Adamson and Lim is a theoretical paper that models Hfq usage and competition, for both the case of a single sRNA-target pair and several. The model developed in this paper heavily inspires the model employed in this section and is a simple symmetric ODE model of the mechanism. The results of the paper theoretically outline how sRNAs, their target mRNAs and Hfq might interact in large networks to ensure efficient regulation in a cell.

One initial important result is that there can be both too little and too much Hfq in a system. In the case of their being too much of the protein, the Hfq forms complexes with individual sRNAs and target mRNAs sequestering them from each other and greatly reducing the efficiency of regulation. Too little Hfq means that there simply is not enough substrate on which the pairing of sRNA and target can occur.

Notably, the paper only looks at the system at the point where the transcription rates of sRNA and target mRNA are equal. In the context of the design of circuits this is a shortcoming, as a circuit employing sRNA regulation will operate within a range of transcription rates that are expected to fluctuate dynamically. In addition, this range does not necessarily need to be centred at the threshold, where the transcription rates are equal.

The third paper [182], by Sagawa *et al.* experimentally verifies the findings of the Adamson and Lim paper. Using the natural inhibiting sRNA-target pairs, that were used in [102], namely RhyB-*sodB*, MicC-*ompC*, OxyS-*fhlA*, and the activating pair GlmZ-*glmUS*, they show that the amount of Hfq in the system can both be insufficient and excessive, in both cases reducing the efficiency of regulation.

In 2005, Večerek *et al.* [212] discovered that Hfq negatively inhibits its own translation. The free RNA-binding protein simply associates with its own mRNA in two places, blocking access by the ribosome, thereby inhibiting translation. Sagawa *et al.* show theoretically that this negative feedback loop increases the total amount of Hfq in a cell when there is a large demand for the protein, increasing the range over which sRNAs regulate efficiently. They do not have experimental results verifying this.

3.2.1 *The Symmetric Model*

In this section a symmetric model used to study the effect of Hfq on sRNA regulation is introduced. This model is almost identical to those found in [1, 182].

As in the case of both sRNA inhibition and activation, where the amount of protein in the system is proportional to the concentration of mRNA that can be translated, in this ODE model the focus is on the RNA species only and the protein is ignored. This facilitates the study of both the activation and inhibition mechanisms through one model.

The model includes the transcription of the target M and sRNA R from their respective genes G_M and G_R. It also includes the expression of Hfq, labelled H, from its gene G_H, modelled as a single step. The model includes the reversible interactions between Hfq and the RNA species, that is: the reversible association of mRNA and sRNA to Hfq forming Hfq-mRNA, HM, and Hfq-sRNA complexes, HR, respectively. It also includes the reversible association of mRNA to the Hfq-sRNA complex and the reversible association of sRNA to the Hfq-mRNA, which both form the three part complex consisting of the Hfq and the sRNA-mRNA target pair, labelled HMR. The release of the heteroduplex formed of the mRNA-sRNA pair, labelled MR, from the Hfq is also included.

Degradation of all of the above species is modelled. The chemical equations for all of these reactions are:

$$
\begin{aligned}
G_M &\xrightarrow{\alpha_M} G_M + M, & M &\xrightarrow{\delta_M} \emptyset,\\
G_R &\xrightarrow{\alpha_R} G_R + R, & R &\xrightarrow{\delta_R} \emptyset,\\
G_H &\xrightarrow{\alpha_H} G_H + H, & H &\xrightarrow{\delta_H} \emptyset,\\
H + M &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} HM, & HM &\xrightarrow{\delta_{HM}} \emptyset,\\
H + R &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} HR, & HR &\xrightarrow{\delta_{HR}} \emptyset,\\
HM + R &\underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} HMR, & HMR &\xrightarrow{\delta_{HMR}} \emptyset,\\
HR + M &\underset{k_{-4}}{\overset{k_4}{\rightleftharpoons}} HMR, & MR &\xrightarrow{\delta_{MR}} \emptyset,\\
HMR &\xrightarrow{K} H + MR,
\end{aligned}
\tag{3.9}
$$

where the transcription rates of mRNA and sRNA are α_M and α_R respectively. The expression rate of Hfq is denoted α_H. The degradation and dilution rates of the species M, R, H, HM, HR, HMR and MR are denoted δ_M, δ_R, δ_H, δ_{HM}, δ_{HR}, δ_{HMR} and δ_{MR}. Dissociation rate of the heteroduplex from Hfq is denoted K and the forward and reverse rates of the interactions between Hfq, and the RNA species in the process of duplex formation are k_i, where $i \in \{1, -1, 2, -2, 3, -3, 4, -4\}$, where the positive sign (+) signifies forwards and the negative sign (−) signifies reverse.

Given that the model is of sRNA regulation, the states in which mRNA is translated and therefore active must be decided. In the case of the inhibition mechanism, the active mRNAs are those that are not in the heteroduplex with sRNA. The number of these is defined to be $M_{free} = M + HM$, which assumes that mRNA is still translated when singularly bound to Hfq. Mirroring this the number of free sRNAs is defined to be $R_{free} = R + HR$. The heteroduplex is required for translation in the case of activation, so the total number of heteroduplex is defined at $M_{reg} = HMR + MR$, which in turn assumes that translation is activated as soon as the remaining half of the heteroduplex binds an Hfq complex.

For ease of analysis the following ODE model of the system is used:

$$\begin{aligned}
\dot{m} &= \alpha_M g_M - \delta_M m - k_1 h m + k_{-1} h_M - k_4 h_R m + k_{-4} h_{MR},\\
\dot{r} &= \alpha_R g_R - \delta_R r - k_2 h r + k_{-2} h_R - k_3 h M r + k_{-3} h_{MR},\\
\dot{h} &= \alpha_H g_H - \delta_H h - k_1 h m + k_{-1} h_M - k_2 h r + k_{-2} h_R + K h_{MR},\\
\dot{h}_M &= -\delta_{HM} h_M + k_1 h m - k_{-1} h_M - k_3 h_M r + k_{-3} h_{MR},\\
\dot{h}_R &= -\delta_{HR} h_R + k_2 h r - k_{-2} h_R - k_4 h_R m + k_{-4} h_{MR},\\
\dot{h}_{MR} &= -\delta_{HMR} h_{MR} + k_3 h_M r - k_{-3} h_{MR} + k_4 h_R m - k_{-4} h_{MR} - K h_{MR},\\
\dot{m}_R &= -\delta_{MR} m_R + K h_{MR},
\end{aligned} \tag{3.10}$$

where m is the concentration of unbound mRNA, r is the concentration of unbound sRNA, h is the concentration of Hfq. The concentrations of the genes encoding these species are denoted g_M, g_R and g_H respectively. The variable h_M is the concentration of the Hfq-mRNA complex, h_R is the concentration of the Hfq-sRNA complex, h_{MR} is the concentration of the Hfq-mRNA-sRNA complex and m_R is the concentration of the heteroduplex formed of mRNA and sRNA.

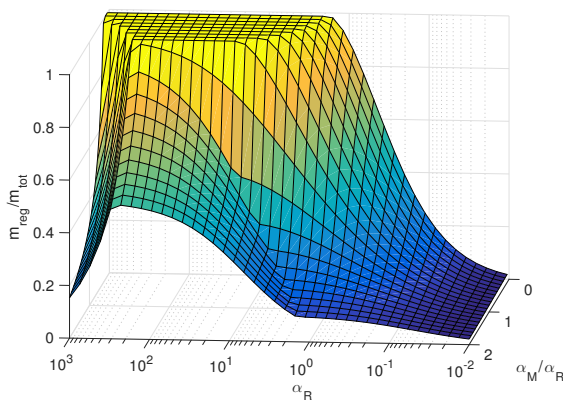
The parameters used for the simulation of this system are presented in Table 3.4. These values assume that Hfq and all complexes with Hfq are not actively degraded and only dilute and so have a half-life equal to the doubling time of a cell, set to about 28min in this case. The rates of association of the Hfq complexes are chosen to be high such that they are not rate limiting. The dissociation rates of the Hfq complexes assume that there is enough RNA capable of binding Hfq in the cell that cycling occurs [68]. According to [210], this is a reasonable assumption. This gives these complexes a half-life of the order of one minute. Free RNA in the cell is assumed to degrade fast, with a turn over of about three minutes.

The steady state of the model is studied numerically, for a range of values of α_M and α_R. The continuous versions of the lumped states are $m_{free} = m + h_M$, representing the concentration of unregulated mRNA; $r_{free} = r + h_R$ is the equivalent for sRNA. The total concentration of regulated mRNA is $m_{reg} = h_{MR} + m_R$. The total mRNA concentration in all states is $m_{tot} = m + h_M + h_{MR} + m_R$. A method of visualising the efficiency of sRNA regulation is by studying the bound ratio, m_{reg}/m_{tot}, i.e. the fraction of mRNA bound in the heteroduplex.

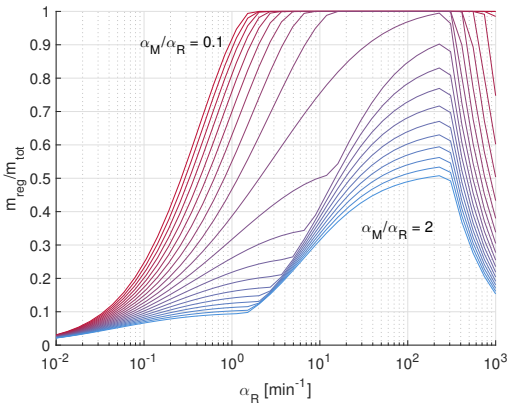
Results showing the efficiency of regulation can be seen in Figure 3.13. As can be seen in Figure 3.13a and Figure 3.13b, the efficiency of regulation by the sRNA drops off for very high ($\alpha_R > 500\text{min}^{-1}$) and low ($\alpha_R < 0.5\text{min}^{-1}$) rates of transcription of sRNA. These are the regimes discussed in [1, 182], where Hfq is limiting for high transcription and in excess for low transcrip-

Table 3.4 – Parameters to simulate the symmetric model of sRNA regulation described in Equation 3.10 and the extended model described by Equation 3.12.

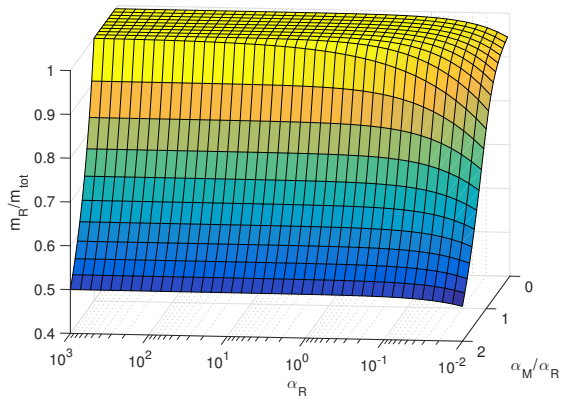
Param.	Value	Param.	Value
α_M	0 to 2000 min^{-1}	k_1	$10nM^{-1}min^{-1}$
δ_M	$0.25min^{-1}$	k_{-1}	$0.7min^{-1}$
α_R	0 to 2000 min^{-1}	k_2	$10nM^{-1}min^{-1}$
δ_R	$0.25min^{-1}$	k_{-2}	$0.7min^{-1}$
α_H	$25min^{-1}$	k_3	$10nM^{-1}min^{-1}$
δ_H	$0.025min^{-1}$	k_{-3}	$0.7min^{-1}$
δ_{HM}	$0.025min^{-1}$	k_4	$10nM^{-1}min^{-1}$
δ_{HR}	$0.025min^{-1}$	k_{-4}	$0.7\ min^{-1}$
δ_{HMR}	$0.025min^{-1}$	g_M	28.2nM
δ_{MR}	$0.25min^{-1}$	g_R	28.2nM
K	$5min^{-1}$	g_H	1.7nM



(a) With Hfq



(b) With Hfq



(c) Without Hfq

Figure 3.13 – The steady state bound ratio, i.e. m_{reg}/m_{tot} in figures (a) and (b) show results from the ODE model of sRNA regulation with Hfq (see Equation 3.10). In (c), the equivalent metric $m_R/(m + m_R)$ is presented for sRNA regulation without Hfq modelled using Equation 3.5b to Equation 3.5d. The parameters used are found in Table 3.4 and Table 3.5. The transcription rate of mRNA, α_R, is swept from 0.01 to $1000min^{-1}$. The ratio α_M/α_R is also varied from 0.1 to 2 to show behaviour around the threshold.

tion. Despite this, there is a large range on transcription rates for which the bound ratio achieves a value of almost one, signifying that nearly all mRNA in the system is bound to sRNA.

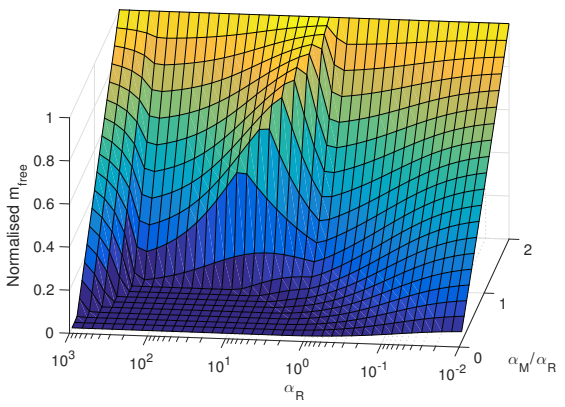
For clarity, the model of regulation with Hfq is compared to sRNA regulation without Hfq, modelled using Equation 3.5b to Equation 3.5d of the sRNA activation model detailed in Section 3.1.2. In this case, $m_{tot} = m + m_R$ and the concentration of heteroduplex is simply the state m_R. The result of the equivalent simulation using the parameters in Table 3.5 is presented in Figure 3.13c. Here it can clearly be seen that there is efficient regulation over the full spectrum except at very low levels of transcription, where the rate of formation of heteroduplex is very low due to the low concentrations of mRNA and sRNA. This effect leads to a smoother transition at the threshold.

Of interest to the design of circuits using sRNA regulation is the behaviour of the concentration of free mRNA, m_{free}, in the case of inhibition, and the concentration of heteroduplex, m_{reg} in the case of activation, as these are directly proportional to the concentration of expressed protein, which is usually used as the output of such a system. In the simple case presented in Section 3.1, the steady state behaviour observed is threshold-linear as presented in Figure 3.1 and Figure 3.5.

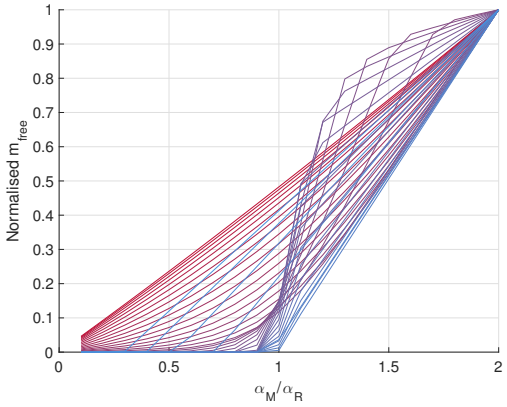
Results showing the behaviour of m_{free} in the case of the symmetric model including Hfq are presented in Figure 3.14a and Figure 3.14b. These figures are obtained by normalising the steady state values of m_{free} at each individual value of α_R. This facilitates a visual comparison of the steady state behaviour of m_{free} around the threshold of $\alpha_R = \alpha_M$. The low efficiency of regulation at high and low values of α_R can be observed as there is mRNA that can be translated far below the threshold of $\alpha_R = \alpha_M$. This is expected, given Figure 3.13. In addition to this, it can be seen in particular that for mid range values of α_R of between 1min^{-1} and 200min^{-1}, that while there is tight regulation for $\alpha_M < \alpha_R$, the behaviour above the threshold is not linear, as it is in for the model without Hfq (see Figure 3.14c). This is in particular visible in Figure 3.14b, where for certain values of α_R, the steady state behaviour has three sections: 1) the inhibited regime, where there is hardly any m_{free} present in the system. 2) a very steep section at and just above the threshold. 3) a flatter section for higher values of α_M/α_R.

As the model studied here treats the sRNA and mRNA symmetrically and the parameter values used for each RNA species are the same, the free concentration of sRNA, r_{free}, behaves in an identical manner to m_{free} and therefore need not be presented.

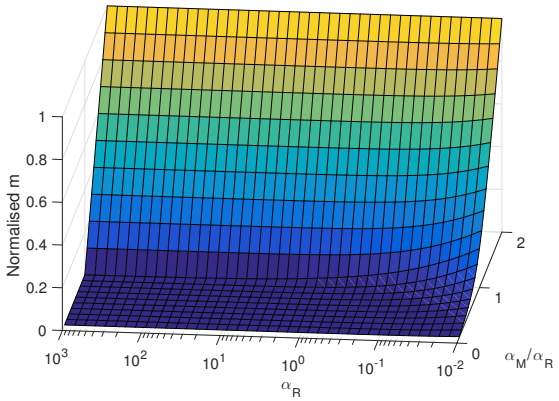
Interestingly enough, the non-linear behaviour observed for m_{free} is not seen in the steady state concentration of m_{ref}, as seen in Figure 3.15. Again, low efficiency in regulation can be seen for high and low levels of α_R. For low α_R, the production of heteroduplex is very slow due to the low concentration of the RNA species; the threshold is very soft and smooth. In the case of high α_R, the maximum amount of heteroduplex is reached well below the threshold, as



(a) With Hfq

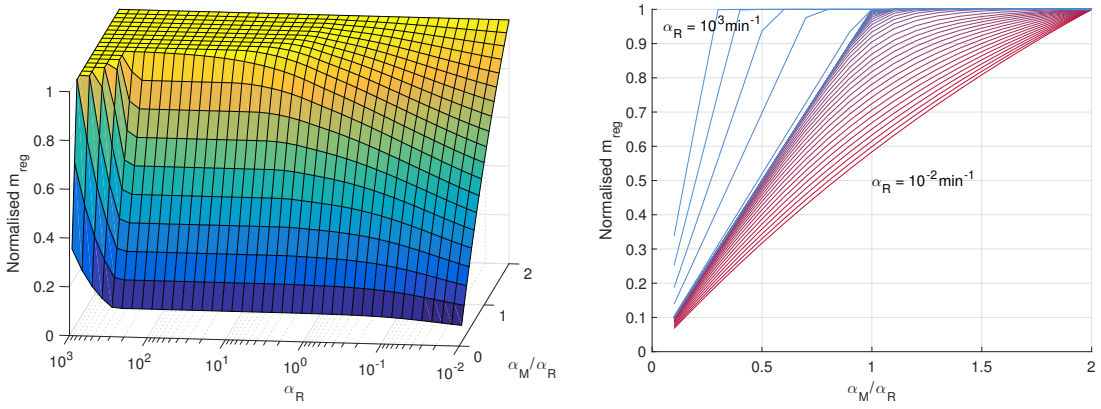


(b) With Hfq



(c) Without Hfq

Figure 3.14 – The steady state normalised free mRNA concentration, m_{free} is presented in figures (a) and (b) which show results from the ODE model of sRNA regulation with Hfq (see Equation 3.10). In (b) the colour scale is from low to high values of α_R. For comparison, in (c) the normalised steady state concentration of m is presented, modelled using Equation 3.5b to Equation 3.5d. The parameters used are found in Table 3.4 and Table 3.5. The transcription rate of mRNA, α_R, is swept from 0.01 to 1000min^{-1}. The ratio α_M/α_R is also varied from 0.1 to 2 to show behaviour around the threshold. The data has been normalised at each value of α_R, to be able to compare the steady state behaviour across many orders of magnitude of transcription rate.



(a) (b)

Figure 3.15 – The steady state normalised regulated mRNA, i.e. mRNA bound in the heteroduplex, m_{reg} in figures (a) and (b) which show results from the ODE model of sRNA regulation with Hfq (see Equation 3.10). For comparison, the equivalent plot for the model of sRNA regulation without Hfq has the exact same shape as Figure 3.13c. The parameters used are found in Table 3.4. The transcription rate of mRNA, α_R, is swept from 0.01 to 1000min^{-1}. The ratio α_M/α_R is also varied from 0.1 to 2 to show behaviour around the threshold. The data has been normalised at each value of α_R, to be able to compare the steady state behaviour of many orders of magnitude of transcription.

Table 3.5 – Parameters used to compare the sRNA system without Hfq defined by equations Equation 3.5b to Equation 3.5d to the symmetric model of sRNA regulation with Hfq described by Equation 3.10.

Param.	Value	Param.	Value
α_M	0 to 2000 min^{-1}	δ_{MR}	0.25min^{-1}
δ_M	0.25min^{-1}	g_M	28.2nM
α_R	0 to 2000 min^{-1}	g_R	28.2nM
δ_R	0.025min^{-1}	K	5nM^{-1}min^{-1}

there are too many RNAs in the system for Hfq to efficiently fulfil its role. This limits the possible heteroduplex concentration.

If indeed correct, the nonlinearity seen in the steady state behaviour of the free mRNA concentration undermines the use of the basic model of sRNA inhibition presented in Section 3.1.1. The source of this nonlinearity turns out to be the ratio between the degradation rates of the two species included in m_{free}. Figure 3.16 shows the steady state behaviour of m_{free} for various values of δ_M. When $\delta_M = \delta_{HM}$ at 0.025min^{-1}, the threshold-linear behaviour is recovered.

3.2.2 *Extension of the Model*

The above model treats the target mRNA and sRNA symmetrically, but experimental results hint that in a real system this is not the case. There is firm agreement that Hfq acts as a chaperone for many sRNAs, protecting them from degradation [57, 77, 99, 102, 136, 211]. On the other hand,

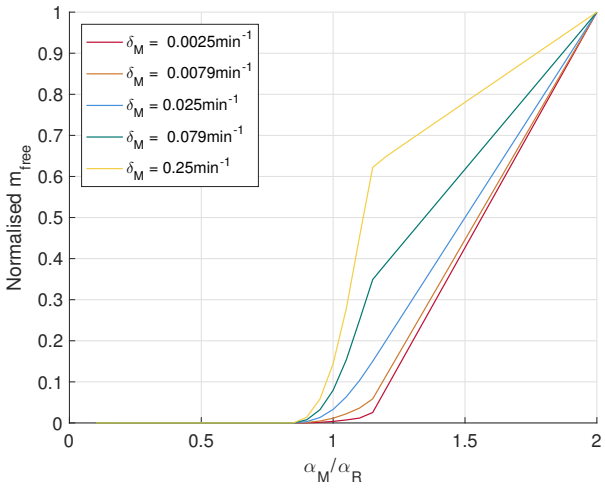


Figure 3.16 – The effect of varying δ_M on the steady state behaviour of the symmetric sRNA regulation model including Hfq detailed in Equation 3.10. The parameters used are those in Table 3.4, except that δ_M is swept from 0.0025min^{-1} to 0.25min^{-1}. The transcription rate of sRNA is fixed at $\alpha_R = 10\text{min}^{-1}$.

Hussein and Lim showed that mRNA concentrations in the cell do not change significantly in the presence of Hfq [102]. This points to asymmetric behaviour in the interaction between these two species of RNA and Hfq.

To incorporate this in the model, assume that the sRNA is protected when bound to Hfq and therefore is not degraded by cellular machinery, but is diluted out of the system and therefore has a half-life equal to the doubling time of the cell, as a protein such as Hfq does. On the other hand, assume the opposite for mRNA, that it is not protected by Hfq and is degraded by the degradosome, whether it is bound to Hfq, in the heteroduplex or free.

A mechanism, that could explain this, is that the sRNA is small in comparison to Hfq and therefore, when it is bound to Hfq all sites targeted by the degradation machinery are protected. mRNAs are much longer, and therefore the Hfq only covers a small section of the entire strand. The unbound loose sections of the mRNA can easily be accessed by the degradation machinery of the cell to begin the process of degradation.

The current model assumes that the Hfq complexes including mRNA only dilute: $\delta_{HR} = \delta_{HMR} = 0.025\text{min}^{-1}$, which gives these species the same half-life as the doubling time of the cell, about 28min. The two chemical equations that need to be added to the symmetric model of sRNA regulation detailed in Equation 3.9 to take the rapid degradation of mRNA into account are:

$$\begin{aligned} \text{HM} &\xrightarrow{\delta_M} \text{H}, \\ \text{HMR} &\xrightarrow{\delta_M} \text{HR}, \end{aligned} \tag{3.11}$$

where the first reaction describes the degradation of mRNA when singularly bound to Hfq, and the second describes the degradation of mRNA when in complex with both Hfq and the sRNA. Note that instead of introducing new rate parameters, the degradation rate of free mRNA, δ_M, is used as the rate constant for these reactions.

These two equations add extra terms to the system of ODEs, extending them to

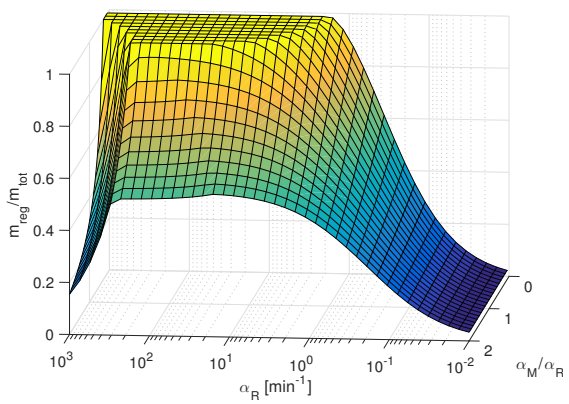
$$\begin{aligned}
\dot{m} &= \alpha_M g_M - \delta_M m - k_1 h m + k_{-1} h_M - k_4 h_R m + k_{-4} h_{MR},\\
\dot{r} &= \alpha_R g_R - \delta_R r - k_2 h r + k_{-2} h_R - k_3 h M r + k_{-3} h_{MR},\\
\dot{h} &= \alpha_H g_H - \delta_H h - k_1 h m + k_{-1} h_M - k_2 h r + k_{-2} h_R + K h_{MR} + \delta_M h_M,\\
\dot{h}_M &= -(\delta_{HM} + \delta_M) h_M + k_1 h m - k_{-1} h_M - k_3 h_M r + k_{-3} h_{MR},\\
\dot{h}_R &= -\delta_{HR} h_R + k_2 h r - k_{-2} h_R - k_4 h_R m + k_{-4} h_{MR} + \delta_M h_{MR},\\
\dot{h}_{MR} &= -(\delta_{HMR} + \delta_M) h_{MR} + k_3 h_M r - k_{-3} h_{MR} + k_4 h_R m - k_{-4} h_{MR} - K h_{MR},\\
\dot{m}_R &= -\delta_{MR} m_R + K h_{MR},
\end{aligned} \tag{3.12}$$

where an extra term is added to each of the equations for h, h_M, h_R and h_{MR} to account for the above chemical reactions.

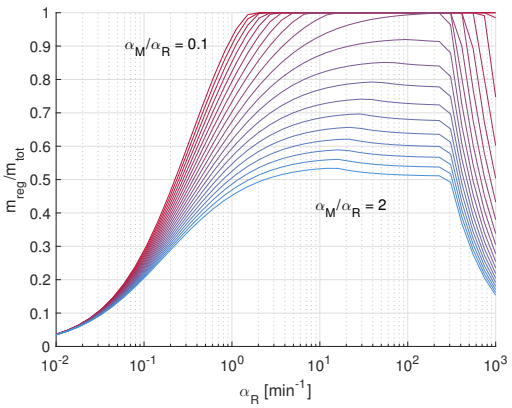
Steady state results of the extended model, calculated using the parameters in Table 3.4, are presented in Figure 3.17. Comparing Figure 3.17a and Figure 3.17b to the equivalent plots in Figure 3.13, the extended model shows marginally more efficient regulation over a larger range of transcription rates. This extension has also removed the nonlinear behaviour from the steady state of m_{free} around the threshold, as can be seen when comparing Figure 3.17c and Figure 3.17d to the equivalent plots in Figure 3.14. The extension of the model has slightly softened the threshold of the heteroduplex (see Figure 3.17e) compared to that of the symmetric model presented in Figure 3.15, but the change is slight, and there is no change to qualitative behaviour.

3.2.3 *Binding Order Preference*

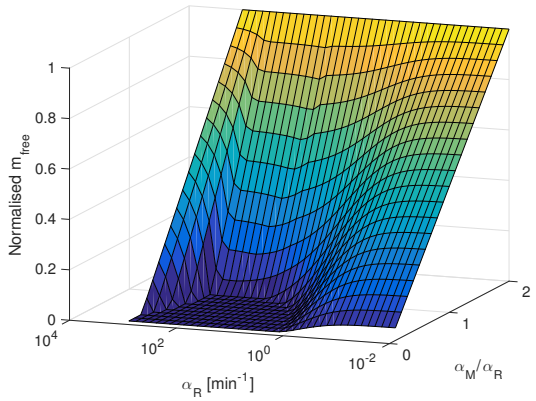
In the above models, the heteroduplex between sRNA and target mRNA can form via two routes: either the sRNA binds Hfq first, or the mRNA binds Hfq first. Adamson and Lim predicted [1] and Sagawa *et al.* showed [182] that this random order binding reduces the efficiency of heteroduplex formation. Particularly at low concentrations of total RNA, if both mRNA and sRNA bind to Hfq with high affinity, the efficiency of regulation drops substantially due to sequestration of the two species. The RNA species are simply trapped in complex with Hfq, where they are singu-



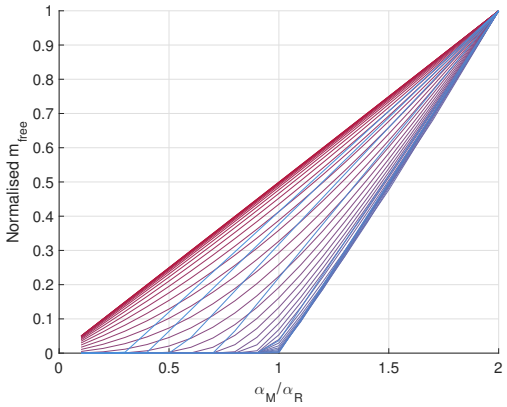
(a) Bound Ratio



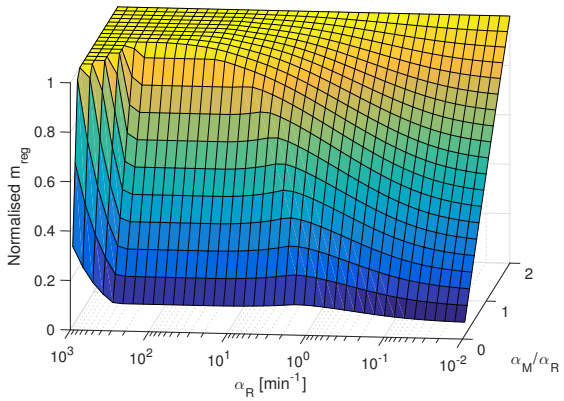
(b) Bound Ratio



(c) Free mRNA



(d) Free mRNA



(e) Heteroduplex

Figure 3.17 – Steady state results of the extended asymmetric sRNA regulation model described by the ODEs in Equation 3.12. The transcription rate of mRNA, α_R, is swept from 0.01 to 1000min^{-1}. The ratio α_M/α_R is varied from 0.1 to 2 to show behaviour around the threshold. In (d) the colour scale is from low to high values of α_R. The other parameters used are in Table 3.4.

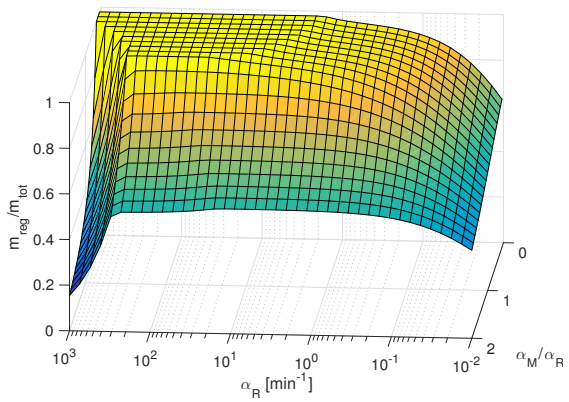
larly bound and therefore cannot find each other. This is observed to occur in Figure 3.13 and Figure 3.17 where there is a sharp drop off in the efficiency of the regulation below $\alpha_R \approx 1\text{min}^{-1}$.

So far in this chapter, all results modelling sRNA regulation with Hfq assume that the forward reaction rates of the interactions between Hfq and the RNA species are equal (see Table 3.4). In most natural systems, such as those studied in [102, 129, 182, 211], it has been shown that both the sRNAs and the target mRNAs used in experiments have sequences that actively and efficiently bind Hfq. Synthetic systems studied in these papers use the naturally occurring sRNA and the 5′ UTR and the first section of the gene coding region of the natural target mRNA fused to a reporter, such as GFP, as the sRNA-target pair. For example in the case of Levine *et al.*, the synthetic system used consisted the sRNA RhyB with a target consisting of the 5′ UTR and the first 11 codons of its target *sodB* fused to *gfp* [129]. In these cases, it is fair to assume that if the forward binding rates are not exactly equal, they are at least comparable and high.

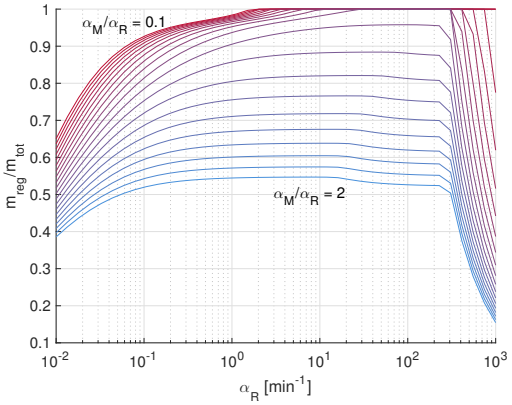
In a synthetic system using sRNA regulation, it can be expected that bias is introduced into the binding order of the RNA species with the Hfq protein. An example of this is the protocol laid out in [224]. The paper describes how an efficient synthetic sRNA inhibition mechanism can be implemented. The mechanism described uses the scaffold of a natural sRNA. The scaffold is defined as all of the natural sRNA apart from the section that is complementary to the target mRNA. Crucially, it does not use a natural scaffold for the target mRNA, such as is employed in [102, 129, 182, 211]. The sRNA is simply designed with a 24 nucleotide section that is perfectly complementary to a section of the target mRNA. See Section 2.1.15.4 for more detail.

As only the sRNA is constructed using a natural scaffold with a known and strong affinity for Hfq, the rate of association of sRNA to Hfq is expected to be much higher than that of the target mRNA creating a bias in the order of binding towards the sRNA binding first. Steady state analysis of the extended model including this bias is performed using the parameters in Table 3.4, with the only difference being that the binding between free Hfq and target mRNA has been lowered to $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$. This assumes that the association rate between free Hfq and the target is low, but due to the complementarity between sRNA and target, and the general ability of Hfq to bind RNA, the rate of association between the Hfq-sRNA complex and the free mRNA remains high.

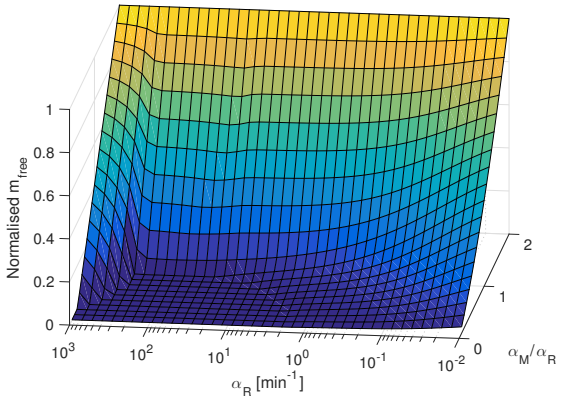
The steady state results are presented in Figure 3.18. As can be seen, the introduction of bias substantially increases the regulation efficiency of the sRNA regulation mechanism, extending it to much lower transcription rates (see Figure 3.18a and Figure 3.18a in particular). The behaviours of the total free mRNA, m_{free}, and the heteroduplex, m_{ref}, continue to have the desired threshold-linear shape. In addition, as the regulation efficiency has dramatically increased for low transcription rates, the behaviour of m_{free} and m_{ref} in this regime has also improved.



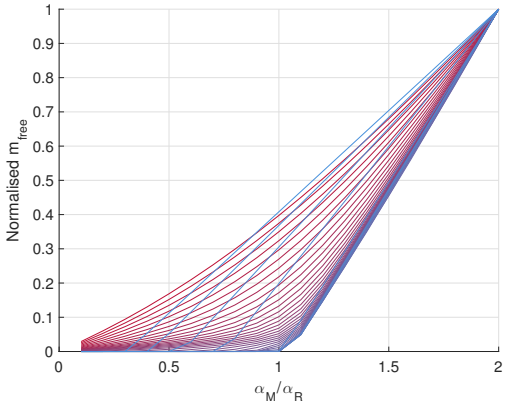
(a) Bound Ratio



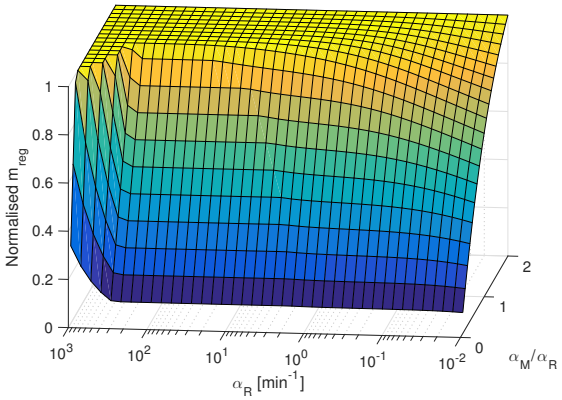
(b) Bound Ratio



(c) Free mRNA



(d) Free mRNA



(e) Heteroduplex

Figure 3.18 – Steady state results when including biased binding order in the extended sRNA regulation model described by the ODEs in Equation 3.12. The transcription rate of mRNA, α_R, is swept from 0.01 to 1000min^{-1}. The ratio α_M/α_R is also varied from 0.1 to 2 to show behaviour around the threshold. In (d) the colour scale is from low to high values of α_R. The other parameters used are in Table 3.4 with the change that $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$.

3.2.4 *Stochastic Simulation*

As with the basic model of sRNA regulation, it is of interest to explore stochastic simulations of a model including Hfq. The simulation results presented in this section employ the extended model of sRNA regulation with Hfq, specifically the reactions in Equation 3.9 combined with the reactions in Equation 3.11. To these reactions are added reactions that govern the dynamics of the expressed protein.

In the case of sRNA inhibition, these reactions are

$$\begin{aligned} M &\xrightarrow{\alpha_P} M + P_{in}, \\ HM &\xrightarrow{\alpha_P} HM + P_{in}, \\ P_{in} &\xrightarrow{\delta_P} \emptyset, \end{aligned} \tag{3.13}$$

where P_{in} denotes the protein expressed in the case of the sRNA inhibition mechanism with translation rate α_P and degradation rate δ_P. As can be seen, protein is only translated from the mRNAs not bound to sRNAs. These are included in $M_{free} = M + HM$.

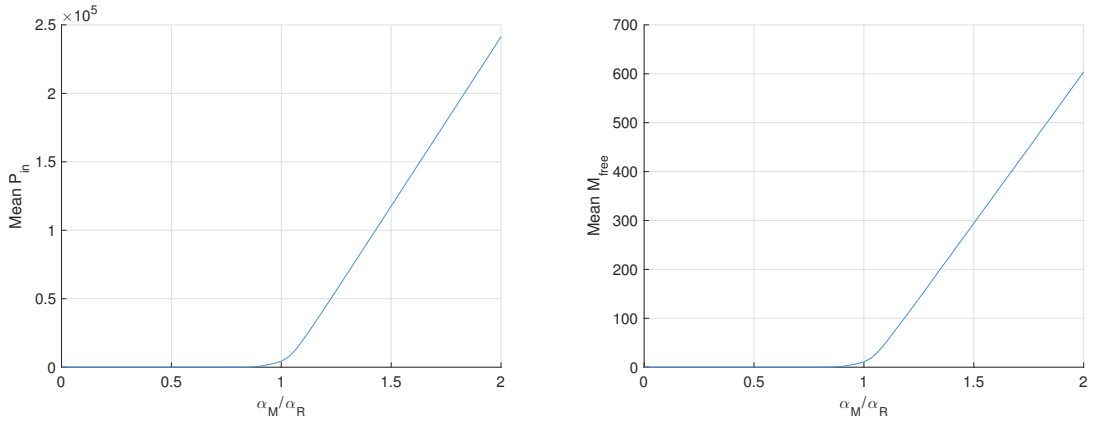
Similarly, in the case of sRNA activation, the additional reactions are

$$\begin{aligned} HMR &\xrightarrow{\alpha_P} HMR + P_{ac}, \\ MR &\xrightarrow{\alpha_P} MR + P_{ac}, \\ P_{ac} &\xrightarrow{\delta_P} \emptyset, \end{aligned} \tag{3.14}$$

where P_{ac} is the protein expressed due to sRNA activation, with the same translation and degradation rate as used in the case of P_{in}. The reactions describe translation of the heteroduplex, as expected.

The results presented are from a 10000 run ensemble simulation of the models of sRNA inhibition and activation using the parameters in Table 3.4 with $\alpha_P = 5\text{min}^{-1}$, $\delta_P = 0.025\text{min}^{-1}$, including the bias term $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$, $\alpha_R = 10\text{min}^{-1}$ and sweeping α_M/α_R from 0 to 2. The choice of sRNA transcription rate ensures that the system is in a regime of efficient regulation. The results for the symmetric model and the extended model without bias were checked and found to be qualitatively identical.

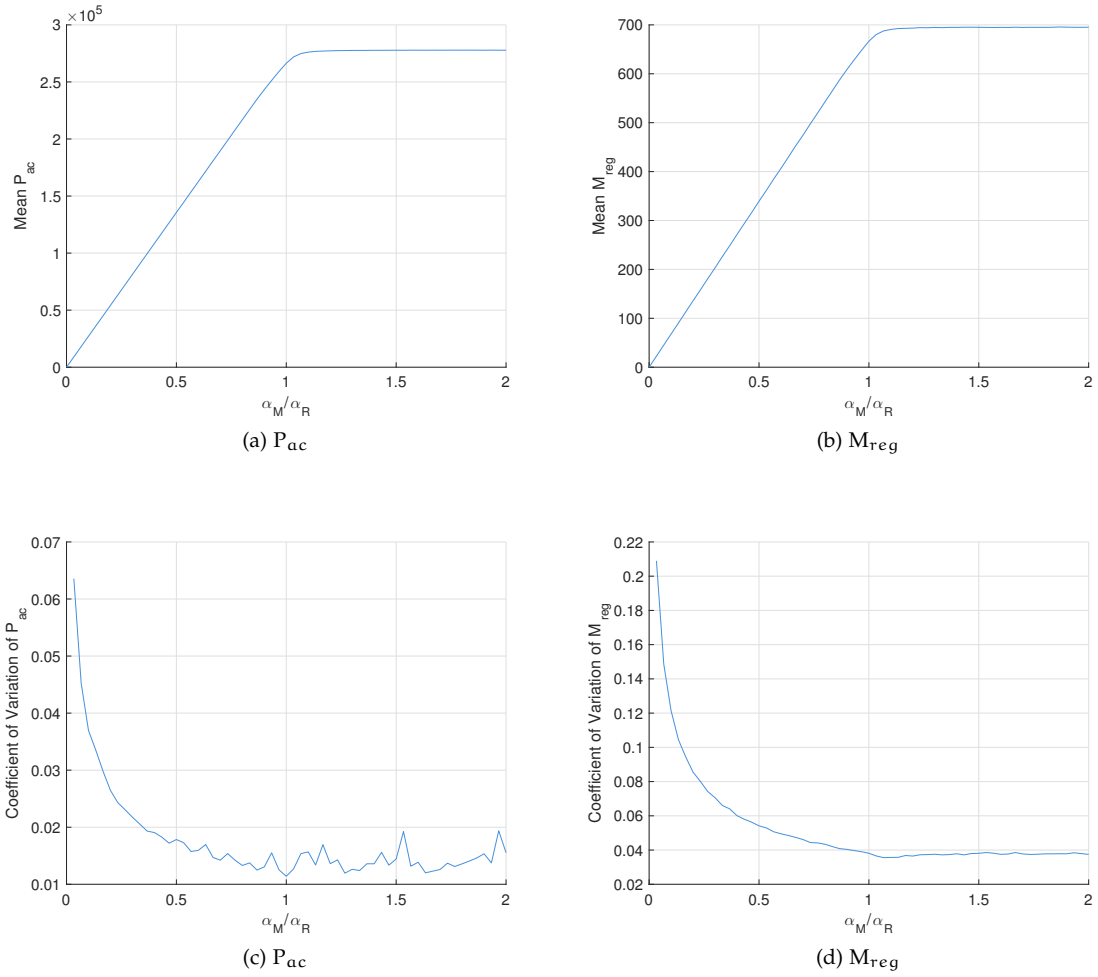
Studying the results of sRNA inhibition with Hfq presented in Figure 3.19, and comparing these results to the results of stochastic simulation of the basic model of sRNA inhibition presented in Figure 3.2 and Figure 3.3, the peak in noise expected just below the threshold is barely visible in



(a) P_{in} (b) M_{free}

(c) P_{in} (d) M_{free}

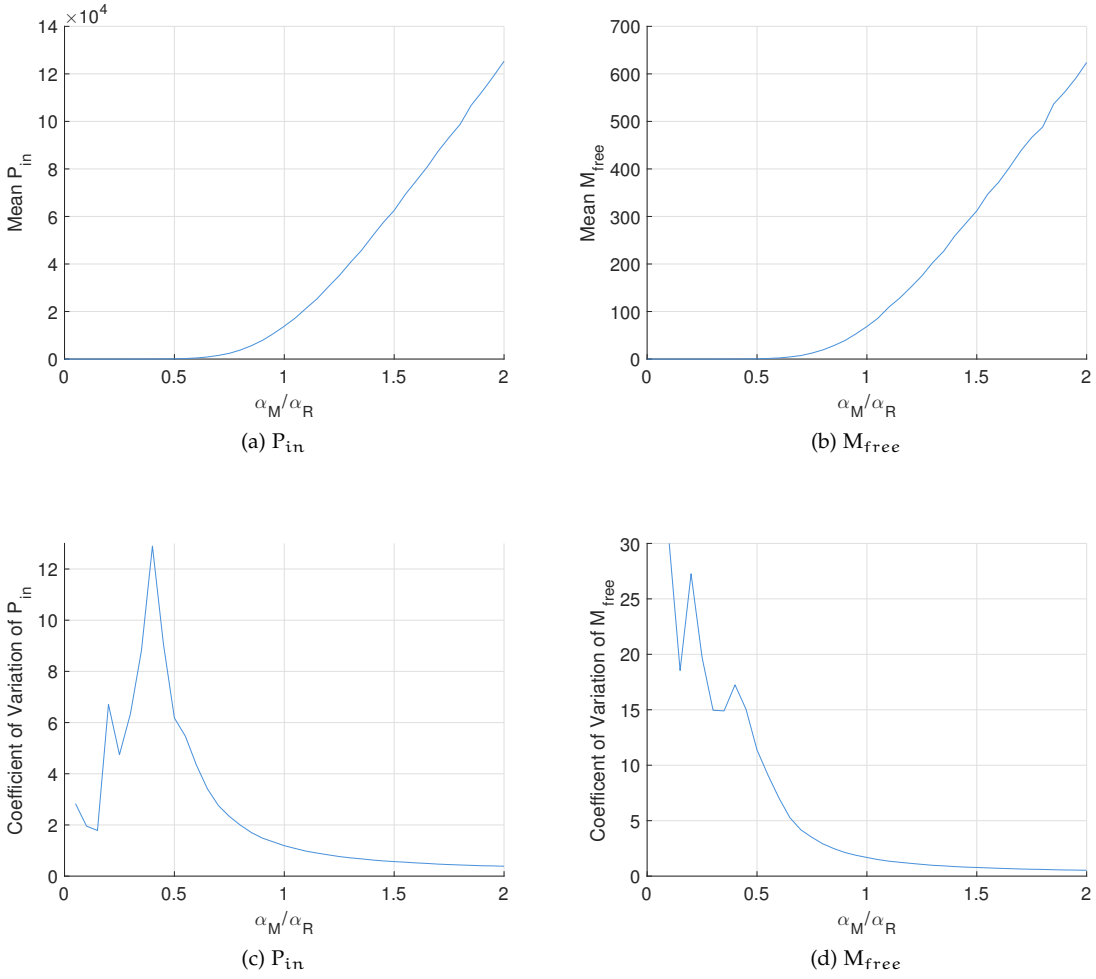
Figure 3.19 – Results from stochastic simulation of the extended model of sRNA inhibition including Hfq and protein translation. The parameters used are in Table 3.4 with $\alpha_P = 5\text{min}^{-1}$, $\delta_P = 0.025\text{min}^{-1}$ and including the bias term $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$, $\alpha_R = 10\text{min}^{-1}$ and sweeping α_M/α_R from 0 to 2.



(a) P_{ac} (b) M_{reg}

(c) P_{ac} (d) M_{reg}

Figure 3.20 – Results from stochastic simulation of the extended model of sRNA activation including Hfq and protein translation. The parameters used are in Table 3.4 with $\alpha_P = 5\text{min}^{-1}$, $\delta_P = 0.025\text{min}^{-1}$, including the bias term $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$, $\alpha_R = 10\text{min}^{-1}$ and sweeping α_M/α_R from 0 to 2.



(a) P_{in} (b) M_{free}

(c) P_{in} (d) M_{free}

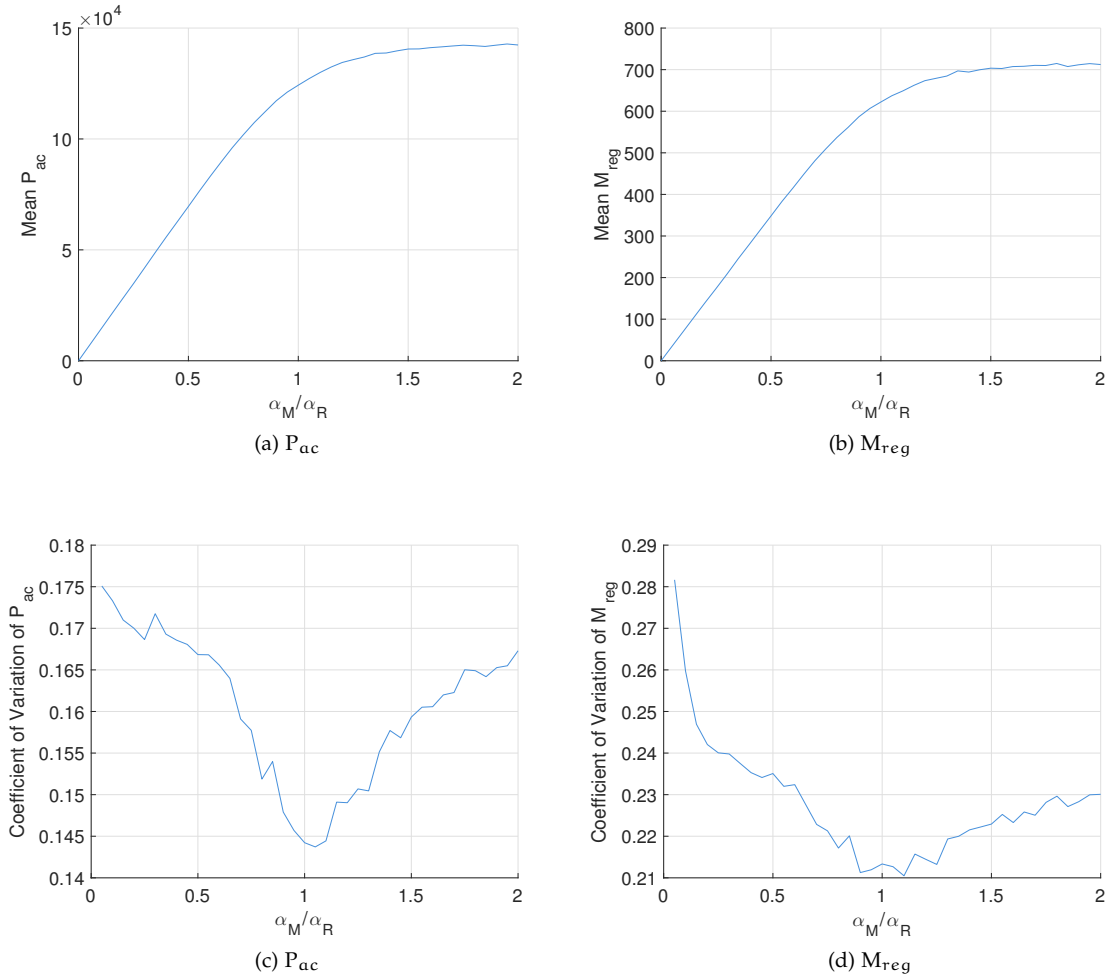
Figure 3.21 – Results from stochastic simulation of the extended model of sRNA inhibition including Hfq, assuming that sRNA and target genes are encoded on separate plasmids. The parameters used are in Table 3.4 with $\alpha_P = 5\text{min}^{-1}$, $\delta_P = 0.025\text{min}^{-1}$ and including the bias term $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$, $\alpha_R = 10\text{min}^{-1}$ and sweeping α_M/α_R from 0 to 2. The production rates of both plasmids are $\alpha_{G_M} = \alpha_{G_R} = 0.44\text{min}^{-1}$ and the degradation rates are $\delta_{G_M} = \delta_{G_R} = 0.025\text{min}^{-1}$.

the case of the protein, P_{in}, and is not visible in the case of the mRNA, though the levels of noise predicted at and above the threshold are very similar.

In the case of sRNA activation with Hfq presented in Figure 3.20, the results agree well with the basic model. This can be seen when comparing results to those in Figure 3.6 and Figure 3.7. The noise levels of the expressed protein P_{ac} are also below 2% for most of the range of α_M/α_R.

