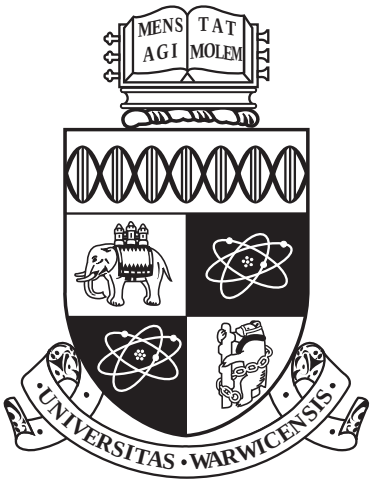

Design of genetic controllers to manage translational resource allocation in synthetic gene networks

by

Alexander P. S. Darlington

Thesis

Submitted to the University of Warwick

for admission to the degree of

Doctor of Philosophy

School of Engineering

February 2018

WARWICK
THE UNIVERSITY OF WARWICK

Contents

List of Figures

List of Tables

Acknowledgements

I am grateful to my supervisor Prof. Declan Bates for his invaluable support and guidance throughout the course of my Ph.D.

I am indebted to our experimental collaborators Dr Juhyun Kim and Dr José Jiménez firstly for providing the experimental validation of significant portions of this thesis and secondly for all of their insightful discussions.

I must thank Prof. Orkun Soyer for the numerous professional development opportunities he has given me over the last three and a half years. I also thank Dr Graham Jones for his invaluable support using WISB's HPC.

I am grateful to have been supported by a Ph.D. studentship from the UK's Synthetic Biology CDT (funded by the University of Warwick and the EPSRC & BBSRC Centre for Doctoral Training in Synthetic Biology (grant EP/L01649/1)) and for use of the High Performance Computing facilities provided by Warwick Integrative Synthetic Biology Centre, funded by BBSRC/EPSRC (grant BB/M017982/1).

Finally, I must thank my Parents and Jessica for all of their support over the last three and a half years and all of the years previously.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented in this thesis was carried out by the author in collaboration with Dr Juhyun Kim of Dr José I. Jiménez's group (University of Surrey). All modelling, computational design and analysis was carried out by the author and all experimental work was carried out by Dr Kim. The author contributed to the analysis and interpretation of the experimental results.

Parts of this thesis have been published (or submitted for publication) by the author:

1. Darlington, A.P.S. and Bates, D.G. (2016) "Host-aware modelling of a synthetic genetic oscillator", in the *Proceedings of 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Orlando, USA.

2. Darlington, A.P.S ., Kim, J., Jiménez, J.I. and Bates, D.G. (2017) "Design of a translation resource allocation controller to manage cellular resource limitations", in the *Proceedings of 20th IFAC World Congress*, Toulouse, France

3. Darlington, A.P.S., Kim, J., Jiménez, J.I. and Bates, D.G. (2018) "Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes." *Nature Communications*, ***9***, 695 *A draft of this paper was deposited on the community pre-print server* bioRχiv *and is available at* https://doi.org/10.1101/138362

4. Darlington, A.P.S., and Bates, D.G. "On the design of integral translational resource allocation controllers for synthetic cellular circuitry" (to be published in the Preceedings of the 57$^{\text{th}}$ IEEE Conference on Decision and Control)

5. Darlington, A.P.S., Kim, J., Jiménez, J.I. and Bates, D.G. "Engineering translational resource allocation controllers: Mechanistic models, design rules and biological implementations." (*under review*)
   A draft of this paper was deposited on the community pre-print server bioRχiv *and is available at* https://doi.org/10.1101/248948

(See Chapter 1 for how this published material relates to the thesis.)

Abstract

To control cellular processes synthetic biologists and biotechnologists often use regulation of gene expression; by regulating transcription it is assumed protein levels will follow. However the use of a common pool of gene expression resources results in the emergence of hidden interactions, couplings, between genes which are not immediately apparent from circuit topologies. This can result in a breakdown in the relationship between transcriptional regulation (input) and protein levels (output). Current evidence suggests that it is the number of free ribosomes which limits protein synthesis capacity and therefore creates these non-regulatory linkages.

In this thesis, the feasibility of dividing the cell's translational resources to reverse the breakdown in input-output responses and decouple co-expressed genes is demonstrated. A model of microbial growth which captures key aspects of host-circuit interactions is developed and demonstrates the feasibility of using orthogonal (circuit-specific) ribosomes for circuit gene expression; showing that by allocating circuit genes to both host and orthogonal translational pools these genes can be decoupled and the flux through a model biotechnological pathway can be improved.

The design of negative feedback controllers to allocate translational resources is investigated. These act to increase supply of orthogonal ribosomes as the demand for translational resources by the circuit increases. The stability of a number of controller architectures is investigated. An experimental prototype of the best performing controller architecture is able to reduce gene coupling by 50%.

The best controller architecture is carried forward for further analysis, and a detailed mechanistic model which can be used as a design tool is developed. A hard trade-off between gene expression and decoupling activity is identified and designs which manage this trade-off suggested. The controller is shown to ameliorate resource-mediated failures of modularity in a range of synthetic biology circuits.

Finally, a discussion on ways to produce a second generation of translational resource allocation controllers is provided.

Abbreviations

AHL	N-Acyl homoserine lactone
ATP	Adenosine triphosphate
DNA	Deoxyribonucleic acid
FBA	Flux balance analysis
GFP	Green fluorescent protein
GTP	Guanosine triphosphate
h-X	host translated protein X
IPTG	Isopropyl β-D-1-thiogalactopyranoside
min	Minute
mRNA	Messenger RNA
o-ribosome	quasi-orthogonal ribosome
o-rRNA	orthogonal 16S rRNA
o-X	orthogonally translated protein X
QSS	Quasi-steady state
RBA	Resource balance analysis
RBS	Ribosome binding site
RFP	Red fluorescent protein
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
S.D.	Standard deviation
sRNA	Small RNA
tRNA	Transfer RNA

Chapter 1

Introduction to this thesis

Synthetic biology is a nascent discipline which aims to make 'biology easier to engineer' [1]. By introducing synthetic gene circuits into microbial hosts, synthetic biologists and biotechnologists are able to control cell function. However, often these initial designs fail due to the effect of unforeseen interactions between the circuit and host cell or the effect of host constraints. This thesis is concerned with ways to mitigate the impact of host constraints on synthetic circuits.

Chapter 2 consists of an introduction to the field of synthetic biology and the key concepts/background used in this thesis. We discuss how host-circuit interactions can lead to the unpredictability, or outright failure, of synthetic circuit designs upon their implementation *in vivo*. We review recent advances in the study of the 'cellular economy' and how intrinsic limitations in the synthesis capacity of cells can impact circuit function. We introduce the concept of 'host-aware' modelling in comparison to other modelling frameworks used in biology and show, using a motivating example, that the inclusion of host factors can dramatically alter the potential behaviour of a synthetic circuit. We propose the use of so-called 'orthogonal' resources as a means of reducing the intrinsic limitations of the host by allowing the partitioning of synthesis capacity into 'host' and 'circuit' specific components.

In Chapter 3, we begin by developing a simple model of microbial growth and gene expression which takes account of key mediating factors in host-circuit interactions. We introduce the necessary changes to model the production and utilisation of orthogonal ribosomes. Using this model we demonstrate how this circuit-specific pool of ribosomes can act as simple allocators of translational activity between host and circuit genes and within circuits. We present experimental data from our collab-

orators which validates our observations. We show how these simple observations can be extended to a simple biotechnologically relevant pathway. We show how resource limitations can lead to counter intuitive decreases in metabolite output with increased pathway expression and show how controlling the allocation of ribosomes can be used to increase the yield.

In Chapter 4 we use this host model to design and investigate the ability of using feedback control to dynamically manipulate the size of the orthogonal ribosome pool so that it increases in size as demand from the synthetic circuit rises. We present the design of a number of different feasible feedback mechanisms and show how they can be realised *in vivo* in the form of simple gene networks. The design of a prototype controller is demonstrated *in vivo*.

Chapter 5 presents a detailed mechanistic model of the best feedback controller architecture. This model is then reduced to highlight key species and behaviours required for the controller to function. A trade-off between gene expression and resource-mediated gene-gene coupling is identified and parameter regimes which manage this proposed. The action of the controller is evaluated for a range of potential biological implementations. The controller is shown to successfully allocate resources across a range of previously published gene circuits.

Chapter 6 summarises the main findings from the work conducted and suggests further theoretical and experimental work required to produce a new generation of ribosomal allocation controllers. We propose future work to investigate controllers to manage transcriptional limitations and controllers which function at the 'interface' of host and circuit functions.

1.1 Publications arising from this work

The motivating example in Chapter 2 was published in:

> Darlington, A.P.S. and Bates, D.G. (2016) "Host-aware modelling of a synthetic genetic oscillator", *Proceedings of 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Orlando, USA.

The work in Chapter 3 and negative feedback controller from Chapter 4 has been published as:

> Darlington, A.P.S., Kim, J., Jiménez, J.I. and Bates, D.G. (2018) "Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes." *Nature Communications*, ***9***, 695

The discussion of the integral feedback control architectures in Chapter 4 has been accepted for presentation and subsequent publication in the conference Proceedings:

> Darlington, A.P.S., and Bates, D.G. (2018) "On the design of integral translational resource allocation controllers for synthetic cellular circuitry" to appear in the *Proceedings of the* 57^{th} *IEEE Conference on Decision and Control*, Miami, USA.

The initial work on the mechanistic model in Chapter 5 was published in:

> Darlington, A.P.S ., Kim, J., Jiménez, J.I. and Bates, D.G. (2017) "Design of a translation resource allocation controller to manage cellular resource limitations", *Proceedings of* $20^{\boldsymbol{th}}$ *IFAC World Congress on Automatic Control*, Toulouse, France.

The completed work on the mechanistic model in Chapter 5 has been submitted:

> Darlington, A.P.S., Kim, J., Jiménez, J.I. and Bates, D.G. "Engineering translational resource allocation controllers: Mechanistic models, design rules and biological implementations." (under review)

Chapter 2

Resource allocation in synthetic gene circuits

2.1 A brief history of synthetic biology

The modern field of synthetic biology encompasses a range of modern biological and biotechnological disciplines; all of which aim to apply engineering principles and theory to the construction of useful novel biological components. These can range from novel protein structures that have been designed *de novo* to the creation of new microbial consortia capable of producing fine chemicals.

In this thesis, we focus on the design and implementation of information processing gene- and protein-based 'circuits'. These molecular intracellular control systems have a range of applications in biomedicine, environmental science and bioprocessing/biotechnology (for examples see [2]). Since 2000, there has been a range of successes in production of a variety of modular functional motifs such as switches [3], oscillators [4, 5], bandpass filters [6] and logic gates [7]. These subjects have been reviewed extensively elsewhere (e.g. [2, 8]). These motifs can be assembled into devices and systems to perform complex tasks, such as counters [9] and multi-input logical processing units [10]. These systems can be implemented within individual cells or distributed across multiple cells in synthetic microbial consortia [10–12].

Synthetic biology grew out of a desire to make biology 'easier to engineer' [1] and so a number of concepts from engineering have been co-opted; especially those of predictive modelling, standardisation, modularity, abstraction, and the separation of design and manufacture (all implemented with different levels of success) [13].

The aim of synthetic circuit designers is the creation of a large catalogue of well characterised genetic sequences – 'parts' – encoding promoters, ribosome binding sites and functional proteins [14]. These standardised parts can be assembled into simple devices (such as gene expression cassettes), which in turn can be combined into modules of defined function (such as logic gates). These modules can themselves be connected into systems, in a manner reminiscent of electrical circuit board design. These systems carry out complex functions, such as information processing in which an input signal is taken, some form of computation takes place (e.g. comparison to a threshold value) and an output produced (often this is a fluorescent protein, but ideally this would be an enzyme or activator of some downstream cellular function). At each stage the components should have known, defined and predictable function (i.e. they are modular). This abstraction hierarchy would allow circuit designers to assemble functional circuits in the conventional 'plug-and-play' approach used by engineers without the need for the circuit designer to understand the assembly methods or detailed sequence-level information. It would also allow part designers to work without paying heed to desired circuit function or final dynamics. The Registry of Standard Biological Parts ('BioBricks') [14, 15] and the Synthetic Biology Open Language [16] aim to impose this hierarchical structure to create the required levels of abstraction. There is increasing success at designing circuits in this manner without the need for iterative (and expensive/time consuming) cycles of design and testing (e.g. [10]) but all too often circuits require extensive 'debugging' by tweaking and re-tweaking components (although these methods themselves are improving, e.g. [17]).

2.2 Failure of synthetic gene circuits

Often synthetic biologists view the host organism as a static chassis or hardware into which a plasmid-based software can be 'loaded and executed' [13]. However, this analogy to computer science does not hold. Synthetic circuits rely on their hosts to supply the necessary building blocks and gene expression machinery for the plasmid-encoded program to be successfully 'run', i.e. expressed. A better analogy is that synthetic circuit genes parasitise their host cell: the circuit genes compete with the host genes for metabolic building blocks as well as the machinery needed to interpret and implement the 'encoded program'. This can result in circuit failure; which we define as an un-desirable (un-designed) circuit behaviour.

Here we briefly review the commonly observed causes for synthetic circuit failure.

These are reviewed extensively elsewhere (e.g. [8, 18–20]) but we provide a brief summary as a means to highlight where this thesis makes a contribution.

At present most gene circuits are implemented in bacteria using genetic elements themselves derived from bacteria. These components can form an interface between the host genome and synthetic circuit leading to unexpected interactions. For example, the commonly used repressor LacI is an *Escherichia coli* gene and is itself often used in *E. coli* circuits. In this case the natural LacI gene must be deleted from the host to prevent interference. In more complex cases synthetic components can have off target effects due to sequence or structural homologies. Careful design of sequence and the use of components from distant relatives can increase orthogonality; i.e. they do not interact with host sequences or processes [21]. Distribution of circuit modules across different strains allows component reuse as compartmentalisation prevents off-target effects (e.g. [7, 22]). In Section 2.6 we develop this concept further by considering the use of orthogonal gene expression machineries.

The DNA-sequence flanking a DNA part can have significant effects on its function. This can affect the function of promoters, ribosome binding sites (RBSs) and transcriptional terminators. Thermodynamic models based on DNA or RNA folding can be used to assess potential sequence effects and aid design to remove them (e.g. the RBS calculator developed in [23]). Post-translational RNA processing by the use of internal self-cleaving RNA sequences have been used to remove the 5'-untranslated region which can remove undesirable effects on the RBS [24].

The connection of separately characterised circuits can result in unexpected interactions due to titration of upstream components by downstream DNA sequences. This feedback between modules is referred to as retroactivity [25] and can have significant effects on the function of the upstream module. The introduction of insulation motifs (such as transcriptional feedback and kinase-phosphatase cycles) in between the modules can be used to mitigate these effects [26, 27]. Negative transcriptional feedback minimises the impact of retroactivity by simultaneously amplifying the input signal and using negative feedback to remove the disturbance caused by the retroactivity signal. Kinase-phosphatase cycles utilise an excess of protein which is posphorylated by a kinase and dephosphorylated by a phosphatase to create the amplification signal. In one state, e.g. the phosphorylated form, the protein activates the downstream module; the proteins' activities are regulated by varying the concentration of kinase or phosphatase. The input signal is passed to the insulation device by increasing or decreasing the concentration of the kinase.

Increasing awareness of host and environmental context dependencies has led to systematic testing of simple circuits in a variety of host backgrounds (e.g. [28–30]). It has been shown that circuit expression can vary up to 1,000 fold depending on chassis selection [29]. Together these studies do not identify a trend between circuit behaviour, strain genotype and/or growth conditions suggesting host-circuit interactions are highly dependent upon the specific conditions. In [30], Moser *et al.* express a previously characterised set of logic gates in an industrial strain of *E coli.* under different medium conditions and volumes. They find that one of the circuits fails when expressed in the lab strain under minimal media conditions but that function is rescued upon transfer to an industrial strain. They also find that whilst the qualitative circuit function can be maintained there are large differences in quantitative behaviour across different conditions.

2.3 The effects of resource limitations in synthetic biology

2.3.1 Limited gene expression resources in bacteria

A key cause of the strain and environmental context dependency is that both host genotype and growth conditions lead to differences in the gene expression resources in bacteria. The model bacterium *E. coli* contains some 5,000 genes which encode information needed for multiple core ('essential') and condition-specific life processes [31]. Cells exhibit behaviours through changes in their intracellular make-up; often by producing new proteins which confer new function in response to an external cue. When a specific genetic program is activated, RNA polymerases first bind a recognition sequence upstream of the main gene in what is known as the *promoter*. They then *transcribe* the DNA-encoded information from the genome (or plasmid) into an intermediate *messenger* RNA (mRNA). This mRNA is bound by ribosomes at the *ribosome binding site* (RBS) and *translated*. The ribosome progresses along the mRNA 'reading' the order of the bases and using this information to assemble a chain of amino acids (a polypeptide). This complex multistep process utilises a number of protein-based and RNA-based accessory factors. Upon completion of reading the mRNA, the polypeptide chain is released and undergoes further processing and folding to form a functional protein.

During balanced exponential growth (when cells are growing maximally) the number of RNA polymerases and ribosomes is constant [32, 33] – with the activity, and total

number of complexes, increasing with increasing growth rate. This creates a set of fixed limited resources for total gene expression which are distributed across all host genes. Experimental evidence suggests that, whilst both RNA polymerase and ribosomes represent limited resources, in most cases it is the level of *free ribosomes* that sets a limit on gene expression [33–39]. The level of RNA polymerase becomes limiting in certain circumstances (for example, where circuit RBS strength is low, and hence ribosome sequestration by circuit mRNAs for translation is also low) [38, 40].

There is also some evidence to suggest that in addition to ribosomal limitations, other translational resources such as transfer RNAs (which are involved in interpreting the mRNA sequence into an amino acid sequence) and amino acid metabolism also form limited resources. For example, the expression of these cellular components increases upon foreign gene expression, although individual changes appear to be circuit-specific [41]. Note that interactions at this level of gene expression are unavoidable if synthetic genes are to be expressed in microbial hosts as all foreign protein production will necessarily use the host supplied amino acids.

Saturation of the cell's protease activity, i.e. protein degradation machinery, can impact circuit function showing this can also form a limited resource in specific cases where circuits require high degradation rates (discussed further below) [42, 43].

2.3.2 Intra-circuit gene coupling

A key impact of limited gene expression resources is the emergence of 'hidden' interactions between genes. Through the sharing of resources, changes in the expression of one gene results in changes in the expression of other genes; creating effective regulatory linkages although no *direct* regulation between the modules is present. We term this phenomenon 'gene coupling' as the expression of one gene becomes connected to the expression of others.

This gene coupling has been demonstrated both theoretically and experimentally in a number of studies [38, 40, 44]. At present in most synthetic circuits protein levels are varied by tuning mRNA levels through the use of small molecule inducers, which via receptor proteins, act to increase or decrease transcription from genes. By inducing one gene and observing the effect on a constitutively expressed gene, coupling can be observed (Figure 2.1a). As the first gene (here RFP) is induced, the concentration of its mRNA rises. This increases the co-option of ribosomes for its own expression and so more protein is produced (as desired from the increasing

induction). However, as during exponential growth, resources are constant, the increase in translation of one gene necessitates the decrease in expression of others (here GFP). This means that as ribosomes are co-opted to the induced gene, the level of translational resources available for the constitutively expressed gene falls and so its expression falls (depicted in Figure 2.1b). The final protein concentration of neither gene follows from its respective mRNA (induction) profile creating a disconnect between desired function (mRNA induction/input) and observed function (protein/output) (Figure 2.1). This coupling effect can be quantified by calculating the slope of the resulting protein-protein so-called 'isocost line' plot. The isocost line represents the potential combinations in which the two proteins can be produced given the fixed budget, p_{budget} [38]. For example, for two proteins (p_1 and p_2) which equally demand cellular resources (e.g. same RBS strength, same mRNA length) then:

$$p_1 + p_2 = p_{budget} \tag{2.1}$$

Therefore, the value of p_2 in terms of p_1 is:

$$p_2 = p_{budget} - p_1 \tag{2.2}$$

In the absence of p_1, the concentration of p_2 is equal to the protein budget. As p_1 is induced, p_2 levels fall. In the simplest case where both proteins equally sequester gene expression resources, as p_2 falls by one protein, p_1 increases by one protein; hence the gradient of the isocost line is -1. See [38] for greater discussion of the implications on the slope of the isocost line if the genes do not equally compete for resources.

In an example of a more complex circuit, Qian *et al.* show both theoretically and experimentally that the action of an activation cascade can be inverted due to resource limitations [45]. They characterise two independent modules which both show monotonically increasing responses to an activator. Upon connecting these modules by making the output of the first the input of the second, they show that some designs show decreasing output from the downstream module as the upstream module is induced. The monotonically increasing response is inverted. Mathematical analysis shows that this is the result of increasing sequestration of ribosomes by the first module. Hence as the first is induced, fewer resources are available for expression of the downstream module. In a separate publication, the same authors show theoretically that resource limitations can cause an inhibition cascade to become bistable [46].

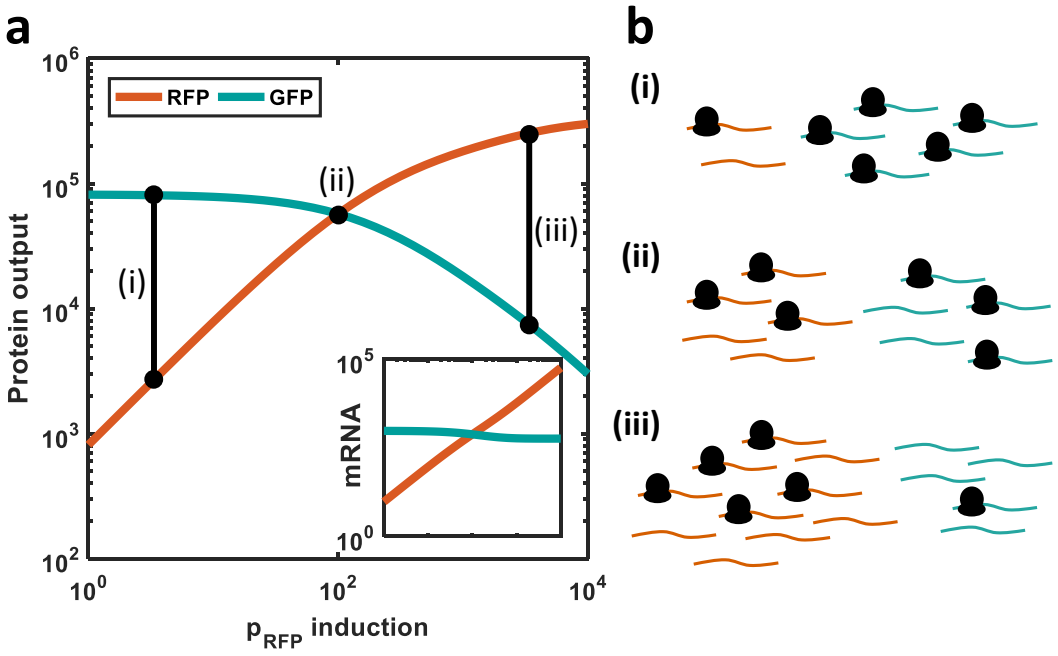

Figure 2.1: **Resource mediated intra-circuit gene coupling.** The protein and mRNA levels in a simple two-gene circuit consisting of the fluorescent reporter genes RFP and GFP are shown. The GFP is constitutively expressed (i.e. constant mRNA induction) while the RFP is induced. **(a)** Protein levels of the reporters as RFP is induced. Inset, mRNA levels. **(b)** Example of ribosome allocation across the mRNAs. Roman numerals relate to the points highlighted in (a). As RFP is induced there are more RFP mRNAs (red curved lines) and so ribosomes (black shapes) are reallocated. Note that the number of GFP mRNAs (green lines) is constant.

Cookson *et al.* show that coupling can be mediated by proteases [42]. They implement a simple two gene circuit oscillatory motif which shares a protease with a third inducible gene. Upon induction, this third gene begins to oscillate. At high levels of induction the clock's behaviour fails. The third gene shares no regulatory linkages with the clock motif but rather the interactions arise due to usage of the common, saturated, protease pool. Coupling mediated by translational competition in dynamic circuits such as these has not yet been shown experimentally. This is largely because most work in this circuit design area focuses on design and construction of genetic oscillators *themselves* without attention paid to surrounding genes.

2.3.3 Global effects on gene expression

As described above in Section 2.3.1, the levels of cellular resources correlate with cell exponential growth rate; with increasing growth rates coinciding with higher resources. This results in global changes in gene expression with changes in host growth rate (reviewed fully in [47]). For example, the output from a constitutive promoter will decrease with increasing host growth rate as the ribosomal mass fraction increases with growth rate; hence decreasing the available synthesis capacity for non-ribosomal genes. Therefore, the expression from a constitutive gene is dependent upon growth conditions; with expression increasing with increasing nutrient quality (and hence growth rate). For example, Tan *et al.* expressed a bacteriophage RNA polymerase in *E. coli* which transcribes itself and the reporter green fluorescent protein (GFP) expecting that the positive feedback loop would cause all cells to express GFP [48]. However, in reality the system appears bistable; with cells either expressing or not expressing GFP. This bistability in a non-cooperative system appears because the expression of the foreign RNA polymerase is somewhat toxic and results in growth retardation, which in turn results in decreased expression due to a fall in cellular resources.

2.3.4 Burden

By definition, the use of host resources by synthetic circuit reduces the availability of those resources for host gene expression. This results in a so-called 'burden', a growth defect observed in host cells (quantified thoroughly in [37]). In addition to the use of cellular resources creating a growth defect, burden can also be imposed by the production of a toxic metabolite or protein (e.g. [48]). There is considerable new found interest in managing these gross physiological effects (e.g. it was recently

proposed that cell free assays could be used to discriminate between designs of different burden [49]). The presence of this growth defect creates an evolutionary pressure which results in circuit function loss over time [50, 51].

2.4 Ameliorating the effects of resource limitations

Here we review the recent advances which have been developed to reduce the impact of resource limitations on synthetic circuits. These include new circuit design rules and the development of modelling frameworks which take account of host constraints.

2.4.1 Circuit redesign in light of host constraints

In [37], Ceroni *et al.* develop and validate a mechanistic model of translation in which they consider both the movement of a ribosome along an mRNA codon by codon and also competition between mRNAs for ribosome binding. They use this model to determine circuit designs to maximise gene expression at a minimum perturbation to a simulated host gene (a proxy for the host as a whole). Consistent with the aim of minimising load on the ribosome pool they show that a design where a gene is expressed from a high copy number plasmid with a weak RBS is better tolerated than a gene expressed from a medium copy plasmid with a strong RBS, even if the parameters are tuned to produce the same output protein. This model allows the simulation of different codon usages, enabling the impact of different codon selection to be included in designs. Their model suggests that the use of rare codons only limits gene expression when a strong RBS is used. They do not consider coupling between circuit genes (which could be incorporated with minor modifications) or circuit dynamics (which would require more extensive modifications). The authors' main conclusions are that RBSs should be weakened to minimise ribosome sequestration by circuit genes.

In [38], Gyorgy *et al.* develop and validate mechanistic models of transcription and translation which allow the design of two-gene circuits which show reduced levels of coupling; i.e. the expression of one gene is maintained as a second is induced. They show both theoretically and experimentally that if the constitutive gene has a strong RBS it will be less sensitive to the effects of other genes in the circuit. (Note the apparent contradiction with the work of Ceroni *et al.* who recommend using a

weaker RBS. In their work, Ceroni *et al.* are concerned with mimising impact on the host cell rather than removing coupling between circuit genes.)

Qian and del Vecchio develop mechanistic models of gene activation and repression which take account of competition for RNA polymerase and ribosomes producing new Hill function-type expressions which take account of resources and introducing metrics which describe each gene's propensity for resources [45, 46]. Using these models, the authors design and experimentally implement an activation cascade by using components which make best use of the resources available.

2.4.2 Internal circuit feedback

It is well known that the incorporation of negative feedback into a process increases robustness. Whilst negative feedback loops have been used extensively in synthetic biology to control cellular processes, here we review the three cases where they have been implemented to remove or reduce the effects of resource limitations.

In a theoretical work, Hamadeh and del Vecchio consider a mechanistic model of two genes which interact through both RNA polymerase and ribosome resource pools [52]. They evaluate the impact of three feedback loops (i) mRNA inhibiting transcription, (ii) proteins inhibiting translation, and (iii) proteins inhibiting transcription. They show both algebraically and by simulation that (iii) is the most efficient (note that this is the motif most present in the *E. coli* genomic network [53]) and highlight that a combination of (i) and (ii) is equally effective.

In [44], Shopera *et al.* demonstrate that incorporating negative feedback loops into the circuits results in decoupling of the two independent circuit modules. The negative feedback that they employ mimics the type (iii) feedback of Hamadeh and del Vecchio but with an intermediate activator protein, i.e. the module protein is co-expressed with an activator protein which itself activates a third downstream protein which inhibits the first module protein. However, due to this topology they find that as the strength of the negative feedback increases, coupling increases due to sequestration of resources by the complex feedback loop.

In [54], Qian and del Vecchio mathematically investigate a similar negative feedback loop to that implemented by Shopera *et al.*; modelling a system whereby a circuit protein activates expression of an inhibitory small RNA. They show that inclusion of this feedback mechanism can remove the resource-mediated inversion of an activation cascade (as discussed in Section 2.3.2) allowing the monotonically increasing

response to the inducer to be restored regardless of resource sequestration by each gene of the network.

2.4.3 Balancing couplings

As discussed previously, limitations in the protein decay machinery in the form of protease enzymes can cause couplings to occur. In [55], McBridge and del Vecchio show both theoretically and numerically that as a common pool of proteases becomes saturated protein levels increase. They show that this increase can compensate for the decrease in protein concentration due to ribosome sharing at biologically reasonable protease concentrations. However, they also observe 'positive' coupling, where simulated expression of a constitutively expressed gene increase in response to the increase in an induced gene (i.e. a positively sloped isocost line) due to protease saturation.

2.4.4 Host-aware design frameworks

The models described above begin to include RNA polymerase and ribosomes as limited resources which allows circuit (re-)design taking account of gene coupling effects (Section 2.3.2), however, these models do not take into account the potential effects of global changes in resource levels and burden (Sections 2.3.3 and 2.3.4). By incorporating additional model species representing 'host factors' and metabolic effects, we can take account of these effects. This has led to the development of 'host-aware' modelling frameworks. As with all design tools, the complexity of these host-aware models must balance their ease of use with their accuracy.

Carrera *et al.* develop expressions relating resource levels to specific growth rate which they parameterise based on a series of induction experiments in *E. coli* [36]. Their model is able to capture the effects of foreign gene expression and agrees with the data in [33] that cells can tolerate foreign protein production up to $\sim 30\%$ of their proteome. However, the empirical expressions relating resource production to growth rate precludes detailed insights into how host resources are regulated.

Weiße *et al.* develop a simplified chemical reaction network model of bacterial life processes; including a simple metabolism, transcriptome and proteome [56]. Unlike in the Carrera *et al.* model, the resources within that of Weiße *et al.* are produced within the model allowing greater mechanistic insight. This model forms the basis of the host-aware model used in this thesis and so is described fully in Chapter 3.

Development of these models is an on-going area of research. Recently, Liao *et al.* have published a more detailed model of host physiology which takes account of multiple cellular resources, including explicit inclusion of RNA polymerase and a simple amino acid metabolism [57]. This model is much more complex than that derived by Weiße *et al.* potentially complicating its use as a design tool.

The 'gold-standard' of host-circuit interaction modelling is the creation of a large scale dynamic model of a whole bacterial cell or cells. Purcell *et al.* used the prototype whole-cell model of *Mycoplasma genitalium* to demonstrate the use of whole cell modelling in synthetic biology design [58, 59]. A variety of simulations of a negative feedback loop, including different codon usage regimes and plasmid copy numbers, show good agreement with published experimental data, including prediction of burden effects. However, whilst the *M. genitalium* model is the most complete whole cell model to date this is not an industrially relevant microbe and so is not used as a chassis for synthetic biology applications. An *E. coli* whole cell model is currently being developed (see intermediate results in [60]) providing a potential design tool. Whole cell modelling is a young field and a significant number of challenges to their development remain [61]. These models are time consuming to simulate and their high complexity complicates their use as design tools.

2.4.5 Other methods for ameliorating resource limitation effects

For completeness, here we briefly discuss other approaches which can be used to allow for design in light of resource limitations.

Flux balance analysis (FBA) is a genome-wide constraint based numerical method for predicting steady state fluxes through a microbe's metabolic network during exponential growth. In [62], Goelzer *et al.* develop an extension to FBA which they term Resource Balance Analysis (RBA) which accounts for allocation of a 'resource budget' across various cellular systems. This allows the determination of how raw metabolites are partitioned to enzymes (which catalyse the flux through the network) and other cellular machinery (including ribosomes). However, like FBA this method does not allow the simulation of dynamic circuit behaviour but has use in the design of new biochemical pathways.

To overcome the difficulties involved in implementing gene circuits *in vivo*, cell-free transcription-translation (Tx-Tl) systems are being investigated as a potential rapid prototyping platform. In these systems circuit genes are expressed *in vitro* in a cytoplasm derived cocktail of metabolites, RNA polymerases, ribosomes and

accessory factors. These systems are easier and less time consuming to work with than bacteria and so multiple circuit designs can be implemented and tested before transformation into bacteria for final testing [63]. The lack of host factors and growth removes the complexities of implementing circuits in hosts for rapid assessment and the level of different gene expression resources can be well controlled. Provisional results suggest that the results determined in *in vitro* systems translate well to *in vivo* systems and can be used to separate the effects of metabolic loading, resource limitations and genotypic interactions [49]. However, these systems are expensive, require extensive further validation and still require large amounts of wet laboratory time. Recent results suggest that they still experience resource limitation problems. In [64], Gygory and Murray develop models of a simple gene circuit expressed in a cell-free system and find that the experimental observations can only be recapitulated *in silico* upon the addition of competition for the gene expression machinery.

2.5 The need for host-aware approaches in synthetic biology: A motivating example

Here we investigate modelling the effects of resource limitations in the context of the design of an oscillator circuit implemented *in vivo* by Stricker *et al.* [5] (topology depicted in Figure 2.2a). This gene circuit consists of an activator and repressor, both under mutual control of each other. The activator (A) triggers its own expression and that of the repressor (I), while the repressor inhibits the expression of both A and I.

We model promoter regulation in the oscillator circuit using Hill functions. Given that both transcription factors affect the same promoter, the regulatory effect is the product of the two transcription factors, and thus we have that:

$$\mathcal{R}(p_A, p_I) = \frac{\alpha}{1 + \alpha + \beta + \alpha \cdot \beta} \tag{2.3}$$

where α is the positive effect of the activator $(p_A/k_A)^{h_A}$ and β is the inhibitory effect of the inhibitor $(p_I/k_I)^{h_I}$. The variables p_A and p_I are the number of activator and repressor molecules respectively. The parameters k_A and k_I are the number of molecules for half the binding sites to be occupied for the activator and repressor respectively while h_A and h_I are the respective Hill coefficients. The transcription rate of each gene is the product of the maximal rate and the regulation effect.

Derivation of an isolated circuit model First we apply conventional techniques to develop a simple model of the circuit which tracks only the circuit mRNAs and proteins (e.g. [4, 5], discussed extensively in [65]). Here mRNAs are created spontaneously and then converted into proteins by a process modelled as a first order rate equation, with constant k_{eff} (equivalent to the proxy chemical reaction mRNA $\rightarrow$ protein). This neglects transcription, translation and energy consumption. All species are degraded and we account for dilution at cell division due to growth λ. Applying the Law of Mass Action to these reactions allows us to develop the following differential equations for the dynamics of mRNA (m) and protein (p) for each gene:

$$\dot{m} = \omega_0 + \omega \cdot \mathcal{R}(p_A, p_I) - (\delta_m + \lambda_{eff}) \cdot m \tag{2.4}$$

$$\dot{p} = k_{eff} \cdot m - (\delta_p + \lambda_{eff}) \cdot p \tag{2.5}$$

Growth rate and translation rate are global parameters linked to the internal status of the host cell. To represent the behaviour of the host we need to estimate fixed values for these parameters *a priori*. We *assume* a cell doubling time of approximately 30 minutes and we estimate an effective global protein production rate (k_{eff}) at 4 molecules per minute (based on the general assumption that a protein is 300 amino acids long and produced at 20 amino acids per second [31]).

Derivation of a host-aware circuit model Our starting point for the inclusion of more complex host-circuit interactions in the oscillator circuit is the microbial trade-off model developed by Weiße *et al.* [56]. This model is discussed extensively in Chapter 3 and so will not be discussed here save to say that it consists of a series of ordinary differential equations which track the time evolution of a simple metabolism, tanscriptome and proteome and includes the utilisation of both an 'energy' substrate and ribosomes. Here mRNAs are produced only when sufficient energy is present. They are then bound by ribosomes (R). Proteins are produced from translation complexes (c). Protein production liberates mRNA and free ribosomes. All components are diluted at the cell's growth rate (λ) which itself is calculated within the model such that $\lambda \propto \Sigma(c)$. We also modify the host equations as needed to model the effect on host metabolism and sequestration of free ribosomes by the circuit genes. We use the parameters determined by Weiße *et al.* for the implementation of the host components.

To model the oscillator circuit in this framework, we introduce the following equa-

Gene expression reactions	Isolated circuit	Host-aware circuit
$\emptyset \xrightarrow{\omega_0} m$ [a]	w_0	w_0
$\emptyset \xrightarrow{\omega \cdot \mathcal{R}} m$ [b]	$\omega \cdot \mathcal{R}(p_A, p_I)$	$\omega \cdot \mathcal{R}(p_A, p_I) \cdot T_X(e)$ [c]
$m + R \xrightarrow{b} c$	n/a	$b \cdot R \cdot m$
$c \xrightarrow{u} R + m$	n/a	$u \cdot c$
$c \xrightarrow{k} R + m + p$	n/a	$k \cdot c = \frac{\gamma(e)}{n} \cdot c$ [c]
$m \xrightarrow{k} p$	$k_{eff} \cdot m$	n/a
$m \xrightarrow{\delta_m} \emptyset$	$\delta_m \cdot m$	$\delta_m \cdot m$
$p \xrightarrow{\delta_p} \emptyset$	$\delta_p \cdot p$	$\delta_p \cdot p$
$X \xrightarrow{g} \emptyset$ [d]	$\lambda_{eff} \cdot X$	$\lambda \cdot X$, $\lambda \propto \Sigma c$

Table 2.1: **Comparison of isolated circuit and host-aware model reactions.**
[a] Transcription due to promoter leakage.
[b] Transcription due to regulation by transcription factors.
[c] $T_X(e)$ and $\gamma(e)$ are functions developed by Weiße *et al.* Both functions scale reactions according to the cell's internal energy status. $T_X(e)$ scales the rate at which mRNAs are produced and $\gamma(e)$ scales the protein production rate. Rates are scaled by $e/(o+e)$ where o is the threshold value. As $e \to 0$ both expressions tend to 0 so that translation and transcription cease at zero energy and as $e \to \infty$ both tend to 1, so at maximal cellular energy transcription and translation are maximal. See [56] for full derivation.
[d] The dilution reaction of the generic species X. For the host-aware model this is λ and is determined by the model. For the isolated and ribosome models the growth rate is estimated as λ_{eff}.

tions to the host model to describe the activator and repressor:

$$\dot{m} = \omega_0 + \omega \cdot \mathcal{R}(p_A,\ p_I) \cdot T_X(e) - b \cdot R \cdot m + u \cdot c + T_L(c,\ e) - (\delta_m + \lambda) \cdot m \quad (2.6)$$

$$\dot{c} = b \cdot R \cdot m - u \cdot c - T_L(c,\ e) - \lambda c \quad (2.7)$$

$$\dot{p} = T_L(c,\ e) - (\delta_p + \lambda) \cdot p \quad (2.8)$$

For the purposes of comparison, all reactions and their corresponding rate equations are collected in Table 2.1.

Different modelling frameworks suggest different designs for implementation. As briefly discussed, the use of the isolated circuit model which neglects host processes requires the estimation of effective protein synthesis and growth rates. To investigate the effect of these estimations, we select a parameter set which produces oscillations in the host-aware framework and vary k_{eff} and λ_{eff} in the isolated

circuit, resourceless, model. The emergence or stabilization of oscillations is highly dependent upon choice of the 'lumped' effective host parameter values (Figure 2.2b). By simulating the same circuit in the host-aware model we find that the agreement between the two models is highly dependent upon the chosen value of k_{eff}. Estimating this value from the literature is highly complex as often a range of average translation rates measured under a variety of growth conditions are reported. The need for these 'effective host' parameters can introduce significant inaccuracies into model predictions and hence extend the circuit design cycle.

We identified a number of parameter sets which lead to oscillations in both the isolated circuit and host-aware model. However, we see large differences in the period and amplitude of the oscillations predicted between the two models (Figure 2.2d). In only one parameter set do the two models produce approximately the same period ($\sim$120 minutes) (Figure 2.2d, *left*) whilst in the others we see the host-aware model producing much longer periods. Comparing the phase planes of the isolated circuit model and host-aware models (Figure 2.2c) we can see two main effects of host-circuit interactions: in some instances the addition of host factors acts to dampen the oscillations observed in the isolated model, while in others, the host factors stabilize the decaying oscillations observed in the isolated model (Figure 2.2c).

A key aspect of synthetic biology is the identification of parameters which determine circuit behaviour. By carrying out a simple local sensitivity analysis we assessed the effect each parameter has on the period. Whilst both models produce similar profiles there are large differences in the sensitivities of the decay parameters (Figure 2.2e). We see generally smaller changes in the period of oscillations in the host-aware model for most parameters, implying that host-circuit interactions may increase circuit robustness. (This observation is consistent with the observations of Stricker *et al.* who found that upon implementation of their proposed design *in vivo*, the circuit oscillates in a larger parameter range, as determined by inducer concentration, than predicted.)

This simple case study shows that guidance on appropriate choices of parameter values determined by simple models which neglect host processes for *in vivo* implementation may lead to non-functional designs. Explicit consideration of models containing resource limitations during the design cycle of synthetic circuits has the potential to significantly increase the efficiency and robustness with which such circuits can be implemented *in vivo*.

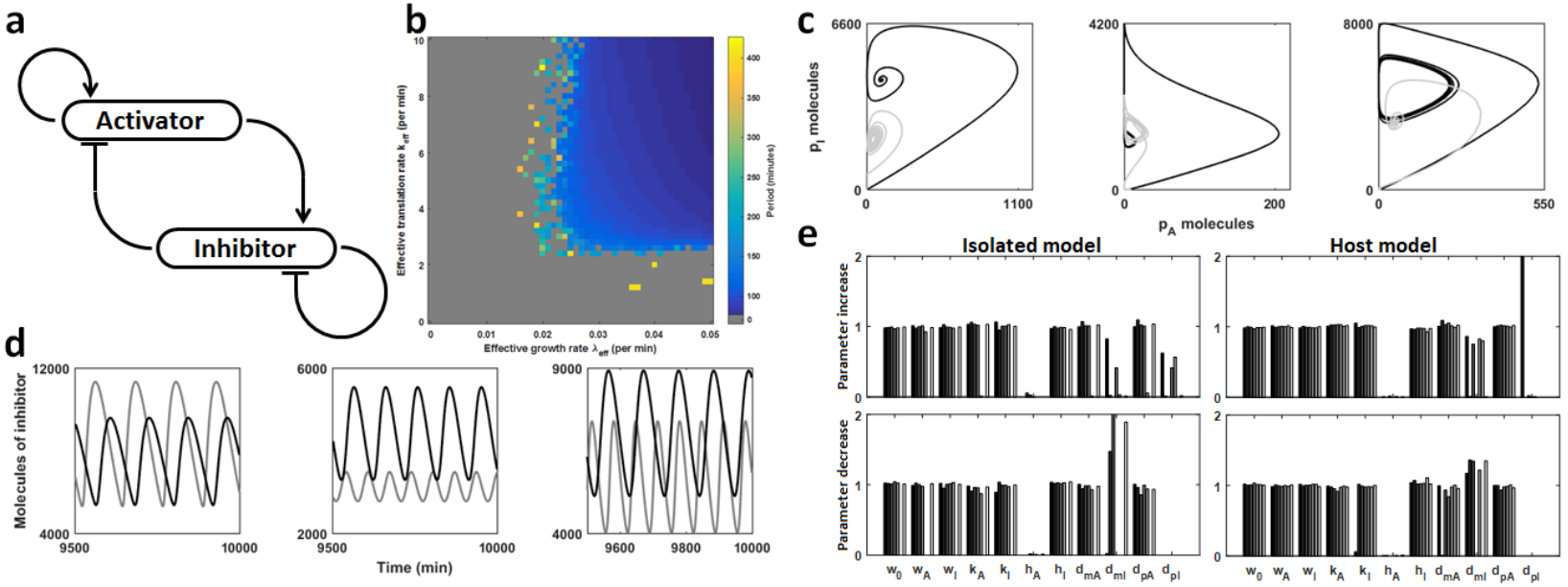

Figure 2.2: **Comparison of simulations in the isolated circuit and host-aware models. (a)** Topology of the Stricker *et al.* oscillator. **(b)** The period of oscillations in the isolated circuit model is highly dependent upon the approximated host parameters. Circuit parameters are kept constant while k_{eff} and λ_{eff} are varied as shown. The period of the resulting oscillations is shown. **(c)** Phase plane analysis of the classical (grey line) and host-aware (black line) models. *left*, the host-aware model is stable, the classical model produces a stable limit cycle. *middle*, both models produce stable limit cycles. *right*, the host-aware model produces a stable limit cycle whilst the classical model is stable. **(d)** Comparison of the oscillations produced by the classical model (grey lines) and host-aware model (black lines). The oscillations vary significantly (although there is good agreement in *left* panel.) **(e)** Comparison of sensitivity analysis. Parameters common to both models were perturbed in turn for all sets which produce oscillations in both models. Results (y-axis) are the ratio of the new period to the original period. The y-axis is limited to 2. Colour of bar corresponds to each parameter set. *Upper* panels, the parameter is increased from its nominal value. *Lower* panels, the parameter is decreased from its nominal value.

2.6 Current approaches to resource allocation management

In addition to circuit redesign and new modelling frameworks to reduce the effects of resource limitation, there is the potential to manage the division of the cell's gene expression machinery between host and circuit function through the use of 'orthogonal' components. As previously discussed circuits are composed of orthogonal components – i.e. the circuit proteins do not (should not) interact with the host – although they still make use of host resources for their expression. This concept of orthogonality can be extended to the gene expression machinery. Here we review the potential methods to create a circuit specific pool of gene expression resources.

As briefly discussed, protein-encoding genes are first transcribed into mRNA by RNA polymerase. Transcription is initiated by an RNA polymerase complex, the *holoenzyme*, which contains a promoter-targeting σ factor. This dissociates during elongation. These σ factors are involved in large scale responses which require co-regulation of multiple genes; most genes are transcribed by the σ_{70} 'general house keeping' factor but some, such as those expressed in response to sudden heat stress or nitrogen starvation, have specific condition-dependent σ factors. These promoters have a different promoter architecture to those using the σ_{70} [66]. Using the σ factors from sufficiently distinct species can create an orthogonal pool of RNA polymerase holoenzymes which are circuit specific (see e.g. [67]). Note the core RNA polymerase complex is not orthogonal and will still bind to host σ factors, and therefore host genes, upon termination of transgene expression, making these σ factor-based orthogonality schemes only *quasi*-orthogonal. Some bacterial viruses (bacteriophage) contain within their small genomes a highly specific small RNA polymerase, often only composed of one subunit, which have high affinity for the viral promoters of their genome. These truly orthogonal polymerases can be co-opted for circuit gene expression (extensively reviewed in [68]).

Proteases are relatively small proteins and have diversified over the period of bacterial evolution. Proteases, much like transcription factors, can be exchanged between species to produce an orthogonal protein decay system [69]. By adding or changing the C-terminal amino acids of the protein of interest it can be targeted to a specific host or orthogonal protease.

As discussed throughout this chapter the key mediator of gene coupling is the limited number of ribosomes in most contexts. However, due to the universality and

complexity of the cell's translational machinery, there does not exist a sufficiently distinct ribosome which can be co-opted into *E. coli* to create a truly orthogonal ribosome pool. However, translational capacity can be divided into host and circuit specific functions by the use of synthetic ribosomal RNA (rRNA) components to create a *quasi*-orthogonal ribosome ('o-ribosome') system [70–73]. The binding interactions between an mRNA and the 16S rRNA of the small ribosomal subunit are known to be a key regulator of translation initiation [23], and thus o-ribosomes can be created by expressing a synthetic 16S rRNA. Evolving or designing the 5' sequences at and around the ribosome binding site of circuit mRNAs to interact with this synthetic sequence allows the creation of an orthogonal translation system. (Note that during the rest of this thesis we refer to the synthetic 5' sequence as an 'orthogonal RBS' although some modifications required may be outside of the true o-RBS). To date these o-ribosomes have mainly been used to probe various aspects of ribosome structure-function relationships (see e.g. [74]).

2.6.1 Resource allocators and controllers

These orthogonal components can be used as parts of conventional synthetic gene circuits as components in logic gates or for protein production. For example, in [75] they divide the T7 RNA polymerase into two fragments which must be co-expressed and assembled into a complex in order to function and so create a simple AND gate with the polymerase used as a transcriptional activator. In [76], the authors combine orthogonal transcription and translation to produce logic gates; for example both the o-RNA polymerase and o-ribosome are required to produce an output protein. In both these cases the orthogonal parts form components of the circuit itself much the same way as transcription factors or receptors; they do not take the form of a resource allocator (see below).

A recent work, published in [77], does utilise an o-ribosome pool as an expression system for multiple pathway enzymes. However the authors focus on the use of an orthogonal translation system as a biocontainment strategy for a biosynthetic pathway. If the circuit plasmid is passed to non-transgenic 'wild' bacteria the pathway cannot be expressed in the receiving organism as it lacks the corresponding o-ribosome translation system.

However, we argue that orthogonal resources should not simply be used as modules within a synthetic circuit (e.g. as a transcriptional activator in [75]) but rather should be used as orthogonal gene expression machines for circuit gene expression

as a whole. The constitutive expression of an orthogonal resource creates a subdivision of the cell's labour. Circuit genes can be targeted to utilise this circuit-specific orthogonal resource as in [77]. This forms a simple resource allocator. By incorporating demand-based feedback we propose the creation of resource allocation controllers which act to match resource supply to circuit demand.

Here we review the little work that has been carried out in this area.

In [78], Segall-Shapiro *et al.* utilise an RNA polymerase from the bacteriophage T7 which they divide into subunits such that the polymerase is only active when all subunits are expressed and assembled correctly. They use this system to create a resource allocator whereby a key component (denoted α) of the polymerase is used to modulate the transcriptional activity of a circuit as a whole. The authors propose that copy number effects can be removed by expressing the α fragment on the same plasmid as the circuit genes and modulating its expression by varying its promoter and RBS strength. In this way all the circuit genes can be tuned by the modification of only two components (rather than tuning each gene individually). In a work with similar goals, Kushwaha and Salis develop the Universal Bacterial Expression Resource (UBER) based on the T7 RNA polymerase [79]. This consists of a T7 RNA polymerase which activates its own expression and expression of a negative feedback loop which can be used to tune the expression capacity of the system as a whole. They demonstrate the function of this device across a range of bacterial species. Note that neither Segall-Shapiro *et al.* nor Kushwaha and Salis assess the ability of their resource allocators to dynamically respond to demand, rather these systems are designed to create an orthogonal transcription budget which can be used by circuit genes and offers a means of tuning the circuit expression as a whole with minimal changes.

Recently, Lillacci *et al.* have proposed the use of σ factor inhibiting proteins (anti-σ factors) as a means to control circuit gene expression [80]. The circuit genes are expressed by an RNA polymerase using an orthogonal σ factor. An anti-σ factor is co-produced with the circuit output. This acts to bind to and sequester the σ factor, so inhibiting its action and hence reducing circuit output. If circuit activity increases then the production of the anti-σ factor increases, inhibiting circuit expression. This controller mechanism is discussed in more detail in Chapter 4.

In addition to these orthogonal RNA polymerase resource allocators, recently two resource allocators have been proposed which are not based on orthogonal gene expression machineries. We discuss these briefly for completeness.

In [81], the authors express a ribonuclease MazF which degrades 96% of host mRNAs and so reduces competition for the gene expression machinery. In the presence of MazF expression, the output from a reporter gene increases due to reduced competition. By introducing the MazF recognition site into its own mRNA they 'close' the feedback loop such that MazF concentration is proportional to the need for mRNA degradation.

In a recent work, Ceroni *et al.* propose the use of a burden-based feedback device which acts to tune circuit function (here a simple metabolic pathway) in response to the cell's 'global health' [82]. They take advantage of the dCas9 protein which can inhibit promoters by steric hindrance when targeted to them by a short RNA sequence (the guide RNA, gRNA). The authors identify burden-sensitive promoters whose expression is activated when a burdensome circuit is expressed and place these upstream of a gRNA whose sequence is complementary to the circuit promoter. As the circuit genes are expressed, the cells experience burden hence the gRNA's are produced, taken up by the dCas9 protein which binds to, thereby inhibiting, the circuit promoter. Note that this is not a form of resource allocation management but rather demand management. This method introduces the dCas9 as an additional cellular resource; the dCas9 inhibition complexes are shared across the circuit promoters and so as circuits become larger additional regulatory interactions as the dCas9 itself becomes a limiting resource.

2.7 Contributions of this thesis

Given the experimental observations that it is free ribosome number that limits gene expression, in this work we consider the use of orthogonal ribosomes as a means to partition host and circuit gene expression activities. We build on this simple resource allocation scheme by considering the implementation of demand-based feedback mechanisms which act to increase the production of o-ribosomes as demand for translational capacity by the circuit increases. In this way we aim to complement the currently proposed RNA polymerase resource allocators and provide additional means of mediating host-circuit and circuit-circuit interactions.

Chapter 3

Resource allocation by use of orthogonal ribosomes

3.1 Introduction

In this chapter, we investigate the ability of orthogonal ribosomes to function as a separate translational resource for synthetic circuit gene expression. We first develop an ordinary differential equation model of microbial growth which includes the production of o-ribosomes. We use this model to assess the feasibility of using o-ribosomes for gene expression and consider how circuit genes can be allocated between host and orthogonal ribosome pools to reduce the effects of resource limitations. These findings are experimentally validated in *E. coli*. We extend the utility of our results by considering how targeting different genes encoding enzymes in a biochemical pathway to different translational pools can be used to increase flux through the pathway.

The experimental data in this chapter were provided by our collaborators Dr Juhyun Kim and Dr José I. Jiménez (University of Surrey). Analysis was carried out by the author in conjunction with Dr Juhyun Kim. For completeness, an overview of their experimental methods can be found in Section 3.10.

3.2 Development of the host model including o-ribosome production

Given that our work focuses on influencing the host ribosome pool, we take a host-aware design approach by developing a simple model of microbial physiology. This allows us to assess the effect of orthogonal ribosome production and usage on host physiology. We based our model on the ordinary differential equation model of microbial growth and gene expression trade-offs recently developed by Weiße *et al.* [56].

This model captures the three fundamental trade-offs in bacterial gene expression:

1. internal anabolic capacity ('energy') is limited by substrate import and enzyme activity
2. ribosomes are autocatalytic and compete with other genes for their own expression resulting in a finite translational capacity
3. finite proteome size creates competition for space (i.e. total number of proteins)

We refined this previously published model by extending the ribosome biosynthesis reactions to include the separate production of ribosomal protein and ribosomal RNA components. We introduced the necessary species to describe the production of orthogonal rRNAs and partition the ribosome pool. This model allows us to characterise the impact of dividing the cell's translational capacity between host and circuit genes.

The final model represents a simplified microbe with:

1. a minimal metabolism which converts an external substrate (s_e) into energy (e), via an internal substrate (s_i)
2. a minimal transcriptome containing mRNAs (denoted m_X, where X is the gene being encoded) and ribosomal rRNAs (r)
3. a minimal proteome with proteins representing the four main classes of protein function: metabolite transport/import (T), metabolic enzymes (E), ribosomal proteins (R) and 'generic' host proteins (H, the q-proteins of Scott *et al.* [33]). We refer to these host genes as the set $X \in \{T, E, H, R\}$.
4. a simplified ribosome biosynthesis scheme which describes the formation of ribosomes (R) from host rRNA (r) and ribosomal proteins (p_R, called the 'empty' ribosome below)

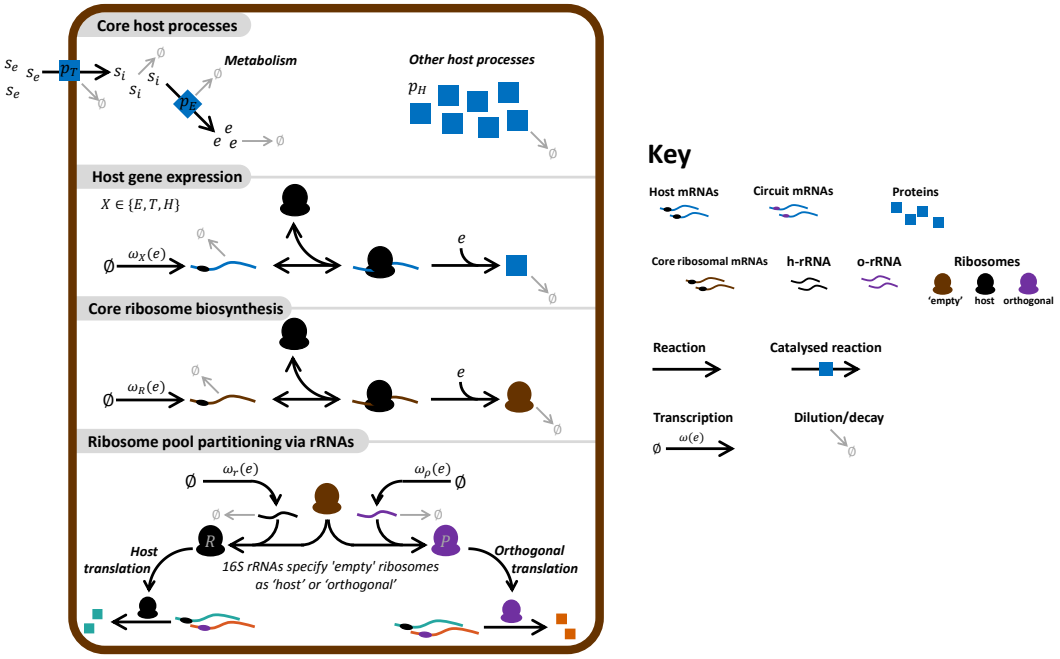

Figure 3.1: **Model schematic.** Main host processes captured by our model are metabolism, gene expression and ribosome biosynthesis. *NB:* Not all reactions are shown to simplify the schematic.

5. an orthogonal translation system made up of o-rRNA (ρ) and o-ribosomes (P).

Inclusion of the additional orthogonal translation system results in an additional trade-off due to competition for ribosomal proteins (p_R).

A schematic of the model is shown in Figure 3.1. The model is parameterised in terms of molecules per cell but note that, for simplicity, we do not consider stochastic effects, rather choosing to focus on population-level behaviours as revealed by ordinary differential equations regardless of molecule number.

Metabolism The metabolism is made up of an extracellular substrate (s_e) which is imported by the transport enzyme (p_T) to become the intracellular species s_i. The intracellular species is converted to the universal energy substrate (e) by enzyme (p_E). Both of these enzyme-mediated reactions are described by Michaelis-Menten kinetics. In addition to these processes, the intracellular substrate is diluted by cell growth ($\lambda \cdot s_i$ term in Equation 3.1).

$$\frac{ds_i}{dt} = \frac{v_T \cdot p_T \cdot s_e}{k_T + s_e} - \frac{v_E \cdot p_E \cdot s_i}{k_E + s_i} - \lambda \cdot s_i \tag{3.1}$$

Each molecule of intracellular substrate creates φ_e molecules of energy. This is a measure of the nutrient efficiency, which we maintain as high throughout (see [56] for a detailed discussion of energy implementation in the original model). Translation of each protein-encoding mRNA by the translation complexes consumes energy ($\sum(\dots)$ term). This energy consumption is proportional to both protein length (n_X) and number of translation complexes of that gene (c_X in the T_L function). The energy species is also diluted due to cell growth (λ).

$$\frac{de}{dt} = \varphi_e \cdot \frac{v_E \cdot p_E \cdot s_i}{k_E + s_i} - \sum_{X \in \{T,E,H,R\}} \Big(n_X \cdot \ T_L(c_X, e) \Big) - \lambda \cdot e \tag{3.2}$$

Host protein production The minimal mRNA-transcriptome consists of transporter genes (denoted by T), enzymes (E), additional host proteins (H) and ribosomal proteins (R). These host proteins form the set $X \in \{T,\ E,\ H,\ R\}$. Transcription of mRNAs is modelled as a spontaneous birth process scaled by the cell's current energy status - i.e. we do not account for RNA polymerase limitation or binding kinetics but we account for the energy dependence of these processes by modifying the rate by e levels. Messenger RNAs are born at the maximal rate ω_X which is scaled by (i) any regulatory interactions $\mathcal{R}$ and (ii) energy status ($e/(o_X + e)$), where o is an energy threshold. The regulatory interactions are described by Equation 3.3 for the respective species.

$$\mathcal{R} = 1, \quad X \in \{T, E, R\} \qquad \mathcal{R} = \frac{1}{1 + (p_H/k_H)^{h_H}}, \quad X = H \tag{3.3}$$

Upon birth, host mRNAs can reversibly bind/unbind free host ribosomes (R) to form translation complexes (c_X). Upon the termination of translation, mRNAs are released ($T_L(c_X, e)$ term, see Equation 3.6). mRNAs are also subject to decay (δ_{m_X}) and dilution due to growth (λ).

$$\frac{dm_X}{dt} = \omega_X \cdot \mathcal{R} \cdot \Big(\frac{e}{o_X + e} \Big) + T_L(c_X, e) - b_X \cdot R \cdot m_X + u_X \cdot c_X + (\delta_{m_X} + \lambda) \cdot m_X \tag{3.4}$$

The dynamics of the translation complex (c_X) follow from the description of the ribosome-mRNA interactions above, with the added degradation and dilution terms:

$$\frac{dc_X}{dt} = b_X \cdot R \cdot m_X - u_X \cdot c_X - T_L(c_X, e) - (\delta_R + \lambda) \cdot c_X \tag{3.5}$$

The T_L function describes the rate of translation of gene X. This expression, derived

in [56], relates the translation rate of individual genes to the cell's global translation rate (γ_{max} term), the gene's length (n_X) and the number of complexes currently translating that gene's mRNA (c_X).

$$T_L(c_X, e) = \frac{1}{n_X} \cdot \left(\frac{\gamma_{max} \cdot e}{k_\gamma + e}\right) \cdot c_i \tag{3.6}$$

Host proteins are born from translation complexes and are subject to decay and dilution:

$$\frac{dp_X}{dt} = T_L(c_X, e) - (\delta_{p_X} + \lambda) \cdot p_X \tag{3.7}$$

Note that when $X = R$ this expression only describes the production of the ribosomal protein component (p_R) and not its refinement into functional host ribosomes, which is outlined in Section 3.2.1.

Note also we have not explicitly included expressions describing the binding/unbinding of proteins to promoters to exert their action but rather use Hill functions to scale the mRNA production rate. For example, when we consider the auto-inhibitory effect of host proteins (i.e. when $X = H$) the derivative dp_H/dt follows the same form as dp_X/dt, and the interaction effect is only seen in dm_H/dt, i.e. the effect of the $\mathcal{R}$ term in Equation 3.4. There are no terms describing the interaction of the transcription factor or RNA polymerase recruitment promoter as there would be in a detailed mechanistic model (as discussed in Chapter 5) but rather a Hill-function is used.

Determination of growth rate Rather than being estimated from empirical/ prototype data, the growth rate is calculated within the model. This also allows effects on host physiology to be assessed by observing the change in this one value ('burden'). Growth rate (λ) is proportional to the product of the global translation rate (γ_{max} term) and number of translating complexes ($\sum(c_X)$). See [56] for a full derivation in terms of protein production and dilution.

$$\lambda = \frac{1}{M} \cdot \left(\frac{\gamma_{max} \cdot e}{k_\gamma + e}\right) \cdot \sum_{X \in \{T,E,H,R\}} \left(c_X\right) \tag{3.8}$$

3.2.1 Modelling ribosome biosynthesis and o-ribosome formation

We refined the original ribosome synthesis reactions to include the separate production of protein and rRNA based components. For the protein components, we

considered the production of a single large protein which represents the small and large ribosomal subunits and any accessory protein complexes.

We assume that host rRNAs are born spontaneously at a maximal rate scaled by the cell's internal energy, in a manner similar to the host mRNAs. This rRNA (r) binds empty ribosomes (p_R). The rRNA is subject to degradation and dilution. The dynamics of the host rRNA are therefore:

$$\frac{dr}{dt} = \omega_r \cdot \left(\frac{e}{o_r + e}\right) - b_r \cdot p_R \cdot r + u_r \cdot R - (\delta_r + \lambda) \cdot r \tag{3.9}$$

As described above, the protein component of the ribosome (p_R) follows the same dynamics as for other host proteins, as outlined in Equation 3.7 when $X = R$. In addition to the production and decay dynamics described in Equation 3.7, p_R undergoes processing by binding with the host rRNA. (We account for ribosome complex disassembly into different subunits by allowing this reaction to be reversible).

$$\frac{dp_R}{dt} = T_L(c_R, e) - (\delta_{p_R} + \lambda) \cdot p_R - b_r \cdot p_R \cdot r + u_r \cdot R \tag{3.10}$$

Free host ribosomes (R) are produced by the binding of host rRNA and empty ribosomes. They also take part in translation of each protein coding gene ($\sum(\dots)$ term). Free ribosomes bind mRNAs to form translation complexes (the m_X term) and are produced when these complexes dissociate before or upon the termination of protein synthesis; the $u_X \cdot c_X$ term and T_L function term respectively. Concurrently translation complexes are lost through dilution (λ) and degradation (δ_R, which we set to zero throughout but its inclusion allows processes such as inhibition by antibiotics to be included).

$$\frac{dR}{dt} = b_r \cdot p_R \cdot r - u_r \cdot R - \sum_{X \in \{T,E,H,R\}} \Big(T_L(c_X, e) - b_X \cdot R \cdot m_X + u_X \cdot c_X\Big) - (\delta_R + \lambda) \cdot R \tag{3.11}$$

By modelling ribosome biosynthesis in this way it incorporates the two important feedback loops which determine ribosome number: (i) host ribosomes are autocatalytic (i.e. R translates p_R) and (ii) ribosomal mRNA and rRNA transcription rates fall as 'energy' is consumed by protein production. This also incorporates the *quasi*-orthogonal nature of the o-ribosome pool by linking the two pools through competition for the protein-based components p_R (see below).

To model the production of orthogonal ribosomes (P) we consider the expression of a

new orthogonal 16S rRNA (ρ) which interacts with the empty ribosome. We assume that the orthogonal 16S rRNA follows the same dynamics as the host rRNA being produced in an energy dependent manner and reacting with empty ribosomes. We assume that this plasmid-carried gene will respond to energy changes in a manner similar to host genes and its binding kinetics will be the same as the host's and hence we parameterised it as outlined in Table 3.2.

$$\frac{d\rho}{dt} = \omega_\rho \cdot \left(\frac{e}{o_\rho + e}\right) - b_\rho \cdot p_R \cdot \rho + u_\rho \cdot P - (\delta_\rho + \lambda) \cdot \rho \tag{3.12}$$

We modify Equation 3.11 to take account of the o-rRNA – empty ribosome interactions which mirror the host rRNA interactions (blue terms in Equation 3.13).

$$\frac{dp_R}{dt} = T_L(c_R, e) - (\delta_{p_R} + \lambda) \cdot p_R - b_r \cdot p_R \cdot r + u_r \cdot R - b_\rho \cdot p_R \cdot \rho + u_\rho \cdot P \tag{3.13}$$

In the same manner as host ribosomes, free orthogonal ribosomes are produced by reversible binding of orthogonal rRNA and empty ribosomes. Functional orthogonal ribosomes bind and translate mRNAs which are specified to them (i.e. circuit genes, denoted by Y, see below) in the same manner as host ribosomes translate host mRNAs. Free o-ribosomes are subject to degradation and dilution due to growth.

$$\frac{dP}{dt} = b_\rho \cdot p_R \cdot \rho - u_\rho \cdot P - \sum_Y \left(T_L(c_Y, e) - b_Y \cdot P \cdot m_Y + u_Y \cdot c_Y\right) - (\delta_R + \lambda) \cdot P \tag{3.14}$$

3.2.2 Introduction of circuit genes

We introduce circuit genes by introducing new species and equations describing the production of mRNA, translation complexes and proteins. We assign these genes to the set Y. These take the same form as in Equations 3.4, 3.5 and 3.7 with the ribosome pool specified as either host or orthogonal as appropriate. Orthogonal ribosomes are specified by modifying R to P (see Section 3.2.1).

We modify Equation 3.2 to take account of the additional energy demand due to the circuit protein production ($\sum(\dots)$ term).

$$\frac{de}{dt} = \varphi_e \cdot \frac{v_E \cdot p_E \cdot s_i}{k_E + s_i} - \sum_{X \in \{T,E,H,R\}} \left(n_X \cdot \ T_L(c_X, e)\right) - \lambda \cdot e - \sum_Y \left(n_Y \cdot \ T_L(c_Y, e)\right) \tag{3.15}$$

We also modify Equation 3.11 to take into account the additional host ribosome

utilisation by circuit genes, if any, ($\sum(\dots)$ term). If circuit genes are translated by the orthogonal ribosome pool then Equation 3.14 is modified instead.

$$\begin{aligned}\frac{dR}{dt} = b_r \cdot p_R \cdot r - u_r \cdot R \; - \sum_{X \in \{T,E,H,R\}} \Big(T_L(c_X, e) - b_X \cdot R \cdot m_X + u_X \cdot c_X\Big) \dots \\ - \sum_{Y} \Big(T_L(c_Y, e) - b_Y \cdot R \cdot m_Y + u_Y \cdot c_Y\Big) \dots \\ - (\delta_R + \lambda) \cdot R\end{aligned} \tag{3.16}$$

We modify the growth rate Equation 3.8 to include the effect of circuit gene translation complexes ($\sum(c_Y)$ term):

$$\lambda = \frac{1}{M} \cdot \left(\frac{\gamma_{max} \cdot e}{k_\gamma + e}\right) \cdot \left(\sum_{X \in \{T,E,H,R\}} \Big(c_X\Big) + \sum_{Y} \Big(c_Y\Big) \right) \tag{3.17}$$

3.2.3 Parameterisation

We used the parameters derived by Weiße *et al.*, using the original protein-only ribosome production parameters for the new empty ribosome (p_R) species. We assumed that the transcription of the host ribosomal rRNA showed the same energy dependence as the protein component (i.e. $o_r = o_R$). See Table 3.2 for full details. Using the growth rate data reported in [33], we optimised additional parameters as described below.

Scott *et al.* grew *E. coli* in different carbon sources and across a range of different anti-ribosomal antibiotic concentrations [33]. They report growth rate (λ) and ribosomal mass fraction (Φ). To model this experimental set up, we incorporate the presence of inhibited ribosomes (c_X^ε) by modifying Equation 3.5 to include the new interactions with an antibiotic ε to produce Equation 3.18.

$$\begin{aligned}\frac{dc_X}{dt} = b_X \cdot R \cdot m_X - u_X \cdot c_X - T_L(c_X, e) - (\delta_R + \lambda) \cdot c_X \dots \\ - k_\varepsilon \cdot \epsilon \cdot c_X\end{aligned} \tag{3.18}$$

$$\frac{dc_X^\varepsilon}{dt} = k_\varepsilon \cdot \varepsilon \cdot c_X - (\delta_R + \lambda) \cdot c_X^\varepsilon \tag{3.19}$$

By simulating the model across a range of nutrient efficiencies and antibiotic concentration values we can estimate the values of λ and Φ. We take the nutrient efficiencies ($\varphi_e = [0.080, 0.115, 0.167, 0.240, 0.347, 0.500]$) and k_ε from [56]. The antibiotic concentrations $\varepsilon = [0, 2, 4, 8, 12]$ nM are taken from [33].

We used MATLAB's *genetic algorithm* to minimise the following cost function:

$$\text{cost} = \sum_{\varphi_e,\ c_0} \left((\lambda_{sim} - \lambda_{exp})^2 + (\Phi_{sim} - \Phi_{exp})^2 \right) \tag{3.20}$$

This equally weights the contributions of both growth rate and ribosomal mass faction estimates as the former is crucial for accurate prediction of circuit 'burden' and the latter is crucial for estimating the available cellular resources.

We optimised the host rRNA gene maximum transcription rate (ω_r) using MATLAB's using 0 and 10^5 rRNAs per min as the lower and upper bounds respectively. We chose our upper bound by assuming that maximal gene expression for a protein encoding gene is on the order of 10^3 RNAs per min and as rRNAs are present in multiple copy number and also driven from strong promoters we allowed ω_r to vary significantly above that ω_X for a protein coding gene [53].

The full host model shows good quantitative agreement with the experimental data published in [33] across most growth conditions (Table 3.1). The model shows best quantitative agreement with the cellular growth rate data (Figure 3.2). The simulations of the ribosomal mass fraction replicate the general trends observed in the data with ribosomal fraction increasing with antibiotic concentration and nutrient quality. Replication of this general trend is sufficient for initial testing of resource allocation schemes. There is good quantitative agreement with the experimental data at high levels of nutrient quality ($\phi_e \geq 0.347$) and low antibiotic concentration ($\varepsilon \leq 2$) which were used by our experimental collaborators (Section 3.10).

The parameters used for circuit genes and those describing the o-rRNA dynamics are shown in Table 3.3.

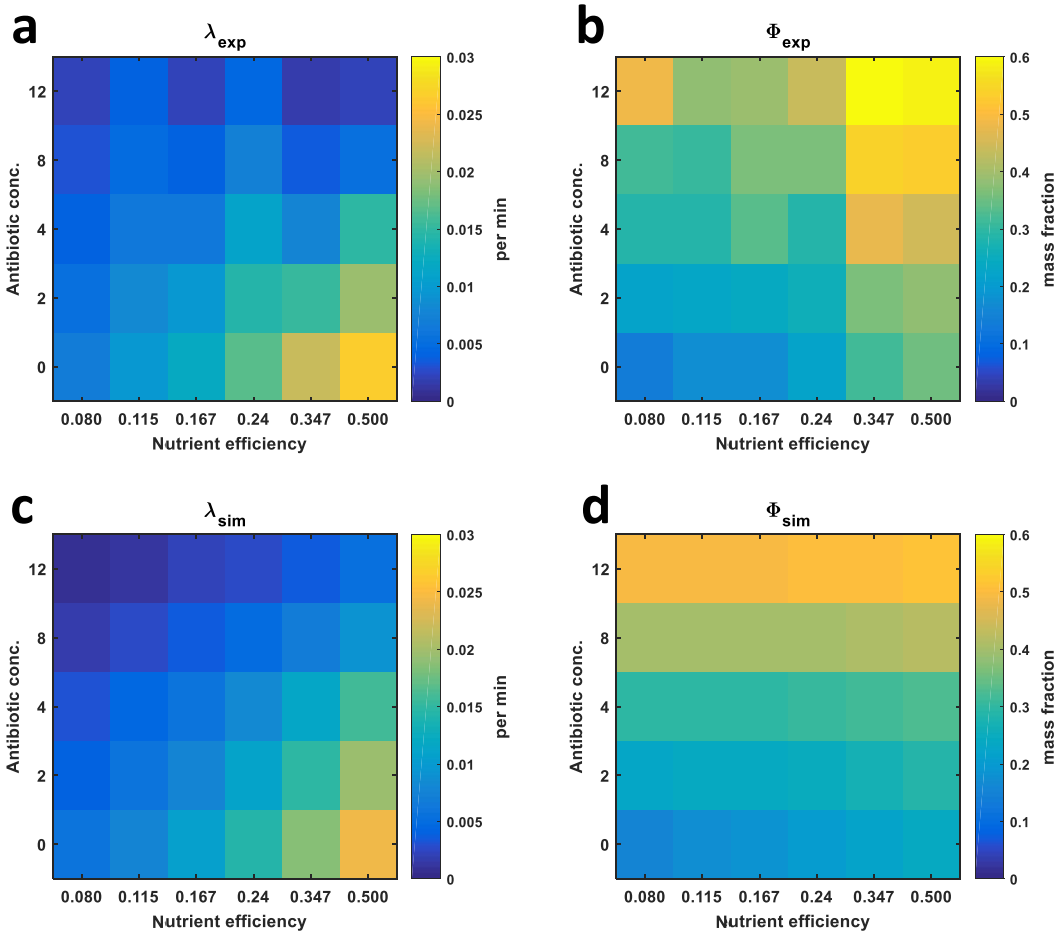

Figure 3.2: **Comparison of host model with Scott *et al.* experimental data.** Data were produced by growing cells at different nutrient qualities (which are quantified here as nutrient efficiency, ϕ_e) and ribosome-toxic antibiotic concentrations. Proteins and RNA were harvested and used to calculate ribosomal mass fraction (see [33] for details). **(a)** λ_{exp} Growth rate determined by Scott *et al.* **(b)** Φ_{exp} Ribosomal mass fraction determined by Scott *et al.* **(c)** λ_{sim} Simulated growth rate. **(d)** Φ_{sim} Simulated ribosomal mass fraction.

		Nutrient efficiency					
		0.080	0.115	0.167	0.240	0.347	0.500
Antibiotic conc.	0	0.0005	0.0000	0.0002	0.0003	0.0091	0.0129
	2	0.0001	0.0000	0.0000	0.0000	0.0093	0.0096
	4	0.0000	0.0002	0.0010	0.0005	0.0245	0.0140
	8	0.0064	0.0085	0.0015	0.0020	0.0185	0.0122
	12	0.0002	0.0138	0.0111	0.0036	0.0089	0.0059

Table 3.1: **Quality of final model fit.** The value of cost function (Equation 3.20) for different concentrations of antibiotic and different nutrient qualities. Simulation parameters are listed in Table 3.2. Experimental data were from Scott *et al.* [33] as discussed in Section 3.2.3.

Parameter	Value	Units	Notes	Ref
s_E	10^4	mc	External substrate	[56]
φ_e	0.5	–	Nutrient efficiency	[56]
v_T	728	mc·min^{-1}	Maximal nutrient import	[56]
v_E	5800	mc·min^{-1}	Maximal substrate-to-energy conversion	[56]
k_T	1000	mc	Transporter Michaelis-Menten constant	[56]
k_E	1000	mc	Enzyme Michaelis-Menten constant	[56]
$\omega_{\{T,E\}}$	4.14	mc·min^{-1}	Maximal transport and enzyme transcription rate	[56]
ω_H	948.93	mc·min^{-1}	Maximal general host protein transcription rate	[56]
ω_R	930	mc·min^{-1}	Maximal ribosomal mRNA transcription rate	[56]
ω_r	3170	mc·min^{-1}	Maximal rRNA transcription rate	$*$
$o_{\{T,E,H\}}$	4.38	mc	Host genes transcription energy threshold	[56]
$o_{\{R,r\}}$	426.87	mc	Ribosomal genes transcription energy threshold	[56],$*$
$b_{\{T,E,H,R\}}$	1	mc^{-1}·min^{-1}	mRNA-ribosome binding rate	[56]
$u_{\{T,E,H,R\}}$	1	min^{-1}	mRNA-ribosome unbinding rate	[56]
b_r	1	mc^{-1}·min^{-1}	rRNA-empty ribosome binding rate	$*$
u_r	1	min^{-1}	rRNA-empty ribosome unbinding rate	$*$
$\delta_{m_{\{T,E,H,R\}}}$	0.1	min^{-1}	Host mRNA degradation rate	[56]
δ_r	0.1	min^{-1}	rRNA degradation rate	$*$
$\delta_{p_{\{T,E,H\}}}$	0	min^{-1}	Protein degradation rate	[56]
δ_{p_R}	0	min^{-1}	Ribosome degradation rate	[56]
$n_{\{T,E,H\}}$	300	aa	Average *E. coli* protein-encoding gene length	[56]
n_R	7459	aa	Ribosomal protein component length	[56]
k_H	152219	mc	Host protein transcription threshold	[56]
h_H	4	–	Host protein transcription Hill constant	[56]
γ_{max}	1260	aa·min^{-1}·(e mc)$^{-1}$	Maximal elongation rate	[56]
k_γ	7	(e mc)	Elongation energy threshold	[56]
M	10^8	aa	Size of proteome	[56]

Table 3.2: **Host model parameters** Most model rate constants were derived in [56] by fitting to the data in [33]. We assume that the ribosomal protein complex (denoted R) has the same parameters as the ribosome as a whole in [56]. We assume that the transcriptional threshold o_r is the same as for the protein component. We assume that the rRNA has the same stability as host mRNAs (i.e. $\delta_r = \delta_{m_X}$) and that the binding/unbinding rates are diffusion limited (i.e. $b_r = u_r = 1$). We fit ω_r to the data in [33].
$*$, Fit in this study. Units: aa, amino acids, mc, number of molecules.

Parameter	Value	Units	Rational
ω_Y	Varied	mc·min^{-1}	Varied between biologically feasible values. 0 and 10^3 [56]
o_Y	4.38	mc	Assumed to be the same as the host genes
b_Y	1	mc^{-1}·min^{-1}	Assumed to strong $\approx$ 1 see Section 3.4.1
u_Y	1	min^{-1}	
n_Y	300	aa	Average length of *E. coli* protein
δ_{m_Y}	0.1	min^{-1}	Assumed to be the same as the host's mRNAs
δ_{p_Y}	0	min^{-1}	Assumed dilution only
ω_ρ	Varied	mc·min^{-1}	Varied between biologically feasible values. 0 and 10^3 [56]
o_ρ	4.38	mc	Assumed to be the same as the host genes
b_ρ	1	mc^{-1}·min^{-1}	Assumed to be the same as host rRNA
u_ρ	1	min^{-1}	Assumed to be the same as host rRNA
δ_ρ	0.1	min^{-1}	Assume same as host rRNA

Table 3.3: **Parameter values for circuit genes and orthogonal 16S rRNA production** Units: aa, amino acids, mc, number of molecules.

3.3 Simulating host responses to orthogonal ribosome usage

To assess the impact of orthogonal ribosome production on host physiology and to test the ability of our model to capture previously reported qualitative behaviour of o-ribosome producing cells, we simulated the production of o-rRNAs and the use of o-ribosomes for gene expression to a simulated environment of high nutrient efficiency and no antibiotic (Figure 3.3).

We initially consider the production of o-ribosomes on the host cell in the absence of circuit gene expression (Figure 3.3a). We simulated the production of o-rRNAs over a number of orders of magnitude. Our results recreate previous experimental results that o-ribosome production has little effect on growth rate (e.g. [70, 71], and the results of our collaborators) (Figure 3.3a, *right*). Quantifying this result we estimate that in these environmental conditions up to 20% of the ribosome pool can be co-opted with only a $\sim$ 10% fall in growth rate. Analysis of the ribosomal biosynthesis reactions (assessing rRNAs, free ribosomes etc.) shows that the 'empty ribosome' (p_R) fraction decreases significantly (Figure 3.3a, *centre*). Concurrently, host 16S rRNAs and mRNAs needed for ribosome production and assembly increase in response to ribosome sequestration by orthogonal rRNAs. The cell is able to compensate for o-ribosome production by increasing the total number of ribosomes by 20%. This allows the number of orthogonal ribosomes to rise by 40% of the original ribosome number (i.e. number of host ribosomes when $\omega_\rho = 0$ rRNAs per min) whilst the number of host ribosomes falls by only 20%.

We next simulated the expression of a simple one gene circuit which is induced at $\omega_A = 100$ mRNAs per min and utilises the *host* ribosomes for its gene expression (Figure 3.3b). At this level of expression, protein output p_A is approximately 25% of the total proteome and the growth rate is predicted to be 0.024 per hour (corresponding to a slow doubling time of approximately 80 minutes) (Figure 3.3b, *right*). This is comparable to the observations of Scott *et al.* [33]. We assess the impact of orthogonal ribosome production on this host translated gene by increasing ω_ρ over several orders of magnitude. Our model predicts a negligible fall in expression as orthogonal ribosome production increases (e.g. $< 5\%$ at $\omega_\rho = 10^3$ rRNAs per min) (Figure 3.3b, *left*). We do not see an increase in orthogonal ribosome production at low ω_ρ as induction of o-rRNA increases. Given the high level of predicted protein expression, there is a concurrent decrease in transcription rate due to its energy dependence (the $(e/(o_\rho + e))$ term in Equation 3.12). This energy usage by protein synthesis moves the o-rRNA curve to the right and also raises the production of host rRNA and empty ribosome mRNA (Figure 3.3b, *centre*). If the transcription rate is increased beyond values which are biologically feasible then we observe similar effects on o-rRNA production (Figure 3.3b, *left* inset).

We simulated the use of the orthogonal ribosome pool for circuit gene expression (Figure 3.3c). For a constitutively expressed gene (induction held at $\omega_A = 100$ mRNAs per min) we show that protein levels, p_A, increase as o-rRNA transcription increases, demonstrating that the size of the orthogonal ribosome pool acts as another 'dial' for controlling transgene expression (Figure 3.3c, *left*). The ribosomal species, including host ribosome (R) and empty ribosomes (p_R), are most sensitive to orthogonal ribosome expression and use (Figure 3.3c, *centre*). Comparison of the host proteome in response to o-ribosome production (Figure 3.3a, *left*) and use (Figure 3.3c, *left*) demonstrates that use of the orthogonal ribosome pool causes more perturbation than production alone. Analysis of the ribosome biosynthesis shows that o-ribosome production rises until $\omega_\rho = 40$ rRNAs per min before falling as overall ribosome fraction falls as the transgene p_A mass fraction significantly increases (Figure 3.3c, *centre*). Additionally we observe that total ribosome number falls as ω_ρ rises. This is due to the higher protein production ($> 33\%$ of the total proteome) brought about by high o-ribosome number. Translation acts to stabilise o-ribosomes, and therefore, prevents them from dissociating and releasing empty ribosomes (p_R). This means that empty ribosomes are not free to be converted into host ribosomes which results in a decrease in the host's translational capacity, and in turn results in less translation of ribosomal mRNAs.

In our simulations, we find that at an o-rRNA induction $\omega_\rho \approx 32$ rRNAs per min and circuit expression of $\omega_A = 100$ mRNAs per min the protein production utilising the o-ribosome is the same as when utilising the host ribosome pool (Figure 3.3c, *left*). At these equivalent levels of gene expression there is negligible change ($< 1\%$) in host protein expression ($p_T + p_E,\ p_H$) and growth rate (change $< 1\%$) (Figure 3.3c, *right*). We find significant reallocation of translational capacity with up to 25% of the ribosome pool rendered orthogonal and only negligible changes in the total number of ribosomes.

Note that these simulations have been carried out at a single antibiotic concentration $\varepsilon = 0$ and high nutrient conditions $\phi_e = 0.5$; these results vary with different growth conditions.

3.4 *In vivo* validation of an orthogonal gene expression system

Having shown in Section 3.3, that cells are likely to tolerate a significant use of the orthogonal ribosome pool for circuit gene expression our collaborators implemented a previously described o-16S rRNA system *in vivo* (see Section 3.10 for an overview of their experimental methods). This contains an o-16S rRNA under the control of P_{lac}, thus allowing its levels to be controlled by IPTG [83]. Our circuit is carried on a second plasmid and consists of RFP under the control of the P_{lux} promoter. Translation by either the host (h-RFP) or orthogonal (o-RFP) ribosome pools is controlled by selection of the RBS (Figure 3.4a). RFP mRNA production is induced with N-acyl homoserine lactone (AHL) via LuxR which is constitutively expressed from the circuit plasmid and utilises host ribosomes for its expression.

To assess the impact of o-ribosome production on the host growth rate and gene expression, gene induction was maintained using a constant concentration of AHL in the presence of increasing IPTG concentration. The production of o-ribosomes alone has no effect on growth rate demonstrating that their presence is non-toxic with growth rate being proportional to protein production regardless of the method of translation (Figure 3.4). This is constant with the phenomenological definition of growth rate in our model (Equation 3.17): Briefly, our model growth rate is derived by considering how the cell mass in amino acids (M) changes with time. This is

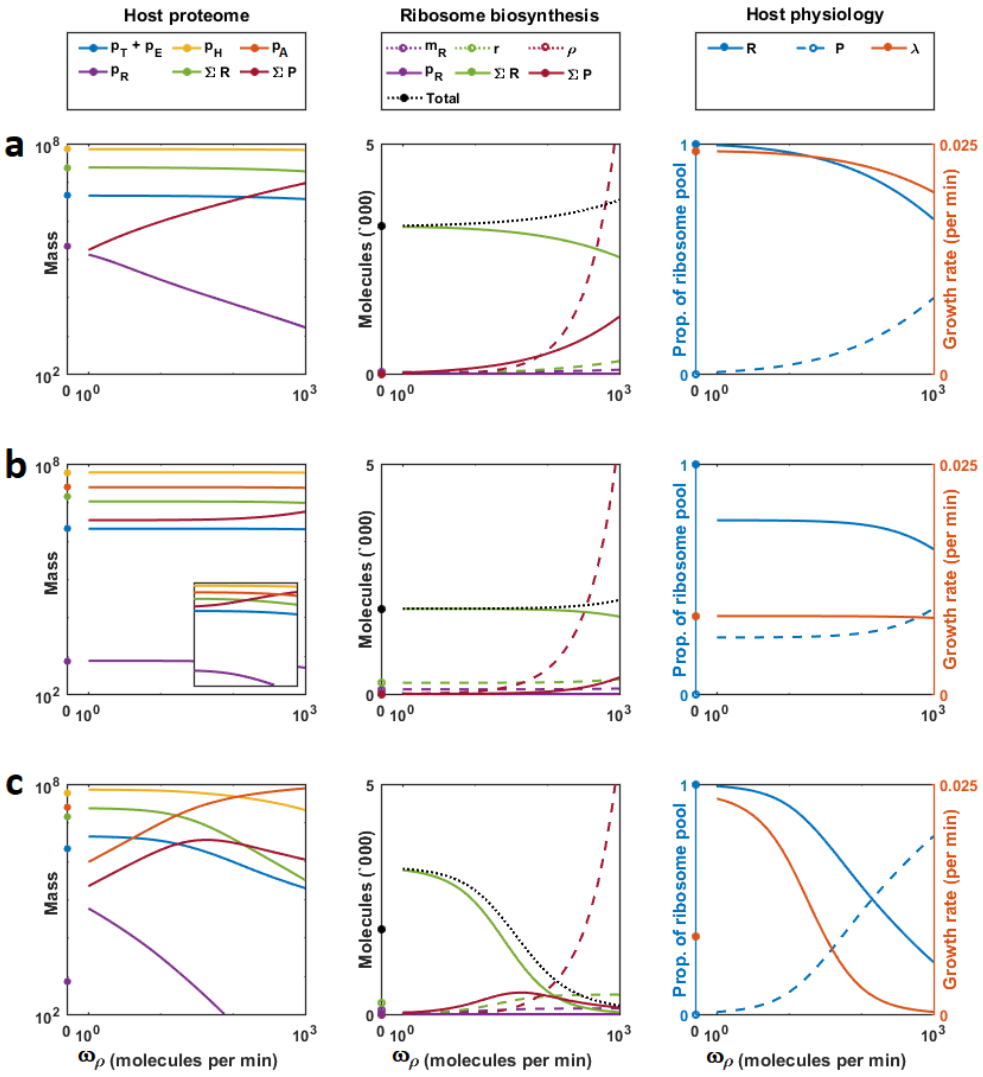

Figure 3.3: **Simulations of the impact of orthogonal ribosomes on host physiology.** The impact of orthogonal ribosome production was assessed by simulating the response to increasing o-rRNA transcription rate (ω_ρ). Single markers at 0 rRNAs per min show the results when using the host ribosome pool in the absence of orthogonal ribosome production. Host effects are assessed by observing changes in the proteome, ribosome biosynthesis and macroscopic effects such as ribosome distribution and growth rate. Legend explanation: $p_T + p_E$, metabolic enzymes; p_H, host proteins; p_A, reporter protein; p_R, 'empty' ribosome; ΣR, free and translating host ribosomes; ΣP, free and translating o-ribosomes; m_R, mRNA of the ribosomal protein; r, host 16S rRNA; ρ, o-16S rRNA; Total, sum of all ribosomes; λ, growth rate. **(a)** Impact of orthogonal ribosomes production on host genes in the absence of circuit genes (i.e. $\omega_A = 0$ mRNAs per min). **(b)** Impact of orthogonal ribosome production on a circuit gene which utilises the host ribosome pool (i.e. $\omega_A = 100$ mRNAs per min). Inset in host proteome: Simulations across an o-rRNA induction of 100x ω_ρ. **(c)** Impact of using the orthogonal ribosome pool for circuit gene expression (i.e. $\omega_A = 100$ mRNAs per min).

simply the sum of protein production minus the effect of dilution due to growth:

$$\frac{dM}{dt} = \left(\gamma \cdot \sum_{X} \left(c_X\right)\right) - \lambda M \tag{3.21}$$

where γ is the global translation rate and X is the set of translated genes (i.e. $X \in \{T,\ E,\ H,\ R,\ Y\}$). The rate of protein production depends only upon the translation rate and number of translation complexes; not on the nature of the ribosomes in those translation complexes.

Experimentally increasing the size of the o-ribosome pool with IPTG acts to dramatically increase o-RFP expression, demonstrating the utility of o-ribosomes to set a 'circuit-specific protein budget' (Figure 3.4c, inset). This is consistent with our model prediction (Figure 3.3c, *left*). However, unlike we see in our model, we do observe a 50% increase in h-RFP fluorescence ($p < 0.05$, t-test, 0 mM *vs.* 0.5 mM IPTG) (Figure 3.4b).

3.4.1 Incomplete orthogonality leads to crosstalk

Experimentally, we observe increased expression of host-translated h-RFP when the o-ribosome pool is induced. If this were due to a global increase in ribosome biosynthesis in response to o-ribosome production we are confident that this would have been captured by our model where we observe an increasing in ribosomal synthesis which does not lead to increased host expression (as captured in Section 3.3).

Instead, we propose this increase in expression is likely due to incomplete orthogonality, i.e. translation of h-RFP by o-ribosomes, leading to 'interference'. To explore this further, we introduced bi-directional crosstalk into our model, allowing o-ribosomes to translate mRNAs with host RBSs and vice versa. To incorporate crosstalk into our model we allow mRNAs to bind to their non-target ribosome at rate β to produce the translation complex c'. The unbinding rate is μ, which we set at 1 throughout. The translation complexes c' follow the same dynamics described in Equation 3.5.