3.2.4.1 *Plasmid Copy Number Variation*

The next step is to include copy number variation in the model. As in the case of the basic models of sRNA regulation, when the sRNA and target genes are assumed to be on the same plasmid, there is no change in behaviour, as the relative transcription rates of the sRNA and target are always the same. Having target and sRNA genes on separate plasmids is modelled by including the reactions



(a) P_{ac} (b) M_{reg}

(c) P_{ac} (d) M_{reg}

Figure 3.22 – Results from stochastic simulation of the extended model of sRNA activation including Hfq, assuming that sRNA and target genes are encoded on separate plasmids. The parameters used are in Table 3.4 with $\alpha_P = 5\text{min}^{-1}$, $\delta_P = 0.025\text{min}^{-1}$, including the bias term $k_1 = 0.1\text{nM}^{-1}\text{min}^{-1}$, $\alpha_R = 10\text{min}^{-1}$ and sweeping α_M/α_R from 0 to 2. The production rates of both plasmids are $\alpha_{G_M} = \alpha_{G_R} = 0.44\text{min}^{-1}$ and the degradation rates are $\delta_{G_M} = \delta_{G_R} = 0.025\text{min}^{-1}$.

in Equation 3.8, as is done in the case of the models without Hfq. The same parameters are used in this case: The mean copy numbers of both plasmids are set to 17.5 and with degradation rates of $\delta_{G_M} = \delta_{G_R} = 0.025\text{min}^{-1}$, this yields production rates of $\alpha_{G_M} = \alpha_{G_R} = 0.44\text{min}^{-1}$.

The results of the sRNA inhibition model are presented in Figure 3.21 and the results of the sRNA activation model are presented in Figure 3.22. When comparing these to the models including Hfq, but without plasmid copy number variation (see Figure 3.19 and Figure 3.20), the thresholds, in both the case of inhibition and activation are vastly softened, leading to a very smooth transition between the two regimes.

As stated, the models of sRNA regulation without plasmid copy number variation, both with and without Hfq agree closely both qualitatively and quantitatively. On the other hand, though the models of sRNA regulation including plasmid copy number variation agree qualitatively, they disagree quantitatively. Compare Figure 3.21 and Figure 3.22, results from the models with both Hfq and plasmid copy number variation, to Figure 3.9, Figure 3.10 and Figure 3.12, results of the model without Hfq but with plasmid copy number variation. The results of the model with Hfq have a greatly smoothed threshold transition and notably larger amounts of noise, particularly above the threshold. In the case of sRNA activation, this is particularly visible, with the coefficient of variation predicted to be about 17% in the case of the model with Hfq as opposed to 9% in the model without. The greatly increased smoothing of the threshold, in particular, supports the theory that the difference in behaviour reported by Levine *et al.* in [129] and Hussein and Lim in [102], shown in Figure 3.11, is due to the differing set-up of the two experiments.

3.3 COMPARISON OF THE DYNAMIC RESPONSES

The final missing piece in the study of sRNA regulation and the role of Hfq in the mechanism is the comparison of the responses of the two types of model: with and without Hfq. The models used in this section are basic models of sRNA regulation and the extended model of sRNA regulation with Hfq. The studied species in this case is the protein produced. This is because in a large system, this is most likely to be the species used to transfer information in the network.

For clarity, the ODE models used are presented again. The models of sRNA regulation for both the inhibition and activation mechanisms are almost identical. Using this fact, the models presented include both the dynamics of the protein produced in each of these cases. This is possible as the dynamics of the proteins produced do not affect other species in the model.

The basic model used is:

$$\begin{aligned}
\dot{p}_{in} &= \alpha_P m - \delta_P p_{in},\\
\dot{p}_{ac} &= \alpha_P m_R - \delta_P p_{ac},\\
\dot{m} &= \alpha_M g_M - \delta_M m - K m r,\\
\dot{r} &= \alpha_R g_R - \delta_R r - K m r,\\
\dot{m}_R &= K m r - \delta_{MR} m_R,
\end{aligned} \tag{3.15}$$

where the states and parameters are as those in Equation 3.5, with the addition of p_{in} and p_{ac}. The new state p_{in} is the concentration of protein produced by the sRNA inhibition mechanism. Note that the ODE governing it is dependent only on m and p_{in}, as is required in the case of inhibition. The other new state, p_{ac}, is the concentration of protein produced by the sRNA activation mechanism. As expected, its ODE only depends m_R and p_{ac}, as it is produced from the heteroduplex. For simplicity, the protein states use the same parameters α_P and δ_P.

The ODE model of the system including Hfq is:

$$\begin{aligned}
\dot{p}_{in} &= \alpha_P(m + h_M) - \delta_P p_{in},\\
\dot{p}_{ac} &= \alpha_P(h_{MR} + m_R) - \delta_P p_{ac},\\
\dot{m} &= \alpha_M g_M - \delta_M m - k_1 h m + k_{-1} h_M - k_4 h_R m + k_{-4} h_{MR},\\
\dot{r} &= \alpha_R g_R - \delta_R r - k_2 h r + k_{-2} h_R - k_3 h M r + k_{-3} h_{MR},\\
\dot{h} &= \alpha_H g_H - \delta_H h - k_1 h m + k_{-1} h_M - k_2 h r + k_{-2} h_R + K h_{MR} + \delta_M h_M,\\
\dot{h}_M &= -(\delta_{HM} + \delta_M) h_M + k_1 h m - k_{-1} h_M - k_3 h_M r + k_{-3} h_{MR},\\
\dot{h}_R &= -\delta_{HR} h_R + k_2 h r - k_{-2} h_R - k_4 h_R m + k_{-4} h_{MR} + \delta_M h_{MR},\\
\dot{h}_{MR} &= -(\delta_{HMR} + \delta_M) h_{MR} + k_3 h_M r - k_{-3} h_{MR} + k_4 h_R m - k_{-4} h_{MR} - K h_{MR},\\
\dot{m}_R &= -\delta_{MR} m_R + K h_{MR},
\end{aligned} \tag{3.16}$$

where the states and parameters are the same as in Equation 3.12, with the addition, similar to the basic model, of p_{in} and p_{ac}. Again p_{in} is the concentration of the protein produced in the case of inhibition and it is produced from $m_{free} = m + h_M$. Similarly p_{ac} is the concentration of the protein produced in the case of activation and it is produced from the RNAs in heteroduplex form: $m_{reg} = h_{MR} + m_R$. Again for simplicity, the protein states use the same parameters for transcription and degradation.

The systems are simulated using the parameters in Table 3.5 for the basic model and Table 3.4 for the model with Hfq, with the addition of $\delta_P = 0.025 \text{min}^{-1}$, giving the protein a degradation

rate with a half-life of about 28min therefore only accounting for dilution, and a translation rate of $\alpha_P = 1\text{min}^{-1}$.

Responses are presented for three regimes around the threshold at $\alpha_M = \alpha_R$. The systems always begin at steady state. At $t = 0$min the transcription rate of mRNA is increased, and then at $t = 300$min it is decreased again to its original value. This gives the transcription rate of mRNA the shape of a square wave with respect to time. The transcription rate of sRNA is fixed to $\alpha_R = 5\text{min}^{-1}$.

The three regimes simulated are:

1. Across the threshold: the systems start at steady state at $\alpha_M = 3\text{min}^{-1}$, the transcription rate is increased to $\alpha_M = 7\text{min}^{-1}$ and then decreased back to $\alpha_M = 3\text{min}^{-1}$ again.

2. Below the threshold: the systems start at steady state at $\alpha_M = 2\text{min}^{-1}$, the transcription rate is increased to $\alpha_M = 4\text{min}^{-1}$ and then decreased back to $\alpha_M = 2\text{min}^{-1}$ again.

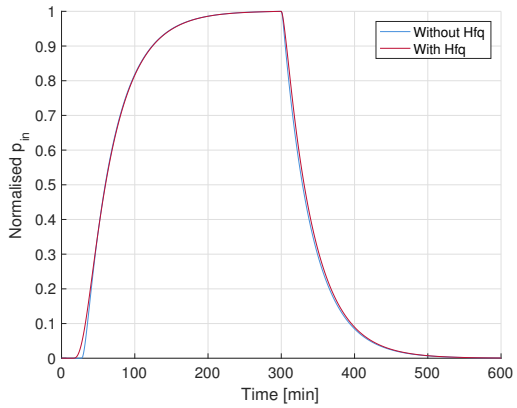
3. Above the threshold: the systems start at steady state at $\alpha_M = 6\text{min}^{-1}$, the transcription rate is increased to $\alpha_M = 8\text{min}^{-1}$ and then decreased back to $\alpha_M = 6\text{min}^{-1}$ again.

The responses to these steps up and down are normalised such that the maximum value of protein concentration is 1 and the minimum value is 0. The dynamics of the responses are then compared directly in Figure 3.23. The responses are very similar, which indicates a good agreement of the dynamics of the two models. This assumes that the systems are operating in a regime where regulation is efficient, i.e. in the case of the model with Hfq, there is neither too much, nor too little Hfq present. In this regime, as is shown in Figure 3.23, there is little to be gained by using the much more complex model with Hfq.

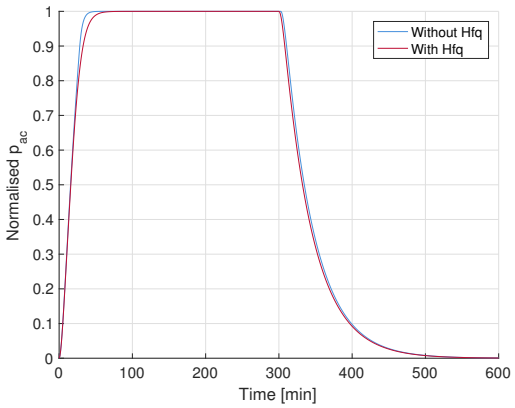
When crossing the threshold from below, p_{in} exhibits the temporal delay noted by Legewie *et al.* in [127] (see Figure 3.23a). This is seen as the delay in the rise of protein concentration after in step change in α_M. Interestingly enough, the responses of the activated protein is faster to rise in all cases.

The plot shown of the normalised p_{in} below the threshold, Figure 3.23c, skews the picture slightly. The total change between maximum and minimum in this case is about 0.6nM and the maximum of p_{in} in that plot is 0.8nM. This is to be expected as the systems are both operating below the threshold where there is very little mRNA in the system.

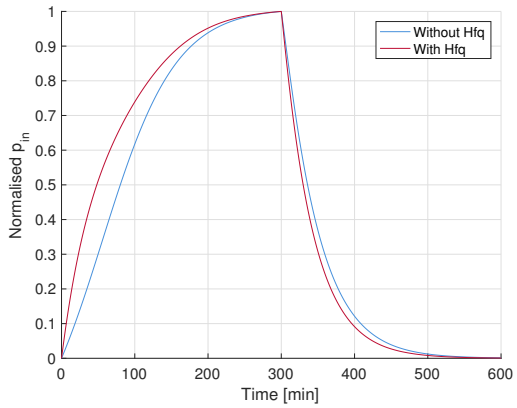
A similar argument is valid for Figure 3.23f; The systems are both now operating above the threshold, where the amount of mRNA available to translate hardly changes. The change shown in the normalised figure represents a change in steady state of less than 4% in the case of the model with Hfq and much less in the case of the model without.



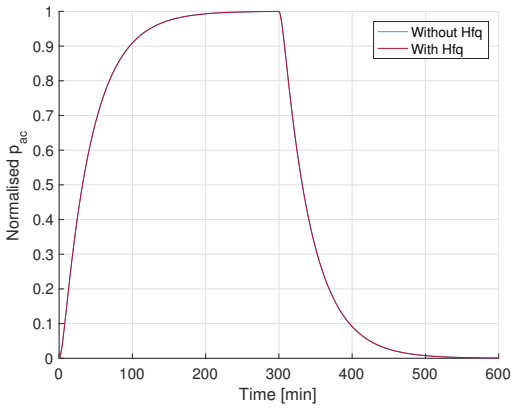
(a) Across Threshold



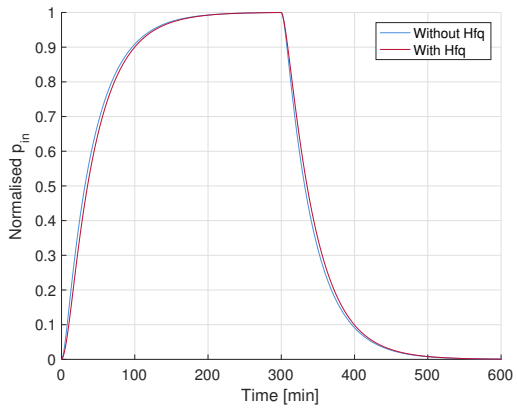
(b) Across Threshold



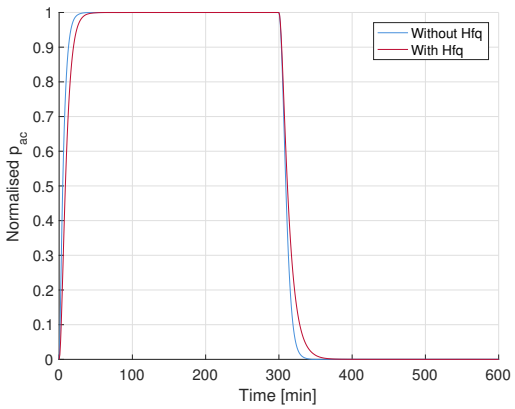
(c) Below Threshold



(d) Below Threshold



(e) Above Threshold



(f) Above Threshold

Figure 3.23 – A comparison of normalised step responses comparing the basic model of sRNA regulation detail in Equation 3.15 and the model including Hfq detailed in Equation 3.16. See text for details and discussion.

3.4 CONCLUSION

The main objective of this chapter is to compare a basic model of sRNA regulation with a more complex model including the Hfq protein and its interactions with RNA, to estimate via theoretical means whether, during the design and modelling of circuits incorporating sRNA regulation, it is important to include the Hfq dynamics.

Given the modelling presented above, the conclusion drawn is that the inclusion of Hfq dynamics in the ODE model of a system employing a single sRNA-target pair is not necessary. This conclusion is drawn as it is shown that, for a wide range of parameters, the symmetric model of sRNA regulation including Hfq behaves qualitatively similarly at steady state to the basic model without Hfq. The behaviour of the modelling including Hfq is improved by introducing asymmetric terms that ensure degradation of the target mRNA in all states, and a further extension including preferential binding order. Finally the responses of the two models are compared and found to be very similar, confirming that unless the system under study operates in a very extreme regime, the basic model of sRNA regulation, in both the case of inhibition and activation should suffice. This is a boon, as the simpler ODE model is much more easily manipulated analytically than the full model including Hfq.

In the case of stochastic modelling of the system, the inclusion of Hfq is necessary in particular, if the system being modelled separates the target mRNA and sRNA gene across several plasmids. The varying copy numbers of the plasmids is predicted to greatly soften the transition at the threshold and increase the overall noise observed in an experimental system. It is suggested that this may be the reason for the difference in results presented by Levine *et al.* in [129] and by Hussein and Lim in [102]. In the case where the target and sRNA genes are not separated and in the case of sRNA inhibition, the inclusion of Hfq in stochastic models should be considered, as the basic model exaggerates the noise peak below the threshold.

That said, to ensure reliable adherence to the behaviour described in this chapter, it is worth keeping a few things in mind. Firstly, the sRNA-target pair must be expressed at a level considerably higher than the background of regulating RNAs present in the host. This is almost certainly the case for any synthetic system as expression is usually high compared to the background, such as those presented in [102, 127, 129, 151, 211, 224]. In the case where two sRNA-target pairs are need in a circuit and therefore would be competing for Hfq, further modelling is required. This is an obvious direction for theoretical extension of this work. Competition for Hfq is briefly studied by Adamson and Lim in [1], but the results do not delve into behaviour for varying

transcription rates, which is of interest in the context of design of circuits, but their work is a solid starting point.

As synthetic systems usually express at relatively high levels, sequestration of sRNAs from their targets by an excess of Hfq is not likely to be an issue experimentally. This is reinforced by the natural negative feedback loop employed by Hfq [212]. If expression is so high that the quantity of Hfq in the cell becomes a limiting factor, it has been shown that synthetically expressing more can solve this problem [102]. As Hfq is usually abundant in a cell, entering into a regime where Hfq is a limiting factor probably also risks over burdening the host and therefore must be done with care in an experimental context.

As a route for experimental work, the noise profiles presented for sRNA regulation from data resulting from stochastic simulations can be verified using a set-up similar to those employed in [102, 127, 129], where the transcription rate of either sRNA or target is controlled with inducer while the other is held constant. Flow cytometry measurements of cultures in exponential growth phase would yield the data necessary, though measurements below the threshold might be a challenge due to low expression. This type of set-up could also be used to verify the effect of spreading the two genes across two separate plasmids.

4 LINEAR DESIGN FRAMEWORK FOR GENE REGULATORY NETWORKS

This chapter presents an application of classical linear design techniques introduced in Section 2.3 to the design of gene regulatory networks. The majority of this chapter is published in [93].

For decades control engineers have successfully been implementing feedback controllers for complex mechanical and electrical systems such as aircraft and sports cars. By incorporating feedback around systems, it is possible to improve their performance and robustness properties to meet pre-specified design objectives. Natural biological systems use feedback extensively for regulation and adaptation with [95, 124, 190, 204, 223] being just a few of the many examples, but apart from the most basic designs, there is no systematic framework for designing feedback controllers in Synthetic Biology. Progress has been made using feedback design techniques to implement DNA based control circuits, such as those in [46, 157, 225]. By closing the loop, the advantages of feedback, such as improved robustness and reduced sensitivity to noise, can be harnessed.

The question that arises is what types of feedback can be designed within a biological context and what such controllers look like. Previous works have implemented feedback at the genetic level in the seminal synthetic biological circuits: the toggle switch [76] and repressilator [66], but also at the protein level [42]. However, to date very few have designed specific controllers and the size of these synthetic systems is restricted to just a few genes [3]. In fact, the design of more complicated controllers such as the ones found in engineered mechanical or electrical circuits still remains a challenge [144], as well as the implementation of simple components such as a summing junction [157]. The use of accurate mechanistic models [91] in conjunction with the reliable design of feedback and controllers are key elements that will allow the size of synthetic systems to increase.

In this chapter we aim to design bio-controllers using ideas borrowed from Control Engineering, building on our previous work in [60]. The approach taken is to investigate what configurations would result in dynamics resembling traditional engineering controllers and use them to achieve a desired objective through optimisation.

Initially the problem is presented and the linear design framework is laid out and applied to classical examples such as positive and negative autoregulation, referred to as autoactivation

and autorepression respectively. The framework is developed to include promoters regulated by several Transcription Factors (TFs) of varying types. A block diagram framework is presented, which adheres to the same rules as the classical engineering block diagram framework. The biological topologies of phase lead and lag controllers are then presented and the phase lag controller is used in a design example presented with a proposed biological topology. Finally, the framework is extended to include sRNA Inhibition as it is modelled in Section 3.1.1.

4.1 PROBLEM SETUP

The problem setup here is based on the ODE modelling techniques presented in Section 2.2.3. The transcription rate functions in the cases of activation and repression as described in Equation 2.7 are used and protein expression is modelled as a single step using the timescale separation presented in Section 2.2.3.1.

For modelling and design in this chapter, it is worth noting the derivatives of Equation 2.7 with respect to the TF concentration, x, and how these are related, as this simplifies notation:

$$\frac{d\Gamma_A}{dx} = \frac{n_X \alpha_X K_X^{n_X} x^{n_X-1}}{(K_X^{n_X} + x^{n_X})^2} =: \gamma_A, \tag{4.1a}$$

$$\frac{d\Gamma_R}{dx} = -\frac{n_X \alpha_X K_X^{n_X} x^{n_X-1}}{(K_X^{n_X} + x^{n_X})^2} =: -\gamma_R, \tag{4.1b}$$

where the subscript X denotes the associated TF, x denotes the concentration of TF, K_X is the associated dissociation constant, n_X is the associated Hill Coefficient and α_X is the maximal rate of transcription or expression due to the associated operator.

Note that the derivatives are identical in form but for the sign which allows us to write:

$$\gamma_X := \frac{n_X \alpha_X K_X^{n_X} x^{n_X-1}}{(K_X^{n_X} + x^{n_X})^2},$$

where X can either be an activating or repressing TF. The function γ_X describes the derivative of the expression rate contribution from a promoter which is positive in the case of an activating TF and negative in the case of a repressing TF.

Having covered promoter regulation, the basal level of expression from a down regulated promoter is included in this work and is denoted α_0. Degradation is modelled as discussed in Section 2.1.10 and degradation rates are denoted δ_X, where X is the species being degraded.

It is worth remembering that the parameters α_X, α_0, K_X, n_X and δ_X and all the concentrations are non-negative, and thus $\Gamma_A, \Gamma_R, \gamma_X \geqslant 0$.

(a) Activated Expression (b) Repressed Expression

Figure 4.1 – These are simple mechanistic diagrams of gene expression for (a) activated expression, shown by the blue arrow with the TF A, and (b) repressed expression, shown by the red line with the bar with the TF R. They include transcription and translation as a single step, producing protein Y from the gene *Y*.

The mechanisms discussed above are now used to model the two simplest cases of regulation: a protein whose expression is activated and a protein whose expression is repressed.

4.1.1 *Activated Expression*

A protein Y, with concentration y, whose expression is activated by the TF A, has the following rate of change of concentration over time

$$\frac{dy}{dt} = \alpha_0 + \Gamma_A - \delta_Y y \tag{4.2}$$

where α_0 is the basal expression rate, the function Γ_A yields the contribution to expression by the activator A, and δ_Y is the degradation rate. In the term Γ_A, the dependence on the concentration on A is implied, but not included explicitly to simplify notation.

A diagram of this genetic circuit can be seen in Figure 4.1a, where the incoming activating TF A interacts with its promoter and regulates the expression of gene *Y* producing protein Y.

The letter Y is generally used to denote the output protein of the system following the standard engineering convention. This is what would be measured in a real biological system through for example tagging with a reporter protein, such as those discussed in Section 2.1.17.

4.1.2 *Repressed Expression*

Analogous to the case above of activated expression, the rate of change of the concentration of a protein, denoted y, whose expression is repressed by the TF R is

$$\frac{dy}{dt} = \alpha_0 + \Gamma_R - \delta_Y y \tag{4.3}$$

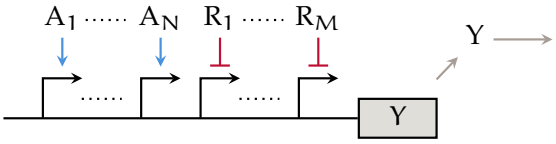


Figure 4.2 – Regulation of a gene by multiple TFs. It is assumed that the promoter order is not important, and the order chosen above is just the order in which the operator regions appear in Equation 4.4. Activation is represented by the blue arrow and repression is represented by red line with the bar. The gene regulated is denoted Y with the output protein Y.

where now the function Γ_R yields the contribution to expression by the repressor R. A diagram of this genetic circuit can be seen in Figure 4.1b, where the incoming repressing TF R interacts with its promoter and regulates the expression of gene Y producing protein Y.

4.1.3 *Expression Regulation by Multiple Transcription Factors*

As previously mentioned, it is common for the expression of a gene to be regulated by several TFs. In this case, it is assumed that these can be modelled through a product of the rate functions presented in Equation 2.7 as is done in [8, 104]. In such a way, a gene's expression can be regulated by an arbitrary number of TFs, yielding the ODE for the concentration of the output protein as follows:

$$\frac{dy}{dt} = \alpha_0 + \prod_{i=1}^{N} \Gamma_{A_i} \prod_{j=1}^{M} \Gamma_{R_j} - \delta_Y y, \tag{4.4}$$

where y is the concentration of the output protein Y and the products are over the N activating TFs, $A_1, \ldots, A_N$, and the M repressing TFs, $R_1, \ldots, R_M$, with their contributions to expression supplied by the activation and repression functions Γ_{A_i} and Γ_{R_j} respectively employing the functional forms in Equation 2.7.

Using the above equation, the regulation of gene expression by any finite number of TFs can be described. A genetic circuit diagram of this system can be seen in Figure 4.2, where the TFs interact with their specific operator regions and regulate the expression of the protein Y.

4.2 FREQUENCY DOMAIN REPRESENTATION

In order to apply classical linear control theoretical analysis and design methodologies to genetic circuits, the dynamics of activated and repressed expression must first be transformed into the frequency domain. To do so requires linearisation and Laplace transformation of the ODEs

Equation 4.2, Equation 4.3, and the more general case Equation 4.4. The methodology followed is that found in Section 2.3.

4.2.1 *Single Input Systems*

Starting with activated expression described in Equation 4.2 and denoting a derivative with respect to time using $\dot{-}$, first differentiate it with respect to the two protein concentrations: y, the concentration of output protein and a, the concentration of activator protein:

$$\begin{aligned} \frac{d\dot{y}}{dy} &= -\delta_Y, \\ \frac{d\dot{y}}{da} &= \gamma_A, \end{aligned}$$

where δ_Y is the degradation rate and γ_A is as defined in Equation 4.1, with x replaced by a. Keeping the state-space representation in Equation 2.14 in mind, y can be though of as both the state and output, and a can be thought of as the input. To that end, perform the coordinate transformations $\bar{y} = y - y^*$ and $\bar{u} = a - a^*$, where $\bar{y}$ and $\bar{u}$ are now small perturbations around the equilibrium values of y^* and a^*. Evaluating the above derivatives at the equilibrium and putting this on the form of Equation 2.14a yields

$$\frac{d\bar{y}}{dt} = -\delta_Y \bar{y} + \gamma_A^* \bar{u}, \tag{4.5}$$

where γ_A^* indicates the function γ_A evaluated at the equilibrium.

For the case of repressed expression, the method is identical to that of the activated expression. Beginning with Equation 4.3, first differentiate with respect to the protein concentrations: y for output protein and r for the repressing TF.

$$\begin{aligned} \frac{d\dot{y}}{dy} &= -\delta_Y, \\ \frac{d\dot{y}}{dr} &= -\gamma_R, \end{aligned}$$

where δ_Y is the degradation rate and γ_R is as defined in Equation 4.1. Performing the same coordinate transformations as above, $\bar{y} = y - y^*$ and $\bar{u} = r - r^*$, where $\bar{y}$ and $\bar{u}$ are now small

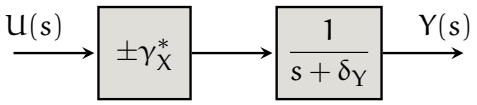


Figure 4.3 – Block diagram representation for expression of a single gene. The input is denoted $U(s)$ and the output is denoted $Y(s)$, and the blocks inbetween are a gain block with the constant gain $\pm\gamma_X$, which can be thought of as being linked to the strength of the promoter, and an 'expression' block, which varies in magnitude depending on s. Expression is either activated, with $X = A$ and the gain block takes the positive sign $(+)$, or repressed, with $X = R$ and the gain block takes the negative sign $(-)$.

perturbations around the equilibrium values of y^* and r^* and evaluating the above derivatives at the equilibrium yields

$$\frac{d\bar{y}}{dt} = -\delta_Y \bar{y} - \gamma_R^* \bar{u}, \tag{4.6}$$

where γ_R^* is the function γ_R evaluated at the equilibrium.

Laplace transforms of Equation 4.2 and Equation 4.3 yield transfer functions of a very similar form:

$$H_X(s) = \frac{Y(s)}{U(s)} = \frac{\pm\gamma_X^*}{s + \delta_Y} \tag{4.7}$$

where activated expression $(X = A)$ takes the positive sign $(+)$ and repressed expression $(X = R)$ takes the negative sign $(-)$. The functions $Y(s)$ and $U(s)$ are the Laplace transforms of $\bar{y}$ and $\bar{u}$ respectively.

This system can be represented by the block diagram seen in Figure 4.3. Note how the protein expression process is split across two blocks. The left block quantifies the interaction between the TF or input, and the promoter. This can be thought of as a gain block. The second block represents the expression of the protein. These are two of the fundamental blocks used in gene regulatory networks.

4.2.2 *Multiple Input System*

The linearisation of the multiple input system, where gene expression is regulated by a number of TFs, is of a similar form to the single input cases, though there are some important differences.

Beginning with Equation 4.4, first differentiate with respect to the output protein concentration, y:

$$\frac{d\dot{y}}{dy} = -\delta_Y.$$

This is identical to the case of regulation by a single TF. Now differentiate Equation 4.4 first with respect to an activating TF concentration, labelled a_p and then with respect to a repressing TF concentration, labelled r_q:

$$\begin{aligned}\frac{d\dot{y}}{da_p} &= \gamma_{A_p}\prod_{i=1}^{p-1}\Gamma_{A_i}\prod_{i=p+1}^{N}\Gamma_{A_i}\prod_{j=1}^{M}\Gamma_{R_j},\\ \frac{d\dot{y}}{dr_q} &= (-\gamma_{R_q})\prod_{i=1}^{N}\Gamma_{A_i}\prod_{j=1}^{q-1}\Gamma_{R_j}\prod_{j=q+1}^{M}\Gamma_{R_j}.\end{aligned}$$

Let us now construct the input vector

$$u = [a_1, a_2, \ldots, a_N, r_1, r_2, \ldots, r_M]^T$$

i.e. a vector of all regulating TF concentrations, whether activating or repressing. The dimension of the input now equals $m = N + M$. The state vector is just y and therefore has only a single dimension, so, remembering the linear state space representation given by Equation 2.14, the input matrix $\mathbf{B}^* \in \mathbb{R}^{1\times N+M}$ and has the form

$$\mathbf{B}^* = \left[\frac{d\dot{y}}{da_1}, \frac{d\dot{y}}{da_2}, \ldots, \frac{d\dot{y}}{da_N}, \frac{d\dot{y}}{dr_1}, \frac{d\dot{y}}{dr_2}, \ldots, \frac{d\dot{y}}{dr_M}\right]\Bigg|_{y^*,u^*} \tag{4.8}$$

where y^* is the equilibrium of the output with the equilibrium input vector u^* and the asterisk indicates that $\mathbf{B}$ is evaluated at equilibrium. Performing the change of variables $\bar{y} = y - y^*$ and $\bar{u} = u - u^*$, the form of the linearisation for generalised regulation is then

$$\frac{d\bar{y}}{dt} = -\delta_Y \bar{y} + \mathbf{B}^*\bar{u} \tag{4.9}$$

where $\bar{y}$ and $\bar{u}$ are now small perturbations to the equilibrium of the system.

Laplace transforming the linearisation yields the following equation:

$$Y(s) = \frac{1}{s+\delta_Y}\mathbf{B}^*U(s), \tag{4.10}$$

where $U(s)$ is now the Laplace transform of $\bar{u}$. The term $\mathbf{B}^*U(s)$ can be thought of as a weighted sum of the inputs as is demonstrated by the block diagram in Figure 4.4. The inputs are weighted by their corresponding term in the input matrix $\mathbf{B}^*$ and are collected at the summing junction before being fed to the gene, thereby regulating expression. In this way the promoters of a gene process the inputs prior to expression.

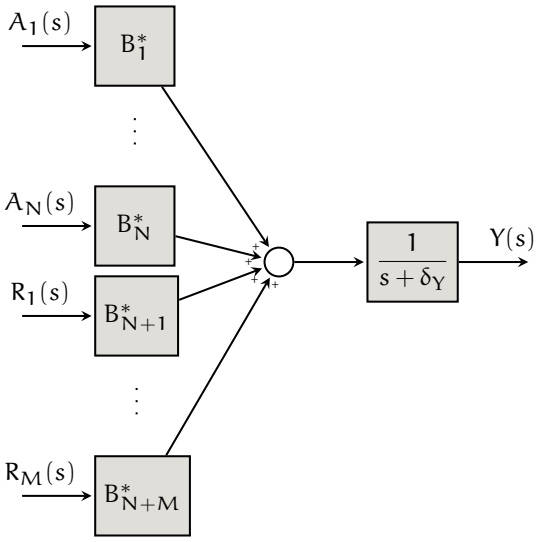


Figure 4.4 – A block diagram of protein expression regulated by multiple transcription factors. The functions $A_1(s)$ to $A_N(s)$ and $R_1(s)$ to $R_M(s)$ are the Laplace transforms of the input concentrations and the elements of $U(s)$. These inputs are weighted by their relevant element of the $\mathbf{B}^*$ matrix and then summed using the summing junction.

4.2.3 *Block Diagram Framework*

The analysis of the above systems motivates a block diagram framework for gene circuit analysis that facilitates the extraction of the transfer function of a system and enables the direct application of linear design principles. This section explains the basic rules of converting a standard genetic circuit diagram such as the ones seen in Figure 4.1 and Figure 4.2 into a block diagram and extracting the relevant transfer function.

1. First, the number of genes involved in the circuit is counted. An expression block with the transfer function $\frac{1}{s+\delta_i}$ pertains to each gene, where δ_i is the degradation rate of the expressed protein labelled with subscript i.

2. The inputs to each gene are then processed through a single or several promoter gain block(s) whose form depends on the types (activating and/or repressing) and number of input TFs going to a specific gene. The gains are calculated using Equation 4.1 in the case of a single input to a gene and Equation 4.8 in the case of multiple inputs.

3. Once processed through the promoter gain blocks the inputs are then summed and fed to the gene expression block, which then outputs the protein expressed at the specific gene.

4. The final step is to link up the genes using the topology provided by the genetic circuit diagram.

This methodology facilitates the composition of the block diagram from which the transfer function can be extracted using standard techniques.

4.2.4 *Example: Autorepression*

This example demonstrates the use of the framework in the context of protein expression with autorepression, also known as negative autoregulation. This is a commonly used system and is widely studied [16, 23, 62, 155, 179]. In this system, the gene codes for a protein that represses the expression of a gene.

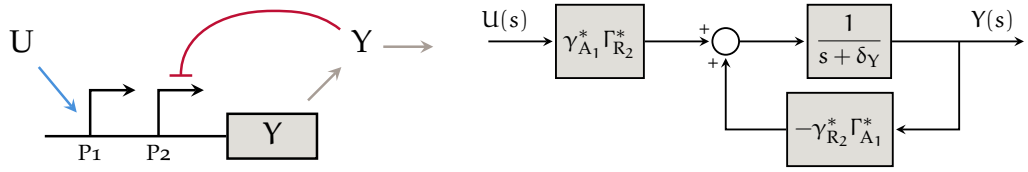
A simple biological example of this would be a gene encoding the protein TetR being driven by the P_{tet} promoter. When the gene is expressed, TetR is produced and the expression rate drops because of the interaction between TetR and the P_{tet} promoter. The specific mechanisms involved are introduced in Section 2.1.14.2.

Another activating TF and promoter have been added to the system to provide an external input and the gene circuit diagram is presented in Figure 4.5a. The block diagram representation is then developed based on the gene circuit diagram following the procedure laid out in Section 4.2.3. The external input is activating and interacts with promoter P1 resulting the label A_1 of its Γ function. The feedback from the expressed protein is repressing and interacts with P2 resulting in the label R_2. The block diagram is presented in Figure 4.5b. As can be seen, the gene has two inputs: U and the expressed protein Y. The promoter-input pairs each have a promoter gain block associated with them. This is calculated using Equation 4.8. They are then summed and fed into the gene expression block.

By standard methods described in Section 2.3.5, the transfer function from input to output of this system can now be calculated:

$$H(s) = \frac{\gamma^*_{A_1}\Gamma^*_{R_2}}{s + \delta_Y + \gamma^*_{R_2}\Gamma^*_{A_1}}.$$

As we know that $\delta_Y, \gamma^*_{A_1}, \gamma^*_{R_2}, \Gamma^*_{A_1}, \Gamma^*_{R_2} \geqslant 0$, certain characteristics of the system can be predicted from the transfer function. The 3dB bandwidth of this circuit has the value $\delta_Y + \gamma^*_{R_2}\Gamma^*_{A_1}$, which is greater than that of a single gene expression. This provides the faster response time as is observed in systems where autorepression occurs [8, 104, 179], and is presented in Figure 4.6.



(a) Gene Circuit (b) Block Diagram

Figure 4.5 – (a) A genetic circuit diagram of protein expression with autorepression. (b) A block diagram of protein expression with autorepression.

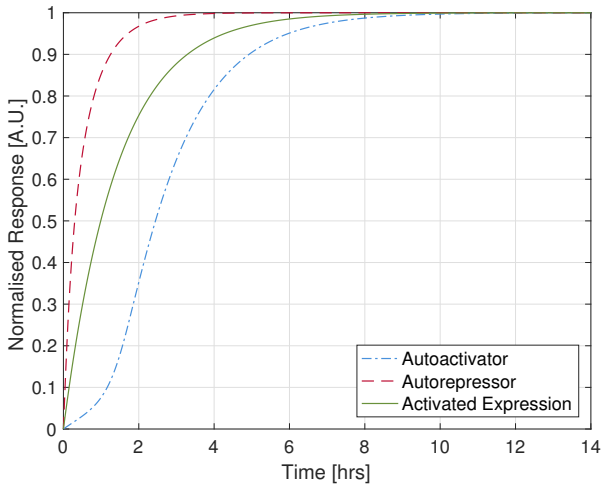


Figure 4.6 – Normalised step responses of the autorepressor and autoactivator are compared to each other and that of activated expression found in Section 4.1.1. As expected, the autorepressor is fastest, while the autoactivator is slowest.

To quantify the blocks in the block diagram, the equilibrium must be known. To find this, the ODE model of the system must be used. The ODE model derived using methods described in Section 2.2.3 yields:

$$\frac{dy}{dt} = \alpha_0 + \Gamma_{A_1}\Gamma_{R_2} - \delta_Y y, \tag{4.11}$$

where Γ_{A_1} provides the dependence on the activating input TF and Γ_{R_2} supplies the feedback of the output protein. This system has only one stable equilibrium in the positive quadrant regardless of parameter values, and calculating this for a certain set of parameter values and an input value allows the quantification of all blocks in the block diagram. This system can also be simulated and exhibits the desired reduced response time compared to single gene expression.

4.2.5 *Example: Autoactivation*

Autoactivation, otherwise known as positive autoregulation, occurs when a gene activates its own expression. This example demonstrates the use of the framework and is interesting as it gives insight into the possibility of a genetic integrator. The genetic circuit diagram can be seen in Figure 4.7a. An example of a biological system that would behave as such would be a gene encoding the activator protein RhaS with its associated promoter P_{rhaBAD} in front of it. As RhaS is expressed it promotes its own expression through interaction with its associated promoter. These are introduced in Section 2.1.14.4. A block diagram produced using the methods discussed is exhibited in Figure 4.7b. The input to output transfer function extracted from this block diagram is

$$H(s) = \frac{\gamma^*_{A_1}\Gamma^*_{A_2}}{s + \delta_Y - \gamma^*_{A_2}\Gamma^*_{A_1}}. \tag{4.12}$$

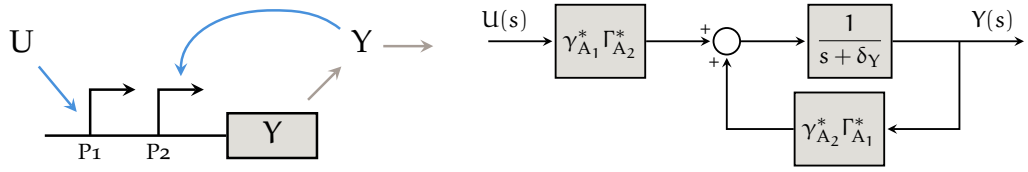
The position of the pole of this transfer function, where the function blows up, is at $s = -\delta_Y + \gamma^*_{A_2}\Gamma^*_{A_1}$. If $\delta_Y < \gamma^*_{A_2}\Gamma^*_{A_1}$, this pole is in the right half plane and causes instability. If $\delta_Y > \gamma^*_{A_2}\Gamma^*_{A_1}$, then the system is stable. If $\delta_Y = \gamma^*_{A_2}\Gamma^*_{A_1}$, the system has a pole at the origin providing the required $\frac{1}{s}$ term of an integrator.

The ODE model of the system provides further insight to its behaviour and takes the form

$$\frac{dy}{dt} = \alpha_0 + \Gamma_{A_1}\Gamma_{A_2} - \delta_Y y,$$

where Γ_{A_1} provides the dependence on the input TF and Γ_{A_2} supplies the feedback from the output protein. This system has up to three equilibria and only one of these is stable. It is around the stable equilibrium that the linearisation is performed. In the stable case of $\delta_Y > \gamma^*_{A_2}\Gamma^*_{A_1}$, the response time of the system is increased as the 3dB bandwidth has now been reduced to $\delta_Y - \gamma^*_{A_2}\Gamma^*_{A_1}$. This is in agreement with the literature [8, 104] and also with simulation of the ODE model as shown in Figure 4.6.

The implementation of a biological integrator would be a powerful tool as it would provide a means to track a reference with no steady state error. Given the form of Equation 4.12, one might think that the autoactivator provides a way of achieving this by setting $\alpha = \gamma^*_{A_2}\Gamma^*_{A_1}$. Unfortunately, this equality is not practically achievable, and even a close approximation risks the system ending up with a pole in the right half plane, resulting in unstable behaviour. The parameters in biological systems are far too uncertain to be able to tune to the accuracy required



(a) Genetic Circuit (b) Block Diagram

Figure 4.7 – (a) A genetic circuit diagram and (b) a block diagram of protein expression with autoactivation.

for the system to behave reliably as an integrator. In reality what would occur in this case, is that the linearisation of the system would break down.

4.3 DESIGNING A FEEDBACK CONTROLLER

Using the modelling framework set up in Section 2.2.3 and the frequency domain analysis and block diagram framework laid out in Section 4.2, the task of designing a genetic feedback controller can now be approached. The standard implementation of such controllers is through unit negative feedback and comparison with a reference producing an error, which is fed to the controller. First let us review possible controllers.

The *ideal* Proportional Integral Derivative (PID) controller has a transfer function of the form

$$K_{PID}(s) = K_P\left(1 + \frac{1}{T_I s} + T_D s\right)$$

where K_P, T_I and T_D are parameters defining the proportional, integral and derivative gains of the controller. Whilst PID controllers offer many advantages, it is discussed in the case of the autoactivator in Section 4.2.5, that building a genetic integrator is problematic. Furthermore, in order to build a PID controller with a non-zero K_D term, for any system, not necessarily a biological one, a filter is required in order to make the transfer function proper. This added complexity makes constructing an ideal genetic PID controller difficult at present.

Alternatively, phase lead and lag controllers can be considered. These are described by the transfer function

$$K_{PL}(s) = K\frac{s+z}{s+p}, \tag{4.13}$$

where K is the gain, $z > 0$ defines the position of the zero and $p > 0$ defines the position of the pole. When $|z| < |p|$, it is said to be a phase lead controller, and when $|z| > |p|$, it is a phase lag controller.

By tuning the parameters K, z and p of the phase lead or lag controller it is possible to change the frequency response of the system. Phase lead controllers typically

1. improve the phase margin (provide robustness to un-modelled time delays),
2. damp out transient oscillations and reduce settling time.

Phase lag controllers typically

1. improve disturbance rejection (provide robustness to noise at the input),
2. reduce steady state error.

An advantage of the lead-lag architecture is that the cascade connection of two such controllers allows the designer to shape the frequency response as desired. This means that over certain frequency ranges the benefits of a phase lead controller can be achieved, whilst in another frequency range the phase lag controller dominates. Such a controller is referred to as a lead-lag controller and takes the form:

$$K_{PLL}(s) = K\frac{(s + z_{lead})(s + z_{lag})}{(s + p_{lead})(s + p_{lag})},$$

where K is the gain. Of course there are conflicting design trade-offs: Phase lead controllers increase bandwidth making the system vulnerable to noise and phase lag controllers can deteriorate the transient response.

The block diagram framework laid out above allows the combination of different blocks in an attempt to achieve a certain transfer function. With this in mind the block diagram in Figure 4.8 is developed which has the transfer function:

$$H(s) = B_2^*\frac{s + \delta_X + \frac{\gamma_C^* B_1^*}{B_2^*}}{s + \delta_X}, \tag{4.14}$$

where γ_C^* is the promoter gain of the controller, the associated gains of the downstream promoters are elements of the downstream input matrix $\mathbf{B}^* = [B_1^*, B_2^*]$ derived from Equation 4.8.

This transfer function has the form of a phase lag controller, as $|z| > |p|$. A genetic circuit diagram of the phase lag controller is simple and takes the form seen in Figure 4.9a, where the error fed into the controller is denoted E and activates expression of an activating state protein X, which in turn activates a downstream promoter.

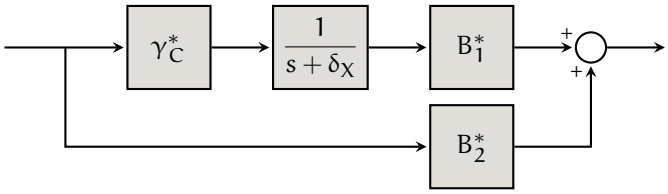
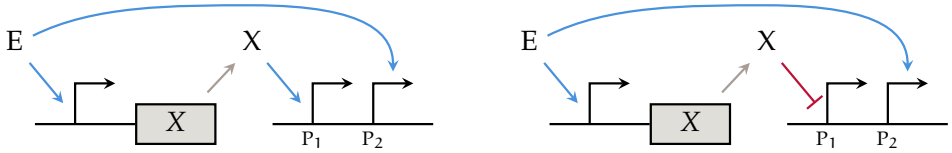


Figure 4.8 – Block diagram of a genetic phase lead (+) and phase lag (-) controller.



(a) Phase Lag Controller

(b) Phase Lead Controller

Figure 4.9 – Genetic circuit diagrams of genetic phase lag and phase lead controllers.

When the state protein X is a repressing transcription factor, on the other hand, the gain of the downstream promoter associated with the state protein $B_1^* < 0$. For this case, the transfer function is then on the form $|z| < |p|$, which is the equivalent of a phase lead controller. The genetic circuit diagram for this controller is presented in Figure 4.9b.

Both the phase lead and lag controllers are therefore implementable genetic controller modules which can be used separately or together in the design of larger circuits.

4.4 EXAMPLE

This section describes the detailed design of a genetic phase lag controller in feedback with a single gene acting as plant. The biological topology is achieved using a protease to supply the desired negative feedback. The parameter values found during the design process might be obtained by engineering ribosome binding sites, promoter strengths, plasmid copy number, as well as degradation rates of the species involved. It must be underlined that the engineering process of accurately tuning such an implemented system is not trivial with current available techniques.

The above described mathematical framework is used to complete this task, as well as the MATLAB optimisation toolbox, and the design objectives are to:

- Minimise the rise time of the closed loop system.
- Minimise steady state error.

- Retain a reasonable phase margin to ensure robustness.

4.4.1 *Modelling*

All Γ functions in the following model activated expression rates and are therefore of the form of Equation 2.7b. In addition to this, their dependency has been stated explicitly. For simplicity, they are just labelled by number. Using control engineering jargon, the plant, i.e. the system to be control, is the activated expression of a gene modelled as

$$\frac{dy}{dt} = \alpha_{Y0} + \Gamma_1(x) - \delta_Y y,$$

where $\Gamma_1(x)$ is the expression rate function with its dependency on the concentration of the activator X explicitly stated. The initial term, α_{Y0} is the basal expression rate of the gene encoding the protein Y and δ_Y is its degradation rate.

The controller used is a phase lag compensator as presented in Figure 4.8, which is put in cascade with the plant. The genetic phase lag controller requires a second promoter in front of the downstream gene to transfer information from the controller to the plant so that both the error and state protein can regulate it, as is seen in Figure 4.9a. This results in the following model for the full system:

$$\frac{dy}{dt} = \alpha_{Y0} + \Gamma_1(x)\Gamma_2(e) - \delta_Y y, \tag{4.15a}$$

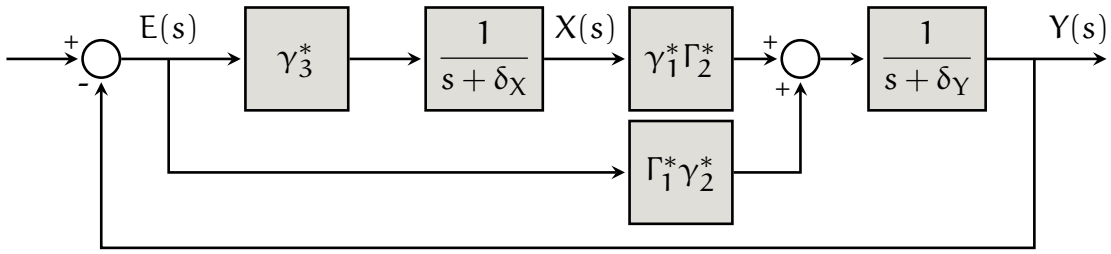
$$\frac{dx}{dt} = \beta_{X0} + \Gamma_3(e) - \delta_X x, \tag{4.15b}$$

where α_{Y0} and α_{X0} are basal expression rates of the plant (protein Y) and controller (protein X) respectively and δ_Y and δ_X are the degradation rates of the plant and controller respectively. It is through the error that unit feedback is supplied:

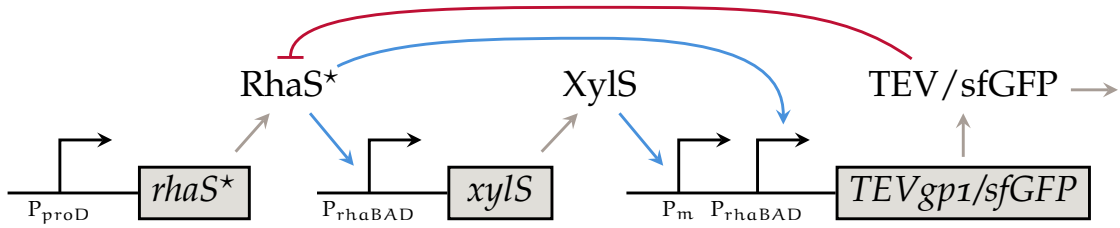
$$e = \begin{cases} y_{ref} - y & \text{if } e \geqslant 0 \\ 0 & \text{otherwise} \end{cases}$$

where the external reference to the system is y_{ref}.

A full block diagram of this model can be seen in Figure 4.10a, where the functions in the promoter gain blocks have been calculated using Equation 4.8. This diagram allows us to calculate



(a) DNA



(b) Genetic Circuit

Figure 4.10 – (a): Block diagram showing the theoretical design example with the phase lag controller in cascade with the plant around which unit negative feedback is applied. (b): Biological implementation of the designed circuit using a protease to supply negative feedback.

the transfer function from the error signal $E(s)$ to the output $Y(s)$, which can then be used to design the system. The full form of this open loop transfer function is

$$H(s) = \frac{Y(s)}{E(s)} = \Gamma_1^*\gamma_2^* \frac{s + \delta_X + \frac{\gamma_1^*\Gamma_2^*\gamma_3^*}{\Gamma_1^*\gamma_2^*}}{(s+\delta_X)(s+\delta_Y)}. \tag{4.16}$$

For later reference, it is worth noting that the zero of transfer function shown in Equation 4.16 is

$$z = -\delta_X - \frac{\gamma_1^*\Gamma_2^*\gamma_3^*}{\Gamma_1^*\gamma_2^*}. \tag{4.17}$$

4.4.1.1 *Uncontrolled System*

The uncontrolled system that is used as a benchmark for comparison is similar to the controlled system excepting the removal of the controller gene. The error is simply fed straight to the plant without being processed. The ODE model of this system is

$$\frac{dy}{dt} = \alpha_{Y0} + \Gamma_2(e) - \delta_Y y,$$

which is identical to Equation 4.15a with the exception of the removal of $\Gamma_1(x)$ representing the promoter that is regulated by the controller. The feedback is included through the error e.

4.4.2 *Optimisation Framework*

This section details the process used to achieve the design objectives for the system. This process derives the parameters that define the controller and a certain input-output response. The optimisation that is performed is formulated as follows:

$$\begin{aligned} \max_{\theta} \quad & H_\theta(0) \\ \text{s.t.} \quad & \mathbf{c}_{eq}(\theta) = \mathbf{0} \\ & \mathbf{c}(\theta) \leqslant \mathbf{0} \\ & \underline{\theta}_i \leqslant \theta_i \leqslant \overline{\theta}_i, \quad i = 1, \ldots, m \end{aligned}$$

The cost function used is the open loop low frequency gain of the open loop system given by:

$$H_\theta(0) = \frac{Y(0)}{E(0)} = \frac{\Gamma_1^* \gamma_2^* \delta_X + \gamma_1^* \Gamma_2^* \gamma_3^*}{\delta_X \delta_Y}. \tag{4.19}$$

The equality constraints are given by:

$$\mathbf{c}_{eq}(\theta) = \begin{bmatrix} \alpha_{Y0} + \Gamma_1^* \Gamma_2^* - \delta_Y y^* \\ \beta_{X0} + \Gamma_3^* - \delta_X x^* \end{bmatrix}.$$

These ensure that the optimisation is performed around the equilibrium of the ODE model of the system in Equation 4.15. As the Laplace transform and linearisation are performed about the equilibrium and the equilibrium is dependent on the parameters, this needs to be recalculated at every iteration of the optimisation.

The inequality constraints on the system aid the placement of the zero of the transfer function (4.17) with the aim of pushing it to the left, far away from the origin, without degrading the phase too much. The constraint is given by

$$\mathbf{c} = \lambda \delta_Y - \delta_X - \frac{\gamma_1^* \Gamma_2^* \gamma_3^*}{\Gamma_1^* \gamma_2^*},$$

where $\lambda \delta_Y$ fixes an upper bound on the position of the zero in the transfer function.

The decision variables are

$$\theta = \begin{bmatrix} y^* & x^* & \alpha_2 & \alpha_3 & \delta_X \end{bmatrix}^T,$$

Table 4.1 – The bounds on the decision variables employed in the optimisation in Section 4.4.2.

Variable	Lower bound	Upper bound
y^*	0 nM	100 nM
x^*	0 nM	100 nM
β_2	0 nMh^{-1}	15 nMh^{-1}
β_3	0 nMh^{-1}	15 nMh^{-1}
α_C	0.7 h^{-1}	13 h^{-1}

where the first two y^* and x^* are the equilibrium, α_2 is the strength of the promoter at the plant with which the error signal interacts, α_3 is the strength of the promoter at the controller with which the error signal interacts and δ_X is the degradation rate of the controller protein X. The bounds on these decision variables are given in Table 4.1, and have been chosen so as to be biologically feasible.