The dynamics of host-translated mRNAs m_X become (with the crosstalk modifica-

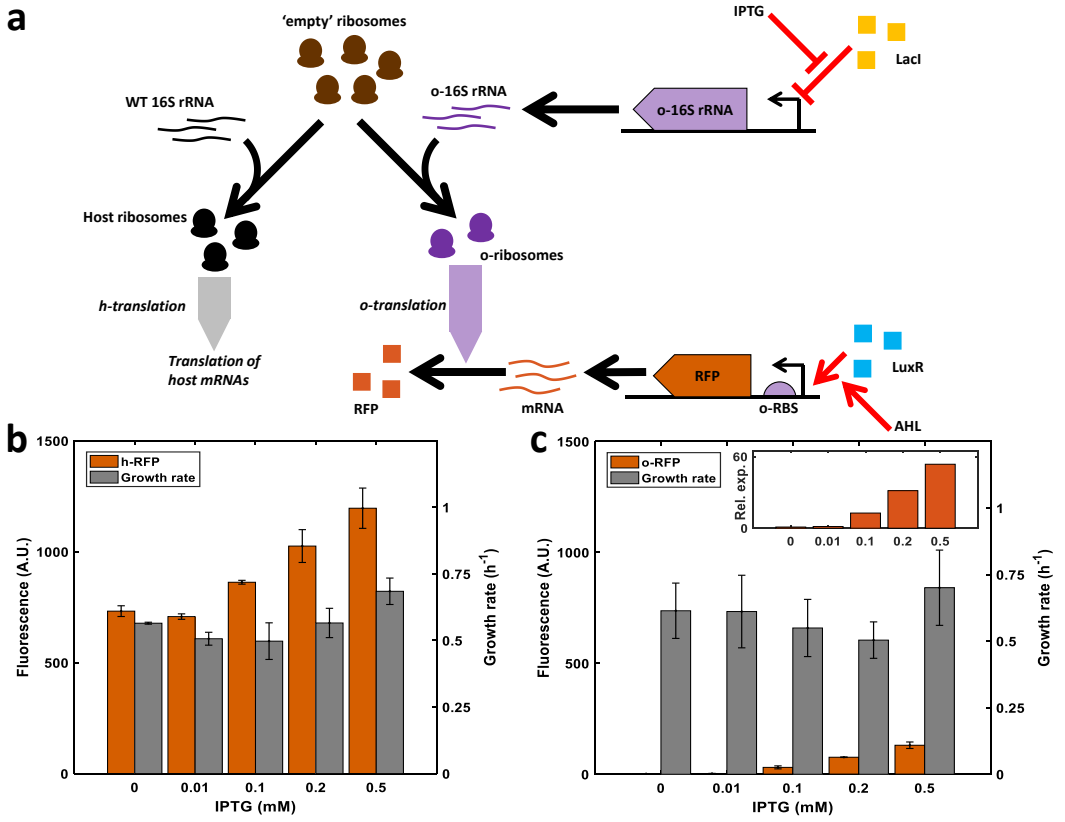

Figure 3.4: **Developing the o-ribosome pool as an expression system.** RFP expression using either the host or o-ribosome pool. o-ribosomes are produced by inducing o-16S rRNA with increasing concentrations of IPTG. (o/h)-RFP is induced with 20 nM AHL throughout. Bars represent means ± 1 S.D. $N = 3$. **(a)** Schematic of the orthogonal translation system. **(b)** Expression of h-RFP by the host ribosome pool. **(c)** Expression of o-RFP by the o-ribosome pool. Inset, o-RFP expression relative to the value at 0 mM IPTG. x-axis, IPTG concentration (mM).

tions shown in blue):

$$\begin{aligned}\frac{dm_X}{dt} = \omega_X \cdot \mathcal{R} \cdot \left(\frac{e}{o_X + e}\right) \ldots \\ + T_L(c_X, e) - b_X \cdot R \cdot m_X + u_X \cdot c_X \ \ldots \\ + T_L(c'_X, e) - \beta \cdot P \cdot m_X + \mu \cdot c'_X \ \ldots \\ + (\delta_{m_X} + \lambda) \cdot m_X\end{aligned} \tag{3.22}$$

The dynamics of the proteins p_X becomes:

$$\frac{dp_X}{dt} = T_L(c_X, e) + T_L(c'_X, e) - (\delta_{p_X} + \lambda) \cdot p_X \tag{3.23}$$

We also modified the energy (Equation 3.15) and growth rate (Equation 3.17) equations to include the energy consumption and translation rates of the new complexes, c'. We included the additional mRNA-ribosome interactions by updating the equations describing the free ribosome dynamics (Equation 3.16, 3.14) with the modifications shown in Equation 3.22 with signs inverted.

Note that experimentally, we do not observe translation of o-RFP by host ribosomes demonstrating that this appears to be a one-way interference (*data not shown*) and so we set β to zero for host genes.

To assess the impact of circuit demand on crosstalk we simulated the production of a single protein which utilises the host ribosome pool for its expression while varying β to control the propensity of o-ribosomes to translate host and host-translated circuit mRNAs. We varied the circuit demand by changing both the translation rate (ω_A) and RBS strength (b_A). Our simulations show that in the presence of a low demand circuit, crosstalk can have a significant effect (Figure 3.5a). When $\beta = b_A = 0.1$, i.e. there is no distinction between the two ribosome pools, the protein production increases fourfold at maximum o-rRNA transcription ($\omega_\rho = 10^3$ rRNAs per min). As o-ribosome production is increased by increasing ω_ρ there is a net increase in ribosome number (as seen in Figure 3.3b). These results suggest that in our hands the h-RFP circuit represents a 'low demand' circuit and so crosstalk results in a clear effect (i.e. the increased h-RFP expression as o-ribosomes are produced, Figure 3.4c).

Increasing the circuit demand by increasing both ω_A and b_A to their maximum feasible values effectively removes the impact of crosstalk with protein levels falling negligibly (Figure 3.5b). Comparing our high demand simulations to experimental investigations of a high demand circuit where two genes are expressed from a high

copy plasmid (the h-RFP, h-GFP circuit of Section 3.5) shows that our gene coupling assessments are carried out in a context of high enough competition which allows crosstalk to be considered negligible throughout (Figure 3.5d).

These simulations suggest that the rise in h-RFP expression in response to o-ribosome induction is likely due to an 'interference' as o-ribosomes have a low propensity to translate the h-RFP. As host ribosomes are co-opted for o-ribosome production, the total ribosome pool increases (as discussed in Section 3.3) increasing translation of the RFP. Increasing the ability of circuit genes to sequester h-ribosomes, for example by increasing RBS strength and expression level, abolishes the impact of this interference as ribosome recycling (and hence o-ribosome formation) decreases as host ribosomes are stabilised during translation of the h-RFP. This is more similar to the experimental conditions in the following section.

3.4.2 Inefficient o-ribosome assembly explains poor o-RFP expression

Our simulations predicted that for a given gene induction, utilisation of the orthogonal ribosome pool results in higher protein production due to the lack of competition (if ω_ρ is tuned appropriately) (Figure 3.3c). However, experimentally we find significantly reduced expression with expression of o-RFP being on the order of ten-fold smaller than that of the h-RFP even at the maximum IPTG induction (Figure 3.4b).

We propose that this may be due to inefficient o-rRNA production or inefficient o-ribosome assembly resulting in a smaller number of available ribosomes or due to the difference in strengths of the o-RBS.

To assess these effects we simulated the production of a single protein first using the host ribosome (simulating an induction of $\omega_A = 100$ mRNAs per min and a maximal RBS strength $b_A = 1$) to set the host value. We then simulated the translation of the gene by orthogonal ribosomes and normalised by this host value so that 1 represents equal production. Values greater than 1 indicate higher production than the host case and values less than 1 indicate poorer production in the orthogonal case.

We varied the two parameters governing o-ribosome production (ω_ρ and b_ρ) and ribosome-mRNA binding (b_A). We leave the unbinding rates as maximal, i.e. $u_A = 1$ and $u_\rho = 1$.

As we have previously shown increasing, rRNA transcription rate (ω_ρ) increases protein production. Even at middling production values of $\omega_\rho \approx 100$ mRNA per

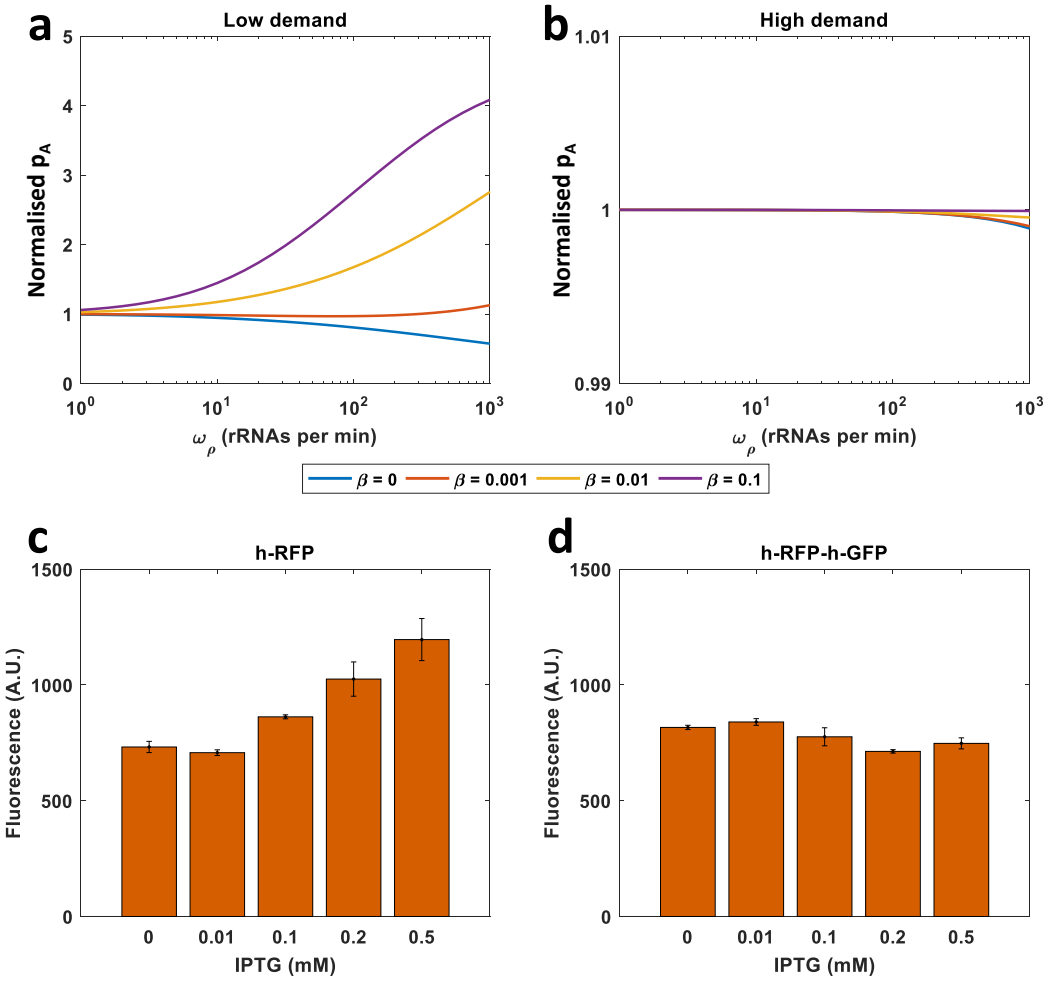

Figure 3.5: **Crosstalk has little effect in circuits with high demand.** Assessment of the impact of crosstalk on circuits with low and high demand. β indicates the level of crosstalk. **(a)** Simulations of a low demand circuit. A single gene is expressed utilising the host ribosome pool with low transcription rate $\omega_{RFP} = 10$ mRNAs per min and weak RBS, $b_{RFP} = 0.1$. **(b)** Simulations of a high demand circuit. A single gene is expressed using the host ribosome pool with high transcription rate, $\omega_{RFP} = 1000$ mRNAs per min and strong RBS, $b_{RFP} = 1$. **(c)** Experimental low demand circuit. Steady state mean RFP expression $\pm$ 1 S.D. using the host ribosome pool. $N = 3$. **(d)** Experimental high demand circuit. Steady state mean $\pm$ 1 S.D. h-RFP from the h-RFP-h-GFP circuit. $N = 3$.

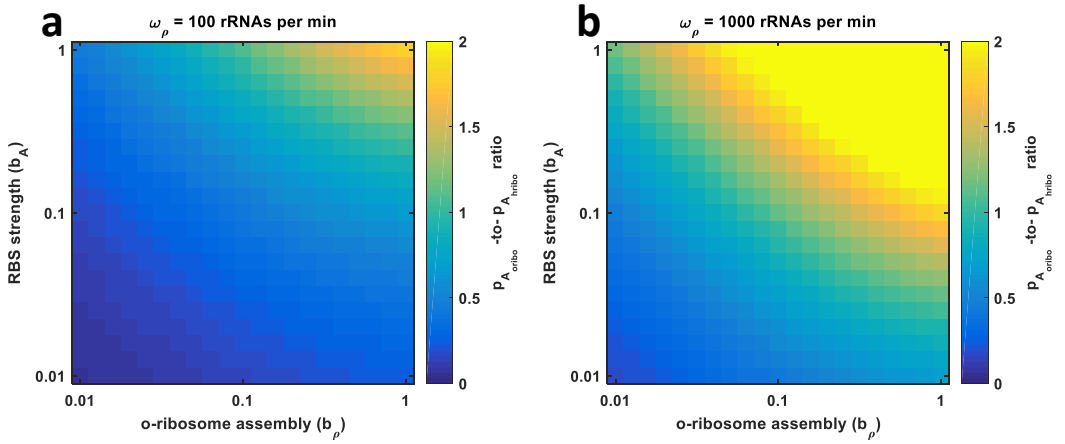

Figure 3.6: **Orthogonal ribosome assembly may be inefficient.** Simulations of the steady state protein outputs for different RBS strengths (b_A-to-u_A ratio) and assembly of the orthogonal ribosome (b_ρ-to-u_ρ ratio) for the given orthogonal ribosome induction (ω_ρ, o-rRNAs per min). Circuit gene induction is constant at $\omega_A = 100$ mRNAs per min throughout. Results are reported as a ratio of the production utilising the host ribosome pool for the same circuit, with values greater than 2 truncated to 2. **(a)** Medium o-rRNA induction of 100 rRNAs per min. **(b)** High o-rRNA induction of 1000 rRNAs per min.

min we see increased production of the output protein (at $b_\rho = 1$ and $b_A = 1$) (Figure 3.6a). Varying each of the other parameters alone is sufficient to cause a decrease in gene expression to less than 40% of the production achieved using the host pool. If both b_A and b_ρ are significantly weaker than those of the host ($b_A < 0.1$ and $b_\rho < 0.1$) then the ratio of the protein production using the orthogonal to the host pool falls to ranges we see experimentally. This is more stark at high rRNA induction (Figure 3.6b). Therefore, we propose that our experimental observations may be due to a combination of inefficient translation (i.e. the orthogonal RBS is weak) or inefficient orthogonal ribosome assembly. See Chapter 6 for further discussion of improving the experimental system.

3.5 Gene coupling in circuits with access to only one ribosome pool

Here we consider a new circuit consisting of two genes; the original RFP cassette and a new GFP cassette. GFP transcription was constitutively driven from the P_{tet} promoter and the host or o-ribosome pool utilised for translation, controlled by se-

lection of RBS as described above. We determined the level of coupling between the two circuit genes by observing the slope of the isocost line of circuit gene expression during exponential growth. As discussed in Section 2.3.2, the isocost line quantifies the change in GFP as RFP is induced (Figure 3.7).

Gene coupling in the h-RFP, h-GFP circuit (utilising the host ribosome pool) results in a slope of -3.3; for every unit of RFP gained, $\sim$3 units of GFP are lost (Figure 3.7c). In this case the isocost line is non-linear at maximum RFP expression; in the following analysis we neglect this small non-linearity in favour of fitting a straight line through all the points. This is consistent with the isocost line theory discussed in Section 2.3.2 and with previous work in this area (such as [38, 44]). Tuning the o-ribosome pool when utilising the host ribosomes as the translational resource has no effect on the coupling observed, consistent with our model predictions (Figure 3.7b). Replacing the host RBS sequences with the o-RBS to produce the o-RFP, o-GFP circuit and utilisation of the orthogonal translation system results in the coupling being reduced to 30% of that observed when using the host pool (Figure 3.7f). As predicted by our model, increasing IPTG increases gene expression (due to increasing the size of the o-ribosome pool) but has negligible effect on coupling (as the circuit genes compete equally for o-ribosomes), (Figure 3.7e).

3.6 Utilising both host and orthogonal ribosome pools reduces coupling of co-expressed genes

Having successfully partitioned the ribosome pool both theoretically and experimentally, we tested the ability of these pools to act as a simple distribution mechanism for translational capacity. Maintaining the original circuit topology and function, we altered the RBS of each gene to create two new circuit variants; o-RFP, h-GFP and h-RFP, o-GFP. Our model predicts that placing the constitutively expressed GFP under control of the o-ribosome pool (h-RFP, o-GFP arrangement) acts to insulate the gene from competition with RFP and so significantly reduces gene coupling, over a range of o-ribosome pool sizes (Figure 3.8b). Experimental validation of these predictions showed near complete abolition of the isocost line slope with coupling falling by over 95% (Figure 3.8c). Varying IPTG levels shows this decoupling is highly robust, with IPTG acting only to tune expression consistent with our model prediction (Figure 3.8b, c).

The inverse arrangement, where the constitutively expressed GFP utilises the host

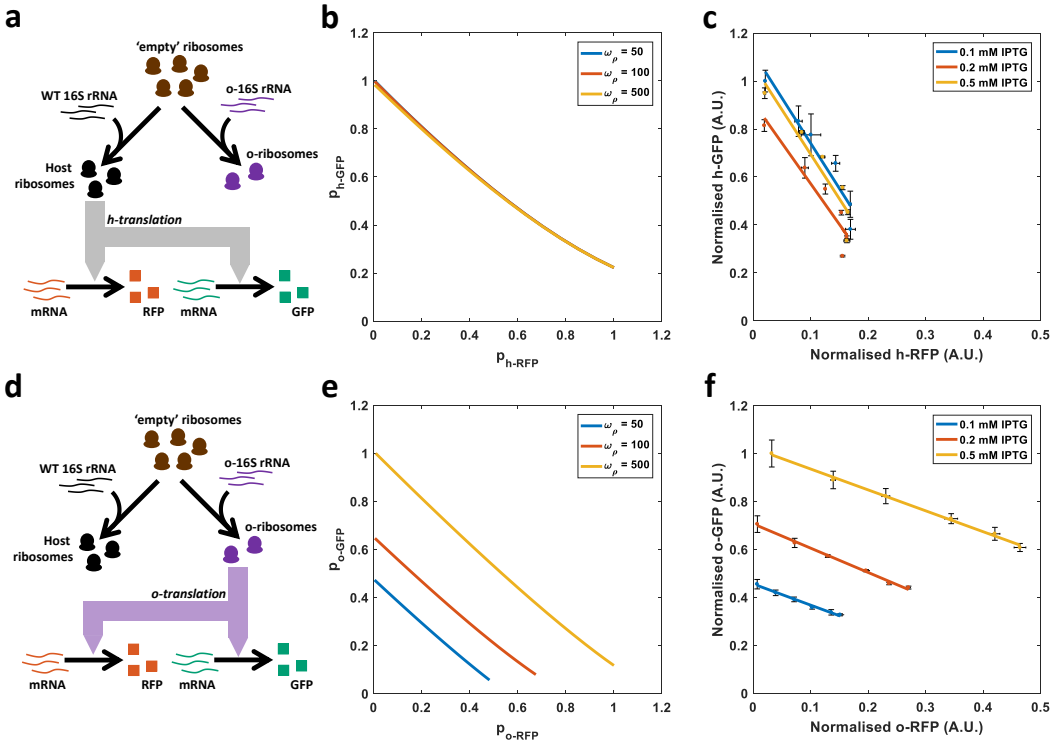

Figure 3.7: **Gene coupling in gene circuits utilising either the host or orthogonal ribosome pool.** Simulations of the steady state concentrations of RFP and GFP normalised by the maximum protein production achieved across the o-rRNA transcription rates tested. $\omega_{GFP} = 100$ and $\omega_{RFP} = 1$ to 10^3 mRNAs per min. o-rRNA production (ω_ρ) was simulated at the RNAs per min as shown. Experimental data were produced by inducing RFP using AHL from 0 to 20 nM. Points are the mean steady state fluorescence $\pm$ 1 S.D normalised by maximum GFP expression obtained across different levels of IPTG treatment. $N = 3$. The isocost line is fit to the mean fluorescence as determined by FACS from cultures during mid-exponential growth (between 3-5 hours post-induction dependent on the strain and circuit). o-ribosome production was induced using three different IPTG concentrations 0.1, 0.2 and 0.5 mM as shown. **(a)** Allocation of ribosomes in panels (b) and (c) where both circuit genes share the host ribosome pool. **(b)** Simulations of circuit using the host ribosome pool. **(c)** *In vivo* protein productions of circuit proteins using the host ribosome pool. **(d)** Allocation of ribosomes in panels (e) and (f) where both circuit genes share the o-ribosome pool. **(e)** Simulations of the circuit using the o-ribosome pool for translation. **(f)** *In vivo* coupling observed in the circuit using the o-ribosome pool.

ribosome pool and the induced RFP utilises the o-ribosome pool (the o-RFP, h-GFP circuit) is predicted by our model to show increased gene coupling (Figure 3.8e). In the experimental system the isocost line gradient increases by over six times in comparison to coupling in circuits using the host ribosomes (Figure 3.8f).

Given the increase in coupling observed in the o-RFP, h-GFP arrangement, we assessed the change in protein components theoretically (Figure 3.9). As the induction of o-RFP is increased (simulated by increasing ω_{RFP}), it forms translation complexes with the o-ribosomes. This stabilises their formation, preventing release of 'empty' ribosomes, p_R. We observe marked declines in empty ribosomes, i.e. there is high competition for the protein core which is not specified until bound by an rRNA. This results in a concurrent fall in host ribosome number (both free, R and translation complexes $h\text{-}c_X$). This results in significant perturbation of host protein levels when the production of o-ribosomes is high (Figure 3.9c). This in turn results in a concurrent fall in expression of the host ribosome-utilising GFP by over 80% at medium levels of RFP induction ($\omega_{RFP} = 100$, $\omega_\rho = 500$ RNAs per min). Therefore we propose the increased coupling we observe experimentally is due to competition for ribosomal proteins or other accessory translation factors. In the h-RFP, o-GFP arrangement this is not seen as the o-GFP is constitutively expressed and so the demand for o-ribosomes does not increase as RFP is increased.

We extended the approach of using multiple ribosome pools *in silico* by simulating the use of one o-ribosome pool for each circuit gene (Figure 3.10). Biologically these could be created by expressing multiple different synthetic 16S rRNAs [70]. We implement a second orthogonal ribosome pool by replicating the changes outlined in Section 3.2. At high levels of o-rRNA expression, competition for empty ribosomes (p_R) results in high levels of gene coupling. We optimised the production of the two orthogonal 16S rRNAs to minimise gene coupling (resulting in o-rRNA production rate of $\omega_{\rho_1} = \omega_{\rho_2} = 10$ molecules per min) which yields near complete decoupling with p_B falling less than 10% as p_A is induced. However, the induced p_A still shows a saturating response to increasing induction due to the finite o-ribosome pool size (Figure 3.10a).

However this assumes that there is no crosstalk between the two orthogonal ribosome pools - i.e. that each pool translates its, *and only its*, target mRNAs. In reality, to date, orthogonal rRNA sequences are often not sufficiently distinct to achieve this and each orthogonal pool will translate mRNAs targeted for the other pool due to crosstalk - i.e. erroneous interactions between one 16S rRNA and its non-target mRNA [70]. To test the impact of crosstalk we introduce interactions between

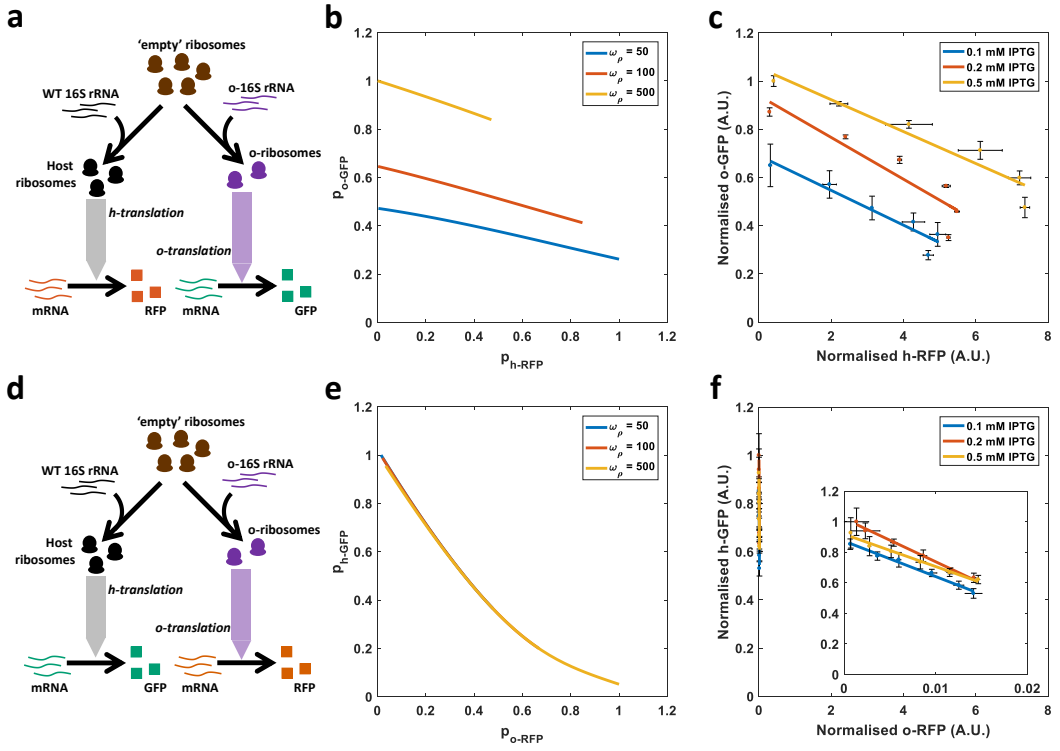

Figure 3.8: **Gene coupling in circuits utilising different ribosome pools.** Simulations of the steady state concentrations of RFP and GFP normalised by the maximum protein production achieved across the o-rRNA transcription rates tested. $\omega_{GFP} = 100$ and $\omega_{RFP} = 1$ to 10^3 mRNAs per min. o-rRNA production (ω_ρ) was simulated at the rates shown. Experimental data were produced by inducing RFP using AHL from 0 to 20 nM. Points are mean steady state fluorescence $\pm$ 1 S.D. normalised by maximum GFP expression obtained across different levels of IPTG treatment. $N = 3$. The isocost line is fit to the mean fluorescence as determined by FACS from cultures during mid-exponential growth (between 3-5 hours post-induction dependent on the strain and circuit). o-ribosome production was induced using three different IPTG concentrations 0.1, 0.2 and 0.5 mM as shown. **(a)** Allocation of ribosomes in panels (b) and (c). **(b)** Simulations of the circuit using host ribosomes for RFP expression and o-ribosome for GFP expression. **(c)** *In vivo* coupling of the h-RFP-o-GFP circuit. **(d)** Allocation of ribosomes in panels e and f. **(e)** Simulations of the circuit using the o-ribosome for RFP expression and host pool for GFP. **(f)** *In vivo* coupling of the o-RFP-h-GFP circuit. The inset shows the data on an expanded x-axis.

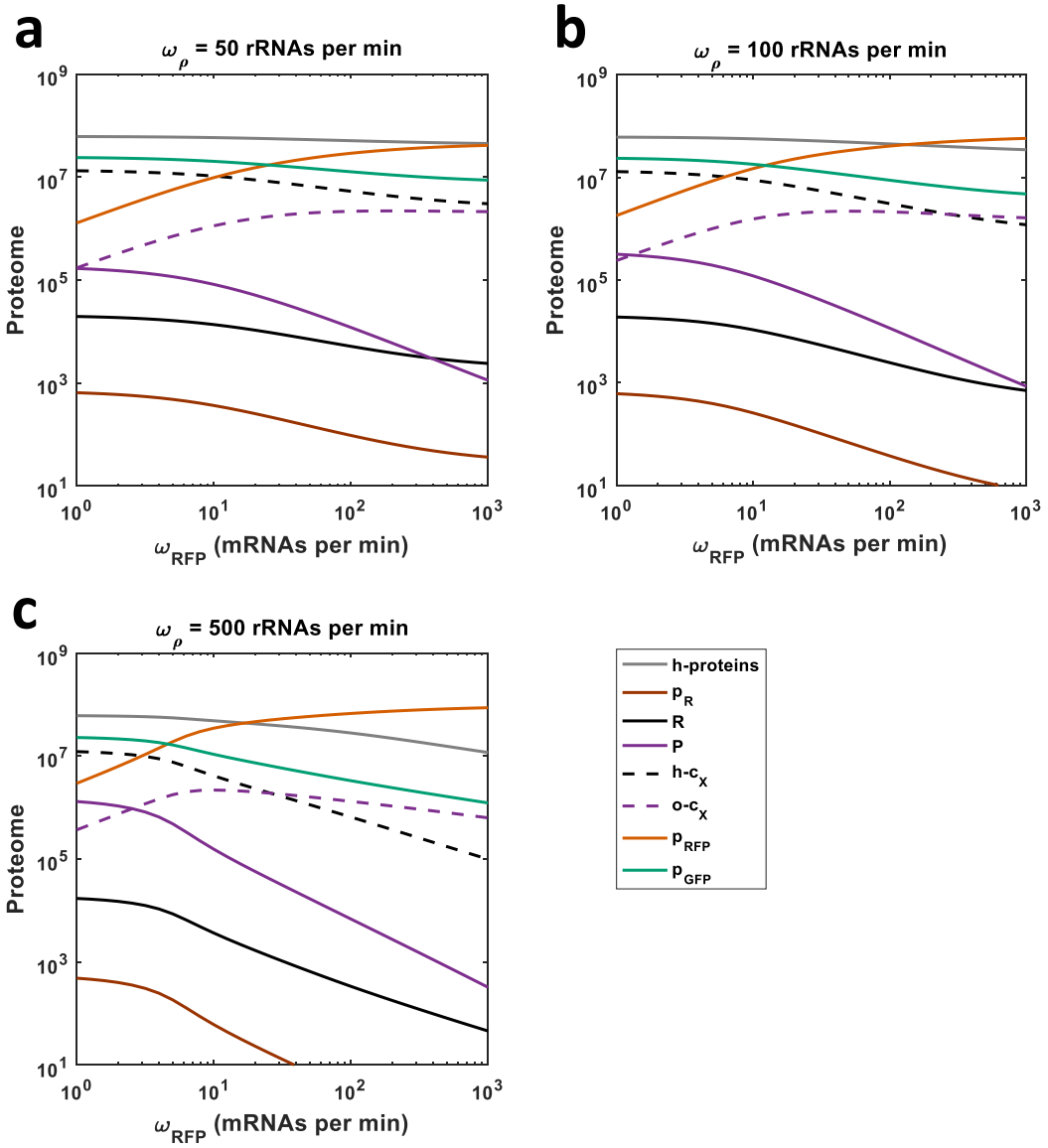

Figure 3.9: **Competition for 'empty' ribosomes explains the increased coupling in o-RFP, h-GFP circuits.** Simulations of the changing distribution of proteome components as ω_{RFP} is increased. The o-rRNA transcription rate is varied as shown. Note the y-axis is a log scale causing changes in GFP and RFP to not be as apparent. Legend explanation: h-proteins, host proteins; p_R, empty non-specific ribosome; R, host ribosomes; P, orthogonal ribosomes; $h\text{-}c_X$, translating host ribosomes; $o\text{-}c_X$, translating o-ribosomes; p_{RFP}, RFP protein; p_{GFP}, GFP protein. **(a)** Low o-rRNA induction. **(b)** Medium o-rRNA induction. **(c)** High o-rRNA induction.

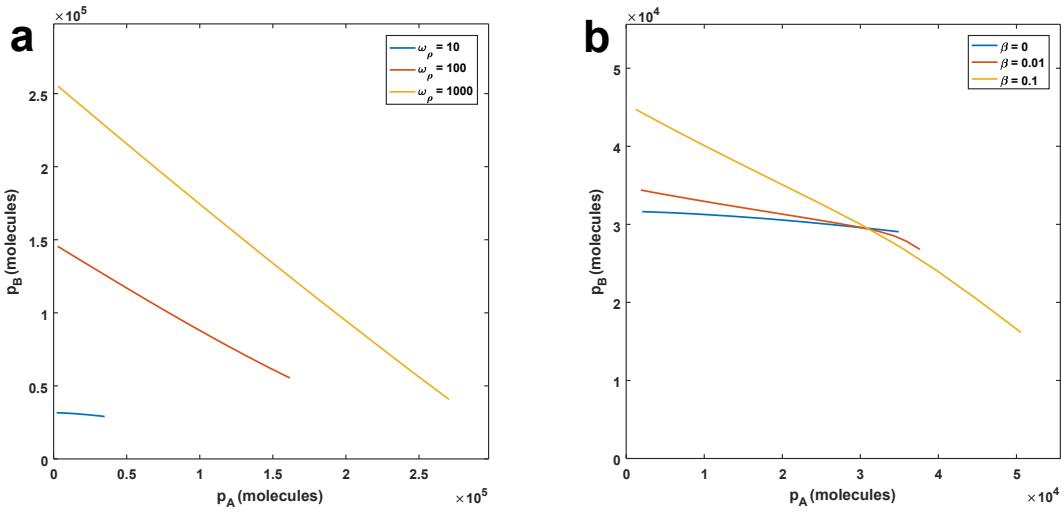

Figure 3.10: **Gene coupling in gene circuits utilising multiple orthogonal ribosome pools.** Simulations of the steady state protein outputs of a simple two-gene circuit where each gene utilises its own orthogonal ribosome pool. **(a)** Isocost lines at different orthogonal ribosome production rates (ω_ρ). **(b)** Simulations of the optimal o-rRNA production rates ($\omega_\rho = 10$ molecules per min) with the introduction of cross talk (β).

each mRNA and its non-target o-ribosome, as in Section 3.4.1 (but we neglect host ribosome, circuit mRNA crosstalk as this is a high expression condition). Messenger RNAs bind to their non-target o-ribosome at rate β to produce the translation complex c_1'. The unbinding rate is μ. The dynamics of an mRNA m_1 which is designed to be translated by P_1 and erroneously translated by P_2 (the effect of crosstalk is highlighted in blue) are:

$$\begin{aligned}\frac{dm_1}{dt} = \omega_1 \cdot \left(\frac{e}{o_1 + e}\right) + T_L(c_1, e) - b_1 \cdot P_1 \cdot m_1 + u_1 \cdot c_1 ... \\ + T_L(c_1', e) - \beta_1 \cdot P_2 \cdot m_1 + \mu_1 \cdot c_1' - (\delta_{m_1} + \lambda) \cdot m_1\end{aligned} \tag{3.24}$$

We vary the level of crosstalk between none ($\beta = 0$) and high ($\beta = 0.1$, approximately 10% of the interaction between the cognate pair). Coupling increases seven fold as crosstalk increases from 0 to 0.1. As crosstalk increases, gene coupling (as measured by the slope of the isocost line) also increases (Figure 3.10b). Whilst our simulations suggest that some level of crosstalk can be tolerated, this must be relatively small ($\beta \leq 0.01$), and may not be achievable with currently available o-ribosomes.

3.7 Simulations of coupling in circuits using tethered ribosomes

As highlighted previously, ribosomes dissociate upon termination of translation into their small and large subunits. This creates competition for the large subunit between the host and orthogonal small subunits and is modelled in our system as competition for the protein component p_R. These two subunits can be permanently linked *in vivo* by either (i) expressing a synthetic RNA containing both the (orthogonal) 16S and 23S rRNA or (ii) by expression of an additional synthetic RNA which binds to both ribosomal subunits and acts as a 'molecular tether' [72, 73]. In these semi-synthetic ribosomes the two subunits are permanently linked and so are not able to dissociate. We extend our study by considering the use of such tethered ribosomes *in silico*

Ribosome tethering can be modelled by assuming that the o-rRNA and protein components could not dissociate upon creation of the functional o-ribosome, i.e. we set u_ρ to 0. We set $b_\rho = 0.45$ to take into account the experimental observation that tethered ribosomes show reduced expression in comparison to host ribosomes [72]. We model the same circuits as before; simulating the constitutive expression of a protein (p_B) while inducing a second p_A. The expression of tethered ribosomes results in a significant decrease in expression of circuits using the host ribosome pool due to the lack of recycling of the ribosomal proteins p_R (Figure 3.11a). Utilising the tethered o-ribosome pool for gene expression significantly increases gene expression from the circuit (Figure 3.11b). Simulating a circuit when the two circuit genes are distributed across both host and tethered o-ribosome pools replicates results observed with orthogonal ribosomes. When the constitutively expressed gene is translated by the tethered o-ribosome pool and the induced gene by the host pool we observe complete decoupling and increased gene expression at all levels of the o-ribosome pool (Figure 3.11c). The opposite arrangement significantly increases coupling (Figure 3.11d).

3.8 Distribution of translational resources increases production of violacein

We have shown that by targeting the distribution of translational resources we can successfully reduce gene-gene couplings in simple circuits composed of fluorescent

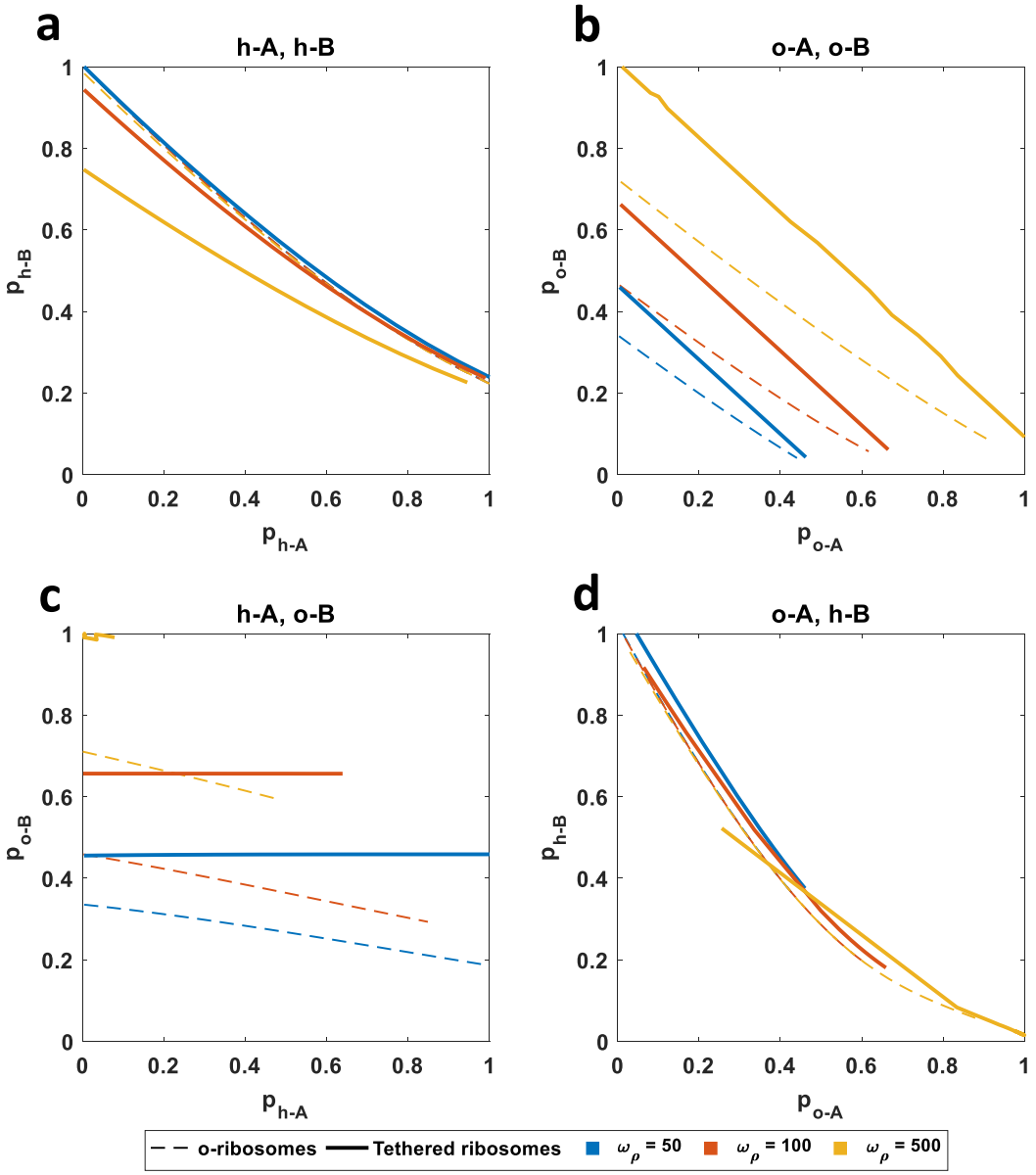

Figure 3.11: **Tethered o-ribosomes may increase yields and increase decoupling in some contexts.** Simulations of the steady state concentrations of p_A and p_B normalised by the maximum protein production achieved across the o-rRNA transcription rates tested where translation is carried out by tethered orthogonal ribosomes (solid line, $b_\rho = 0.45$ and $u_\rho = 0$). For comparison simulations of non-tethered orthogonal ribosomes are also shown (dashed lines, $b_\rho = u_\rho = 1$). $\omega_A = 1$ to 10^3 mRNAs per min while $\omega_B = 100$ mRNAs per min. o-rRNA production (ω_ρ) was simulated at the RNAs per min as shown. **(a)** Both genes translated by the host ribosome pool. **(b)** Both genes translated by the tethered o-ribosome pool. **(c)** p_B translated by the tethered o-ribosome pool while p_A is translated by the host ribosome pool. **(d)** p_A is translated by the tethered o-ribosome pool while p_B is translated by the host ribosome pool.

reporter genes. These genes do not have useful biological function beyond allowing visualisation. To demonstrate the utility of manipulating resource allocation in a broader biotechnological context, we consider the production of a metabolite from a multienzyme pathway.

As an example pathway we selected the well characterised *vio* pathway from *Chromobacterium violaceum* [84]. This five enzyme pathway produces violacein from L-tryptophan. Violacein has been shown to have anticancer and antibacterial properties, and has previously proved difficult to produce [84]. Additionally, the second enzyme in the pathway has previously been shown to induce a high cellular burden due to its large size and consequently large ribosome sequestration ability [37]. Due to violacein's purple colour its production can be easily tracked experimentally.

To demonstrate how resource-mediated competition can impact pathway function, we incorporated the violacein pathway into our model. This highlights one of the advantages of this 'abstraction level' of modelling: that we are able to concurrently (and easily) consider gene expression and metabolism.

We consider a simple five-part linear biochemical pathway which converts the intracellular 'energy' metabolite, e, to an output substrate, s_5, via four intermediates (Equation 3.25).

$$e \xrightarrow{v_1,\ k_1} s_1 \xrightarrow{v_2,\ k_2} s_2 \xrightarrow{v_3,\ k_3} s_3 \xrightarrow{v_4,\ k_4} s_4 \xrightarrow{v_5,\ k_5} s_5 \tag{3.25}$$

We model the conversion of each metabolite, s_i, into the next as an enzyme catalysed reaction by protein p_{i+1} with the reaction dynamics described by Michaelis-Menten kinetics.

The dynamics of the first metabolite are:

$$\frac{ds_1}{dt} = \frac{v_1 \cdot e \cdot p_1}{k_1 + e} - \frac{v_2 \cdot s_1 \cdot p_2}{k_2 + s_1} - \lambda \cdot s_1 \tag{3.26}$$

The dynamics of the intermediate metabolites are (where $i = \{2...4\}$):

$$\frac{ds_i}{dt} = \frac{v_i \cdot s_{i-1} \cdot p_i}{k_i + s_{i-1}} - \frac{v_{i+1} \cdot s_i \cdot p_{i+1}}{k_{i+1} + s_i} - \lambda \cdot s_i \tag{3.27}$$

The dynamics of the final metabolite, the output, are defined by:

$$\frac{ds_5}{dt} = \frac{v_5 \cdot s_4 \cdot p_5}{k_5 + s_e} - \lambda \cdot s_5 \tag{3.28}$$

We also modify Equation 3.15 to account for the conversion of e into s_1 (highlighted in blue). This represents the drain on the host metabolic flux of the new pathway. φ_s represents the number of molecules of e required to make one molecule of s_1.

$$\frac{de}{dt} = \varphi_e \cdot \frac{v_E \cdot p_E \cdot s_i}{k_E + s_i} - \sum_{X \in \{T,E,H,R,Y\}} \Big(n_X \cdot T_L(c_X, e) \Big) - \lambda \cdot e \; -\varphi_s \cdot \frac{v_1 \cdot e \cdot p_1}{k_1 + p_1} \quad (3.29)$$

We assumed that pathway enzymes have the same kinetics as the host's lumped enzyme function ($v_i = 5800$ molecules per min and $k_i = 1000$ molecules). We set φ_s to be 0.01 to model the impact of the additional flux through the pathway on the central metabolism. We also increase the nutrient efficiency φ_e to 1 to model the additional media supplementation as the amino acid tryptophan is added to the medium in the experiments. We take the enzyme sizes, in amino acids, for the components of the violacein producing pathway from UniProt: $n_1 = 418, n_2 = 998, n_3 = 429, n_4 = 373, n_5 = 196$. Other parameters were set as found in Table 3.3. Note that for simplicity of language and to allow the same nomenclature with the experimental implementation of the violacein pathway, the enzymes are referred to by their corresponding letter below, such that $i = 1$ becomes A, 2 becomes B etc.

We model the production of the enzymes as we modelled the reporter genes before. Whilst, experimentally, *vioBCDE* is expressed as an operon from one promoter, for simplicity we model these genes as four separate mRNAs (i.e. one mRNA for *vioB*, one for *vioC* etc) to maintain the number of RBSs and hence competition. We ensure the parameters representing the mRNA birth-death processes are equal (i.e. ω_i, o_i and δ_{m_i} for $i = \{B,\ C,\ D,\ E\}$) to model the fact these genes are co-regulated.

Experimentally, expression of the pathway enzymes was divided between two operons with the first enzyme of the pathway *vioA* placed under the control of P_{tet} and the other genes *vioB, C, D* and E in one polycistonic operon under the control of P_{lux}, making expression inducible with AHL. For simplicity, we refer to this as the downstream cassette. In this arrangement, the flow of mass through the metabolic pathway mimics the information flow in an activation cascade which has previously been shown to be highly sensitive to the effects of resource competition (as discussed in Chapter 2 and observed in [45]).

In the absence of resource competition, it would be expected that increasing AHL would increase the pathway flux due to increased expression of the downstream pathway enzymes. However, simulation of our model shows that when a single pool of ribosomes is used for the expression of both cassettes competition emerges

(illustrated in Figure 3.12a). As the downstream cassette is induced, the first constitutively expressed enzyme falls (Figure 3.12c). This results in a concurrent decrease in metabolite production. We have shown above that the expression of a constitutive gene can be maintained if this gene is targeted to the orthogonal ribosome pool. We apply this observation to the pathway by allocating the first enzyme to the o-ribosome pool (illustrated in Figure 3.12b). The decrease of the first enzyme in response to the induction of the downstream cassette is significantly reduced in comparison to using only one ribosome pool (3.12d). This change in enzyme distribution results in a concurrent increase in metabolite production across the ranges of induction simulated. It should be noted that metabolite production does begin to plateau at high induction. Our simulations predict that the use of this resource allocation scheme can improve final metabolite production significantly.