The optimisation is then performed with the parameter values given in Table 4.2 held constant. The results produced a system with the equilibrium at $(y^*, x^*) = (0.89\text{nM}, 0.46\text{nM})$ with the parameters $\alpha_2 = \alpha_3 = 15\text{nMh}^{-1}$ and $\delta_X = 1.85\text{h}^{-1}$. The low frequency gain of the controlled system is 21.5dB higher than in uncontrolled case and the controlled system has a phase margin of 84.6°.

The bode plots of the controlled and uncontrolled systems can be seen in Figure 4.11 and step responses to a 10% increase in reference can be seen in Figure 4.12. The reduced steady state error can clearly be seen in the step response, and the other obvious effect of the phase lag controller is the slight overshoot and the underdamped response. The system has a reasonable phase margin and the rise time of the system has also been reduced.

An obvious next line of inquiry would be to investigate how the linearisation compares to the nonlinear system. Using Lyapunov theory [116, 161] and sum-of-squares programming [162] the worst case difference (over all initial conditions) between the two approaches was shown to be small. Figure 4.12 compares simulations of the linear and nonlinear models and shows that they are in good agreement. For specific instances see Figure 4.13, where the nonlinear system has been subjected to perturbations of an order of magnitude greater than those valid for the linear analysis. The perturbations to the chosen parameters are ±20% and ±50%.

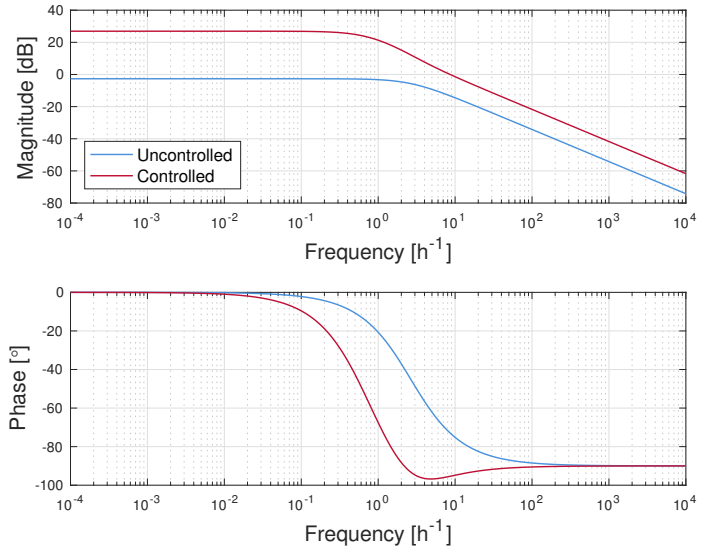


Figure 4.11 – Bode magnitude and phase plots showing both the controlled and uncontrolled system.

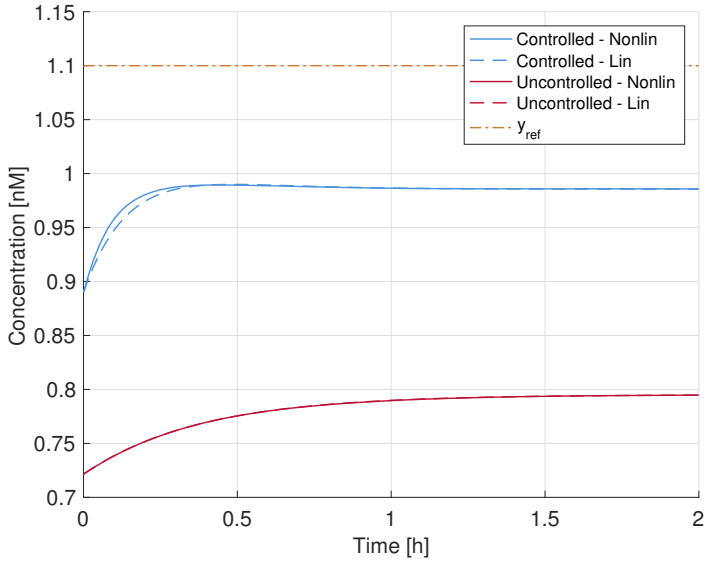
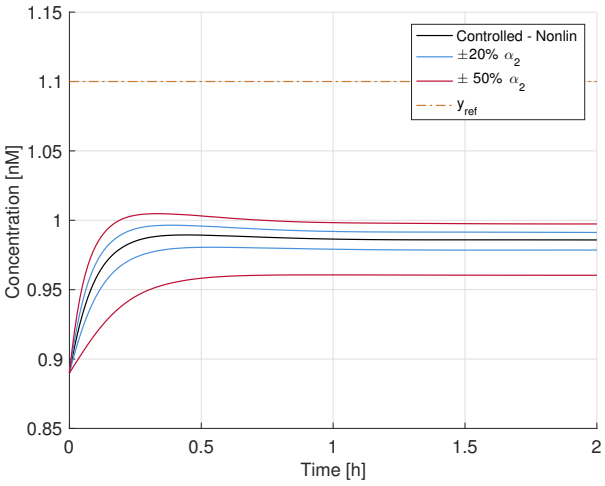


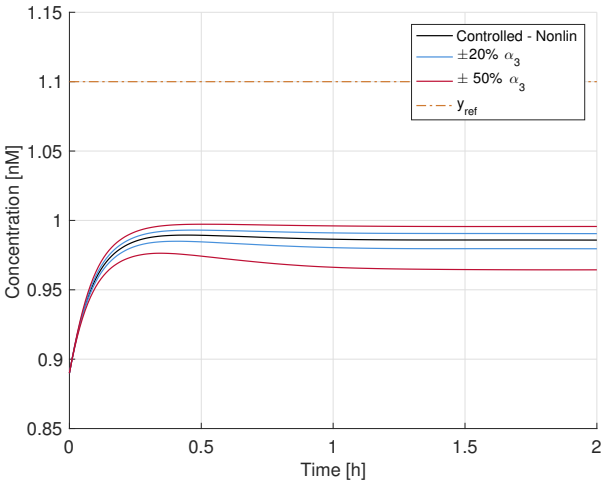
Figure 4.12 – Step responses of the controlled and uncontrolled systems both linear and nonlinear to a 10% increase in reference, y_{ref}.

Table 4.2 – The the values of the constant parameters in the optimisation in Section 4.4.2.

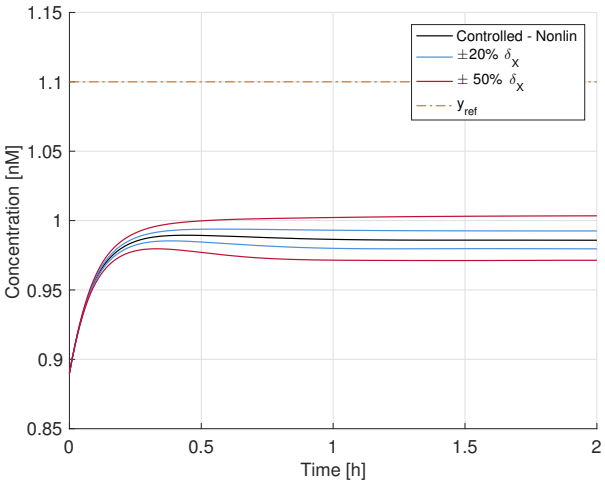
Parameter	Value	Parameter	Value
β_{P0},β_{C0}	0.15 nMh^{-1}	α_P	0.7 h^{-1}
β_1	15 nMh^{-1}	y_{ref}	1 nM
K_1,K_2,K_3	0.5	λ	5
n_1,n_2,n_3	2		



(a) Perturbations to α_2



(b) Perturbations to α_3



(c) Perturbations to δ_X

Figure 4.13 – Step response of the optimised and controlled example with large deviations to the designed parameters.

4.4.3 *Biological Implementation*

A biological implementation of this theoretical system, consisting of three genetic modules is proposed; see Figure 4.10b. The *TEVgp1/sfGFP* gene encodes a fusion polypeptide between the *Tobacco etch virus* P1 protease (TEV protease; the plant) [7] and sfGFP [164]. Expression of this gene is induced through two promoters: the meta-cleavage pathway operon promoter (P_m) from *Pseudomonas putida*, which is activated by the activating TF XylS in the presence of benzoate or m-toluate [143]; and the L-rhamnose inducible promoter (P_{rhaBAD}), which is activated by the RhaS [79, 204, 206].

TEV protease is a cysteine protease with a molecular mass of 49 kDa [39, 40]. Its native role in the life cycle of the virus is to cleave the large 346 kDa polyprotein at specific sites into its mature protein products [7, 193]. The consensus recognition site for this enzyme is the sequence of amino acids: Glu-Asn-Leu-Tyr-Phe-Gln-(Gly/Ser), with proteolysis occurring between the Gln and Gly/Ser residues [40]. It is also regularly used to direct proteolysis of various targeted proteins [109, 148]. To provide the desired negative feedback, the TEV-protease recognition site is inserted into a region of RhaS. Specifically, the cleavage site is inserted at residue 167 of RhaS, which lies between the N-terminal domain responsible for L-rhamnose and dimerisation and the C-terminal domain responsible for DNA and σ^{70} binding [218]. This should ensure minimal disruption to the function of RhaS and accessibility of the recognition site to TEV protease.

The *rhaS⋆* gene is the source of this engineered version of the activator RhaS (denoted RhaS⋆ to indicate that it has been engineered) and is expressed constitutively at a high level using the engineered *proD* promoter (P_{proD})[52]. The *xylS* gene is the controller module and expresses the activator XylS, through induction of P_{rhaBAD} by the unprocessed RhaS⋆ variant. The reference signal input to the system, mathematically denoted y_{ref}, can be thought of as the activator RhaS⋆ prior to digestion.

Once constructed and tested, this system would be optimised through parameter estimating, changing and tuning parts and so on, until the desired output is reached.

4.5 EXTENSION TO INCLUDE SRNA INHIBITION

This section extends the linear design framework presented in this chapter to include the sRNA inhibition mechanism as it is modelled in Section 3.1.1. An equivalent approach can be employed to include the sRNA activation mechanism modelled in Section 3.1.2.

Here, as opposed to previous sections, transcription and translation must be modelled as separate processes. To this end the model describes the output protein concentration, the target mRNA concentration and the sRNA concentration. For simplicity, it is chosen that the transcription of both mRNA and sRNA is activated by two TFs whose concentrations are u_1 and u_2 respectively. These act as the inputs to the system. The ODE model is then:

$$\begin{aligned}
\dot{y} &= \alpha_y m - \delta_y y, \\
\dot{m} &= \alpha_{m0} + \Gamma_{A1} - \delta_m m - K m r, \\
\dot{r} &= \alpha_{r0} + \Gamma_{A2} - \delta_r r - K m r,
\end{aligned} \tag{4.20}$$

where y is the concentration of output protein, m is the concentration of target mRNA and r is the concentration of sRNA, α_y is the translation rate, α_{m0} and α_{r0} are the basal rates of transcription from the promoters controlling mRNA and sRNA respectively. The transcription due to activation by the input proteins with concentration u_1 and u_2 is governed by Γ_{A1} and Γ_{A2} respectively, δ_y, δ_m and δ_r are the degradation rates of output protein, mRNA and sRNA respectively and K is the rate of heteroduplex formation between mRNA and sRNA.

The system is now linearised around an equilibrium denoted $x^* = [y^*, m^*, r^*]$. This is done in terms of perturbations around the equilibrium $\bar{x} = x - x^*$ for the states and $\bar{u} = u - u^*$ for the input vector $u = [u_1, u_2]$:

$$\begin{aligned}
\dot{\bar{y}} &= \alpha_y \bar{m} - \delta_y \bar{y}, \\
\dot{\bar{m}} &= \gamma_{A1} \bar{u}_1 - (K r^* + \delta_m)\bar{m} - K m^* \bar{r}, \\
\dot{\bar{r}} &= \gamma_{A2} \bar{u}_2 - (K m^* + \delta_r)\bar{r} - K r^* \bar{m}.
\end{aligned}$$

Laplace transforming the linearised equations then yields:

$$\begin{aligned}
Y(s) &= \frac{\alpha_y}{s + \delta_y} M(s), \\
M(s) &= \frac{\gamma_{A1} U_1(s) - K m^* R(s)}{s + K r^* + \delta_m}, \\
R(s) &= \frac{\gamma_{A2} U_2(s) - K r^* M(s)}{s + K m^* + \delta_r},
\end{aligned}$$

where $Y(s)$, $M(s)$, $R(s)$, $U_1(s)$ and $U_2(s)$ are the Laplace transforms of the output protein concentration, the mRNA concentration, the sRNA concentration, and the input TF concentration u_1

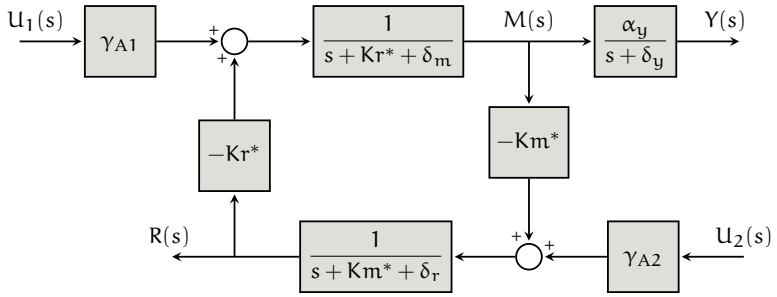


Figure 4.14 – Block diagram of open loop sRNA inhibition mechanism

and u_2 respectively. From these the transfer functions from the inputs $U_1(s)$ and $U_2(s)$ to the output $Y(s)$ can be derived:

$$H_1(s) = \frac{Y(s)}{U_1(s)} = \frac{\alpha_y \gamma_{A1}(s+\delta_{r*})}{(s+\delta_y)(s-s_+)(s-s_-)},$$
$$H_2(s) = \frac{Y(s)}{U_2(s)} = \frac{-K\alpha_y m^* \gamma_{A2}}{(s+\delta_y)(s-s_+)(s-s_-)},$$

where $H_1(s)$ and $H_2(s)$ are the transfer functions to $Y(s)$ from $U_1(s)$ and $U_2(s)$ respectively, and

$$s_\pm = -\frac{1}{2}(\delta_{m*}+\delta_{r*}) \pm \frac{1}{2}\sqrt{d},$$
$$d = (\delta_{m*}-\delta_{r*})^2 + 4K^2 m^* r^*,$$
$$\delta_{m*} = \delta_m + Kr^*,$$
$$\delta_{r*} = \delta_r + Km^*.$$

This can be represented by the block diagram Figure 4.14. This shows that there is a positive feedback loop involved in sRNA inhibition. This makes sense intuitively; more mRNA means more mRNA free to be translated.

With the required extension to the block diagram framework, the frequency domain representation for sRNA inhibition and the equivalent for sRNA activation can be used in a similar manner to the design example above to design GRNs including sRNA regulation.

4.6 CONCLUSION

The aim of this chapter is to adapt the classical linear design framework to the context of gene regulatory networks. Standard modelling techniques used to study genetic circuits are introduced, linearised and Laplace transformed resulting in the development of such a framework focussing on the biological mechanisms of transcriptional control. The biological topologies of

phase lead and lag controllers are presented. These are well known in the literature as incoherent and coherent feedforward motifs, and are found in natural systems [9, 135]. It is then shown how the use of such a framework facilitates the design of feedback in the context of gene regulatory networks through an example implementing the phase lag controller. The example is designed and optimised using standard techniques and an implementable biological topology for the designed circuit is proposed. Finally the framework is extended to include the sRNA inhibition mechanism as modelled in Section 3.1.1, which introduces a few new types of block in the block diagram framework.

An equivalent approach can be used to extend the model to include sRNA activation and also to include inducers, which act through gain blocks much like those seen in the case of activated and repressed expression.

The obvious next step would be the experimental implementation of the designed circuit. Though, the initial construction of the circuit might take up to a month, the task of extracting the step response of the system is a considerable challenge and has not been done before. This would require the system to be at steady state before being perturbed by some means. In a biological context, what this means is not necessarily completely clear. I think this would require cells containing the circuit to be held in exponential phase till the distribution of fluorescence across the population of cells becomes stationary. Only once this has occurred can the cells then be perturbed. The perturbation presented in the example is simple enough to implement in this case, as it is just an increase is the expression of the activator RhaS.

A second avenue of further research would be to locate the phase lead and lag controller motifs in natural regulatory network and search for negative feedback around the motifs. The aim of this would be to determine whether biological systems use these controllers in a similar manner to way to its use in engineering.

5 MODELLING NEGATIVE FEEDBACK CIRCUITS

To engineer gene regulatory networks, a threshold must be crossed whereafter it is possible to model and predict the behaviour of circuits prior to implementation. The two seminal papers in the field of Synthetic Biology detailing the Toggle Switch [76] and the Repressilator [66], and the seminal paper on the Autorepressor, also referred to as the negative autoregulator [23] are early examples of this. In these three basic examples simple modelling, both stochastic and deterministic is used to understand and enrich the experimental data yielded from the systems studied. Despite these early successes, progress since has been slow and the size and complexity of genetic circuits studied has not increased greatly [3].

The aim of this chapter is to approach this challenge and build on it through modelling and comparing the simplest forms of negative feedback possible in gene regulatory networks. The Autorepressor is used as a model system as it is well studied [23, 62, 155, 179]. A further three novel circuits are presented. The metrics of the circuits presented are aimed specifically to be of a type that is easy to test experimentally, such as dose responses and noise profiles.

The approach used for each circuit is to present the biological implementation of the circuit early on in the process so that this can guide the modelling and parameter choices. All four circuits described consist of well-known and documented biological parts which should assist in the modelling process, increasing the predictive power of the models.

The chosen repressor is probably the best understood Transcription Factor (TF) mechanism: the Tetracycline Repressor (TetR) (see Section 2.1.14.2). The chosen activator is Rhamnose Activator (RhaS), which has been proven as a reliable tight activator that even has a non-metabolisable inducer [112] (see Section 2.1.14.4). A novel feature of these circuits is the incorporation of sRNA regulation.

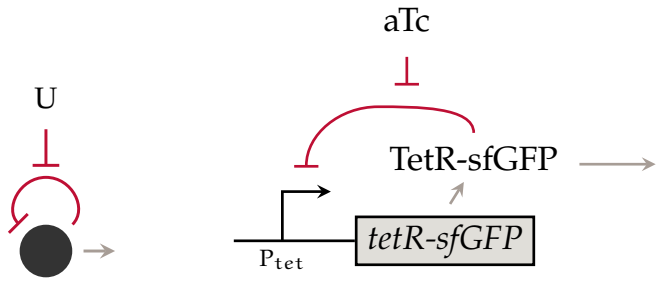
The use of sRNAs in larger designed gene circuits provides potential benefits as they are fast acting and come at a low energy cost to a host. This is due to the fact that they are short functional RNAs and therefore only require transcription. The sRNA inhibition mechanism, as opposed to the activation mechanism, is used as two protocols describing the reliable implementation of sRNA inhibition with a non-natural target are available [134, 224]. The examples of synthetic systems using sRNA activation employ a native activating sRNA targeting a section of a native target fused to GFP [103, 182]. Though it is most likely possible to design sRNA activation targets that are

non-native, it was decided to avoid this challenge as a means of de-risking the project. Based on the in-depth modelling performed in Chapter 3, modelling of similar circuits employing sRNAs is straight forward.

Four circuits are presented; the first, the Autorepressor is used as a starting point. The second circuit, sRNA Circuit I is based on the Autorepressor but includes an sRNA inhibition mechanism. The third circuit, sRNA Circuit II provides the activator RhaS with negative feedback through an inhibiting sRNA and the final circuit, the Repressor-Activator Circuit provides negative feedback through another TF. In this way, the circuits presented represent negative feedback at three levels: through autoregulation where the gene acts on itself, through an sRNA, and through another TF.

Models of the circuits are first described by a set of biochemical reactions guided by the presented biological implementation. The reaction networks are then translated into two ODE models: a reduced model employing Hill Functions to model the interactions between TFs, inducers and genes, and a higher order model that explicitly models each reaction in the biochemical model. Parameters used to study these models are then carefully selected from the literature and the steady state behaviours of the models are predicted with respect to their inputs. Interestingly enough, in some cases this process shows rather large discrepancies between the reduced Hill Function models and the higher order models. The circuits are then simulated stochastically using the Gillespie Algorithm to produce expected noise profiles of the circuits. Various implementation architectures are taken into account at this point, such as gene copy number fluctuation and the possibility of having the two genes of a circuit on separate plasmids. Modelling gene copy number fluctuation is important as it has a notable effect on the expected noise profiles. It is found, in the case of circuits employing sRNA inhibition in particular, that a two plasmid set-up is much noisier than if the system is on a single plasmid. On the other hand, the Repressor-Activator Circuit, consisting of two TFs does not exhibit a substantial increase in noise as a result of the two plasmid set-up. It is suggested that the translation and the longer half-lives of the TFs act as an additional filter to the noise resulting from gene copy number fluctuations. Finally, the step response dynamics of the four circuits are compared.

To follow on from the modelling performed in this chapter, the genetic circuits studied here were then implemented biologically and studied through experiment. The details of these experiments and the results are described in Chapter 6.



(a) Min. Rep. (b) Biological Circuit

Figure 5.1 – (a) The minimal representation of the Autorepressor. The black dot represents expression of a TF, the barred red lines represent repression or inhibition. The inducer is labelled U and inhibits the expressed TF and the grey arrow represents an output of the system. A more detailed biological diagram is presented in (b). This shows the biological parts used in the circuit constructed for this project, with the actual names of the gene and inducer used.

5.1 AUTOREPRESSOR

The Autorepressor, also commonly referred to as the negative autoregulator, is a common genetic motif consisting of a gene encoding a protein repressing the expression of that same gene. It is well studied in the literature [9, 16, 23, 62, 155, 179] and is ubiquitous in biological systems as a means of controlling expression of repressors in both prokaryotes [179] and eukaryotes [126]. As a standard regulatory unit, the Autorepressor functions as a benchmark in this project, to which other genetic circuits and motifs are compared. In addition to this, as the Autorepressor has been studied to such depth, it is an ideal case study of the combination of modelling and experimental techniques crucial to the field of Synthetic Biology.

As such, the aim of this section is to present the Autorepressor implemented for this project and then to introduce, compare and discuss a number of different modelling techniques.

The topology of the Autorepressor is presented through a minimal representation in Figure 5.1a. For large systems more classical circuit diagrams such as that seen in Figure 5.1b become cumbersome and are not clear. The minimal representation presents an easily understandable visual of the circuit topology without going into detail about the various biological parts.

5.1.1 *Biological Implementation*

The implemented Autorepressor topology with biological parts is exhibited in Figure 5.1b and consists of a gene encoding a fusion protein, consisting of the TetR fused to super-folder Green Fluorescent Protein (sfGFP), with the *tet*-promoter (P_{tet}) upstream of it. The fusion protein is

constructed by removing the Stop Codon from the end of *tetR* and then inserting the coding sequence for sfGFP immediately after the penultimate codon of *tetR*. More can be read about TetR in Section 2.1.14.2 and sfGFP in Section 2.1.17. Translation is controlled by an optimised RBS produced by the RBS Calculator [184]. The *tet*-repressor system can be induced experimentally by Anhydrotetracycline (aTc), an inefficient antibiotic that acts by interfering with the ribosome. The usual range of induction for aTc is between 0 ng/ml and 100 ng/ml [125, 132]. The plasmid backbone used is pJ404, a commercial backbone from DNA2.0, which provides resistance to ampicillin and its analog carbenicillin. The origin of replication is pBR322, which has a copy number of about 15-20 [194, p.230]. This circuit is very similar to those implemented in [16, 23, 62, 155, 179].

5.1.2 *Basic Biochemical Model*

Given the experimental implementation of the Autorepressor described in the previous section, the next step in studying it is to develop a biochemical model of the system. Here, the reactions that need to be taken into account are considered.

As described in Section 2.1.14.2, TetR is only active in dimer form and dimerises with a very high affinity [18, 97]. Assume therefore that TetR only exists in dimer form, and therefore the dimerisation reaction is neglected [91]. Assume then that the free TetR dimer is the only species that can reversibly interact with the operator region at the gene. This is reasonable as the association rate between the TetR-aTc complex and the operator region at the gene is low [125]. The half-life of the complex between TetR and the gene has been experimentally measured to be about 12 min [97]. Now assume that TetR reversibly binds to the inducer aTc. Including the necessary degradation reactions, the complete model is then:

$$\begin{aligned} M &\xrightarrow{\alpha_P} M + P, & P &\xrightarrow{\delta_P} \emptyset, \\ G &\xrightarrow{\alpha_M} G + M, & PU &\xrightarrow{\delta_P} \emptyset, \\ P + U &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PU, & PG &\xrightarrow{\delta_P} G, \\ P + G &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} PG, & M &\xrightarrow{\delta_M} \emptyset, \end{aligned} \tag{5.1}$$

where the species P, M, G and U are the free TetR dimer protein, the mRNA, gene and inducer respectively, the complexes PU and PG are the TetR-inducer and TetR-gene complexes respectively, the rate of translation is α_P, the rate of transcription is α_M, and the rate of degradation of species

including the repressor protein is δ_P and the degradation rate of the mRNA is δ_M. The pair k_1, k_{-1} are the forward and reverse rate constants of the interaction between the inducer and repressor protein, and k_2, k_{-2} are the forward and reverse rate constants of the interaction between the gene and repressor protein. The inducer U is the external input to the system, and is therefore fixed.

All the species including repressor protein degrade (P, PU and PG), but note that conservation of the gene is ensured. The total amount of gene is $G_{tot} = G + PG$, where G_{tot}, the total gene copy number, is constant.

For the biological implementation of this system, the experimental output is a fluorescence intensity measurement, which is used as a proxy of the concentration of expressed protein. This means that for the above model, the output is a combination of all the species including the repressor protein, as the repressor protein is tagged with the fluorescent protein sfGFP:

$$Y = P + PE + PG. \tag{5.2}$$

In effect, what is being observed is the TetR-sfGFP fusion protein in a number of different states.

5.1.3 *ODE Model with Hill Function*

The most common model of the Autorepressor is a single state model employing a Hill Function to model the inhibition mechanism [8, 23, 179]. The previously published models do not include the interaction with inducer, but this is included here as it is the most straight forward experimental input to the system. The single state model employed here, is as follows:

$$\dot{y} = \alpha g_{tot} \frac{K_G(K_I + u)}{K_G(K_I + u) + K_I y} - \delta_P y, \tag{5.3}$$

where y is the concentration of repressor protein and u is the concentration of inducer, $\alpha = \frac{\alpha_P \alpha_M}{\delta_M}$ is the maximal expression rate, g_{tot} is the total concentration of gene, and δ_P is the degradation rate of the repressor protein, $K_G = \frac{k_{-2}}{k_2}$ and $K_I = \frac{k_{-1}}{k_1}$ are the dissociation constants of the interactions between the repressor and gene, and the repressor and inducer respectively.

The model includes all reactions in Equation 5.1 by using some mathematical tricks. Expression is modelled in one step using the time scale separation argument presented in Section 2.2.3.1, which yields the above expression for α. The interactions between the repressor and the gene, and the interactions between the inducer and the repressor, both presented in the chemical re-

actions in Equation 5.1, are modelled using Hill Functions. How this is done is presented in Section 2.1, which results in the expression function on the form of Equation 2.11. Here, the example of the literature is followed [23, 179] and the Hill Coefficient is set to $n = 1$.

To begin with, the steady state protein concentration as a function of inducer is calculated explicitly, and found to be:

$$y^{ss}(u) = \frac{K_G(K_I + u)}{2K_I}\left[\sqrt{1 + \frac{4\alpha g_{tot} K_I}{\delta_P K_G(K_I + u)}} - 1\right], \tag{5.4}$$

where the superscript ss denotes the steady state. As this solution is relatively simple, some basic analysis can be performed with ease. In the limit of large u, $y^{ss} = \frac{\alpha g}{\delta}$, which is the value expected in the case where there is no autorepression. This steady state estimate is also true in the case of the limit of large K_G, which is the mechanistic equivalent of very weak or no binding between TetR and the operator regions. Again, this is the equivalent of no negative feedback, so this result agrees with what is expected intuitively.

5.1.4 *Choosing Parameters*

This section discusses the choice of parameters for the Autorepressor in particular, but this discussion generalises to all circuits modelled in this chapter. To begin with, the simpler parameters can be dealt with; the concentration of gene is $g_{tot} = 28.2\text{nM}$, which assumes a gene copy number of 17, consistent with the pBR322 *ori*-region in the implemented system [194, p.230]. The only effect assumed to remove protein from these systems is dilution, which is consistent with the literature [23, 62, 179]. This sets the half-life of the protein to the doubling time of the cell. Assume that this is about 28min, yielding $\delta_P = 0.025\text{min}^{-1}$. Assume that the degradation rate of mRNA is ten times this: $\delta_M = 0.25\text{min}^{-1}$, which gives the mRNA a half-life of about 2.8min, consistent with [87]. The transcription and translation rates are set to $\alpha_M = \alpha_P = 1\text{min}^{-1}$, which is a reasonable biological value and yields the overall expression rate of $\alpha = 4\text{min}^{-1}$.

The question turned to now is how to model the *tet*-repressor mechanism. The rich literature provides a lot of insight into parameter values involved. Initially the dissociation constants of the two interactions involving TetR are selected. These are the dissociation constants for the interaction with the operator region of the gene, K_G, and the interaction with the inducer aTc, K_I.

To assist in comparison with the literature, a model of repressed expression is constructed, for which the form of the dose response is well known [132]. The biological implementation of

this would have TetR repress the expression of a downstream reporter protein and its topology is presented in Figure 4.1b.

The ODE model of repressed expression of a gene used here is:

$$\dot{y} = \frac{\alpha g_{tot} K_G (K_I + u)}{K_G (K_I + u) + K_I p_{tot}} - \delta_P y, \tag{5.5}$$

where y in this case is the concentration of reporter protein, p_{tot} is the concentration of the repressor TetR, α, g_{tot}, and δ_P are the expression rate, gene concentration and degradation rate of the reporter protein and K_G and K_I are the dissociation rates of the interaction between TetR, and the operator region at the reporter gene and aTc respectively. This model is very similar to the single state model of autorepression presented in Equation 5.3 and yields a steady state of:

$$y^{ss}(u) = \frac{\alpha g_{tot} K_G (K_I + u)}{\delta_P (K_G (K_I + u) + K_I p_{tot})}, \tag{5.6}$$

where the symbols mean the same as above.

Estimates for the two parameters vary considerably. Scholz *et al.* [186] state that $K_I < K_G$, i.e. that the interaction between TetR and tetracycline is stronger than the interaction between TetR and its operator region, while giving an estimate of $K_I \approx 10^{-12}$M. There seems to be a general consensus that $K_I < K_G$ [95, 97, 186] as it is expected that the signalling system must be incredibly sensitive to tetracycline in the cell as it is an antibiotic, but this is not consistent with parameter estimates in the literature. Further estimates of K_I include [54], where Degenkolb *et al.* study the interaction between TetR and a number of tetracycline derivatives. They estimate $K_I \approx 10^{-11}$M for aTc and between 10^{-9}M and 10^{-10}M for other derivatives. Takahashi *et al.* find $K_I = 3.6 \times 10^{-10}$M [201]. The K_G values in the literature are generally smaller that the estimates of K_I contradicting the statement that $K_I < K_G$. For example, Hillen *et al.* provide an estimate of $K_G \approx 10^{-12}$ to 10^{-13}M [95] and Lederer *et al.* provide estimates up to $K_G \approx 10^{-15}$. All of these values are estimated based on *in vitro* kinetic assays. With an overview of the literature and ignoring $K_I < K_G$, general estimates of the dissociation constants of $K_I \approx 10^{-2}$nM and $K_G \approx 10^{-3}$nM to 10^{-4}nM seem reasonable.

Full induction of P_{tet} occurs at around 100 ng/ml of aTc [125, 132], which is the equivalent of 235 nM or about 138 molecules per cell calculated using a molecular weight of aTc of 426g/mol. This is considerably higher than the gene copy number encoding TetR. If the gene is chromosomal, the copy number is one per cell, or if it is on a plasmid with a pBR322 *ori*-region, such as is the case for the construct used here, the copy number is between 15 and 20 [194, p.230]. As a result,

Table 5.1 – The parameters used to calculate the dose response of the Autorepressor using Equation 5.4 and also the dose response of repressed expression using Equation 5.6, while discussing the values of K_I and K_G.

Param.	Value
α	4min^{-1}
δ_P	0.025min^{-1}
g_{tot}	28.2nM

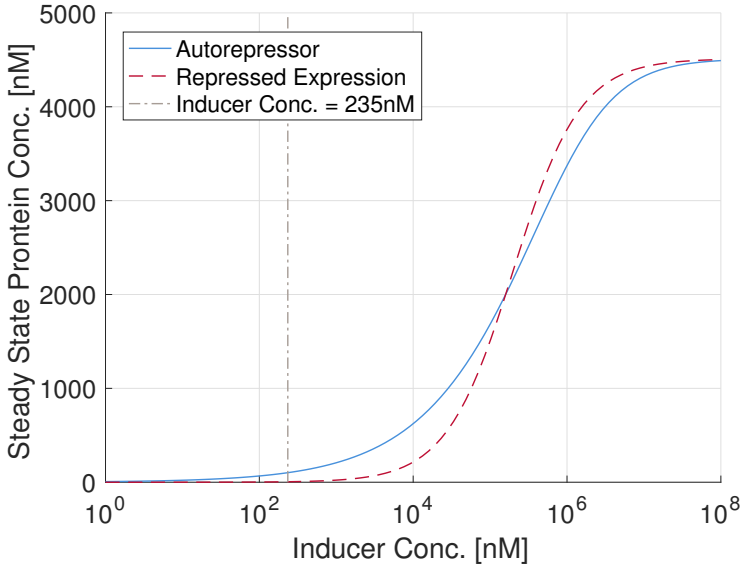


Figure 5.2 – The dose responses of the Autorepressor and repressed expression for a range of inducer concentrations calculated using Equation 5.4, Equation 5.6 and parameter values from the literature. The line corresponds to the point of expected maximal induction of 100ng/ml ($10^{2.4}$nM) of aTc [125, 132]. The parameters used are in Table 5.1, with $K_I = 10^{-2}$nM, $K_G = 10^{-4}$nM and, in the case of repressed expression, $p_{tot} = 2000$nM.

there is a large discrepancy in numbers between the two targets of TetR at high induction. In any case, the dose response of the modelled repressed expression is expected to saturate at around 100ng/ml of aTc.

Using the parameters in Table 5.1 the steady state value of protein concentration for a range of inducer values is calculated. The dose responses of the Autorepressor, calculated using Equation 5.4, and repressed expression, calculated using Equation 5.6, are presented in Figure 5.2. These dose responses are similar and show that maximal induction only occurs above inducer concentrations of about 10^7nM, which is much higher than the level of maximum induction of 100 ng/ml of aTc observed experimentally [125, 132], equivalent to a concentration of 235nM = $10^{2.4}$nM.

As the dose responses in Figure 5.2 calculated using the parameter in Table 5.1 do not show the expected saturation point of around 100ng/ml of aTc, what set of K_I and K_G does? In fact,

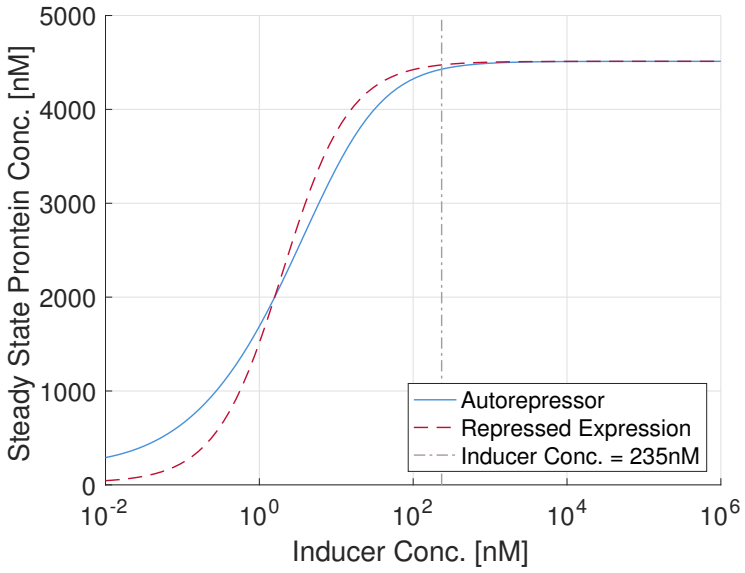


Figure 5.3 – The dose responses of the Autorepressor and repressed expression for a range of inducer concentrations calculated using Equation 5.4, Equation 5.6 and adjusted parameters that produce the desired response. The line corresponds to the point of expected maximal induction of 100ng/ml ($10^{2.3}$nM) of aTc [125, 132]. The parameters used are in Table 5.1, with $K_I = 10^{-2}$nM, $K_G = 10$nM and, in the case of repressed expression, $p_{tot} = 2000$nM.

Table 5.2 – These are the parameters used to calculate the steady state output of the full three state models of autorepression (Equation 5.7) and repressed expression (Equation 5.8. The values of k_1 is from [201] and the value of k_{-2} is from [97].

Param.	Value	Param.	Value
α_P	1min^{-1}	k_1	$1.98 \times 10^{-2}\text{nM}^{-1}\text{min}^{-1}$
δ_P	0.025min^{-1}	k_{-1}	$1.98 \times 10^{-4}\text{min}^{-1}$
α_M	1min^{-1}	k_2	$5.8 \times 10^{-3}\text{nM}^{-1}\text{min}^{-1}$
δ_M	0.25min^{-1}	k_{-2}	$5.8 \times 10^{-2}\text{min}^{-1}$
K_I	10^{-2}nM	g_{tot}	28.2nM
K_G	10nM	α	4min^{-1}

as long as $K_G \ll \frac{4\alpha g}{\delta}$, which using the parameters here is equivalent to $K_G \ll 18048$, what is important is the ratio between the two dissociation constants: $\frac{K_G}{K_I}$. Setting $K_I = 10^{-2}$nM as suggest in the literature and $K_G = 10$nM, leading to a ratio between the two values of 1000, yields the dose responses presented in Figure 5.3. These show saturation at around the desired level of aTc induction of 100ng/ml. Interestingly enough, this supports the statement in [186] that $K_I < K_G$, signifying tighter binding between the inducer and repressor than between the repressor and the gene.

To be able to perform a stochastic simulation of the system of chemical reactions in Equation 5.1 using the Gillespie algorithm, the dissociation constants must be split into their component parts: $K_I = \frac{k_{-1}}{k_1}$ and $K_G = \frac{k_{-2}}{k_2}$. Again, the literature is turned to facilitate this process.

Takahashi *et al.* found $k_1 = 3.3 \times 10^5 M^{-1}s^{-1} = 1.98 \times 10^{-2} nM^{-1}min^{-1}$ [201] and Hillen *et al.* estimated the TetR-operator complex, PG in the model, to have half-life of about 12min [97]. This translates to $k_{-2} = 0.058 min^{-1}$. Taking the values of $K_I = 10^{-2} nM$ and $K_G = 10 nM$, this fixes the two remaining parameters: $k_{-1} = 1.98 \times 10^{-4} min^{-1}$ and $k_2 = 5.8 \times 10^{-3} nM^{-1}min^{-1}$. These values for the rate constants of the reversible interactions between the repressor and gene, and the repressor and inducer are now used to study the predicted dose responses when using larger ODE models.

The behaviour of the steady state of the stochastic simulation performed using the Gillespie algorithm is identical to that of the full ODE model of the Autorepressor detailed in the chemical reactions of Equation 5.1. This ODE model is used to check the steady state and is as follows:

$$\begin{aligned} \dot{p} &= \alpha g - \delta_P p - k_1 p u + k_{-1} p_u - k_2 p g + k_{-2}(g_{tot} - g), \\ \dot{p}_u &= k_1 p u - (k_{-1} + \delta_P) p_u, \\ \dot{g} &= -k_2 p g + (k_{-2} + \delta_P)(g_{tot} - g), \end{aligned} \tag{5.7}$$

where the parameters are represented by the same symbols as in Equation 5.1 and p is the concentration of free repressor protein P, p_u is the concentration of protein-inducer complex PU, and g is the free unbound gene. The state p_g, representing the concentration of protein bound to gene has been removed by using the relationship $g_{tot} = g + p_g$. The only notable change is that expression is modelled in one step with the required rescaled expression rate $\alpha = \frac{\alpha_P \alpha_M}{\delta_M}$. The output of the system, i.e. what is observed experimentally, is the total concentration of the TetR-sfGFP fusion protein, in this case $y = p + p_u + p_g = p + p_u + g_{tot} - g$. This is the species for which the steady state is calculated. The steady state of the above system of equations is calculated explicitly, but the result is not presented here.

As the step response of the repressed expression, with the topology shown in Figure 4.1b and which is modelled above using Hill Functions in single state, is known to saturate at 100ng/ml aTc [125, 132], this again serves as the comparison. The equivalent full ODE system for repressed expression is as follows:

$$\begin{aligned} \dot{y} &= \alpha g - \delta_P y, \\ \dot{p} &= -k_1 p u + k_{-1}(p_{tot} - p - g_{tot} + g) - k_2 p g + k_{-2}(g_{tot} - g), \\ \dot{g} &= -k_2 p g + k_{-2}(g_{tot} - g), \end{aligned} \tag{5.8}$$

where the states and parameters are as they are in the model of the Autorepressor, except the additions of y, which is now the regulated downstream reporter protein, and p_{tot} which is the

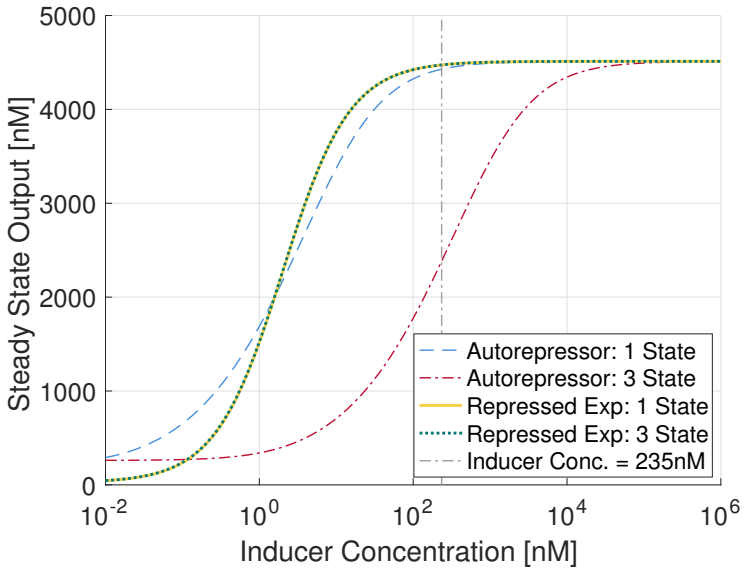


Figure 5.4 – Comparison of dose responses: the Autorepressor single state model (Equation 5.3) and full three-state full model (Equation 5.7), and repressed expression single state model (Equation 5.5) and three-state full model (Equation 5.8). The single state models use the parameters given in Table 5.1, with $K_I = 10^{-2}$nM, K_G10nM, and the full three state models use the parameters in Table 5.2 with the addition of $p_{tot} = 2000$nM for the models of repressed expression. Note that the plots of both one and three state models of repressed expression are on top of each other.

total amount of repressor available in the system. Again, the steady state is found explicitly, but not presented here.

The dose responses of these three-state full models of autorepression and repressed expression are compared to the single state models of the equivalent systems in Figure 5.4. The parameters used to calculate the dose responses of the three-state models are in Table 5.2. The dose responses of the single state models are those presented in Figure 5.3. The result is interesting as the single and three-state models of repressed expression agree, indicating that the two ODE models predict the same dose response. On the other hand, in the case of the Autorepressor, there is a large discrepancy. The three-state ODE predicts saturation at a much higher concentration of inducer. Saturation occurs at about 10^5nM of aTc, the equivalent of 4.3×10^5ng/ml. This level of aTc would almost certainly be incredibly toxic to bacteria. Apart from this discrepancy in level of saturation, the qualitative and quantitative behaviour of the dose responses is identical; they have the same minimum and maximum values. The use of the Hill Functions in the single state model of autorepression changes the predicted level of induction necessary for both the increase in steady state and saturation in the case of the Autorepressor when compared to the three-state model, whereas it does not in case of repressed expression.

This difference in behaviour indicates that, in the case of repressed expression, the single state model is a true reduced version of the full model. This is not the case for the Autorepressor; the full model cannot be reduced to the single state model using the Hill Functions.

If the result is considered intuitively, it makes sense that the dose response of the Autorepressor might saturate at a higher level than repressed expression. This is due to the fact that aTc in effect sequesters (or titrates) TetR out of the cytoplasm [96, 186]. As the aTc enters the cell, it binds TetR tightly, inactivating the repressor. In this process the aTc and TetR in the cell can be thought of as being 'consumed'. In the three-state model of the Autorepressor, the total local concentration of aTc greatly exceeds the level of externally applied inducer, as this includes both the free aTc and that bound to TetR. This mechanism is not included in the Hill Function model by its very construction.

The fact that this discrepancy exists is worth keeping in mind. It is of interest when comparison to experimental data is made as it shows a difference between models that use Hill Functions and those that do not.

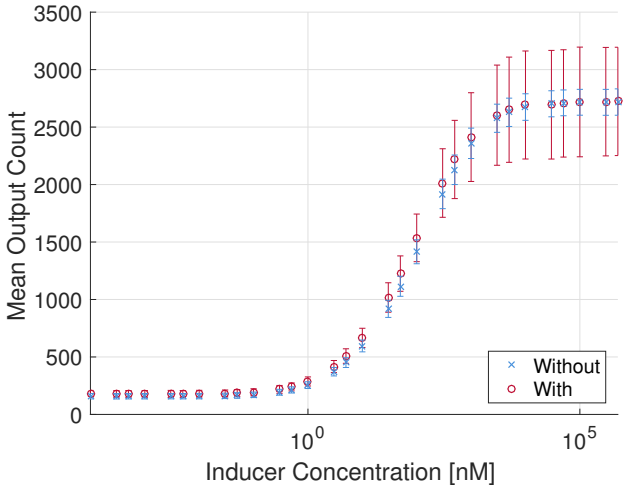
The main focus of this section was to demonstrate the difficulties of choosing parameters. As has been shown, parameters of the *tet*-repressor mechanism provided by the literature through *in vitro* experiments[1] do not necessarily result in expected behaviour when inserted into models. Different models also behave quantitatively and qualitatively differently despite modelling the same system and employing equivalent parameters. As this is the case, the focus of the modelling in the remainder of this chapter is on providing qualitative prediction of behaviour of the systems studied, rather than quantitative. To this end, the choice of parameters is not crucial, unless the choice affects qualitative behaviour, such as through a bifurcation.

5.1.5 *Stochastic Models and Simulation*

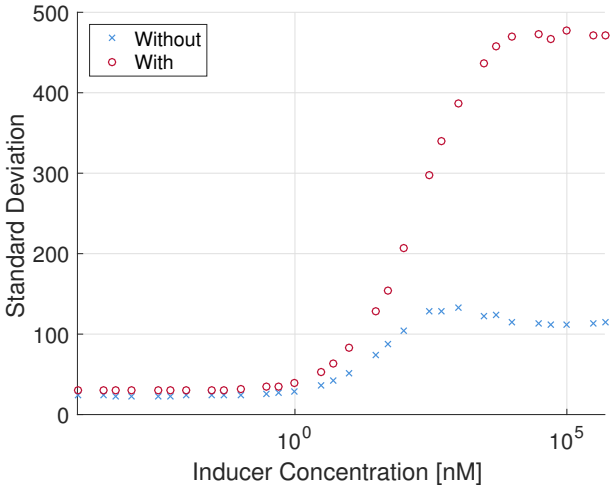
Given that the dose response of the Autorepressor behaves qualitatively as shown in Figure 5.4, stochastic simulations employing the Gillespie Algorithm, as is detailed in Section 2.2.4, can shed light on the expected noise profile expected of the Autorepressor. This is of particular interest as this can be confirmed experimentally through single cell fluorescence collected through flow cytometry. This is similar to the work found in [62], and provides simulation results that can be used to validate models and characterise novel circuits presented later in this chapter.

Two slightly different biochemical reaction networks are employed to model the Autorepressor using the Gillespie Algorithm. The first model is fully described by the reactions in Equation 5.1.

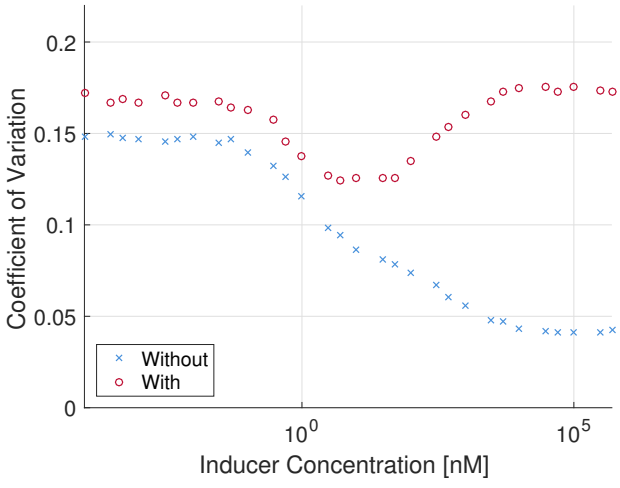
1 It is worth noting that *in vivo* methods of estimating kinetic rates of transcription factors are being developed [90].



(a) Dose Response



(b) Standard Deviation



(c) Coef. of Var.

Figure 5.5 – Stochastic simulation results of the Autorepressor, modelled without and with gene (plasmid) copy number fluctuation. The simulations use the parameters listed in Table 5.2, remembering that some of these need rescaling to have the correct units and with the addition of $\alpha_G = 0.425\text{min}^{-1}$ and $\delta_G = 0.025\text{min}^{-1}$ for the model with gene copy number fluctuation.

The second has been extended to include plasmid copy number fluctuation. This extension is simple and only requires the addition of the following two chemical reactions:

$$\emptyset \xrightarrow{\alpha_G} G \xrightarrow{\delta_G} \emptyset,$$
$$PG \xrightarrow{\delta_G} P,$$

where the first reactions model the production and degradation of the gene (plasmid) with rates α_G and δ_G respectively, and the second reaction ensures the degradation of the gene, when bound to the repressor protein.

The parameter values used to model this assume an average gene copy number of 17 and that the only mechanism removing genes from the system is dilution. As such the half-life of the gene is the same as that of the protein, which here has been set to 28min resulting in a degradation rate of $\delta_G = 0.025\text{min}^{-1}$. As the average gene copy number is fixed by the ratio $\frac{\alpha_G}{\delta_G}$, the production rate can be fixed and is set to $\alpha_G = \delta_G * 17 = 0.425\text{min}^{-1}$.

The parameters used to simulate the Autorepressor employing the Gillespie Algorithm are those presented in Table 5.2, remembering that k_1 and k_2 require rescaling to have the correct units. These are rescaled to $k_1 = 9.6 \times 10^{-3}\text{min}^{-1}$ and $k_2 = 3.29 \times 10^{-2}\text{min}^{-1}$. Each individual simulation in the ensemble used is run for 300min in simulation time, long enough to ensure that the system has reached a point where data represents fluctuations around the steady state mean. For the data analysis, only the end point of each of the 5000 runs in each ensemble is used.

The results from the simulation are presented in Figure 5.5. There is a clear difference between the qualitative behaviour of the model with and without gene copy number fluctuation. Interestingly enough, apart from being noisier at high levels of induction, the model with gene copy number fluctuation predicts a drop and then a rise again in the coefficient of variation. This dip occurs over the course of the transition between the minimum and maximum levels of induction. With this in mind, copy number fluctuation is included in further modelling as it changes the qualitative behaviour of the predicted coefficient of variation. These models can be verified though collection of experimental data using the construct implemented and a data collection method such at flow cytometry.

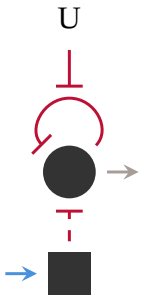


Figure 5.6 – The minimal representation of the topology of sRNA Circuit I. The black square indicates expression of an sRNA, the dashed red line with a bar indicates an interaction at RNA level. The blue arrow indicates external activation or in this case that the transcription rate of the sRNA can be varied. The other symbols are identical to those used in the case of the Autorepressor.

5.2 SRNA CIRCUIT I

This is the first of three novel circuits that are studied in this chapter. The topology of this circuit is presented in Figure 5.6, and as can be seen, it is the basic Autorepressor with an added sRNA inhibiting the expression of the repressor. This gives the system, in effect, two inputs: the inducer interacting with the repressor, and the transcription rate of the sRNA.

5.2.1 *Biological Implementation*

sRNA Circuit I requires two genes. These can be treated as compossible subunits: the Autorepressor gene and the sRNA gene. The implementation of sRNA Circuit I uses the Autorepressor implementation on the pJ404 backbone as described in Section 5.1.1. This consists of a TetR-sfGFP fusion gene with the P_{tet} promoter controlling its expression.

The initial implementation of this circuit has the Autorepressor and the sRNA gene split across two plasmids. The Autorepressor on the pJ404 backbone with pBR322 *ori*-region, and the sRNA gene on a pSUtat backbone with a p15A *ori*-region [112]. More details about the plasmid backbones can be found in Table 2.2. The original implementation of the sRNA employs the natural MicC scaffold as is recommended in [151, 224], which leaves space for 24 bases for the target sequence.

There are two suggested target regions presented in the literature. These can be seen in Figure 2.11. Na *et al.* find that the most efficient regulation occurs when the target region begins with the Start Codon on the target mRNA, whereas Man *et al.* suggest that the most efficient regulation occurs when the sRNA targets a region beginning with the Shine-Dalgarno sequence. More information is available in Section 2.1.15.4.