To validate these predictions, our collaborators implemented the *vioABCDE* pathway in *E. coli.* as described in Section 3.10. Cell growth and violacein production were monitored for the cases where (i) *vioA* is produced from the h-ribosome pool and (ii) *vioA* is produced form the o-ribosome pool. When using the host ribosome pool for expression of all of the pathway components, we see a peak of production of violacein at intermediate concentrations of AHL (Figure 3.12e). Both low and high concentrations of AHL yielding poor production, as expected from the predicted couplings derived from ribosomal competition (Figure 3.12e). Insulating vioA by using the orthogonal ribosome system allows for higher production yields per cell that increase monotonically with AHL concentration (Figure 3.12e). Despite these higher yields no significant decrease in growth rate was observed. Moreover, there was no emergence of mutants which do not produce violacein in either of the strains tested, suggesting the decrease in production yields is due to the competition for translational resources between different segments of the pathway rather than evolution to reduce 'burden'.

3.9 Conclusions

In this chapter, we have developed and validated a simple model of microbial growth which takes account of multiple host factors, ribosome biosynthesis and the formation of orthogonal ribosomes. We demonstrated with our model that translation by o-ribosomes is tolerable by cells but upon validation found significantly reduced expression. Using our model we proposed that this is likely due to inefficient formation of orthogonal translation complexes. We show that despite this, our simple model

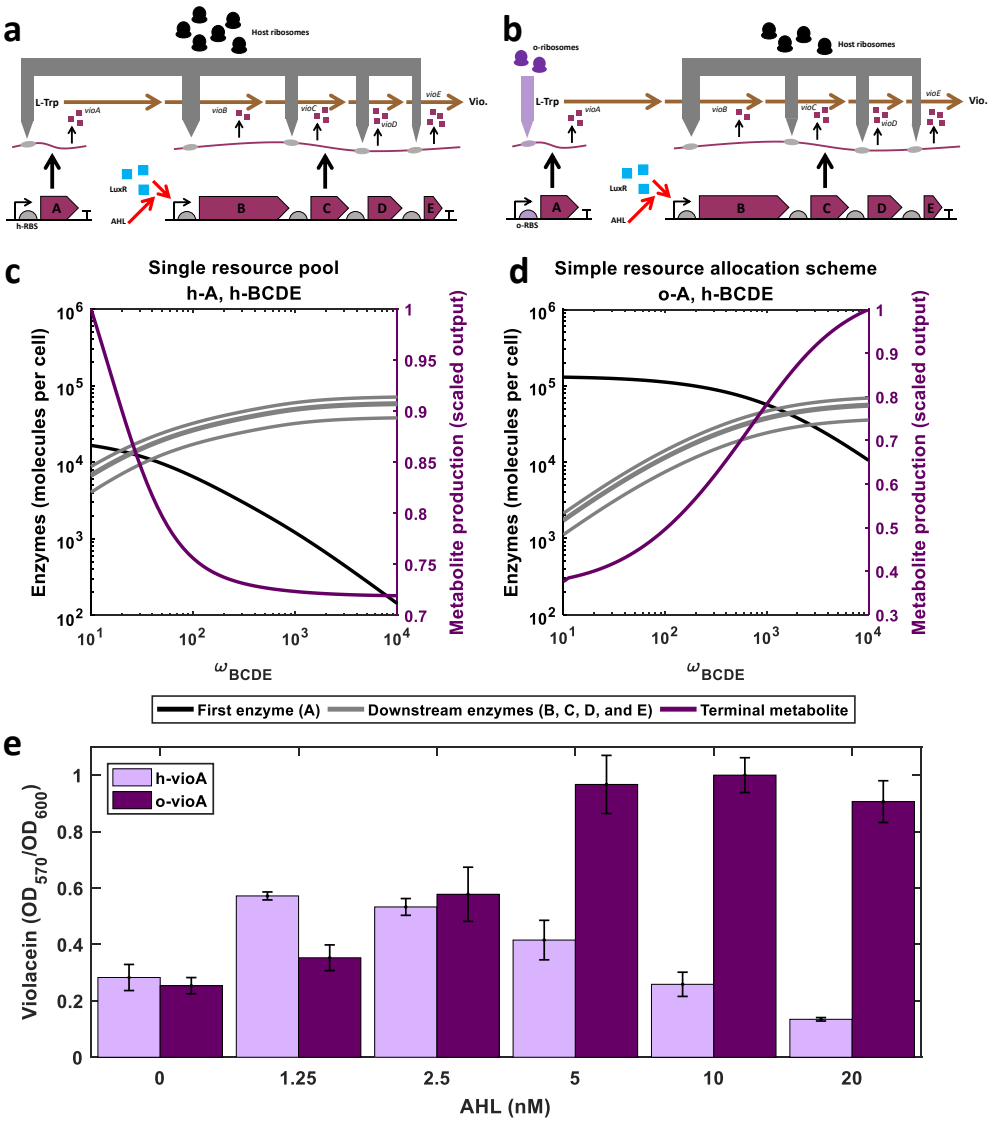

Figure 3.12: **Resource allocation control increases production of violacein.** The pathway is divided into two cassettes. *vioA* is constitutively expressed. The *vioBCDE* cassette is inducible with AHL. Simulations show the steady state concentrations of the pathway enzymes. Note that the enzymes are grouped by induction mechanism so that the downstream enzymes are depicted in the same colour. Variation in their levels is determined by protein size only. Scaled metabolite production represents the steady state amount of the final metabolite in the pathway, scaled by the highest amount achieved across the induction. Throughout $\omega_A = 25$ and $\omega_\rho = 500$ RNAs per min. The downstream cassette is induced as shown by varying ω_{BCDE}. **(a)** Single resource pool. **(b)** Simple resource allocation scheme. **(c)** Simulation results of the circuit depicted in (a). **(d)** Simulation results of the circuit depicted in (b). **(e)** Violacein production per cell at 24 h post induction. Violacein concentration was determined as outlined in Section 3.10. Normalised by the largest production per cell achieved across all conditions. Bars represent means ±1 S.D. $N = 3$.

is able to predict allocation schemes between the host and orthogonal ribosome pool which reduced coupling; namely that constitutively expressed genes should be targeted for translation by the orthogonal pool. We theoretically explored the possibility of using tethered o-ribosomes and predicted that their use would result in similar coupling profiles for each expression scheme (albeit with increased expression due to reduced recycling). We extended these simple observations to the design of a prototype biochemical pathway. We showed both theoretically and experimentally that by controlling the distribution of pathway genes across the two translational pools, total flux through the pathway could be increased.

3.10 Appendix: An overview of the experimental methods

The experimental data in this Chapter were provided by our collaborators Dr Juhyun Kim and Dr José I. Jiménez (University of Surrey). Analysis was carried out by the author in conjunction with Dr Juhyun Kim. Here we provide a brief overview of the experimental methods.

Plasmids were constructed using standard protocols [85]. The two gene circuit from [38] was excised by restriction enzymes and ligated into a standard vector from the SEVA collection [86]. Where appropriate the host ribosome binding sites were replaced by PCR using primers containing the orthogonal RBS sequence prior to ligation into the vector.

For experiments to test the impact of resource allocation on coupling between the two genes, the circuit and o-rRNA carrying plasmids were transformed into *E.coli* MG1655 [87]. In line with standard practise, *E. coli* strains were grown in Luria-Bertani (LB) medium at 37°C supplemented with the appropriate antibiotics overnight before being diluted into the same medium and grown to mid-exponential phase as determined by optical density at 600 nm (OD_{600}). These rapidly growing cells were then subcultured into fresh medium, supplemented with antibiotics and the inducers IPTG or AHL as appropriate (the function of these inducers is discussed earlier in this Chapter).

Growth (as OD_{600}) and population-level fluorescence of these cultures was monitored over 7 hours using a microplate reader. Periodically, samples were collected for single cell fluorescence measurements. Medium volume was maintained by addition of fresh medium with appropriate supplements. These samples were diluted and their fluorescence assessed by flow cytometry which separates individual cells and assesses their individual fluorescences. GFP and RFP expression, was determined by emission fluorescence, after excitation with the appropriate lasers. For each sample, 20,000 cells were measured and population means and standard deviations estimated by Juhyun Kim using the default instrument software and provided to the author.

To test our initial conclusions in a potential biotechnological application, the violacein production pathway from *Chromobacterium violaceum* ATCC 12472 was reconstituted in *E. coli* [84]. The *vio*-cassette used was designed *de novo* by Juhyun Kim and chemically synthesised commercially before being ligated into the SEVA

backbone and transformed in *E. coli* MG1655. Strains were cultured overnight and then subcultured in the manner described above with the addition of the amino acid tryptophan to the medium. Samples were taken periodically and the amount of violacein present was determined by spectrophotometry at 570 nm after lysis and removal of cell debris by centrifugation. Cell growth was determined by re-suspending the cell debris in water and measuring its optical density determined at 600 nm.

Chapter 4

Control of orthogonal ribosome allocation

4.1 Introduction

In Chapter 3, we developed a model of a simple microbe, which included the production and utilisation of a pool of orthogonal ribosomes, and demonstrated both theoretically (and experimentally) that using separate host and o-ribosome pools can significantly reduce coupling between co-expressed genes. However, we still observe a saturated input-output response profile; as RFP induction increases the output levels begin to plateau (both experimentally and in simulations). To assess the potential reasons for this saturation effect, we assessed the changing internal components for the circuit confirmations in Chapter 3 *in silico*. We find that the internal metabolite, e, does not fall below 85% across all inductions (Figure 4.1). This metabolite represents the outputs of metabolism; such as 'energy' (approximating the function ATP/GTP) but also the precursors needed for protein synthesis (such as amino acids). As this metabolite is maintained well across the induction levels, we propose that this is not the cause of the saturation effect. The transcription rates in our model are not limiting with the constitutive gene being maintained while the induced gene increases linearly with induction. Whilst our model does not explicitly include RNA polymerase sharing across genes, the propensity for gene expression to fall (mediated by falling RNA polymerase number) with decreasing internal energy is included via the transcriptional thresholds (o_X terms). Therefore we propose that the saturation effects observed in our experimental system remain due to competition for finite (host or orthogonal) ribosomes rather than other factors.

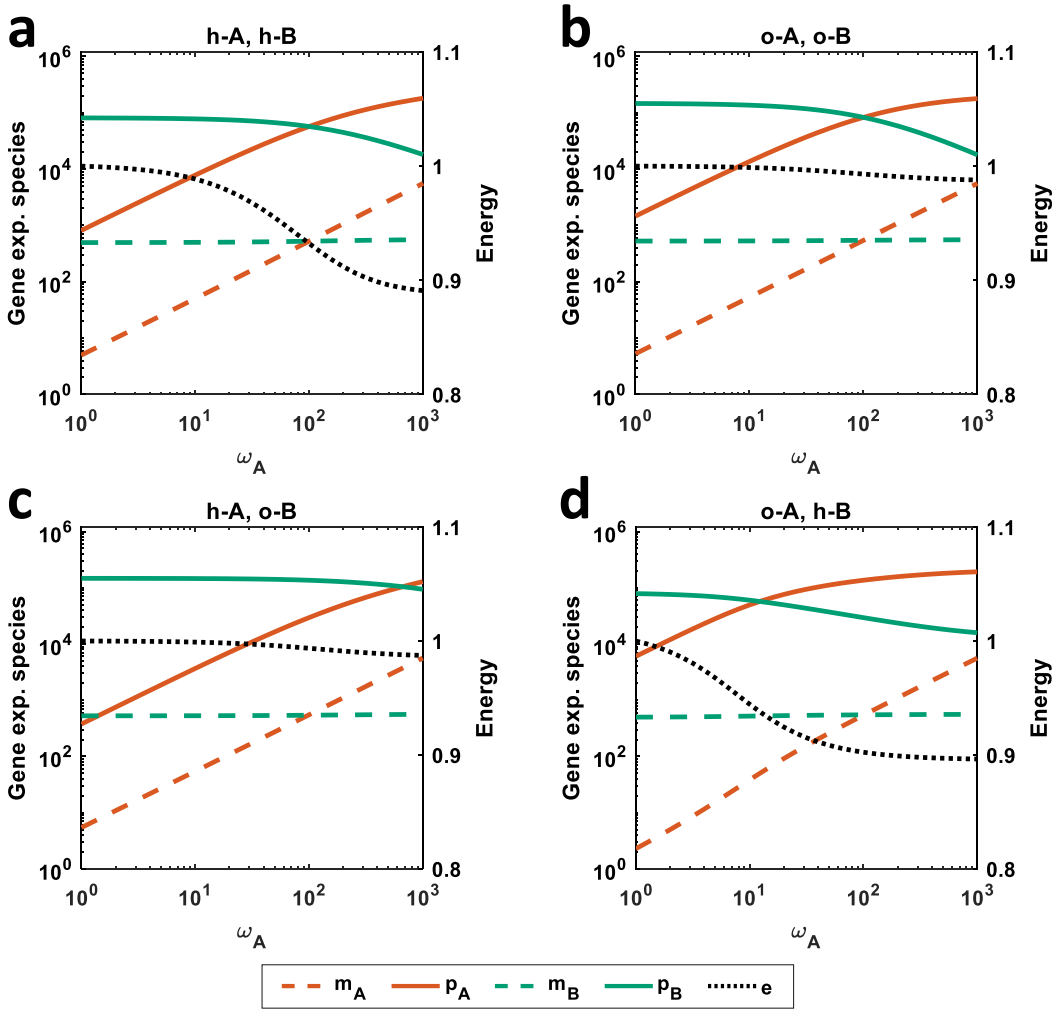

Figure 4.1: **Gene expression plateauing is mediated by saturation of the ribosome pool.** Simulations of the steady state concentrations of mRNA and protein levels from simple two-gene circuits. p_A is induced from $\omega_A = 1$ to 10^3 mRNAs per min while p_B is constitutively expressed $\omega_B = 100$ mRNAs per min. o-rRNA transcription is held at $\omega_\rho = 100$ rRNAs per min throughout. Normalised levels of the intermediate 'energy' metabolite (e) are also shown. **(a)** Both genes using the host ribosome pool. **(b)** Both genes using the o-ribosome pool. **(c)** p_A is translated by the host ribosome pool while p_B is translated by the o-ribosome pool. **(d)** p_A is translated by the o-ribosome pool while p_B is translated by the host pool.

Parameters	Reaction description	Lower bound	Upper bound	Rounding
ω	Maximal transcription rate	0	1,000	-2
k	Hill function threshold	10^3	10^6	-3
h	Hill constant	1	4	0
o	Transcription energy threshold	4.38		2
b	mRNA-ribosome binding rate	1		1
u	mRNA-ribosome unbinding rate	1		1
$\delta_{m,r}$	mRNA (m) or sRNA (r) degradation rate	0.1		2
δ_p	Protein degradation rate	0		2
n	Protein length (amino acids)	300		0

Table 4.1: **Controller parameter ranges** Rounding is to the nearest N digits using the MATLAB inbuilt *round* function. 0 rounds to the nearest integer, < 0 rounds to the left of decimal place and > 0 rounds to the right of the decimal place.

Given that these so-called 'open' loop confirmations do not remove resource mediated saturation effects, in this chapter we propose to use principles from feedback control to 'close' the loop. By introducing a demand-sensing mechanism and using this error signal to control o-ribosome production, we can develop controllers which act to match o-ribosome supply to circuit demand. Note that increasing ribosome biosynthesis is difficult to achieve experimentally and so the proposed controllers act to change the ratio of orthogonal to host ribosomes. By implementing additional chemical reactions into our host-aware model we are able to assess the function of simple genetic feedback mechanisms.

4.2 Controller design process

To develop the models of our proposed translational controllers, the ordinary differential equations describing the controller species were implemented in the host-aware model of o-ribosome production developed in Chapter 3. As described in that chapter, we also modified the expressions for the energy species de/dt, orthogonal ribosomes dP/dt and growth rate λ to take account of the additional energy usage and translation complexes of the new protein(s) required by the controller species. This fixes the controller topology.

Key controller parameters were allowed to vary within biologically feasible ranges (see Table 4.1). To achieve specific behaviours, we utilised the *genetic algorithm* function from MATLAB's Global Optimisation Toolbox (version 3.4), utilising the Parallel Computing Toolbox (version 6.8) where appropriate. Parameters were optimised to minimise the value of the cost function (Equation 4.2). We discretised the parameter space by rounding values as described in Table 4.1.

Resource-mediated coupling acts to perturb input (mRNA)–output (protein) relationships. Resource limitation results in saturation of the response and coupling between genes means that all inputs affect all outputs. We wish to design controllers which restore the modularity of the input–output response by removing resource-mediated saturation and coupling. We therefore desired the behaviour of the protein outputs to mirror that of the mRNA inputs. During the design process we scored coupling between two genes by simulating the increasing production of one gene (by increasing ω_1 from 1 to 10^4 mRNAs per min) while maintaining the production of a second (ω_2 is constant). We initially simulate the controller to steady state before inducing the two circuit genes. We then simulate the action of the circuit over a relatively long period ($0...10^4$ minutes). Upon termination of the simulation we assess the maximum absolute value of the derivative. If this value is greater than 1 (negligible on the scale of protein production in this model, as total protein at steady state is 10^8 amino acids) then the simulation is deemed not at steady state and the value is not returned to the scoring function. This acts to prevent design of controllers which do not reach steady state within 10^4 simulated minutes and so removes parameter sets which produce oscillation but also biases the selection scheme against controllers which reach steady state very slowly (although our interests here are in steady state behaviour not response time arguably slow controllers are also undesirable).

We quantified coupling by taking logs (base 10) of the induction of the first gene (ω_1) and the protein levels (p_1 and p_2) to produce the transformed data w, x and y. Using the in-built *polyfit* function we fit straight lines through the points w v x and w v y. We assessed the effect of increased w (i.e. increasing x) on y by observing the change in the gradient of the line w v y ($y_{gradient}$). We assessed the effect of the resource competition imposed by the constitutive gene on the induced gene by observing the deviation of the simulated values x (x_{sim}) from those expected by fitting a line through the points w v x (x_{fit}). We fit straight lines as this reimposes the desired input-output response; where protein levels mimic the behaviour of mRNA levels (the latter of which are approximately linear, see for example insets in Figure 4.3)

Individual simulations were scored as follows:

$$\text{score}(\omega_{\mathbf{1}}, \omega_2) = \sum \left((x_{fit} - x_{sim})^2 \right) + \sum \left((y_{gradient})^2 \right) \tag{4.1}$$

We give each component of the score equal weights as we do not wish to bias the optimisation towards selection of designs which factor the first component (which acts to reduce resource mediated saturation as measured by the linearity of x) over

or below the second (which acts to reduced coupling as measured by $y_{gradient}$).

This scoring function can be used to optimise the controller parameters for one level of expression for the constitutive protein p_2. To ensure the controller can function across a range of these values we develop a cost function which designs controllers over a range of protein p_2 inductions (vectors of induction, $\omega_\mathbf{2}$):

$$\text{cost}(\omega_\mathbf{1}, \omega_\mathbf{2}) = \sum_{1}^{N} \text{score}(\omega_\mathbf{1}, \omega_2) \tag{4.2}$$

4.3 Development of a negative feedback controller

Negative feedback utilises the output of a system to modify the input such that disturbances or fluctuations can be rejected to some extent. Here we consider the simplest implementation of such a loop, which converts a disturbance in the o-ribosome pool brought about by translation of a circuit (in effect the error signal) into a change in o-ribosome production, such that demand is approximately matched by supply.

We exploit the constitutive production of a repressor protein which utilises the o-ribosome pool for its own translation and inhibits the expression of the o-16S rRNA. Constitutive production of the repressor mRNA means that repressor protein levels act as a sensor for translational demand. We call this simple architecture the F-controller.

4.3.1 Development of the model

To implement the feedback controller in the model, we introduce the new equations and species required to describe the production of the controller protein's mRNA, its translation and the action of the protein. We denote these new components with F. To implement (i.e. 'close') the feedback loop, we modify Equation 3.12 to include the inhibitory action of the controller protein p_F. As before we use a simple Hill function to describe the inhibition (blue term).

$$\frac{d\rho}{dt} = \omega_\rho \cdot \left(\frac{1}{1 + (p_F/k_F)^{h_F}} \right) \cdot \left(\frac{e}{o_\rho + e} \right) - b_\rho \cdot p_R \cdot \rho + u_\rho \cdot P - (\delta_\rho + \lambda) \cdot \rho \tag{4.3}$$

4.3.2 Negative feedback successfully decouples two genes

We optimised the key controller parameters (ω_ρ, ω_F, k_F, h_F and b_F) using the cost function as described in Section 4.2.

To demonstrate the ability of the controller to increase o-ribosome production in response to demand, we consider the induction point when approximately equal amounts of protein are produced ($\omega_A = \omega_B = 100$ mRNAs per min) (Figure 4.2). We consider the constitutive expression of one circuit gene and simulate the response of the system to the stepped induction of a second gene (Figure 4.2b, c, d). When circuit demand is low (depicted in Figure 4.2a, *left*), before the induction of the second gene, competition between the circuit and the controller is low. This results in high expression of the controller protein and therefore high repression of the o-rRNA, meaning that few ribosomes are co-opted from the host. Upon induction of the second gene (depicted in Figure 4.2a, *right*), the demand for o-ribosomes increases (Figure 4.2b). The repressor mRNAs will remain largely unaffected, but their translation falls due to increased competition (Figure 4.2c). The decrease in repressor production results in relief of the inhibition of the o-16S rRNA gene and so increased o-rRNA production and increased co-option of host ribosomes (Figure 4.2b). This results in the maintenance of circuit protein production as other circuit genes are induced (Figure 4.2d), although it should be noted that a small steady state error of $\sim 2\%$ remains. Note that whilst the free o-ribosome is not maintained in the presence of the controller there is only a 5% fall (Figure 4.2b) but in the absence of the controller this fall is 50%.

We simulate the increasing induction of one protein p_A and assess the fall in a constitutive protein p_B. The negative feedback controller successfully slows the fall in p_B while removing the resource-mediated saturation effect on p_A (Figure 4.3). Note that the mRNA levels in both open and closed loop confirmations are equivalent. To determine the effect of the controller, we simulate the performance of the same circuit in the presence of the o-ribosome system but in the absence of the controller. As we found that protein levels were a large determinant of host response in Chapter 3 (specifically Figure 3.3), we vary the value of ω_ρ so that both the open loop (i.e. without the feedback mechanism) and closed loop (i.e. with the feedback mechanism) circuits produce the same protein outputs when the induction of the two genes is matched (i.e. when $\omega_A = \omega_B = 100$ mRNAs per minute) (Figure 4.3). Comparison of the protein levels with their respective mRNAs shows the controller successfully restores the input–output mapping. In the open loop confirmation p_B

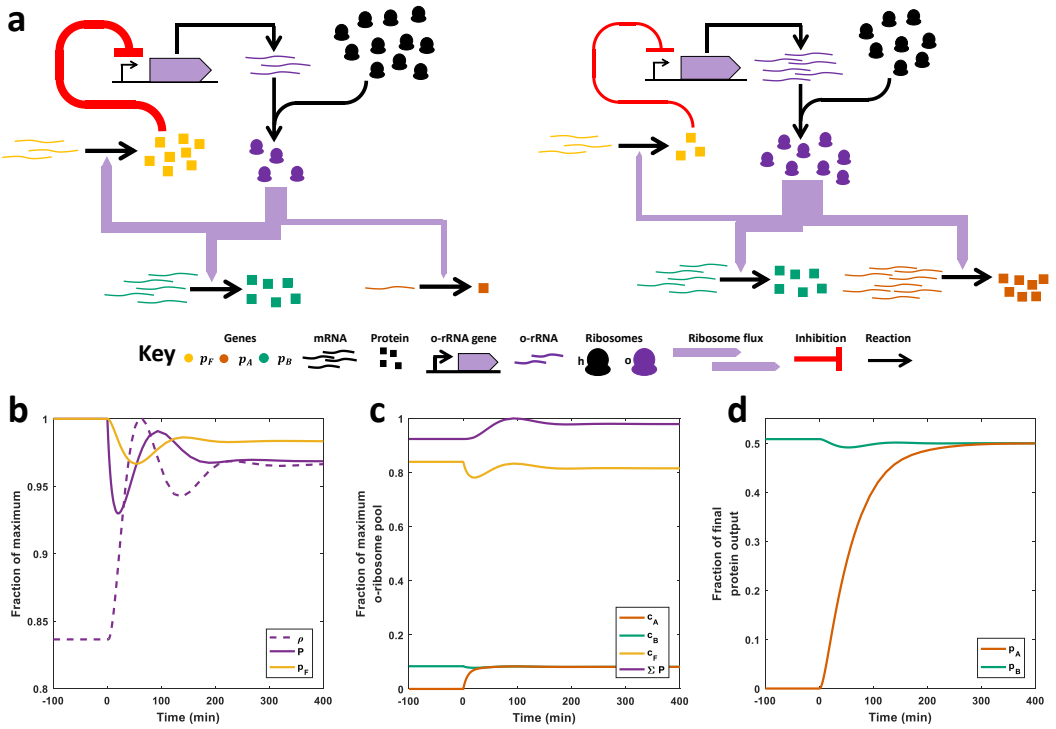

Figure 4.2: **Operation of the F-controller.** Simulations of a two-gene circuit were carried out as described in the main text. p_A is induced at $t = 0$. The changing distribution of controller and circuit components is shown in (b), (c) and (d). **(a)** Structure and function of the F-controller. *Left,* Low demand circuit. When competition is low p_F expression is high and so o-rRNA production is low and hence co-option of host ribosomes is low. *Right,* High demand circuit. As circuit demand increases (as p_A is induced), the o-ribosome pool redistributes across circuit and controller genes (width of purple ribosome flux lines) due to competition between mRNAs. This reduces translation the constitutively expressed p_F. This reduces the repression of o-rRNA production, allowing more co-option of ribosomes to the orthogonal pool. This maintains ribosome flux for m_B translation despite the increase in m_A. **(b)** Changing distribution of the controller components. ρ, o-16S rRNA; P, free o-ribosome; p_F, controller protein. Normalised by their maximum value. **(c)** Changing distribution of the translation complexes over time in response to p_A induction. c_Y, translation complex of gene Y, Σ P, sum of all o-ribosomes. Normalised by maximum Σ P. **(d)** Protein output over time normalised by sum of the final circuit protein concentration.

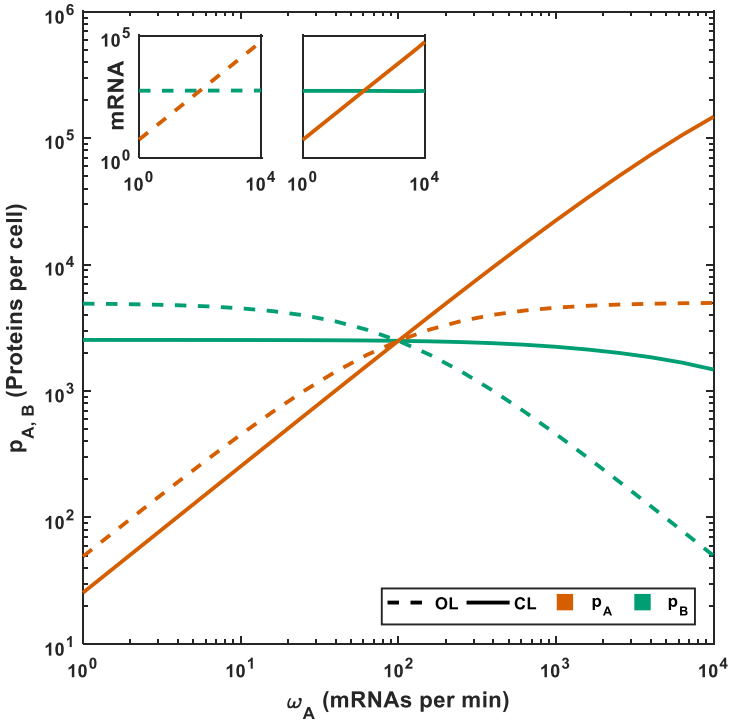

Figure 4.3: **The simple F-controller decouples co-expressed genes.** Simulation of the action of the negative feedback controller. ω_A is varied between 1 and 10^4 mRNAs per minute. w_B is held constant at 100 mRNAs per minute throughout. The simulation time span is increased until it reaches steady state. Controller parameters: $\omega_\rho = 350$ rRNAs per min; $\omega_F = 10^3$ mRNAs per min; Hill function parameters $k_F = 10^4$ and $h_F = 4$. Open loop $\omega_\rho = 1.6$ rRNAs per min. Insets, *left,* mRNA concentration in the open loop; *right,* mRNA concentration in the closed loop.

falls 50% over the first two orders of magnitude of ω_A induction, whilst in the presence of the controller the equivalent fall is negligible at less than 2%. p_B is maintained within 10% of its initial value over three orders of magnitude of ω_A induction. Over the whole range of ω_A, p_B only falls 40% (compared to 99% in the open loop). In the presence of the controller p_A increases linearly and is not subjected to the saturation effects of the open loop controller.

4.3.3 Robustness analysis and extension to different circuits

To test the robustness of the F-controller design to uncertainty or variability arising from potential experimental implementations, we carried out a simple robustness analysis around the optimal solution identified. We took a simple Monte Carlo

approach by drawing 10,000 random parameter sets where all parameters of the o-ribosome production system and controller protein could be varied between ±50% from the optimal value. All parameter sets were simulated to steady state. This analysis indicates that the feedback topology is highly robust, with all designs tested showing a better mRNA-protein mapping than the circuit using the o-ribosome pool without control, including those parameters which are difficult to design such as the 'transcriptional energy threshold' o_ρ (Figure 4.4a). Approximately 60% of controllers produce results which fall within 50% of the original controller function across all ω_A inductions.

To identify if oscillatory behaviour of the controlled system occurs, we simulated 10^6 random controllers, drawing parameters from a uniform distribution between the bounds shown in Table 4.1. We simulate the controller alone without a circuit protein (to determine the controller's apparent stability) and then in the presence of a circuit protein ($\omega_A = 100$ mRNAs per min) (to determine if the disturbance caused by the circuit reveals a hidden instability in the system). Whilst these controllers, as expected, show significantly different behaviours, all converge to a fixed point, supporting the argument that the stability of this controller is highly robust.

We designed our controller to decouple circuit genes over a range of ω_B values. We simulated the circuit as before by holding ω_B constant and increasing ω_A but considered different ω_B values for each set of simulations (Figure 4.4b). The controller acts to decouple genes across all ω_B values tested. In all cases the controller function begins to break down at $\omega_A > 10^3$ mRNAs per min. As ω_B increases, the efficiency of the controller also increases. Simulations of the open loop confirmation at these levels of induction also demonstrate decoupling. At very large values of ω_B, the ribosome sequestration ability of mRNA B increases due to their large number and so this gene becomes more insensitive to resource limitations. Concurrently, there are fewer ribosomes for mRNA A expression resulting in linearisation of the p_A response. These results are consistent with those reported in [38].

4.3.4 Design principles

Having demonstrated that the controller functions as expected and is largely robust, we carried out a sensitivity analysis around the optimal controller parameter set to determine how key parameters contributed to its behaviour. For consistency with Chapter 3 and to inform the construction of the prototype by our experimental collaborators we formulate our design in the form of isocost lines.

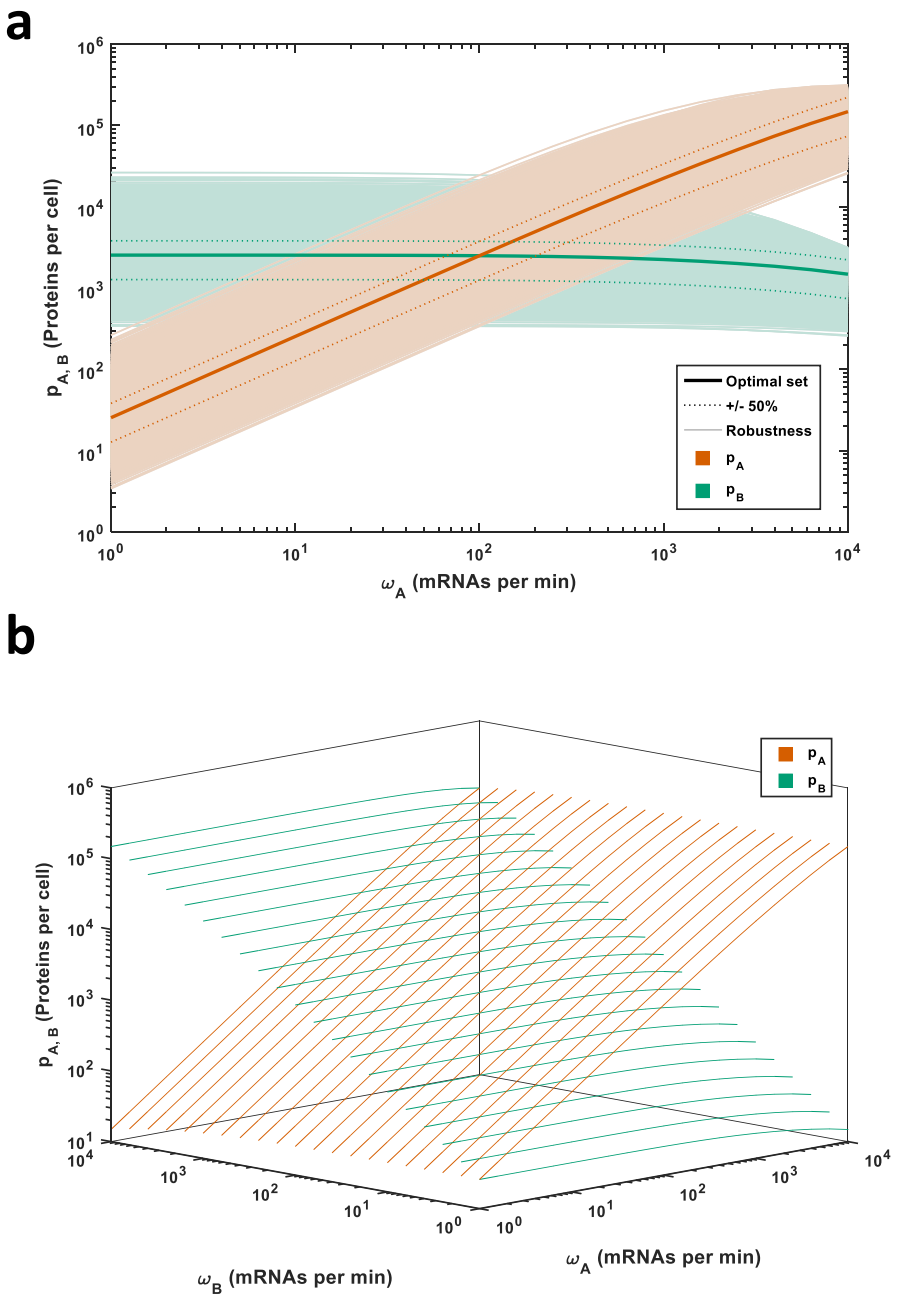

Figure 4.4: **Robustness of the simple negative feedback controller (a)** Function and robustness of the controller decoupling two genes. The optimal controller parameters were perturbed by drawing random values between ±50% of the original value ($N = 10,000$). All parameters controlling o-rRNA and controller protein were allowed to vary. For ease of analysis the steady state values within 50% of the optimal set are shown as dotted lines. **(b)** Function of the controller over a range of p_A (ω_A) and p_B (ω_B) inductions.

As found in Chapter 3, the coupling is largely independent of the size of the o-ribosome pool (i.e. varying ω_ρ has negligible effect, Figure 4.5a). However, in this case the small o-ribosome pool ($\omega_\rho = 10$ rRNAs per min) shows increased coupling due to the presence of the additional load on the o-ribosome pool brought about by the controller protein expression. Our sensitivity analysis shows that the controller protein needs to have a high ability to sequester o-ribosomes; i.e. decoupling is achieved when controller transcription (ω_F) or RBS strengths (b_F) are high (Figure 4.5b, c). At low ω_F or b_F, representing a weak closing of the feedback loop, there is no effect of the feedback. Regarding the characteristics of the controller protein itself, the sensitivity analysis shows that the protein should be multimeric (with increasing h_F increasing decoupling) and tightly binding (low k_F) (Figure 4.5d, e).

4.4 Development of an integral controller

Increasingly there is interest in realising integral control in synthetic circuits. In addition to 'acting against the perturbation' to reduce the steady state error, the advantage of integral action causes the system to undergo perfect adaptation (i.e. no steady state error). This would produce advantages over the F-controller implemented in Section 4.3 where a small steady state error remains. A genetic integral controller topology composed of multiple activatory and inhibitory transcription factor-based feedback loops was proposed by Ang *et al.* [88]. A further topology based on a smaller number of proteins was proposed by Briat *et al.* in [89]. Recently, the experimental realisation of this prototype has been reported [80]. These designs take the form of a process which produces an activator which itself activates an inhibitor which inhibits the process. If the process output is increased by a disturbance then the process is subjected to greater inhibition. If the process output falls, the level of inhibitor production falls and so the process input increases allowing adaptation.

These proposed controllers are implemented through the use of multiple proteins. Given that additional protein production would impose greater competition for o-ribosomes (and, of course, greater burden), we leverage the advantage of small RNAs in our proposed controller. The expression of small RNAs requires only transcription by RNA polymerase and not translation by ribosomes. As discussed in Chapter 2, the current experimental evidence is that in most cases RNA polymerase is not limiting. Production of RNAs is also less energy intensive than the production of proteins so further reducing the whole cell burden of their expression [90]. Taking

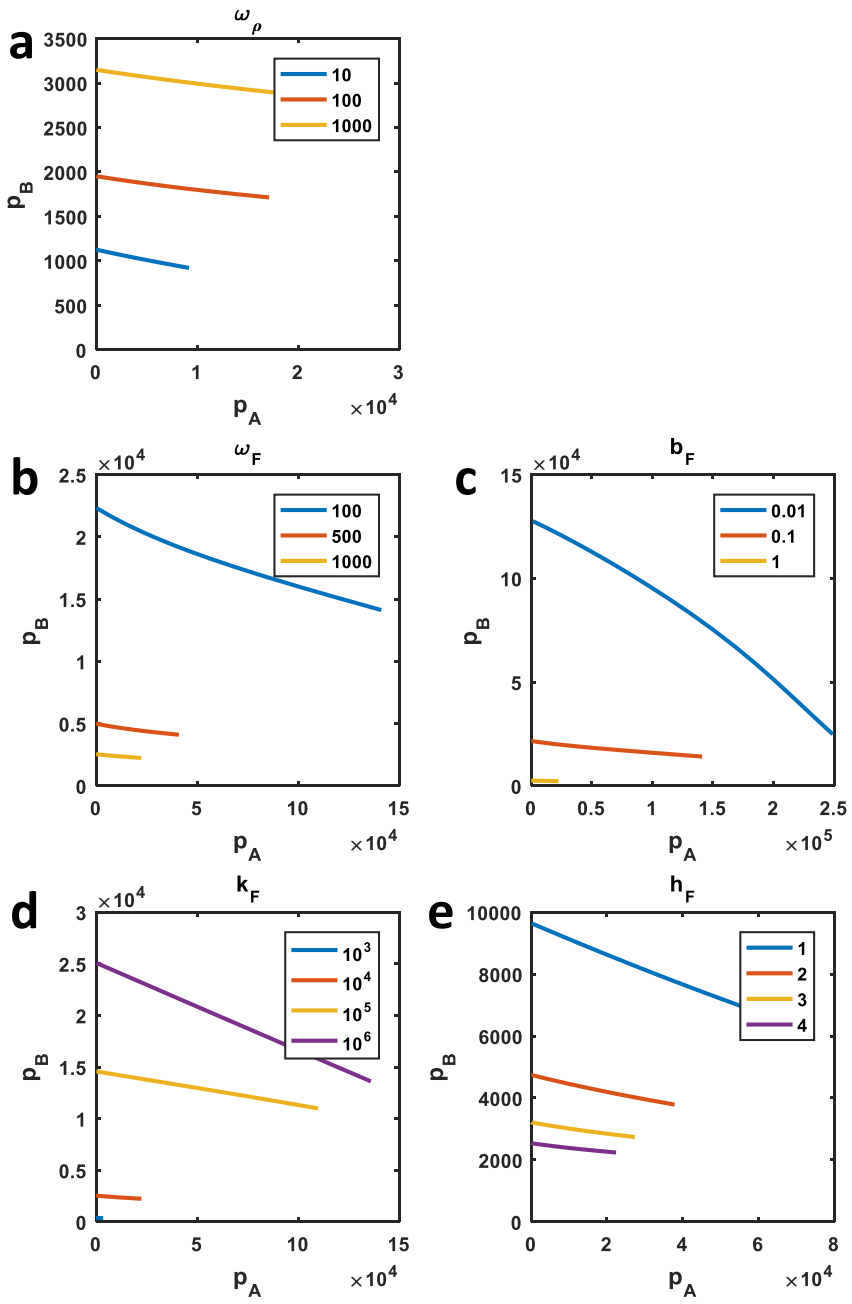

Figure 4.5: **Design of the simple negative feedback controller.** The impact of each varying parameter on the optimal controller. Isocost lines were plotted by keeping $\omega_B = 100$ and varying ω_A between 1 and 10^3 mRNAs per minutes. The resulting steady state protein concentrations are plotted. The impact of varying the following parameters is shown: **(a)** The maximal o-rRNA transcription rate. **(b)** The maximal transcription rate of the controller protein mRNA. **(c)** The strength of the controller protein RBS. **(d)** The dissociation constant of the repressor. **(e)** The Hill function coefficient of the repressor.

inspiration from [54, 89], we propose the inhibitory action takes the form of sequestration of the o-rRNA by a synthetic small RNA. Briat *et al.* propose the following reaction network:

Some output X is produced in response to the input reference species Z_1 by some function f (Equations 4.4 and 4.5). This activates the production of Z_2 by some function g and therefore the output is measured via the concentration of Z_2, the sensor species (Equation 4.6). The action of the controller is through a 'comparison' reaction in which both Z_1 and Z_2 are eliminated (Equation 4.7). (This elimination reaction is the proposed source of the integral action [89]:

$$\emptyset \xrightarrow{\mu} Z_1 \tag{4.4}$$

$$\emptyset \xrightarrow{f(Z_1)} X \tag{4.5}$$

$$\emptyset \xrightarrow{g(X)} Z_2 \tag{4.6}$$

$$Z_1 + Z_2 \xrightarrow{\eta} \emptyset \tag{4.7}$$

This chemical reaction network acts to reject disturbances to the species X.

We propose the following topology for our putative controller:

The o-rRNA takes the place of the reference species Z_1 but can only act through the o-ribosome pool; hence the transcription of the o-rRNA, the co-option of host ribosomes and translation of the species X by the o-ribosome pool form the chemical reactions which enact the $f(\cdot)$ function. The function of species X (i.e. the function $g(\cdot)$) is carried out by an activatory transcription factor p_G. The level of p_G is a measure of circuit induced disturbance. p_G activates the production of a small RNA (r_i) which takes the place the sensor species (Z_2). The elimination/comparison reaction (equivalent to Equation 4.7) takes the form of a biomolecular reaction between ρ and r_i to form the RNA duplex r_d which is degraded:

$$\rho + r_d \underset{u_i}{\overset{b_i}{\rightleftharpoons}} r_d \xrightarrow{\delta_{r_d}} \emptyset \tag{4.8}$$

4.4.1 Development of the model

To implement the proposed controller in the model we introduce the new equations and species required to produce the constitutively expressed activator's mRNA, translation complex and protein (denoted G).

We model the sequestration reaction by applying the Law of Mass Action to Equa-

tion 4.8 and modify the o-rRNA production ordinary differential equation (Equation 3.12) accordingly:

$$\frac{d\rho}{dt} = \omega_\rho \cdot \left(\frac{e}{o_\rho + e}\right) - b_\rho \cdot p_R \cdot \rho + u_\rho \cdot P... \\ - (\delta_\rho + \lambda) \cdot \rho - b_i \cdot \rho \cdot r_i + u_i \cdot r_d \tag{4.9}$$

The small RNAs are born spontaneously as previously described for mRNAs. The rate of production is also scaled by the activatory Hill function (p_G term). The interaction with the o-rRNA is described by the Law of Mass Action:

$$\frac{dr_i}{dt} = \omega_i \cdot \left(\frac{(p_G/k_G)^{h_G}}{1 + (p_G/k_G)^{h_G}}\right) \cdot \left(\frac{e}{o_i + e}\right) - b_i \cdot \rho \cdot r_i + u_i \cdot r_d - (\delta_{r_i} + \lambda) \cdot r_i \tag{4.10}$$

The production dynamics of RNA duplex and its degradation/dilution are given by:

$$\frac{dr_d}{dt} = b_i \cdot \rho \cdot r_i - u_i \cdot r_d - (\delta_{r_d} + \lambda) \cdot r_d \tag{4.11}$$

4.4.2 Function of the controller

We optimise the design of the G-controller as outlined in Section 4.2, allowing the following parameters to vary ω_ρ, ω_i, ω_G, b_G, k_G and h_G. We assume the interaction between the o-rRNA and the small RNA is diffusion limited and reversible (such that $b_i = u_i = 1$, Equation 4.8). We account for the rapid degradation of double stranded RNA by setting the decay constant (δ_{r_d}) to 0.5 molecules per minute (~ 5 times greater than single stranded RNAs). During initial controller design it was noted that a small but significant number of parameter sets resulted in oscillations (discussed further below).

As above, we first consider the action of the controller by comparing it to an open loop confirmation with equivalent protein production where p_A and p_B intersect (Figure 4.6). Over the first two orders of magnitude of ω_A the controlled ω_B falls only 7%, while in the absence of the controller this fall is 45%. At inductions of p_A over around 500, p_B suffers significant perturbation with p_B falling over 35%. The controller restores the input–output mapping of p_A between $\omega_A = 1$ to 10^3 mRNAs per min.

To demonstrate the changing distribution of controller components, as before, we consider the induction point when approximately equal amounts of protein are made

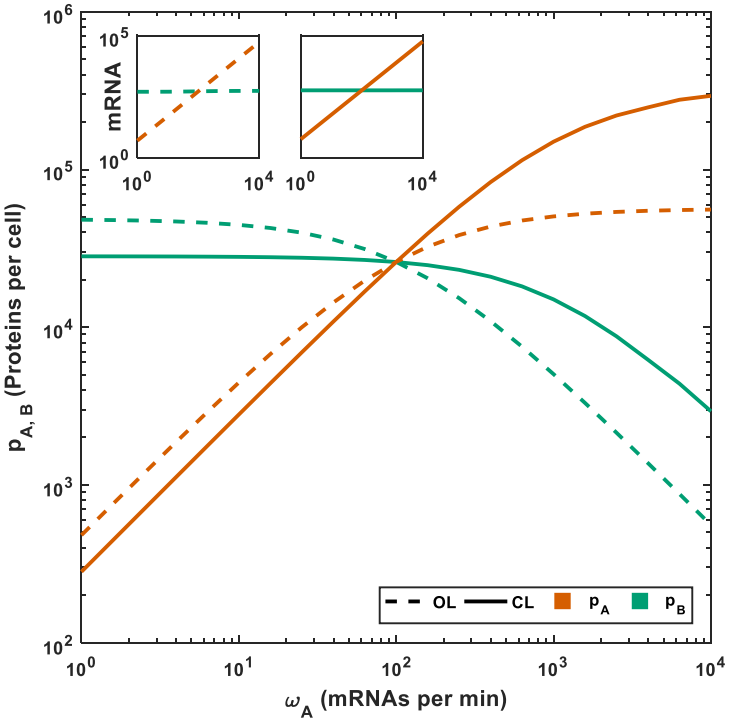

Figure 4.6: **The proposed integral controller decouples co-expressed genes.** Simulation of the action of new controller architecture. ω_A is varied between 1 and 10^4 mRNAs per minute. ω_B is held constant at 100 mRNAs per minute throughout. The simulation time span is increased until it reaches steady state. Controller parameters: $\omega_\rho = 220$ rRNAs per min; $\omega_i = 930$ sRNAs per min; $\omega_G = 990$ mRNAs per min; Hill function parameters $k_F = 8 \times 10^4$ and $h_F = 4$. Open loop $\omega_\rho = 15.825$ rRNAs per min. Insets, *left,* mRNA concentration in the open loop; *right,* mRNA concentration in the closed loop.