To test this, two different sRNAs are designed: one targeting each of the suggested regions. In the original implementation, to ensure that the sRNA is expressed in excess in relation to its target, transcription is controlled by two strong constitutive promoters in tandem: the engineered *proD* promoter (P_{proD}) [52] and the P_r promoter [224].

Later implementations of sRNA Circuit I migrate the sRNA gene so that both the Autorepressor and the sRNA gene are on the same plasmid. In addition to this, various versions of the circuit are produced where transcription of the sRNA is controlled by a number of means. In the first instance, constitutive expression of the sRNA in controlled through individual or combinations of constitutive promoters of various strengths. The transcription of the sRNA is also placed under the control of the P_{rhaBAD} promoter, which is activated by the activating TF RhaS and induced by L-rhamnose or analogs such at L-mannose. More can be read about this mechanism in Section 2.1.14.4.

5.2.2 *Basic Biochemical Model*

As this circuit includes an sRNA, the guidelines given in Chapter 3 are followed. They state that ODE models of the sRNA mechanism need not include Host-Factor of Bacteriophage Qβ (Hfq), the chaperone protein that catalyses the sRNA inhibition mechanism. On the other hand, stochastic simulations should include this catalytic mechanism. To this end, an initial biochemical reaction model used to study sRNA Circuit I using ODEs is put together without including the Hfq mechanism.

This model is based on reasoning presented in Section 5.1.2 for the Autorepressor subunit of the circuit, and the reactions included in the sRNA subunit are described in detail in Section 3.1.1, which covers the basic model of the sRNA inhibition mechanism. These two biochemical models are composed to produce the following model of sRNA Circuit I.

$$
\begin{aligned}
M &\xrightarrow{\alpha_P} M + P, & P &\xrightarrow{\delta_P} \emptyset,\\
G_P &\xrightarrow{\alpha_M} G_P + M, & PU &\xrightarrow{\delta_P} \emptyset,\\
G_R &\xrightarrow{\alpha_R} G_R + R, & PG_P &\xrightarrow{\delta_P} G_P,\\
P + U &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PU, & M &\xrightarrow{\delta_M} \emptyset,\\
P + G_P &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} PG_P, & R &\xrightarrow{\delta_R} \emptyset,\\
M + R &\xrightarrow{K} \emptyset,
\end{aligned}
\tag{5.9}
$$

where P is the repressor protein, M is the mRNA encoding the repressor, and this mRNA is targeted by the sRNA R. The mRNA is transcribed from the free gene encoding the repressor G_P, and the sRNA is encoded from the gene G_R. The complex between the repressor and gene is PG_P, which is not transcribed. The inducer is denoted U and binds to the repressor forming the complex PU. The translation rate of protein is α_P and the transcription rates of mRNA and sRNA are α_M and α_R respectively. The association and dissociation rates of the protein-inducer interaction are k_1 and k_{-1} respectively. The association and dissociation rates of the protein-gene interaction are k_2 and k_{-2} respectively. The mRNA and sRNA interaction occurs at rate K and is treated as irreversible. All species including protein degrade at the rate δ_P. The mRNA and sRNA degrade at rates δ_M and δ_R respectively.

It is worth noting that the species observed experimentally through the reporter protein is the all species including P. This yields $Y = P + PU + PG_P$. In addition to this, although the gene encoding the protein is involved in a few of the reactions, its total quantity is conserved in the above model $G_{Ptot} = G_P + PG_P$.

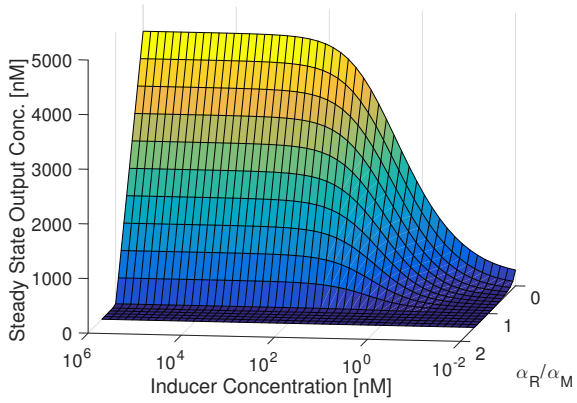
5.2.3 *ODE Model with Hill Function*

Similar to the case of the Autorepressor, this first ODE model studied uses Hill Functions to model the interaction between the protein and inducer, and the protein and gene. The form of this function and the means of its derivation are found in Section 2.2.3.3. The three-state ODE model is as follows:

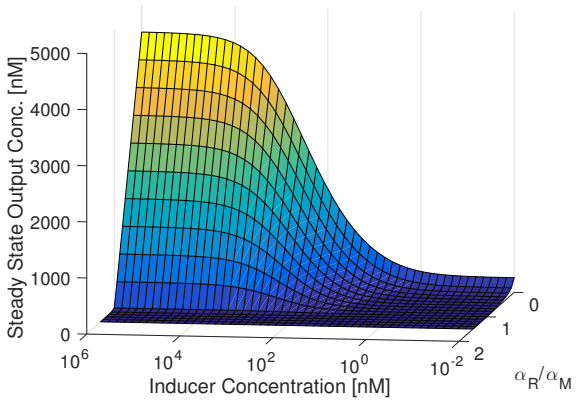
$$\begin{aligned}
\dot{y} &= \alpha_P m - \delta_P y, \\
\dot{m} &= \alpha_M g_{Ptot} \frac{K_G(K_I + u)}{K_G(K_I + u) + K_I y} - \delta_M m - K m r, \\
\dot{r} &= \alpha_R g_{Rtot} - \delta_R r - K m r,
\end{aligned} \tag{5.10}$$

where y is the total protein concentration, m is the concentration of mRNA and r is the concentration of sRNA. The total concentration of the genes encoding the protein and sRNA are g_{Ptot} and g_{Rtot} respectively. This is set to 28.2nM based on the copy number of 17 per cell. The dissociation constants are defined as $K_I = \frac{k_{-1}}{k_1}$ and $K_G = \frac{k_{-2}}{k_2}$. Apart from that, the rate constants and their significance are given in Equation 5.9.

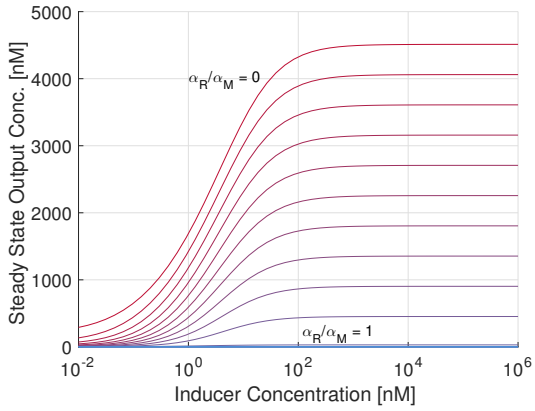
The steady state of this system can be found explicitly and there is one solution in the positive quadrant. The explicit result is not presented here as it is complex and not dealt with analytically.



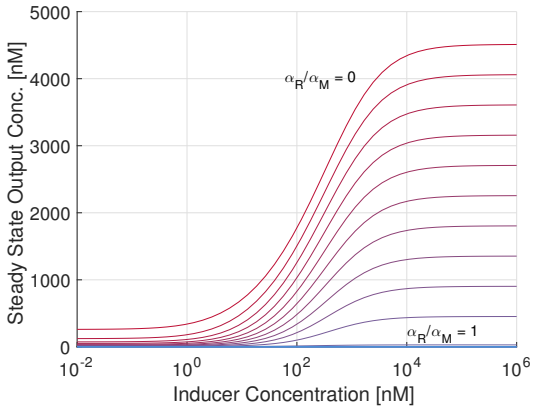
(a) Hill Func. Model



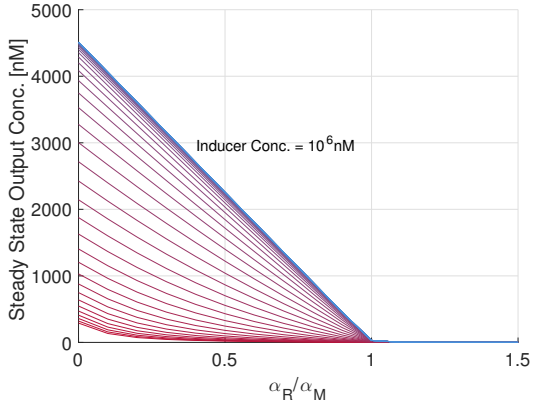
(b) Full Model



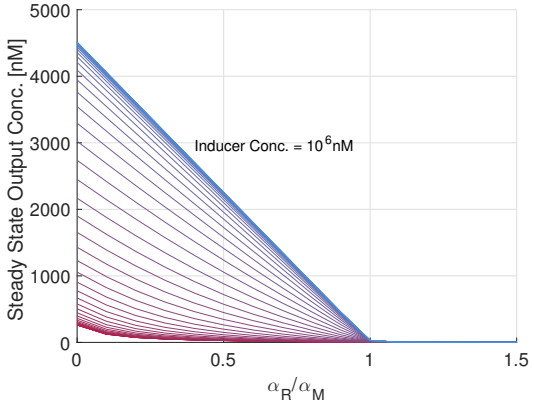
(c) Hill Func. Model



(d) Full Model



(e) Hill Func. Model



(f) Full Model

Figure 5.7 – Results studying the steady state output (y) for both the three-state ODE model with Hill Functions (Equation 5.10) and the five-state full ODE model (Equation 5.11) of sRNA Circuit I. These are found using the parameters in Table 5.3. The concentration of inducer u is varied from 10^{-2}nM to 10^{6}nM and the ratio between the translation rates of sRNA and mRNA, $\frac{\alpha_R}{\alpha_M}$, is varied from 0 to 1.5.

5.2.4 *Full ODE Model*

As in the case of the Autorepressor, a further ODE model is produced including every reaction in Equation 5.9 explicitly. This results in the following five-state full model:

$$\begin{aligned}
\dot{p} &= \alpha_P m - \delta_P p - k_1 p u + k_{-1} p_u - k_2 p g_P + k_{-2}(g_{Ptot} - g_P),\\
\dot{p}_u &= k_1 p u - (k_{-1} + \delta_P) p_u,\\
\dot{m} &= \alpha_M g_P - \delta_M m - K m r,\\
\dot{g}_P &= (k_{-2} + \delta_P)(g_{Ptot} - g_P) - k_2 p g_P,\\
\dot{r} &= \alpha_R g_{Rtot} - \delta_R r - K m r,
\end{aligned} \tag{5.11}$$

where p is the free repressor concentration, u is the inducer concentration, m is the concentration of mRNA, r is the concentration of sRNA, g_P is the concentration of free gene encoding the repressor, p_u is the repressor bound to inducer, g_{Ptot} and g_{Rtot} are the total concentrations of gene encoding repressor and sRNA respectively. The rate constants are as given in Equation 5.9.

As is mentioned in the case of the biochemical reaction network, what is observed experimentally is the combination of all species including the protein: $y = p + p_u + p_g = p + p_u + g_{Ptot} - g_P$, where p_g is the concentration of repressor bound to gene PG_P, which is removed through the conservation of gene: $g_{Ptot} = g_P + p_g$. The lumped output variable y is what is of interest in an experimental context and is therefore what is studied here. The steady state values for the five-state full model are found numerically.

5.2.5 *Steady State Output Results*

The plots of the steady state output of the three-state model, using Hill Functions, and the five-state full model are calculated and presented in Figure 5.7 using the parameters presented in Table 5.3 while varying the concentration of inducer u from 10^{-2}ng/ml to 10^{6}ng/ml and the ratio between the translation rates of sRNA and mRNA, $\frac{\alpha_R}{\alpha_M}$ from 0 to 1.5.

The parameters used to model the repression mechanism are identical to those used in the case of the Autorepressor in the previous section. The parameters modelling the sRNA inhibition mechanism follow those given in Section 3.1.1, in particular the assumption that the sRNA degrades at the same rate as protein as it is guarded by the Hfq chaperone.

Table 5.3 – Parameters used to calculate the steady state output of the three-state ODE model with Hill Functions (Equation 5.10) and the five-state full ODE model (Equation 5.11) of sRNA Circuit I. The parameters for the repression mechanism are identical to those used for the Autorepressor.

Param.	Value	Param.	Value
α_P	1min^{-1}	α_R	1min^{-1}
δ_P	0.025min^{-1}	δ_R	0.025min^{-1}
α_M	1min^{-1}	K_G	10nM
δ_M	0.25min^{-1}	K_I	10^{-2}nM
k_1	$1.98 \times 10^{-2}\text{nM}^{-1}\text{min}^{-1}$	K	$5\text{nM}^{-1}\text{min}^{-1}$
k_{-1}	$1.98 \times 10^{-4}\text{min}^{-1}$	g_{Ptot}	28.2nM
k_2	$5.8 \times 10^{-3}\text{nM}^{-1}\text{min}^{-1}$	g_{Rtot}	28.2nM
k_{-2}	$5.8 \times 10^{-2}\text{min}^{-1}$		

The behaviour seen is similar to that of the Autorepressor with the addition that the sRNA adds a means of controlling the level of maximum induction of the Autorepressor. Particularly, looking at Figure 5.7c and Figure 5.7d for $\alpha_R/\alpha_M = 0$, this is equivalent to having no sRNA inhibition, i.e. just the Autorepressor on its own. The difference between the model with Hill Functions and full model is the same as in the case of the Autorepressor. The Hill Function model retains maximum induction at a concentration of aTc of about 100ng/ml ($10^{2.3}$nM), whereas in the case of the full model the level of aTc required for maximum induction is about 10^5nM.

At higher levels of induction, the ratio between the transcription rates α_R/α_M decides the level of expression with the familiar threshold linear response. At low levels of induction, the rate of transcription of mRNA is heavily repressed and therefore does not exceed the transcription of sRNA and yields a very low response. It is expected that when the expression of the sRNA is controlled by the tandem constitutive promoters, the system is in the regime $\alpha_R > \alpha_M$. This can then be used to test the tightness and efficiency of the regulation by the sRNA.

Though not presented here, an ODE model of sRNA Circuit I including the Hfq mechanism, modelled as in Section 3.2.2, was studied and found to behave almost identically to the full model in a regime where the sRNA regulation was efficient. This is as expected given the results in Chapter 3.

5.2.6 *Stochastic Models*

This section sheds light on the expected noise properties of sRNA Circuit I at steady state. Following the recommendation of Chapter 3, a stochastic model of a gene regulatory circuit including the sRNA inhibition mechanism should include the Hfq regulation mechanism.

With this in mind the following set of biochemical reactions are put together to model sRNA Circuit I:

$$
\begin{aligned}
M &\xrightarrow{\alpha_P} M + P, & P &\xrightarrow{\delta_P} \emptyset,\\
HM &\xrightarrow{\alpha_P} HM + P, & PU &\xrightarrow{\delta_P} \emptyset,\\
G_P &\xrightarrow{\alpha_M} G_P + M, & PG_P &\xrightarrow{\delta_P} G_P,\\
G_R &\xrightarrow{\alpha_R} G_R + R, & M &\xrightarrow{\delta_M} \emptyset,\\
G_H &\xrightarrow{\alpha_H} G_H + H, & HM &\xrightarrow{\delta_M} H,\\
P + U &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PU, & HMR &\xrightarrow{\delta_M} HR,\\
P + G_P &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} PG_P, & MR &\xrightarrow{\delta_M} \emptyset,\\
H + M &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HM, & R &\xrightarrow{\delta_R} \emptyset,\\
H + R &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HR, & H &\xrightarrow{\delta_H} \emptyset,\\
HM + R &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HMR, & HM &\xrightarrow{\delta_H} \emptyset,\\
HR + M &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HMR, & HR &\xrightarrow{\delta_H} \emptyset,\\
HMR &\xrightarrow{K} H + MR, & HMR &\xrightarrow{\delta_H} \emptyset,
\end{aligned}
\tag{5.12}
$$

where the symbols and rates are equivalent to those in Equation 5.9, with the addition of H which denotes a free Hfq protein, the complex between mRNA and Hfq is denoted HM, the complex between sRNA and Hfq is denoted HR, and the complex between Hfq and the forming mRNA-sRNA heteroduplex is HMR. The gene encoding the Hfq protein is G_H. It is assumed that this gene is on the chromosome and the total copy number is therefore $G_{Htot} = 1$. The forward and reverse reaction rates for all interactions between Hfq and the RNA species are assumed to be the same and are denoted k_f and k_r respectively. The Hfq protein is expressed at the rate α_H, and Hfq and all complexes in which it partakes degrade at the rate δ_H. The rate of heteroduplex formation is again K, but it has a slightly different significance to that in Equation 5.9, as the required reactants for the reaction are different. In addition to this, it is assumed that the total gene copy numbers of the genes encoding sRNA and mRNA are $G_{Rtot} = G_{Ptot} = 17$.

The above models the Autorepressor analogously to the chemical reactions in Equation 5.9, but is then extended to include the asymmetric model of sRNA inhibition as it is presented in Section 3.2.2. The reactions that make this model of Hfq asymmetric are those that allow for the rapid degradation of mRNA whether it is bound to Hfq or not. The assumption is that the sRNA,

though an RNA, is not degraded rapidly when bound to Hfq as the chaperone protects it from the degradation machinery.

This model assumes no gene copy number fluctuation for any of its species. As the implemented versions of this circuit have each of the two subunits, the Autorepressor gene and sRNA gene, on the same plasmid and split across two separate plasmids, it is of interest to study both these set-ups.

Assuming first the circuit has both genes required on a single plasmid, this can be added to the model by the inclusion of the following reactions in Equation 5.12:

$$\emptyset \xrightarrow{\alpha_G} G_P \xrightarrow{\delta_G} \emptyset, \tag{5.13a}$$

$$G_P \xrightarrow{\alpha_R} G_P + R, \tag{5.13b}$$

$$PG_P \xrightarrow{\alpha_R} PG_P + R, \tag{5.13c}$$

$$PG_P \xrightarrow{\delta_P} \emptyset, \tag{5.13d}$$

where G_P now is the species indicator for both the gene encoding the protein and the gene encoding the sRNA. In this way it can be thought of as the plasmid as a whole. Equation 5.13a is simply included in the model. Equation 5.13b and Equation 5.13c replace the reaction governing the transcription of sRNA: $G_R \xrightarrow{\alpha_R} G_R + R$. Finally, Equation 5.13d replaces the equation governing the degradation of the gene-repressor complex: $PG_P \xrightarrow{\delta_P} G_P$.

The sRNA is now expressed from the plasmid from its constitutive promoter, whether or not the repressor protein is bound to its operator region elsewhere on the plasmid. The average gene copy number again assumed to be 17, and with a degradation rate of plasmid equal to the dilution rate of protein and the doubling time of the cell, $\delta_G = 0.025\text{min}^{-1}$, this yields a production rate of the plasmid of $\alpha_G = 0.425\text{min}^{-1}$.

Lastly, it is assumed that the two genes encoding the circuit are on separate plasmids. This is done by including the following equations in the chemical reaction model given in Equation 5.12:

$$\emptyset \xrightarrow{\alpha_G} G_P \xrightarrow{\delta_G} \emptyset, \tag{5.14a}$$

$$\emptyset \xrightarrow{\alpha_G} G_R \xrightarrow{\delta_G} \emptyset, \tag{5.14b}$$

$$PG_P \xrightarrow{\delta_P} \emptyset, \tag{5.14c}$$

where Equation 5.14a and Equation 5.14b are simply added to the reactions in Equation 5.12 and govern the production and degradation of the two separate plasmids denoted G_P and G_R

Table 5.4 – The parameters used to perform stochastic simulations of sRNA Circuit I. Results of these are presented in Figure 5.8 and Figure 5.9. The values used to simulate the repression mechanism are identical to those used for the Autorepressor. The parameter k_1 is from [201] and k_{-2} is from [97].

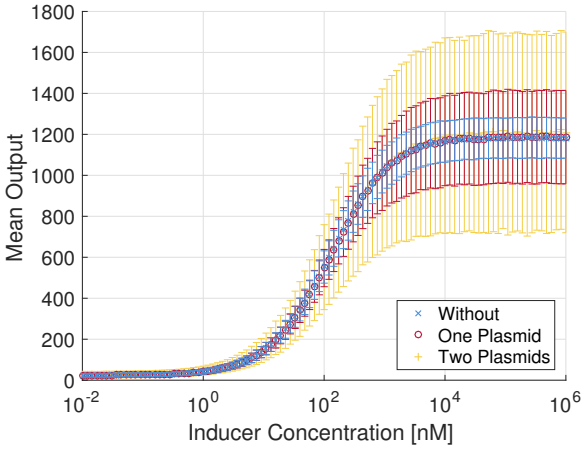
Param.	Value	Param.	Value
α_P	1min^{-1}	k_1	$3.29 \times 10^{-2}\text{min}^{-1}$
δ_P	0.025min^{-1}	k_{-1}	$1.98 \times 10^{-4}\text{min}^{-1}$
α_M	1min^{-1}	k_2	$9.63 \times 10^{-3}\text{min}^{-1}$
δ_M	0.25min^{-1}	k_{-2}	$5.8 \times 10^{-2}\text{min}^{-1}$
α_R	0.5min^{-1}	k_f	16.6min^{-1}
δ_R	0.25min^{-1}	k_r	0.7min^{-1}
α_H	1min^{-1}	K	5min^{-1}
δ_H	0.025min^{-1}	α_G	0.425min^{-1}
G_{Ptot}	17	δ_G	0.025min^{-1}
G_{Rtot}	17	U	69000
G_{Htot}	1		

respectively. Again Equation 5.14c replaces the equation governing the degradation of the gene-repressor complex: $PG_P \xrightarrow{\delta_P} G_P$. It is assumed that the two plasmids have the same average copy number of 17, and therefore use the same rates of production and degradation. Though this is not correct for the system implemented here, where the Autorepressor gene is on a plasmid with copy number 15-20 and the sRNA gene is on a plasmid with copy number 10-12, this facilitates the comparison of behaviour at various transcription rates while still including the copy number fluctuation mechanism.

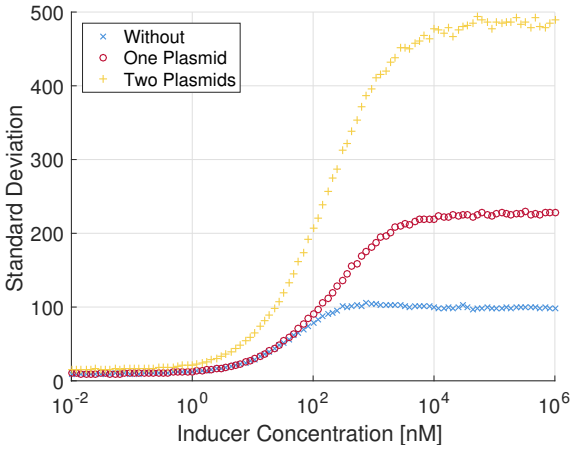
5.2.7 *Stochastic Results*

Instead of simulating the stochastic model for a mesh of inducer concentrations and relative transcription rates, as is done in the deterministic case, slices are taken. The first slice is taken for $\alpha_R = 0.5\text{min}^{-1}$ while sweeping the concentration of inducer from 10^{-2}nM to 10^6nM. This is done employing the parameters presented in Table 5.4, which use identical parameters to model the repression mechanism as those used to model the Autorepressor, remembering that k_1 and k_2 require rescaling to have the correct units for a stochastic simulation.

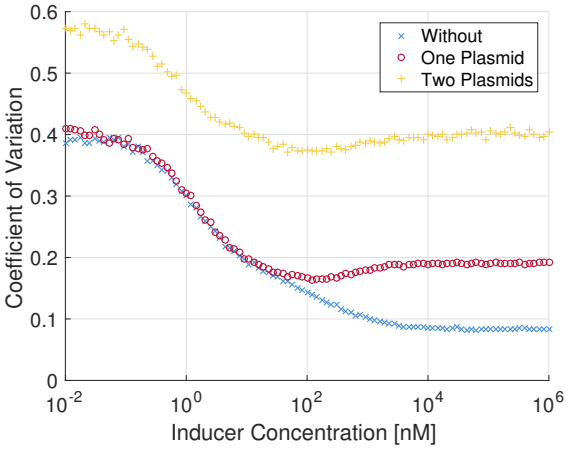
Ensembles of 5000 runs are executed for each inducer concentration, and each simulation is run for 300min in simulation time to ensure that the system has reached fluctuations around steady state prior to termination. Only end point data is used to provide the mean and standard deviation of the population.



(a) Mean

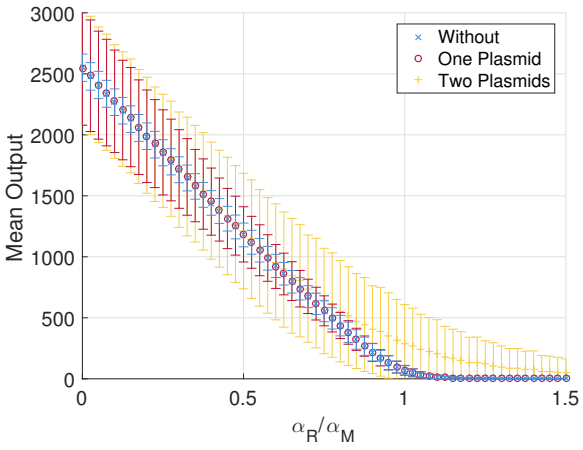


(b) Standard Deviation

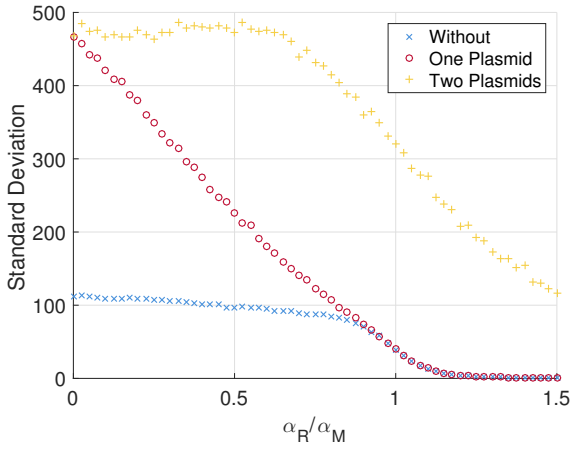


(c) Coefficient of Variation

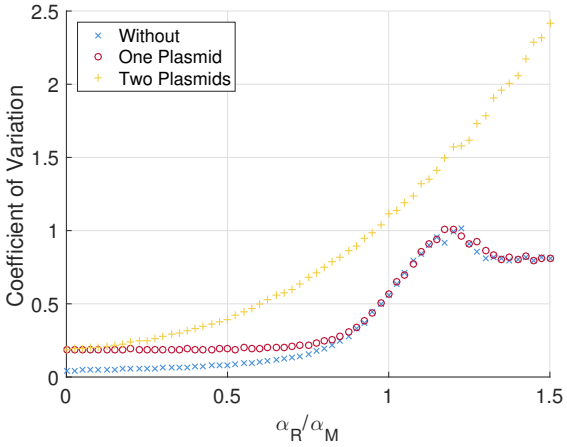
Figure 5.8 – Stochastic simulation of the three models of sRNA Circuit I: the two plasmid model, one plasmid model and the model with no gene copy fluctuation. This simulation fixes $\alpha_R = 0.5\text{min}^{-1}$ while sweeping the concentration of inducer from 10^{-2}nM to 10^{6}nM. The rest of the parameters are detailed in Table 5.4.



(a) Mean



(b) Standard Deviation



(c) Coefficient of Variation

Figure 5.9 – Stochastic simulation of the three models of sRNA Circuit I: the two plasmid model, one plasmid model and the model with no gene copy fluctuation. This simulation fixes the inducer concentration to 5×10^4nM, equivalent to U= 69000 and then varies the relative transcription rate α_R/α_M from 0 to 1.5. The rest of the parameters are detailed in Table 5.4.

The results from these simulations for the three models of sRNA Circuit I can be seen in Figure 5.8. The mean output values of the three models agree closely. It is shown that the system spread across two plasmids is considerably more noisy than both the model without copy number fluctuation and the model with a single plasmid. This is to be expected given the results in Chapter 3. As with the Autorepressor, the coefficient of variation is at its lowest at mid levels of induction, around 100nM.

The second slice taken fixes the inducer concentration to 5×10^4nM, equivalent to $U = 69000$ and then varies the relative transcription rate α_R/α_M from 0 to 1.5. This value of inducer concentration is chosen as it is around maximum induction. Again, the rest of the parameters are detailed in Table 5.4.

The results for the three models of sRNA Circuit I are presented in Figure 5.9. The mean values show the expected threshold linear response and agree well, though as expected from results in Chapter 3, the transition at the threshold is softened for the system where the genes for the Autorepressor and sRNA are split across two plasmids. It is interesting how the standard deviation for the model without copy number fluctuation and the model with two plasmids behaves qualitatively similarly, remaining relatively flat for $\alpha_R/\alpha_M < 1$ and then declining above the threshold. Note how the coefficient of variation grows continuously for the two plasmid model, whereas it levels off for the one plasmid case at the same level as the model without plasmid copy number fluctuation.

Through experimentation with the implemented versions of this system, varying the amount of inducer employed and the transcription rate of the sRNA the above results can be validated. Data collection must be performed using flow cytometry while the cells are in exponential phase, following a protocol much like that in [62].

5.3 SRNA CIRCUIT II

The circuit studied in this section, called sRNA Circuit II, employs two genes and negative feedback, like sRNA Circuit I, but in this case the negative feedback is supplied through an inhibiting sRNA. The topology of the circuit is presented in Figure 5.10. Similarly to sRNA Circuit I, this circuit has two inputs: the inducer, labelled U in the figure, and in addition to this, it is possible to control the transcription rate of the mRNA encoding the activator.

This topology is interesting as a similar topology is known to exist in *E. coli*. This circuit topology regulates the expression of the *rpoE* gene that encodes the sigma-factor σ24 through

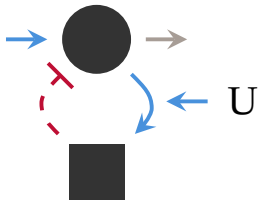


Figure 5.10 – This is the topology of sRNA Circuit II. The circuit consists of two genes, where negative feedback is supplied through an inhibiting sRNA. The black circle represents the expression of an activating TF, in this specific case RhaS, which activates the expression of an sRNA, the black square. The sRNA then in turn inhibits the translation of RhaS. As in the case of sRNA Circuit I, the dashed line indicates an interaction at RNA level. The activator is induced by the inducer U, in this case l-rhamnose or an analog such as l-mannose. The second input to the system is the controllable transcription rate of the mRNA encoding the activator. The output of the system is indicated by the grey arrow, and in this case is a fluorescence measurement of the sfGFP fused to the activator RhaS.

the sRNA RybB, which increases the speed of degradation of the *rpoE* mRNA [190]. The factor σ24 is the factor required for expression of the sRNA.

5.3.1 *Biological Implementation*

The implemented systems representing this circuit all use a RhaS-sfGFP fusion protein as both activator and output. This fusion protein is constructed by removing the Stop Codon from the end of *rhaS* and cloning *sfGFP* so that it begins immediately after the penultimate codon of *rhaS*. This is identical to the TetR-sfGFP fusion. More can be read about the RhaS activation mechanism in Section 2.1.14.4. Translation of this gene controlled by the native RBS found in the *E. coli* strain MG1655, from which both the *rhaS* gene and the native RBS are cloned. This RBS yields a score of 20k from the RBS Calculator [184]. Transcription of the RhaS-sfGFP fusion gene is controlled by constitutive promoters of varying strengths, such as P_{proD} [52] and P_r [224], but there is also the option of placing it under the control of an inducible promoter such as P_{tet}, which would allow an experimentalist to vary the transcription rate *in vivo*.

The sRNA used is very similar to that used in sRNA Circuit I. The MicC scaffold is used again as recommended in [151, 224], which leaves a recommended 24 base pairs for the target binding region of the sRNA. As with sRNA Circuit I, two implementations of sRNA are made, one for each of the two target binding regions suggested in Figure 2.11. The expression of this sRNA is controlled by the P_{rhaBAD} promoter, which is activated by RhaS. This completes the circuit.

The activator RhaS is induced by l-rhamnose, but this is metabolised in *E. coli*. For sustained induction, it has been shown that the non-metabolisable analog l-mannose can be used instead [113].

There are two basic set-ups for the circuit that are similar to the case of sRNA Circuit I. The first is a two plasmid system, where the gene encoding the RhaS-sfGFP fusion protein and its associated RBS and promoter are encoded on the pJ404 backbone [113], which has a pBR322 *ori*-region and a copy number of 15-20. The sRNA and associated promoter is encoded on the pSUtat backbone [112], which has a p15A *ori*-region and a copy number of 10-12. The second set-up has both genes on the pJ404 plasmid.

A last requirement for the implementation of this circuit is the use of a Δ*rhaS* strain of *E. coli*, so that the circuit and its behaviour is not influenced by the RhaS expressed from the chromosome of wild-type strains such as MG1655. The Keio Collection strain JW3876 is used for this [17]. In this strain *rhaS* is replaced with a gene encoding a kanamycin resistance protein.[2]

5.3.2 *Basic Biochemical Model*

Kinetic details of the RhaS activation mechanism, describing the interaction of the activator with its operator regions and inducers, have not been found experimentally. This is due to the fact that the protein is difficult to purify due to the exposure of hydrophobic amino acids when not induced. This leads to aggregation making the protein very hard to deal with experimentally [64], resulting in a lack of kinetic parameters.

Though RhaS is an activator of a metabolic pathway, and therefore is not nearly as sensitive to inducer as TetR, the two systems are to some extent similar. To begin with, both transcription factors are dimers, where each subunit binds an inducer molecule. The assumption that the native state is a dimer is supported by the literature to some extent [26, 63, 64, 213], and therefore the dimerisation step is ignored. RhaS is activated on binding its inducer and the active form binds the target operator regions of the DNA. These two processes are modelled here similarly to the case of the *tet*-repressor mechanism. As this is an activator, it is assumed that expression only occurs from a promoter that is bound to the active TF.

When constructing the model, the guidelines given in Chapter 3 are followed. As these biochemical reactions are used to construct sets of ODEs, the sRNA inhibition mechanism is modelled as in Section 3.1.1.

2 The full genotype of JW3876 is F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, ΔrhaS765::kan, hsdR514.

The basic biochemical reaction model for sRNA Circuit II is therefore:

$$\begin{aligned}
M &\xrightarrow{\alpha_P} M + P, & P &\xrightarrow{\delta_P} \emptyset,\\
G_P &\xrightarrow{\alpha_M} G_P + M, & PU &\xrightarrow{\delta_P} \emptyset,\\
P + U &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PU, & G_RPU &\xrightarrow{\delta_P} G_R,\\
G_R + PU &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} G_RPU, & M &\xrightarrow{\delta_M} \emptyset,\\
G_RPU &\xrightarrow{\alpha_R} G_RPU + R, & R &\xrightarrow{\delta_R} \emptyset,\\
M + R &\xrightarrow{K} \emptyset,
\end{aligned} \tag{5.15}$$

where the P is the activator protein, M is the mRNA encoding the activator protein, R is the sRNA, U is the inducer, G_P is the gene encoding P, G_R is the free gene encoding R, PU is the active TF complex, and G_RPU is the complex between PU and the gene G_R, from which transcription can occur. Transcription of M and R occur at rates α_M and α_R respectively, and translation of P occurs at rate α_P. The forward and reverse reaction rate constants for the protein-inducer interaction are k_1 and k_{-1} respectively. The forward and reverse reaction rate constants for the active protein's interaction with the gene encoding the sRNA are k_2 and k_{-2} respectively. Formation of the heteroduplex between the RNA species occurs at rate K, all species containing the protein degrade at rate δ_P, and the mRNA and sRNA degrade at rates δ_M and δ_R respectively.

As in previous examples, it is worth noting that what is observed experimentally is all species including the protein, as this is what is tagged with the reporter sfGFP: $Y = P + PU + G_RPU$. In addition to this, the total amount of gene encoding the sRNA is conserved, yielding $G_{Rtot} = G_R + G_RPU$.

5.3.3 *ODE Model with Hill Function*

This section lays out the simplest ODE model of sRNA Circuit II. This model uses a Hill Function to model the interactions between the activator, inducer and gene. The derivation of this expression rate function is very similar to that seen in Section 2.2.3.3, the major difference being that an

activator is now being modelled. Using the basic principles laid out in Section 2.2.3, the following system of ODEs can be produced:

$$\begin{aligned}
\dot{y} &= \alpha_P m - \delta_P y, \\
\dot{m} &= \alpha_M g_{Ptot} - \delta_M m - K m r, \\
\dot{r} &= \alpha_R g_{Rtot} \frac{yu}{K_G(K_I + u) + yu} - \delta_R r - K m r,
\end{aligned} \tag{5.16}$$

where y is the total concentration of protein and the observed output, m is the concentration of mRNA, r is the concentration of sRNA, u is the concentration of inducer, and g_{Ptot} and g_{Rtot} are the total concentrations of the genes encoding the protein and sRNA respectively. The rate constants used are identified in Equation 5.15 with the addition of the dissociation constants $K_I = \frac{k_{-1}}{k_1}$ and $K_G = \frac{k_{-2}}{k_2}$, which are for the activator-inducer interaction and the activator-gene interaction respectively.

5.3.4 *Full ODE Model*

In a similar vein to the Autorepressor and sRNA Circuit I, a second ODE model is proposed that models all the reactions in Equation 5.15 explicitly. This model has five states and is of the following form:

$$\begin{aligned}
\dot{p} &= \alpha_P m - \delta_P p - k_1 p u + k_{-1} p_u, \\
\dot{p}_u &= k_1 p u - (k_{-1} + \delta_P) p_u - k_2 p_u g_R + k_{-2}(g_{Rtot} - g_R), \\
\dot{m} &= \alpha_M g_{Ptot} - \delta_M m - K m r, \\
\dot{g}_R &= (k_{-2} + \delta_P)(g_{Rtot} - g_R) - k_2 p_u g_R, \\
\dot{r} &= \alpha_R (g_{Rtot} - g_R) - \delta_R r - K m r,
\end{aligned} \tag{5.17}$$

where p is the concentration of free activator protein, p_u is the concentration of activator protein bound to inducer, m is the concentration of mRNA, g_R is the concentration of free sRNA gene, and r is the concentration of sRNA. The total concentrations of gene encoding the protein and sRNA are g_{Ptot} and g_{Rtot} respectively.

The observed output of the system is the combined concentration of all species including the protein: $y = p + p_u + p_g = p + p_u + g_{Rtot} - g_R$, where p_g is the activator-inducer complex bound to the gene, which is denoted $G_R PU$ in Equation 5.15. This state is removed by using the conservation of total gene concentration: $g_{Rtot} = g_R + p_g$.

5.3.5 *Steady State Output Results*

The hardest parameters to choose when modelling this system are the parameters quantifying the RhaS activation mechanism. As there is very little literature on the matter due to the experimental difficulties of handling the pure protein [118], there are no values to guide the choice. To choose the parameters k_1, k_{-1}, k_2, k_{-2} and through them K_I and K_G, a similar process is used to that demonstrated in the case of the *tet*-repressor mechanism in Section 5.1.4. Two models, this time of activated expression such as is presented in Figure 4.1a, are used to choose the values. The first model employs a Hill Function and the second models the on and off reactions between the activator, inducer and gene explicitly. With the knowledge that maximum induction of the RhaS mechanism occurs at about 1mg/ml of L-rhamnose, the parameters are chosen to yield a suitable dose response and are presented in Table 5.5. The value $k_{-2} = 5.8 \times 10^{-2}\text{min}^{-1}$ from [97], that gives the dissociation rate between TetR and P_{tet}, is used as a guiding value. A half-life of 12min for the activator-gene complex is perfectly reasonable with no further information at hand. As is clear from the value and units of k_1, the chosen parameters show that the RhaS mechanism is much less sensitive to inducer than TetR, which is expected as RhaS regulates a metabolic pathway; it would not make sense for bacteria to express proteins required to digest L-rhamnose if it only were present at very low concentrations.

With the parameters chosen, the steady state values of the output y are calculated for both ODE models of sRNA Circuit II: the three-state model using a Hill Function, Equation 5.16, and the five-state full model, Equation 5.17. A full list of the parameters employed is presented in Table 5.5. The steady state results for the two models are presented in Figure 5.11, where both the inducer concentration and the ratio between the transcription rates of the protein and the sRNA have been varied.

In contrast to the case of the Autorepressor and sRNA Circuit I, there is good, but not complete, agreement between the two ODE models. This is an interesting system, as counter intuitively, the steady state output is predicted to drop when RhaS is induced. This is because with no inducer present, none of the inhibiting sRNA is transcribed, so the activator is freely expressed from the constitutive promoter. The models predict that there is only very low output for the case of a high inducer concentration, $u > 1\text{mg/ml}$, and a low ratio between the transcription rates $\alpha_M/\alpha_R < 1$. This ideal operating region is most likely one where $\alpha_M \approx \alpha_R$ as this region allows for both a relatively high maximum output and almost an almost zero minimum.

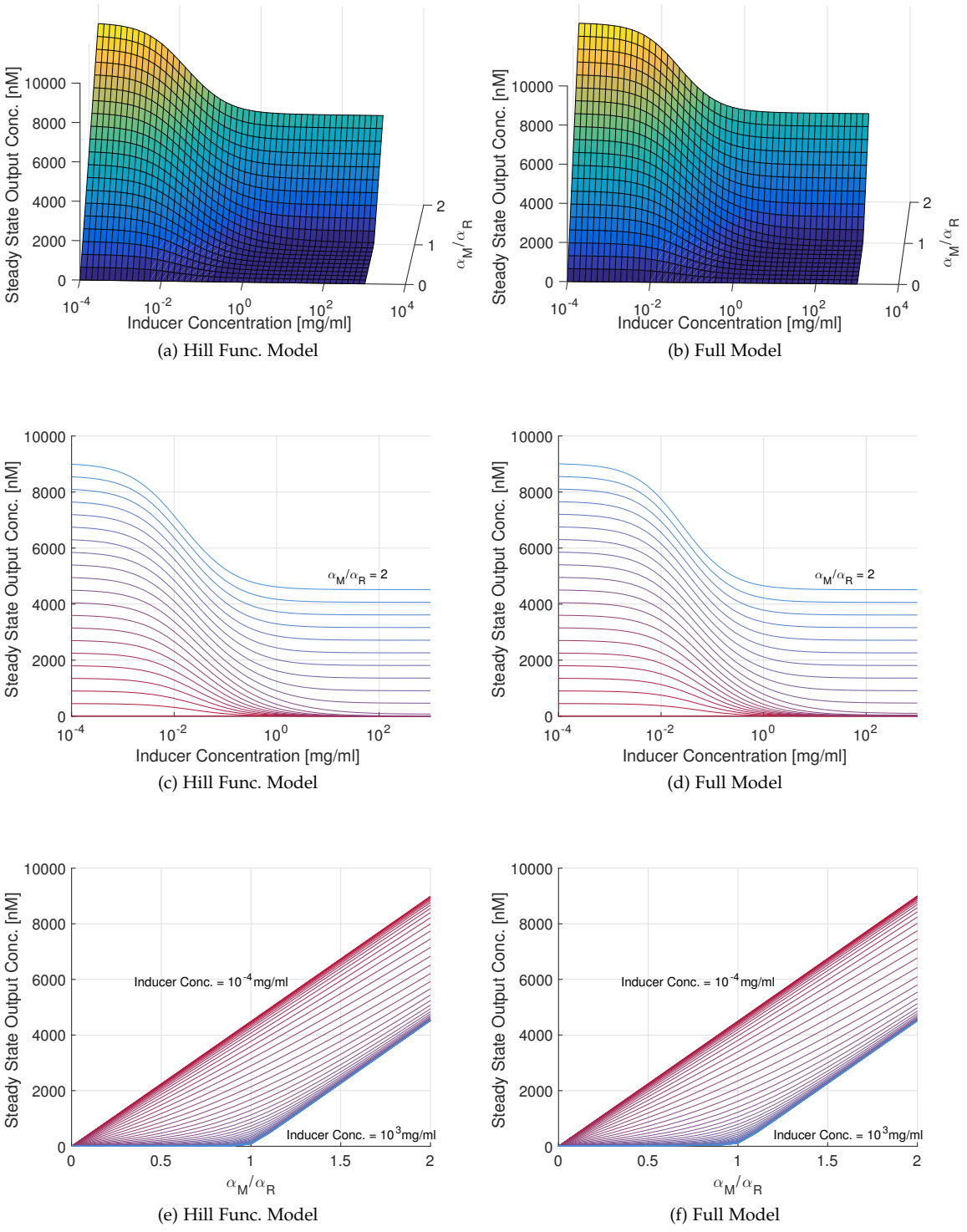
Table 5.5 – Parameters used to calculate the steady state out put of the three-state ODE model with Hill Functions (Equation 5.16) and the five-state full ODE model (Equation 5.17) of sRNA Circuit II.

Param.	Value	Param.	Value
α_P	1min^{-1}	α_R	1min^{-1}
δ_P	0.025min^{-1}	δ_R	0.025min^{-1}
α_M	1min^{-1}	K_G	1nM
δ_M	0.25min^{-1}	K_I	100gl^{-1}
k_1	$10^{-2}\text{lg}^{-1}\text{min}^{-1}$	K	$5\text{nM}^{-1}\text{min}^{-1}$
k_{-1}	1min^{-1}	g_{Ptot}	28.2nM
k_2	$5.8 \times 10^{-2}\text{nM}^{-1}\text{min}^{-1}$	g_{Rtot}	28.2nM
k_{-2}	$5.8 \times 10^{-2}\text{min}^{-1}$		

As with the case of sRNA Circuit I, an ODE model of sRNA Circuit II including Hfq was checked and found to behave no differently to the five-state full model in a regime where the sRNA regulation is efficient. This is as expected given results in Chapter 3.

5.3.6 *Stochastic Models*

Stochastic simulation techniques are now employed as these can predict the expected noise profile expected from a circuit. Following the guidelines given in Chapter 3, stochastic simulation including an sRNA regulation mechanism should include a model of the Hfq interactions.



(a) Hill Func. Model (b) Full Model

(c) Hill Func. Model (d) Full Model

(e) Hill Func. Model (f) Full Model

Figure 5.11 – Results studying the steady state output (y) for both the three-state ODE model with the Hill Function (Equation 5.16) and the five-state full ODE model (Equation 5.17) of sRNA Circuit II. These are found using the parameters in Table 5.5. The concentration of inducer u is varied from 10^{-4}mg/ml to 10^{3}mg/ml and the ratio between the translation rates of mRNA and sRNA, $\frac{\alpha_M}{\alpha_R}$, is varied from 0 to 2.

To that end, the following biochemical reaction model of sRNA Circuit II is employed:

$$
\begin{aligned}
M &\xrightarrow{\alpha_P} M + P, & P &\xrightarrow{\delta_P} \emptyset,\\
HM &\xrightarrow{\alpha_P} HM + P, & PU &\xrightarrow{\delta_P} \emptyset,\\
G_P &\xrightarrow{\alpha_M} G_P + M, & G_RPU &\xrightarrow{\delta_P} G_R,\\
G_RPU &\xrightarrow{\alpha_R} G_RPU + R, & M &\xrightarrow{\delta_M} \emptyset,\\
G_H &\xrightarrow{\alpha_H} G_H + H, & HM &\xrightarrow{\delta_M} H,\\
P + U &\underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PU, & HMR &\xrightarrow{\delta_M} HR,\\
PU + G_R &\underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} G_RPU, & MR &\xrightarrow{\delta_M} \emptyset,\\
H + M &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HM, & R &\xrightarrow{\delta_R} \emptyset,\\
H + R &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HR, & H &\xrightarrow{\delta_H} \emptyset,\\
HM + R &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HMR, & HM &\xrightarrow{\delta_H} \emptyset,\\
HR + M &\underset{k_r}{\overset{k_f}{\rightleftharpoons}} HMR, & HR &\xrightarrow{\delta_H} \emptyset,\\
HMR &\xrightarrow{K} H + MR, & HMR &\xrightarrow{\delta_H} \emptyset,
\end{aligned}
\tag{5.18}
$$

where the species and parameters are as in Equation 5.15 with the additions of the free Hfq protein, H, the mRNA-Hfq and sRNA-Hfq complexes, HM and HR respectively, the RNA heteroduplex in complex with Hfq, HMR, the expression rate of Hfq, α_H, the degradation rate of Hfq, δ_H, and the forward and reverse reaction rates of the Hfq interactions with the RNA species, k_f and k_r respectively. The rate of duplex formation is still labelled K, but its significance is different in the above model as it models the dissociation from Hfq of the formed heteroduplex. Again, note that the output of the system, that which is observed experimentally is the sum of all species including the protein:$Y = P + PU + G_RPU$.

The above model is a hybrid of the basic biochemical model of sRNA Circuit II, detailed in Equation 5.15, extended to include the asymmetric model of sRNA regulation detailed in Section 3.2.2. This model assumes that, whereas the sRNA is protected by the Hfq protein, the mRNA degrades fast, whether it is bound or not.

The experimental set-up of this system employs plasmids as vectors for the genes of the circuit. As with sRNA Circuit I, it is interesting to study what effect these have on the performance of the system.

With some additions to the chemical reaction model in Equation 5.18, using one plasmid as a vector for both of the genes that comprise the circuit can be included in the model. To do this the following reactions are required:

$$\emptyset \xrightarrow{\alpha_G} G_R \xrightarrow{\delta_G} \emptyset, \tag{5.19a}$$

$$G_R \xrightarrow{\alpha_M} G_R + M, \tag{5.19b}$$

$$G_R PU \xrightarrow{\alpha_M} G_R PU + M, \tag{5.19c}$$

$$G_R PU \xrightarrow{\delta_P} \emptyset, \tag{5.19d}$$

where G_R now signifies the plasmid. Equation 5.19a is simply added to the model and describes the dynamics of the plasmid with production rate α_G and degradation rate δ_G. Equation 5.19b and Equation 5.19c replace the reaction for the transcription of mRNA: $G_P \xrightarrow{\alpha_M} G_P + M$. These reactions model the transcription of the mRNA whether or not it is bound by the activator. Equation 5.19d replaces the reaction for the degradation rate of the gene-activator complex: $G_R PU \xrightarrow{\delta_P} G_R$.

As no active degradation of the plasmid occurs, the only mechanism removing it from the system is dilution. Therefore the degradation rate of the plasmid is set to the same as that of the protein: $\delta_G = 0.025 \text{min}^{-1}$. The plasmid copy number is assumed to be 17 which results in a production rate of $\alpha_G = 0.425 \text{min}^{-1}$.

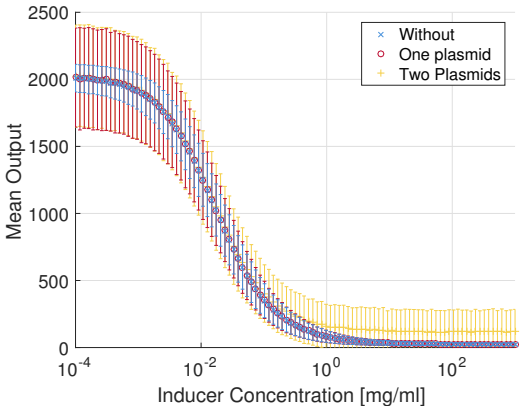
The reactions required to model the gene on two separate plasmids are:

$$\emptyset \xrightarrow{\alpha_G} G_P \xrightarrow{\delta_G} \emptyset, \tag{5.20a}$$

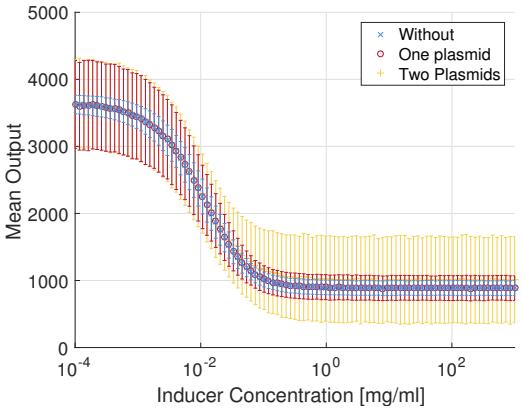
$$\emptyset \xrightarrow{\alpha_G} G_R \xrightarrow{\delta_G} \emptyset, \tag{5.20b}$$

$$G_R PU \xrightarrow{\delta_P} \emptyset. \tag{5.20c}$$

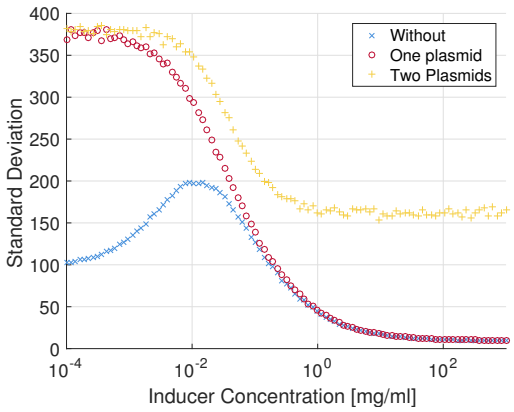
The reactions in Equation 5.20a and Equation 5.20b are simply added to Equation 5.18 and model the dynamics of the two plasmids, and Equation 5.20c replaces the reaction for the degradation rate of the gene-activator complex: $G_R PU \xrightarrow{\delta_P} G_R$. As in the case of sRNA Circuit I, the two plasmids are assumed to have the same copy number of 17.



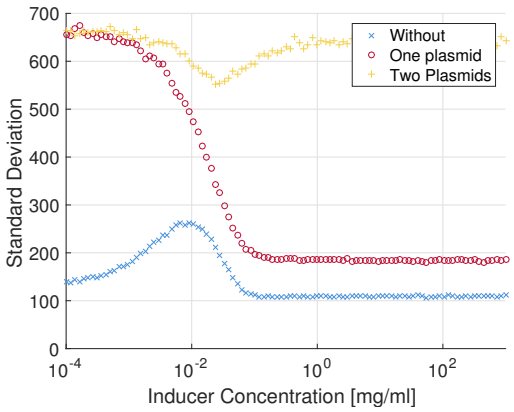
(a) Low α_M/α_R



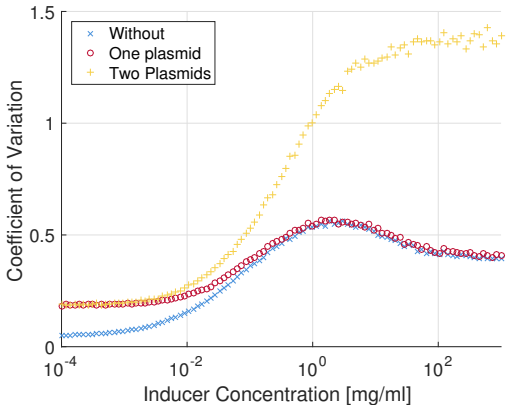
(b) High α_M/α_R



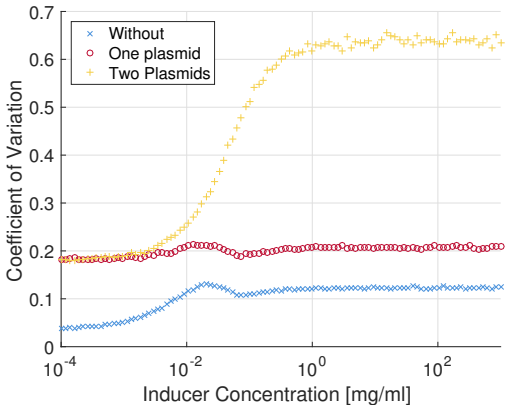
(c) Low α_M/α_R



(d) High α_M/α_R

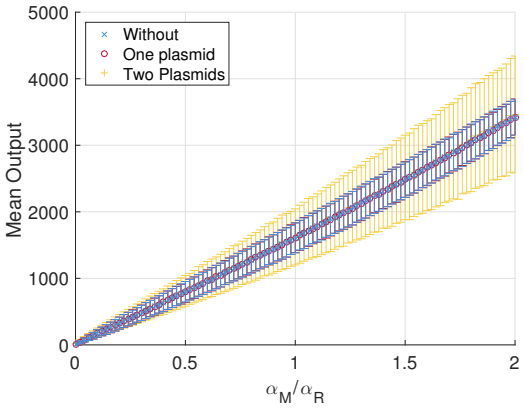


(e) Low α_M/α_R

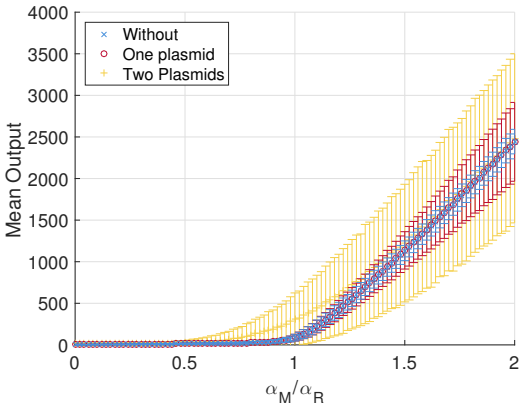


(f) High α_M/α_R

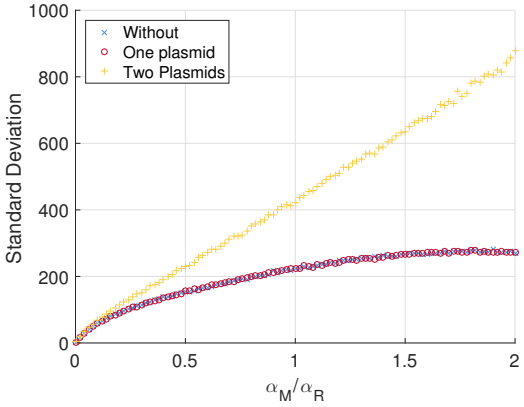
Figure 5.12 – Stochastic simulation of sRNA Circuit II for the three models: not including plasmids, with one plasmid and two plasmids. The ratio between the translation rates of mRNA and sRNA, $\frac{\alpha_M}{\alpha_R}$ is 0.8 when low and 1.4 when high. The concentration of inducer is varied from 10^{-4}mg/ml to 10^{3}mg/ml.



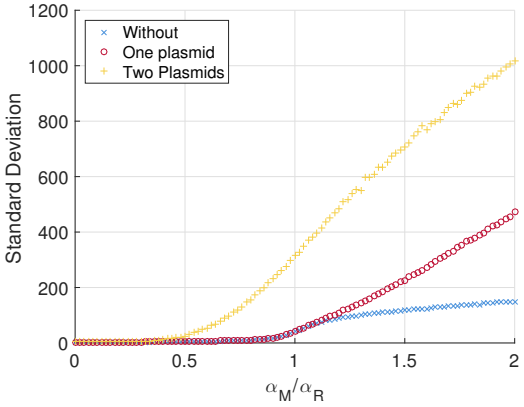
(a) Low Inducer Concentration



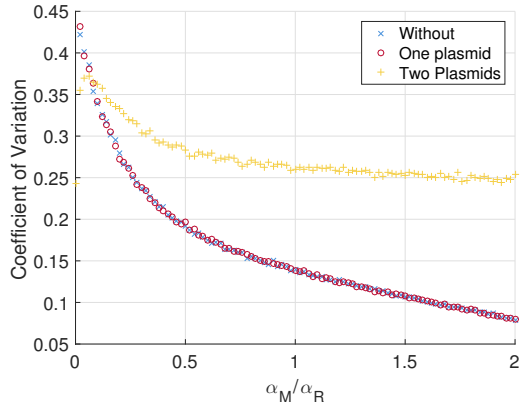
(b) High Inducer Concentration



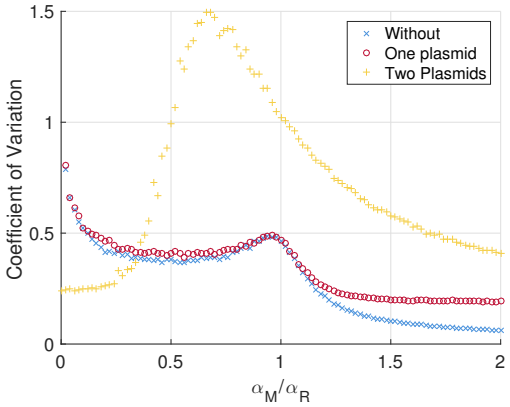
(c) Low Inducer Concentration



(d) High Inducer Concentration



(e) Low Inducer Concentration



(f) High Inducer Concentration

Figure 5.13 – Stochastic simulation of sRNA Circuit II for the three models: not including plasmids, with one plasmid and two plasmids. The concentrations of inducer chosen are 10^{-2}mg/ml when low and 10^{2}mg/ml when high. The ratio between the translation rates of mRNA and sRNA, $\frac{\alpha_M}{\alpha_R}$, is varied from 0 to 2.

Table 5.6 – The parameters used to perform stochastic simulations of sRNA Circuit II. Results of these are presented in Figure 5.12 and Figure 5.13.

Param.	Value	Param.	Value
α_P	1min^{-1}	k_1	$2.7 \times 10^{-9}\text{min}^{-1}$
δ_P	0.025min^{-1}	k_{-1}	10^{-2}min^{-1}
α_M	$0.8\text{or}1.4\text{min}^{-1}$	k_2	$9.63 \times 10^{-2}\text{min}^{-1}$
δ_M	0.25min^{-1}	k_{-2}	$5.8 \times 10^{-2}\text{min}^{-1}$
α_R	1min^{-1}	k_f	16.6min^{-1}
δ_R	0.25min^{-1}	k_r	0.7min^{-1}
α_H	1min^{-1}	K	5min^{-1}
δ_H	0.025min^{-1}	α_G	0.425min^{-1}
G_{Ptot}	17	δ_G	0.025min^{-1}
G_{Rtot}	17	U	3.7×10^4 or 3.7×10^8
G_{Htot}	1		

5.3.7 *Stochastic Results*

Using these three models of sRNA Circuit II, the system is simulated using the parameters presented in Table 5.6. Note that the value of k_1 here shows the huge insensitivity of RhaS to inducer. At 1mg/ml, the number of inducer molecules expected in a cell with a volume of 10^{-15}l is 3.7×10^6, so this insensitivity is not surprising.

Due to the behaviour seen in the deterministic models, it is decided to take four slices in inducer and transcription rate ratio space. The first two slices are defined by fixing the values 0.8 and 1.4 for the ratio $\frac{\alpha_M}{\alpha_R}$ and sweeping the inducer concentration from 10^{-4}mg/ml to 10^3mg/ml. The values for the transcription rate ratio are chosen as they represent a value above and below the threshold $\alpha_M = \alpha_R$. The results from these slices can be seen in Figure 5.12. The second two slices fix the values 0.01mg/ml and 100mg/ml for the inducer concentration and sweep the value of the transcription rate ratio from 0 to 2. The inducer concentrations represent an almost un-induced system and a highly induced system. These results can be seen in Figure 5.13.

All the results presented are derived from end point data of ensemble simulations of 5000 individual runs. These simulations run for 300min in simulation time to ensure that the moments studied provide data around the steady state.

The means of the three systems are in good agreement, and also compare well with the deterministic results. As with sRNA Circuit II, the two plasmid system is by far the most noisy, but in general the noise level decreases approaching that of the one plasmid system as the output level increases.

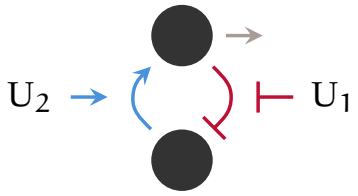


Figure 5.14 – The topology of the activator repressor circuit. The interactions supplying the negative feedback are through transcription factors. A repressor, used as the output, inhibits the expression of an activator that in turn activates expression of the repressor. The two inducers, used as inputs to the system are presented. The inducer interacting with the repressor is labelled U_1, and the inducer interacting with the activator is labelled U_2.

The noise levels in the case of a low ratio of transcription rates is considerably higher at high levels of induction than in the case of a high level of transcription rates. This is because in the prior case the level of expression of output protein approaches zero for high levels of induction. Given the fact that the biologically implemented system uses a relatively strong constitutive promoter to control the expression of the output protein, it is expected that the implemented system is in the regime of high $\frac{\alpha_M}{\alpha_R}$.

The results in Figure 5.13b shows the threshold linear response expected when the inducer concentration is high. The coefficient of variation shows a slight bump in the one and no plasmid results at the threshold in this system (see Figure 5.13f). This is a trait that might be measurable experimentally. The noise is again much higher in the two plasmid case, and the threshold is softened.

5.4 THE REPRESSOR-ACTIVATOR CIRCUIT

This genetic circuit is the final system studied in this chapter. The circuit consists of negative feedback using two genes that interact through two transcription factors. The topology of the circuit is presented in Figure 5.14. As can be seen, the circuit consists of a repressor, inhibiting the expression of an activator, that in turn activates the expression of the repressor. The system has two natural inputs, namely the inducers belonging to each of the transcription factor mechanisms used. In the experimental set-up, the repressor is used as the output.

5.4.1 *Biological Implementation*

Given the above discussions about the repressor TetR and the activator RhaS, it is not surprising that these are the two TFs used to realise this circuit. The genes encoding the two TFs are simply

linked using their associated promoters: P_{tet} is placed just upstream of *rhaS* with the native *rhaS* RBS in between, and P_{rhaBAD} is placed just upstream of *tetR* with the optimised RBS used in the Autorepressor in between.

As the Autorepressor was implemented prior to this circuit, it was decided to use the TetR-sfGFP fusion protein present in the Autorepressor as the output of this system. Either of the two TFs could have been chosen as the experimental output of the system. It is even possible to use them both as outputs simultaneously using other fluorophores, such as YFP, CFP or mCherry that emit light with different spectra. One of these could be fused to RhaS to enable this. For more information on reporters, see Section 2.1.17.

The experimental implementation of this circuit has both genes on the pJ404 backbone [112], which has a copy number of 15-20.

5.4.2 *Biochemical Model*

The reaction network below describes the model of the Repressor-Activator Circuit. This combines the two TF mechanisms in accordance with the topology in Figure 5.14, providing the desired negative feedback:

$$
\begin{aligned}
M_1 &\xrightarrow{\alpha_{P1}} M_1 + P_1, & P_1 &\xrightarrow{\delta_{P1}} \emptyset,\\
G_1P_2U_2 &\xrightarrow{\alpha_{M1}} G_1P_2U_2 + M_1, & P_1U_1 &\xrightarrow{\delta_{P1}} \emptyset,\\
M_2 &\xrightarrow{\alpha_{P2}} M_2 + P_2, & G_2P_1 &\xrightarrow{\delta_{P1}} G_2,\\
G_2 &\xrightarrow{\alpha_{M2}} G_2 + M_2, & M_1 &\xrightarrow{\delta_{M1}} \emptyset,\\
P_1 + U_1 &\underset{k_{-11}}{\overset{k_{11}}{\rightleftharpoons}} P_1U_1, & P_2 &\xrightarrow{\delta_{P2}} \emptyset,\\
G_2 + P_1 &\underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} G_2P_1, & P_2U_2 &\xrightarrow{\delta_{P2}} \emptyset,\\
P_2 + U_2 &\underset{k_{-21}}{\overset{k_{21}}{\rightleftharpoons}} P_2U_2, & G_1P_2U_2 &\xrightarrow{\delta_{P2}} G_1,\\
G_1 + P_2U_2 &\underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} G_1P_2U_2, & M_2 &\xrightarrow{\delta_{M2}} \emptyset,
\end{aligned}
\tag{5.21}
$$

where P_i denotes a TF protein and G_i, M_i and U_i are its associated gene, mRNA and inducer respectively, $i = 1$ denotes the repressor and $i = 2$ denotes the activator, P_iU_i denotes the TF-inducer complex, $G_1P_2U_2$ denotes the complex between the repressor gene and the activator bound to inducer, and G_2P_1 denotes the activator gene-repressor complex. The translation rates of the TFs are α_{Pi}, the transcription rates of the associated mRNAs are α_{Mi}, the forward and

reverse reaction rates of the interaction between a TF and its associated inducer are k_{i1} and k_{-i1} respectively, and the forward and reverse reaction rates for the interaction between a TF and the associated gene are k_{i2} and k_{-i2} respectively, the TFs and all associated complexes degrade at rate δ_{Pi}, the mRNAs degrade at rate δ_{Mi}, where $i = 1$ refers to the repressor and $i = 2$ refers to the activator.

As with the other systems studied in this chapter, the experimental output is what is of interest. The output of this system can be both or either of the two total amounts of transcription factor. These are denoted $Y_1 = P_1 + P_1U_1 + G_2P_1$ and $Y_2 = P_2 + P_2U_2 + G_1P_2U_2$.

5.4.3 *ODE Model with Hill Functions*

As with the other circuits studied in this chapter, the first ODE model presented employs Hill Functions. This model requires two Hill Functions one modelling induced repression and the other modelling induced activation. These are derived using the methods presented in Section 2.2.3.3. The two-state model is as follows:

$$\begin{aligned} \dot{y}_1 &= \alpha_1 g_{1tot} \frac{y_2 u_2}{K_{G2}(K_{I2} + u_2) + y_2 u_2} - \delta_{P1} y_1 \\ \dot{y}_2 &= \alpha_2 g_{2tot} \frac{K_{G1}(K_{I1} + u_1)}{K_{G1}(K_{I1} + u_1) + K_{I1} y_1} - \delta_{P2} y_2, \end{aligned} \tag{5.22}$$

where y_1 is the total concentration of repressor, y_2 is the total concentration of activator, u_1 is the concentration of inducer that interacts with the repressor, and u_2 is the concentration of inducer that interacts with the activator. The maximal expression rates $\alpha_1 = \frac{\alpha_{P1}\alpha_{M1}}{\delta_{M1}}$ and $\alpha_2 = \frac{\alpha_{P2}\alpha_{M2}}{\delta_{M2}}$ are derived from model reduction by time scale separation presented in Section 2.2.3.1, and the dissociation constants are defined as $K_{Ii} = \frac{k_{-i1}}{k_{i1}}$ and $K_{Gi} = \frac{k_{-i2}}{k_{i2}}$, where $i = 1$ refers to the constants associated with the repressor and $i = 2$ refers to those associated with the activator. The other parameters are presented in Equation 5.21.

Table 5.7 – These are the parameters used to study the ODE models of the Repressor-Activator Circuit. The parameters modelling the TF mechanisms are identical to those used in the above circuits.

Param.	Val.	Param.	Val.
α_{P1}	1min^{-1}	k_{11}	$1.98 \times 10^{-2}\text{nM}^{-1}\text{min}^{-1}$
δ_{P1}	0.025min^{-1}	k_{-11}	$1.98 \times 10^{-4}\text{min}^{-1}$
α_{M1}	1min^{-1}	k_{12}	$5.8 \times 10^{-3}\text{nM}^{-1}\text{min}^{-1}$
δ_{M1}	0.25min^{-1}	k_{-12}	$5.8 \times 10^{-2}\text{min}^{-1}$
α_{P2}	1min^{-1}	k_{21}	$10^{-2}\text{lg}^{-1}\text{min}^{-1}$
δ_{P2}	0.025min^{-1}	k_{-21}	1min^{-1}
α_{M2}	1min^{-1}	k_{22}	$5.8 \times 10^{-2}\text{nM}^{-1}\text{min}^{-1}$
δ_{M2}	0.25min^{-1}	k_{-22}	$5.8 \times 10^{-2}\text{min}^{-1}$
g_{1tot}	28.2nM	g_{2tot}	28.2nM

5.4.4 *Full ODE Model*

Again, a full ODE model is presented, that explicitly models all interactions in Equation 5.21. Due to the conservation of total gene concentration, this model has eight states and is as follows:

$$\begin{aligned}
\dot{p}_1 &= \alpha_{P1} m_1 - \delta_{P1} p_1 - k_{11} p_1 u_1 + k_{-11} p_{1u} - k_{12} p_1 g_2 + k_{-12}(g_{2tot} - g_2),\\
\dot{p}_{1u} &= k_{11} p_1 u_1 - (k_{-11} + \delta_{P1}) p_{1u},\\
\dot{m}_1 &= \alpha_{M1}(g_{1tot} - g_1) - \delta_{M1} m_1,\\
\dot{g}_2 &= (k_{-12} + \delta_{P1})(g_{2tot} - g_2) - k_{12} p_1 g_2,\\
\dot{p}_2 &= \alpha_{P2} m_2 - \delta_{P2} p_2 - k_{21} p_2 u_2 + k_{-21} p_{2u},\\
\dot{p}_{2u} &= k_{21} p_2 u_2 - (k_{-21} + \delta_{P2}) p_{2u} - k_{22} p_{2u} g_1 + k_{-22}(g_{1tot} - g_1),\\
\dot{m}_2 &= \alpha_{M2} g_2 - \delta_{M2} m_2,\\
\dot{g}_1 &= (k_{-22} + \delta_{P2})(g_{1tot} - g_1) - k_{22} p_{2u} g_1,
\end{aligned} \tag{5.23}$$

where p_i is the concentration of free TF protein, p_{iu} is the concentration of TF-inducer complex, m_i is the concentration of mRNA encoding a TF, g_i is the concentration of free gene encoding a TF, and u_i is the inducer that interacts with a TF, where i indicates the specific TF, with $i = 1$ indicating the repressor and $i = 2$ indicating the activator. The parameters of the model are presented in Equation 5.21.

The observed output is the total concentration of either repressor or activator or both. The total concentration of repressor is $y_1 = p_1 + p_{1u} + g_{2tot} - g_2$, and the total concentration of activator is $y_2 = p_2 + p_{2u} + g_{1tot} - g_1$. These quantities are studied at steady state, as they are the easiest species to relate to experimental output.

5.4.5 *Steady State Output Results*

Using the inducer concentrations u_1 and u_2 as inputs, the Hill Function and full models are studied. The steady states of the two possible outputs y_1 and y_2 are calculated numerically using the parameter values in Table 5.7, and the results are presented in Figure 5.15 and Figure 5.16 respectively.

These results fit with a general biochemical intuition. For example, it is expected and observed that at low levels of u_2, i.e. no L-rhamnose in the system, the concentration of TetR (y_1) is low as it is not induced, and therefore the concentration of RhaS (y_2) is high, as it is not repressed at all. Also, for high levels of u_1, i.e. aTc, the RhaS gene is not repressed, therefore a high level of y_2 is observed and the level of y_1 simply follows the expected dose response of the RhaS activator system.

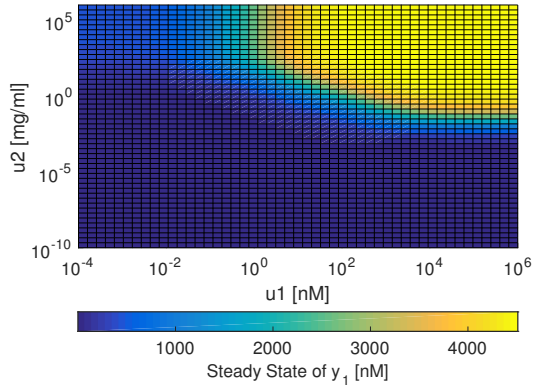
One surprise is the huge range of inducer values required to search to observe the full steady state behaviour of the system. The first inducer concentration u_1 ranges from 10^{-4}nM to 10^{6}nM, while the second u_2 ranges from 10^{-10}mg/ml to 10^{6}mg/ml. This is due in particular to the behaviour of the two state model employing Hill Functions; see the left column of Figure 5.16 in particular. It is clear that the saturation in the dose response for y_2 is only obtained at the very limits of these ranges. The eight state full model predicts all behaviour with in a much smaller range of inducer concentrations.

Another striking difference between the two models is the behaviour of y_1 at high levels of u_2 and low levels of u_1. The Hill Function model predicts a steady state of around 400nM whereas the full model predicts a steady state of around 3000nM, much reducing the step between maximum and minimum values. This is a feature of the system that can be checked experimentally.

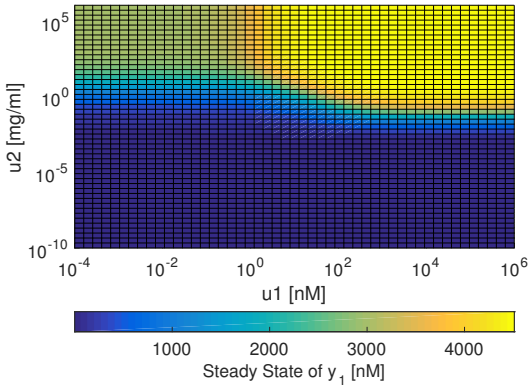
5.4.6 *Stochastic Models*

As with the other circuits, stochastic simulations are desired to shed light on expected noise properties at steady state. The system is modelled in three slight variations: naïvely using the model detailed in Equation 5.21, incorporating one plasmid, and finally incorporating two plasmids.

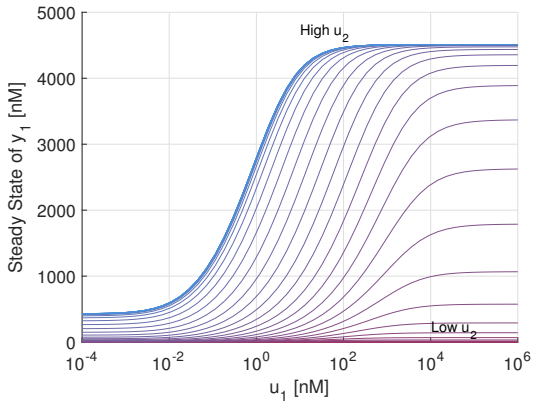
As the one plasmid model requires a way of modelling interactions between the two genes and the two TFs interacting with them, large changes to the model are required to incorporate



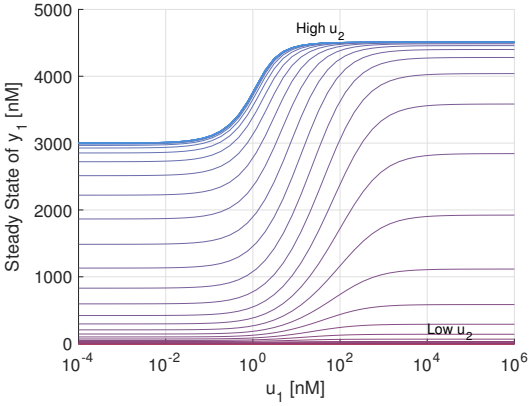
(a) Hill Function Model



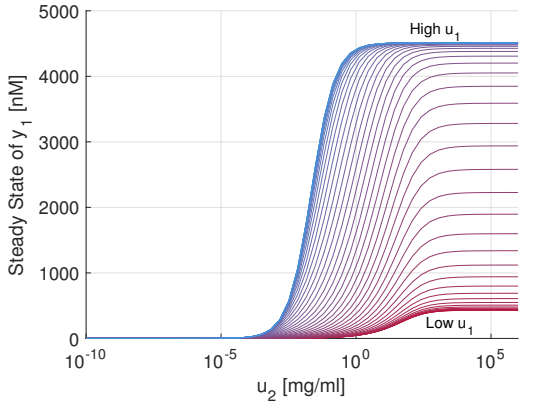
(b) Full ODE Model



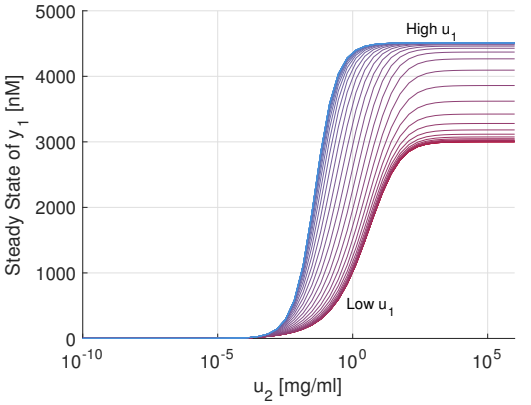
(c) Hill Function Model



(d) Full ODE Model

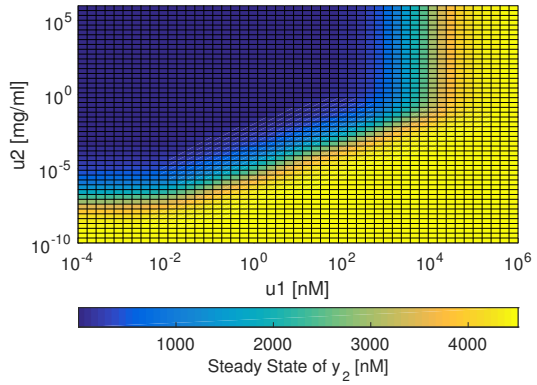


(e) Hill Function Model

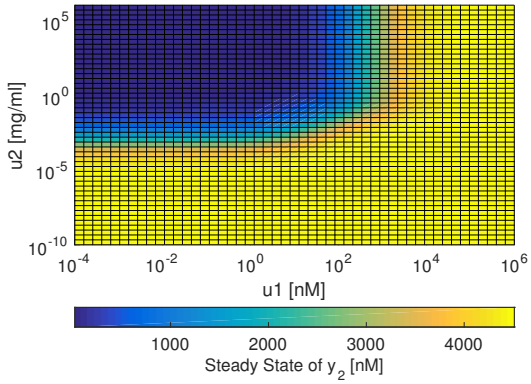


(f) Full ODE Model

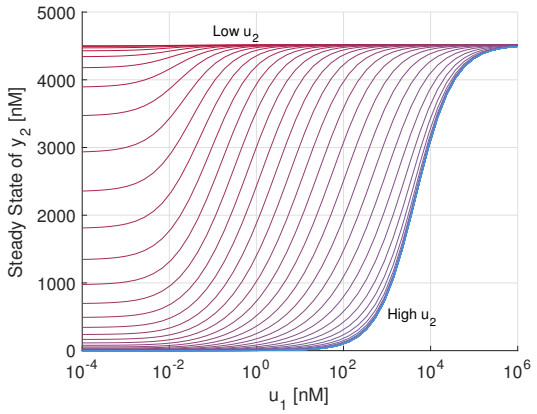
Figure 5.15 – Steady state values of y_1 found using the parameters presented in Table 5.7. The left and right columns are calculated numerically using Equation 5.22 and Equation 5.23 respectively. The inducer concentrations are used as inputs to the system and varied. The inducer u_1 is varied from 10^{-4}nM to 10^{6}nM, and the second inducer u_2 is varied from 10^{-10}mg/ml to 10^{6}mg/ml.



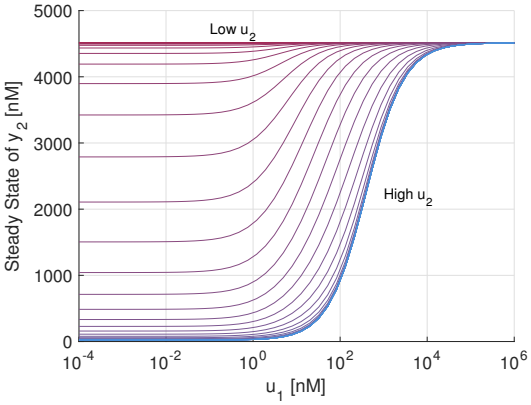
(a) Hill Function Model



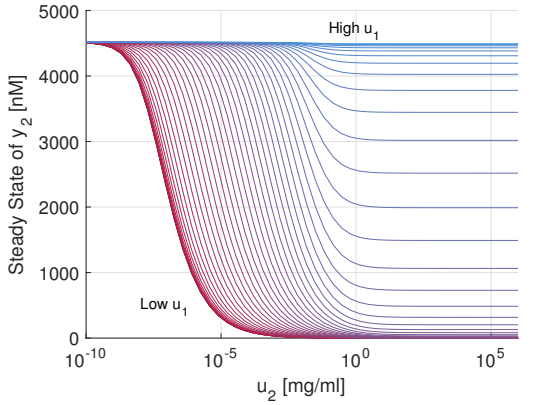
(b) Full ODE Model



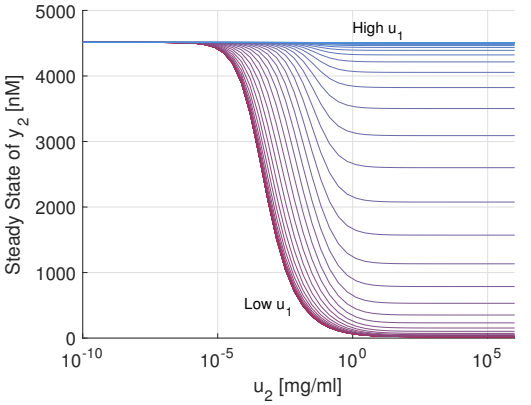
(c) Hill Function Model



(d) Full ODE Model



(e) Hill Function Model



(f) Full ODE Model

Figure 5.16 – Steady state values of y_2 found using the parameters presented in Table 5.7. The left and right columns are calculated numerically using Equation 5.22 and Equation 5.23 respectively. The inducer concentrations are used as inputs to the system and varied. The inducer u_1 is varied from 10^{-4}nM to 10^{6}nM, and the second inducer u_2 is varied from 10^{-10}mg/ml to 10^{6}mg/ml.

this. The model is therefore written out in full again with the necessary changes made. To model both genes on a single plasmid, the following model is used:

$$
\begin{aligned}
M_1 &\xrightarrow{\alpha_{P1}} M_1 + P_1, & & \\
G_{10} &\xrightarrow{\alpha_{M1}} G_{10} + M_1, & P_1 &\xrightarrow{\delta_{P1}} \emptyset, \\
G_{11} &\xrightarrow{\alpha_{M1}} G_{11} + M_1, & P_1U_1 &\xrightarrow{\delta_{P1}} \emptyset, \\
M_2 &\xrightarrow{\alpha_{P2}} M_2 + P_2, & M_1 &\xrightarrow{\delta_{M1}} \emptyset, \\
G_{00} &\xrightarrow{\alpha_{M2}} G_{00} + M_2, & P_2 &\xrightarrow{\delta_{P2}} \emptyset, \\
G_{10} &\xrightarrow{\alpha_{M2}} G_{10} + M_2, & P_2U_2 &\xrightarrow{\delta_{P2}} \emptyset, \\
P_1 + U_1 &\underset{k_{-11}}{\overset{k_{11}}{\rightleftharpoons}} P_1U_1, & M_2 &\xrightarrow{\delta_{M2}} \emptyset, \\
G_{00} + P_1 &\underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} G_{01}, & G_{00} &\xrightarrow{\delta_G} \emptyset, \\
G_{10} + P_1 &\underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} G_{11}, & G_{10} &\xrightarrow{\delta_G} \emptyset, \\
P_2 + U_2 &\underset{k_{-21}}{\overset{k_{21}}{\rightleftharpoons}} P_2U_2, & G_{01} &\xrightarrow{\delta_G} \emptyset, \\
G_{00} + P_2U_2 &\underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} G_{10}, & G_{11} &\xrightarrow{\delta_G} \emptyset, \\
G_{01} + P_2U_2 &\underset{k_{-12}}{\overset{k_{12}}{\rightleftharpoons}} G_{11}, & &
\end{aligned}
\tag{5.24}
$$

where the species and parameters are as in Equation 5.21, with the additions of G_{00}, G_{10}, G_{01}, G_{11}, which represent the plasmid in its four possible states. This first index denotes the G_1 position, the gene encoding the repressor, and the second index denotes the G_2 position, the gene encoding the activator. Then G_{00} is the plasmid not bound by of the TFs, G_{10} has the P_2U_2 complex bound to G_1, G_{01} has P_1 bound to G_2, and G_{11} has TFs bound at both sites. Note that M_1 can only be transcribed when the P_2U_2 complex bound to G_1, i.e. from G_{10} and G_{11}, whereas M_2 can only be transcribed when the site at G_2 is free, i.e. from G_{00} and G_{10}. The rates of production and degradation of plasmid, assuming a copy number of 17 are $\alpha_G = 0.425\text{min}^{-1}$ and $\delta_G = 0.025\text{min}^{-1}$ respectively. This completes the model of the Repressor-Activator Circuit on one plasmid.

Table 5.8 – These are the parameters used perform stochastic simulations of the Repressor-Activator Circuit. The parameters modelling the TF mechanisms are identical to those used in the above circuits.

Param.	Val.	Param.	Val.
α_{P1}	1min^{-1}	k_{11}	$3.3 \times 10^{-2}\text{min}^{-1}$
δ_{P1}	0.025min^{-1}	k_{-11}	$1.98 \times 10^{-4}\text{min}^{-1}$
α_{M1}	1min^{-1}	k_{12}	$9.6 \times 10^{-3}\text{min}^{-1}$
δ_{M1}	0.25min^{-1}	k_{-12}	$5.8 \times 10^{-2}\text{min}^{-1}$
α_{P2}	1min^{-1}	k_{21}	$2.7 \times 10^{-9}\text{min}^{-1}$
δ_{P2}	0.025min^{-1}	k_{-21}	1min^{-1}
α_{M2}	1min^{-1}	k_{22}	$9.6 \times 10^{-2}\text{min}^{-1}$
δ_{M2}	0.25min^{-1}	k_{-22}	$5.8 \times 10^{-2}\text{min}^{-1}$
α_G	0.425min^{-1}	G_{1tot}	17
δ_G	0.025min^{-1}	G_{2tot}	17
U_1	138	U_2	3.7×10^6

The case where the activator and repressor genes are on separate plasmids is the simpler of the two cases. This feature can be incorporated in the biochemical model detailed in Equation 5.21 using the following chemical equations:

$$\emptyset \xrightarrow{\alpha_G} G_1 \xrightarrow{\delta_G} \emptyset, \tag{5.25a}$$

$$\emptyset \xrightarrow{\alpha_G} G_2 \xrightarrow{\delta_G} \emptyset, \tag{5.25b}$$

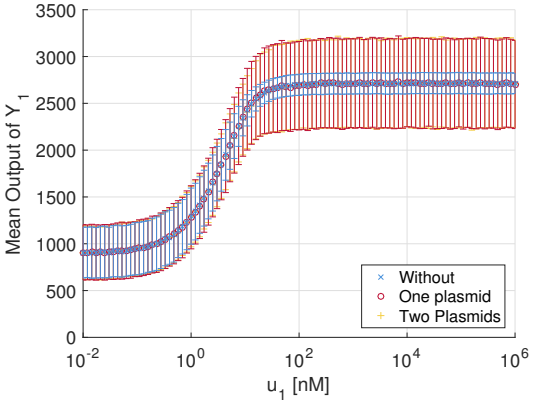
$$G_1P_2U_2 \xrightarrow{\delta_{P2}} \emptyset, \tag{5.25c}$$

$$G_2P_1 \xrightarrow{\delta_{P1}} \emptyset, \tag{5.25d}$$

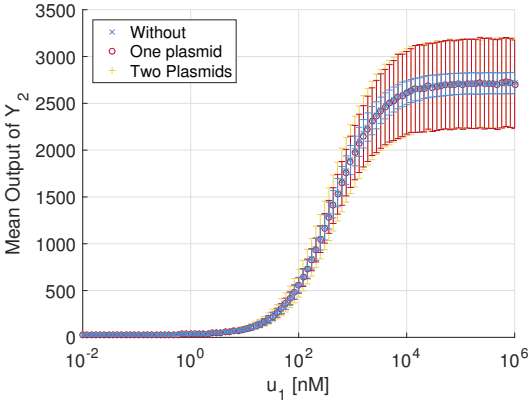
where Equation 5.25a and Equation 5.25b are simply added to the reaction network and model the production and degradation of the two plasmids with rates α_G and δ_G respectively. The rates of these are the same as in the one plasmid case. The reaction in Equation 5.25c replaces $G_1P_2U_2 \xrightarrow{\delta_{P2}} G_1$ and the reaction in Equation 5.25d replaces $G_2P_1 \xrightarrow{\delta_{P1}} G_2$. This completes the model incorporating two plasmids.

5.4.7 *Stochastic Results*

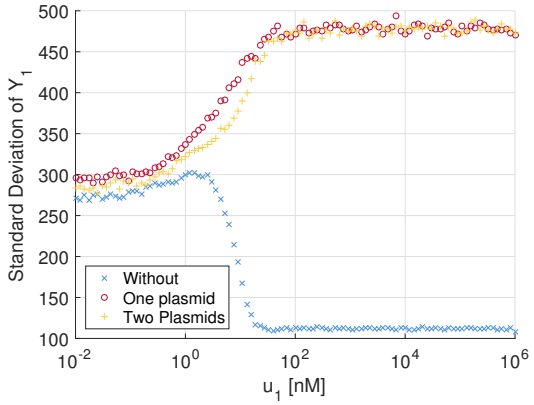
With the above three models, including a model without gene copy number fluctuation, a model with both genes on one plasmid and a model with the two genes on separate plasmids, stochastic simulations employing the Gillespie Algorithm as detailed in Section 2.2.4 are performed. The simulations execute ensembles of 5000 runs each and only use end point data, collected after 300min in simulation time, to ensure the observed behaviour is around steady state. The para-



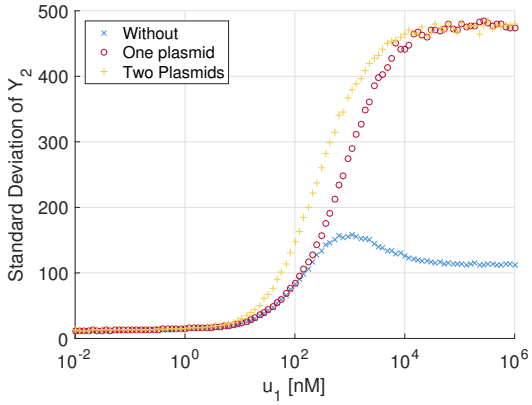
(a) Y_1: Mean



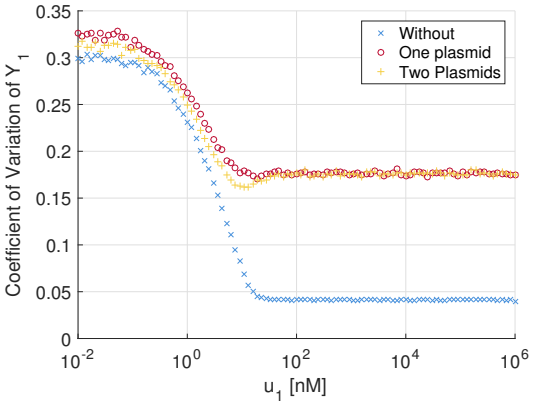
(b) Y_2: Mean



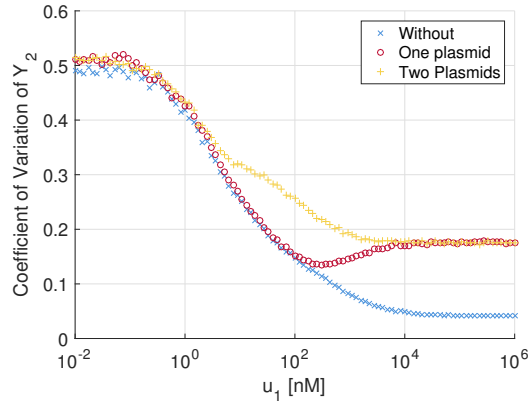
(c) Y_1: Standard Deviation



(d) Y_2: Standard Deviation

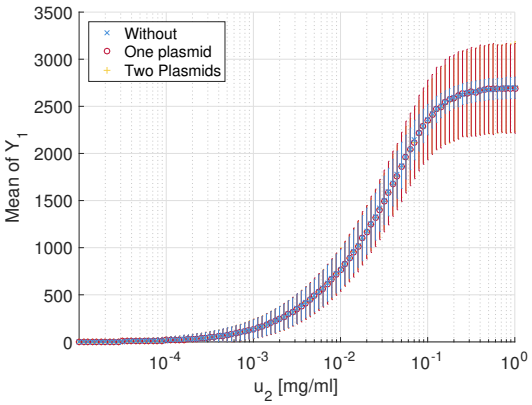


(e) Y_1: Coef. of Var.

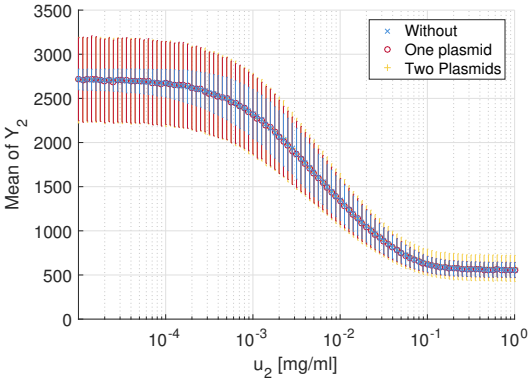


(f) Y_2: Coef. of Var.

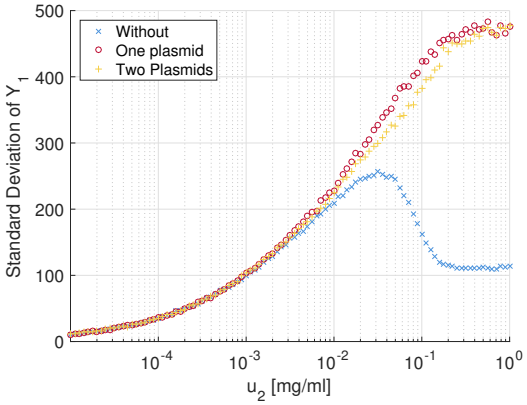
Figure 5.17 – Results from stochastic simulation of the three models of the Repressor-Activator Circuit: the one plasmid model, the two plasmid model and a model without gene copy number fluctuation. The results simulated using the parameters in Table 5.8, while sweeping the concentration of the first inducer u_1.



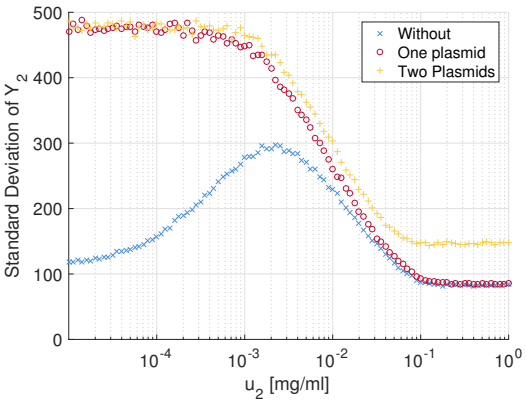
(a) Y_1: Mean



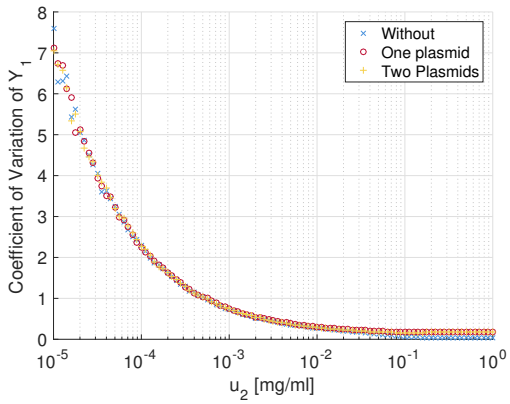
(b) Y_2: Mean



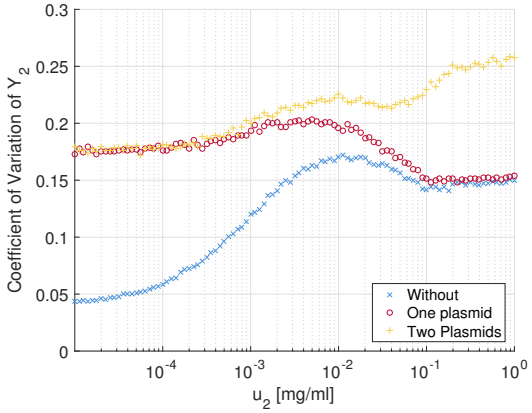
(c) Y_1: Standard Deviation



(d) Y_2: Standard Deviation



(e) Y_1: Coef. of Var.



(f) Y_2: Coef. of Var.

Figure 5.18 – Results from stochastic simulation of the three models of the Repressor-Activator Circuit: the one plasmid model, the two plasmid model and a model without gene copy number fluctuation. The results simulated using the parameters in Table 5.8, while sweeping the concentration of the second inducer u_2.

meters used are those presented in Table 5.8 which are identical to those in Table 5.7, but have been rescaled to have the correct units, with the additions of the parameters governing plasmid production and degradation.

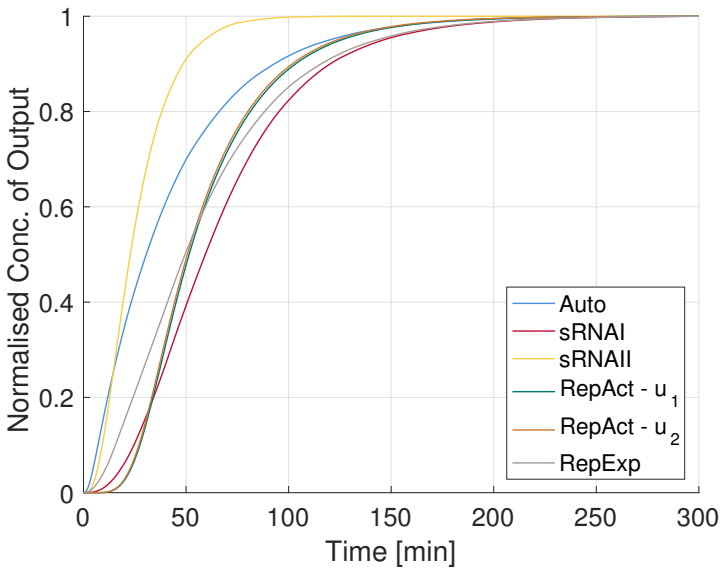
The results of these simulation are presented in Figure 5.17 and Figure 5.18. For the results in the first figure, the number of the free second inducer is fixed to $U_2 = 3.7 \times 10^6$, which is the equivalent of a concentration of 1mg/ml of L-rhamnose, while the concentration of the first inducer is varied from 10^{-2}nM to 10^6nM. For the second figure, the number of the free first inducer is fixed to $U_1 = 138$, which is the equivalent of a concentration of 100ng/ml of aTc, while the concentration of the second inducer is varied from 10^{-5}mg/ml to 1mg/ml. Note that in these figures, the left column presents results for Y_1 and the right column presents results for Y_2.

The overall behaviour of the means is in good agreement with the behaviour demonstrated by the full ODE model, as expected. For this circuit, the noise levels predicted do not substantially increase in the two plasmid case, as it does for the circuits including sRNA regulation. The two plasmid model tracks the noise expected of the one plasmid model closely. This comparison can be made experimentally. The reason for this might be that translation and the long half-lives of the TFs act as two additional filters to the intrinsic noise created by the relative fluctuations between the copy numbers of the two genes in the two plasmid case. This is not a luxury afforded to sRNAs, that are active just after transcription and have a target with a short half-life.

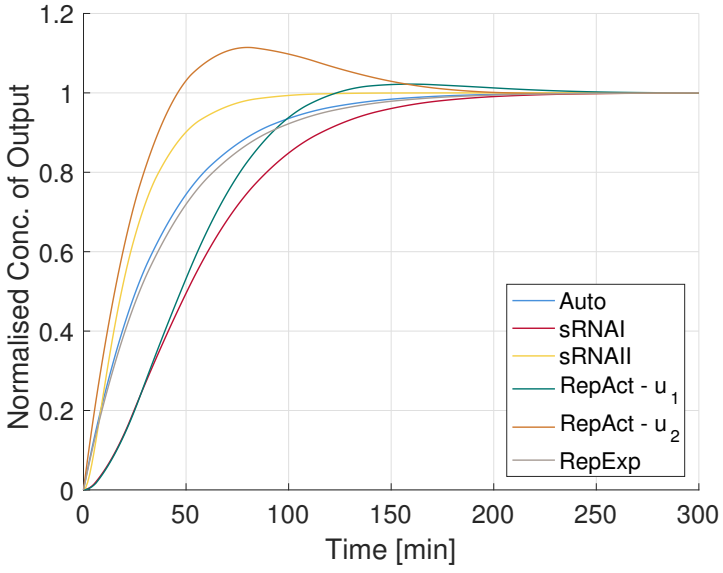
The results from stochastic modelling of the system show a region in which the Repressor-Activator Circuit is sensitive to both the concentration of U_1 and U_2. This is from about 1nM to 10^3nM of U_1 or aTc and 10^{-3}mg/ml and 1mg/ml of U_2 or L-rhamnose. These values correspond well to the inducer concentration used when these TF mechanisms are experimented with. These estimates may well be off due to the rough selection of the parameters used to model these systems, particularly in the case of the RhaS activator system. But it is this region that is of interest experimentally, as this is the region in which the noise profiles can be studied and verified.

5.5 COMPARISON OF DYNAMIC RESPONSES

The final aspect of these circuits to study is their expected dynamic response to a change in input. For these, the inducer concentrations have been used exclusively as inputs, as these are the easiest to implement experimentally. These dynamic responses are presented in Figure 5.19, and are simulated using both the reduced Hill Function models and the full ODE models. Repressed



(a) Full Models



(b) Hill Function Models

Figure 5.19 – A comparison of the dynamic responses of the circuits to a step change of 10% in inducer concentration. (a) The Autorepressor (Auto) is modelled using Equation 5.7, sRNA Circuit I (sRNAI) is modelled using Equation 5.11, sRNA Circuit II (sRNAII) is modelled using Equation 5.17, the Repressor-Activator Circuit (RepAct) is modelled using Equation 5.23, and Repressed Expression (RepExp) is modelled using Equation 5.8. (b) Auto is modelled using Equation 5.3, sRNAI is modelled using Equation 5.10, sRNAII is modelled using Equation 5.16, RepAct is modelled using Equation 5.22, RepExp is modelled using Equation 5.5. In the case of RepAct, the output presented is y_1 and the two time courses represent a step change to each of the two inputs. For the circuits that have aTc as an inducer input (Auto, sRNAI and RepAct), the starting concentration used is 10nM, for the circuits which have L-rhamnose as an inducer input (sRNAII and RepAct), the starting concentration used is 0.5mg/ml. The parameters used are in Table 5.2, Table 5.3, Table 5.5 and Table 5.7 for the four circuits respectively. For RepExp $p_{tot} = 2000$nM.

expression is included for comparison to an open-loop system. The change in input is a step of 10% in inducer level. For all systems except sRNA Circuit II, this is an increase in inducer concentration. For sRNA Circuit II this is a 10% decrease to ensure the output increases. In the case of the Repressor-Activator Circuit, both inducers are perturbed with the 10% increase separately to produce two time courses: one when varying each input. The output of the Repressor-Activator Circuit presented in Figure 5.19 is y_1 corresponding to the concentration of TetR-sfGFP fusion. For the cases where the inducer represents aTc the starting concentration is 10nM. When the inducer in the experimental system is L-rhamnose, the starting concentration is 0.1mg/ml. The time courses are normalised with respect to their steady state such that they can be compared with ease. The responses have been checked for other values of inducer concentrations and though the responses do change to some extent, the overall picture remains similar.

For both the Hill Function models and the full models, the Autorepressor and sRNA Circuit II have very similar and short rise times[3], and sRNA Circuit II has the shortest settling time[4] of all the circuits. sRNA Circuit I consistently has the longest rise time.

Looking at the behaviour of the Repressor-Activator Circuit, it is clear that the two models provide different predictions of the response. The Hill Function model predicts an overshoot in the case of perturbing both inputs. This overshoot is not present in the simulations produced with the full model, substantially reducing the settling time of the system. The Hill Function model for this circuit predicts the fastest rise time for this circuit when u_2 is perturbed. Though despite this, the Hill Function model of this circuit just about ties sRNA Circuit I for settling time.

As such, use of sRNA regulation in a manner such as is done in sRNA Circuit II might be able to provide fast responses to perturbations. This fast response might come at the price of increased noise, depending on the experimental set-up. These responses provide another measurable metric of these systems that can be studied.