($\omega_A = \omega_B = 100$ mRNAs per min) (Figure 4.7). As before, we consider the response of a constitutively expressed gene (p_B) to the induction of second (p_A) (Figure 4.7). When circuit demand is low (depicted in Figure 4.7a, *left*), the expression of p_G is high. This results in high expression of the sRNA (r_i) and so high sequestration of the o-rRNA (Figure 4.7c). Upon induction of the second gene (and so increase in demand, depicted in Figure 4.7, *right*) the expression of p_G falls due to a decrease in translation because of resource competition. This results in a large decrease in sRNA production and so liberates o-rRNA, whose levels rise by nearly 40%. This results in greater co-option of host ribosomes to the o-ribosome pool and therefore decouples the co-expressed genes, although note that a negligible error (0.2%) persists at steady state (Figure 4.7d). The o-ribosome pool is not maintained in the presence of the disturbance but the perturbation is reduced from 50% (in the absence of control) to only 6% (in the presence of the controller).

4.4.3 Robustness analysis and extension to different circuits

As before, we investigated the controller robustness by taking a Monte Carlo approach. We draw 10,000 random parameter sets where we allowed all controller parameters to vary by up to 50%. Taking into account the putative instability we observed during initial model design, we ensured that we started with 10,000 controllers which initially appeared stable (i.e. in the absence of the circuit) by rejecting those that did not reach a steady state during their initial simulation. Of these 10,000 sets, when we simulated the circuit in Figure 4.6, only 5,789 controllers produced stable results for all ω_A values tested - a failure rate of $\sim 40\%$. Controllers which result in greater protein production (i.e. where $p_B(\omega_A = 1)$ is much greater than the value produced by the optimal controller) seem to be more likely to be unstable. Of the controllers which function as desired nearly all fall below 50% of the original controller (Figure 4.8a). Only 246 combined controller and circuit ω_A values simulated fall within 50% of the original function. Note that this number rises to nearly 800 (still $< 0.1\%$) if only results less than $\omega_A = 10^3$ are analysed.

To investigate this instability further we simulated 10^6 random controllers, drawing parameters from a uniform distribution and simulated as described in Section 4.3.3. We find a number of unstable controllers which we discuss further in Section 4.4.4.

We designed the G-controller to function over a range of ω_B values. We assessed the function of the controller over a number of ω_A and ω_B values (Figure 4.8b). It successfully decouples genes across a large proportion of the circuit design space

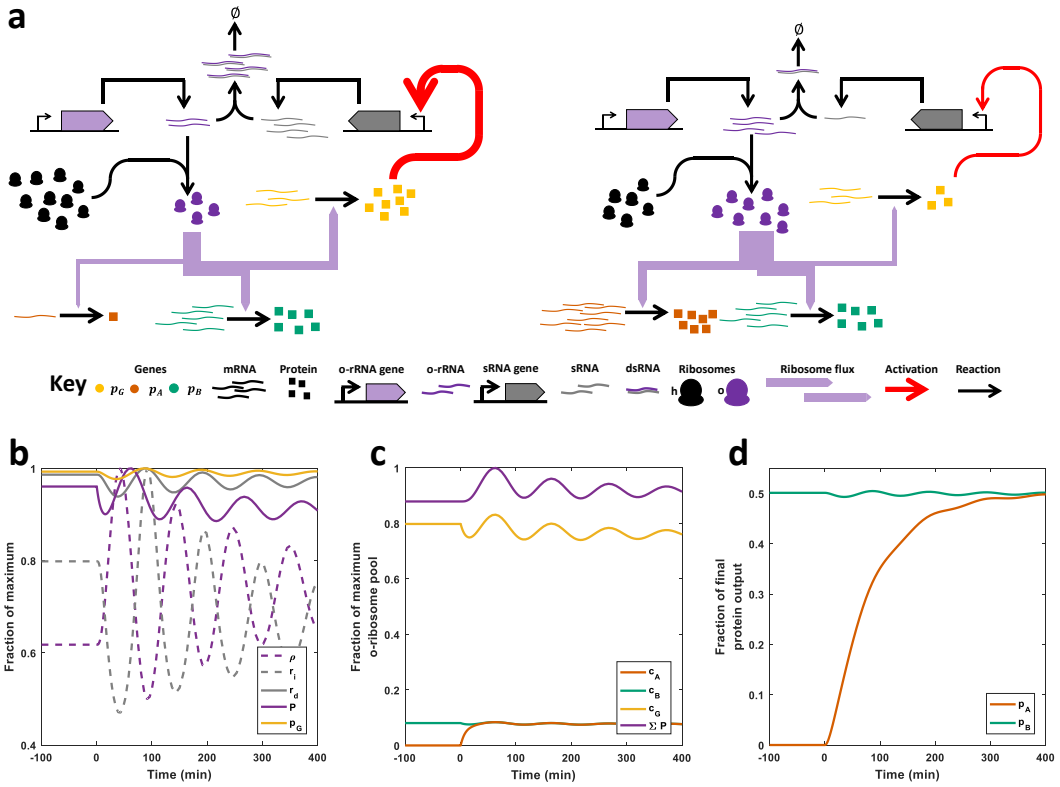

Figure 4.7: **Operation of the G-controller.** Simulations of a two-gene circuit were carried out as described in the main text. p_A is induced at $t = 0$. The changing distribution of controller and circuit components is shown in (b), (c) and (d). The system reaches steady state at $t = 10^3$ min. **(a)** Structure and function of the G-controller. *Left,* Low demand circuit. When competition is low, p_G expression is high and so sRNA transcription is activated resulting in high sequestration of, and hence degradation of, the o-rRNA. Therefore co-option of ribosomes to the o-ribosome pool is low. *Right,* High demand circuit. As circuit demand increases (as p_A is induced), the o-ribosome pool redistributes across circuit and controller genes (width of purple ribosome flux lines) due to competition between mRNAs. This reduces translation of p_G and hence reduced activation of r_i transcription. This results in less sequestration of the o-rRNA and so increased co-option of ribosomes to the orthogonal pool. This maintains ribosome flux for m_B translation despite the increase in m_A. **(b)** Changing distribution of the controller components. ρ, o-16S rRNA; r_i, sRNA; r_d, RNA duplex (dsRNA); P, free o-ribosome; p_G, controller protein. Normalised by their maximum value. **(c)** Changing distribution of the translation complexes over time in response to p_A induction. c_Y, translation complex of gene Y, $\Sigma\ P$, sum of all o-ribosomes. Normalised by maximum $\Sigma\ P$. **(d)** Protein output over time normalised by sum of the final circuit protein concentration.

tested. However, at high ω_A the response begins to saturate and we see greater perturbation to p_B. As before as ω_B increases, p_B becomes less sensitive to resource competition due to its superior ribosome sequestration ability. Instability emerges in these resource intensive circuits (u/s points in Figure 4.8b).

4.4.4 Design principles

Having demonstrated that the optimal G-controller functions as expected (although with poor performance) we next considered how each parameter contributed to the behaviour. Again, we consider how each contributes to the slope of the isocost line before we consider the issue of the instability uncovered by our robustness analysis.

By beginning with our optimal controller we vary each of the experimentally tunable parameters individually (Figure 4.9). We find that in addition to setting the size of the protein budget in this controller, the maximal o-rRNA transcription rate (ω_ρ) also determines the slope of the isocost line, with intermediate values showing the best decoupling. If $\omega_\rho > \omega_i$ then decoupling fails and the slope of the isocost line increases. Consistent with this observation ω_i must be large and its activator (i.e. p_G) must be highly expressed (large ω_G and b_G, Figure 4.9c, d). The slope of the isocost line is largely insensitive to the properties of the activator itself; with the slope being largely consistent for $k_G \leq 10^5$ and $h_G > 2$ (Figure 4.9d, e). Here, as opposed to the F-controller, increasing co-operativity increases protein expression.

Having found that a number of G-controllers are unstable in Section 4.4.3, we further investigated the parameter regimes where the controller is unstable. We simulated the controllers in the absence and presence of circuit genes as the instability of some controllers is only revealed upon the addition of the input. For example, in Figure 4.10b a controller which initially appears stable is revealed to be unstable forming sustained oscillations upon the induction of a circuit gene. A pairwise comparison of parameters for these unstable controllers reveals distinct regions of parameter space where controllers are likely to be unstable. Simulating 10^6 random controllers from within this region reveals that whilst the majority are stable, 6% are unstable in the absence of the circuit, with this value rising to 10% upon inclusion of the circuit gene. This provides more evidence to show that controllers which initially appear stable can in fact be unstable when subjected to circuit inputs. Analysis of the parameters which give rise to oscillations does not allow further refinement. As the region is poorly defined at present it complicates the design process. Implementation *in vivo* relies on the identification of biological components which have the desired dynamics.

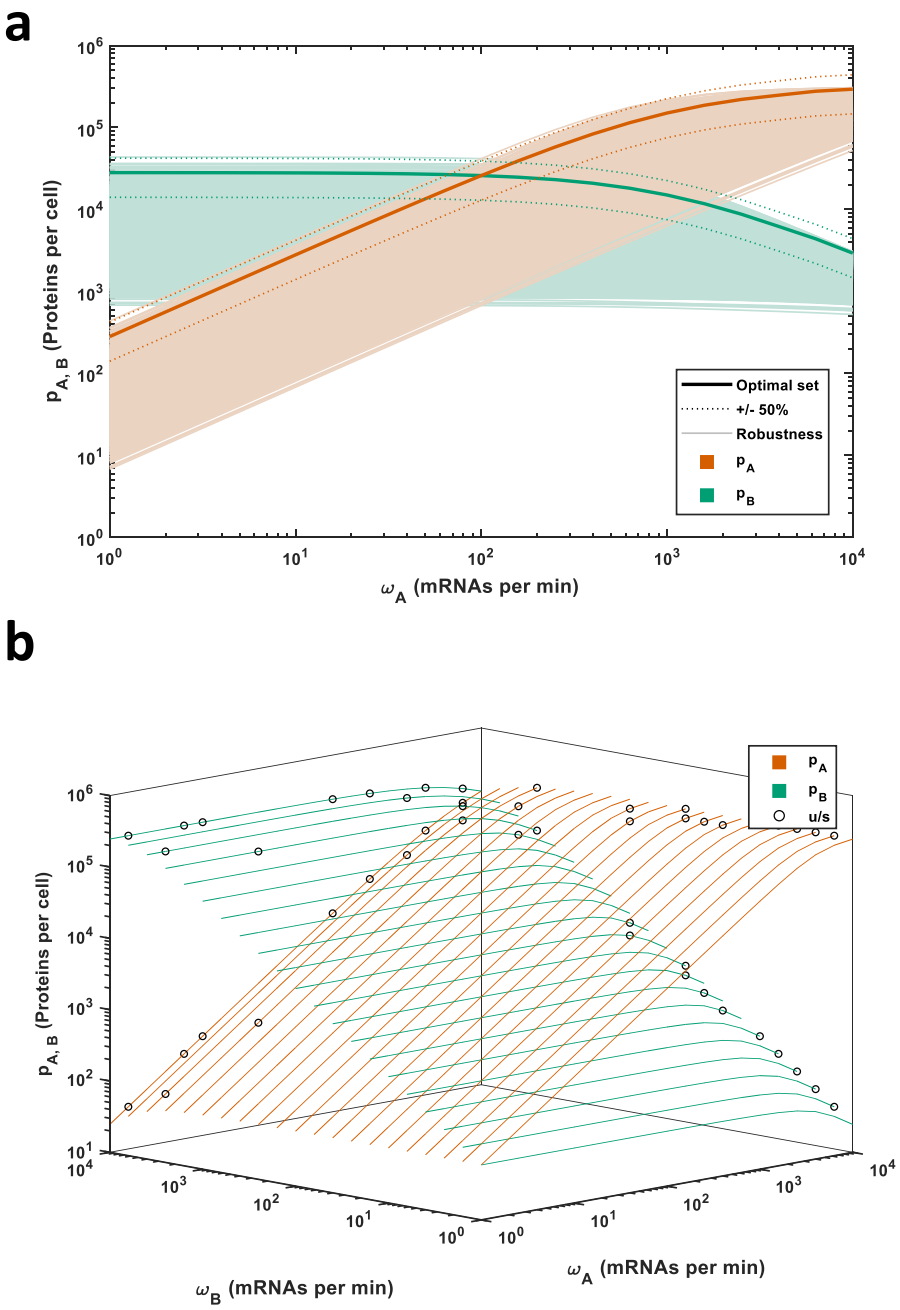

Figure 4.8: **Robustness of the integral controller. (a)** Function and robustness of the controller decoupling two genes. The optimal controller parameters were perturbed by drawing random values between ±50% of the original value ($N = 5,934$ reach a steady state with the time span across all ω_A tested). All parameters controlling o-rRNA and controller protein were allowed to vary. For ease of analysis the steady state values within 50% of the optimal set are shown as dotted lines. **(b)** Function of the controller over a range of p_A and p_B inductions. u/s, simulations do not reach steady state so protein level at time $t = t_{max}$ is taken as an approximation for long term behaviour.

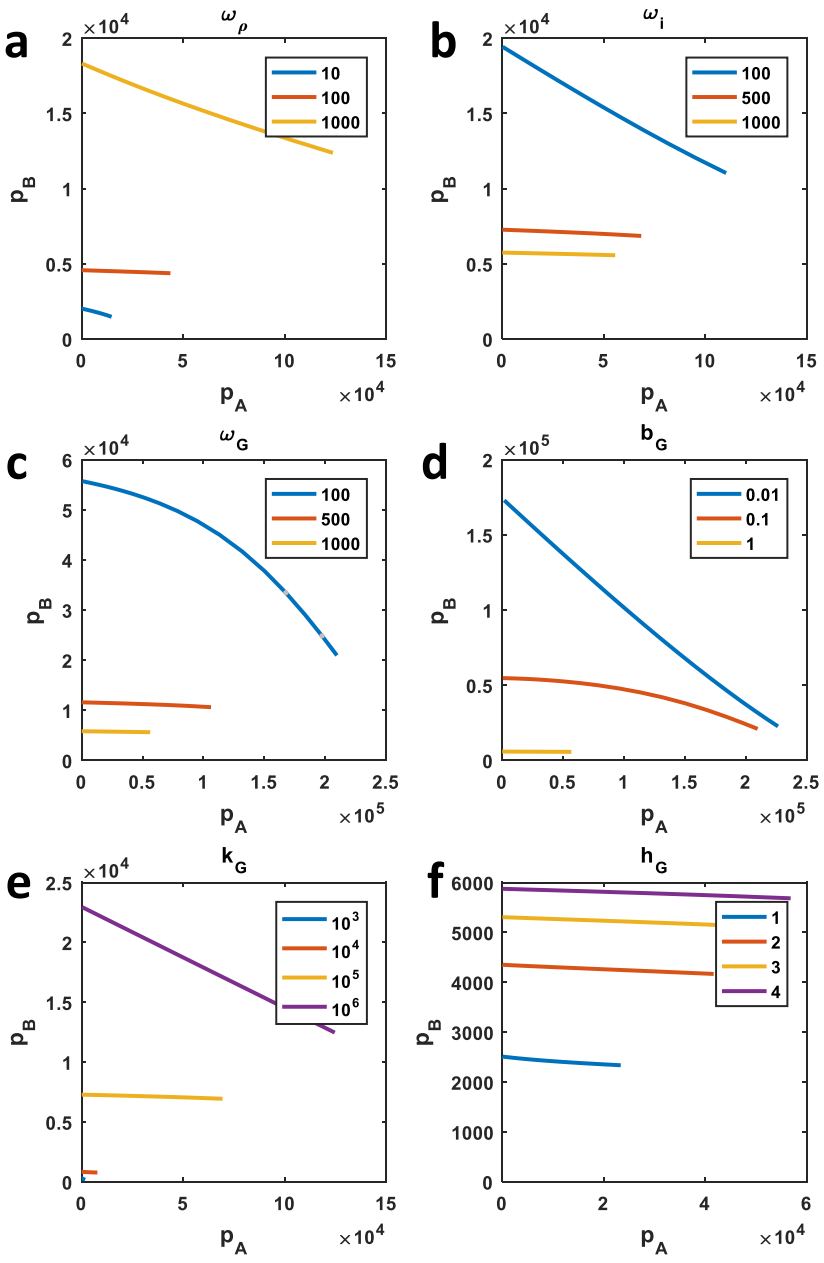

Figure 4.9: **Design of the integral G-controller** The impact of each varying parameter on the optimal controller. Isocost lines were plotted by keeping $\omega_B = 100$ and varying ω_A between 1 and 10^3 mRNAs per minutes. The resulting steady state protein concentrations are plotted. Where simulations failed to reach steady the point is highlighted in grey and the long term behaviour is plotted instead. The impact of varying the following parameters is shown: **(a)** The maximal o-rRNA transcription rate. **(b)** The maximal transcription rate of the o-rRNA sequestering small RNA. **(c)** The maximal transcription rate of the sRNA activator protein. **(d)** The activator protein RBS strength. **(e)** The dissociation constant of the activator. **(f)** The Hill function coefficient of the activator.

Parameter	Lower bound	Upper bound
ω_ρ	100	800
ω_i	500	1000
ω_G	0	1000
k_G	10^4	10^5
h_G	3	4
b_G	0.01	1

Table 4.2: **Unstable region of parameter space for the proposed integral controller.** The approximate bounds which define the region of reduced stability for the integral controller. Values were taken from the analysis in Figure 4.10a.

Often only second choice components with similar, but not optimal dynamics, are available, but as this analysis shows, controllers may have similar values but one may be stable and the other form oscillations.

4.4.5 Investigating the loss of perfect adaptation via model reduction

From Figure 4.7, we can see that after the induction of the second gene, the first does not fully recover – a small steady state error remains, indicating the G-controller is not functioning as a perfect integral controller in this instance. We propose that this loss of integral control may be due to the inclusion of additional host factors in our model. To test this hypothesis we gradually remove host factors to derive a simple model which *does* show perfect integral action.

We initially removed the host enzymes and other proteins to reduce competition and metabolism-based feedback (the latter through holding the internal energy levels (e) constant). We removed the mRNA and translation complexes (m_R and c_R) by introducing the spontaneous resource-free production of p_R (born at rate γ_R in the reduced model).

We parameterised the new model using the same values as before and we scaled the transcription rates (ω terms) by the $(\tilde{e}/(\tilde{e}+o))$ where $\tilde{e}$ is the value of the internal energy molecule in the host model. Similarly, we calculated γ_R as $(\gamma(\tilde{e})/n_R)\tilde{c}_R$, where $\tilde{c}_R$ is the steady state concentration of the translation complex of the 'empty' ribosomes in the host model. We also removed competition for the p_R species by setting the host rRNA transcription rate ω_r to zero throughout. These modifications do not restore perfect integral action.

Concurrently, a series of design rules for creating *quasi*-integral action have been

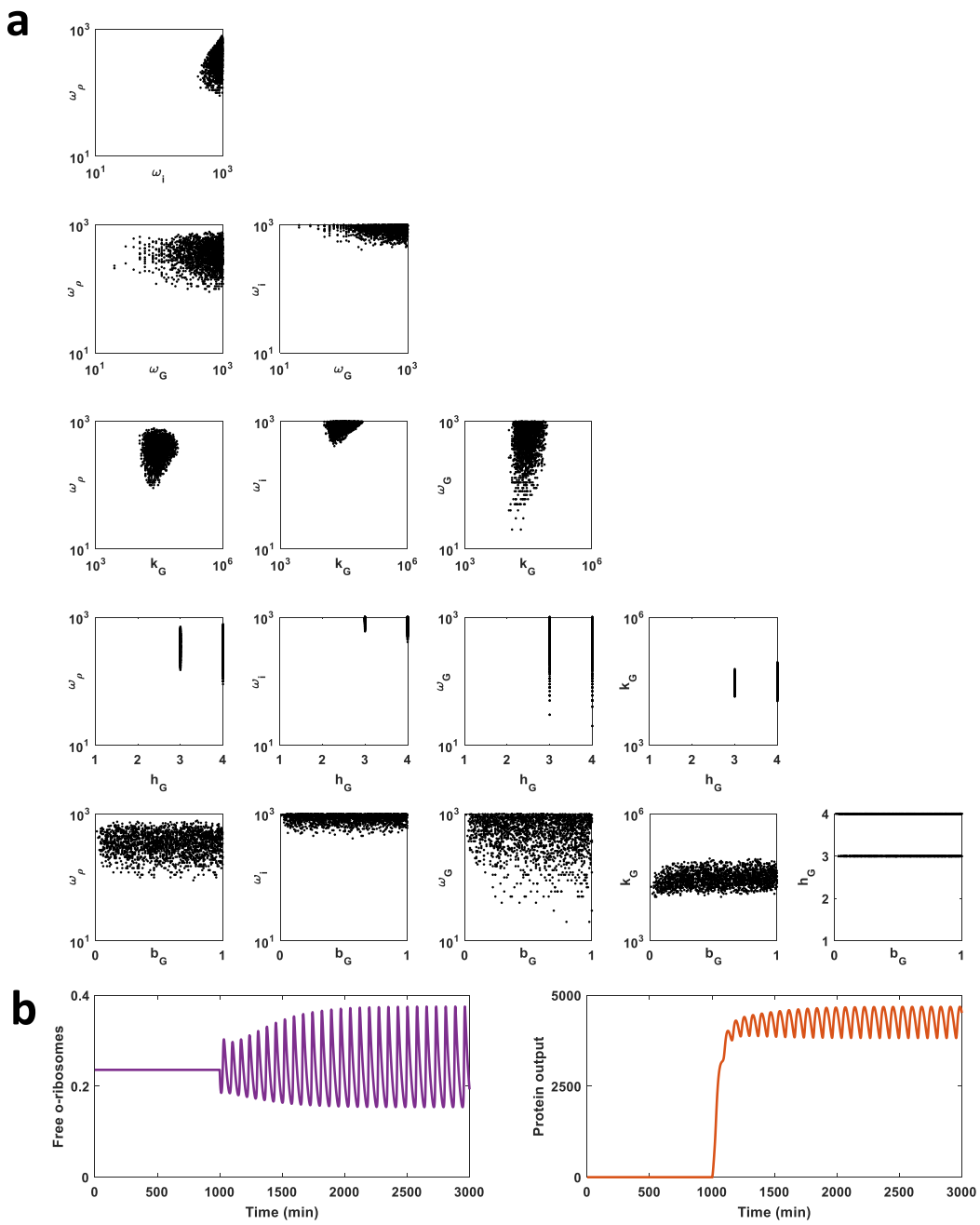

Figure 4.10: **Unstable regions of the proposed integral controller. (a)** The parameters describing the unstable controllers were identified and plotted pairwise against one another. **(b)** An unstable controller. The controller is simulated before a circuit gene is induced at $t = 10^3$ minutes with $\omega_A = 100$ mRNAs per min. Before induction the controller is stable but after the combined system of the controller and circuit shows sustained oscillations. *Left*, free o-ribosome number. *Right*, circuit protein expression. Controller parameters: $\omega_\rho = 500$ rRNAs per min; $\omega_i = 10^3$ RNAs per min; $\omega_G = 250$ mRNAs per min; Hill function constants $k_G = 10^4$ and $h_G = 4$.

published [91]. The authors show that the dilution of the sequestration components (here the o-rRNA and sRNA) results in a 'leak' effect which destroys the integral action of our proposed topology; as it results in decreases in o-rRNA and sRNA concentration that are not as a result of the sequestration/elimination reaction and so the 'comparison' function does not perfectly track the concentrations of the two species as they are not solely removed in the elimination reaction. They show that this controller topology functions as an integral controller in the absence of this dilution effect. As cell growth, and hence dilution are unavoidable, they develop a series of design rules to create what they term *quasi*-integral action. They demonstrate that increasing the transcription rates of the controller components and increasing the affinity of the RNA sensor for the reference species both act to increase sequestration of the target species and hence can partially restore the integral action.

To apply these observations in our context we first remove λ and set the protein decay rates $\delta_{p_R} = \delta_R = \delta_{p_G} = \delta_{p_A} = \delta_{p_B} = 0.022$ per min (this modification maintains the removal of the protein species). We then set the decay rates of the o-rRNA and sRNA to zero ($\delta_\rho = \delta_{r_i} = 0$). To prevent 'leak' of the controller species due to the reversibility of the reactions they are involved in we also set $u_\rho = u_i = 0$. In this context simply removing the loss of components due to dilution did not fully restore integral action and we still need to increase the transcription rate of the sequestering sRNA. (Figure 4.11b). (Note that if we re-consider competition for p_R by relaxing the $\omega_r = 0$ condition, sustained oscillations are observed (Figure 4.11b, *inset*))

If the assumption that the sRNA–o-rRNA and o-rRNA–'empty' ribosome binding reactions are reversible is restored (i.e. $u_\rho = u_i = 1$) then integral action at the higher transcription rate is not lost (Figure 4.11c). (Again, simulations containing competition for 'empty' ribosome by including the host rRNA results in sustained oscillations (Figure 4.11c, *inset*)).

Reinstating the RNA decay rates ($\delta_\rho = \delta_{r_i} = 0.1$ per min) results in loss of perfect integral action (Figure 4.11d)

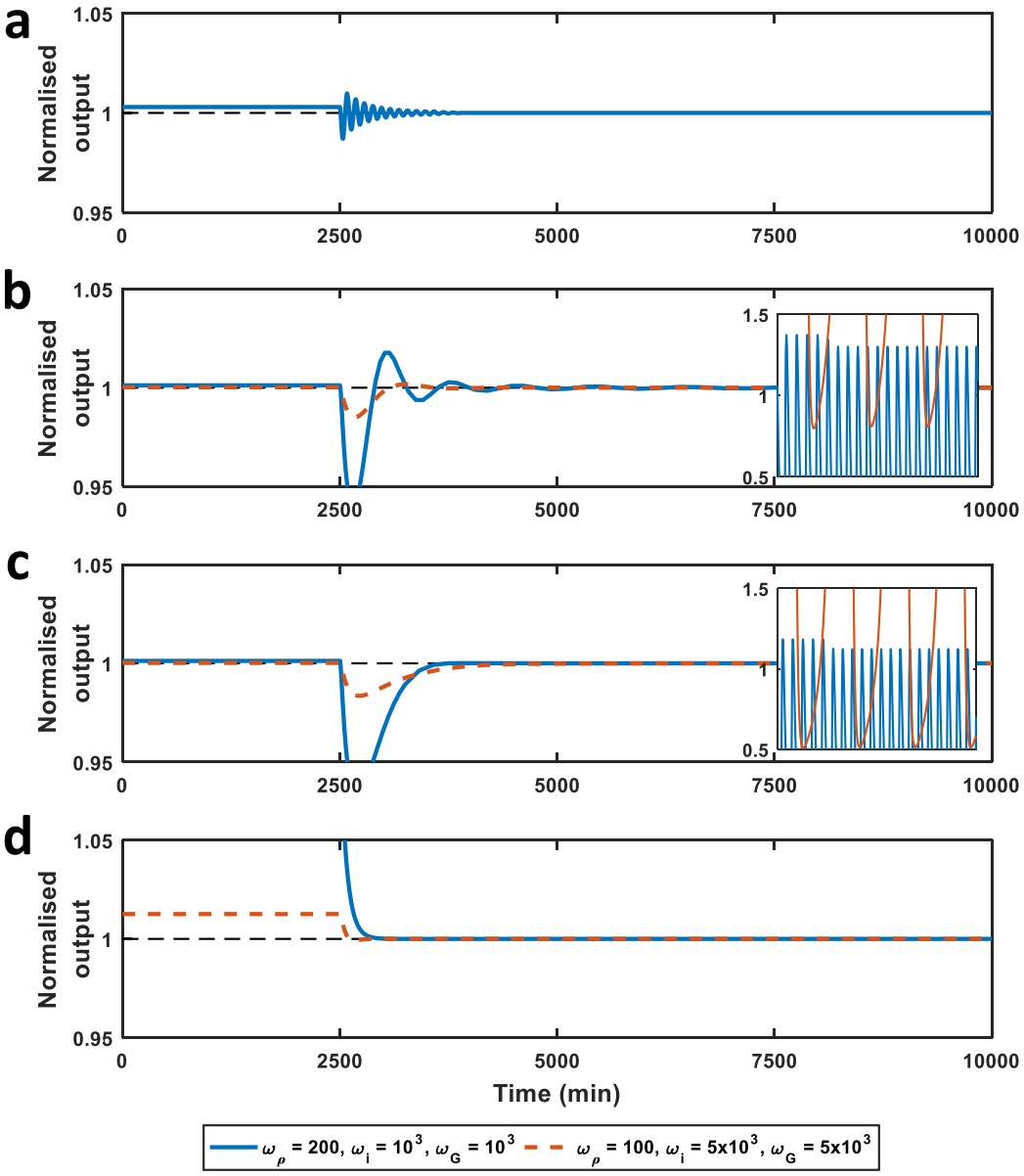

Figure 4.11: **Integral action is lost due to host-circuit interactions.** Simulations showing the change in p_A in response to the induction of a second gene p_B. Both genes induced at $\omega_A = \omega_B = 100$ mRNAs per min. p_B is induced at 2500 minutes. **(a)** Simulation of the full model with all host reactions. **(b)** Simulations of the reduced model. Host processes are reduced as described in the main text. λ approximated for protein species. Reversible reactions removed (i.e. $u_i = u_\rho = 0$). Inset, $\omega_r = 3,1750$ h-rRNAs per min. **(c)** Same simulation in (b) but with reversible reactions reinstated ($u_\rho = u_i = 1$). Inset, $\omega_r = 3,1750$ h-rRNAs per min. **(d)** Same simulations as in (b) but this 'leak' reactions reintroduced ($\delta_\rho = \delta_{r_i} = 0.1$ per min.)

The final reduced model which recreates integral action is given by:

$$\frac{d\tilde{p}_R}{dt} = \gamma_R - \delta_{p_R} \cdot \tilde{p}_R - b_r \cdot \tilde{p}_R \cdot \tilde{r} + u_r \cdot \tilde{R} - b_\rho \cdot \tilde{p}_R \cdot \tilde{\rho} + u_\rho \cdot \tilde{P} \tag{4.12}$$

$$\frac{d\tilde{r}}{dt} = \tilde{\omega}_r - \delta_{p_R} \cdot \tilde{r} - b_r \cdot \tilde{p}_R \cdot \tilde{r} + u_r \cdot \tilde{R} \tag{4.13}$$

$$\frac{d\tilde{R}}{dt} = b_r \cdot \tilde{p}_R \cdot \tilde{r} - u_r \cdot \tilde{R} - \delta_{p_R} \cdot \tilde{R} \tag{4.14}$$

$$\frac{d\tilde{\rho}}{dt} = \omega_\rho - \delta_\rho \cdot \tilde{\rho} - b_\rho \cdot \tilde{p}_R \cdot \tilde{\rho} + u_\rho \cdot \tilde{P} - b_i \cdot \tilde{\rho} \cdot \tilde{r}_i + u_i \cdot \tilde{r}_d \tag{4.15}$$

$$\frac{d\tilde{P}}{dt} = b_\rho \cdot \tilde{p}_R \cdot \tilde{\rho} - u_\rho \cdot \tilde{P} - \delta_{p_R} \cdot \tilde{P} ... + \sum_{X \in \{G,\ A,\ B\}} \Big(T_L(\tilde{c}_X,\ \tilde{e}) - b_X \cdot \tilde{P} \cdot \tilde{m}_X + u_X \cdot \tilde{c}_X \Big) \tag{4.16}$$

$$\frac{d\tilde{r}_i}{dt} = \omega_i \cdot F(\tilde{p}_G) - \delta_{r_i} \cdot \tilde{r}_i - b_i \cdot \tilde{r}_i \cdot \tilde{\rho} + u_i \cdot \tilde{r}_d \tag{4.17}$$

$$\frac{d\tilde{r}_d}{dt} = -\delta_{r_d} \cdot \tilde{r}_d + b_i \cdot \tilde{r}_i \cdot \tilde{\rho} + u_i \cdot \tilde{r}_d \tag{4.18}$$

where $F(p_G)$ symbolises an activatory Hill function. Note that $d\tilde{r}/dt$ and $d\tilde{R}/dt$ can be neglected if ω_r is set to zero.

The equations describing the protein-encoding genes ($X \in \{G,\ A,\ B\}$) are:

$$\frac{d\tilde{m}_X}{dt} = \tilde{\omega}_X - \delta_{m_X} \cdot \tilde{m}_X + T_L(\tilde{c}_X,\ \tilde{e}) - b_X \cdot \tilde{P} \cdot \tilde{m}_X + u_X \cdot \tilde{c}_X \tag{4.19}$$

$$\frac{d\tilde{c}_X}{dt} = -\delta_{p_R} \cdot \tilde{c}_X - T_L(\tilde{c}_X,\ \tilde{e}) + b_X \cdot \tilde{P} \cdot \tilde{m}_X - u_X \cdot \tilde{c}_X \tag{4.20}$$

$$\frac{d\tilde{p}_X}{dt} = T_L(\tilde{c}_X,\ \tilde{e}) - \delta_{p_X} \cdot \tilde{p}_X \tag{4.21}$$

4.5 Development of an integral controller with improved robustness

Given the instability observed in the G-controller design, we revised our proposed integral controller design. In Section 4.4, our o-rRNA sequestration scheme requires the activation of gene expression. This can result in a threshold and then amplification effect due to the action of the Hill function used to model p_G transcription factor action. The amplification effect can potentially lead to instabilities. We reconsidered the integral architecture to develop a controller design which does not include this effect by not using gene activation.

RNAs can be bound by small RNA binding proteins. Biologically these proteins have a range of functions in post-transcriptional processing such as modulating mRNA stability or occluding the RBS to modulate translation (reviewed in [92, 93]). We developed a new controller architecture (which we called K) by considering the use of RNA binding proteins as a means of o-rRNA sequestration (Equation 4.22). Again the design relies on the constitutive expression of this protein, making its levels indicative of the circuit demand due to resource-mediated coupling.

$$\rho + p_K \underset{u_i}{\overset{b_i}{\rightleftharpoons}} d_K \tag{4.22}$$

4.5.1 Development of the model

As before we introduced the necessary equations to describe the interaction between the o-rRNA and RNA binding proteins (the latter denoted K). We model the binding reaction by applying the Law of Mass Action. We assume this reaction can be reversible and is diffusion-limited (i.e. the maximal binding/unbinding rate is 1). We modify the o-rRNA equation to include the new sequestration reaction (highlighted in blue):

$$\begin{aligned}\frac{d\rho}{dt} = \omega_\rho \cdot \left(\frac{e}{o_\rho + e}\right) - b_\rho \cdot p_R \cdot \rho + u_\rho \cdot P... \\ - (\delta_\rho + \lambda) \cdot \rho \textcolor{blue}{- b_i \cdot \rho \cdot p_K + u_i \cdot d_K}\end{aligned} \tag{4.23}$$

The mRNA and translation complex associated with the new RNA binding protein follow the same dynamics as for other protein-coding genes (Equations 3.7). The dynamics of the RNA binding protein follow those of other proteins but are modified to include the association with the sRNA:

$$\frac{dp_K}{dt} = T_L(c_K, e) - (\delta_{p_K} + \lambda) \cdot p_K - b_i \cdot \rho \cdot p_K + u_i \cdot d_K \tag{4.24}$$

The dynamics of the rRNA-RNA binding protein complex, d_K, is given by:

$$\frac{dd_K}{dt} = b_i \cdot \rho \cdot p_K - u_i \cdot d_K - (\delta_{d_K} + \lambda) \cdot d_K \tag{4.25}$$

Note that we assume this complex is stable (hence $\delta_{d_K} = 0$ as for all other proteins) and so is removed only by dilution due to growth (λ term).

4.5.2 The K-controller successfully decouples two genes

We optimised the o-rRNA and RNA binding protein transcription rates (ω_ρ and ω_K) using the cost function as outlined in Section 4.2.

To demonstrate the function of the controller we assessed its ability to maintain the expression level of one gene in response to the induction of a second considering the point where approximately equal amounts of protein are produced ($\omega_A = \omega_B = 100$ mRNAs per min). Before induction of the second gene, competition between the circuit and controller is low so controller expression is high (depicted in Figure 4.12a, *left*). This results in high sequestration of o-rRNA in low demand circuits (Figure 4.12b). Upon induction of the second gene (depicted in Figure 4.12a, *right*), competition for o-ribosomes increases which reduces expression of the RNA binding protein (p_K) despite it being a constitutively expressed gene (Figure 4.12b). This results in decreased sequestration of the o-rRNA. The levels of o-rRNA increase, increasing co-option of ribosomes from the host to the orthogonal pool. Therefore as circuit mRNAs increase due to additional gene induction the demand is matched by the supply of o-ribosomes and so the levels of the original protein are maintained (Figure 4.12c). Note that this does not function as a perfect integral controller as an error of 1% remains. The free o-ribosome falls less than 5% upon induction of p_A.

To assess the function of the controller, we simulate its ability to decouple co-expressed genes in a simple two gene circuit as before (where p_A is induced and p_B induction is held constant). As before we tune ω_ρ in the absence of the controller to create an open loop comparison. The controller successfully restores the mRNA–protein relationship for the majority of ω_A values. In the absence of the controller, p_B falls by nearly 50% across the first two orders of magnitude of ω_A whilst in the controlled system this fall is less than 2%. From $\omega_A = 10$ mRNAs per min to $\omega_A = 10^3$ mRNAs per min the constitutive gene falls only 2% while the induced gene increases linearly with increasing induction (Figure 4.13a). However, at high p_A induction ($\omega_A > 10^3$ mRNAs per min) the controller fails with both protein levels rising dramatically. Closer inspection of the simulations reveals that at these levels of induction the growth rate falls to zero (Figure 4.13b, *right*). In this case p_K is saturated by the o-rRNAs and so the high value of ω_ρ means that the number of o-ribosomes increases. Orthogonal ribosomes get sequestered into circuit translation complexes (i.e. c_A and c_B) (Figure 4.13b, *left*). Therefore the number of free ribosomes falls to zero, as 'empty' ribosomes are not recycled (Figure 4.13b, *left*).

This increases competition between host and orthogonal rRNAs which is likely to account for the strong coupling observed at $\omega_A > 10^3$ mRNAs per min (Figure 4.13b, *right*).

4.5.3 Robustness analysis and extension to different circuits

We carried out a simple robustness analysis as before by drawing 10,000 random numbers from a uniform distribution with ±50% of the value produced by the optimisation routine. All parameter sets were successfully simulated to steady state. The topology is highly stable for $\omega_A < 10^3$ with most parameter sets showing significant decoupling ability. However, across all levels of p_A induction only 35% of the controllers tested fall within 50% of the original function (although note this increases to 50% for $\omega_A \leq 10^3$ mRNAs per min).

We supplemented this analysis by simulating 10^6 random controllers (created by drawing parameter designs from between the minimum and maximum permissible values). As before we simulated these controllers in the presence and absence of a simple circuit consisting of one gene induced at 100 mRNAs per min. In all cases these controllers produce stable, although as expected highly variable, results. This suggests that the topology is highly stable and designs are unlikely to be unstable or form sustained oscillations. We can, however, expect poor levels of robust performance from this design.

Our controller design process aims to find the optimum controller over a range of ω_A and ω_B values (see Section 4.2). As before we simulated a range of ω_B levels, holding ω_B constant at each level while varying ω_A. The controller successfully decouples genes across the range of values tested (Figure 4.14b). However, as observed in both Figures 4.13 and 4.14a, there are significant growth rate effects at high p_A induction ($\omega_A > 10^3$ mRNAs per min) and high p_B induction ($\omega_B > 10^3$ mRNAs per min). At high demand the growth rate falls to near zero resulting in accumulation of protein and increased coupling due to the falling levels of free ribosomes (Figure 4.13b).

Varying the design parameters of the controller acts to move the region of zero-growth to the left (i.e. making poor growth occur at lower inductions). Lowering growth rate significantly *in vivo* is highly problematic (for example, small populations are undesirable in a biotechnological setting due to the associated poor product yield). In light of these initial observations, and as we do not yet have a complete solution, we did not carry this controller forward for further sensitivity analysis.

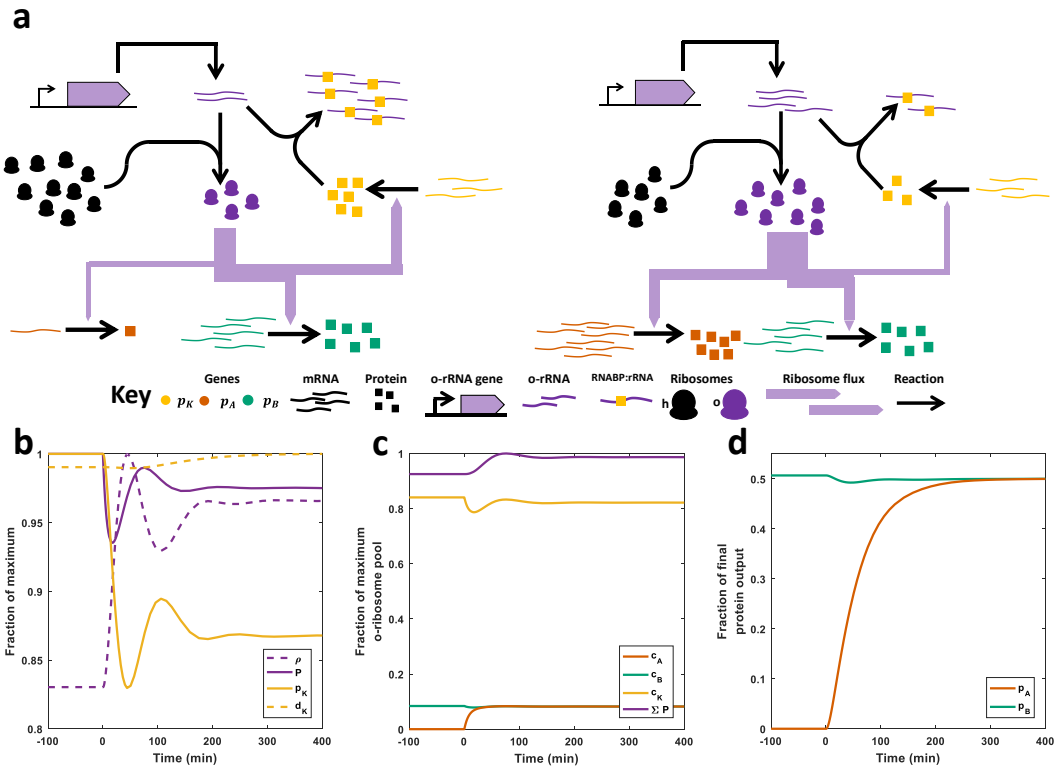

Figure 4.12: **Operation of the K-controller.** Simulations of a two-gene circuit were carried out as described in the main text. p_A is induced at $t = 0$. The changing distribution of controller and circuit components is shown in (b), (c) and (d). **(a)** Structure and function of the K-controller. *Left,* Low demand circuit. When competition is low, p_K expression is high and so o-rRNA sequestration by the RNA binding protein is high. Therefore co-option of ribosomes to the o-ribosome pool is low. *Right,* High demand circuit. As circuit demand increases (as p_A is induced), the o-ribosome pool redistributes across circuit and controller genes (width of purple ribosome flux lines) due to competition between mRNAs. This reduces translation of p_K and hence sequestration of the o-rRNA. This increased co-option of ribosomes to the orthogonal pool and maintains ribosome flux for m_B translation despite the increase in m_A. **(b)** Changing distribution of the controller components. ρ, o-16S rRNA; P, free o-ribosome; p_k, controller RNA binding protein; d_K, o-rRNA sequestered by controller protein. Normalised by their maximum value. **(c)** Changing distribution of the translation complexes over time in response to p_A induction. c_Y, translation complex of gene Y, Σ P, sum of all o-ribosomes. Normalised by maximum Σ P. **(d)** Protein output over time normalised by sum of the final circuit protein concentration.

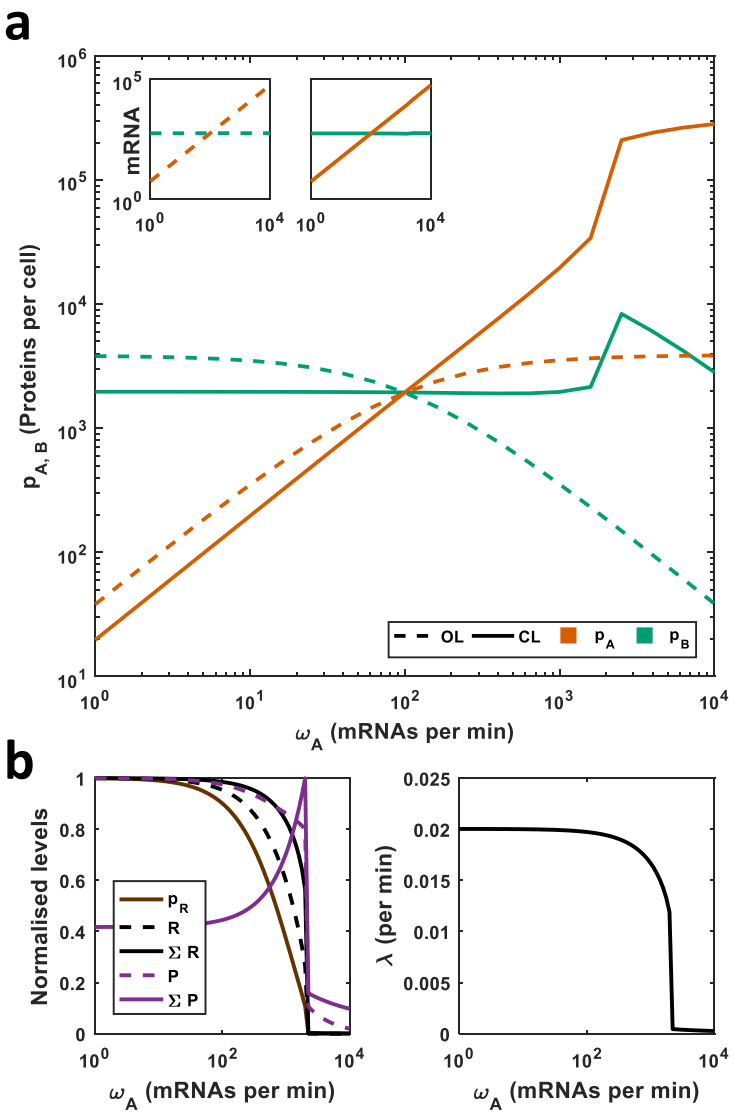

Figure 4.13: **The K-controller decouples co-expressed genes below $\omega_A = 10^3$ mRNAs per min.** Simulation of the action of new controller architecture. ω_A is varied between 1 and 10^4 mRNAs per minute. ω_B is held constant at 100 mRNAs per minute throughout. The simulation time span is increased until it reaches steady state. Controller parameters: $\omega_\rho = 500$ rRNAs per min; $\omega_K = 10^3$ mRNAs per min. Open loop $\omega_\rho = 1.245$ rRNAs per min. **(a)** The controller successfully decouples co-expressed genes. Insets, *left,* mRNA concentration in the open loop; *right,* mRNA concentration in the closed loop. **(b)** *Left*, Changing distribution of ribosomes, normalised by their initial level at $\omega_A = 10$ mRNAs per min. Legend explanation: p_R, 'empty' protein component of the ribosomes; R, free host ribosomes; $\Sigma\ R$, translating host ribosomes; P, free o-ribosomes; $\Sigma\ P$, translating o-ribosomes. *Right*, growth rate.

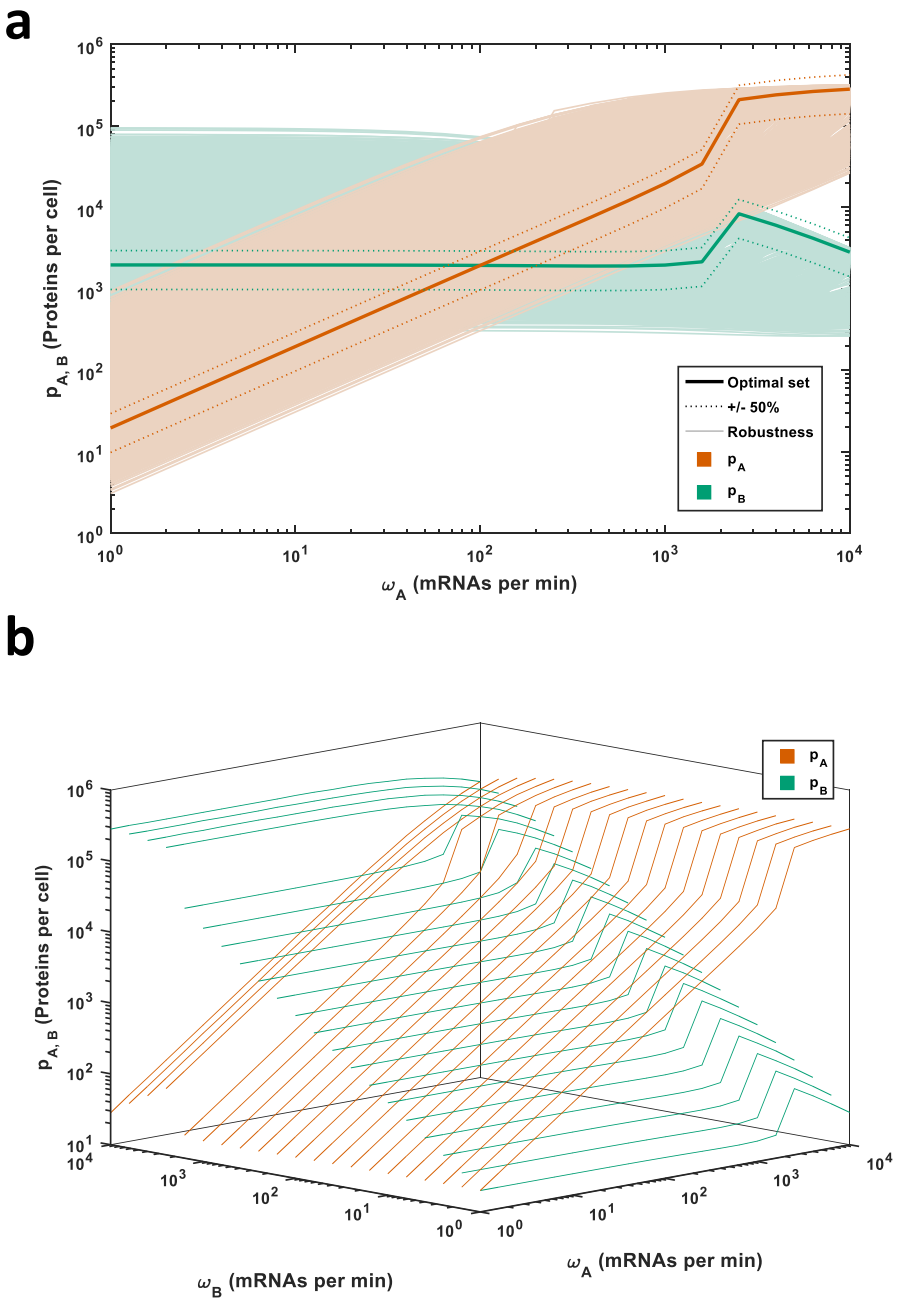

Figure 4.14: **Robustness of the new integral controller.** **(a)** Function and robustness of the controller decoupling two genes. The optimal controller parameters were perturbed by drawing random values between ±50% of the original value $N = 10,000$. All parameters controlling o-rRNA and controller protein were allowed to vary. For ease of analysis the steady state values within 50% of the optimal set are shown as dotted lines. **(b)** Function of the controller over a range of p_A and p_B inductions.

4.6 *In vivo* implementation of a prototype controller

Comparison of the three controllers shows that the F-controller (from Section 4.3) is more stable than the integral controller (the G-controller from Section) 4.4) and has higher performance than the modified integral controller (the K-controller from Section 4.5). Therefore, our experimental collaborators implemented a prototype of the F-controller *in vivo*. In addition to having the best predicted performance, this controller is the easiest to implement biologically as it consists simply of a protein regulating the o-rRNA promoter and its behaviour can be tuned with commonly available parts (e.g. by varying promoter strengths). Implementation of the first integral G-controller is complicated by the need to design *de novo* the sequence of the sRNA as well as the potential instabilities identified by our analysis. There are also more parameters which require tuning; whilst advantageous for detailed design this complicates production of a prototype. Simulations of the RNA binding protein K-controller highlight its poor performance and the production of an RNA binding protein specific to the o-rRNA of interest is likely to require significant protein design, further complicating the prototyping phase.

Using our sensitivity analysis as a guide, the design was based on the strongly binding LacI repressor ($k_D \approx 0.02$ nM, [94]), which also shows a highly non-linear mode of action due to the dimerisation steps required to produce the functional tetramer (Figure 4.5c, d). As described in Chapter 3, in the experimental system the production of the o-16S rRNA is inducible with IPTG via the host bacterium's endogenous LacI protein. To implement the negative feedback loop in *E. coli* the endogenous *lacI* gene was deleted using conventional techniques and an exogenous copy of the *lacI* gene was inserted into the o-16S rRNA carrying plasmid. From our results in Figure 4.5b, the strong P_{LacI^q} promoter was selected to drive LacI transcription [95]. Whilst the use of a stronger RBS is desirable, experimental limitations mean the same orthogonal RBS was used. Isocost lines for different IPTG concentrations were produced as described in Section 3.10.

To evaluate the function of the prototype controller we compared the open and closed loop experimental designs which produce equivalent initial GFP number (i.e. they have the same initial demand for ribosomes). The 0.2 mM IPTG treatment of the o-RFP, o-GFP circuit (i.e. the open loop system) has an isocost line gradient of -1.0, so one GFP unit is lost for each RFP gained. Upon addition of the controller (i.e. the 0.5 mM IPTG treatment of the close loop system) this slope increases to -0.5 (Figure 4.15). This represents a halving in the amount of coupling. Tuning

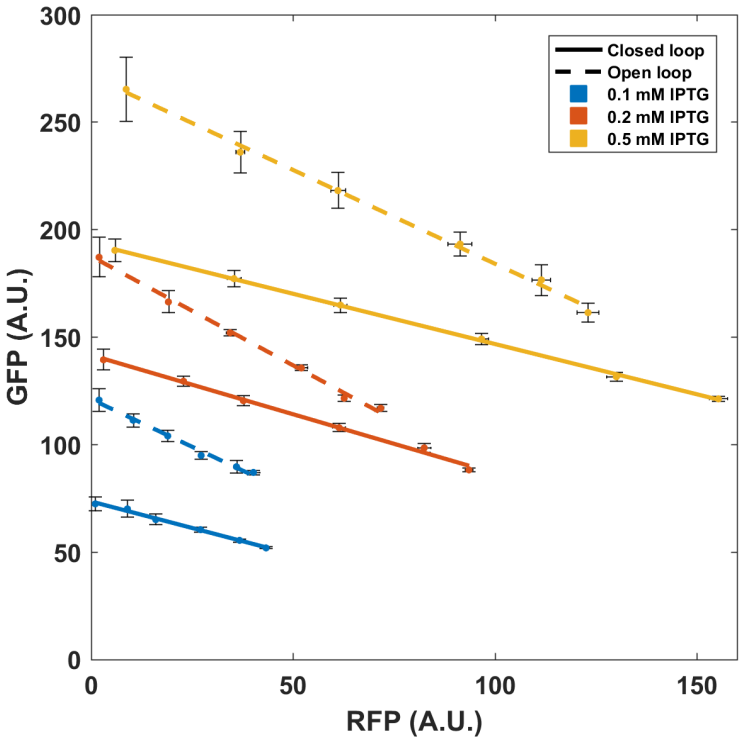

Figure 4.15: ***In vivo* implementation of the prototype F-controller.** Response of constitutively expressed o-GFP as o-RFP is induced. o-RFP was induced using AHL from 0 to 20 nM. Points are the mean steady state fluorescence ± 1 S.D. as determined by FACS from cultures during mid-exponential growth (between 3-5 hours post-induction dependent on the strain and circuit). $N = 3$. The level of gene expression from the controller can be tuned with IPTG as shown (CL, closed loop). For comparison the o-RFP, o-GFP results which produce comparable initial GFP levels are also shown (OL, open loop). Isocost lines are fit to mean values using the *polyfit* function in MATLAB.

the controller threshold with IPTG allows the tuning of protein levels at no cost to decoupling (consistent with the model prediction in Figure 4.5a).

4.7 Conclusions

In this chapter, we have used our model developed in Chapter 3 to test the feasibility of different controller systems, which act to dynamically distribute ribosomes between host and orthogonal pools depending upon circuit demand for o-ribosomes.

We first considered the simplest method of implementing a feedback controller, called the F-controller, by modelling a protein which both uses o-ribosomes for its translation and inhibits o-rRNA production. Our analysis suggests this controller is highly stable, shows reasonable performance and is able to function across a range of two-gene circuits. However, the circuit does not fully decouple co-expressed genes meaning that a small steady state error persists.

Addition of integral action to a controller provides a means of removing steady state errors and so increasing circuit performance. We therefore implemented two proposed integral controllers (G and K) based on a recently published controller topology. However, we find niether controller demonstrates perfect integral action and analysis of the G-controller suggests this may be due to decay of the o-rRNA due to RNA turnover. Our robustness analysis demonstrates that the G-controller is unstable, especially in the presence of high demand circuits. Analysis of the second integral K-controller shows that whilst it is stable it shows poor performance as at high circuit induction, the growth rate becomes negligible suggesting the action of this controller may lead to over co-option of ribosomes from the host and therefore toxic growth effects.

Working with our experimental colleagues we designed a prototype negative feedback F-controller using the results of our sensitivity analysis. When implemented *in vivo* this prototype is able to decouple co-expressed genes by approximately 50% (in comparison to the open loop confirmation of equivalent initial protein level). This decoupling ability is robust with the slope of the isocost line remaining constant even as protein output is tuned with IPTG.

Chapter 5

Mechanistic modelling and design of an optimal negative feedback controller

5.1 Introduction

In this chapter, we re-consider the mechanism of the F-type negative feedback controller and develop a detailed model of its function. Having demonstrated the feasibility of a translational controller using a host-aware model at a high level of abstraction and shown that a prototype controller functions as expected, we desire a model which can be used for component selection and detailed design. Given the biological insights we have gleaned from the prototype *in vivo* implementation (such as the host growth rate, cell size, that controller action does not cause significant changes in growth rate etc.), we can now better estimate the effect of the circuit on the host and are able to parameterise a more detailed mechanistic model. We can now estimate the values of previously unknown parameters; for example, we find across the circuits tested experimentally that all have an approximate growth rate of 1 h^{-1}. This allows us to determine, for example, ribosome concentration, from previously published works.

This new model is of an intermediate level of abstraction between an isolated circuit model (presented in Section 2.5), which neglects all host resources, and the host-aware framework, which is a high-dimensional and highly non-linear model (as depicted in Chapter 3). This new model also re-expresses concentrations in the more

conventional units of nanomolar rather than molecules per cell. The advantage of using this modelling approach is that all the parameters relate directly to biological 'parts' or components, such as promoter strengths and copy number (which are lumped into the ω terms in Chapter 3). Additionally, on/off rates are reformulated as dissociation constants (see discussion in Section 5.3). This allows specific components to be identified by conducting an extensive literature search. Additionally, smaller models of this type are more amenable to analysis using conventional techniques from control and systems theory. These models of intermediate complexity are increasingly used in this field as resource limitations become apparent [1, 38, 96].

We first re-derive a model of the F-controller and then apply techniques from systems biology to reduce the complexity of the model and demonstrate its utility as a design tool. We identify hard trade-offs between gene coupling, gene expression and dynamics and find simple design rules which can be used to manage them. We demonstrate how our negative feedback controller functions when allocating resources in a range of dynamic circuits.

We reformulate our circuit and controller as a block diagram to allow the identification of key components of the controller (Figure 5.1). This is referred to below.

5.2 Derivation of a mechanistic model

We first derive the mechanistic gene expression model of the circuit and controller system. Note the renaming of certain components from the nomenclature in Chapters 3 and 4. See Table 5.1 for new notation.

We assume that each circuit promoter (g_i) can be bound by a multimeric transcription factor (u_i) to form a promoter complex (κ_i) capable of recruiting a free RNA polymerase (σ) to form a transcription complex. When transcription occurs, an mRNA (m_i) is produced, and the original RNA polymerase and promoter complex are released. The above interactions are described by the following chemical reactions:

$$g_i + \eta_i \cdot u_i \underset{\alpha_{r_i}}{\overset{\alpha_{f_i}}{\rightleftharpoons}} \kappa_i \qquad \kappa_i + \sigma \underset{\xi_{r_i}}{\overset{\xi_{f_i}}{\rightleftharpoons}} x_i \qquad x_i \xrightarrow{\tau_i} m_i + \sigma + \kappa_i \tag{5.1}$$

The mRNA is bound by a free (orthogonal) ribosome, R, to form a translation complex (c_i). Upon translation, a protein (p_i) is produced and the mRNA and R

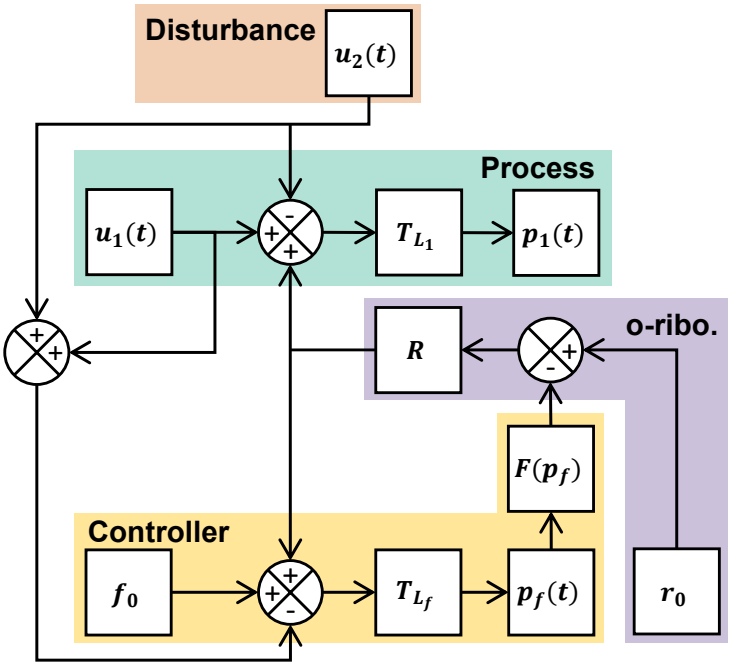

Figure 5.1: **Block diagram of the controller.** The process, highlighted in green, converts the input u_1 into protein output p_1 utilising the o-ribosome pool R. The input into a second process (output not shown) u_2 acts as a disturbance to the first process which is ameliorated by the effect of the controller. The controller protein is constitutively expressed (f_0 signal) so the output p_f is dependent upon R. As inputs u_i disturb R the level of p_f changes (i.e. as u_i increases, p_f decreases). As p_f is a repressor the disturbance signal is inverted in the $F(p_f)$ block.

Species	Symbol
RNA polymerase	σ
Orthogonal ribosome	R
Host ribosome	R_h
Orthogonal 16S rRNA	r
Circuit input transcription factor	u_i
Promoter conc. of gene i	g_i
Gene i promoter - transcription factor complex	κ_i
Gene i transcription complex	x_i
mRNA of gene i	m_i
mRNA of gene i translation complex	c_i
Protein i	p_i

Table 5.1: **Notation for molecular species in Chapter 5**

are released:

$$m_i + R \underset{\beta_{r_i}}{\overset{\beta_{f_i}}{\rightleftharpoons}} c_i \qquad\qquad c_i \xrightarrow{\gamma_i} m_i + R + p_i \tag{5.2}$$

Additionally, both mRNAs and proteins degrade at rates δ_{m_i} and δ_{p_i}, respectively:

$$m_i \xrightarrow{\delta_{m_i}} \emptyset \qquad\qquad p_i \xrightarrow{\delta_{p_i}} \emptyset \tag{5.3}$$

Applying the Law of Mass Action we derive the following ODEs describing the time evolution of the circuit components:

$$\dot{g}_i = -\alpha_{f_i} \cdot g_i \cdot {u_i}^{\eta_i} + \alpha_{r_i} \cdot \kappa_i \tag{5.4}$$

$$\dot{\kappa}_i = \alpha_{f_i} \cdot g_i \cdot {u_i}^{\eta_i} - \alpha_{r_i} \cdot \kappa_i - \xi_{f_i} \cdot \kappa_i \cdot \sigma + \xi_{r_i} \cdot x_i + \tau_i \cdot x_i \tag{5.5}$$

$$\dot{x}_i = \xi_{f_i} \cdot \kappa_i \cdot \sigma - \xi_{r_i} \cdot x_i - \tau_i \cdot x_i \tag{5.6}$$

$$\dot{m}_i = \tau_i \cdot x_i - \beta_{f_i} \cdot m_i \cdot R + \beta_{r_i} \cdot c_i + \gamma_i \cdot c_i - \delta_{m_i} \cdot m_i \tag{5.7}$$

$$\dot{c}_i = \beta_{f_i} \cdot m_i \cdot R - \beta_{r_i} \cdot c_i - \gamma_i \cdot c_i \tag{5.8}$$

$$\dot{p}_i = \gamma_i \cdot c_i - \delta_{p_i} \cdot p_i \tag{5.9}$$

This represents a simple single-input-single-output (SISO) motif and forms the basis of our model. Complex circuits can be constructed by letting the output from one module form the input to another, as discussed in Section 5.8.

To implement our controller we first consider the conversion of host ribosomes (R_{host}) into circuit-specific orthogonal ribosomes (R). The orthogonal 16S rRNA

gene promoter (g_r) recruits σ to form a translation complex (x_r) which produces the orthogonal rRNA (r):

$$g_r + \sigma \underset{\xi_{r_r}}{\overset{\xi_{f_r}}{\rightleftharpoons}} x_r \qquad\qquad x_r \xrightarrow{\tau_r} r + \sigma + g_r \tag{5.10}$$

The orthogonal 16S rRNA binds host ribosomes, R_h, and so recruits ribosomes to the circuit-only orthogonal pool, R:

$$R_h + r \underset{\varrho_r}{\overset{\varrho_f}{\rightleftharpoons}} R \tag{5.11}$$

In the presence of the controller the orthogonal rRNA gene is regulated by the repressor p_f. The η_f repressor monomers bind the free g_r promoter and prevents the binding of RNA polymerase and associated factors (σ in our model). (Note that in reality repressor monomers will assemble into a functional multimeric complex before binding but this does not significantly affect the structure of the ODE model and so we ignore it.)

$$g_r + \eta_f \cdot p_f \underset{\alpha_{f_r}}{\overset{\alpha_{f_f}}{\rightleftharpoons}} \kappa_f \tag{5.12}$$

We model expression of the regulator protein by considering the constitutive expression of its mRNA from an unregulated promoter, g_f:

$$g_f + \sigma \underset{\xi_{f_r}}{\overset{\xi_{f_f}}{\rightleftharpoons}} x_f \tag{5.13}$$

We model the transcription and translation of the repressor's mRNA and protein in the same manner as the circuit genes, as described above.

Applying the Law of Mass Action results in the following ordinary differential equations describing the production of the repressor and intermediate complexes:

$$\dot{g}_f = -\xi_{f_f} \cdot g_f \cdot \sigma + \xi_{r_f} \cdot x_f + \tau_f \cdot x_f \tag{5.14}$$

$$\dot{x}_f = \xi_{f_f} \cdot \kappa_f \cdot \sigma - \xi_{r_f} \cdot x_f - \tau_f \cdot x_f \tag{5.15}$$

$$\dot{m}_f = \tau_f \cdot x_f - \beta_{f_f} \cdot m_f \cdot R + \beta_{r_f} \cdot c_f + \gamma_f \cdot c_f - \delta_{m_f} \cdot m_f \tag{5.16}$$

$$\dot{c}_f = \beta_{f_f} \cdot m_f R - \beta_{r_f} \cdot c_f - \gamma_f \cdot c_f \tag{5.17}$$

$$\dot{p}_f = \gamma_f \cdot c_f - \delta_{p_f} \cdot p_f - \eta_f \cdot \alpha_{f_r} \cdot g_r \cdot {p_f}^{\eta_f} + \eta_f \cdot \alpha_{r_r} \kappa_r \tag{5.18}$$

Applying the Law of Mass Action to the o-rRNA promoters and ribosome species yields:

$$\dot{g}_r = -\xi_{f_r} \cdot g_r \cdot \sigma + \xi_{r_r} \cdot x_r + \tau_r \cdot x_r - \alpha_{f_r} \cdot g_r \cdot {p_f}^{\eta_f} + \alpha_{r_r} \cdot \kappa_r \tag{5.19}$$

$$\dot{x}_r = \xi_{f_r} \cdot g_r \cdot \sigma - \xi_{r_r} \cdot x_r - \tau_r \cdot x_r \tag{5.20}$$

$$\dot{\kappa}_r = \alpha_{f_r} \cdot g_r \cdot {p_f}^{\eta_f} + \alpha_{r_r} \cdot \kappa_r \tag{5.21}$$

$$\dot{r} = \tau_r \cdot x_r - \delta_r \cdot r - \varrho_f \cdot r \cdot R_H + \varrho_r \cdot R \tag{5.22}$$

$$\dot{R}_{host} = -\varrho_f \cdot r \cdot R_{host} + \varrho_r \cdot R \tag{5.23}$$

5.3 Model reduction

This full model is highly complex containing many species and many forward and reverse reaction rates.

On inspection of the model we can see that whilst the promoters change state they are not created or destroyed and so their total number for each gene remains constant (i.e. total promoter concentration is conserved) such that:

$$g_{i,T} = g_i + \kappa_i + x_i \tag{5.24}$$

$$g_{r,T} = g_r + x_r + \kappa_r \tag{5.25}$$

$$g_{f,T} = g_f + x_f \tag{5.26}$$

The total number of ribosomes is also conserved as, again, ribosomes only change state such as by binding mRNA or being converted into an orthogonal ribosome:

$$R_{Total} = R_h + R + c_f + \sum_{i=1}^{N} \Big(c_i \Big) \tag{5.27}$$

By applying these conservation laws we are able to reduce the number of species which must be tracked in the model. For example, from Equation 5.26, we need only simulate the behaviour of g_f because $x_f = g_{f,T} - g_f$ as the total promoter concentration is conserved.

Individual binding and unbinding rates for genetic components and proteins are rarely reported in the literature, due to the difficulty in determining their values experimentally because of the effects of confounding factors (such as additional fluxes along the reaction pathway). However, by considering the time-scale separation of

the reactions involved, we are able to reduce our model and formulate the forward and reverse reactions in terms of 'lumped' dissociation constants. By defining the model in terms of effective dissociation constants, we are then able to determine suitable genetic components with the desired dynamics from a search of previously published data. The effective dissociation constants of the RNA polymerase for the promoter (k_X), of the ribosome for the RBS (k_L) and of the protein for its binding site (μ) are defined as follows:

$$k_X = \frac{\xi_r + \tau}{\xi_f} \qquad k_L = \frac{\beta_r + \gamma}{\beta_f} \qquad \mu = \frac{\alpha_r}{\alpha_f} \tag{5.28}$$

This time-scale separation results from the fact that different biological processes occur over a range of different time spans with binding/unbinding reactions occurring on the order of milliseconds, transcription and translation taking minutes and protein degradation/dilution occurring over tens of minutes or hours [31]. This effectively 'separates' reactions in time and allows us to assume that 'fast' species, such as mRNAs, reach a (quasi)-steady state (QSS) instantaneously. By calculating the QSS concentrations of intermediate species and substituting as appropriate, we are able to remove the majority of intermediate gene expression species from our model. Below, we denote the QSS complex of species y as $\bar{y}$.

Since current experimental evidence suggests that competition for RNA polymerases does not significantly limit gene expression, we remove RNA polymerase mediated competition by considering each gene to have access to its own small local pool of RNA polymerase (e.g. [33, 38]). Additionally, we assume that the dissociation constant for RNA polymerase is much higher than the concentration of free polymerase, consistent with experimental observations [97]. This allows us to reduce the complexity of the expressions by assuming that:

$$\sigma + k_X \approx k_X \tag{5.29}$$

This follows a similar approach used in [38] to reduce complexity of the transcriptional expressions.

By applying these assumptions we can simplify Equations 5.4 to 5.9 as follows:

The QSS of the activated circuit gene ($\bar{x}_i$) which gives rise to mRNAs is solely a function of the input (u_i):

$$\bar{x}_i = \frac{\sigma_{Total}}{1 + (1/\hat{x}_i)} \tag{5.30}$$

where $\hat{x}_i$ can be considered as a measure of demand for RNA polymerase by gene i and is defined as follows:

$$\hat{x}_i = \frac{g_{i,T}}{k_{X_i}} \frac{{u_i}^{\eta_i}}{\mu + {u_i}^{\eta_i}} \tag{5.31}$$

The ODE describing the time-evolution of the protein species is:

$$\dot{p}_i = \gamma_i \left(\frac{\tau_i}{\delta_{m_i}} \frac{R}{k_{L_i}} \bar{x}_i \right) - \delta_{p_i} \cdot p_i \tag{5.32}$$

We can also define the constant $\hat{c}_i$:

$$\hat{c}_i = \frac{1}{k_{L_i}} \frac{\tau_i}{\delta_{m_i}} \bar{x}_i \tag{5.33}$$

This functions as a measure of demand for ribosomes by gene i as it incorporates both a proxy RBS strength ($1/k_{L_i}$) and mRNA QSS concentration. The latter is the steady state promoter concentration $\bar{x}_i$ multiplied by the net mRNA production rate (τ_i/δ_{m_i}).

If a single unregulated pool of ribosomes is used for circuit expression then the number of free (host/orthogonal) ribosomes, R, available for circuit translation is given by:

$$R = \frac{R_{Total}}{1 + \sum_{i=1}^{N} \left(\hat{c}_i \right)} \tag{5.34}$$

Therefore the response of p_1 depends not only on the input u_1 but also the demand for ribosomes by other genes $\hat{c}_{i,\ i \neq 1}$. This forms the basis of our circuit SISO 'process' model (the green process block in Figure 5.1).

Applying the same assumptions to the equations describing the production of the regulator p_f we have that:

$$\dot{p}_f = \gamma_f \left(\frac{\tau_f}{\delta_{m_f}} \frac{R}{k_{L_f}} \bar{x}_f \right) - \eta_f \cdot \bar{g}_r \cdot {p_f}^{\eta_f} + \eta_f \cdot \mu_f \cdot \bar{\kappa}_r - \delta_{p_f} \cdot p_f \tag{5.35}$$

where $\bar{x}_f$ follows the same form as Eq. 5.30 and $\hat{x}_f = g_{f,T}/k_{X_f}$ (f_0 block of Figure 5.1). Additionally, for further simplicity, we assume the transcription factor binding rate (α_{f_f}) is 1 h^{-1} allowing the unbinding rate α_{r_r} to be equal to the dissociation constant μ_f. Equation 5.35 forms the basis of the 'controller module' shown in Figure 5.1.

The QSS of the three o-16S rRNA promoter states are: (i) the open free promoter

($\bar{g}_r$), (ii) the promoter when bound by σ being actively transcribed ($\bar{x}_r$, calculated using $\hat{x}_r$), or (iii) the promoter bound by the regulator and therefore inhibited ($\bar{\kappa}_r$):

$$\bar{g}_r = g_{r,T} - \bar{x}_r - \bar{\kappa}_r \tag{5.36}$$

$$\hat{x}_r = \frac{g_{r,T}}{k_{X_r}}\Big(\frac{\mu_f}{\mu_f + {p_f}^{\eta_f}}\Big) \tag{5.37}$$

$$\bar{\kappa}_r = (g_{r,T} - \bar{x}_r)\Big(\frac{{p_f}^{\eta_f}}{\mu_f + {p_f}^{\eta_f}}\Big) \tag{5.38}$$

$\bar{x}_r$ determines the rate of host ribosome co-option, via the o-16S rRNA (r, see Equation 5.22) with $g_{r,T}/k_{X_r}$ determining the maximal rate (r_0 block, Figure 5.1) and $\mu_f/(\mu_f + {p_f}^{\eta_f})$ representing the inhibitory action of the controller p_f (F block, Figure 5.1).

The rate of change of the orthogonal 16S rRNA is as described in Equation 5.22 and co-option of the host ribosomes is described in Equation 5.34.

Finally, the number of free orthogonal ribosomes is given by:

$$R = \frac{R_{Total} - R_h}{1 + \hat{c}_f + \sum_{i=1}^{N}\Big(\hat{c}_i\Big)} \tag{5.39}$$

Note that this follows the same form as Equation 5.34, with the total number of o-ribosomes available to the circuit being the total number of all ribosomes (R_{Total}) minus the number of host ribosomes (R_{host}).

5.3.1 Numerical testing of the model reduction

Using the specific binding and unbinding rates of cellular components reported in [52], and calculating their respective dissociation constants as needed, we can compare the behaviour of the full mechanistic model and reduced model. Simulations demonstrate that the reduced model accurately captures the transient and steady-state behaviour of the full model, for both simple circuits based on activation of multiple genes and more complex circuits including oscillatory inputs (Figure 5.2). Crucially, the model reduction process preserves the rapidly changing closed-loop dynamics produced by the non-linear controller (Figure 5.2c).

Initially, we simplified the model to track only protein dynamics - whose control is the main subject of this thesis. This can be achieved by assuming the equations describing the dynamics of the o-16S rRNA (Equation 5.22) and host ribosomes

(Equation 5.23) are also at steady state. However, this model no longer captures the transient dynamics of the system, although it does still successfully recapitulate the steady state behaviour of the full model to static inputs (Figure 5.2d, *left*). In the presence of oscillatory inputs this additional reduction acts to hide the induction of oscillations in other genes due to the sharing of cellular resources (Figure 5.2d, *right*). This may imply a given controller design may lead to complete decoupling when it will not. Analysis of the parameters shows that o-ribosome assembly is slow ($\varrho_f = 0.9$ (nM·h)$^{-1}$) which violates the assumption that these species are at quasi-steady state (see Section 5.4 for the discussion of parameters).

5.4 Review of model parametrisation

To parameterise our model we conducted an extensive literature search to identify potential biological components.

σ_T, R_{Total}, Cellular resources Numerous studies into translational capacity show that as growth rate falls, translational capacity, in the form of ribosome number, also falls. Tadmor and Tlusty utilise a simple model of microbial growth, ribosome biosynthesis and previously published data to produce a model of how cellular resources vary with growth rate [98]. We utilise their results to approximate the level of free ribosomes in the cells (at slow growth rate as demonstrated by our prototype, $\sim$ 1 doubling per hour). This yields a free ribosome concentration of around 5000 nM. We account for competition between host and circuit genes by assuming that only half of this pool is available to the synthetic circuit; this is consistent with the observation that up to 50% of the cell's translational capacity can be used for heterologous gene expression [78]. This yield results in a similar range in the number of ribosomes per cell as in [37]. Applying similar assumptions to RNA polymerase and the data in [32] we set σ_T for each gene to be 250 nM.

g_T, promoter concentration (i.e. plasmid copy number) Assuming a slow growth rate ($\sim$ 0.6 doublings per hour) and therefore a small cellular volume ($\sim$ 0.34 μm^3, [32]), we can calculate the approximate concentrations for a range of plasmids. High copy number plasmids (e.g. $>$ 100 copies per cell) correspond to around 500 nM. Commonly used medium copy number plasmids ($\sim$ 30 copies per cell) correspond to around 100 nM. Very low copy plasmids ($<$ 5 copies per cell) correspond to a concentration of around 10 nM. Concentrations of less than 10 nM

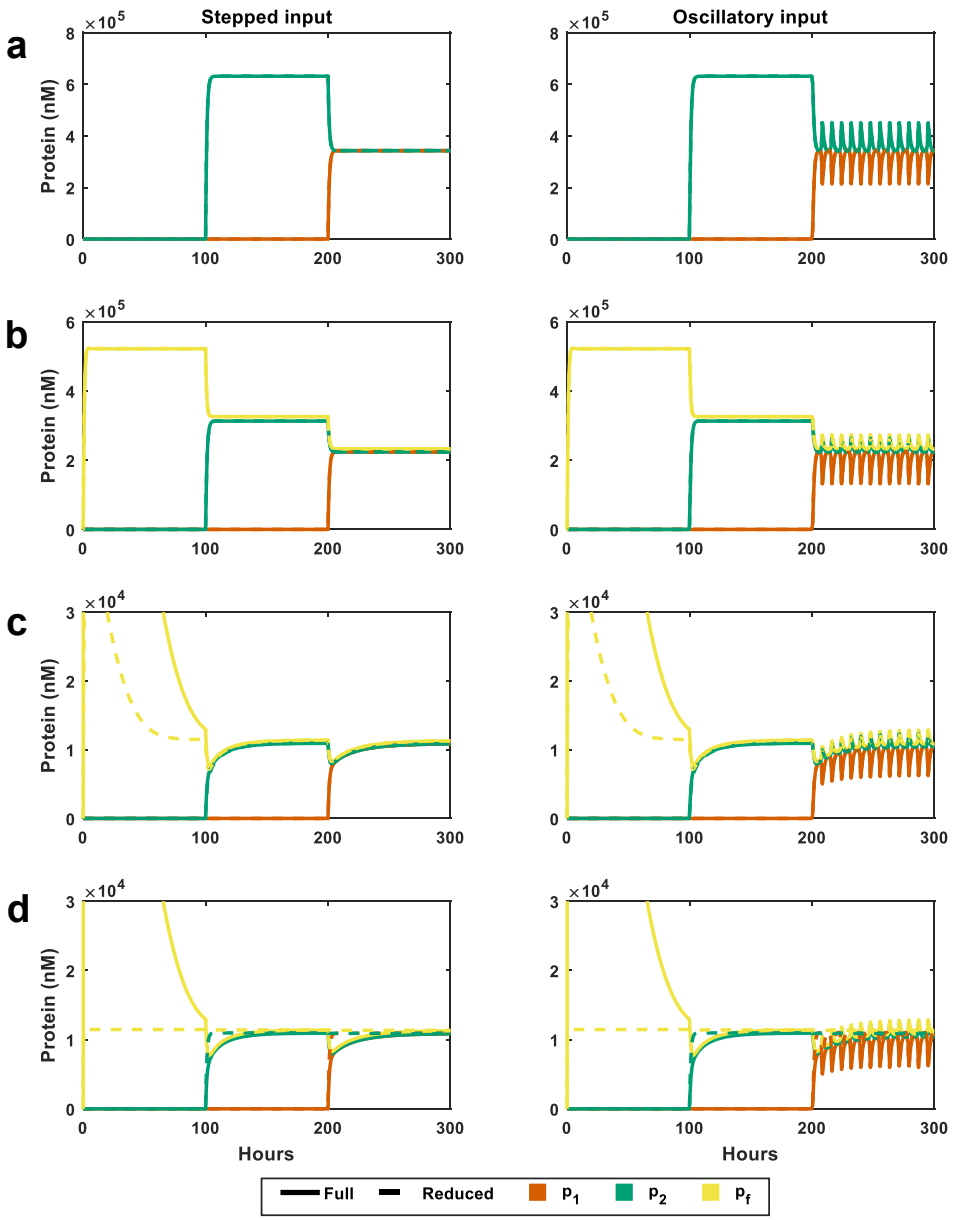

Figure 5.2: **Comparison of the full and reduced models.** The reduced model (dashed line) successfully captures the behaviour of the full model (solid line). Parameters used are derived from [52], modified to account for multiple resource use (see Section 5.4). In the stepped input examples: $u_1 = u_2 = 0$ nM except $u_1(t > 100) = 500$ nM and $u_2(t > 200) = 500$ nM. In the oscillator input examples: $u_1 = u_2 = 0$ except $u_1(t > 100) = 500(\cos(0.8t) + 1)$ nM and $u_2(t > 200) = 500(\cos(0.8t)+1)$ nM. A range of controller prototypes are shown: **(a)** open loop with no controller $g_{f,T} = 0$ nM. **(b)** a linear controller $\eta_f = 1, g_{f,T} = 500$ nM. **(c)** a non-linear controller $\eta_f = 2, g_{f,T} = 500$ nM. **(d)** The same simulation as in part (c) but using the highly reduced, protein-only model.

would indicate around one copy per cell, which could be achieved by chromosomal integration.

Accounting for multiple RNA polymerases and ribosomes per gene Our model assumes that each gene is bound by one RNA polymerase (σ) and each mRNA bound by one ribosome (R). However, *in vivo*, an RNA polymerase can initiate a new transcription event at the promoter before the previous polymerase has cleared the downstream gene. This leads to multiple RNA polymerases translating the same gene at once, each at a different stage. Similarly, multiple ribosomes can translate each mRNA leading to polysome complexes. We account for this in our model by increasing the copy number of each gene by n_σ (i.e. the number of RNA polymerase per gene) and increasing the mRNA transcription rate τ by n_R (i.e. the number of ribosomes per mRNA). By increasing these parameters we maintain the levels of competition whilst allowing the simple one-to-one relationship used to derive our model. Therefore we weight the literature values of $g_{i,r,f}$ and $\tau_{i,f}$ by these scaling constants for the simulations such that:

$$g_{i,r,f}^{\text{simulation}} = n_\sigma \cdot g_{i,r,f}^{\text{estimated}} \qquad \tau_{i,f}^{\text{simulation}} = n_R \cdot \tau_{i,f}^{\text{estimated}} \tag{5.40}$$

This approach is also used in [38, 45].

In [97], n_σ is between 1 and ~ 50 dependent upon the specific gene and ~ 13 for a constitutively expressed gene. In [99], the authors report values of n_σ up to 38 and n_R up to 40, with a median value of 28. We used a conservative estimate of $n_\sigma = 10$ and $n_R = 20$ throughout. (Note that τ_r is not increased as the o-rRNA is not a protein encoding gene and so does not compete for ribosomes. Each ribosome contains only one 16S rRNA).

k_X, RNA polymerase–promoter dissociation constant Brewster *et al.* provide a model which allows the calculation of binding energies from DNA sequences [100]. They experimentally validate their model against a series of degenerate promoters. They calculate the binding energies of an RNA polymerase binding across these sequences to range from -7 to -1 $k_B \cdot T$. In [101], Bintu *et al.* relate binding energy ($\Delta\varepsilon$) to dissociation constant k_D using the simple expression:

$$\Delta\varepsilon = k_B \cdot T \cdot \ln\Big(\frac{K_D}{K'_D}\Big) \tag{5.41}$$

where k_B is the Boltzmann constant, T is the temperature, K_D is the dissociation constant of the RNA polymerase from its promoter, and K'_D is the non-specific binding of the polymerase to DNA. The authors estimate K'_D to be 10^4 nM. Using these results we estimate the dissociation constants of RNA polymerase for promoters to be between 3 nM (for strong promoters) and 1000 nM (for the weakest promoters). Additionally, experimental measurements of the P_{lac} promoter show it has a dissociation constant of 560 nM [102] and P_{PR} bound by cI has a dissociation constant of 100 nM [103]. This demonstrates that the results of the model developed by Brewster *et al.* produce results in a realistic range.

Where promoter dissociation constants could not be identified in the literature, the -41 to -1 sequence of the promoter was retrieved and used to calculate the binding energy using the model described by Brewster *et al.* [100]. This was converted to a dissociation constant as outlined above using the formula provided by Bintu *et al.*

A visual inspection of the reduced model shows that promoter concentration (plasmid copy number) and promoter dissociation constant are found exclusively as a ratio g_T/k_X. This means that the same transcription dynamics can be created by different combinations of copy number and dissociation constant (Figure 5.3).

k_L, Ribosome-RBS dissociation constant The ribosome binds the mRNA at the ribosome binding site in a multistep process which we model here as a single step. In [104], Na *et al.* use a thermodynamic model to estimate the dissociation constant for the ribosome-RBS dissocation constant. Using this data, we estimate the wild-type RBS for the host ribosome to be in the order of $k_L \approx 10^4$ nM. This is the same order of magnitude as that determined in [98]. In Chapter 3, we proposed that the association between the orthogonal ribosome and orthogonal RBS is much weaker than the association between host ribosome and host RBS. We assume the orthogonal RBS dissociation is on the order of $k_L \approx 10^5$ nM. To extend our analysis to include weaker or stronger RBSs which could be produced experimentally, we allow k_L to vary between 10^3 and 10^6 nM.

$\mu_{i,f}$, $\eta_{i,f}$, Transcription factor parameters $\mu_{i,f}$ represents the dissociation constant of the transcription factor from the DNA. Where both the on-rate, α_f, and off-rate, α_r, are present in the equation we set $\alpha_f = 1$ h^{-1} to allow the off-rate to be equal to the dissociation constant (as $\mu = \alpha_r/\alpha_f$). $\eta_{i,f}$ represents the co-operativity due to multimerisation and binding to the promoter. A literature search shows the range of dissociation constants falls between 10^{-2} nM (e.g. LacI) and 10^3

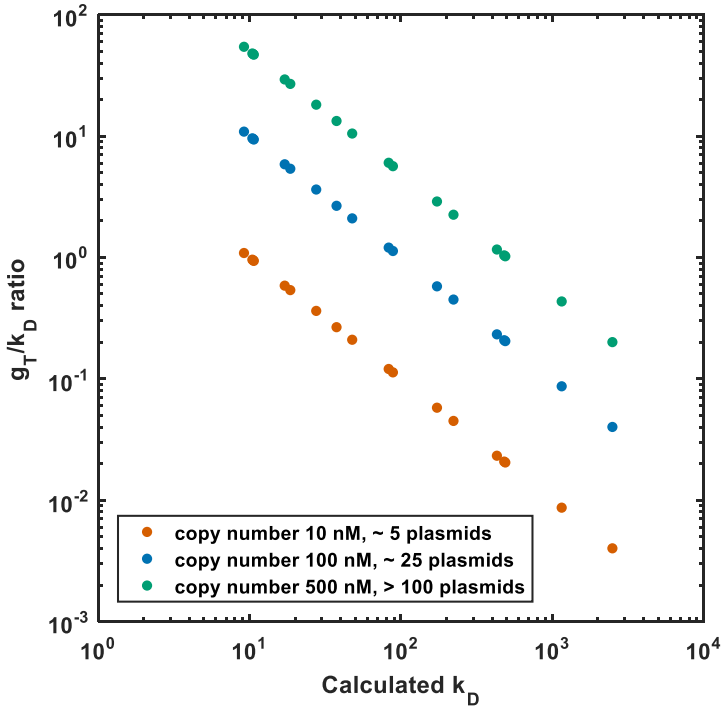

Figure 5.3: **Transcriptional g_T/k_X ratio.** Transcription rates are determined by the copy number-to-promoter strength ratio. Each ratio is obtainable by a different combination of plasmid copy number and RNA polymerase dissociation constant. The polymerase–promoter dissociation constants (k_X) are those calculated from the energetics reported in [100]. Plasmid copy numbers are calculated as described in the promoter concentration section.

nM (e.g. Cro.). The transcription factors investigated in this chapter are shown in Table 5.5.

μ_r, 16S rRNA-ribosome dissociation constant In [105], the authors report the *in vitro* ϱ_f and ϱ_r values for wild type *E. coli* ribosomes as 0.9 $(\text{nM.h})^{-1}$ and 25 h^{-1}. At present, due to technical limitations, we have not determined the efficiency of o-ribosome assembly in our collaborators' experimental system so we take these values as a proxy for ribosome assembly and disassembly of orthogonal ribosomes. Our key assumption is that assembly cannot be faster than that of the wild type host ribosome.

Reaction rates Utilising common values for *E. coli* collated in [31], we calculate the production and decay rates of the RNA and protein species. Assuming a 300

amino acid protein, such as a fluorescent reporter, and an amino acid incorporation rate $\sim$ 20 aa s^{-1} yields $\gamma_{i,f} = 240$ h^{-1}. A 300 amino acid protein is encoded by an mRNA of approximately 900 nucleotides. Assuming a nucleotide incorporation rate of $\sim$ 80 nt s^{-1} yields an mRNA elongation rate of $\tau_{i,f} = 320$ h^{-1}. We calculate $\tau_r = 190$ h^{-1}, assuming the orthogonal 16S rRNA is around 1500 nucleotides in length. We estimate the decay constant for RNA species ($\delta_{m_{i,f}}$ and δ_r) to be around 20 h^{-1} assuming an RNA lifetime of around 3 minutes. We estimate the protein decay rate to be around 1 h^{-1} for dilution due to cell growth. Protein decay rates can be increased by utilising amino acid sequences which target the protein to the degradation machinery, in which case $1 < \delta_{p_{i,f}} < 40$ h^{-1} (half life between 60 minutes and 1 minute).

Biologically reasonable protein outputs From [31], the *E. coli* proteome has a concentration of approximately $\sim 4.7 \times 10^6$ nM. In their study of translational capacity, Scott *et al.* find that cells can tolerate transgenes making up to 30% of their proteome. We assume that the maximal expression of circuit proteins can be up to $\sim 1.4 \times 10^6$ nM.

5.5 Model analysis reveals a trade-off between gene expression and level of decoupling

We initially considered a simple two-gene circuit; where one gene is induced and a second induced some time later. As discussed previously, due to resource competition, the first falls upon induction of the latter (e.g. Figure 5.2). To investigate the action of the controller we simulated a range of controllers by sampling parameter values from a biologically feasible range. This crude analysis suggested there was a trade-off between gene expression and the decoupling ability of the controller; with decoupling coming at a loss of gene expression.

To assess the function of the controller we extended this rudimentary analysis more formally as described below.

5.5.1 Assessing coupling dynamically in a simple two-gene circuit

We simulate the induction of p_1 by $u_1 = 500$ nM. At $t = 12$ h (called θ_{ind} below), we simulate the increase of u_2 from 0 to 500 nM. The difference in expression of

p_1 due to the u_2 disturbance is what we term 'coupling'. The final output level of p_2 at the end of the simulation, θ_{ss}, is compared to the output p_{target} which is the protein level using the host ribosome pool. We call this 'expression' below. (Note that at θ_{ss}, $p_1 = p_2$ by definition as there are only two circuit genes with identical parameters as the two circuit genes will co-opt equal numbers of ribosomes if they are identical as we assume).