5.6 CONCLUSION

In this chapter four negative feedback circuits were studied in depth with the aim of predicting their behaviour with a specific focus on metrics that can be extracted from experiment. These four circuits are simple and the implementations of each consist of tried and tested mechanisms. The Autorepressor has been previously studied in depth [23, 62, 179] and is used as a solid starting point for the rest of the chapter. The three other circuits presented are novel. Each circuit is

3 Defined to be the time taken to get from 10% to 90% of the steady state.
4 Defined to be the time taken to reach and stay with 5% of the steady state.

presented with the description of a biological implementation. This is aimed to guide the choice of models and parameters used. The modelling for each circuit begins with the construction of a biochemical reaction network that is then transformed into a reduced ODE model employing Hill Functions to model interactions between TF, inducers and genes. A second ODE model is then developed which explicitly models these interactions. The steady state behaviour of the experimental output, the total concentration of an expressed transcription factor, is then calculated for the two ODE models, and these results are compared. Once this has been done, a stochastic simulation is used to predict the noise profiles expected of the circuits in an experimental context.

To this end, the next step is to perform the relevant experiments using the implemented circuits described in this chapter and to analyse the results of these in the context of the predictions made here. The steady state and noise data can be collected from cultures in exponential phase containing the correct vectors and a range of inducer concentrations using flow cytometry. This measures the fluorescence individual cells, giving a description of the heterogeneity across a population.

At the theoretical end, modelling techniques and methods such as Hill Functions need to be investigated in more detail, along the lines of [91]. It is clear from the above modelled circuits that in some situations the Hill Function approximation agrees closely with un-reduced models of higher order. But in other situations, these two models do not agree. The reasons for this must be understood in more detail, as model reduction is a key to being able to perform analysis on complex systems.

6 EXPERIMENTAL DATA

The aim of this chapter is to present and analyse data collected on the circuits described in Chapter 5. The chapter begins with a materials and methods section which briefly presents the experimental techniques used. The plasmids used to produce the data in this chapter are listed in Table 6.2 and Synthetic Biology Open Language (SBOL) diagrams of these plasmids are appended to the chapter [74]. The DNA sequences of ordered gblocks are presented in Table 6.3. A break-down of all the parts and their sequences is also provided. The genes encoding proteins used in the circuits are listed in Table 6.4. RBSs and promoters used in the circuits are listed in Table 6.5. The parts used to construct the sRNAs, including the cassette and target regions are listed in Table 6.6. The parts making up the two plasmid backbones used, including terminators, resistance cassettes and *ori*-regions are listed in Table 6.7 for the pJ404 backbone and Table 6.8 for the pSUtat backbone.

Initially, culture level data is presented for the Autorepressor, and sRNA Circuits I and II. This mainly consists of time course data collected in a plate reader. Then, a time course of the Autorepressor has an ODE model fit to it. Finally, as stochastic modelling techniques are used in Chapter 5, cell population data for the Autorepressor, sRNA Circuit II and the Repressor-Activator Circuit is presented studying the mean and noise profiles of the fluorescence across cell populations.

6.1 MATERIALS AND METHODS

As this chapter presents experimental results, it begins with a quick run through of the techniques used.

The strains of *E. coli* used are DH5α for cloning and the K12 strain MG1655 for all assays, except those testing circuits including RhaS. For these, the Δ*rhaS* strain JW3876 is used.[1] In this strain, the chromosomal gene *rhaS* has been removed, ensuring that no *rhaS* is expressed from the chromosome. The gene is replaced with a gene encoding a kanamycin resistance protein.

1 The full genotype of JW3876 is F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, ΔrhaS765::kan, hsdR514.

This ensures that chromosomal RhaS does not interfere with the synthetic circuit. The strain used is a member of the Keio Collection [17].

The antibiotics used in cultures are ampicillin (AMP) at 125 µg/ml and chloramphenicol (CML) at 25 µg/ml. On agar plates carbenicillin (CARB) is also used as an analogue to AMP as it does not degrade as fast.

Transformations are done by heat shock at 42°C using a standard transformation buffer. Cloning is performed using Gibson Cloning [80] and DNA amplification is done using the Polymerase Chain Reaction method.

The sources of the pJ404 backbone, sfGFP, RhaS and P_{rhaBAD} are described in [113]. The translation of RhaS is always controlled by the naturally occurring RBS, that scored 20k using the 'RBS Calculator' [184]. The source of the plasmid backbone, pSUtat, is described in [112]. Three gblocks were ordered from IDT. The sequences of these are in Table 6.3. The biggest block was the autorepressor cassette, which consists of P_{tet}, an optimised RBS with an 'RBS Calculator' score of 55k [184], and the *tetR* gene fused to the first few codons of sfGFP. At the beginning of the block there is an overhang that matches the pJ404 backbone. The two other blocks contain the two versions of the TetR targeting sRNA. These consist of overhangs at either end to facilitate cloning into pSUtat, the 24-base target binding regions followed by the *micC* sRNA cassette and a terminator. The two gene-blocks only differ in the 24-base target binding region. The sRNAs that target the RhaS mRNA were created using PCR, with long primers used to change the targeting regions of the sRNA gene-blocks. These pieces are then combined to create the circuits.

Three different media are used for growth of cells: Luria Broth (LB), M9 Minimal Media (M9) and EZ Rich Media (EZ). EZ rich media is commercially available from Teknova [153]. It is a rich media, so cells grow nearly as rapidly as they do in LB, but it has the great benefit that it is defined, so the constituents for each experiment are the same. It is also clear and therefore does not interfere with measurements. M9 is useful as it greatly slows down the growth of the cells, so their response is also greatly slowed.

For all experiments, MG1655 or JW3876 are transformed with the relevant plasmids and plated. Single cells are then selected for pre-cultures in 3ml of LB and the relevant antibiotics are incubated for 16 hours at 37°C with shaking. For the experiments using EZ or M9, to prepare for inoculation, the cells are washed and then resuspended in the relevant media.

All the data is taken with at least three independent biological repeats during the same experiment. That is to say that the data is collected from three identical experiments, usually run simultaneously. The sole difference between these is that each of the three experiments is inocu-

lated with cells picked from each of three single colonies transformed with the same plasmid. The mean of the three biological repeats is the quantity presented.

6.1.1 *Culture Level Assays*

Nearly all the culture level data is collected in a plate reader, similarly to the data in [100, 179]. A plate reader is a machine that incubates and mixes samples in a plate, classically with 96 wells, and takes regular measurements of each well. The measurements are of the concentration and fluorescence intensity of the cells in each well. The data collected is therefore fluorescence and growth over time, yielding a time course. Two plate readers are used here, both from BGM LABTECH: a FLUOstar OMEGA and a CLARIOstar.

The quantity of cells is measured through the absorbance of light at 600nm through a known path length of sample. This quantity is known as the optical density and is often referred to as 'OD600', referencing the wavelength of the light used. The range over which this measurement is proportional to concentration of cells is up to an OD600 value of about one. Above this value, the relationship between the OD600 reading and the cellular concentration is non-linear and therefore measurements taken at OD600 values above one cannot be used for quantitative analysis.

The fluorescence intensity measurement is taken by first exciting the sample, and then measuring the intensity of the emitted light at a certain wavelength. In the case of sfGFP used here, the excitation wavelength is 485nm and the emission peak is around 510nm [164]. Photobleaching is not an issue as measurements are taken using a quick succession of flashes. Bleaching only occurs after prolonged illumination of the order of an hour [164]. The gain of the plate reader used is set so that the maximum fluorescence measurement gave a reading at about 80% of the saturation level of the machine.

To produce a measurement of OD600 and the fluorescence intensity for a well, the plate reader takes a number of measurements of each, from three to ten, and averages them for a reading.

The 96 well plates are pipetted out to a total volume of 200µl per well and are covered in a breathable clear film. This limits evaporation in the plate reader when incubating at 37°C. The plates used are flat bottom plates which are black to stop interference by radiation from neighbouring wells. The bottom of the plate is clear and the readings are taken through the plate.

When studying the time course data produced by these experiments, the value of interest in the fluorescence intensity per cell. For a particular time point, this is calculated by dividing the fluorescence intensity measured for a well by the OD600. This process is called normalisation and is performed on all culture-level data presented.

6.1.2 *Cell Population Assays*

The cell population data is collected with a ThermoFisher Scientific Attune Flow Cytometer. The standard pre-culture protocol is followed for transformation, picking and overnight incubation. Overnight cultures were subcultured into 300 µl of fresh EZ in deep well plates, to a final OD600 of 0.05. The plates were then incubated at 37°C with rapid shaking. A sample is taken and diluted 1:10 into phosphate-buffered saline to a total volume of 200µl. The samples are all collected in a plate and inserted in the flow cytometer which then takes the reading. The dilution ensures that the cells are less concentrated, which increases the quality of the data. The flow cytometer passes the sample through a thin capillary, exciting and taking measurements from a single cell at a time. With this technique, it is possible to take measurements on thousands of cells per second.

The initial data produced includes many events that are not desired for later analysis as they do not represent live healthy cells. This includes debris in the solution and dead cells. This is then gated, using forward and side scatter, to include about 10000 events and the moments are then calculated. The data presented consists of the mean fluorescence intensity per cell and the standard deviation and coefficient of variation of this across the population. Again, all of this is performed with three biological repeats taken simultaneously.

6.2 CULTURE LEVEL DATA

The first half of this chapter focuses on data collected through fluorescence measurements of the entire bacterial culture. As a proxy for protein concentration, the emission intensity of a reporter protein, in this case sfGFP (see Section 2.1.17), is measured. The intensity is assumed to be proportional to the concentration of the reporter protein in the sample. As the reporter proteins used in the circuits studied here are fused to TFs, the intensity gives a measure of the concentration of TF. This is the species that is the focus of modelling. This is a standard assumption in the literature [16, 19, 23, 62, 66, 76, 100, 155, 179].

6.2.1 *Autorepressor Data*

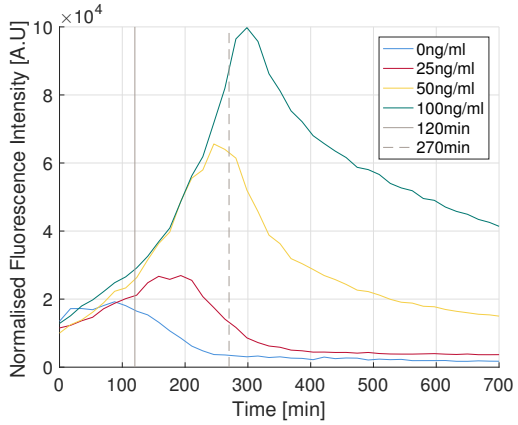
Time course data from two plate reader experiments employing the Autorepressor (pAH01) in MG1655 are presented in Figure 6.1. Both these experiments use EZ with ampicillin, and were set up almost identically, with various levels of the inducer aTc.

Looking first at the OD600 data in Figure 6.1(c) and (d), it is clear to see that the cells grow almost identically. There is a clear lag phase, prior to the exponential growth phase marked by the vertical guides. It is not surprising that the cells with no aTc grow faster than the other experiments. These cells have the benefit of not having to express large quantities of the TetR-sfGFP fusion in addition to not having to cope with the aTc. After the exponential phase, the cells are starved of nutrients and move into what is called stationary phase.

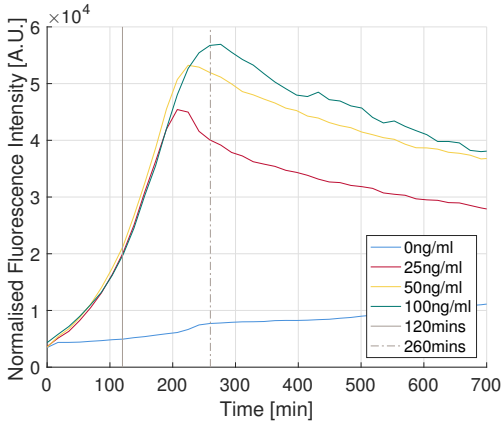
Figure 6.1(a) and (b) show the fluorescence intensity data from each experiment. This has been normalised by OD600 to give the fluorescence per cell. As opposed to the x-axis, which has a well defined scale with the units of minutes, the scale on the y-axis is somewhat arbitrary and not important and is therefore labelled Arbitrary Unit (AU). This is due to the fact that data from two different experiments is never quantitatively compared, so the gain and exact machine employed are not vital. The level of expression increases with the amount of inducer present. This is in agreement with intuition and also with the steady state modelling results (see Section 5.1), though the experimental system does not reach steady state. All the traces including inducer rise through to the end of the exponential phase and then drop thereafter as the cells enter stationary phase. Apart from this, the normalised fluorescence intensity traces for the two plate reader experiments presented in Figure 6.1a and Figure 6.1b differ substantially. The largest difference being the drop in normalised fluorescence intensity after the end of the exponential phase; this occurs much faster in Figure 6.1a than in Figure 6.1b.

Figure 6.1 gives a good impression of the type of data commonly found in plate readers. The reasons for the differences could be due to experimental variability as these experiments rely on a large amount of manual pipetting with small volumes, which is prone to error even in experienced hands. The growth rates are very similar, which indicates healthy cells.

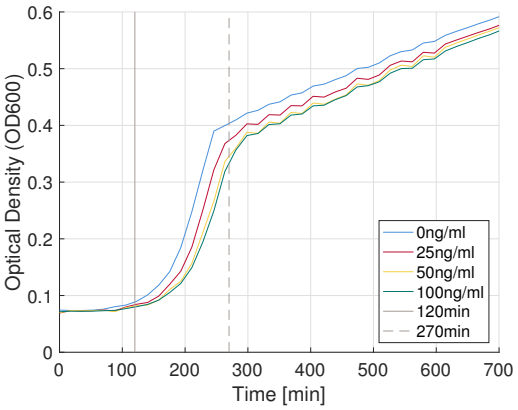
Plate reader experiments often provide inconsistent data. The examples in Figure 6.1 are two of many plate reader experiments with the Autorepressor that all exhibit slightly different behaviour. This makes plate readers a challenging tool to use when attempting to model and characterise genetic circuits. To engineer in Synthetic Biology, experimental tests of GRNs would need produce repeatable and similar results that with ease could be quantitatively studied and



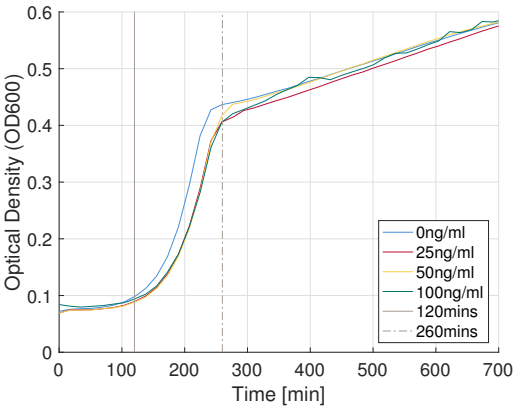
(a) Exp. 1: Fluorescence Intensity



(b) Exp. 2: Fluorescence Intensity



(c) Exp. 1: OD600



(d) Exp. 2: OD600

Figure 6.1 – Two plate reader experiments of the Autorepressor (pAH01) in EZ with various amounts of aTc. The two experiments were prepared identically. The vertical guides approximate the beginning and end of the exponential growth phase. See text for a discussion.

compared to modelling such as that presented in Section 5.5. In the case of the Autorepressor, this criteria is not fulfilled.

6.2.2 *sRNA Circuit I Data*

When studying sRNA Circuit I, the first aspect that required testing was the sRNA inhibition mechanism. Two separate methods of regulation were proposed: Yoo *et al.* proposed that the sRNA should target a region of the target mRNA beginning with the Start Codon [224], and Man *et al.* proposed that the region targeted should begin with the Shine-Dalgarno sequence [134]. Both these designs were implemented with a two plasmid set-up. The two circuits both contain pAH01, containing the Autorepressor, and a second plasmid, containing the sRNA gene. The plasmid pAH09 contains the gene encoding the sRNA targeting a region beginning with the Start Codon, and pAH06 has the sRNA that targets a region beginning with the Shine-Dalgarno. Apart from this difference, these two plasmids are identical with expression controlled by two tandem constitutive promoters: P_r and P_{proD}.

In the case of both designs, the two required plasmids were transformed into the *E. coli* strain MG1655. Both the antibiotics ampicillin and chloramphenicol were added to the media to ensure that the cells are not cured of either plasmid. The standard pre-culture protocol was followed and a plate reader experiment was performed in EZ with a dilution of 1:100 into the wells with a total final volume of 200µl.

Plate reader data from this experiment is presented in Figure 6.2. Only data up to the end of the exponential growth phase of the cells is shown. The Autorepressor, on its own, is included for comparison. As expected the circuits corresponding to sRNA Circuit I show lower fluorescence than the Autorepressor, which indicates that the sRNA mechanism is working. This is expected as the tandem constitutive promoters are particularly strong. The circuits in which the sRNA targets a region beginning with the Shine-Dalgarno are regulated more tightly than those targeting a region beginning with the Start Codon. This can be observed as the fluorescence intensity from these circuits is lower. This supports what is reported by Man *et al.* [134]. For this reason, all further experiments performed with sRNA Circuit I use the sRNA binding a region beginning with the Shine-Dalgarno sequence.

The next data set, presented in Figure 6.3, is from a plate reader experiment containing cells in LB, which was used as it is optimised for growth rate. Though is is not defined, as opposed

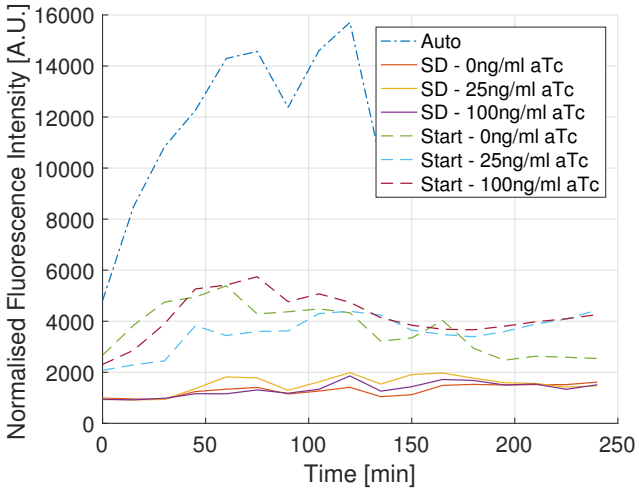
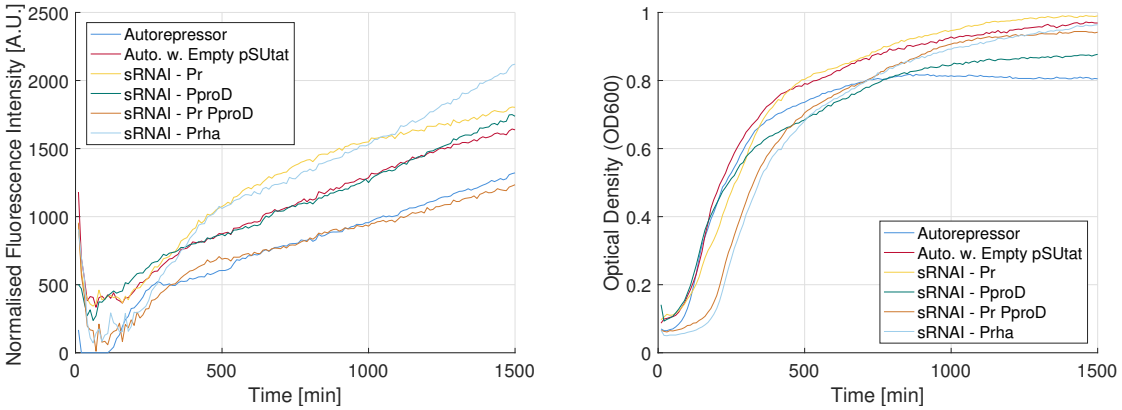


Figure 6.2 – Experimental results comparing the two implementations of sRNA Circuit I. Auto denotes cells with pAH01 only. SD denotes cells with sRNA Circuit I, where the sRNA's target region begins with the Shine-Dalgarno (pAH01 and pAH06). Start denotes cells with the circuit, where the target region begins with the the Start Codon (pAH01 and pAH09). Expression of the sRNAs is constitutive from P_r and P_{proD}. The respective concentrations of aTc are listed. Only data up to the end of the exponential growth phase is presented. The Autorepressor on its own is presented for comparison.



(a) Fluorescence (b) Optical Density

Figure 6.3 – Data from a plate reader experiment focussing on sRNA Circuit I using LB. The plasmids used: Autorepressor is pAH01, Auto. w. Empty pSUtat is pAH01 and pAH23, sRNAI - Pr is pAH01 and pAH07, sRNAI - PproD is pAH01 and pAH08, sRNAI - Pr PproD is pAH01 and pAH06, sRNAI - Prha is pAH01 and pAH17. Prha here is short for P_{rhaBAD}. See text for discussion.

to the two other choices of media, and therefore the precise make-up varies from batch to batch. The wild type strain MG1655 is used again and the standard pre-culture protocol is followed.

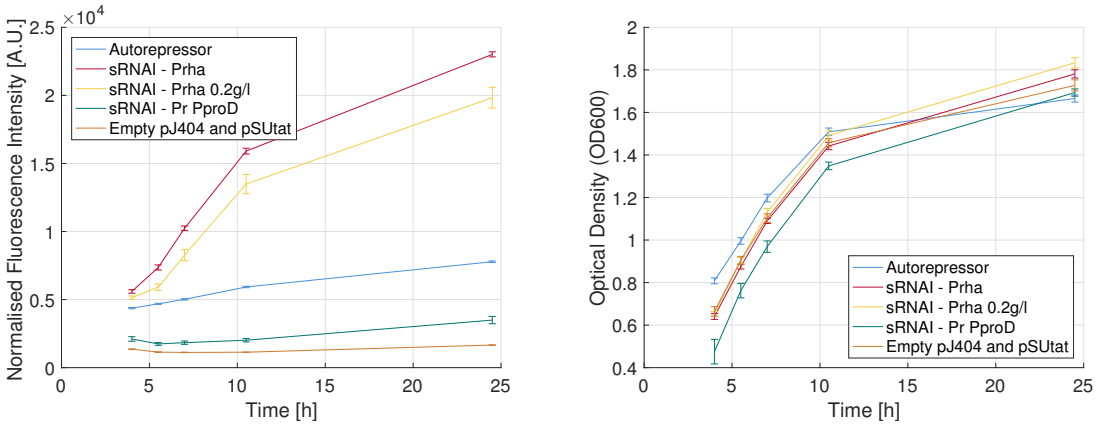
The aim of the experiment was to compare versions of sRNA Circuit I with various expression rates of sRNA. The promoters used to control the expression of the sRNA, in order of expected strength from high to low, are the two tandem P_r and P_{proD}, P_{proD} only, P_r only and then P_{rhaBAD}. No L-rhamnose is used to induce the P_{rhaBAD} promoter, so it would be expected that the trace shown by this circuit is similar to that of the Autorepressor alone. To improve the comparison, two versions of the Autorepressor are used. The first is cells containing just pAH01, the Autorepressor plasmid. The second is cells containing both pAH01 and pAH23. The second of these plasmids is the empty pSUtat plasmid, including only the *ori*-region and the antibiotic resistance gene.

Again, beginning with the OD600 growth curves presented in Figure 6.4b, the growth of the cells is similar. Turning to the fluorescence intensity data presented in Figure 6.4a, the traces for the Autorepressor with the empty pSUtat would be expected to be at the same level as the Autorepressor. Experimentally, it is observed to be considerably higher than the Autorepressor. The traces for sRNA Circuit I with sRNA expression controlled by un-induced P_{rhaBAD} and P_r only are also considerably higher. This is not what would be expected. The presence of the sRNA is expected to reduce the level of fluorescence (see Section 5.2).

From the above, it seems that the addition of the plasmids containing the pSUtat backbone and the additional antibiotic, chloramphenicol, that is then required has an effect on the Autorepressor. The TetR repression mechanism might be weakened by the presence of these. Another possible option, is that the pSUtat backbone and the chloramphenicol fluoresce independently at similar wavelengths to the sfGFP.

To probe this further another experiment was performed with sRNA Circuit I. It was decided that growth in a larger volume in a falcon tube might be more consistent and therefore provide clearer data than that provided by the plate reader. From pre-cultures that stood overnight, the cultures were grown in 3ml of EZ in falcon tubes incubated at 37°C and with shaking. The data was sampled manually and pipetted into a plate to a total volume of 150µl per well. A plate reader was then used to take each measurement. The data for this experiment is presented in Figure 6.4.

To test whether the empty pSUtat and chloramphenicol fluoresces, the combination of both empty pJ404 and empty pSUtat backbones is used as a control. The version of sRNA Circuit I where expression of the sRNA is controlled by the inducible promoter P_{rhaBAD} is included, both with and without inducer. The inducer concentration used is 0.2mg/ml, which is a relatively high



(a) Fluorescence Intensity (b) Optical Density

Figure 6.4 – The is time course data for variations of sRNA Circuit I grown in an incubator. The plasmids used: Autorepressor is pAH01, sRNAI - Prha is pAH01 and pAH17, sRNAI - Pr PproD is pAH01 and pAH06, empty pJ404 and pSUtat is pAH24 and pAH23. The version of sRNA Circuit I with expression of the sRNA controlled by P_{rhaBAD} is including both un-induced and with 0.2g/l of L-rhamnose.

level of induction. The plain Autorepressor and sRNA Circuit I, with sRNA expression controlled by both P_r and P_{proD}, are also included.

Studying Figure 6.4, the growth data is as expected, with tight error bars indicating the standard error between the three biological repeats. The transition between exponential and stationary phase occurs at around 11 hours. This experiment has therefore provided a much longer period of growth than the plate reader experiments do. Hence, growth conditions were improved substantially by this protocol. It must be kept in mind that the OD600 measurements for much of this experiment are outside the strictly linear range. The data points with OD600 measurements above one are kept as this point is reached early on in the time course due to favourable growth conditions.

The fluorescence intensity data as a result of this protocol is clearer than the plate reader data in Figure 6.3, and the results from the plate reader are repeated. The versions of sRNA Circuit I, containing the pSUtat plasmid with sRNA expression controlled by P_{rhaBAD}, show much higher levels of fluorescence than the Autorepressor plasmid alone. The induced version of this circuit has a lower level of fluorescence than the un-induced circuit, which confirms that the inducible promoter is functioning. This is also expected from modelling (see Figure 5.9), and is the equivalent of increasing the ratio α_R/α_M. Interestingly enough the induced circuit has a much higher standard error on the data.

A reason why the fluorescence, after induction with L-rhamnose, remains high could be that the circuit only relies on chromosomally expressed RhaS. Due to this, the concentration of RhaS might not be at levels high enough to substantially activate expression of the sRNA. This can be

mitigated by synthetically expressing RhaS constitutively, increasing the concentration to improve activation. Another option might be that P_{rhaBAD} is simply too weak a promoter to be of use. The second issue could be tackled by increasing gene copy number by using another *ori*-region, or using another activator system such as *araC* or *xylS*.

This data supports the results in Figure 6.3 suggesting that fluorescence is increased due to the addition of the pSUtat plasmid with chloramphenicol. These components seem to have an effect on the *tet*-repressor mechanism. If the pSUtat and chloramphenicol fluoresced, this would be seen in the trace of the empty pJ404 and empty pSUtat backbones, but the fluorescence intensity for this stays low. The reason why the trace of sRNA Circuit I, with expression of the sRNA controlled by both P_r and P_{proD}, remains low is that the combination of these two strong constitutive promoters is strong enough to overcome the effect of the pSUtat and the chloramphenicol on the *tet*-repressor mechanism.

6.2.3 *sRNA Circuit II Data*

This section presents plate reader data for sRNA Circuit II. The first aspect that needed testing was the two different sRNA mechanisms, similarly to the case of sRNA Circuit I. In the left column of Figure 6.5, the two versions of sRNA Circuit II, containing the two different sRNAs, are compared. Cells transformed with the plasmid pAH12 alone, containing the gene encoding the RhaS-sfGFP fusion protein constitutively expressed under control by P_{proD}, and a negative control are also included for comparison. The experiment is performed using EZ with the standard pre-culture protocol and the Δ*rhaS* strain JW3876.

The growth curves for the three circuits expressing the RhaS-sfGFP fusion are as expected. The growth rates for cells with the two-plasmid circuits are lower than the cells with only pAH12. Interestingly enough, the cells containing the sRNA targeting a region starting with the Shine-Dalgarno grow faster than the cells containing the sRNA targeting the a region that begins with the Start Codon. This is also seen in data from a similar experiment run using M9, but is not presented here. This might be due to off-target effects of the sRNA targeting the Start Codon, though nothing substantial came up when this was checked using the Target RNA Tool [114].

This theory is supported by the growth curve for the negative control. The negative control consists of cells transformed with pAH15 only. This is the pSUtat backbone with the sRNA targeting a region beginning with the Start Codon. The growth of the negative control is very slow, indicating that the sRNA might interfere with the cell somehow. In the case of the full circuit

spread across two plasmids with the sRNA targeting a region beginning with the Start Codon, the sRNA has both the target RhaS-sfGFP and native targets to bind, so its effect is less severe. The negative control in this case just expresses the sRNA with no synthetic target, so it is not surprising that it has a much more marked effect on growth.

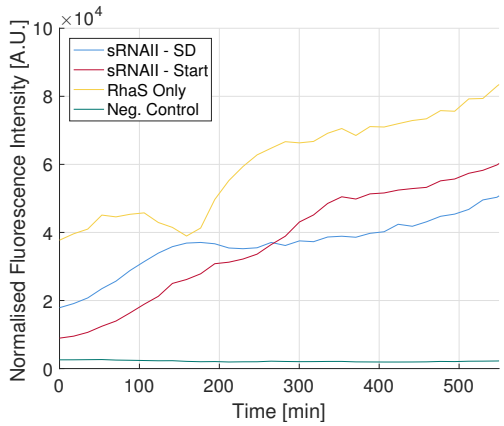
The fluorescence intensity data for this comparison shows that the circuits including the sRNA have lower fluorescence than pAH12 expressing RhaS on its own. This indicates that the sRNAs are functioning. Deciding which of the sRNAs to use based on the fluorescence intensity data is difficult: the sRNA targeting the Start Codon regulates more efficiently at first, but at later times the sRNA targeting the Shine-Dalgarno seems to result in tighter regulation. This results was repeated in M9, but not presented here.

In conclusion, the version of the sRNA targeting a region beginning with the Shine-Dalgarno was chosen as the preferred sRNA, as growth with this sRNA is better. The growth of the negative control clearly indicates some sort of interference with the host. But this example shows that there is no clear-cut method for designing sRNAs, and that different systems will require designs using different mechanisms. Therefore implementing and testing a few of the suggestions made by both Yoo *et al.* [224] and Man *et al.* [134] are important when implementing sRNA regulation.

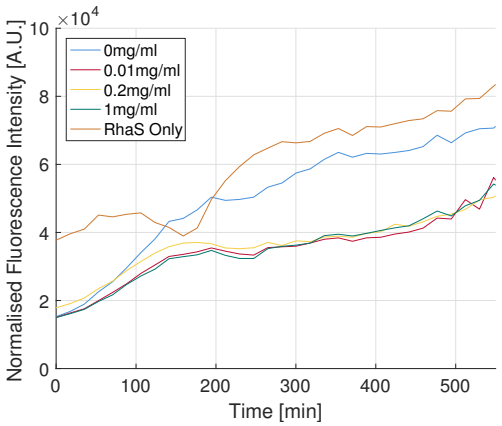
The data in the right column of Figure 6.5 compares sRNA Circuit II at various levels of induction. The data was collected in EZ using the standard pre-culture protocol. The sRNA used targets a region beginning with the Shine-Dalgarno. The circuit is compared to pAH12, which constitutively expresses RhaS, on its own. The growth for all systems is as expected, with the lone pAH12 growing fastest. This is due to lower burden and a single antibiotic as opposed to two.

Turning to the fluorescence data, it would be expected that the un-induced sRNA Circuit II would express at about the same level as pAH12 on its own. The levels observed are close, but there is still a clear difference. This may well be due to the burden of the inclusion of the second plasmid. As expected the induced circuits show lower levels of expression than the un-induced circuit. It would be expected for the various levels of induction to produce a graded expression level, with a decreasing level of output as the inducer level increases, as is predicted in Figure 5.11. This is not seen in the data presented in Figure 6.5b.

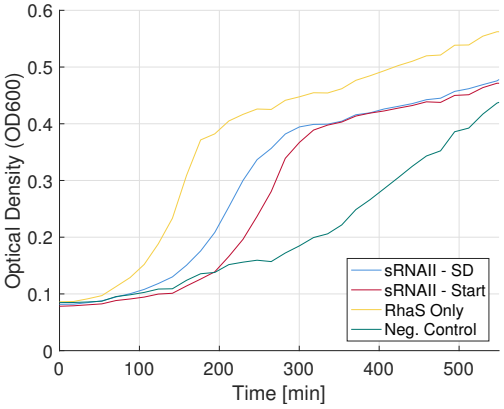
It is possible that the RhaS activation mechanism is already saturated at 0.01mg/ml of L-rhamnose and therefore the system exhibits the lower level of expression seen in Figure 5.11 at inducer concentrations above 1mg/ml. As the protein is very difficult to handle experimentally [118] and therefore the choice of parameters is guided but rough, the focus of the model is to capture the qualitative behaviour. This is successfully done in this case, as the model correctly predicts a drop in expression when the circuit is induced.



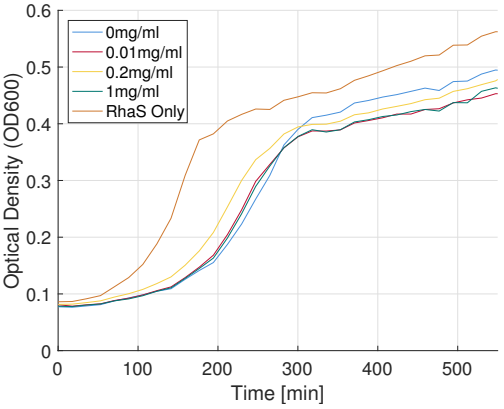
(a) Exp 1: Fluorescence Intensity



(b) Exp 2: Fluorescence Intensity



(c) Exp 1: Optical Density



(d) Exp 2: Optical Density

Figure 6.5 – The results from two experiments studying sRNA Circuit II. In experiment 1 (Exp. 1), the circuits used are: sRNAII - SD is pAH12 and pAH16, sRNAII - Start is pAH12 and pAH15, RhaS Only is pAH12, and Neg. Control is pAH15. All have 0.2g/l of L-rhamnose added. In experiment 2 (Exp. 2) the induced system used are identical to sRNAII - SD in Exp. 1. These are compared to pAH12 (RhaS Only).

6.3 FITTING DYNAMICS TO THE AUTOREPRESSOR

In this section, data collected from the Autorepressor is fitted to an ODE model. This is undertaken to explain the process and discuss the assumptions that it is based on. The modelled output in the case of the Autorepressor is the total concentration of repressor protein in the system. In an experimental context this is equivalent to normalised fluorescence intensity data, which the model is fitted to. This data is presented in Figure 6.1a and Figure 6.1b for the Autorepressor. From this data, the most likely candidate for a good fit must be selected.

The model used to fit the data is a nondimensionalised form of the single state ODE model of the Autorepressor presented in Equation 5.3. For more information on this model, see Section 5.1. The nondimensionalised form of the model is

$$\frac{d\bar{y}}{d\tau} = \frac{1}{1 + C\bar{y}} - \bar{y}, \tag{6.1}$$

where $\bar{y} = \psi y$, is the nondimensionalised total protein concentration, $\tau = \delta t$ is the nondimensionalised time, $C = \frac{\psi K_I}{K_G(K_I+u)}$ is a constant (assuming a constant level of inducer), $\psi = \frac{\alpha g_{tot}}{\delta}$ is a constant, α is the maximal expression rate of the protein, g_{tot} is the total concentration of the gene encoding the protein, δ is the degradation rate of the repressor protein, K_I is the dissociation constant for the interaction between the repressor and inducer, K_G is the dissociation constant for the interaction between the repressor and gene, and finally u is the concentration of inducer, assumed to be constant.

Experimental data used to fit models would ideally closely resemble the simulated responses. In this case, these are presented in Figure 5.19. The simulated time courses are based on a number of assumptions discussed in general in Chapter 5, but one in particular is important here. This is the assumption that protein is not actively degraded, but that the only mechanism removing it from the system is dilution. This is supported experimentally, with proteins shown to by stable with half-lives of up to 20 hours [140], much longer than the doubling time of the cell. From this it can be concluded that when the cells are growing at any reasonable rate, dilution based on cell growth is the sole contributing factor to the removal of stable protein from these systems.

As can be seen in Figure 6.1c and Figure 6.1d, the growth rates vary throughout a plate reader experiment. There is a short period where the growth rate is more or less constant. This is during exponential growth phase. Therefore the section of the time courses studied is that during exponential phase. Looking at the data in Figure 6.1a and Figure 6.1b, the response that looks most like the modelled transient response for the Autorepressor in Figure 5.19 is possibly

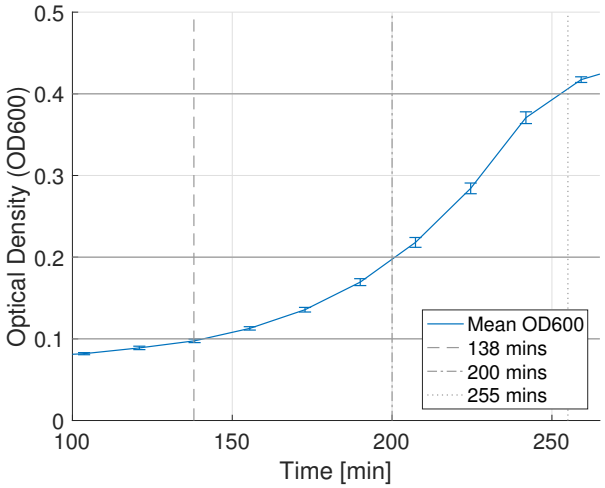


Figure 6.6 – The exponential section of the growth curve for time course data of the Autorepressor. The data shown is that corresponding to the 50ng/ml of aTc trace seen in Figure 6.1b. The guides mark the time at certain levels of cell concentration. Taking the average of the difference between these times yields an estimate for the doubling time of approximately 58.5min.

the response for the autorepressor with 50ng/ml of aTc in Figure 6.1b. This trace is selected for fitting.

To fit the model to the data, the doubling time of the cells during exponential growth phase must be found. This is done in Figure 6.6, and the doubling time is estimated to be 58.5min. This fixes the degradation rate: $\delta = \frac{\ln(2)}{58.5\text{min}} = 0.0118\text{min}^{-1}$, where $\ln(-)$ is the natural logarithm.

The first step in fitting the data is to rescale the time by the factor δ as calculated above. To make the process simpler, the steady state value of the output is estimated and the data is normalised with respect to this value. In this case, the value chosen is the normalised fluorescence intensity measurement at the last data point. The data is then ready for fitting.

When fitting the ODE model, it is assumed that the system starts from a zero initial condition, and the model data is divided by the steady state value so that the steady state is one. The model data is then shifted to the left such that is starts at about 0.4, like the experimental data.

Three possible fits are presented in Figure 6.7, demonstrating the various behaviour of the nondimensionalised model presented in Equation 6.1. As can be seen, low values of C result in a slower rise time, whereas high values of C rise faster. The value $C = 0.0810$ is calculated using the parameters found in Table 5.2, excepting δ which is fixed using the doubling time calculated in Figure 6.6 and remembering to recalculate the concentration of inducer: $u = 50\text{ng/ml} = 118\text{nM}$. The remaining parameters used are the dissociation constants: $K_1 = 10^{-2}\text{nM}$ and $K_2 = 10\text{nM}$.

Technically, α is a free parameter that can be fitted to the data as this is the only parameter that is not known or estimated otherwise. The other two values $C = 1$ and $C = 10^4$ are therefore

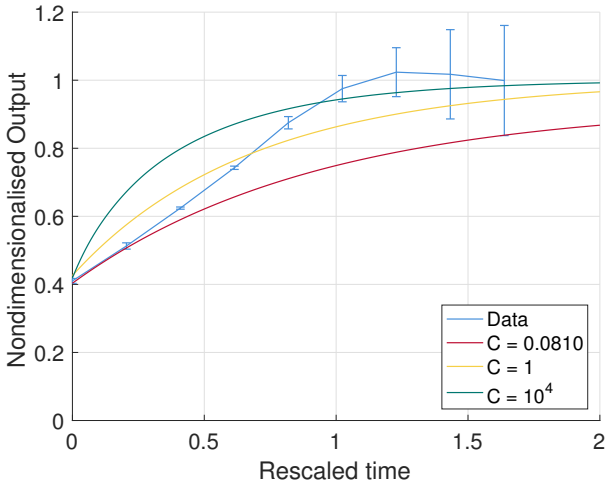


Figure 6.7 – Autorepressor data corresponding to the 50ng/ml of aTc trace seen in Figure 6.1b fitted to the nondimensionalised model in Equation 6.1 for three values of C.

included and correspond to $\alpha = 49.4\text{min}^{-1}$ and $\alpha = 4.94 \times 10^5\text{min}^{-1}$. The $C = 1$ fit is the best of the three, and the corresponding steady state protein concentration is approximately 7.3×10^4nM, this is well above what would be toxic to the cell. For more information on the choosing of parameters for the Autorepressor, see Section 5.1.4.

In summary, the process of fitting the nondimensionalised model to the data has given a possible estimate for the maximal expression rate of $\alpha = 49.4\text{min}^{-1}$, having fixed the degradation rate constant using the growth data from the experiment and assuming that the estimates of the dissociation constants are reasonable.

The fit models presented in Figure 6.7 do not seem to capture the intricacies of the experimental behaviour entirely. Measured expression in Figure 6.1a and Figure 6.1b in exponential phase seems at least linear and, in particular in Figure 6.1a, the expression rate increases in the exponential phase. The current model does not allow for this to occur. The traces also suddenly level off. Generally, this occurs shortly before or when exponential growth comes to an end, but in the case of the 25ng/ml traces in both Figure 6.1a and Figure 6.1b, this occurs in mid-exponential phase. In both these cases, the model predicts that the traces should level out to some steady state, but they actually begin to drop.

6.3.1 *Extending the Autorepressor Model*

One possible explanation for this is that the inducer, aTc, is actually a consumable resource in the cell. The binding between aTc and TetR is so tight that both species in effect begin to accumulate.

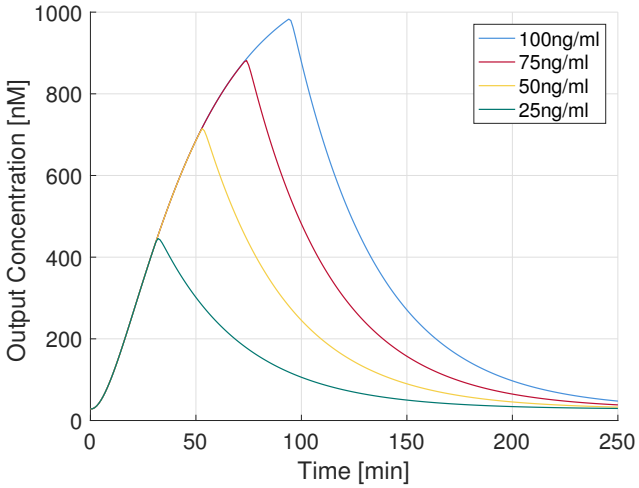


Figure 6.8 – Simulation results for the model of the Autorepressor with aTc consumption given by Equation 6.2. The values of the initial inducer concentrations are given in the legend. These are arbitrary to some extent; they just indicate relative quantity. The parameters used are those in Table 6.1.

This process continues until the aTc is exhausted, and then repression reasserts itself. It could even be argued that such behaviour is essential in the case of an antibiotic resistance mechanism, such as the natural context of TetR. This would ensure the survival of the bacteria by continuously expressing the necessary genes until all the antibiotic is sequestered or removed. Indeed, the Autorepressor motif is present in the natural tetracycline resistance genes [97].

Such a mechanism can be modelled using the chemical reactions in Equation 5.1, but adding an extra state to model the inducer concentration as a finite consumable quantity. The four state ODE model for this is:

$$\begin{aligned}
\dot{p} &= \alpha g - \delta p - k_1 p u + k_{-1} p_u - k_2 p g + k_{-2}(g_{tot} - g),\\
\dot{p}_u &= k_1 p u - (k_{-1} + \delta) p_u,\\
\dot{g} &= -k_2 p g + (k_{-2} + \delta)(g_{tot} - g),\\
\dot{u} &= -k_1 p u + k_{-1} p_u,
\end{aligned} \tag{6.2}$$

where p, p_u, g and u are the concentrations of repressor protein, repressor-inducer complex, free gene and inducer respectively. The parameter α is the expression rate, δ is the degradation rate, k_1 and k_{-1} are the forward and reverse rate constants for the interaction between the repressor and inducer, k_2 and k_{-2} are the forward and reverse rate constants for the interaction between the repressor and gene, and g_{tot} is the total gene concentration.

A simulation of this model is presented in Figure 6.8. These time courses are qualitatively closer to the data presented in Figure 6.1, than the expected dose responses presented in Fig-

Table 6.1 – The parameters used to simulate the Autorepressor with aTc consumption modelled in Equation 6.2. The values of k_1 is from [201] and the value of k_{-2} is from [97].

Param.	Value	Param.	Value
α	4min^{-1}	k_1	$1.98 \times 10^{-2}\text{nM}^{-1}\text{min}^{-1}$
δ	0.025min^{-1}	k_{-1}	$1.98 \times 10^{-4}\text{min}^{-1}$
k_2	$5.8 \times 10^{-3}\text{nM}^{-1}\text{min}^{-1}$	g_{tot}	28.2nM
k_{-2}	$5.8 \times 10^{-2}\text{min}^{-1}$		

ure 5.19, although results will not hold outside the exponential growth phase, as the dilution rate in this regime cannot be assumed to be constant.

This mechanism alone, though an improvement, does not entirely describe the dynamics observed. The changing growth rates in the experimental system are coupled to varying degradation rates in the model. The growth rates, and hence degradation rates, are at their highest and nearly constant in the exponential phase. This varying degradation rate can be included in a model by analysis of the growth curve, as its derivative with respect to time yields the rate of degradation at every point in time. In the case of the data in Figure 6.1, a variable degradation rate would be low in stationary phase, and therefore reduce the speed at which the sequestered aTc-TetR complexes are diluted out of the system.

The combination of these two mechanisms in a model, the consumption of inducer and the varying degradation rate, may well provide an improved transient fit to the experimental data. But the focus here is on qualitative comparison, rather than quantitative, as qualitative comparison remains challenging enough.

In summary, fitting models using the methods described above are, as of yet, not accurate enough to be able to meaningfully fit and distinguish between the transient behaviours of a number of different circuits, such as those modelled in Figure 5.19. This is required before quantitative design of these systems can occur. As of yet, the mechanisms modelled require revising and validation in comparison to experimental data to ensure that they are correct. The suggestion above regarding the Autorepressor in effect consuming the inducer is reasonable in the context of this specific system and the experiment performed. Other circuits using other mechanisms and experimental protocols will require other tailored solutions depending on their parts.

Another elements worth considering are higher level regulatory effects, such as the allocation of ribosomes [84], the burden on the host cell and how transcription and translation rates vary with growth rates.

6.4 CELL POPULATION DATA

In this section, cell population data is presented and used as an alternative to culture level data to facilitate identification and validation of models. Cell population data is powerful as it gives a description of behaviour across a population, and is particularly suited to comparison to the modelling results found in Chapter 5. The is done using the first and second moments of the data to calculate the mean, standard deviation and from there the coefficient of variation.

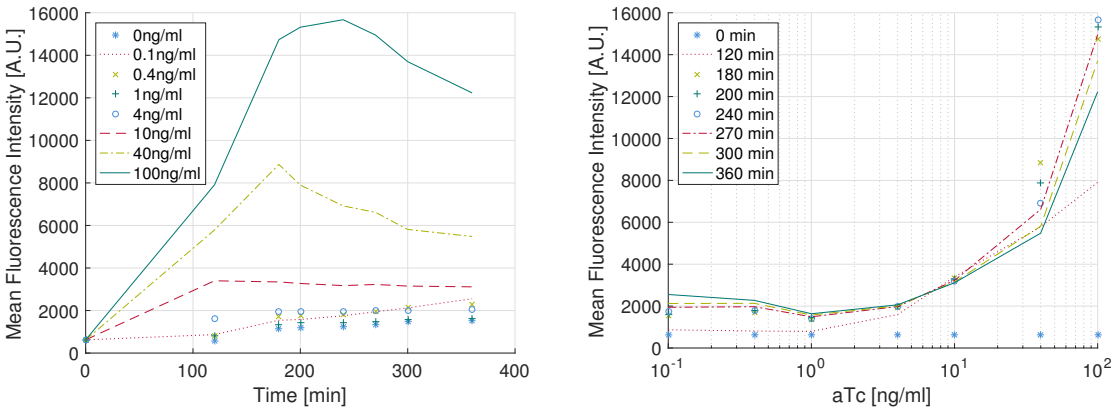
6.4.1 *Autorepressor Data*

The first data presented is for the Autorepressor. This data is collected using pAH01 in MG1655 using EZ media.

The initial data set taken for the Autorepressor is presented in Figure 6.9. The dose response here hints at the fact that induction is only just beginning at 100ng/ml of aTc. For simple repressed expression such as is modelled in Section 5.1.4, maximum induction of the *tet*-repressor system is known to occur around or just below 100ng/ml [132]. This is also predicted by the one state model of the Autorepressor using Hill Functions (detailed in Equation 5.3) when using parameters from the literature. The three state full model, given by Equation 5.7, on the other hand predicts that induction of the Autorepressor should occur at higher concentrations of inducer (see Figure 5.4). This prediction seems to agree with the data presented. To test this, higher concentrations of aTc must be used.

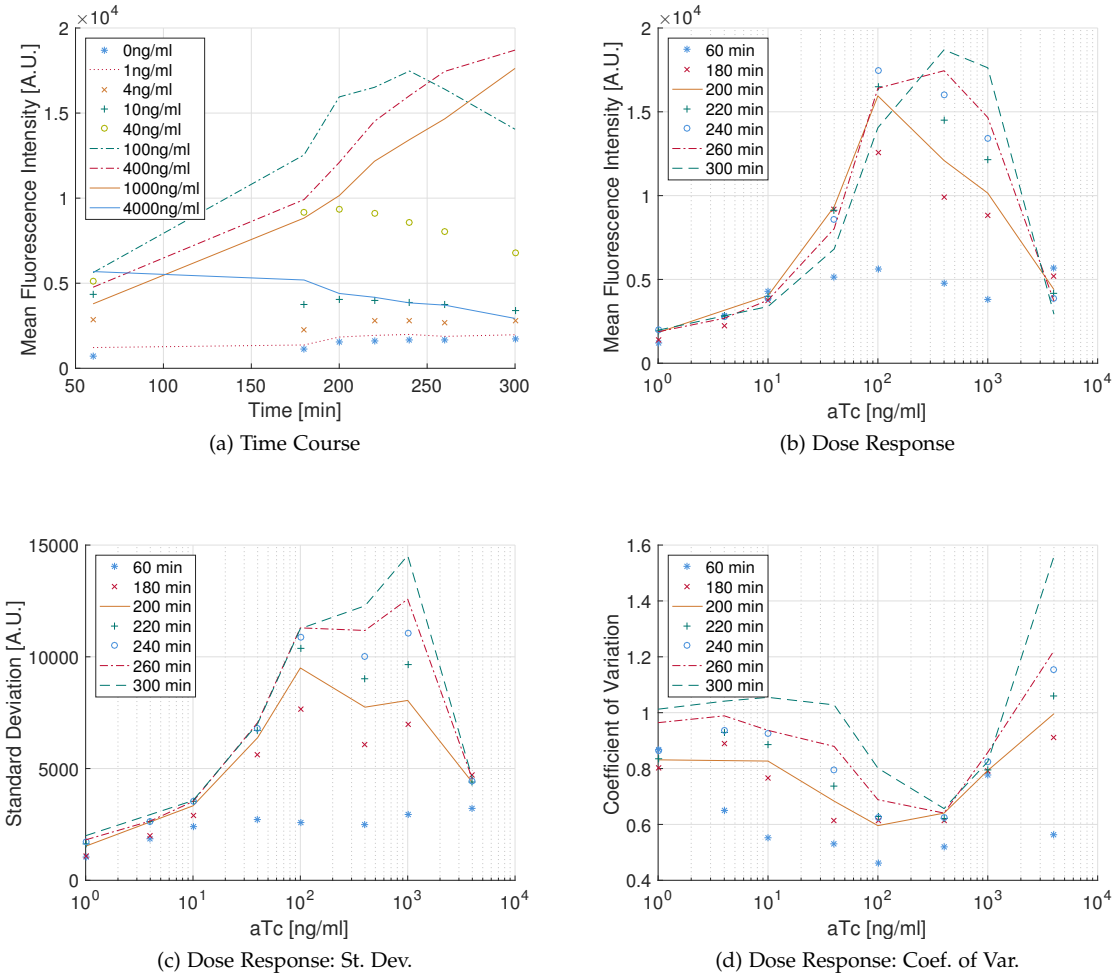
Incidentally, the time course data presented for this experiment, seen in Figure 6.9a, bears resemblance to the plate reader data for the Autorepressor in Figure 6.1 and therefore also the simulated time courses of the model with consumption of aTc included, as presented in Figure 6.8.

The Autorepressor was then tested with higher levels of aTc. This data is presented in Figure 6.10. Looking first at the time course in Figure 6.10a, the trace corresponding to an aTc concentration of 100ng/ml rises to a peak and begins to fall, as it does in Figure 6.9a. The traces for the higher aTc concentrations of 400ng/ml and 1000ng/ml are still growing. Their rise is slower than that of the 100ng/ml trace likely due to the extra burden of the additional aTc with the trace corresponding to 400ng/ml rising faster than the trace for 1000ng/ml. The trace for 4000ng/ml, on the other hand, does not increase. The growth data, which is not shown here, reveals that there is hardly any growth at all in the 4000ng/ml samples. At this level of aTc, it has begun to be toxic to the cells.