The values are normalised for ease of analysis:

$$\text{coupling} = \Big(p_1(t = \theta_{ss}) - p_1(t = \theta_{ind}\Big)/p_1(t = \theta_{ind}) \tag{5.42}$$

$$\text{expression} = \Big(p_2(t = \theta_{ss}) - p_{target}\Big)/p_{target} \tag{5.43}$$

The ideal controller would have minimal coupling (0) and maximal expression (0) as defined by Equations 5.42 and 5.43. Controllers which are poor at decoupling would have a 'coupling' score of less than zero, indicating $p_1(t = \theta_{ind}) > p_1(t = \theta_{ss}$. Controllers which also act to decrease the final protein output will achieve a negative 'expression' score as $p_2(t = \theta_{ss}) < p_{target}$. We carried out a multiobjective optimisation to identify the trade-off between these two characteristics.

5.5.2 Multiobjective optimisation

The multiobjective optimisation was carried out using the in-built MATLAB function *gamultiobj* from the Optimisation Toolbox (Version 7.4, with computational speed increased using the Parallel Computing Toolbox Version 6.8 or 6.10). This function is initialised with a random set of parameter values (the 'population'). The cost function is evaluated at each parameter set in the population and the best 'individuals' carried forward for reassessment (i.e. the next 'generation'). Diversity is introduced into the population by random 'mutation' or by 'crossing over' between different parameter sets. This function continues with this selection scheme, selecting the best parameter sets which minimise the two elements of the given cost function. This produces sets of parameters which lie on the Pareto front, the optimal trade off between the two parts of the cost function where improvement in one dimension results in the detriment of the other.

The population size was set to 200 individuals to allow good searching of the parameter space. We set the Pareto fraction to 0.25 (i.e. 50 individuals), this option ensures that the Pareto front will be well defined as the routine aims to keep approximately 50 individuals on or near the front at each selection step whilst also

Parameter	Scale	Lower bound	Upper bound	Block
$g_{r,T}/k_{X_r}$	$\log_{10}$	10^{-2}	10^{2}	r_0
k_{L_f}	$\log_{10}$	10^{3}	10^{8}	T_{L_f}
η_f	linear [integer]	1	4	$F(\cdot)$
μ_f	$\log_{10}$	10^{-2}	10^{3}	$F(\cdot)$
$g_{f,T}/k_{X_f}$	$\log_{10}$	10^{-2}	10^{2}	f_0

Table 5.2: **Multiobjective parameters.** Parameters varied in the optimisation routine. Parameters were varied on the scale shown between the lower and upper bounds. The block refers to the block diagram in Figure 5.1.

ensuring diversity by rejecting the 150 individuals with the lowest fitness. k_X values were set to 1 allowing the $\hat{x}$ ratios to be investigated by varying $g_{r,T}$ and $g_{f,T}$ only. All parameters and bounds varied are shown in Table 5.2.

The optimisation routine aims to concurrently minimise both elements of the cost function:

$$\begin{aligned}\chi_1 &= \Big(p_1(t=\theta_{ind}) - p_1(t=\theta_{ss})\Big)^2 \\ \chi_2 &= \Big(p_2(t=\theta_{ss}) - p_{target}\Big)^2\end{aligned} \tag{5.44}$$

where θ_{ind} is the time of the induction of p_2 and θ_{ss} is the last time point, t_{max}. If the simulation is not at steady state at t_{max} then the result is given the poorest fitness (specified as 'not a number' data type which is interpreted poorly by the optimisation function). p_{target} is calculated by simulating the action of the circuit in a model utilising the host ribosome pool for gene expression (i.e. this is the maximal gene expression.). Note that the cost function takes a different form to the definitions of coupling and expression above (Equations 5.42 and 5.43). The results of this optimisation are shown in Table 5.3.

5.5.3 Shape of the Pareto front

We assess the impact on protein levels by comparing final protein outputs to those of the same circuit where translation is uncontrolled and mediated by the host ribosome pool. This identified a hard trade-off between these two objectives, with the range of equally optimal solutions (the Pareto-optimal front) showing an inverted concave shape, i.e. decreases in gene coupling are achieved at the expense of gene expression. Note that the Pareto front, as found by our optimisation routine, is not smooth. Our optimisation routine has found the approximate front; our robustness

Point	$g_{r,T}/k_{X_r}(10^x)$	$k_{L_f}(10^x)$ nM	η_f	$\mu_f(10^x)$ nM	$g_{f,T}/k_{X_f}(10^x)$	Exp.	Coup.
1	-0.97	4.30	4	-1.801	0.31	-1.0000	0.0000
2	0.87	6.30	1	0.975	-1.70	-0.0165	-0.2985
3	0.11	7.32	2	1.024	-0.61	-0.6345	-0.0476
4	-1.91	7.60	2	1.513	-1.10	-0.7495	-0.0302
5	0.22	7.70	2	1.246	-1.60	-0.0714	-0.2374
6	0.80	7.46	2	1.325	-1.70	-0.0383	-0.2667
7	0.59	7.69	2	1.516	-1.70	-0.0177	-0.2894
8	-0.03	7.74	2	1.644	-1.35	-0.0922	-0.2221
9	0.04	7.64	3	1.454	-1.60	-0.5097	-0.0557
10	-0.64	7.31	2	1.220	-1.45	-0.5610	-0.0607
11	-0.05	7.82	3	1.092	-1.40	-0.5885	-0.0434
12	0.99	7.54	1	1.777	-1.68	-0.0047	-0.3061
13	0.11	7.92	3	1.534	-1.67	-0.2239	-0.1274
14	0.11	7.48	2	1.124	-1.57	-0.2156	-0.1559
15	0.95	7.75	2	1.449	-1.01	-0.0454	-0.2599
16	0.67	7.68	2	1.348	-1.65	-0.0244	-0.2815
17	0.08	7.42	2	1.223	-1.13	-0.4104	-0.0933
18	0.13	6.60	2	0.981	-0.44	-0.8759	-0.0137
19	0.70	7.81	3	1.305	-1.58	-0.2694	-0.1116
20	0.88	7.69	2	1.270	-1.41	-0.0361	-0.2689
21	-0.96	7.87	3	1.263	-1.54	-0.6413	-0.0360
22	0.76	7.62	2	1.530	-1.69	-0.0169	-0.2903
23	0.17	7.80	3	1.245	-1.70	-0.3523	-0.0880
24	-1.54	7.54	2	1.548	-1.16	-0.6811	-0.0403
25	0.66	7.65	2	1.405	-1.69	-0.0222	-0.2839
26	0.12	7.49	2	1.265	-1.61	-0.1571	-0.1830
27	0.62	5.92	2	1.030	-1.02	-0.9083	-0.0099
28	0.08	7.40	2	1.668	-1.56	-0.1347	-0.1952
29	0.35	7.74	2	1.055	-1.39	-0.1183	-0.2049
30	0.53	6.77	1	1.302	-1.08	-0.0159	-0.2988
31	0.47	6.80	1	1.314	-1.68	-0.0091	-0.3032
32	0.14	7.06	2	1.203	-0.65	-0.6973	-0.0376
33	0.10	5.67	2	1.214	-1.22	-0.9418	-0.0064
34	-0.20	6.72	2	0.891	-1.24	-0.8238	-0.0201
35	0.99	7.54	1	1.777	-1.68	-0.0047	-0.3061
36	0.04	7.64	3	1.454	-1.60	-0.5097	-0.0557
37	0.18	7.22	2	1.266	-1.21	-0.4591	-0.0817
38	-0.83	7.73	3	0.962	-1.48	-0.7543	-0.0225
39	-0.45	7.73	3	0.859	-1.48	-0.7156	-0.0268
40	0.15	7.77	2	1.286	-1.60	-0.0587	-0.2479
41	0.05	7.82	2	1.233	-1.49	-0.0872	-0.2256
42	-0.39	6.73	2	1.164	-1.21	-0.8154	-0.0212
43	0.28	7.14	2	1.318	-1.52	-0.3300	-0.1154
44	0.34	7.33	2	1.351	-1.47	-0.2146	-0.1563
45	0.50	7.65	2	1.312	-0.95	-0.1833	-0.1701
46	0.16	6.77	2	0.906	-1.08	-0.7842	-0.0253
47	-0.97	4.30	4	-1.801	0.31	-1.0000	0.0000
48	0.15	7.77	2	1.284	-1.60	-0.0589	-0.2478
49	0.48	7.82	2	1.211	-1.08	-0.1083	-0.2112
50	0.25	7.66	2	1.175	-1.69	-0.0682	-0.2400

Table 5.3: **Controllers determined by the optimisation routine.** The 50 controllers returned from the multiobjective optimisation which lie on the putative Pareto front. Exp., expression as defined in Equation 5.43. Coup., coupling as defined in Equation 5.42.

analysis (described below) finds the true front which coincides with η_f values of 4 (see Section 5.5.4). Our simulations suggest that coupling can be halved for only a 20% reduction in gene expression, while coupling can be reduced to between 5-10% for a 50% reduction in gene expression (Figure 5.4a).

To determine the robustness of the approximated Pareto front to parameter selection, we varied the optimised parameters by up to 50% for each point on the front. None of these designs result in controller failure where expression is lost but coupling is not abolished - i.e. no designs fall into the lower left quadrant. This suggests that potential parameter variations, when selecting biological components to implement the controller, can act to move the controller along the front, but should *not* result in failure. We did find that a small number of these perturbed designs show slower responses and we discounted these from further analysis. We carried out an additional robustness analysis allowing *all* parameters governing the controller behaviour to vary. This includes parameters which are either difficult to design (e.g. controller translation rate γ_f) or intrinsic properties which cannot be designed (e.g. o-rRNA association rate, μ_r). We find that all of these controllers also fall upon the same front demonstrating that uncertainty in these values does not preclude controller design (Figure 5.4b).

5.5.4 Parameter variation across the Pareto front

To determine how each parameter contributes to the gene expression and coupling trade-off, we analysed how each changes across the front. This highlighted the need for high η_f values. This parameter represents the level of co-operativity in the system brought about, for example, by transcription factor dimerisation or the presence of multiple operator sites at the target promoter. The true Pareto front coincides with a value of $\eta_f = 4$ (Figure 5.5a). For this reason we discount controllers where $\eta_f = 1$ from further analysis in this section as these controllers perform most poorly. We also find that small μ_f values are most often associated with controllers which act to nearly completely decouple genes but at a significant cost to gene expression (Figure 5.5e). Similarly, small k_{L_f} values, corresponding to strong ribosome binding sites (low ribosome-mRNA dissociation constant), are associated with large levels of decoupling at a high cost to gene expression (Figure 5.5d). Simulations suggest $k_{L_f} > 10^5$ nM and $\mu_f > 0.1$ nM in all cases, for the simple two-gene circuit example used here. (Note that for many natural transcription factors co-opted into synthetic gene networks $\mu_f < 0.1$ nM and η_f may be limited. We demonstrate how this can be compensated for in controller design in Section 5.7). A high $g_{f,T}/k_{X_f}$ ratio

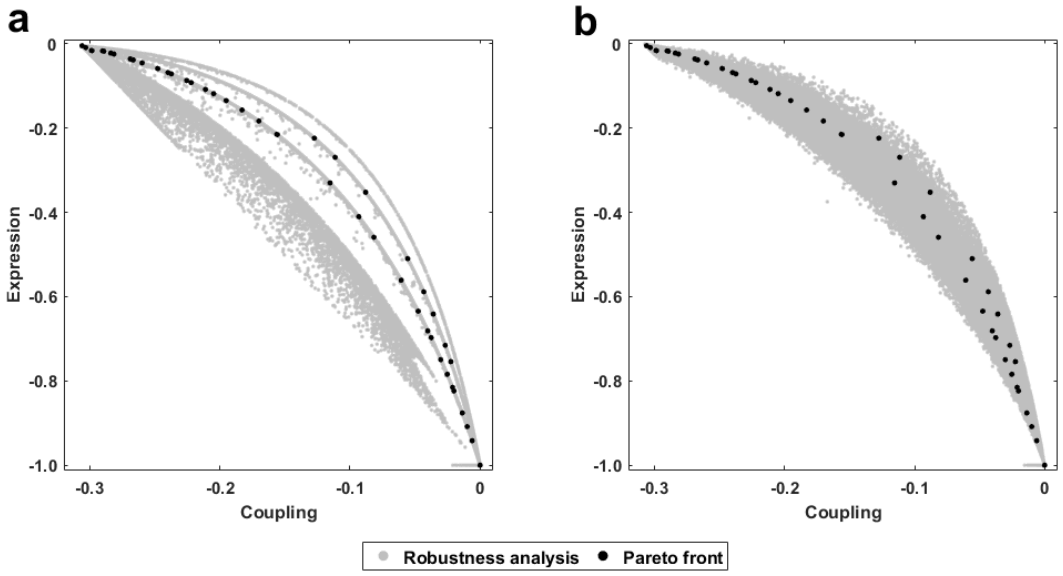

Figure 5.4: **Trade-off between gene expression and decoupling.** The Pareto front demonstrating the trade off between gene expression and coupling. Expression is measured as change in steady state gene expression in comparison to simulations of the circuit using the saturated host ribosome pool. Coupling is measured as the steady-state change in p_1 in response to u_2 as defined in Section 5.5.1. $100,000$ simulations were run and simulations reaching steady state at $\theta_{ind} = 72$ h and $\theta_{ss} = 192$ h are shown. **(a)** Robustness was determined by allowing optimised parameters of the controller to vary by up to $\pm 50\%$. The points shown are those which reached steady state within a reasonable time frame ($89,890$ of $100,000$ simulated). (Parameters corresponding to the controllers which fall on the approximate Pareto front are listed in Table 5.2). **(b)** Robustness was determined by allowing all parameters describing the o-rRNA and controller to vary by up to $\pm 50\%$. The points shown are those which reached steady state within a reasonable time frame ($88,273$ of $100,000$ simulated).

($g_{f,T}/k_{X_f} > 1$, produced by expressing the regulator from a strong promoter carried on a high copy number plasmid) results in complete decoupling and abolition of gene expression (Figure 5.5c). We therefore suggest keeping $g_{f,T}/k_{X_f} < 1$ in all instances to ensure adequate gene expression. We find that the $g_{r,T}/k_{X_r}$ ratio governing maximal o-rRNA transcription rate varies significantly across all behaviours making general guidelines difficult to establish (Figure 5.5b).

5.5.5 Selection of controller parameters for design guidelines

Coupling and expression scores were calculated for each controller as outlined above for all the results of robustness analysis (Equations 5.42 and 5.43). These results were then scaled by their maximum absolute values to ensure both axis were between 0 and 1 (note that for calculating the distance metric we can ignore signs). We calculate the Euclidean distance between each point (x_{scaled}, y_{scaled}) and the point of interest from our numerical optimisation (x_0, y_0):

$$d = \sqrt{\left((x_{scaled} - x_0)^2 + (y_{scaled} - y_0)^2\right)} \tag{5.45}$$

We then define any points within a circle of radius 0.025 centred on (x_0, y_0) as within the local region. Qualitatively these points have the same behaviour and so we group them for further analysis.

To provide specific quantitative design rules, we selected five points across the front and assess the parameter space around these points which give rise to these controller behaviours (Table 5.4). Assessing how parameters correlate in these local clusters shows that k_{L_F} is a key regulator of behaviour. k_{L_f} is inversely correlated with $g_{r,T}/k_{X_r}$ indicating that as the o-ribosome production rate increases a stronger RBS is needed for controller function (i.e. a smaller value of k_{L_f}). We also identify an inverse correlation between k_{L_f} and μ_f in the cluster around (-0.01, -0.9), i.e. the most decoupled parameter set, such that decreases in repression by the transcription factor (μ_f) can be compensated for by increasing the RBS strength (decreasing k_{L_f}). We also find that changes in k_{L_f} and $g_{f,T}/k_{X_f}$ have some compensatory effects such that increases in k_{L_F} (i.e. weakening the RBS) can be mitigated by increasing $g_{f,T}/k_{X_f}$ (e.g. by increasing copy number).

In Figure 5.6, we show the internal dynamics of two controllers which demonstrate the trade-off across the Pareto front. Controller design 47 (Table 5.3) results in complete decoupling while nearly abolishing gene expression. Analysis of the controller

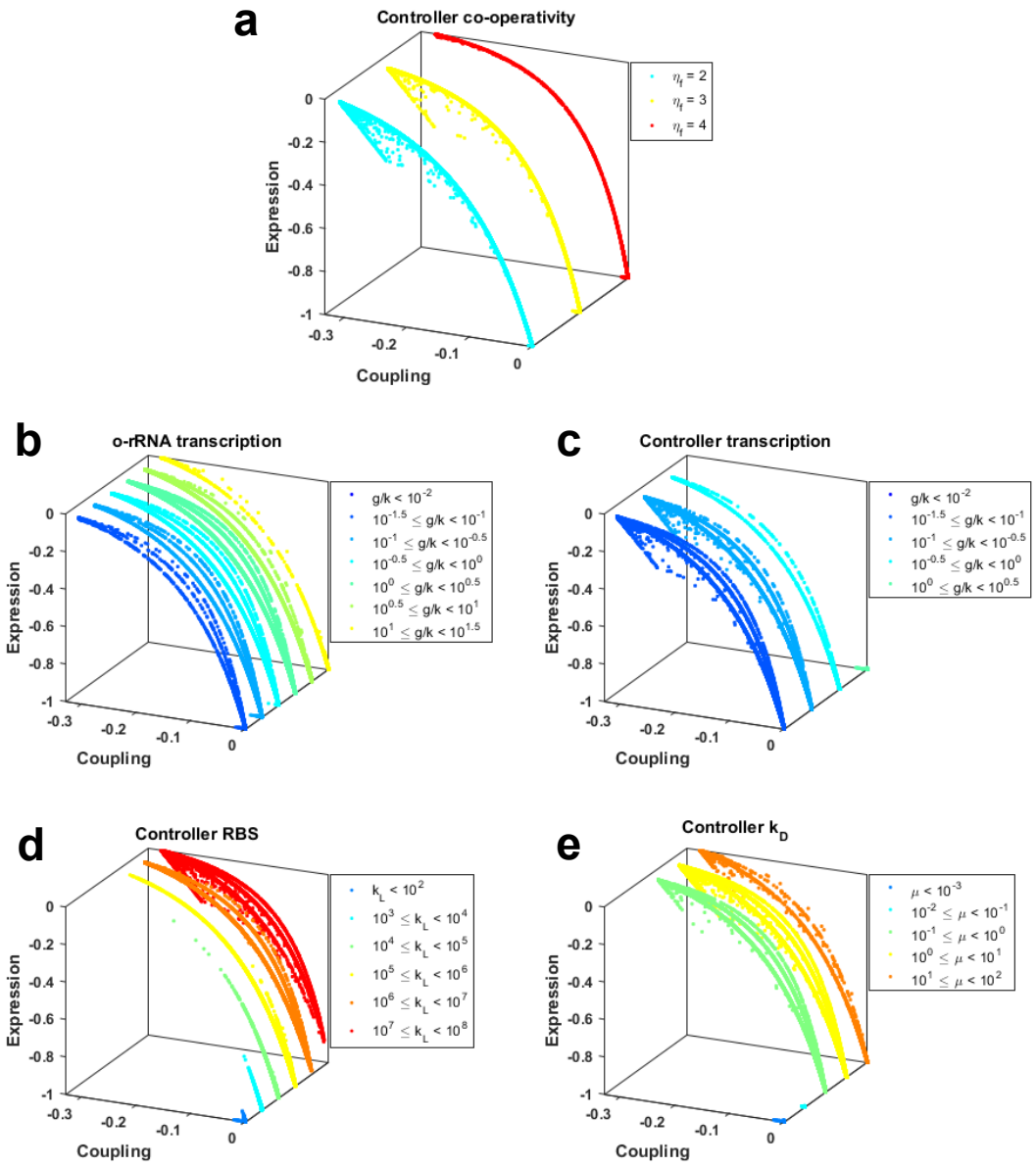

Figure 5.5: **Variation in parameters across the Pareto front.** The values of the parameters for each point shown in Figure 5.5a. As described in the text, controllers where $\eta_f = 1$ are removed for clarity. Also note that the third axis and subsequent separation serves only to aid visualisation and does not represent parameter value which is indicated by the colour and outlined in the figure legend. **(a)** Controller co-operativity as shown by η_f. **(b)** o-rRNA transcription as determined by the $g_{r,T}/k_{X_r}$ ratio. **(c)** Transcription of the controller protein as determined by the $g_{f,T}/k_{X_f}$ ratio. **(d)** Controller mRNA-ribosome binding site strength as measured by the mRNA-ribosome dissociation constant k_{L_f}. **(e)** Controller protein g_r dissociation constant μ_f.

		$g_{r,T}/k_{X_r}$		k_{L_f}		η_f		μ_f		$g_{r,T}/k_{X_r}$	
(co. %, ex. %)	**N**	**LB**	**UB**	**LB**	**UB**	**LB**	**UB**	**LB**	**UB**	**LB**	**UB**
(-13, -22)	125	$10^{-1.91}$	$10^{1.16}$	$10^{7.25}$	$10^{9.25}$	3	3	$10^{0.497}$	$10^{2.34}$	$10^{-2.38}$	$10^{-0.28}$
(-11, -27)	110	$10^{-1.89}$	$10^{0.87}$	$10^{7.01}$	$10^{9.08}$	3	3	$10^{0.471}$	$10^{2.46}$	$10^{-2.47}$	$10^{-0.33}$
(- 9, -35)	100	$10^{-2.28}$	$10^{0.94}$	$10^{7.11}$	$10^{8.80}$	3	3	$10^{0.593}$	$10^{2.27}$	$10^{-2.39}$	$10^{-0.3}$
(- 5.5, - 51)	126	$10^{-2.49}$	$10^{0.8}$	$10^{6.86}$	$10^{8.82}$	3	3	$10^{0.492}$	$10^{2.49}$	$10^{-2.23}$	$10^{-0.32}$
(- 1, -90)	2514	$10^{-2.86}$	$10^{1.42}$	$10^{4.12}$	$10^{8.26}$	2	4	$10^{0.43}$	$10^{2.47}$	$10^{-2.54}$	$10^{-0.23}$

Table 5.4: **Controller designs to manage the coupling expression trade-off.** Regions of parameter space were identified as described in Section 5.5.5 using a distance score of 0.25. The coupling (co. %) and expression cost (ex. %) are reported for each controller from the Pareto front chosen. The number of controllers in the local region is reported as N.

dynamics shows that expression of the controller protein dominates the ribosome pool (Figure 5.6a, *upper left*). This controller shows a slow response taking over 12 h to respond to the disturbance. Controller design 9 shows intermediate behaviour, with incomplete coupling (coupling changes from $> 30\%$ in the open loop confirmation to $\sim 5\%$ in the closed loop confirmation) but a loss of only 50% gene expression (Figure 5.6b, *right*). Analysis of the controller dynamics shows the controller responds significantly faster reaching steady state within 12 h of the disturbance being applied, although there is a brief (and here negligible) overshoot. The circuit dominates the translational resources with only a small number of o-ribosomes needed for controller function.

5.6 Tuning controller parameters to design response times

Analysis of the controllers tested so far has focused on how they are able to correct steady state errors brought about by gene coupling and so has largely ignored the system dynamics bar excluding excessively slow controllers (e.g. penalising those which only reach steady state after > 24 h). However, a controller which decouples genes well but has a slow response time will not be suitable for many applications in synthetic biology. Therefore, we took the five previous candidate controllers and conducted a local sensitivity analysis around each design point to assess the impact of each parameter on the system's speed of response. In addition to the controller parameters varied so far we also varied δ_ρ, δ_{m_f} and δ_{p_f}, which represent the decay of the o-rRNA, controller mRNA and controller protein respectively. These parameters were kept constant in the previous design evaluations to minimise the number of parameters in the optimisation and due to difficulties in engineering them [106], but since decay parameters often have significant affects on speed of response we

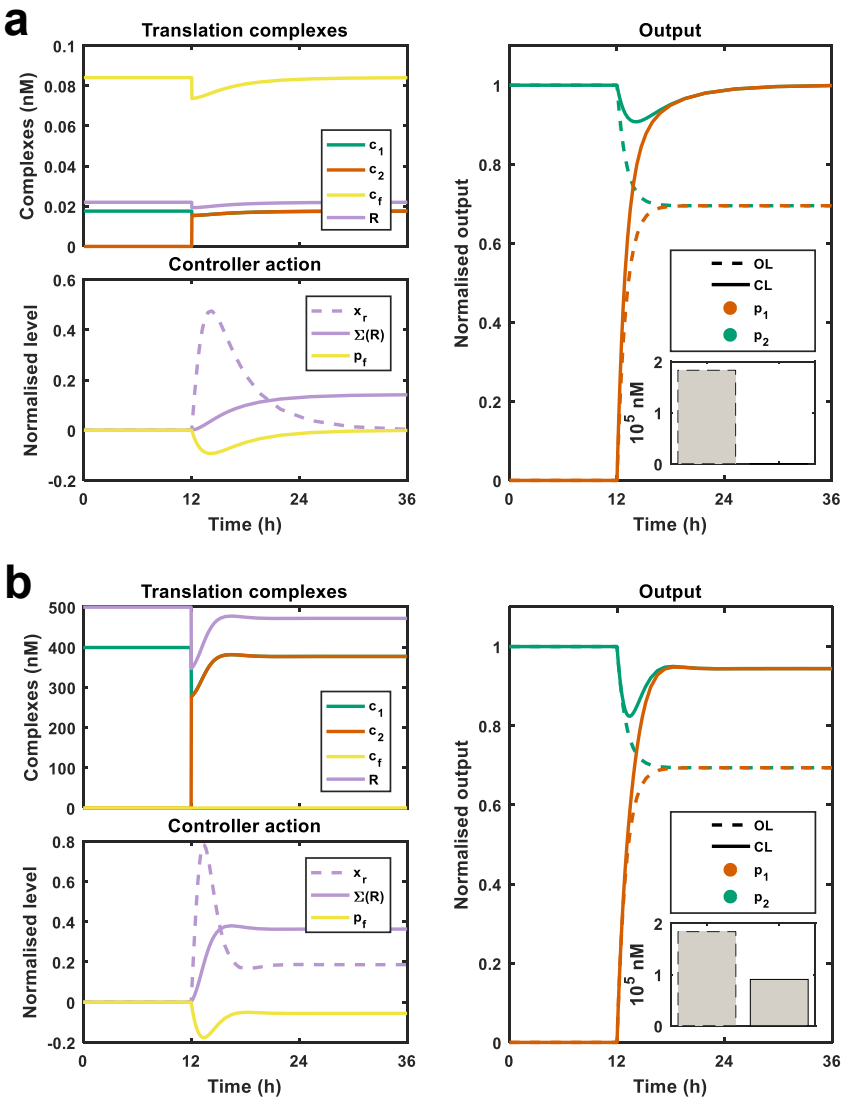

Figure 5.6: **Controller dynamics at different coupling-expression trade-offs.** The first gene p_1 is constitutively expressed, $u_1 = 500$ nM throughout. At 12 h, u_2 rises from 0 to 500 nM. *Upper left,* Translational complexes: Changing the distribution of the orthogonal ribosomes across circuit and controller mRNAs, c_1, c_2 and c_f represent the translation complexes of the mRNAs of gene 1, 2 and regulator f respectively. R represents the free orthogonal ribosomes. c_f acts as the sensor for the disturbance at $t = 12$ h. Levels are normalised by the total number of orthogonal ribosomes at $t = 12$ h. *Lower left,* Controller action: Changes in controller components over time. Levels are normalised by value at $t = 12$ such that 0 indicates no change. x_r, o-16S rRNA gene in the transcribing state; $\Sigma(R)$, number of orthogonal ribosomes; p_f, controller protein. *Right,* Normalised protein output. Protein levels are normalised by $p_1(t = 12)$ and shown in the absence of the controller (OL, $g_{f,T} = 0$ nM) and the presence of the controller (CL). Insets, non-normalised steady state protein levels. **(a)** Controller design 47. **(b)** Controller design 9.

explicitly assess their impact here. We allow the decay parameters δ_r and δ_{p_f} to vary between half and one and half times their nominal value.

Beginning with the four controllers which show intermediate behaviours (i.e. do not show complete decoupling), changing the parameters has significant impact on coupling and expression levels as discussed before. For controllers which have strong decoupling ability we find that there are individual parameters which when varied do not significantly affect the decoupling ability (e.g. $g_{r,T}/k_{X_r}$, μ_f) although these do still affect the expression levels (i.e. the controller gain).

In all five cases, the o-rRNA decay constant (δ_r) and protein controller decay constant (δ_{p_f}) are key to determining the speed of the system response. Figure 5.7 shows typical results for controller designs that result in complete decoupling (Controller 47, Figure 5.7a) or have intermediate behaviour (Controller 9, 5.7b). Increasing both parameters acts to increase the speed of response, at an expense of decoupling ability (Figure 5.7). In the regions tested, varying δ_r is less likely to introduce significant overshoots into the system (as seen at low δ_{p_f} values). However, a greater range of speed-up is achievable by varying the protein decay constant. The latter is also a more experimentally tractable parameter. Increasing both parameters acts antagonistically, with increases in δ_r decreasing gene coupling and increases in δ_{p_f} increasing it, meaning tuning both parameters may be advantageous. We see very little impact from varying the mRNA decay rate (δ_{m_f}).

As previously discussed the value of the controller co-operativity (η_f) is a key determinant of controller decoupling ability (Figure 5.7). This analysis replicates this result and also highlights that, at least in this parameter regime, increasing co-operativity can also act to significantly increase the speed of response.

5.7 Potential biological implementations of the controller designs

We carried out a detailed literature review to identify potentially suitable repressors with which to implement our system, focusing our analysis on (i) the ability of the repressor to be expressed in bacterial hosts (i.e. repressors from bacteria or bacteriophage), (ii) orthogonality (i.e. repressors which are not used in fundamental host processes), (iii) the presence of a known promoter architecture (which could be used to infer the dissociation constant of the RNA polymerase, see Section 5.4) and (iv) detailed characterisation of binding kinetics (ideally dissociation constants

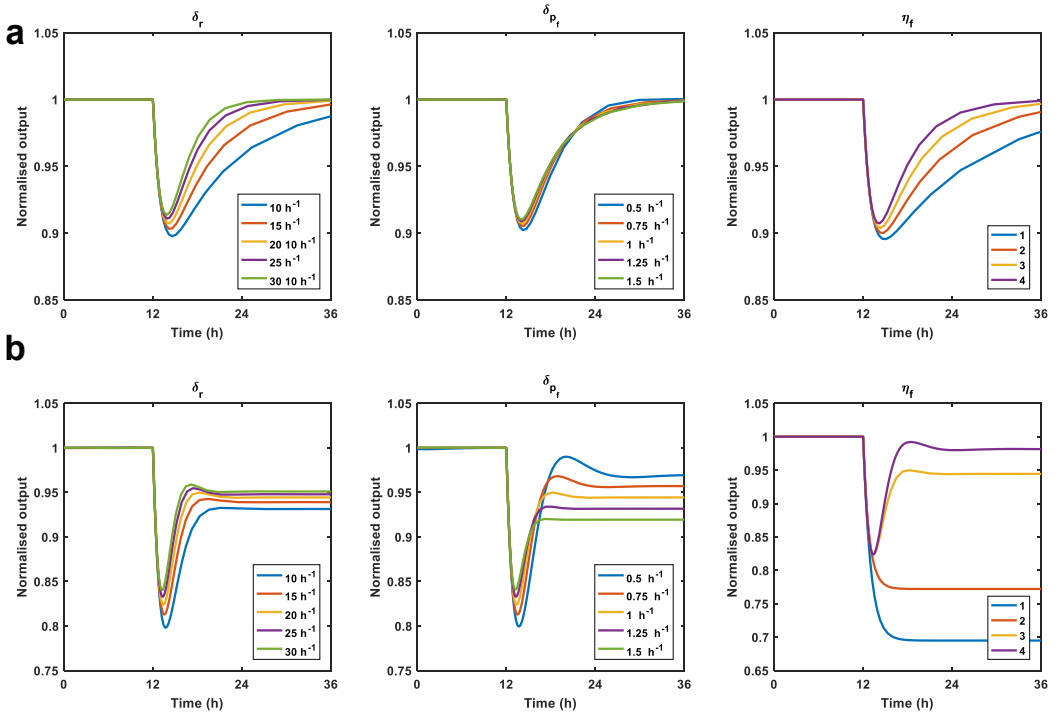

Figure 5.7: **Tuning decay parameters allows design of system dynamics.** The effect of varying the decay parameters δ_r and δ_{p_f} between 0.5 and 1.5 times their nominal value on the response of p_1 to the additional input u_2. The effect of varying the co-operativity η_f is also shown. **(a)** Sensitivity analysis around the parameter set from the high decoupling regime (Point 47). **(b)** Sensitivity analysis around a parameter set from the intermediate decoupling regime (Point 9).

measured in a biochemical assay, rather than a constant inferred from device function such as by fitting a Hill function to induction-fluorescence curves, as is often the case).

We identified six repressors from this literature search, including the commonly used LacI [94], TetR [107] and cI [108] repressors. We also identified putative controller candidates Cro and RstR from bacteriophages PY54 [109] and CTXφ [110], and LmrR, a global regulator of antibiotic resistance from Gram positive *Lactococcus lactis* [111]. See Table 5.5 for calculated dissociation constants and Table 5.6 for promoter sequences used in the analysis below.

Using the results of our sensitivity analysis and additional biological constraints we simulated a number of feasible biological implementations; (i) the o-rRNA and regulator having the same medium copy number (mimicking placement in the same plasmid, such as ColE1), (ii) a high copy number regulator, carried on for example a pUC vector, and a chromosomally integrated regulator and (iii) the effect of a protein degradation motif on designs of type (ii). Note that we did not assess the potential designs requiring the o-rRNA and regulator to be carried on different copy number plasmids, as these would result in high burden on the cells and significantly decreased growth rate as these cells would need to carry at least three plasmids, one containing circuit genes and one each for the o-rRNA gene and regulator. All of these prototype controllers fall along the Pareto front (Figure 5.8). We did not consider candidate designs which require destabilisation of the o-rRNA (i.e. increase δ_r) due to the difficulty in experimentally implementing these designs [106].

We carried forward designs representing a range of trade-offs between gene expression and decoupling for further analysis (Table 5.7). The putative controllers based on tetramers (LacI and RstR) show the fastest dynamics and best decoupling with minimal p_1 settling times and p_2 rise time upon induction of the second gene (Figure 5.9). Designs based on the phage repressor Cro with degradation motifs show the highest gene expression with an acceptable p_2 rise time of < 3 hours but coupling is far from fully abolished (Figure 5.9).

5.8 A dynamic resource allocation controller restores modularity in a range of more complex gene circuits

Having successfully demonstrated the ability of the proposed approach to decouple two independent modules, we analyse the ability of the controller to remove resource

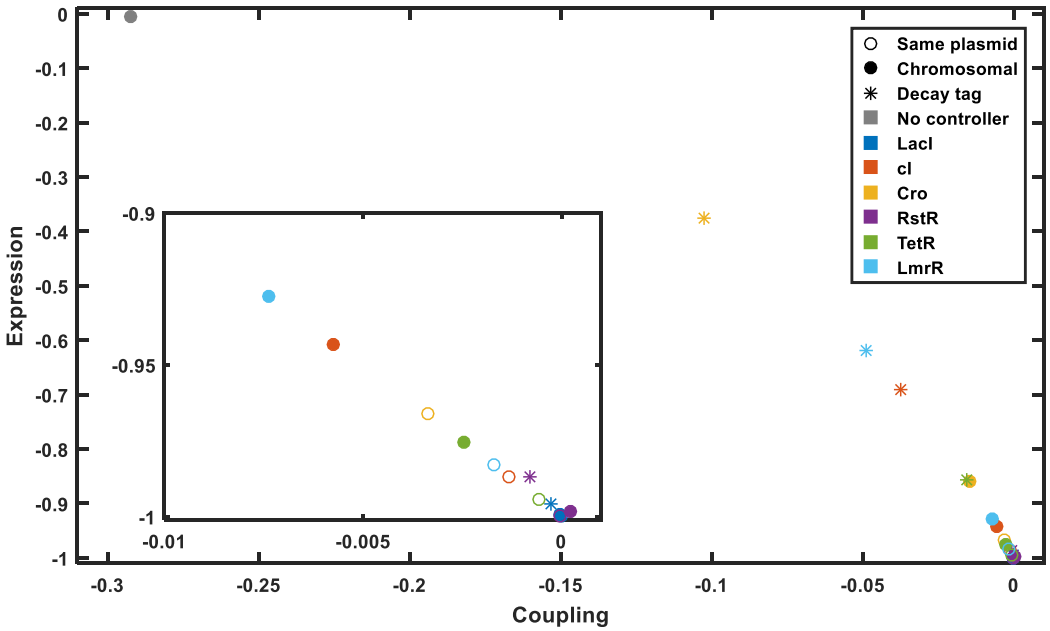

Figure 5.8: **Comparison of biological implementations based on orthogonal repressors.** Simulations of implementations based on the repressors in different plasmid confirmations and with degradation motifs. The positions of the prototype controllers in the coupling-expression space. Inset, expansion of the main figure around point (0,-1). Point colours represent the regulator protein and point style denotes copy number as follows: Same plasmid, $g_{r,T} = g_{f,T} = 100$ nM; Chromosomal, $g_{r,T} = 500$ nM and $g_{f,T} = 10$ nM. Decay tag, $g_{r,T} = 500$ nM, $g_{f,T} = 10$ nM, $\delta_{p_f} = 15$ h^{-1}.

Transcription factor	k_X (nM)	Ref.	μ (nM)	η	Ref.
LacI	550	[102]	0.02	~ 4	[94]
cI	100	[103]	20	~ 2	[108]
Cro	300	Table 5.6	1000	~ 2	[109]
RstR	410	Table 5.6	3.9	~ 4	[110]
TetR	350	Table 5.6	5.6	~ 2	[107]
LmrR	150	Table 5.6	64	~ 2	[112]

Table 5.5: **Transcription factor parameters.**

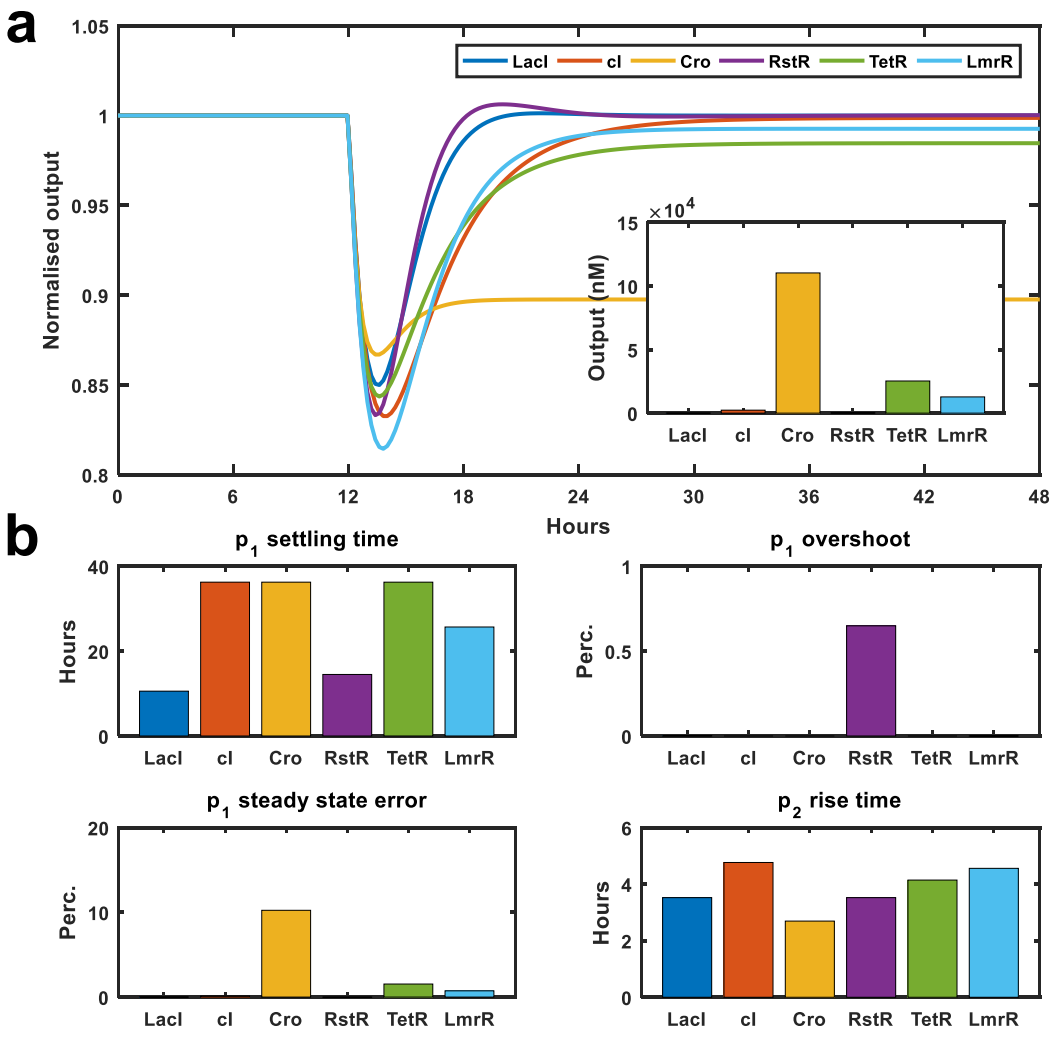

Figure 5.9: **Assessing the dynamics of chosen biological implementations.** **(a)** Circuit dynamics showing the normalised levels of protein 1. Inset, steady state output at t = 48 h. **(b)** Characterisation of the response of p_1 to the disturbance caused by u_2. Settling time, number of hours from the induction until p_1 returns to steady state; Overshoot, transient increase in p_1. Steady state error, difference between $p_1(t = 48)$ and $p_1(t = 12)$. Rise time, time it takes for p_2 to increase from 10% of its steady state to 90% of its steady state. See Table 5.7 for implementations in this figure.

Repressor	Promoter sequence (Position -41 to -1)	Ref.
TetR	TAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCAC	$*$
Cro	GATCTGTTTATAGCTGGCTATATAAAATTTATAAATTATAC	[109]
RstR	TTAGTCTTGAAAGTGCGCATTGGTTGCTGTATTTTAGCTCT	[110]
LmrR	GTACCTTGACTTACATAGTAATGTGAAGTATAATATACTTT	[111]

Table 5.6: **Promoter sequences used to determine k_X** Where promoter dissociation constants were not available from the literature a thermodynamic calculator was used to estimate RNA polymerase dissociation constant as described. $*$, The sequence of the TetR-responsive promoter was retrieved from iGEM Registry of Standard Biological Parts (Part no. BBa_R0040, URL http://parts.igem.org/Part:BBa_R0040, Retrieved 28/06/2017)

Repressor	rRNA copy number	TF copy number	Decay tag
LacI	Medium	Medium	Yes
cI	Medium	Medium	No
Cro	High	Chromosome	Yes
RstR	High	Chromosome	No
TetR	High	Chromosome	Yes
LmrR	High	Chromosome	No

Table 5.7: **Biological designs simulated in Figure 5.9.** Potential biological implementations based on the repressors found in our literature survey. Different genetic arrangements for the rRNA and repressor are as specified. Medium copy number $\approx$ 100 nM, High copy number $\approx$ 500 nM, Chromosomal integrated genes are assumed to have low copy number 10 nM. In the presence of a protein degradation tag we assume $\delta_{p_f} \approx 15$ h^{-1}.

Fig.	i	η	μ (nM)	g_T (nM)	k_X (nM)	τ (h^{-1})	δ_m (h^{-1})	k_L (nM)	γ (h^{-1})	δ_p (h^{-1})
5.10	1...N	1	100	500	500	320	20	10^5	240	1
5.11	1...3	4	0.02	500	500	320	20	10^5	240	1
5.11	4	1	1	500	500	320	20	10^4	240	1
5.12	1	2	200	10...500	275	320	20	10^3...10^7	240	1
5.12	2	2	200	10...500	275	320	20	10^5	240	1

Table 5.8: **Parameters for simulations of more complex circuits in Section 5.8.**

dependent failure in a variety of more complex gene circuits.

5.8.1 Independence of multiple single-input-single-output motifs

We initially simulate multiple SISO modules with new modules being activated at different intervals (Figure 5.10). In the absence of the controller, activation of each additional module has an impact on the previously activated modules. For example, the expression of the first module p_1 falls by over 50% as three additional genes are induced. As shown in Figure 5.10, the controller successfully eliminates this coupling, making p_1 relatively insensitive to the induction of over 10 additional genes. Note, however, that the rise time and settling time increase slightly with the induction of each additional gene due to the dynamics of o-ribosome assembly.

5.8.2 Imposing modularity on an oscillator

A key principle of synthetic biology is that previously characterised components or devices can be introduced into the same cell to form a complex circuit. Here we assess the effect of introducing two separately characterised devices into one complex circuit, i.e. we want to investigate the effect of introducing an additional resource consumer on a previously characterised device. As the production of robust genetic oscillators to create clocks for temporal functions is of fundamental importance in synthetic circuit design, we consider designs for the repressilator clock and an additional SISO module.

The repressilator is composed of a three gene network where each gene (i) inhibits the next ($i + 1$) in turn [4]. The repressilator genes $i = \{1...3\}$ have the following

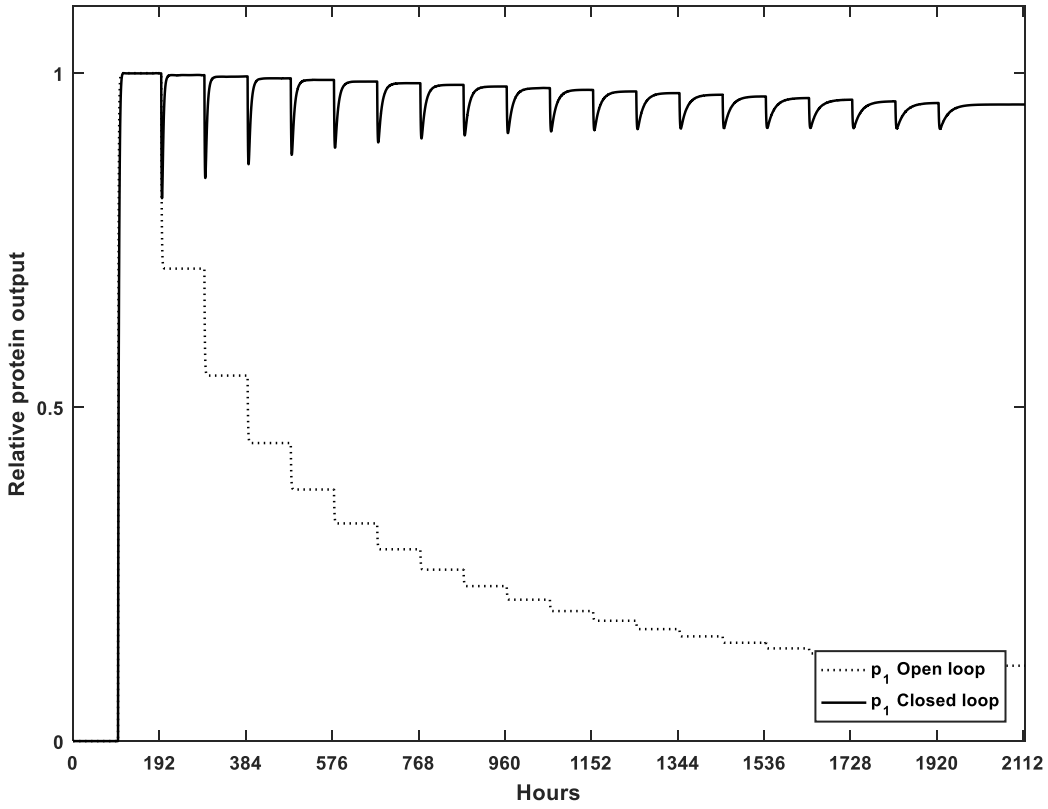

Figure 5.10: **Multiple single-input-single-output motifs.** The controller successfully renders a gene invulnerable to the induction of many additional genes at 100 h intervals. All y-axes are normalised output. The other genes are not shown.

dynamics:

$$\dot{p}_i = \underbrace{\gamma_i \left(\frac{\tau_i}{\delta_{m_i}} \frac{R}{k_{L_i}} \bar{x}_i \right)}_{p_i \text{ production}} - \underbrace{\delta_{p_i} p_i}_{\text{decay}} - \underbrace{\eta_i (g_{i+1,T} - \bar{x}_{i+1} - \bar{\kappa}_{i+1})({p_i}^{\eta_i})}_{\text{binding}} + \underbrace{\eta_i \mu_i \bar{\kappa}_{i+1}}_{\text{unbinding}} \tag{5.46}$$

where the demand for RNA polymerase by each gene i is given by:

$$\hat{x}_i = \frac{g_{i,T}}{k_{X_i}} \frac{\mu_i}{\mu_i + {p_{i-1}}^{\eta_{i-1}}} \tag{5.47}$$

$\bar{x}_i$ is given by substituting $\hat{x}_i$ into Equation 5.30.

The quasi-steady state concentration of the inhibited promoter of gene i is:

$$\bar{\kappa}_i = (g_{i,T} - \bar{x}_i) \Big(\frac{{p_{i-1}}^{\eta_{i-1}}}{\mu_{i_1} + {p_{i-1}}^{\eta_{i-1}}} \Big) \tag{5.48}$$

Note that both $\hat{x}_i$ and $\bar{\kappa}_i$ are functions of the previous protein in the repressilator sequence p_{i-1}, not external inputs u.

The binding and unbinding portions of Equation 5.46 represent the titrating effect as p_i proteins bind to g_{i+1} and inhibit the promoter.

The dynamics of the additional inducible protein (p_4) are given by:

$$\dot{p}_4 = \gamma_4 \left(\frac{\tau_4}{\delta_{m_4}} \frac{R}{k_{L_4}} \bar{x}_4 \right) - \delta_{p_4} p_4 \tag{5.49}$$

where the RNA polymerase demand is given by:

$$\hat{x}_4 = \frac{g_{4,T}}{k_{X_4}} \frac{u^{\eta_4}}{\mu + u^{\eta_4}} \tag{5.50}$$

Note the presence of the external input u which acts to induce p_4 via its effects on $\hat{x}_4$.

These modules are first simulated separately, as shown in (Figure 5.11a, *upper panels*). Upon linking these separate devices through a common pool of resources, i.e. coupled through their competition for ribosomes, we see that p_4 induction destroys the oscillations of the repressilator (Figure 5.11a, *lower panels*).

If, however, we consider the design of these two devices in the presence of the controller and then introduce them into the same resource pool as before, we see

that circuit function is now maintained (Figure 5.11b, *lower panels*). Note that while there is still a small loss in repressilator amplitude upon induction of p_4 this is significantly reduced, thus staying closer to the original device behaviour.

5.8.3 Restoring the input-output response of an activation cascade

It has previously been shown that resource limitations can change the input-output response of a simple genetic activation cascade [45]. Qian *et al.* show that if the upstream module has a stronger ability to sequester ribosomes than the downstream module (a small k_{L_1}-to-k_{L_2} ratio) then the expected response determined from simple Hill-function type modelling (i.e. an increasing output to increasing input in a step-like fashion) can become biphasic or even invert (Figure 5.12, dotted lines).

We modelled the activation cascade as follows. The protein dynamics of the upstream gene ($i = 1$) are:

$$\dot{p}_1 = \gamma_1 \left(\frac{\tau_1}{\delta_{m_1}} \frac{R}{k_{L_1}} \bar{x}_1 \right) - \delta_{p_1} p_1 - \eta_2 (g_{2,T} - \bar{x}_2 - \bar{\kappa}_2)({p_1}^{\eta_2}) + \eta_2 \mu_2 \bar{\kappa}_2 \tag{5.51}$$

where $\bar{x}_1$ and $\bar{x}_2$ are calculated by substituting $\hat{x}_1$ and $\hat{x}_2$ into Equation 5.30.

$\hat{x}_1$ and $\hat{x}_2$ are given by:

$$\hat{x}_1 = \frac{g_{1,T}}{k_{X_1}} \Big(\frac{u^{\eta_1}}{\mu_1 + u^{\eta_1}} \Big) \qquad\qquad \hat{x}_2 = \frac{g_{2,T}}{k_{X_2}} \Big(\frac{{p_1}^{\eta_2}}{\mu_2 + {p_1}^{\eta_2}} \Big) \tag{5.52}$$

The quasi-steady state concentration of the activated second promoter ($i = 2$) is given by:

$$\bar{\kappa}_2 = (g_{2,T} - \bar{x}_2) \Big(\frac{{p_1}^{\eta_2}}{\mu_2 + {p_1}^{\eta_2}} \Big) \tag{5.53}$$

The dynamics of the downstream gene $i = 2$ are given by:

$$\dot{p}_2 = \gamma_2 \left(\frac{\tau_2}{\delta_{m_2}} \frac{R}{k_{L_2}} \bar{x}_2 \right) - \delta_{p_2} p_2 \tag{5.54}$$

We simulated a range of prototype activation cascades in the absence and presence of our controller. In the absence of the controller, no additional resources are available as demand increases and so we see the activation cascade failing in the same manner as found in [45]. In the presence of the controller, the desired behaviour of the activation cascade is restored, as translational capacity is directed to the circuit

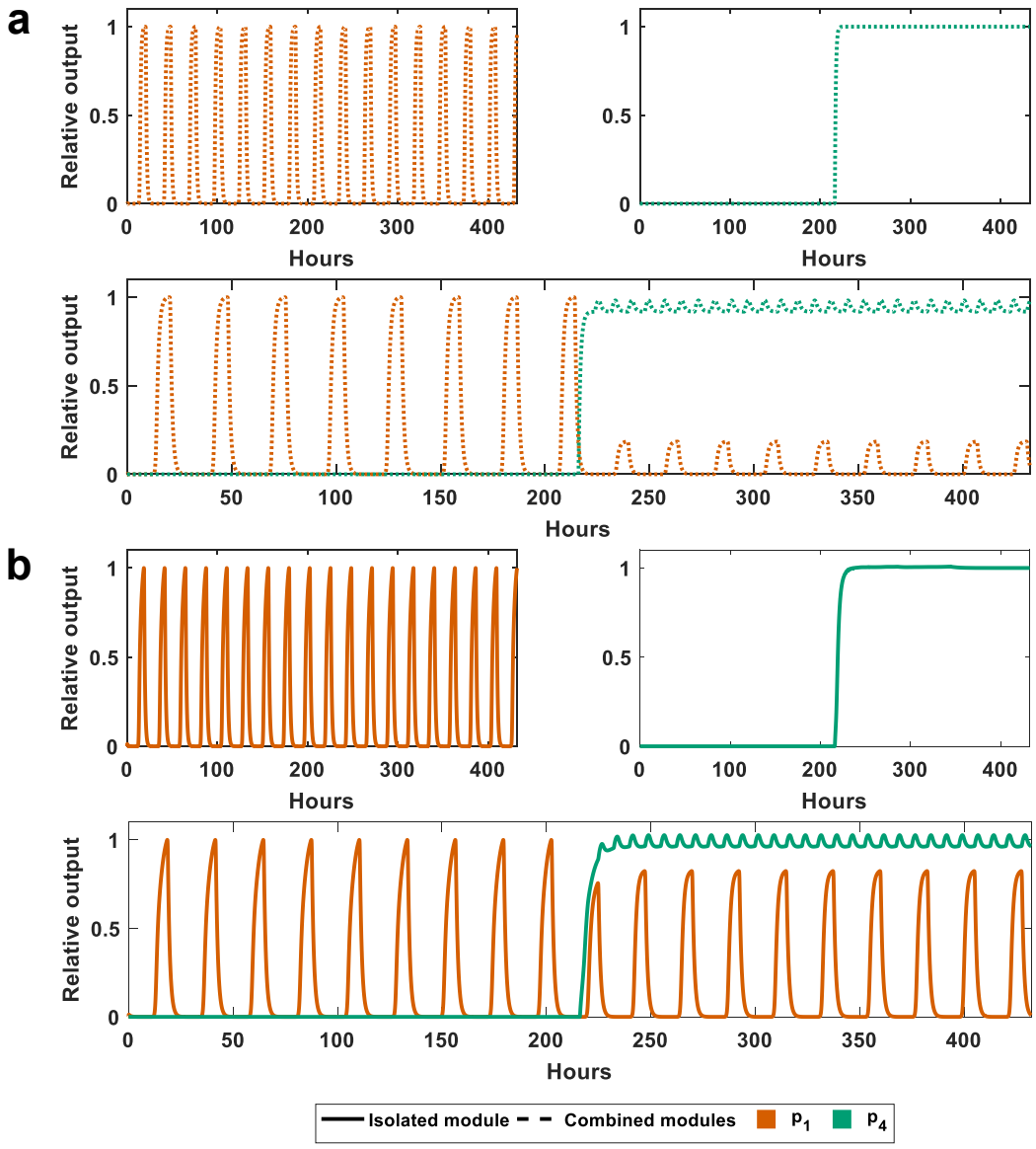

Figure 5.11: **Imposing modularity on an oscillator.** The controller acts to maintain the behaviour of the repressilator in the presence of an induced gene. The repressilator (protein p_1 to p_3, note that only p_1 is shown) is simulated before an additional gene p_4 with a stronger RBS is induced just after 200 h at $u = 500$ nM. *Upper panels*, show the function of the individual modules (repressilator and additional gene respectively) alone. *Lower panel* shows the function of the two modules in one circuit. **(a)** Open loop conformation (no controller). **(b)** Closed loop conformation (with controller).

as demand increases. The controller acts to remove the resource limitation, thus allowing simpler models, which often do not account for limited cellular resources, to be used to produce circuit designs which then function as expected *in vivo* (Figure 5.12).

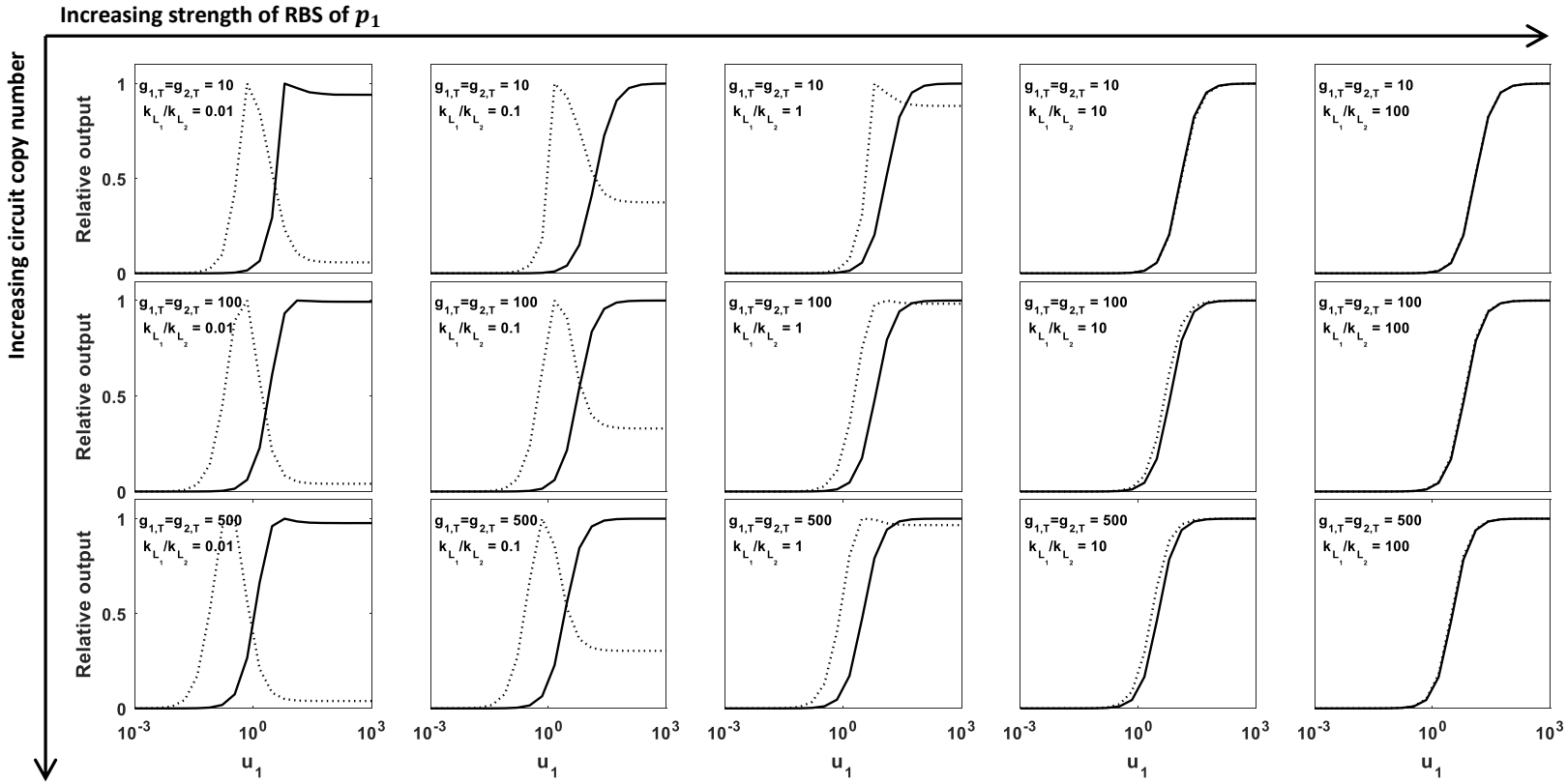

Figure 5.12: **Restoring the behaviour of an activation-cascade.** The controller removes resource limitation induced failure in the design of an activation cascade ($u_1 \rightarrow p_1 \rightarrow p_2$). In the absence of the controller (dotted line) some prototype designs do not show the monotonically increasing output of p_2 to u_1 as desired in an activation cascade. The controller removes these resource limitations allowing the circuit to function as expected across all prototype designs.

5.9 Conclusions

In this chapter, we developed a detailed mechanistic model of gene expression and resource allocation, which when simplified to a tractable level of complexity, allows the rational design of optimal translational controllers. We demonstrated that this new model allows the design of controllers which can dynamically allocate orthogonal ribosomes to synthetic circuits within reasonable timeframes (< 12 hours). Using our model, we identified a fundamental trade-off in controller design; that reducing coupling acts to decrease gene expression. We determined how each controller design parameter affects the overall closed-loop behaviour of the system, leading to a detailed set of design guidelines for optimally managing this trade-off. We found that both controller co-operativity and RBS strength are key parameters in determining the level of decoupling that can be achieved. Based on our designs, we identified and evaluated a number of alternative potential experimental implementations of the proposed system using commonly available biological components. Finally, we showed that our controller is capable of dynamically allocating ribosomes as needed to restore modularity in a number of more complex synthetic circuits composed of multiple interacting modules, such as the repressilator and activation cascades.

Chapter 6

Conclusions and future works

Recently, a number of opinion pieces have highlighted the potential of orthogonal ribosomes for the expression of circuit genes [74, 113–117]. During the course of this work, we have gone some way towards validating the proposals outlined in these articles. Using a host-aware modelling approach we explored the ability of orthogonal ribosomes to function as resource allocators for synthetic circuit gene expression and showed that this can be used as a means of reducing gene coupling mediated by the sharing of cellular resources. Furthermore, we went on to show that this pool of orthogonal ribosomes can be dynamically controlled by simple genetic motifs which act to control the supply of translational capacity to synthetic circuit genes. In this way our controllers go some way to forming the 'virtual machine' proposed by Liu *et al.* as they manage the level of resources that host and circuit processes can utilise [117].

In Chapter 2, we discussed how host constraints can impact circuit function. Specifically, our literature review highlights how limitations in levels of the gene expression machinery, particularly ribosomes, can result in the emergence of unwanted and initially unpredictable *non*-regulatory interactions between genes. We highlighted (and demonstrated via a motivating example) the need for more focused modelling efforts in synthetic biology to 'design around' host-circuit interactions – which we term a 'host-aware' approach which includes basic host factors, energetic limitations and gene expression resources. We proposed that so-called orthogonal ribosomes could be used to reduce these effects by creating a circuit-specific translational capacity.

In Chapter 3, we developed a model of bacterial growth by extending previously published work. Crucially, this host-aware model captures the key trade-offs which

govern host-circuit interactions; namely, competition for space, ribosomes and energy. This model allows the prediction of effects which modulate circuit function such as growth rates and resource levels. Using this model we demonstrated that the use of orthogonal ribosomes for circuit gene expression is tolerated by cells (from a cellular economy point of view, neglecting toxicity etc.). We went on to show that the use of host and orthogonal ribosomes can be employed as a crude resource allocation scheme when circuit or pathway genes are subdivided between the two pools. Comparison with experimental data shows that this model qualitatively recapitulates the observations *in vivo* and can be used to predict the optimal distribution of genes in a simple two-gene circuit and across a metabolic pathway; namely constitutively expressed genes should be translated by the o-ribosome pool and dynamic (induced) genes by the host pool. Crucially, the use of a model at the level of abstraction chosen allows for the incorporation of the effects of metabolic pathways on gene expression (which is lacking from frameworks such as flux balance analysis).

These resource allocators do not respond dynamically to changes in circuit composition (i.e. will not change based on additional gene induction) and so in Chapter 4, we developed three feedback controllers. The first, which we term the F-controller, is composed of a simple negative feedback loop and successfully decouples genes showing a decreased isocost line slope in most parameter ranges. We demonstrated the action of the controller over a number of two-gene circuits with significantly different demands and showed that in all cases the controller successfully decouples co-expressed genes. Our Monte Carlo simulations find that the F-controller always produces a stable output. However, because this controller lacks integral action it will always result in a constant steady state error and so we investigated the possibility of implementing two putative integral controllers based on a recently published chemical reaction network scheme. We showed that the first integral controller, G, does not result in zero steady state error (i.e. it does not function as a perfect integral controller) but 'quasi-integral' action can be achieved by tuning parameters. We reduced the complexity of the model in an attempt to regain integral action and confirmed the results presented in [91] that the decay and dilution of intermediate species causes the loss of the integral action. Analysing the dynamics we observed that in many cases the G-controller shows high levels of overshoot and long term damped oscillations. We made some recommendations for designs which should be avoided. We developed another integral controller, which we called K, which is based on an RNA binding protein that acts to sequester the o-rRNA if it is not needed for o-ribosome formation. Whilst this controller's stability is more robust to parametric uncertainty than the G-controller, our model predicts significant growth

defects which would preclude *in vivo* implementation.

In light of these observations, we liaised with our experimental colleagues to design and produce a prototype of the F-controller. This prototype is able to reduce coupling by 50% in a simple two-gene circuit.

Having successfully implemented the prototype, in Chapter 5 we developed a detailed mechanistic model of the controller to allow for quantitative design. Using the observation that the strains carrying the prototype grew at an intermediate rate, we were able to parameterise our model based on literature values for cell composition and reaction rates at these growth rates. We identified a hard trade-off between decoupling and expression levels; with decoupling coming at the expense of expression. We demonstrated how each design parameter relates to this trade-off and provide potential biological implementations. We showed how the controller reduces resource-mediated failures in modularity without the need for circuit redesign across a range of complex synthetic gene circuits.

6.1 Future work

This thesis has shown that the use of o-ribosomes for circuit gene expression can reduce resource-mediated gene coupling and that orthogonal ribosomes can be dynamically allocated by genetic feedback controllers. The main conclusions in Chapter 3 and the F-controller from Chapter 4 were successfully validated *in vivo*. Significant experimental investment is still required to fully validate the dynamic observations and conclusions presented in Chapter 5.

At present the experimental system leads to poor expression levels. In Chapter 3, we demonstrated using our model that this is likely due to both relatively low assembly of o-ribosomes and poor RBS strength. The experimental system requires further refinement; for example, by further directed evolution or using a thermodynamic model to design complementary 16S rRNA and RBS, which would also minimise crosstalk with the host system. We also showed that the use of tethered ribosomes can increase gene expression while maintaining coupling profiles, which is yet to be validated *in vivo*. In both the trade-off and mechanistic models we have assumed a simple single step co-option of host to orthogonal ribosomes. In Chapter 5, we used the host 16S rRNA:small subunit assembly rates from an *in vitro* reconstitution assay as proxy for those of the orthogonal system. During future characterisation of the o-ribosomal system, the number of orthogonal ribosomes yielded per o-rRNA

transcribed should be experimentally determined to allow (i) the number tolerated with negligible impact on growth to be assessed, and (ii) the approximate rate of the assembly reaction to be determined.

This thesis has focused on three potential implementations of negative feedback controllers. However, by definition, negative feedback is reactive; taking an output error and using this to modify the input. To calculate the error signal in our systems requires the production of a protein by the o-ribosome pool. The expression level of this constitutively expressed protein then becomes a proxy for the error signal which is created when other circuit proteins utilise the o-ribosome pool. This error-calculation necessitates that a fraction of the orthogonal ribosome pool be dedicated to its own controller; although as shown in Chapter 5, if some coupling can be tolerated this may only be small. Given the current poor protein production in the prototype o-ribosome system this represents a waste of protein synthesis capacity. Feedforward control provides a means to remove this constraint. In feedforward control systems the disturbance is measured and corrective action taken before the input signal is fed into the system [118]. To implement feedforward control we are designing novel RNA-based networks which act to sense mRNA levels, rather than translation complex levels, to control o-rRNA production. These networks have the potential to function faster and have reduced impacts on host growth by utilising RNA for their computation [119, 120].

In Chapter 3, we found that competition for the ribosomal protein component was sufficient to cause circuit failure when the induced gene is expressed from the o-ribosome pool. To remove this competition we could develop controllers which function at the 'interface' of the host and orthogonal system. For example, the F-controller architecture from Chapter 4 can be extended to include inhibition of the host rRNA genes. Upon sensing the disturbance caused by the induction of a circuit gene, the controller acts to not only increase the o-rRNA but also decreases host rRNA production. This would reduce competition for 'empty' ribosomes. Preliminary results suggest that this could also allow for increased gene expression. Additionally, we propose the integration of the burden-based feedback mechanism proposed by Ceroni *et al.* [82] into our o-ribosome system as a means to modulate circuit function with minimal components.

At present in the experimental implementation of the controller, we observe a significant remaining level of coupling. This can be improved by controller redesign using the concepts within Chapters 4 and 5. Our mechanistic model neglects the coupling at the RNA polymerase level due to the assumptions outlined in Sections

5.2 and 5.3; rather it uses the RNA polymerase concentration to set the maximum expression for each gene. If we restore RNA polymerase competition by relaxing these assumptions, then a small amount of gene coupling returns. In the future this assumption should be removed to allow for the design of translational controllers which manage any emerging trade-offs at transcription. Transcription is known to be a noisier process than translation and therefore different transcription kinetics (and hence different noise profiles) can lead to vast variation in protein levels across a population of cells [121]. In light of this any ordinary differential equation modelling (which is favoured for design and analysis) of transcription should be supplemented with stochastic approaches (such as simulations using Gillespie's algorithm or the addition of noise by the use of the Chemical Langevin Equation [65]) to allow the assessment of noise.

By incorporating species representing an orthogonal RNA polymerase we can develop controllers which manage transcriptional coupling. By combining these *transcriptional* controllers with our *translational* controller we can create a new generation of 'dual' controllers. Preliminary modelling suggests that combining our negative feedback F-type translational controller with currently published (and novel) RNA polymerase controllers can yield further decoupling than implementing control at translation alone. These dual controllers, and the implications of any noise arising from the transcriptional component, will be fully investigated in future work.

From an industrial biotechnology point of view, the initial results in Chapter 3 suggest that by controlling resource distribution using o-ribosomes we can increase flux through a prototype pathway. The production of complex molecules, such as macrolide antibiotics, requires biosynthetic operons composed of numerous large enzymes (reviewed in [122]). The expression of these pathways in heterologous hosts is often difficult in part due to resource limitations and subsequent competition between genes. Our results show that allocating resources by the use of o-ribosomes may relieve these problems and increase yield. In the future a more thorough investigation of how pathway genes are allocated to each ribosome pool will be carried out by considering all possibilities in turn. By combining translational resources, both host and orthogonal, with genome scale metabolic pathway reconstructions in a Resource Balance Analysis (see e.g. [62]) we can better manage the cellular economy and design novel metabolic pathways.

Bibliography

[1] J.A.J. Arpino, E.J. Hancock, J. Anderson, M. Barahona, G-B. Stan, A. Papachristodoulou, and K. Polizzi. Tuning the dials of synthetic biology. *Microbiology*, 159(7): 1236–1253, 2013.

[2] A.S. Khalil and J.J. Collins. Synthetic biology: Applications come of age. *Nature Reviews Genetics*, 11(5): 367–379, 2010.

[3] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767): 339–342, 2000.

[4] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767): 335–338, 2000.

[5] J. Stricker, S. Cookson, M.R. Bennett, W.H. Mather, L.S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221): 516–519, 2008.

[6] S. Basu, Y. Gerchman, C.H. Collins, F.H. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037): 1130–1134, 2005.

[7] A. Tamsir, J.J. Tabor, and C.A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, 469(7329): 212–215, 2011.

[8] J.A.N. Brophy and C.A. Voigt. Principles of genetic circuit design. *Nature Methods*, 11(5): 508–20, 2014.

[9] A.E. Friedland, T.K. Lu, X. Wang, D. Shi, G. Church, and J.J. Collins. Synthetic gene networks that count. *Science*, 324(5931): 1199–1202, 2009.

[10] A.K. Nielsen, B.S. Der, J. Shin, P. Vaidyanathan, D. Densmore, and C.A. Voigt. Genetic circuit design automation. *Science*, 352(6281): 53, 2016.

[11] T. Danino, O. Mondragón-Palomino, L. Tsimring, and J. Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279): 326–330, 2010.

[12] Y. Chen, J.K. Kim, A.J. Hirning, K. Josie, and M.R. Bennett. Emergent genetic oscillations in a synthetic microbial consortium. *Science*, 349(6351): 986–989, 2015.

[13] E. Andrianantoandro, S. Basu, D.K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2: 2006.0028, 2006.

[14] B. Canton, A. Labno, and D. Endy. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology*, 26(7): 787–793, 2008.

[15] iGEM Foundation. Registry of Standard Biological Parts. URL `http://parts.igem.org/Main_Page`. Accessed 21/02/2018

[16] M. Galdzicki, K.P. Clancy, E. Oberortner, M. Pocock, J.Y. Quinn, C.A. Rodriguez, N. Roehner, M.L. Wilson, L. Adam, J.C. Anderson, B.A. Bartley, J. Beal, D. Chandran, J. Chen, D. Densmore, D. Endy, R. Grünberg, J. Hallinan, N.J. Hillson, J.D. Johnson, A. Kuchinsky, M. Lux, G. Misirli, J. Peccoud, H.A. Plahar, E. Sirin, G-B. Stan, A. Villalobos, A. Wipat, J.H. Gennari, C.J. Myers, and H.M. Sauro. The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology*, 32(6): 545–50, 2014.

[17] T.E. Gorochowski, A. Espah-Borujeni, Y. Park, A.A.K. Nielsen, J. Zhang, B.S. Der, D.B. Gordon, and C.A. Voigt. Genetic circuit characterization and debugging using RNA-seq. *Molecular Systems Biology*, 13(952): 1–16, 2017.

[18] R. Kwok. Five hard truths for synthetic biology. *Nature*, 463: 36–42, 2010.

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

[20] C. Zhang, R. Tsoi, and L. You. Addressing biological uncertainties in engineering gene circuits. *Integrative Biology*, 8(4): 456–464, 2015.

[21] B.C. Stanton, A.A.K. Nielsen, A. Tamsir, K. Clancy, T. Peterson, and C.A. Voigt. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology*, 10(2): 99–105, 2014.

[22] N. Perry, E.M. Nelson, and G. Timp. Wiring together synthetic bacterial consortia to create a biological integrated circuit. *ACS Synthetic Biology*, 5(12): 1421–1432, 2016.

[23] A. Espah-Borujeni, A.S. Channarasappa, and H.M. Salis. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4): 2646–2659, 2014.

[24] C.J. Bashor and J.J. Collins. Insulating gene circuits from context by RNA processing. *Nature Biotechnology*, 30(11): 1061–1062, 2012.

[25] A. Gyorgy and D. Del Vecchio. Modular composition of gene transcription networks. *PLOS Computational Biology*, 10(3): e1003486, 2014.

[26] D. Del Vecchio, A.J. Ninfa, and E.D. Sontag. Modular cell biology: Retroactivity and insulation. *Molecular Systems Biology*, 4(161): 161, 2008.

[27] D. Mishra, P.M. Rivera, A. Lin, D. Del Vecchio, and R. Weiss. A load driver device for engineering modularity in biological networks. *Nature Biotechnology*, 32(12): 1268–1275, 2014.

[28] S. Cardinale, M.P. Joachimiak, and A.P. Arkin. Effects of genetic variation on the *E. coli* host-circuit interface. *Cell Reports*, 4(2): 231–237, 2013.

[29] C. Vilanova, K. Tanner, P. Dorado-Morales, P. Villaescusa, D. Chugani, A. Frías, E. Segredo, X. Molero, M. Fritschi, L. Morales, D. Ramón, C. Peña, J. Peretó, and M. Porcar. Standards not that standard. *Journal of Biological Engineering*, 9: 17, 2015.

[30] F. Moser, N.J. Broers, S. Hartmans, A. Tamsir, R. Kerkman, J.A. Roubos, R. Bovenberg, and C.A. Voigt. Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synthetic Biology*, 1(11): 555–564, 2012.

[31] R. Milo and R. Phillips. *Cell Biology by the Numbers*. Garland Science, New York, NY, USA, 2016.

[32] S. Klumpp and T. Hwa. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 105(51): 20245–50, 2008.

[33] M. Scott, C.W. Gunderson, E.M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007): 1099–1102, 2010.

[34] J. Vind, M.A. Sørensen, M.D. Rasmussen, and S. Pedersen. Synthesis of proteins in *Escherichia coli* is limited by the concentration of free ribosomes: Expression from reporter genes does not always reflect functional mRNA levels. *Journal of Molecular Biology*, 231(3): 678–88, 1993.

[35] A. Goelzer and V. Fromion. Bacterial growth rate reflects a bottleneck in resource allocation. *Biochimica et Biophysica Acta*, 1810(10): 978–988, 2011.

[36] J. Carrera, G. Rodrigo, V. Singh, B. Kirov, and A. Jaramillo. Empirical model and in vivo characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnology Journal*, 6(7): 773–783, 2011.

[37] F. Ceroni, R. Algar, G-B. Stan, and T. Ellis. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nature Methods*, 12(5): 415–423, 2015.

[38] A. Gyorgy, J.I. Jiménez, J. Yazbek, H.-H. Huang, H. Chung, R. Weiss, and D. Del Vecchio. Isocost lines describe the cellular economy of genetic circuits. *Biophysical Journal*, 109(3): 639–646, 2015.

[39] T.E. Gorochowski, I. Avcilar-Kucukgoze, R.A.L. Bovenberg, J.A. Roubos, and Z. Ignatova. A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes. *ACS Synthetic Biology*, 5(7): 710–720, 2016.

[40] M. Carbonell-Ballestero, E. Garcia-Ramallo, R. Montanez, C. Rodriguez-Caso, and J. Marcia. Dealing with the genetic load in bacterial synthetic biology circuits: Convergences with the Ohm's law. *Nucleic Acids Research*, 44(1): 496–507, 2015.

[41] Q. Liu, J. Schumacher, X. Wan, C. Lou, and B. Wang. Orthogonality and burden of heterologous AND gate gene circuits in *E. coli*. *ACS Synthetic Biology*, 7(2): 553-564, 2018.

[42] N.A. Cookson, W.H. Mather, T. Danino, O. Mondragón-Palomino, R.J. Williams, L.S. Tsimring, and J. Hasty. Queueing up for enzymatic processing:

Correlated signaling through coupled degradation. *Molecular Systems Biology*, 7(561), 2011.

[43] A. Prindle, J. Selimkhanov, H. Li, I. Razinkov, L.S. Tsimring, and J. Hasty. Rapid and tunable post-translational coupling of genetic circuits. *Nature*, 508(7496): 387–391, 2014.

[44] T. Shopera, L. He, T. Oyetunde, Y.J. Tang, and T.S. Moon. Decoupling resource-coupled gene expression in living cells. *ACS Synthetic Biology*, 6(8): 1596–1604, 2017.

[45] Y. Qian, H-H. Huang, J.I. Jiménez, and D. Del Vecchio. Resource competition shapes the response of genetic circuits. *ACS Synthetic Biology*, 6(7): 1263–1272, 2017.

[46] Y. Qian and D. Del Vecchio. Effective interaction graphs arising from resource limitations in gene networks. *Proceedings of the American Control Conference*, Chicargo, IL, USA, 4417–4423, 1-3 July 2015.

[47] S. Klumpp and T. Hwa. Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. *Current Opinion in Biotechnology*, 28: 96–102, 2014.

[48] C. Tan, P. Marguet, and L. You. Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chemical Biology*, 5(11): 842–848, 2009.

[49] O. Borkowski, C. Bricio-Garberi, M. Murgiano, G-B. Stan, and T. Ellis. Cell-free prediction of protein expression costs for growing cells. *Nature Communications*, 9: 1457, 2018.

[50] S.C. Sleight, B.A. Bartley, J.A. Lieviant, and H.M. Sauro. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering*, 4(1):12, 2010.

[51] S.C. Sleight and H.M. Sauro. Visualization of evolutionary stability dynamics and competitive fitness of *Escherichia coli* engineered with randomized multigene circuits. *ACS Synthetic Biology*, 2(9): 519–528, 2013.

[52] A. Hamadeh and D. del Vecchio. Mitigation of resource competition in synthetic genetic circuits through feedback regulation. *Proc. of 53rd IEEE Conference on Decision and Control*, Los Angeles, CA, USA, 15-17 December 2014.

[53] A. Martínez-Antonio, S.C. Janga, and D. Thieffry. Functional organisation of *Escherichia coli* transcriptional regulatory network. *Journal of Molecular Biology*, 381(1): 238–247, 2008.

[54] Y. Qian and D. Del Vecchio. Mitigation of ribosome competition through distributed sRNA feedback. *Proc. of 55th IEEE Conference on Decision and Control*, Las Vagas, NA, USA, 12-14 December 2016.

[55] C. McBride and D. Del Vecchio. Analyzing and exploiting the effects of protease sharing in genetic circuits. *Proc. of 20th IFAC World Congress*, Tolourse, France, 9-14 July 2017.

[56] A.Y. Weiße, D.A. Oyarzún, V. Danos, and P.S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9): E1038–E1047, 2015.

[57] C. Liao, A.E. Blanchard, and T. Lu. An integrative circuit-host modelling framework for predicting synthetic gene network behaviours. *Nature Microbiology*, 2: 1658-1666, 2017.

[58] J.R. Karr, J.C. Sanghvi, D.N. MacKlin, M.V. Gutschow, J.M. Jacobs, B. Bolival, N. Assad-Garcia, J.I. Glass, and M.W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2): 389–401, 2012.

[59] O. Purcell, B. Jain, J.R. Karr, M.W. Covert, and T.K. Lu. Towards a whole-cell modeling approach for synthetic biology. *Chaos*, 23(2): 1–8, 2013.

[60] J. Carrera, R. Estrela, J. Luo, N. Rai, and A. Tsoukalas. An integrative, multi-scale, genome-wide model reveals the phenotypic landscape of *Escherichia coli*. *Molecular Systems Biology*, 10(735): 1–12, 2014.

[61] D.N. Macklin, N.A. Ruggero, and M.W. Covert. The future of whole-cell modeling. *Current Opinion in Biotechnology*, 28: 111–115, 2014.

[62] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6): 1210–1218, 2011.

[63] D. Siegal-Gaskins, Z.A. Tuza, J. Kim, V. Noireaux, and R.M. Murray. Gene circuit performance characterization and resource usage in a cell-free "breadboard". *ACS Synthetic Biology*, 3(6): 416–425, 2014.

[64] A. Gyorgy and R.M. Murray. Quantifying resource competition and its effects in the TX-TL system. *Proc. of 55th IEEE Conference on Decision and Control*, Las Vagas, NA, USA, 12-14 December 2016.

[65] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, and R. Herwig. *Systems Biology: A Textbook*. John Wiley & Sons, Berlin, Second edition, 2016.

[66] S.K. Vogel, A. Schulz, and K. Rippe. Binding affinity of *Escherichia coli* RNA polymerase σ_{54} holoenzyme for the *glnAp2*, *nifH* and *nifL* promoters. *Nucleic Acids Research*, 30(18): 4094–101, 2002.

[67] I. Bervoets, M. Van Brempt, K. Van Nerom, B. Van Hove, J. Maertens, M. De Mey, and D. Charlier. A sigma factor toolbox for orthogonal gene expression in *Escherichia coli*. *Nucleic Acids Research*, *Electronic publication ahead of print, available at* doi: 10.1093/nar/gky010, 2018.

[68] R. Sousa and S. Mukherjee. T7 RNA Polymerase. *Progress in Nucleic Acid Research and Molecular Biology*, 73: 1–41, 2003.

[69] D.E. Cameron and J.J. Collins. Tunable protein degradation in bacteria. *Nature Biotechnology*, 32(12): 1276–1281, 2014.

[70] O. Rackham and J.W. Chin. A network of orthogonal ribosome·mRNA pairs. *Nature Chemical Biology*, 1(3): 159–166, 2005.

[71] L.M. Chubiz and C.V. Rao. Computational design of orthogonal ribosomes. *Nucleic Acids Research*, 36(12): 4038–4046, 2008.

[72] C. Orelle, E.D. Carlson, T. Szal, T. Florin, M.C. Jewett, and A.S. Mankin. Protein synthesis by ribosomes with tethered subunits. *Nature*, 524: 119–124, 2015.

[73] S.D. Fried, W.H. Schmied, C. Uttamapinant, and J.W. Chin. Ribosome subunit stapling for orthogonal translation in *E. coli*. *Angewandte Chemie*, 54(43): 12791–12794, 2015.

[74] B.J. Des Soye, J.R. Patel, F.J. Isaacs, and M.C. Jewett. Re-purposing the translation apparatus for synthetic biology. *Current Opinion in Chemical Biology*, 28: 83–90, 2015.

[75] Y. Schaerli, M. Gili, and M. Isalan. A split intein T7 RNA polymerase for transcriptional AND-logic. *Nucleic Acids Research*, 42(19): 12322–12328, 2014.

[76] W. An and J.W. Chin. Synthesis of orthogonal transcription- translation networks. *Proceedings of the National Academy of Sciences*, 106(21): 8477–8482, 2009.

[77] B. Jia, H. Qi, B-Z. Li, S. Pan, D. Liu, H. Liu, Y. Cai, and Y-J. Yuan. Orthogonal ribosome bio-firewall. *ACS Synthetic Biology*, 6(11): 2108–2117, 2017.

[78] T.H. Segall-Shapiro, A.J. Meyer, A.D. Ellington, E.D. Sontag, and C.A. Voigt. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Molecular Systems Biology*, 10(7): 742, 2014.

[79] M. Kushwaha and H.M. Salis. A portable expression resource for engineering cross-species genetic circuits and pathways. *Nature Communications*, 6: 7832, 2015.

[80] G. Lillacci, S. Aoki, D. Schweingruber, and M. Khammash. A synthetic integral feedback controller for robust tunable regulation in bacteria. *Preprint available at doi: 10.1101/170951*, 2017.

[81] O.S. Venturelli, M. Tei, S. Bauer, L.J.G. Chan, C.J. Petzold, and A.P. Arkin. Programming mRNA decay to modulate synthetic circuit resource allocation. *Nature Communications*, 8: 15128, 2017.

[82] F. Ceroni, S. Furini, T.E. Gorochowski, A. Boo, O. Borkowski, Y.N. Ladak, A.R. Awan, C. Gilbert, G-B. V. Stan, and T. Ellis. Burden-driven feedback control of gene expression. *Nature Methods*, 15: 387—393, 2018.

[83] A. Sachdeva, K. Wang, T. Elliott, and J.W. Chin. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *Journal of the American Chemical Society*, 136(22): 7785–7788, 2014.

[84] N. Duran, G.Z. Justo, M. Duran, M. Brocchi, L. Cordi, L. Tasic, G.R. Castro, and G. Nakazato. Advances in *Chromobacterium violaceum* and properties of violacein-Its main secondary metabolite: A review. *Biotechnology Advances*, 34(5): 1030–1045, 2016.

[85] J. Sambrook and D. Russel. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 2001.

[86] R. Silva-Rocha, E. Martínez-García, B. Calles, M. Chavarría, A. Arce-Rodríguez, A. de Las Heras, A.D. Páez-Espino, G. Durante-Rodríguez, J. Kim,

P.I. Nikel, R. Platero, and V. de Lorenzo. The Standard European Vector Architecture (SEVA): A coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research*, 41: D666–75, 2013.

[87] F.R. Blattner, G. Plunkett, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew, J. Gregor, N.W. Davis, H.A. Kirkpatrick, M.A. Goeden, D.J. Rose, B. Mau, and Y. Shao. The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331), 1997.

[88] J. Ang, S. Bagh, B.P. Ingalls, and D.R. McMillen. Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network. *Journal of Theoretical Biology*, 266(4): 723–738, 2010.

[89] C. Briat, A. Gupta, and M. Khammash. Antithetic integral feedback ensures robust perfect adaptation in noisy bimolecular networks. *Cell Systems*, 2(1): 15–26, 2016.

[90] M. Lynch and G.K. Marinov. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*, 112(51): 15690–15695, 2015.

[91] Y. Qian and D. Del Vecchio. Realizing 'integral control' in living cells: How to overcome leaky integration due to dilution? *Journal of The Royal Society Interface*, 15(139), 20170902, 2018.

[92] B.M. Lunde, C. Moore, and G. Varani. RNA-binding proteins: Modular design for efficient function. *Nature Reviews Molecular Cell Biology*, 8(6): 479–490, 2007.

[93] P. Babitzke, C.S. Baker, and T. Romeo. Regulation of translation initiation by RNA binding proteins. *Annual Review of Microbiology*, 63(1):27–44, 2009.

[94] C.M. Falcon and K.S. Matthews. Operator DNA sequence variation enhances high affinity binding by hinge helix mutants of lactose repressor protein. *Biochemistry*, 39(36): 11074–11083, 2000.

[95] P. Penumetcha, K. Lau, X. Zhu, K. Davis, T.T. Eckdahl, and A.M. Campbell. Improving the Lac system for synthetic biology. *Bios*, 81(1): 7–15, 2010.

[96] E.J. Hancock, G-B. Stan, J.A.J. Arpino, and A. Papachristodoulou. Simplified mechanistic models of gene regulation for analysis and design. *Journal of The Royal Society Interface*, 12: 20150312, 2015.

[97] H. Bremer, P. Dennis, and M. Ehrenberg. Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie*, 85(6): 597–609, 2003.

[98] A.D. Tadmor and T. Tlusty. A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLOS Computational Biology*, 4(5), 2008.

[99] D. Kennell and H. Riezman. Transcription and translation initiation frequencies of the *Escherichia coli* lac operon. *Journal of Molecular Biology*, 114(1): 1–21, 1977.

[100] R.C. Brewster, D.L. Jones, and R. Phillips. Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*. *PLOS Computational Biology*, 8(12), 2012.

[101] L. Bintu, N.E. Buchler, H.G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: Models. *Current Opinion in Genetics & Development*, 15(2): 116–24, 2005.

[102] M. Liu, G. Gupte, S. Roy, R.P. Bandwar, S.S. Patel, and S. Garges. Kinetics of transcription initiation at lacP1. Multiple roles of cyclic AMP receptor protein. *Journal of Biological Chemistry*, 278(41): 39755–39761, 2003.

[103] D.K. Hawley, A.D. Johnson, and R. Mcclureii. Functional and physical characterization of transcription initiation complexes. 260(14): 8618–8626, 1985.

[104] D. Na, S. Lee, and D. Lee. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Systems Biology*, 4(1): 71, 2010.

[105] J-P. Ebel, B. Ehresmann, C. Ehresmann, M. Mougel, R. Giegé, and D. Moras. Structure and interactions of ribosomal components and ligands. In B. Hardesty and G. Kramer, editors, *Structure, Function, and Genetics of Ribosomes*, chapter Structure, pages 270–285. Springer New York, NY, USA, 1986.

[106] T.A. Carrier and J.D. Keasling. Controlling messenger RNA stability in bacteria: Strategies for engineering gene expression. *Biotechnology Progress*, 13(6): 699–708, 1997.

[107] A. Kamionka, J. Bogdanska-Urbaniak, O. Scholz, and W. Hillen. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Research*, 32(2): 842–847, 2004.

[108] Y. Wang, L. Guo, I. Golding, E.C. Cox, and N.P. Ong. Quantitative transcription factor binding kinetics at the single-molecule level. *Biophysical Journal*, 96(2): 609–620, 2009.

[109] J.A. Hammerl, N. Roschanski, R. Lurz, R. Johne, E. Lanka, and S. Hertwig. The molecular switch of telomere phages: High binding specificity of the PY54 Cro lytic repressor to a single operator site. *Viruses*, 7(6): 2771–2793, 2015.

[110] H.H. Kimsey and M.K. Waldor. The CTXϕ repressor RstR binds DNA cooperatively to form tetrameric repressor-operator complexes. *Journal of Biological Chemistry*, 279(4): 2640–2647, 2004.

[111] H. Agustiandari, E. Peeters, J.G. de Wit, D. Charlier, and A.J.M. Driessen. LmrR-mediated gene regulation of multidrug resistance in *Lactococcus lactis*. *Microbiology*, 157(5): 1519–1530, 2011.

[112] K. Takeuchi, M. Imai, and I. Shimada. Dynamic equilibrium on DNA defines transcriptional regulation of a multidrug binding transcriptional repressor, LmrR. *Scientific Reports*, 7(1): 267, 2017.

[113] C.J. Glasscock, J.B. Lucks, and M.P. Delisa. Engineered protein machines: Emergent tools for synthetic biology. *Cell Chemical Biology*, 23(1): 45–56, 2016.

[114] I. Avcilar-Kucukgoze and Z. Ignatova. Rewiring host activities for synthetic circuit production: a translation view. *Biotechnology letters*, 39(1): 25–31, 2016.

[115] H. de Jong, J. Geiselmann, and D. Ropers. Resource reallocation in bacteria by reengineering the gene expression machinery. *Trends in Microbiology*, 25(6): 480–493, 2017.

[116] Y. Liu, D.S. Kim, and M.C. Jewett. Repurposing ribosomes for synthetic biology. *Current Opinion in Chemical Biology*, 40: 87–94, 2017.

[117] C.C. Liu, M.C. Jewett, J.W. Chin, and C.A. Voigt. Towards an orthogonal central dogma. *Nature Chemical Biology*, 14(2): 103–106, 2018.

[118] K.J. Åström and R.M. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton University Press, Princeton, NJ, USA, 2008.

[119] J. Chappell, K.E. Watters, M.K. Takahashi, and J.B. Lucks. A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future. *Current Opinion in Chemical Biology*, 28: 47–56, 2015.

[120] M.Kushwaha, W. Rostain, S. Prakash, J. N. Duncan, and A. Jaramillo. Using RNA as molecular code for programming cellular function. *ACS Synthetic Biology*, 5(8): 795809, 2016.

[121] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135(2): 216–226, 2008.

[122] L.B. Pickens, Y. Tang, and Y-H. Chooi. Metabolic engineering for the production of natural products. *Annual Review of Chemical and Biomolecular Engineering*, 2(1): 211–236, 2011.