(a) Time Course (b) Dose Response

Figure 6.9 – The first cell population data for the Autorepressor collected. MG1655 cells are transformed with pAH01. The time courses for a number of different inducer concentrations are presented in (a) and the dose responses at a various times are in (b).



(a) Time Course (b) Dose Response

(c) Dose Response: St. Dev. (d) Dose Response: Coef. of Var.

Figure 6.10 – Cell population data for the Autorepressor with MG1655 cells transformed with pAH01. (a) Time course of the mean fluorescence intensity per cell. (b) Dose response of the mean fluorescence intensity per cell. (c) Standard deviation of dose response. (d) Coefficient of variation of dose response.

The above results are mirrored by the dose response data in Figure 6.10b. The mean fluorescence intensity begins to climb until the aTc level becomes toxic and expression halts. Taking into account both the dose response and time course plots, it seems that the induction of the autorepressor might not saturate before the level of aTc becomes toxic to the cells.

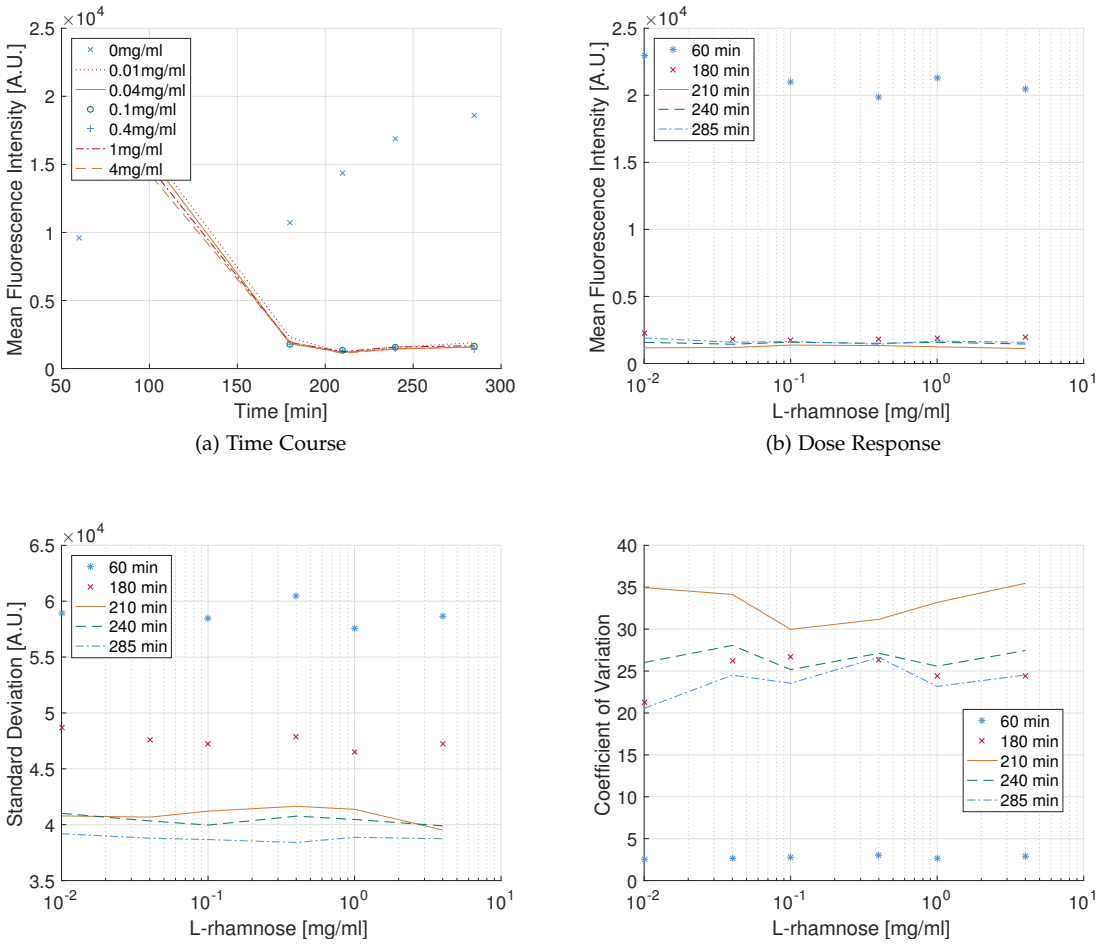
The coefficient of variation of the dose response in Figure 6.10d shows the dip predicted when plasmid copy number fluctuation is included in the model (see Figure 5.5). This supports the fact that plasmid copy number variation plays an important role in these systems and should be included when modelling. Though the plots have a similar shape, the absolute value of the noise in the real data is higher than that predicted by the model. This extra noise is most likely a result of mechanisms not included in the stochastic model in Section 5.1. The model also does not predict the steep increase in the coefficient of variation at high aTc concentrations. A reason for this may be that the cells react to the increased stress in a number of different ways, producing heterogeneous expression across the population.

6.4.2 *sRNA Circuit II Data*

The data collected for sRNA Circuit II uses plasmids pAH12 and pAH16. The first plasmid, pAH12, includes the gene encoding the RhaS-sfGFP fusion protein with expression controlled by P_{proD}. The other plasmid, pAH16, has the sRNA, targeting a region beginning with the Shine-Dalgarno sequence, for which expression is controlled by P_{rhaBAD}. This means that the system should be compared to the two plasmid modelling results in Section 5.3. The experiment is again performed using EZ media and the host bacteria is the Δ*rhaS* strain JW3876.

The results are presented in Figure 6.11. Studying the time course data in Figure 6.11a, the level of fluorescence per cell drops sharply between 60min and 180min for all induced systems. The time courses of the induced circuit are well below the un-induced circuit, indicating that the sRNA inhibition mechanism is functioning, as this agrees with the modelling in Section 5.3. The fast response of the system might support the fast response time observed in the dynamic simulation in Figure 5.19. This data agrees with the plate reader data seen in Figure 6.5b, where the induced samples also cluster tightly. This further supports the theory that the activation mechanism is saturated at an inducer concentration of 0.01mg/ml of L-rhamnose.

As expected, the dose response, presented in Figure 6.11b, shows that the system is likely saturated for all used concentrations of inducer. What is interesting are the standard deviation and coefficient of variation of the circuit. The coefficient of variation is large for all concentrations



(a) Time Course

(b) Dose Response

(c) Dose Response: St. Dev.

(d) Dose Response: Coef. of Var.

Figure 6.11 – Cell population data collected for sRNA Circuit II using the two plasmids pAH12 and pAH16 transformed into JW3876 in EZ. (a) Time course of the mean fluorescence intensity per cell. (b) Dose response of the mean fluorescence intensity per cell. (c) Standard deviation of dose response. (d) Coefficient of variation of dose response.

of inducer. This is due the addition of inducer causing the population distribution of fluorescence to have, not two, but three modes: a completely off mode, a completely on mode, and one somewhere in the middle. This is not the case for the un-induced system, where only one mode is present.

Though the models predict the observed reduction in output fluorescence when induced, they do not predict the trimodal fluorescence distribution. The models used in Section 5.3 only predict single modes, though in some cases the distributions are quite flat. The reason for this qualitative difference in behaviour is not clear, though ideally the model should be able to capture such behaviour. The experimentally observed behaviour might be due to intrinsic stochasticity of the system, or a mechanism that is not modelled.

To delve into this problem, a good starting point is the detailed testing of the sRNA regulation mechanism, studying the population distribution using flow cytometry, along the lines suggested in Chapter 3. This would shed light on the noise behaviour of sRNA regulation and lead to further development of the models. To do this, a range of sRNA regulation circuits such as those found in [102] and [129] could be constructed and then experimented on.

sRNAs are used in stress response mechanisms in natural systems. In this context, a heterogeneous response of a cellular population is beneficial to the population as a whole, as it increases the probability of the survival of the cells. Mechanisms driving this heterogeneity may well be part of the sRNA regulation mechanism not included here.

6.4.3 *Repressor-Activator Circuit Data*

Finally, data for the Repressor-Activator Circuit is presented. The plasmid used here is pAH30, which is transformed into MG1655. All samples include 0.4mg/ml of L-rhamnose (a high level of induction) unless otherwise stated, and are all in grown in EZ. This is a single plasmid system, which is worth keeping in mind when comparing the data to simulation results in Section 5.4. The measured quantity is the fluorescence of the TetR-sfGFP fusion protein; this is equivalent to Y_1 in the models.

The experimental data for the Repressor-Activator Circuit is presented in Figure 6.12. As the Autorepressor showed that higher levels of inducer than expected are required for induction, these higher aTc concentrations are also used when testing this circuit, though the highest concentration of 4000ng/ml is not included as it halted growth.

The time course data, presented in Figure 6.12a, shows relatively consistent dynamics, though at different levels for the various levels of induction. The time courses begin with a dip, and then rise at later times. The highest level of induction, 1000ng/ml, shows continually increasing fluorescence intensity right up until the end of the experiment, indicating that the system has indeed not saturated at 100ng/ml of aTc. This does not quantitatively agree with modelling, as the higher level of inducer required to saturate the system is not predicted (see Figure 5.15). The parameters used to model the RhaS activation mechanism were chosen using the *tet*-repressor parameters as a guide and fitting to the known dose response, as there are no parameters available in the literature. As this is the case, the focus is on the qualitative behaviour, for which in the case of the dose response, the model and data agree.

The time courses for the system with 0ng/ml and 1ng/ml of aTc (with 0.4mg/ml of L-rhamnose) behave as if un-induced, indicating that no induction has occurred at these low concentrations of inducer.

The dose response indicates that the induction of the system is only just beginning for the high levels of aTc used, though this is only visible at late times in the experiment. This is interesting as it might support the slow response time predicted by the modelling and presented in Figure 5.19.

Modelling of the system predicts a drop in the coefficient of variation as induction of the system occurs. This effect might be seen in Figure 6.12d, though the rise in coefficient of variation prior to this is not explained by the model.

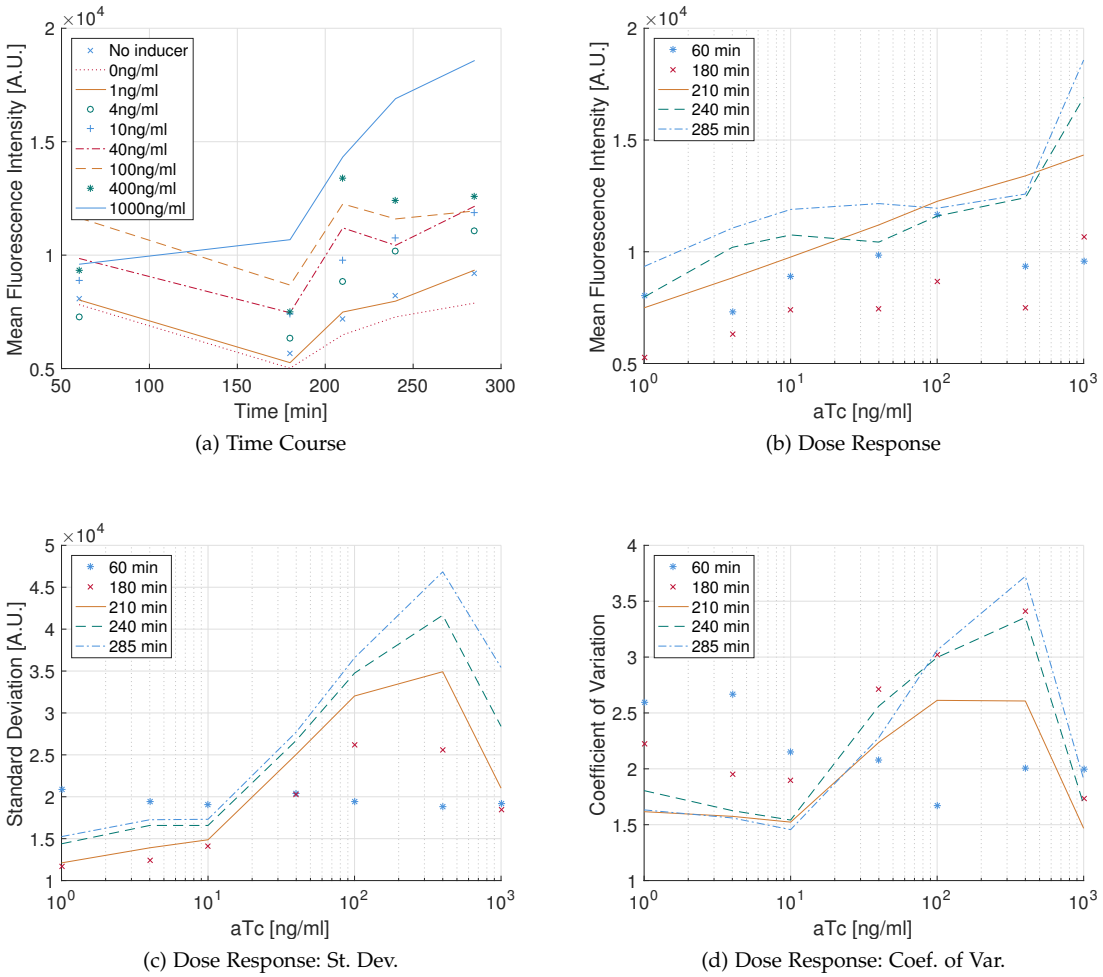
An interesting next step would be to fuse the RhaS, expressed by this circuit, to another reporter such that data on both the expressed proteins can be collected. This would give a better understanding of how the two TFs interact in this circuit.

6.5 CONCLUSION

This chapter presents experimental data collected from implementations of the circuits modelled in Chapter 5. Initially culture level data is presented for the Autorepressor and sRNA Circuits I and II, the majority of which is collected in a plate reader. The data shows that the circuits are functional.

In the case of the Autorepressor, the inconsistency of plate reader data is demonstrated, by comparing two identical experiments.

The data for sRNA Circuit I clearly shows that the sRNA targeting a region beginning with the Shine-Dalgarno sequence, regulates more efficiently than an sRNA targeting a sequence beginning



(a) Time Course

(b) Dose Response

(c) Dose Response: St. Dev.

(d) Dose Response: Coef. of Var.

Figure 6.12 – Cell population data collected for the Repressor-Activator Circuit using the plasmid pAH30 transformed into MG1655 and EZ. (a) Time course of the mean fluorescence intensity per cell. Here, no inducer means no L-rhamnose or aTc. (b) Dose response of the mean fluorescence intensity per cell. (c) Standard deviation of dose response. (d) Coefficient of variation of dose response.

with the Start Codon. The more efficient sRNA is used in all later experiments. It is shown that the addition of the pSUtat backbone, in conjunction with its required antibiotic, seems to weaken the *tet*-repressor mechanism, resulting in much higher levels of expression when this plasmid is included in circuits. Inducible control of the sRNA expression is shown to work, but not very well.

Experiments with sRNA Circuit II show that the choice between sRNA target regions may not be obvious. The reason that the sRNA targeting the Shine-Dalgarno region is chosen over the other design is due to growth being faster in the case of the Shine-Dalgarno targeting sRNA. This is thought to be due to possible off-target effects of the sRNA with the target region beginning with the Start Codon. It is shown that this circuit functions, but interestingly enough, although adding inducer reduces the output as expected from modelling, increasing the inducer concentration does not provide a graded output. This might be because the RhaS activation mechanism is saturated at very low levels at L-rhamnose due to the circuit dynamics.

As the culture level data presents dynamics of the circuits, a single state ODE model of the Autorepressor is fitted to one of the time courses presented. The section of the time course, where the cells are in exponential growth phase, is extracted from the data. The model is nondimensionalised to facilitate the fitting process. Given the variability in the data collected, it is argued that the model does not capture the timescale or the dynamics of the circuit accurately enough to be able to distinguish between the responses of different circuits and be used to study circuit dynamics in detail. The focus is therefore placed on qualitative comparison and prediction.

Following this discussion of the modelled mechanisms, an alternative model for the Autorepressor is proposed, where the inducer aTc is in effect consumed. It is shown to have some level of qualitative similarity to the data presented. A further extension is suggested, where the time-varying degradation rate can be calculated from the experimental growth curve.

Cell population data is then turned to for better understanding of the heterogeneity within the population. This data is ideal for comparison with models in Chapter 5, which study the steady states of systems using ODEs, and the noise profiles using stochastic simulations. The models are shown to accurately predict the observed behaviour of the Autorepressor, in particular the coefficient of variation of the dose response. It is shown that the prediction given by a higher order ODE model, modelling the interactions between the repressor, inducer and gene explicitly, predicts the dose response with a higher accuracy than a single state model employing Hill Functions. The higher order model correctly predicts that the induction of the Autorepressor would occur at higher concentrations of inducer than the single state model.

Table 6.2 – An overview of the plasmids used in the experiments presented in Chapter 6. The names of the circuits have been shortened: Autorepressor (Auto), sRNA Circuit I (sRNAI), sRNA Circuit II (sRNAII), Repressor-Activator (RepAct). The antibiotic resistance indicates the backbone used: ampicillin (AMP) and carbenicillin (CARB) indicates the pJ404 backbone [113], and chloramphenicol (CML) indicates the pSUtat backbone [112]. Synthetic Biology Open Language (SBOL) diagrams of these plasmids are also presented [74]. An SBOL diagram of pAH01 is presented in Figure 6.13. SBOL diagrams of the plasmids needed to complete sRNA Circuit I are presented Figure 6.14 and Figure 6.15. The SBOL diagrams of the plasmids require for sRNA Circuit II are presented in Figure 6.16. The SBOL diagram of the plasmid containing the Repressor-Activator Circuit is presented in Figure 6.17 and finally the SBOL diagrams of the control plasmids are presented in Figure 6.18.

Label	Contents	Circuit	Resistance
pAH01	TetR-sfGFP fusion gene with P_{tet} in front	Auto/sRNAI	AMP/CARB
pAH06	TetR SD sRNA with P_{proD} and P_r in front	sRNAI	CML
pAH07	TetR SD sRNA with P_r in front	sRNAI	CML
pAH08	TetR SD sRNA with P_{proD} in front	sRNAI	CML
pAH09	TetR Start sRNA with P_{proD} and P_r in front	sRNAI	CML
pAH12	RhaS-sfGFP fusion gene with P_{proD} in front	sRNAII	AMP/CARB
pAH15	RhaS Start sRNA with P_{rhaBAD} in front	sRNAII	CML
pAH16	RhaS SD sRNA with P_{rhaBAD} in front	sRNAII	CML
pAH17	TetR SD sRNA with P_{rhaBAD} in front	sRNAI	CML
pAH23	Empty pSUtat, only *ori* and resistance	Control	CML
pAH24	Empty pJ404, only *ori* and resistance	Control	AMP/CARB
pAH30	TetR-sfGFP fusion gene with P_{rhaBAD} in front, and RhaS activator gene with P_{tet} in front.	RepAct	AMP/CARB

Comparison with models in the cases of sRNA Circuit II and the Repressor-Activator Circuit reveal that the time scales of response might be correctly predicted with sRNA Circuit II showing a fast response and the Repressor-Activator showing a much slower response. It is suggested that the trimodal response observed in the case of sRNA Circuit II may be an inherent quality of the sRNA regulation mechanism. A method of exploring this further is presented.

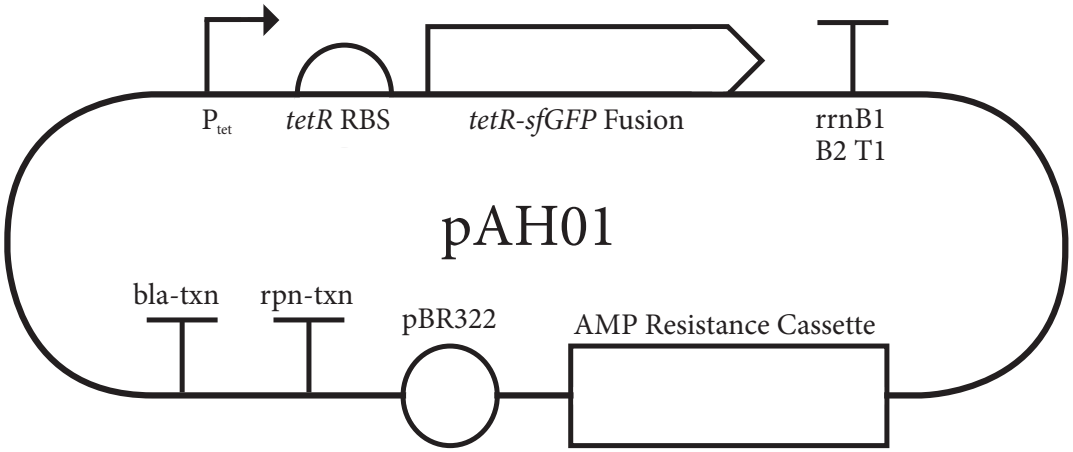
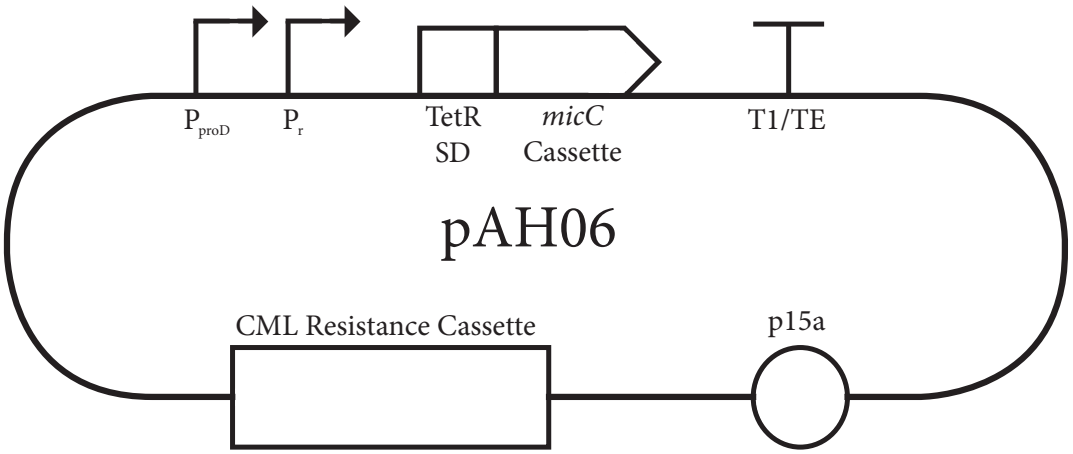


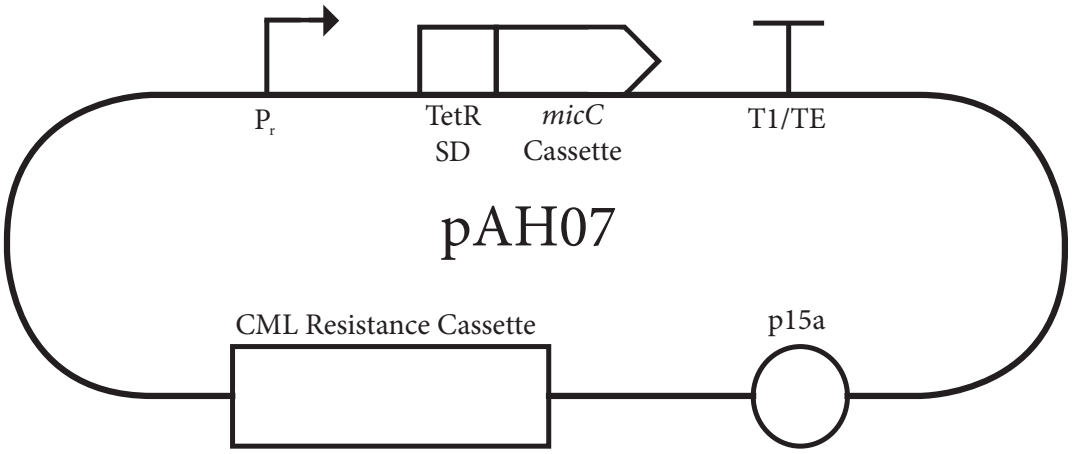
Figure 6.13 – The SBOL diagram of the autorepressor plasmid. This is the complete autorepressor circuit, but is also used in sRNA Circuit I.

Table 6.3 – The DNA sequences of the three gblocks ordered from IDT to create the plasmids listed in Table 6.2.

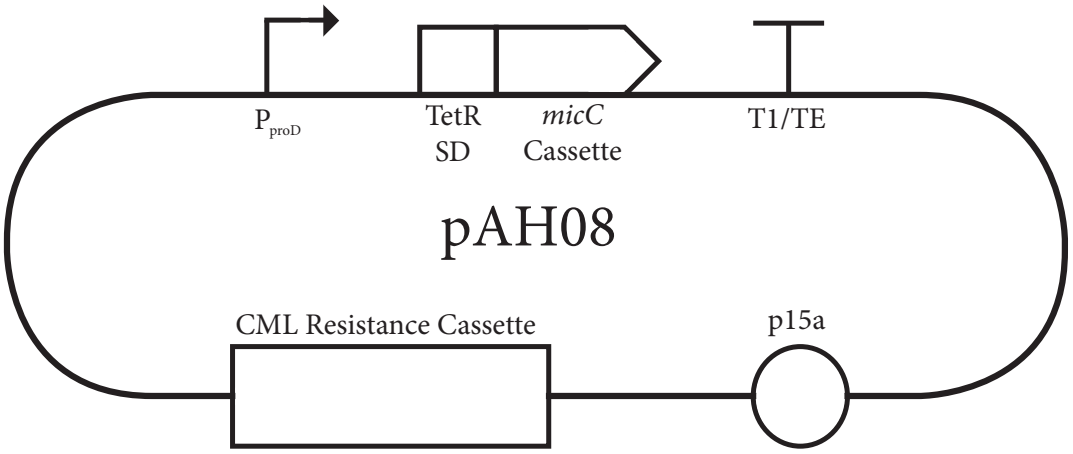
Description	DNA Sequence
Autorepressor cassette: P_{tet}, an optimised RBS to 55k [184], *tetR* and the first few codons of sfGFP. There is an overhang at the beginning to facilitate cloning into pJ404.	tttattatgcatttagaataaattttgtgtcgccctttccctatcagtgatagagattgacatccc tatcagtgatagagatactgagcacatcgataaacacagagaagtaggttcagaatgtcc agattagataaaagtaaagtgattaacagcgcattagagctgcttaatgaggtcggaatcg aaggtttaacaacccgtaaactcgcccagaagctaggtgtagagcagcctacattgtattg gcatgtaaaaaataagcgggctttgctcgacgccttagccattgagatgttagataggcac catactcacttttgccctttagaaggggaaagctggcaagatttttacgtaataacgctaaa agttttagatgtgctttactaagtcatcgcgatggagcaaaagtacatttaggtacacggcct acagaaaaacagtatgaaactctcgaaaatcaattagccttttatgccaacaaggttttca ctagagaatgcattatatgcactcagcgctgtggggcattttactttaggttgcgtattggaa gatcaagagcatcaagtcgctaaagaagaaagggaaacacctactactgatagtatgccg ccattattacgacaagctatcgaattatttgatcaccaaggtgcagagccagccttcttattcg gccttgaattgatcatatgcggattagaaaaacaacttaaatgtgaaagtgggtcccgtaa aggcgaagagctgttcactggtgtcgtccctattctaa
TetR Shine-Dalgarno Targeting sRNA cassette, with P_{proD} and P_r in front, the MicC cassette [44], a terminator and overhangs for cloning into pSUtat.	aaacgacggccagtgaattgcacagctaacaccacgtcgtccctatctgctgccctaggtct atgagtggttgctggataactttacgggcatgcataaggctcgtataatatattcagggaga ccacaacggtttccctctacaaataatttttgtttaacttttaacaccgtgcgtgttgactatttta cctctggcggtgataatggttgctaatctggacattctgaacctacttttctgttgggccattg cattgccactgattttccaacatataaaaaagacaagcccgaacagtcgtccgggctttttttct cgagccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatct gttgtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctg cgtttataattgcgttgcgctcactgccc
TetR Start Codon Targeting sRNA cassette, with P_{proD} and P_r in front, the MicC cassette [44], a terminator and overhangs for cloning into pSUtat.	aaacgacggccagtgaattgcacagctaacaccacgtcgtccctatctgctgccctaggtct atgagtggttgctggataactttacgggcatgcataaggctcgtataatatattcagggaga ccacaacggtttccctctacaaataatttttgtttaacttttaacaccgtgcgtgttgactatttta cctctggcggtgataatggttgctttacttttatctaatctggacattttctgttgggccattgca ttgccactgattttccaacatataaaaaagacaagcccgaacagtcgtccgggctttttttctcg agccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgtt gtttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcg tttataattgcgttgcgctcactgccc



(a)

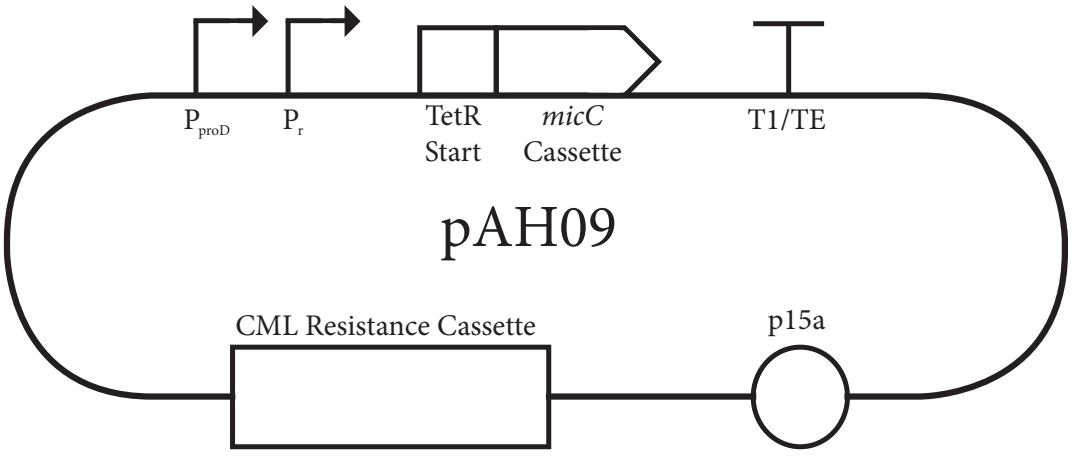


(b)

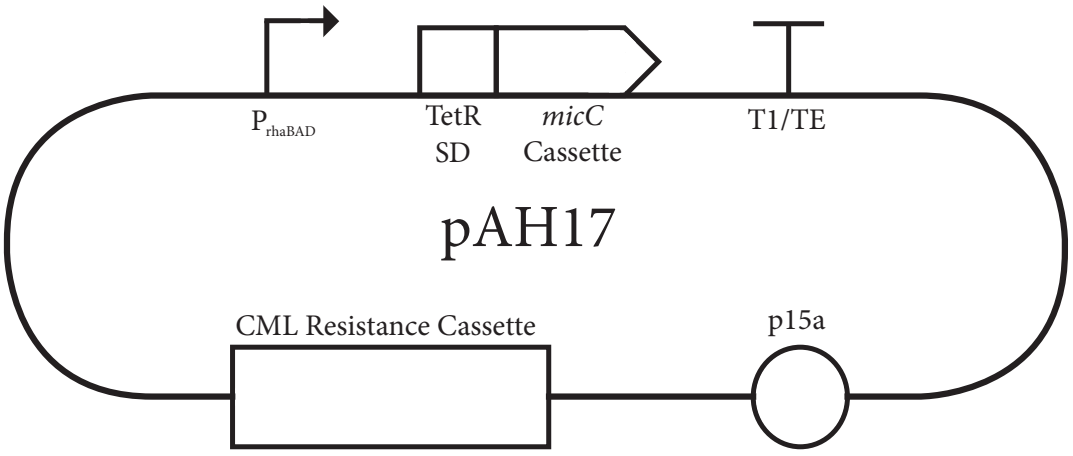


(c)

Figure 6.14 – The SBOL for various plasmids used to construct sRNA Circuit I. One of these is added to pAH01 to form the circuit. These plasmids have different rates of expression of the sRNA, but the sRNA is consistently targeting a region beginning with the Shine-Dalgarno. More plasmids used for sRNA Circuit I are presented in Figure 6.15.

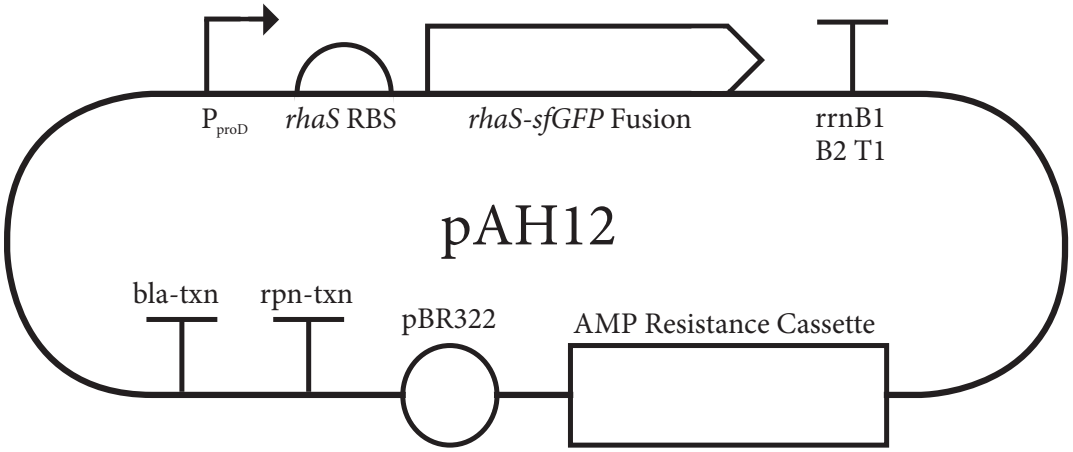


(a)

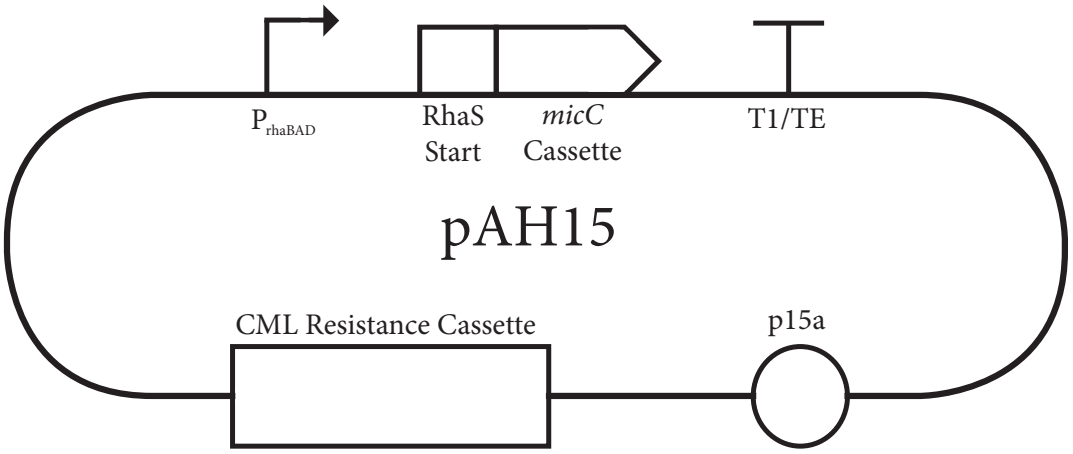


(b)

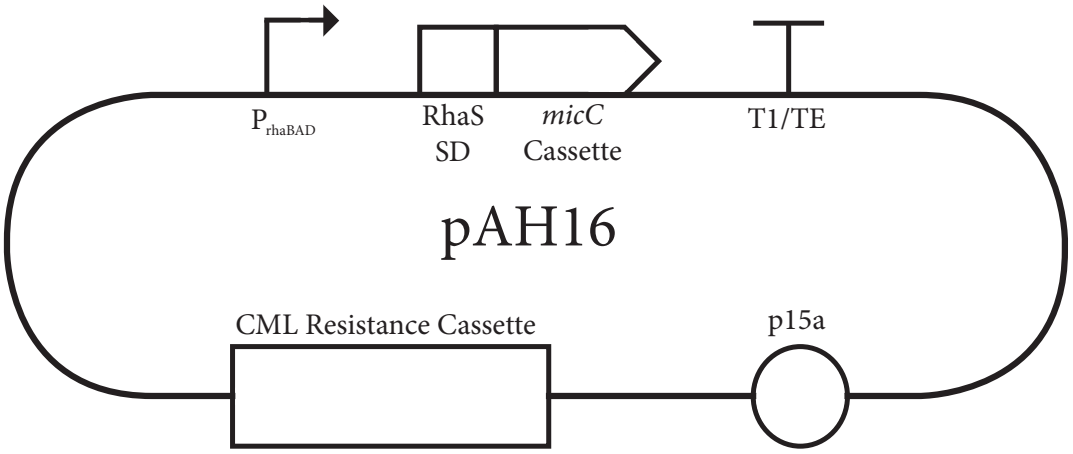
Figure 6.15 – The SBOL for various plasmids used to construct sRNA Circuit I. One of these is added to pAH01 to form the circuit. pAH09 includes an sRNA design, that targets a region beginning with the Start Codon of the target mRNA and is constitutively expressed. The sRNA target of pAH17 is a region beginning with the Shine-Dalgarno, and the promoter is inducible. More plasmids used for sRNA Circuit I are presented in Figure 6.14.



(a)



(b)



(c)

Figure 6.16 – SBOL diagrams of the plasmids required to construct sRNA Circuit II. One of either pAH15 or pAH16 is combined with pAH12 to complete the circuit. pAH15 and pAH16 vary in their target regions: pAH15 targets a region beginning with the Start Codon, pAH16 targets a region beginning with the Shine-Dalgarno, otherwise they are identical.

Table 6.4 – The DNA sequences of the three genes encoding proteins used in the plasmids listed in Table 6.2.

Part	DNA Sequence
tetR [167]	atgtccagattagataaaagtaaagtgattaacagcgcattagagctgcttaatgaggtcggaatcg aaggtttaacaacccgtaaactcgcccagaagctaggtgtagagcagcctacattgtattggcatgta aaaaataagcgggctttgctcgacgccttagccattgagatgttagataggcaccatactcacttttgc cctttagaaggggaaagctggcaagatttttacgtaataacgctaaaagttttagatgtgctttacta agtcatcgcgatggagcaaaagtacatttaggtacacggcctacagaaaaacagtatgaaactctcg aaaatcaattagccttttttatgccaacaaggtttttcactagagaatgcattatatgcactcagcgctgt ggggcattttactttaggttgcgtattggaagatcaagagcatcaagtcgctaaagaagaaaggga aacacctactactgatagtatgccgccattattacgacaagctatcgaattatttgatcaccaaggtgc agagccagccttcttattcggccttgaattgatcatatgcggattagaaaaacaacttaaatgtgaaa gtgggtcctag
sfGFP with HIS tag [113]	cgtaaaggcgaagagctgttcactggtgtcgtccctattctggtggaactggatggtgatgtcaacg gtcataagttttccgtgcgtggcgagggtgaaggtgacgcaactaatggtaaactgacgctgaagtt catctgtactactggtaaactgccggtaccttggccgactctggtaacgacgctgacttatggtgttca gtgctttgctcgttatccggaccatatgaagcagcatgacttcttcaagtccgccatgccggaaggcta tgtgcaggaacgcacgatttcctttaaggatgacggcacgtacaaaacgcgtgcggaagtgaaattt gaaggcgataccctggtaaaccgcattgagctgaaaggcattgactttaaagaagacggcaatatc ctgggccataagctggaatacaattttaacagccacaatgtttacatcaccgccgataaacaaaaaaa tggcattaaagcgaattttaaaattcgccacaacgtggaggatggcagcgtgcagctggctgatcac taccagcaaaacactccaatcggtgatggtcctgttctgctgccagacaatcactatctgagcacgca aagcgttctgtctaaagatccgaacgagaaacgcgatcatatggttctgctggagttcgtaaccgca gcgggcatcacgcatggtatggatgaactgtacaaaggctccggctccggctcccaccatcaccatc accattga
rhaS [113]	atgaccgtattacatagtgtggatttttttccgtctggtaacgcgtccgtggcgatagaaccccggctc ccgcaggcggattttcctgaacatcatcatgattttcatgaaattgtgattgtcgaacatggcacgggt attcatgtgtttaatgggcagccctataccatcaccggtggcacggtctgtttcgtacgcgatcatgat cggcatctgtatgaacataccgataatctgtgtctgaccaatgtgctgtatcgctcgccggatcgatttc agtttctcgccgggctgaatcagttgctgccacaagagctggatgggcagtatccgtctcactggcg cgttaaccacagcgtattgcagcaggtgcgacagctggttgcacagatggaacagcaggaagggg aaaatgatttaccctcgaccgccagtcgcgagatcttgtttatgcaattactgctcttgctgcgtaaaag cagtttgcaggagaacctggaaaacagcgcatcacgtctcaacttgcttctggcctggctggaggac cattttgccgatgaggtgaattgggatgccgtggcggatcaattttctctttcactgcgtacgctacatc ggcagcttaagcagcaaacgggactgacgcctcagcgatacctgaaccgcctgcgactgatgaaa gcccgacatctgctacgccacagcgaggccagcgttactgacatcgcctatcgctgtggattcagcg acagtaaccacttttcgacgctttttcgccgagagtttaactggtcaccgcgtgatattcgccagggac gggatggctttctgcaataa

Table 6.5 – The DNA sequences of the two RBSs and four promoters used in the plasmids listed in Table 6.2. The *tetR* RBS was designed using the Salis 'RBS Calculator' [184].

Part	DNA Sequence
RBS for *tetR*	tcgataaacacagagaagtaggttcaga
RBS for *rhaS* [113]	tatttcgccgtgttgacgacatcaggaggccagt
P_{tet} (BBa_R0040)	tccctatcagtgatagagattgacatccctatcagtgatagagatactgagcaca
P_{rhaBAD} [113]	ccacaattcagcaaattgtgaacatcatcacgttcatctttccctggttgccaatggcccattttcctgtc agtaacgagaaggtcgcgtattcaggcgctttttagactggtcgtaatgaa
P_{proD} [52]	cacagctaacaccacgtcgtccctatctgctgccctaggtctatgagtggttgctggataactttacgg gcatgcataaggctcgtataatatattcagggagaccacaacggtttccctctacaaataattttgttta actttt
P_r [224]	taacaccgtgcgtgttgactattttacctctggcggtgataatggttgc

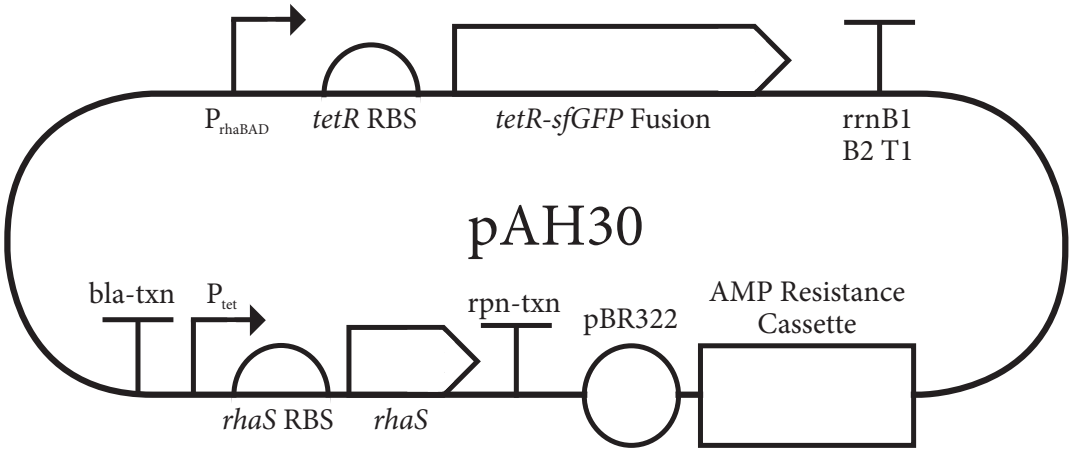
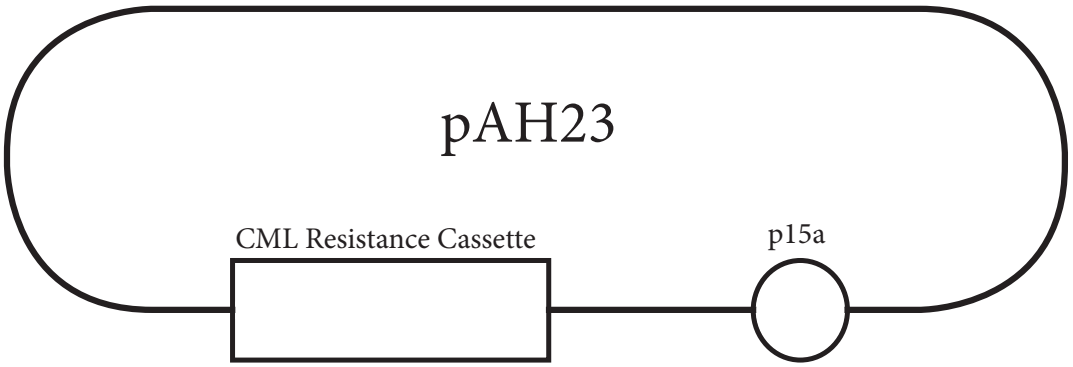


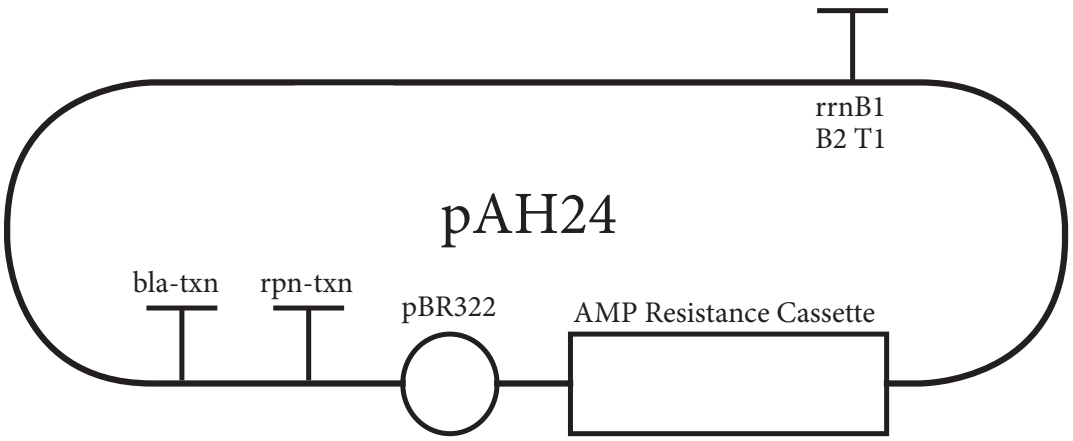
Figure 6.17 – The SBOL diagram of the plasmid with the Repressor-Activator Circuit on it. Note that the directionality of the various parts are presented here following the standard convention: above the top strand is forward, above the bottom strand is reverse.

Table 6.6 – The DNA sequences of the sRNA parts used in the plasmids listed in Table 6.2. The target regions were designed through complementarity with their respective targets.

Part	DNA Sequence
micC sRNA Cassette [44]	tttctgttgggccattgcattgccactgattttccaacatataaaaagacaagcccgaacagt cgtccgggctttttt
sRNA Target Region: TetR SD	taatctggacattctgaacctact
sRNA Target Region: RhaS SD	actggcctcctgatgtcgtcaaca
sRNA Target Region: RhaS Start	atccacactatgtaatacggtcat



(a)



(b)

Figure 6.18 – The SBOL diagrams of the control plasmids used in experiments.

Table 6.7 – The DNA sequences of the terminators, resistance operon and *ori*-region used in the pJ404 plasmids listed in Table 6.2. The source of all these parts are detailed in [113].

Part	DNA Sequence
pBR322 *ori*-region	agatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgctgcttgcaaacaaaaaa accaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggt aactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagttagccc accacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtgg ctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccgga taaggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcga acgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccc gaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcg cacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccac ctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacg ccagcaacgcg
AMP Resistance Cassette	ttcaaatatgtatccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaag gaagaatatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcc tgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgca cgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgccccg aagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtattatcccgtat tgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgag tactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtg ctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggacc gaaggagctaaccgctttttgcacaacatgggggatcatgtaactcgccttgatcgttggg aaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcg atggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaaca attaatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccg gctggctggtttattgctgataaatccggagccggtgagcgtggttctcgcggtatcatcgc agcgctggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtca ggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagca ttggtaa
rpn txn Terminator	cttaaatattctgacaaatgctctttccctaaactccccccataaaaaaacccgccgaagcgg gtttttacgttatttgcggattaacgattactcgttatcagaaccgcccag
bla txn Terminator	tattctaaatgcataataaatactgataacatcttatagtttgtattatattttgtattatcgttga catgtataattttgatatcaaaaactgattttccctttattattttcgagatttatttttcttaattctc tttaacaaactagaaatattgtatatacaaaaaatcataaataatagatgaatagtttaattat aggtgttcatcaatcgaaaaagcaacgtatcttatttaaagtgcgttgctttttttctcatttata aggttaaataattctcatatatcaagcaaagtgaca
rrnB1 B2 T1 Terminator	aattagcccgggcgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggca gttccctactctgcgcatggggagtccccacactaccatcggcgctacggcgtttcacttctga gttcggcatggggtcaggtgggaccaccgcgctactgccgccaggcaaa

Table 6.8 – The DNA sequences of the terminators, resistance operon and *ori*-region used in the pSUtat plasmids listed in Table 6.2. The source of all these parts are detailed in [112].

Part	DNA Sequence
p15a *ori*-region	ttgagatcgttttggtctgcgcgtaatctcttgctctgaaaacgaaaaaaccgccttgcagg gcggtttttcgaaggttctctgagctaccaactctttgaaccgaggtaactggcttggagga gcgcagtcaccaaaacttgtcctttcagtttagccttaaccggcgcatgacttcaagactaac tcctctaaatcaattaccagtggctgctgccagtggtgcttttgcatgtctttccgggttggac tcaagacgatagttaccggataaggcgcagcggtcggactgaacgggggttcgtgcat acagtccagcttggagcgaactgcctacccggaactgagtgtcaggcgtggaatgagac aaacgcggccataacagcggaatgacaccggtaaaccgaaaggcaggaacaggagag cgcacgagggagccgccaggggaaacgcctggtatctttatagtcctgtcgggtttcgcca ccactgatttgagcgtcagatttcgtgatgcttgtcaggggggcggagcctatggaaa
CML Resistance Cassette	tgatcggcacgtaagaggttccaactttcaccataatgaaataagatcactaccgggcgtat tttttgagttatcgagattttcaggagctaaggaagctaaaatggagaaaaaaatcactgg atataccaccgttgatatatcccaatggcatcgtaaagaacattttgaggcatttcagtcagtt gctcaatgtacctataaccagaccgttcagctggatattacggccttttaaagaccgtaaag aaaaataagcacaagttttatccggcctttattcacattcttgcccgcctgatgaatgctcatc cggaatttcgtatggcaatgaaagacggtgagctggtgatatgggatagtgttcacccttg ttacaccgttttccatgagcaaactgaaacgttttcatcgctctggagtgaataccacgacga tttccggcagtttctacacatatattcgcaagatgtggcgtgttacggtgaaaacctggccta tttccctaaagggtttattgagaatatgttttcgtctcagccaatccctgggtgagtttcacca gttttgatttaaacgtggccaatatggacaacttcttcgcccccgttttcaccatgggcaaata ttatacgcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgccgtctgtg atggcttccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtggcaggg cggggcgtaa
T1/TE Terminator	ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgt ttgtcggtgaacgctctctactagagtcacactggctcaccttcgggtgggcctttctgcgttt ata

7 CONCLUSION

The overall aim of this thesis was to improve and develop the engineering approach to the field of Synthetic Biology, with a specific focus on the engineering of Gene Regulatory Networks (GRNs).

To this end, Chapter 3 focussed on sRNAs as a tool for post-transcriptional regulation in a dynamical network of genes. A variety of models of both the mechanism of sRNA inhibition and sRNA activation were studied in depth. A basic model of the mechanism was compared to a more complex model including the Hfq protein, a chaperone protein that acts as a catalyst of the sRNA regulation mechanism. In addition to this, the models studied included mechanisms describing gene copy number fluctuation, as the genes included in the mechanism can either be encoded on the same plasmid [102], or two separate plasmids [129].

The chapter concluded that modelling the sRNA regulation mechanism using ODEs does not require the inclusion of the catalytic Hfq mechanism. This was shown through the novel extension of a model of sRNA regulation with Hfq from [1]. The extension consists of asymmetric terms, modelling the interaction between Hfq and the RNA species, that allow the mRNA species to degrade fast in all complexes. This ensures that the steady state behaviours of the mRNA and the heteroduplex, the species of interest, are identical when modelling with and without Hfq.

In the case of stochastic models, the inclusion of the Hfq mechanism is important, as it greatly changes the expected steady state and noise profiles of the system. This is particularly the case when the genes involved in the mechanism are encoded on separate plasmids. This greatly softens the threshold of the steady state and increases the expected noise. It is suggest that this difference in experimental set-up may explain the difference between the results in [102] and [129].

Chapter 4 adapted classical engineering design techniques to the field of Synthetic Biology. This chapter took linear and frequency domain design techniques and developed a consistent framework for the design of GRNs. The focus was initially on genetic networks consisting of TFs and the framework presents a method of circuit visualisation through block diagrams that facilitates the calculation of transfer functions.

With this framework at hand, the question of whether nature uses any controller designs known in the field of engineering, was approached. Genetic network motifs corresponding to the phase lead and lag controllers were presented as an example of such. These are known to

exist in nature and have elsewhere been referred to as incoherent and coherent feedforward respectively [9, 135]. The phase lag controller was then used in a design example with a presented implementable biological topology. Finally, the framework was extended to include sRNA regulation. A framework such as this is crucial to the scale-up of designed systems as it gives a method by which mathematical abstractions can be made that facilitate the design of feedback.

Next, specific GRNs were modelled. Chapter 5 presented the detailed modelling of four simple genetic circuits harnessing negative feedback. The first, the Autorepressor, is a common and well studied motif [16, 23, 62, 179] in which a gene negatively regulates its own transcription through a TF. The second, sRNA Circuit I, is based on the Autorepressor, but includes an sRNA inhibiting translation. In the third circuit, sRNA Circuit II, the negative feedback is supplied through an sRNA, which inhibits the translation of an activating TF that controls the sRNAs expression. The fourth circuit, the Repressor-Activator Circuit, also features negative feedback through another gene, but as the name indicates, all the regulation is performed by TFs.

Modelling of these circuits was performed based on the specific biological mechanisms used to implement the circuits. The modelling also took the results of Chapter 3 into account, in the cases of circuits involving sRNAs. The focus of the modelling was steady state behaviour and the related noise profiles through both ODEs and stochastic modelling approaches. For circuits consisting of two genes, the behaviour of single and double plasmid set-ups was explored. The double plasmid set-up was found to be incredibly noisy in the case of sRNA Circuits I and II, but not in the case of the Repressor-Activator Circuit. At the end of the chapter, the dynamic responses of each of the studied circuits were compared.

The penultimate Chapter 6 presented and explored experimental data collected on the circuits modelled in Chapter 5. The first section of the chapter focussed on culture level data, most of which was collected in a plate reader, which produces time courses of the fluorescence intensity and cell concentration.

Two data sets of identical Autorepressor experiments were presented to demonstrate the inconsistency of data collected in plate reader experiments. The circuit was found to behave as expected otherwise.

sRNA Circuit I was shown to function as expected. Data from two different implementations of this circuit was presented. The two implementations only differ in the target region of the sRNA: one targets a region beginning with the Shine-Dalgarno sequence, the other targets a region beginning with the Start Codon. It was shown that the sRNA targeting a region beginning with the Shine-Dalgarno regulates more efficiently, and therefore this implementation was used for all further experiments. It was also shown that the inclusion of the pSUtat plasmid and its associated

antibiotic, increase the fluorescence of the Autorepressor substantially, which may be due to interference with the TetR repression mechanism. A time course of different implementations of sRNA Circuit I were compared in an experiment where samples were grown in larger volumes and incubated in falcon tubes. This data was found to be more reliable than the plate reader data.

The third circuit for which culture level data was presented is sRNA Circuit II. As in the case of sRNA Circuit I, two different sRNA implementations were tested. Once again, the sRNA targeting the Shine-Dalgarno is the preferred choice, but not due to tighter regulation. This decision was made due to slow growth in the case of the sRNA, targeting a region beginning with the Start Codon. This shows that there is no optimal sRNA design, and that when implementing circuits including sRNA regulation, several designs must be tested. The circuit was shown to function, though the response seems to saturate at very low levels of inducer.

Plate reader data from the Autorepressor was then fitted to a one state ODE model which had been nondimensionalised. The trace used to fit the model was deliberately cherry-picked to be most similar to the expected response. Despite this, the model does not seem to properly capture the behaviour of the dynamics or the time scale.

A suggestion was made to include the inducer, in the case of the Autorepressor, as a finite resource. Simulating this yields behaviour which qualitatively resembles that seen in the plate reader data, but does not fully capture the time scales of the system. Additional changes to the model could be made by adding a variable degradation rate dependent on the rate of growth.

Further data was taken with a flow cytometer using cultures incubated in falcon tubes. In addition to time course data, this provided noise data which is required for comparison with noise profiles predicted by stochastic modelling.

Data from the Autorepressor was presented first. This data supports a three-state ODE model of the system, as opposed to the one-state model employing a Hill Function. This is because it predicts that the induction of the circuit should require a higher level of inducer than predicted by the simpler model. The system was tested with higher levels of inducer, and as expected fluorescence intensity continued to grow until the inducer concentration was so high that cell growth was almost completely halted. The behaviour of the coefficient of variation was correctly predicted by the stochastic model of the circuit.

Data for sRNA Circuit II was then presented; this agreed with the data taken in the plate reader showing that the circuit functions. The level of fluorescence was reduced when inducer was added, as predicted by the modelling. Again, the circuit seemed to be saturate at very low levels of inducer. In addition to this, the transient response of the circuit is faster than that of the

Autorepressor, which is in line with predictions based on ODE models. The noise predictions in the case of this circuit were inaccurate as it was found that the implemented circuit demonstrates three different population modes: on, off and in between.

Finally, the implementation of the Repressor-Activator Circuit was studied. Levels of inducer were used up to, but not including the concentration shown to halt growth in the case of the Autorepressor. The dynamics of this circuit are slower than the Autorepressor, which is also in agreement with predictions. The dose response presented shows the beginning of the predicted increase in output due to induction, and also the reduction in noise as induction begins.

7.1 DISCUSSION

In engineering, the use of models to predict and then design systems is crucial. This is a major goal of Synthetic Biology. As shown in Chapter 6, the qualitative comparison between models in Chapter 5 and experimental data is somewhat successful. The next step, required to be able to accurately predict and design larger systems is accurate quantitative design.

In this sense, modelling and experiments are coming together, but there is still a way to go yet. One way of improving this synergy is to adapt and improve models. Extensions, such as inducer consumption, presented in the context of the Autorepressor, give possible solutions to these problems. The inclusion of higher level regulatory mechanisms, such as ribosome allocation [84], the effect of burden and toxicity (seen in the case of high aTc) and variation of transcription, translation and degradation rates depending on the growth rate of the cells, the media they are in, their main carbon source and so on, are all areas that would provide mechanisms enriching and improving the models used.

Approaching this from a different angle; it is worth asking the question, is it possible for the experiments to fulfil the assumptions of the current models? The current models are of a relatively low mathematical complexity, which allows for easy manipulation and analysis. Retaining this simplicity would be useful.

Taking the example of this thesis, the majority of the modelling in Chapter 5 focusses on the steady state of the modelled circuits. These must be studied, understood and modelled accurately before the dynamics can be approached. From a modelling perspective, at steady state the left hand side of the ODE models is zero. In a stochastic context, though individual systems are constantly jumping between different states, the distribution across the population at steady state is near static.

What does the steady state of a model mean in an experimental context? All data collected for this thesis contains cells in a finite amount of media. They are taken from a pre-culture, where they are in stationary phase and suddenly placed into new media with an inducer. Their metabolisms rapidly change as they begin to grow and multiply. They reach exponential phase but are only there for a relatively short period of time, before they are starved of nutrients and move into stationary phase again. Not only this, but, as seems to be the case for the Autorepressor, the inducer is in effect consumed and therefore a finite resource. At later times, cells are therefore in effect exposed to less inducer than cells at earlier times. Such an environment is constantly changing. Can a genetic circuit within a cell be a steady state when its environment is constantly changing?

These variations in the environment can be modelled through time varying parameters. In particular, the degradation, transcription and translation rates are prone to vary depending on the growth rate of the cells, the quantity and type of nutrients available and so on. It is not currently known how these vary, and therefore they are generally assumed to be constant.

Can the constancy of environment, required to provide the near constant parameters necessary to observe the steady state of a genetic circuit, be achieved and maintained experimentally? A set-up that might just be capable of doing this was published in May 2016 [145]. In this paper, Milias-Argeitis *et al.* set-up a bioreactor with a tailor made chemostat. This did not only control the temperature, aeration and the pH, but also the phase of the cells. Using an *in silico* feedback controller, OD600 measurements of the cells in the bioreactor are kept constant by continuously diluting the cells in the reactor with new media, while removing excess to keep the volume constant. In this way, it can be ensured that the cells are kept in exponential growth phase. They are constantly fed new media so they can continue to grow. Not only this, but the new media contains the necessary antibiotics and can also contain inducers at a required concentration. Theoretically, this could continue for as long as necessary. With such a set-up, the assumptions used in the modelling suddenly look much more like the actual cellular environment.

The suggestion is to use such a set-up to keep cells, transformed with circuits, in exponential phase, while measurements are taken from the extracted waste until the fluorescence distribution across the population settles down. This can then be defined as the steady state of the dynamical genetic circuit within each cell and this creates the starting point for all further measurement.

The steady state behaviour of the circuits can at this point be studied for various levels of inducer or the like. Once the circuits have been studied in this context and the steady states are well understood, the next step can be taken. This is to introduce carefully controlled perturbations to the system and then observe the transient response of the circuit. The simplest example

of this would be to change the inducer concentration. This mechanism is already included in the models used in Chapter 5. Other perturbations such as temperature change, change of media or carbon source, and varying the antibiotic concentration are equally possible. How to model the effect of such changes can be derived from observing such experiments. These perturbations do not necessarily have to be step-changes; they could oscillate as well.

Using this set-up and applying controlled perturbations, the dynamic response can then be measured and compared to simulated responses such as those presented in Figure 5.19. Being able to perform such experiments and study the transient responses of GRNs in this way would suddenly open up the world of Synthetic Biology to the direct application of a wealth of engineering design tools, such as linear design techniques developed and applied to gene regulatory networks in Chapter 4.

Another benefit to this method of testing is that the circuits would be tested in a context very similar to that expected of a possible final application. An obvious example of this is the use of GRNs to optimise and control metabolic pathways, such as has been done in [226]. In such a context, the circuits would be expected to function in large bioreactors.

Another effective method of keeping cells in such a controlled environment is through microfluidics platforms, where measurements are taken through fluorescence microscopy [25]. This has the benefit that individual cells can be followed over time, but the numbers of cells studying in an experiment is severely limited. Once cells can be kept in exponential phase for prolonged periods of time, an additional challenge arises in the form of evolutionary effects, through mutations to the constructs involved. This will effect the behaviour of the GRNs in unpredictable ways.

In the current paradigm, modelling and experiment agree qualitatively. The leap to quantitative modelling, that accurately describes the designed systems, needs to be taken for the full potential of Synthetic Biology to be harnessed. The challenges face by the field are considerable [121], but the key challenge currently faced is ensuring that models and experiments converge to allow for quantitative design to occur. To tackle this challenge, it is crucial that the experimental and theoretical aspects of the research are brought together, either within individuals with broad skill sets, or in very tight-knit teams, where communication and mutual understanding are central.

BIBLIOGRAPHY

[1] DN Adamson and HN Lim. 'Essential requirements for robust signaling in Hfq dependent small RNA networks'. In: *PLoS Computational Biology* 7.8 (2011).

[2] DN Adamson and HN Lim. 'Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation'. In: *Proceedings of the National Academy of Sciences* 110.32 (2013), pp. 13120–5.

[3] T Afroz and CL Beisel. 'Understanding and exploiting feedback in synthetic biology'. In: *Chemical Engineering Science* 103 (2013), pp. 1–12.

[4] H Aiba. 'Mechanism of RNA silencing by Hfq-binding small RNAs.' In: *Current Opinion in Microbiology* 10.2 (2007), pp. 134–9.

[5] B Alberts, A Johnson, J Lewis, D Morgan, M Raff, K Roberts and P Walter. *Molecular Biology of the Cell*. 5th. Garland Science, 2008.

[6] B Alberts, A Johnson, J Lewis, D Morgan, M Raff, K Roberts and P Walter. *Molecular Biology of the Cell*. 6th. Garland Science, 2014, p. 1464.

[7] R F Allison, J C Sorenson, M E Kelly, F B Armstrong and W G Dougherty. 'Sequence determination of the capsid protein gene and flanking regions of tobacco etch virus: Evidence for synthesis and processing of a polyprotein in potyvirus genome expression.' In: *Proceedings of the National Academy of Sciences* 82.9715 (1985), pp. 3969–3972.

[8] U Alon. *An introduction to systems biology: design principles of biological circuits*. Boca Raton, Florida: Chapman & Hall, 2006.

[9] U Alon. 'Network motifs: theory and experimental approaches.' In: *Nature Reviews. Genetics* 8.6 (2007), pp. 450–61.

[10] S Altuvia. 'Identification of bacterial small non-coding RNAs: experimental approaches.' In: *Current Opinion in Microbiology* 10.3 (2007), pp. 257–61.

[11] S Altuvia, D Weinstein-Fischer, A Zhang, L Postow and G Storz. 'A Small, Stable RNA Induced by Oxidative Stress: Role as a Pleiotropic Regulator and Antimutator'. In: *Cell* 90.1 (1997), pp. 43–53.

[12] W An and JW Chin. 'Synthesis of orthogonal transcription-translation networks.' In: *Proceedings of the National Academy of Sciences* 106.21 (2009), pp. 8477–82.

[13] J Anderson, N Strelkowa, GB Stan, T Douglas, J Savulescu, M Barahona and A Papachristodoulou. 'Engineering and ethical perspectives in synthetic biology'. In: *EMBO Reports* 13.7 (2012), pp. 584–590.

[14] E Andrianantoandro, S Basu, DK Karig and R Weiss. 'Synthetic biology: new engineering rules for an emerging discipline.' In: *Molecular Systems Biology* 2 (2006), p. 2006.0028.

[15] JAJ Arpino, EJ Hancock, J Anderson, M Barahona, GBV Stan, A Papachristodoulou and K Polizzi. 'Tuning the dials of Synthetic Biology.' In: *Microbiology* 159.Pt 7 (2013), pp. 1236–53.

[16] DW Austin, MS Allen, JM McCollum, RD Dar, JR Wilgus, GS Sayler, NF Samatova, CD Cox and ML Simpson. 'Gene network shaping of inherent noise spectra.' In: *Nature* 439.7076 (2006), pp. 608–11.

[17] T Baba, T Ara, M Hasegawa, Y Takai, Y Okumura, M Baba, KA Datsenko, M Tomita, BL Wanner and H Mori. 'Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.' In: *Molecular Systems Biology* 2 (2006), p. 2006.0008.

[18] H Backes, C Berens, V Helbl, S Walter, FX Schmid and W Hillen. 'Combinations of the alpha-helix-turn-alpha-helix motif of TetR with respective residues from LacI or 434Cro: DNA recognition, inducer binding, and urea- dependent denaturation'. In: *Biochemistry* 36.18 (1997), pp. 5311–5322.

[19] E Batchelor, TJ Silhavy and M Goulian. 'Continuous Control in Bacterial Regulatory'. In: *Journal of Bacteriology* 186.22 (2004), pp. 7618–7625.

[20] A Battesti, N Majdalani and S Gottesman. 'The RpoS-mediated general stress response in Escherichia coli.' en. In: *Annual Review of Microbiology* 65 (2011), pp. 189–213.

[21] JR Battista. 'Against all odds: the survival strategies of Deinococcus radiodurans.' In: *Annual Review of Microbiology* 51 (1997), pp. 203–24.

[22] TS Bayer and CD Smolke. 'Programmable ligand-controlled riboregulators of eukaryotic gene expression.' In: *Nature Biotechnology* 23.3 (2005), pp. 337–343.

[23] A Becskei and L Serrano. 'Engineering stability in gene networks by autoregulation.' In: *Nature* 405.June (2000), pp. 590–593.

[24] CL Beisel and G Storz. 'Base pairing small RNAs and their roles in global regulatory networks.' In: *FEMS Microbiology Reviews* 34.5 (2010), pp. 866–82.

[25] Matthew R. Bennett and Jeff Hasty. 'Microfluidic devices for measuring gene network dynamics in single cells'. In: *Nature Reviews Genetics* 10.9 (2009), pp. 628–638.

[26] PM Bhende and SM Egan. 'Amino acid-DNA contacts by RhaS: An AraC family transcription activator'. In: *Journal of Bacteriology* 181.17 (1999), pp. 5185–5192.

[27] E Bianconi et al. 'An estimation of the number of cells in the human body'. In: *Annals of Human Biology* 40.6 (2013), pp. 301–4460.

[28] FR Blattner et al. 'The Complete Genome Sequence of Escherichia coli K-12'. In: *Science* 277 (1997), pp. 1453–1462.

[29] O Borkowski, F Ceroni, GB Stan and Tom Ellis. 'Overloaded and stressed: whole-cell considerations for bacterial synthetic biology'. In: *Current Opinion in Microbiology* 33 (2016), pp. 123–130.

[30] M Bouvier, CM Sharma, F Mika, KH Nierhaus and J Vogel. 'Small RNA binding to 5' mRNA coding region inhibits translational initiation.' In: *Molecular Cell* 32.6 (2008), pp. 827–37.

[31] RW Bradley, M Buck and B Wang. 'Tools and Principles for Microbial Gene Circuit Engineering'. In: *Journal of Molecular Biology* 428.5 (2016), pp. 862–888.

[32] RG Brennan and TM Link. 'Hfq structure, function and ligand binding.' In: *Current Opinion in Microbiology* 10.2 (2007), pp. 125–33.

[33] JAN Brophy and CA Voigt. 'Principles of genetic circuit design.' In: *Nature Methods* 11.5 (2014), pp. 508–20.

[34] JAN Brophy and CA Voigt. 'Antisense transcription as a tool to tune gene expression'. In: *Molecular Systems Biology* 12.1 (2016), pp. 854–854.

[35] G Cambray, JC Guimaraes, VK Mutalik, C Lam, QA Mai, T Thimmaiah, JM Carothers, AP Arkin and D Endy. 'Measurement and modeling of intrinsic transcription terminators'. In: *Nucleic Acids Research* 41.9 (2013), pp. 5139–5148.

[36] DE Cameron, CJ Bashor and JJ Collins. 'A brief history of synthetic biology.' In: *Nature Reviews. Microbiology* 12.5 (2014), pp. 381–90.

[37] DE Cameron and JJ Collins. 'Tunable protein degradation in bacteria.' In: *Nature Biotechnology* November (2014), pp. 1–8.

[38] AJ Carpousis. 'The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E.' In: *Annual Review of Microbiology* 61.1 (2007), pp. 71–87.

[39] JC Carrington and WG Dougherty. 'Processing of the tobacco etch virus 49K protease requires autoproteolysis'. In: *Virology* 160.2 (1987), pp. 355–362.

[40] JC Carrington and WG Dougherty. 'A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing'. In: *Proceedings of the National Academy of Sciences* 85.10 (1988), pp. 3391–3395.

[41] F Ceroni, S Furini, A Stefan, A Hochkoeppler and E Giordano. 'A synthetic post-transcriptional controller to explore the modular design of gene circuits.' In: *ACS Synthetic Biology* 1.5 (2012), pp. 163–71.

[42] YC Chang, JP Armitage, A Papachristodoulou and GH Wadhams. 'A single phosphatase can convert a robust step response into a graded, tunable or adaptive response'. In: *Microbiology* 159.PART7 (2013), pp. 1276–1285.

[43] J Chappell, KE Watters, MK Takahashi and JB Lucks. 'A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future'. In: *Current Opinion in Chemical Biology* 28 (2015), pp. 47–56.

[44] S Chen, A Zhang, LB Blyn and G Storz. 'MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli'. In: *American Society for Microbiology* 186.20 (2004), pp. 6689–6697.

[45] YJ Chen, P Liu, AAK Nielsen, JAN Brophy, K Clancy, T Peterson and CA Voigt. 'Characterization of 582 natural and synthetic terminators and quantification of their design constraints.' In: *Nature Methods* 10.7 (2013), pp. 659–64.

[46] YJ Chen, N Dalchau, N Srinivas, A Phillips, L Cardelli, D Soloveichik and G Seelig. 'Programmable chemical controllers made from DNA.' In: *Nature nanotechnology* 8.10 (2013), pp. 755–62.

[47] GM Church, MB Elowitz, CD Smolke, CA Voigt and R Weiss. 'Realizing the potential of synthetic biology.' In: *Nature Reviews. Molecular Cell Biology* 15.4 (2014), pp. 289–294.

[48] C Cookson. 'Synthetic biology research to get £60m UK government boost'. In: *Financial Times* (2013).

[49] BP Cormack, RH Valdivia and S Falkow. 'FACS-optimized mutants of the green fluorescent protein (GFP)'. In: *Gene* 173.1 (1996), pp. 33–38.

[50] F Crick. 'Central Dogma of Molecular Biology'. In: *Nature* 227 (1970), pp. 561–563.

[51] EA Davidson and AD Ellington. 'Synthetic RNA circuits.' In: *Nature Chemical Biology* 3.1 (2007), pp. 23–8.

[52] JH Davis, AJ Rubin and RT Sauer. 'Design, construction and characterization of a set of insulated bacterial promoters'. In: *Nucleic Acids Research* 39.3 (2011), pp. 1131–1141.

[53] Nicholas De Lay, Daniel J. Schu and Susan Gottesman. 'Bacterial small RNA-based negative regulation: Hfq and its accomplices'. In: *Journal of Biological Chemistry* 288.12 (2013), pp. 7996–8003.

[54] J Degenkolb, M Takahashi, GA Ellestad and W Hillenl. 'Structural requirements of tetracycline-Tet repressor interaction : determination of equilibrium binding constants for tetracycline analogs with the Tet repressor'. In: *Antimicrobial Agents and Chemotherapy* 35.8 (1991), pp. 1591–1595.

[55] D Del Vecchio, AJ Ninfa and ED Sontag. 'Modular cell biology: retroactivity and insulation.' In: *Molecular Systems Biology* 4.161 (2008), p. 161.

[56] Douglas Densmore and Soha Hassoun. 'Design automation for synthetic biological systems'. In: *IEEE Design and Test of Computers* 29.3 (2012), pp. 7–20.

[57] MP Deutscher. 'Degradation of RNA in bacteria: Comparison of mRNA and stable RNA'. In: *Nucleic Acids Research* 34.2 (2006), pp. 659–666.

[58] MP Deutscher. 'How bacterial cells keep ribonucleases under control'. In: *FEMS Microbiology Reviews* 39.3 (2015), pp. 350–361.

[59] RM Dirks, JS Bois, JM Schaeffer, E Winfree and NA Pierce. 'Thermodynamic Analysis of Interacting Nucleic Acid Strands'. In: *SIAM Review* 49.1 (2007), pp. 65–88.

[60] J Dolan, J Anderson and A Papachristodoulou. 'A loop shaping approach for designing biological circuits'. In: *51st IEEE Conference on Decision and Control (CDC)*. Ieee, 2012, pp. 3614–3619.

[61] JC Doyle, BA Francis and AR Tannenbaum. *Feedback Control Theory*. Ed. by John Griffin. New York: Macmillan Publishing Company, 1992.

[62] Y Dublanche, K Michalodimitrakis, N Kummerer, M Foglierini and L Serrano. 'Noise in transcription negative feedback loops: simulation and experimental analysis'. In: *Molecular Systems Biology* 2.1744-4292 (Electronic) (2006), p. 41.

[63] SM Egan and RF Schleif. 'A Regulatory Cascade in the Induction of rhaBAD'. In: *Journal of Molecular Biology* 234 (1993), pp. 87–98.

[64] SM Egan and RF Schleif. 'DNA-dependent renaturation of an insoluble DNA binding protein. Identification of the RhaS binding site at rhaBAD.' In: *Journal of Molecular Biology* 243.5 (1994), pp. 821–9.

[65] H El-Samad, JP Goff and M Khammash. 'Calcium homeostasis and parturient hypocalcemia: an integral feedback perspective.' In: *Journal of Theoretical Biology* 214.1 (2002), pp. 17–29.

[66] MB Elowitz and S Leibler. 'A synthetic oscillatory network of transcriptional regulators.' In: *Nature* 403.6767 (2000), pp. 335–8.

[67] Carola Engler, Romy Kandzia and Sylvestre Marillonnet. 'A one pot, one step, precision cloning method with high throughput capability'. In: *PLoS ONE* 3.11 (2008).

[68] A Fender, J Elf, K Hampel, B Zimmermann and EGH Wagner. 'RNAs actively cycle on the Sm-like protein Hfq'. In: *Genes & Development* 3.Table 1 (2010), pp. 2621–2626.

[69] N Figueroa-Bossi, M Valentini, L Malleret, F Fiorini and L Bossi. 'Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target.' In: *Genes & Development* 23.17 (2009), pp. 2004–15.

[70] A Fire, S Xu, MK Montgomery, SA Kostas, SE Driver and CC Mello. 'Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans'. In: *Nature* 391 (1998), pp. 806–811.

[71] MT Franze de Fernandez, L Eoyang and JT August. 'Factor fraction required for the synthesis of bacteriophage Qβ-RNA'. In: *Nature* 219 (1968), pp. 588–590.

[72] C Fraser, J Gocayne, O White, M Adams, R Clayton, R Fleischmann and JC Venter. 'The Minimal Gene Complement of Mycoplasma genitalium'. In: *Science* 270 (1995), p. 397.

[73] KS Fröhlich and J Vogel. 'Activation of gene expression by small RNA'. In: *Current Opinion in Microbiology* 12.6 (2009), pp. 674–682.

[74] Michal Galdzicki et al. 'The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology'. In: *Nature Biotechnology* 32.6 (2014), pp. 545–550.

[75] MT Gallegos, R Schleif, A Bairoch, K Hofmann and JL Ramos. 'AraC/XylS Family of Transcriptional Regulators'. In: *Microbiology and Molecular Biology Reviews* 61.4 (1997), pp. 393–410.

[76] TS Gardner, CR Cantor and JJ Collins. 'Construction of a genetic toggle switch in Escherichia coli.' In: *Nature* 403.6767 (2000), pp. 339–42.

[77] TA Geissmann and D Touati. 'Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator.' In: *The EMBO Journal* 23.2 (2004), pp. 396–405.

[78] J Georg and WR Hess. 'cis-antisense RNA, another level of gene regulation in bacteria.' In: *Microbiology and Molecular Biology Reviews* 75.2 (2011), pp. 286–300.

[79] MJ Giacalone, AM Gentile, BT Lovitt, NL Berkley, CW Gunderson and MW Surber. 'Toxic protein expression in Escherichia coli using a rhamnose-based tightly regulated and tunable promoter system'. In: *BioTechniques* 40.3 (2006), pp. 355–364.

[80] DG Gibson, L Young, RY Chuang, JC Venter, CA Hutchison III and HO Smith. 'Enzymatic assembly of DNA molecules up to several hundred kilobases.' In: *Nature Methods* 6.5 (2009), pp. 343–5.

[81] DT Gillespie. 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions'. In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434.

[82] AL Goldberg. 'Degradation of abnormal proteins in Escherichia coli (protein breakdown-protein structure-mistranslation-amino acid analogs-puromycin).' In: *Proceedings of the National Academy of Sciences* 69.2 (1972), pp. 422–426.

[83] AL Goldberg. 'Protein degradation and protection against misfolded or damaged protein'. In: *Nature* 426.December (2003), pp. 895–899.

[84] TE Gorochowski, I Avcilar-Kucukgoze, RAL Bovenberg, JA Roubos and Z Ignatova. 'A Minimal Model of Ribosome Allocation Dynamics Captures Trade-offs in Expression between Endogenous and Synthetic Genes.' In: *ACS Synthetic Biology* (2016).

[85] S Gottesman. 'The small RNA regulators of Escherichia coli: roles and mechanisms'. In: *Annual Review of Microbiology* 58 (2004), pp. 303–328.

[86] S Gottesman. 'Micros for microbes: non-coding regulatory RNAs in bacteria'. In: *TRENDS in Genetics* 21.7 (2005), pp. 399–404.

[87] LL Granger, EB O'Hara, RF Wang, FV Meffen, K Armstrong, SD Yancey, Paul Babitzke and S R Kushner. 'The Escherichia coli mrsC gene is required for cell growth and mRNA decay.' In: *Journal of Bacteriology* 180.7 (1998), pp. 1920–1928.

[88] E Guisbert, VA Rhodius, N Ahuja, E Witkin and CA Gross. 'Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in Escherichia coli.' In: *Journal of Bacteriology* 189.5 (2007), pp. 1963–73.

[89] A Haldimann, LL Daniels and BL Wanner. 'Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the Escherichia coli phosphate regulon'. In: *Journal of Bacteriology* 180.5 (1998), pp. 1277–1286.

[90] P Hammar, M Walldén, D Fange, F Persson, O Baltekin, G Ullman, P Leroy and J Elf. 'Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation.' In: *Nature Genetics* 46.4 (2014), pp. 405–8.

[91] EJ Hancock, GB Stan, JAJ Arpino and A Papachristodoulou. 'Simplified mechanistic models of gene regulation for analysis and design'. In: *Journal of the Royal Society, Interface* 12.108 (2015), pp. 1–33.

[92] CB Harley and RP Reynolds. 'Analysis of E. coli promoter sequences'. In: *Nucleic Acids Research* 15.5 (1987), pp. 2343–2361.

[93] AWK Harris, JA Dolan, CL Kelly, J Anderson and A Papachristodoulou. 'Designing Genetic Feedback Controllers'. In: *IEEE Transactions on Biomedical Circuits and Systems* 9.4 (2015), pp. 475–484.

[94] W Hillen and C Berens. 'Mechanisms Underlying Expression of Tn10 Encoded Tetracycline Resistance'. In: *Annual Review of Microbiology* 48 (1994), pp. 345–369.

[95] W Hillen and B Unger. *Binding of four repressors to DS Tet operator region stablizes it against thermal denaturation*. 1982.

[96] W Hillen, G Klock, I Kaffenberger, LV Wray and WS Reznikoff. 'Purification of the TET repressor and TET operator from the transposon Tn10 and characterization of their interaction.' In: *Journal of Biological Chemistry* 257.11 (1982), pp. 6605–6613.

[97] W Hillen, C Gatz, L Altschmied, K Schollmeier and I Meier. 'Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions'. In: *Journal of Molecular Biology* 169.3 (1983), pp. 707–721.

[98] Nathan J Hillson. 'j5 DNA Assembly Design Automation BT - (null)'. In: *Methods in Molecular Biology* 1116.Chapter 17 (2013), pp. 245–269.

[99] J Houseley and D Tollervey. 'The Many Pathways of RNA Degradation'. In: *Cell* 136.4 (2009), pp. 763–776.

[100] V Hsiao, ELC de Los Santos, WR Whitaker, JE Dueber and RM Murray. 'Design and Implementation of a Biomolecular Concentration Tracker.' In: *ACS Synthetic Biology* (2014).

[101] M Hucka et al. 'The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models'. In: *Bioinformatics* 19.4 (2003), pp. 524–531.

[102] R Hussein and HN Lim. 'Disruption of small RNA signaling caused by competition for Hfq.' In: *Proceedings of the National Academy of Sciences* 108.3 (2011), pp. 1110–5.

[103] R Hussein and HN Lim. 'Direct comparison of small RNA and transcription factor signaling'. In: *Nucleic Acids Research* 40.7 (2012), pp. 7269–7279.

[104] B Ingalls. *Mathematical Modelling in Systems Biology: An Introduction*. Cambridge, Massachusetts: The MIT Press, 2013.

[105] FJ Isaacs, DJ Dwyer and JJ Collins. 'RNA synthetic biology.' In: *Nature Biotechnology* 24.5 (2006), pp. 545–54.

[106] FJ Isaacs, DJ Dwyer, C Ding, DD Pervouchine, CR Cantor and JJ Collins. 'Engineered riboregulators enable post-transcriptional control of gene expression.' In: *Nature Biotechnology* 22.7 (2004), pp. 841–7.

[107] M Ishiura, S Kutsuna, S Aoki, H Iwasaki, CR Andersson, A Tanabe, S S Golden, C H Johnson and T Kondo. 'Expression of a gene cluster kaiABC as a circadian feedback process in cyanobacteria'. In: *Science* 281.5382 (1998), pp. 1519–1523.

[108] BD Janssen and CS Hayes. 'The tmRNA ribosome rescue system'. In: *Advances in Protein Chemistry and Structual Biology* 805 (2012), pp. 1–30.

[109] RB Kapust and DS Waugh. 'Controlled intracellular processing of fusion proteins by TEV protease.' In: *Protein Expression and Purification* 19.2 (2000), pp. 312–8.

[110] JR Karr, JC Sanghvi, DN Macklin, MV Gutschow, JM Jacobs, B Bolival, N Assad-Garcia, JI Glass and Markus W Covert. 'A whole-cell computational model predicts phenotype from genotype.' In: *Cell* 150.2 (2012), pp. 389–401.

[111] KC Keiler, PR Waller and RT Sauer. 'Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA.' In: *Science* 271.5251 (1996), pp. 990–3.

[112] CL Kelly, C Pinske, BJ Murphy, A Parkin, F Armstrong, T Palmer and F Sargent. 'Integration of an [FeFe]-hydrogenase into the anaerobic metabolism of Escherichia coli'. In: *Biotechnology Reports* 8 (2015), pp. 94–104.

[113] CL Kelly et al. 'Synthetic chemical inducers and genetic decoupling enable orthogonal control of the rhaBAD promoter'. In: *ACS Synthetic Biology* (2016), acssynbio.6b00030.

[114] MB Kery, M Feldman, J Livny and B Tjaden. 'TargetRNA2: Identifying targets of small regulatory RNAs in bacteria'. In: *Nucleic Acids Research* 42.W1 (2014), pp. 124–129.

[115] AS Khalil and JJ Collins. 'Synthetic biology: applications come of age.' In: *Nature Reviews. Genetics* 11.5 (2010), pp. 367–79.

[116] HK Khalil and JW Grizzle. *Nonlinear Systems*. Second Edi. New Jersey: Prentice Hall, 1996.

[117] A Kolin, V Jevtic, L Swint-Kruse and SM Egan. 'Linker regions of the RhaS and RhaR proteins'. In: *Journal of Bacteriology* 189.1 (2007), pp. 269–271.

[118] A Kolin, V Balasubramaniam, J Skredenske, J Wickstrum and SM Egan. 'Differences in the Mechanism of the Allosteric L-Rhamnose Responses of the AraC/XylS Family Transcription Activators RhaS and RhaR'. In: *Molecular Microbiology* 68.2 (2008), pp. 448–461.

[119] E Kreyszig. *Advanced Engineering Mathematics*. John Wiley & Sons, 2010.

[120] H Kurata, H El-Samad, R Iwasaki, O Hisao, JC Doyle, I Grigorova, CA Gross and M Khammash. 'Module-based analysis of robustness tradeoffs in the heat shock response system'. In: *PLoS Computational Biology* 2.7 (2006), pp. 0663–0675.

[121] R Kwok. 'Five Hard Truths for Synthetic Biology'. In: *Nature* 463.January (2010), pp. 288–290.

[122] MR Lakin, S Youssef, L Cardelli and A Phillips. 'Abstractions for DNA circuit design'. In: *Journal of the Royal Society, Interface* 9.July (2012), pp. 470–486.

[123] D Lalaouna, M Simoneau-Roy, D Lafontaine and E Massé. 'Regulatory RNAs and target mRNA decay in prokaryotes'. In: *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1829.6-7 (2013), pp. 742–747.

[124] PR LeDuc, WC Messner and JP Wikswo. 'How do control-based approaches enter into biology?' In: *Annual Review of Biomedical Engineering* 13 (2011), pp. 369–96.

[125] T Lederer, M Takahashi and W Hillen. 'Thermodynamic analysis of tetracycline-mediated induction of Tet repressor by a quantitative methylation protection assay.' In: *Analytical Biochemistry* 232.2 (1995), pp. 190–6.

[126] TI Lee et al. 'Transcriptional Regulatory Networks in Saccharomyces cerevisiae'. In: *Science* 298 (2002), pp. 799–804.

[127] S Legewie, D Dienst, A Wilde, H Herzel and IM Axmann. 'Small RNAs establish delays and temporal thresholds in gene expression.' In: *Biophysical Journal* 95.7 (2008), pp. 3232–3238.

[128] E Levine and T Hwa. 'Small RNAs establish gene expression thresholds'. In: *Current Opinion in Microbiology* 11.6 (2008), pp. 574–579.

[129] E Levine, Z Zhang, T Kuhlman and T Hwa. 'Quantitative characteristics of gene regulation by small RNA.' In: *PLoS Biology* 5.9 (2007), e229.

[130] JC Liang, RJ Bloom and CD Smolke. 'Engineering biological systems with synthetic RNA molecules.' In: *Molecular Cell* 43.6 (2011), pp. 915–26.

[131] D Liu, T Evans and F Zhang. 'Applications and advances of metabolite biosensors for metabolic engineering'. In: *Metabolic Engineering* 31 (2015), pp. 35–43.

[132] R Lutz and H Bujard. 'Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O , the TetR/O and AraC/I1-I2 regulatory elements'. In: *Nucleic Acids Research* 25.6 (1997), pp. 1203–1210.

[133] N Majdalani, C Cunning, D Sledjeski, T Elliott and S Gottesman. 'DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription.' In: *Proceedings of the National Academy of Sciences* 95.21 (1998), pp. 12462–7.

[134] S Man, R Cheng, C Miao, Q Gong, Y Gu, X Lu, F Han and W Yu. 'Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria.' In: *Nucleic Acids Research* 39.8 (2011), e50.

[135] S Mangan, A Zaslaver and U Alon. 'The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks'. In: *Journal of Molecular Biology* 334.2 (2003), pp. 197–204.

[136] E Massé, FE Escorcia and S Gottesman. 'Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli'. In: *Genes & Development* 17.19 (2003), pp. 2374–2383.

[137] E Massé and S Gottesman. 'A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli.' In: *Proceedings of the National Academy of Sciences* 99.7 (2002), pp. 4620–5.

[138] E Massé, H Salvail, G Desnoyers and M Arguin. 'Small RNAs controlling iron metabolism.' In: *Current Opinion in Microbiology* 10.2 (2007), pp. 140–5.

[139] M Matsumoto and T Nishimura. 'Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator'. In: *ACM Transactions on Modeling and Computer Simulation* 8.1 (1998), pp. 3–30.

[140] MR Maurizi. 'Proteases and protein degradation in Escherichia coli.' In: *Experientia* 48 (1992), pp. 178–201.

[141] R McDaniel and R Weiss. 'Advances in synthetic biology: on the path from prototypes to applications.' In: *Current Opinion in Biotechnology* 16.4 (2005), pp. 476–83.

[142] P Mehta, S Goyal and NS Wingreen. 'A quantitative comparison of sRNA-based and protein-based gene regulation.' In: *Molecular Systems Biology* 4.221 (2008), p. 221.

[143] N Mermod and KN Timmis. 'Construction of a broad host range vector for the regulated expression of cloned genes in a range of gram-negative bacteria'. In: *Experientia* 42.2 (1986), pp. 105–105.

[144] A Milias-Argeitis, S Summers, J Stewart-Ornstein, I Zuleta, D Pincus, H El-Samad, M Khammash and J Lygeros. 'In silico feedback for in vivo regulation of a gene expression circuit.' In: *Nature Biotechnology* 29.12 (2011), pp. 1114–6.

[145] A Milias-Argeitis, M Rullan, SK Aoki, P Buchmann and M Khammash. 'Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth'. In: *Nature Communications* 7.May (2016), p. 12546.

[146] I Moll, S Grill, CO Gualerzi and U Blasi. 'Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control'. In: *Molecular Microbiology* 43.1 (2002), pp. 239–246.

[147] T Møller, T Franch, P Højrup, DR Keene, HP Bächinger, RG Brennan and P Valentin-Hansen. 'Hfq'. In: *Molecular Cell* 9.1 (2002), pp. 23–30.

[148] M Mondigler and M Ehrmann. 'Site-specific proteolysis of the Escherichia coli SecA protein in vivo'. In: *Journal of Bacteriology* 178.10 (1996), pp. 2986–2988.

[149] M Monsalve, B Calles, M Mencía, M Salas and F Rojo. 'Transcription activation or repression by phage psi 29 protein p4 depends on the strength of the RNA polymerase-promoter interactions.' In: *Molecular Cell* 1.1 (1997), pp. 99–107.

[150] K Moon and S Gottesman. 'Competition among Hfq-binding small RNAs in Escherichia coli'. In: *Molecular Microbiology* 82.6 (2011), pp. 1545–1562.

[151] D Na, SM Yoo, H Chung, H Park, JH Park and SY Lee. 'Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs.' In: *Nature Biotechnology* 31.2 (2013), pp. 170–4.

[152] N Nandagopal and MB Elowitz. 'Synthetic biology: integrated gene circuits.' In: *Science* 333.6047 (2011), pp. 1244–8.

[153] FC Neidhardt, PL Bloch and DF Smith. 'Culture medium for enterobacteria'. In: *Journal of Bacteriology* 119.3 (1974), pp. 736–747.

[154] FC Neidhardt and R Curtiss. *Escherichia coli and Salmonella: Cellular and Molecular Biology*. 2nd. Vol. 2. ASM Press, Washington D.C., 1996.

[155] D Nevozhay, RM Adams, KF Murphy, K Josic and G Balázsi. 'Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression.' In: *Proceedings of the National Academy of Sciences* 106.13 (2009), pp. 5123–8.

[156] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore and C. A. Voigt. 'Genetic circuit design automation'. In: *Science* 352.6281 (2016), aac7341–aac7341.

[157] K Oishi and E Klavins. 'Biomolecular implementation of linear I/O systems.' In: *IET Systems Biology* 5.4 (2011), pp. 252–60.

[158] M Olejniczak. 'Despite similar binding to the Hfq protein regulatory RNAs widely differ in their competition performance'. In: *Biochemistry* 50.21 (2011), pp. 4427–4440.

[159] JA Opdyke, JG Kang and G Storz. 'GadY, a small-RNA regulator of acid response genes in Escherichia coli'. In: *Journal of Bacteriology* 186.20 (2004).

[160] JD Orth, I Thiele and BO Palsson. 'What is flux balance analysis?' In: *Nature* 28.3 (2010), pp. 245–248.

[161] A Papachristodoulou and S Prajna. 'A tutorial on sum of squares techniques for systems analysis'. In: *Proceedings of the American Control Conference* (2005), pp. 2686–2700.

[162] A Papachristodoulou, J Anderson, G Valmorbida, S Prajna and PA Parrilo. *SOSTOOLS-Sum of Squares Optimization Toolbox, User's Guide*. 2013.

[163] K Papenfort, V Pfeiffer, F Mika, S Lucchini, JCD Hinton and J Vogel. 'σE-dependent small RNAs of Salmonella respond to membrane stress by accelerating global omp mRNA decay'. In: *Molecular Microbiology* 62.6 (2006), pp. 1674–1688.

[164] JD Pédelacq, S Cabantous, T Tran, TC Terwilliger and GS Waldo. 'Engineering and characterization of a superfolder green fluorescent protein.' In: *Nature Biotechnology* 24.1 (2006), pp. 79–88.

[165] G Peters, P Coussement, J Maertens, J Lammertyn and M De Mey. 'Putting RNA to work: Translating RNA fundamentals into biotechnological engineering practice'. In: *Biotechnology Advances* 33.8 (2015), pp. 1829–1844.

[166] AV Pinheiro, D Han, WM Shih and H Yan. 'Challenges and opportunities for structural DNA nanotechnology.' In: *Nature Nanotechnology* 6.12 (2011), pp. 763–72.

[167] Kathleen Postle, Toai T Nguyen and Kevin P Bertrand. 'Nucleic acids research.' In: *Nucleic acids research* 12.2 (1984), pp. 4849–4863.

[168] L Potvin-Trottier, ND Lord, G Vinnicombe and J Paulsson. 'Synchronous long-term oscillations in a synthetic gene circuit'. In: *Nature* 538.7626 (2016), pp. 1–4.

[169] J Power. 'The L-rhamnose genetic system in Escherichia coli K-12.' In: *Genetics* 55.3 (1967), pp. 557–568.

[170] TP Prescott and A Papachristodoulou. 'Synthetic biology: A control engineering perspective'. In: *European Control Conference*. 2014, pp. 1182–1186.

[171] JC Priscu et al. 'A microbiologically clean strategy for access to the Whillans Ice Stream subglacial environment'. In: *Antarctic Science* 25.05 (2013), pp. 637–647.

[172] WF Prouty and AL Goldberg. 'Fate of abnormal proteins in E. coli accumulation in intracellular granules before catabolism.' In: *Nature New Biology* 240.100 (1972), pp. 147–150.

[173] M Ptashne. *A Genetic Switch: Phage (Lambda) and Higher Organisms*. Cambridge, Massachusetts: Cell Press, 1992.

[174] PEM Purnick and R Weiss. 'The second wave of synthetic biology: from modules to systems.' In: *Nature Reviews. Molecular Cell Biology* 10.6 (2009), pp. 410–22.

[175] A Raj and A van Oudenaarden. 'Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences'. In: *Cell* 135.2 (2008), pp. 216–226.

[176] JL Ramos, M Martı, AJ Molina-henares, W Tera, R Brennan and R Tobes. 'The TetR family of transcriptional'. In: 69.2 (2005), pp. 1–31.

[177] M Resch, H Striegl, EM Henssler, M Sevvana, C Egerer-Sieber, E Schiltz, W Hillen and YA Muller. 'A protein functional leap: How a single mutation reverses the function of the transcription regulator TetR'. In: *Nucleic Acids Research* 36.13 (2008), pp. 4390–4401.

[178] M Riley, AB Pardee, F Jacob and J Monod. 'On the expression of a structural gene'. In: *Journal of Molecular Biology* 2.4 (1960), pp. 216–225.

[179] N Rosenfeld, MB Elowitz and U Alon. 'Negative autoregulation speeds the response times of transcription networks'. In: *Journal of Molecular Biology* 323.5 (2002), pp. 1–7.

[180] LJ Rothschild. 'Life in Extreme Environments.' In: *Ad Astra* 14.September 2000 (2001), pp. 1092–1101.

[181] WC Ruder, T Lu and JJ Collins. 'Synthetic biology moving into the clinic.' In: *Science* 333.6047 (2011), pp. 1248–52.

[182] S Sagawa, JE Shin, R Hussein and HN Lim. 'Paradoxical suppression of small RNA activity at high Hfq concentrations due to random-order binding'. In: *Nucleic Acids Research* 43.17 (2015), pp. 8502–8515.

[183] HM Salis. 'The ribosome binding site calculator'. In: *Methods in Enzymology*. Ed. by Christopher A. Voigt. 1st. Vol. 498. Elsevier, 2011, pp. 19–42.

[184] HM Salis, EA Mirsky and CA Voigt. 'Automated design of synthetic ribosome binding sites to control protein expression.' In: *Nature biotechnology* 27.10 (2009), pp. 946–50.

[185] M Saramago, C Barria, RF dos Santos, IJ Silva, V Pobre, S Domingues, JM Andrade, SC Viegas and CM Arraiano. 'The role of RNases in the regulation of small RNAs'. In: *Current Opinion in Microbiology* 18.1 (2014), pp. 105–115.

[186] O Scholz, P Schubert, M Kintrup and W Hillen. 'Tet repressor induction without Mg2+'. In: *Biochemistry* 39.35 (2000), pp. 10914–10920.

[187] J Shankleman. 'George Osborne: make UK a world leader in energy storage'. In: *The Guardian* (2012).

[188] SJ Sharp, J Schaack, L Cooley, DJ Burke and D Söll. 'Structure and transcription of eukaryotic tRNA genes.' In: *Critical Reviews in Biochemistry* 19.2 (1985), pp. 107–44.

[189] O Shimomura, FH Johnson and Y Saiga. 'Extraction , Purification and Properties of Aequorin , a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea'. In: *Journal of Cellular Physiology* 1353.165 (1962), pp. 223–239.

[190] Y Shimoni, G Friedlander, G Hetzroni, G Niv, S Altuvia, O Biham and H Margalit. 'Regulation of gene expression by small non-coding RNAs: a quantitative view.' In: *Molecular Systems Biology* 3.138 (2007), p. 138.

[191] YJ Shin and L Bleris. 'Linear control theory for gene network modeling.' In: *PloS One* 5.9 (2010), pp. 1–16.

[192] Michael J. Smanski, Hui Zhou, Jan Claesen, Ben Shen, Michael A. Fischbach and Christopher A. Voigt. 'Synthetic biology to access and expand nature's chemical diversity'. In: *Nature Reviews Microbiology* 14.3 (2016), pp. 135–149.

[193] TA Smith and BD Kohorn. 'Direct selection for sequences encoding proteases of known specificity.' In: *Proceedings of the National Academy of Sciences* 88.12 (1991), pp. 5159–5162.

[194] Larry Snyder and Wendy Champness. *Molecular Genetics of Biology*. 3th. ASM Press, Washington D.C., 2007.

[195] P Sobrero and C Valverde. 'The bacterial protein Hfq: much more than a mere RNA-binding factor'. In: *Critical Reviews in Microbiology* 38.4 (2012), pp. 276–299.

[196] T Soper, P Mandin, N Majdalani, S Gottesman and SA Woodson. 'Positive regulation by small RNAs and the role of Hfq'. In: *Proceedings of the National Academy of Sciences* 107.21 (2010), pp. 9602–9607.

[197] JT Stevens and JM Carothers. 'Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways'. In: *ACS Synthetic Biology* 4 (2015), pp. 107–115.

[198] G Storz, JA Opdyke and A Zhang. 'Controlling mRNA stability and translation with small, noncoding RNAs.' In: *Current Opinion in Microbiology* 7.2 (2004), pp. 140–144.

[199] P Stougaard, S Molin and K Nordstrom. 'RNAs involved in copy-number control and incompatibility of plasmid R1'. In: *Proceedings of the National Academy of Sciences* 78.10 (1981), pp. 6008–6012.

[200] J Stricker, S Cookson, MR Bennett, WH Mather, LS Tsimring and J Hasty. 'A fast, robust and tunable synthetic gene oscillator.' In: *Nature* 456.7221 (2008), pp. 516–9.

[201] M Takahashi, L Altschmied and W Hillen. 'Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements'. In: *Journal of Molecular Biology* 187.3 (1986), pp. 341–348.

[202] MK Takahashi et al. 'Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems'. In: *ACS Synthetic Biology* 4.5 (2015), pp. 503–515.

[203] K Terpe. 'Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems'. In: *Applied Microbiology and Biotechnology* 72.2 (2006), pp. 211–222.

[204] JF Tobin and RF Schleif. 'Positive regulation of the Escherichia coli l-rhamnose operon is mediated by the products of tandemly repeated regulatory genes'. In: *Journal of Molecular Biology* 196.4 (1987), pp. 789–799.

[205] JF Tobin and RF Schleif. 'Purification and properties of RhaR, the positive regulator of the L-rhamnose operons of Escherichia coli'. In: *Journal of Molecular Biology* 211.1 (1990), pp. 75–89.

[206] JF Tobin and RF Schleif. 'Transcription from the rha operon psr promoter.' In: *Journal of Molecular Biology* 211.1 (1990), pp. 1–4.

[207] J Tomizawa, T Itoh, G Selzer and T Som. 'Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA.' In: *Proceedings of the National Academy of Sciences* 78.3 (1981), pp. 1421–1425.

[208] DJ Trump. *"You know it, I know it, everybody knows it !"* 2016.

[209] M Ullah and O Wolkenhauer. *Stochastic Approaches for Systems Biology*. 1st. Springer Scientific, 2011.

[210] TB Updegrove, A Zhang and G Storz. 'Hfq: The flexible RNA matchmaker'. In: *Current Opinion in Microbiology* 30 (2016), pp. 133–138.

[211] JH Urban and J Vogel. 'Translational control and target recognition by Escherichia coli small RNAs in vivo'. In: *Nucleic Acids Research* 35.3 (2007), pp. 1018–1037.

[212] B Vecerek, I Moll and U Blasi. 'Translational autocontrol of the Escherichia coli hfq RNA chaperone gene'. In: *RNA* 11 (2005), pp. 976–984.

[213] P Vía, J Badía, L Baldomà, N Obradors and J Aguilar. 'Transcriptional regulation of the Escherichia coli rhaT gene.' In: *Microbiology* 142 (Pt 7.1 996 (1996), pp. 1833–1840.

[214] J Vogel, Jörg Vogel, Ben F. Luisi and BF Luisi. 'Hfq and its constellation of RNA'. In: *Nature Reviews Microbiology* 9.8 (2011), pp. 578–589.

[215] U Vogel and KF Jensen. 'The RNA chain elongation rate in Escherichia coli depends on the growth rate'. In: *Journal of Bacteriology* 176.10 (1994), pp. 2807–2813.

[216] EGH Wagner. 'Cycling of RNAs on Hfq.' In: *RNA Biology* 10.4 (2013), pp. 619–26.

[217] LS Waters and G Storz. 'Regulatory RNAs in bacteria.' In: *Cell* 136.4 (2009), pp. 615–28.

[218] JR Wickstrum and SM Egan. 'Amino acid contacts between sigma 70 domain 4 and the transcription activators RhaS and RhaR'. In: *Journal of Bacteriology* 186.18 (2004), pp. 6277–6285.

[219] JR Wickstrum, JM Skredenske, A Kolin, DJ Jin, J Fang and SM Egan. 'Transcription activation by the DNA-binding domain of the AraC family protein RhaS in the absence of its effector-binding domain'. In: *Journal of Bacteriology* 189.14 (2007), pp. 4984–4993.

[220] MN Win and CD Smolke. 'A modular and extensible RNA-based gene-regulatory platform for engineering cellular function'. In: *Proceedings of the National Academy of Sciences* 104.36 (2007), pp. 14283–14288.

[221] WW Wong, TY Tsai and JC Liao. 'Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator.' In: *Molecular Systems Biology* 3.130 (2007), p. 130.

[222] P Yakovchuk, E Protozanova and MD Frank-Kamenetskii. 'Base-stacking and base-pairing contributions into thermal stability of the DNA double helix'. In: *Nucleic Acids Research* 34.2 (2006), pp. 564–574.

[223] TM Yi, Y Huang, MI Simon and J Doyle. 'Robust perfect adaptation in bacterial chemotaxis through integral feedback control.' In: *Proceedings of the National Academy of Sciences* 97.9 (2000), pp. 4649–53.

[224] SM Yoo, D Na and SY Lee. 'Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli.' In: *Nature Protocols* 8.9 (2013), pp. 1694–707.

[225] B Yordanov, J Kim, RL Petersen, A Shudy, VV Kulkarni and A Phillips. 'Computational design of nucleic acid feedback control circuits'. In: *ACS Synthetic Biology* 3.8 (2014), pp. 600–616.

[226] F Zhang, JM Carothers and JD Keasling. 'Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids.' In: *Nature Biotechnology* 30.4 (2012), pp. 354–9.

[227] Yanxiang Zhou, Juliane Liepe, Xia Sheng, Michael P H Stumpf and Chris Barnes. 'GPU accelerated biochemical network simulation'. In: *Bioinformatics* 27.6 (2011), pp. 874–876.