

# **Superposition and Synthetic Genetic Devices: Framework and Model System to Investigate Linearity in *Escherichia coli***

*Meghdad Hajimorad*



Electrical Engineering and Computer Sciences  
University of California at Berkeley

Technical Report No. UCB/EECS-2011-37

<http://www.eecs.berkeley.edu/Pubs/TechRpts/2011/EECS-2011-37.html>

May 4, 2011

Copyright © 2011, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**Superposition and Synthetic Genetic Devices: Framework and  
Model System to Investigate Linearity in *Escherichia coli***

by

Meghdad Hajimorad

A dissertation submitted in partial satisfaction of the  
requirements for the degree of  
Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA, BERKELEY

Committee in charge:

Professor Jay D. Keasling, Co-Chair  
Professor Paul R. Gray, Co-Chair  
Professor Bernhard E. Boser  
Professor J. Christopher Anderson

Spring 2011

**Superposition and Synthetic Genetic Devices: Framework and Model  
System to Investigate Linearity in *Escherichia coli***

© Copyright 2011  
by  
Meghdad Hajimorad

## Abstract

### **Superposition and Synthetic Genetic Devices: Framework and Model System to Investigate Linearity in *Escherichia coli***

by

Meghdad Hajimorad

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professors Jay D. Keasling and Paul R. Gray, Co-chairs

The ability to compose biological systems from smaller elements that act independently of the other upon assembly may help make the forward engineering of biological systems practical. Engineering biology in this manner is made difficult by the inherent nonlinear response of organisms to genetic devices. Devices are inevitably coupled to one another in the cell because they share the same gene expression machinery for expression. Thus, new properties can emerge when devices that had been characterized in isolation are expressed concurrently. We show in this report that, similar to physical systems, the *Escherichia coli* (*E. coli*) transcriptional system can exhibit linear behavior under “small” perturbation conditions. This, in turn, allows devices to be treated as independent modules. A framework and model system consisting of three devices was developed to investigate linear system behavior in *E. coli*. The framework employed the transfer curve concept to determine the amount of nonlinearity elicited by the *E. coli* transcriptional system in response to the devices. To this effect, the model system was quantitatively characterized using real-time quantitative PCR to produce device transfer curves (DTCs). Two of the devices encoded the bacterial neomycin phosphotransferase II (*nptII*) and chloramphenicol acetyl transferase (*cat*), while the third encoded the jellyfish-originating green fluorescent protein (*gfp*). The *gfp* device was the most nonlinear in our system, with *nptII* and *cat* devices eliciting linear responses. Superposition experiments verified these findings, with independence among the three devices having been lost when *gfp* was present at copy numbers above the lowest one used. Elucidation of the mechanism underlying the nonlinearity observed in *gfp* may lead to design rules that ensure linear system behavior, enabling the accurate prediction of the quantitative behavior of a system assembled from individually characterized devices. This research suggests that biological systems follow principles similar to physical ones, and that concepts borrowed from the latter (such as DTCs) may be of use in the characterization and design of biological systems.

## Table of Contents

<b>ACKNOWLEDGEMENTS .....</b>	<b>III</b>
<b>CHAPTER 1: INTRODUCTION .....</b>	<b>1</b>
Background .....	1
Motivation.....	4
Thesis Organization .....	8
<b>CHAPTER 2: PRELIMINARY INVESTIGATION OF SUPERPOSITION.....</b>	<b>10</b>
Background .....	10
Results .....	11
Superposition improved as copy number was reduced.....	11
Discussion.....	15
Methods.....	17
<b>CHAPTER 3: INVESTIGATION OF LINEARITY VIA DEVICE TRANSFER CURVES.....</b>	<b>20</b>
Background .....	20
Results .....	20
Copy number of genetic device varied with plasmid origin of replication.....	20
Linear device transfer curves obtained in <i>E. coli</i> .....	23
Introduction of <i>gfp</i> genetic device led to nonlinear device transfer curves .....	27
Superposition lost at higher expression levels with the addition of <i>gfp</i> .....	30
Discussion.....	34
Methods.....	37
<b>CHAPTER 4: PLASMID COPY NUMBER CHARACTERIZATION OF A SET OF BGLBRICK-COMPATIBLE VECTORS .....</b>	<b>43</b>
Background .....	43

<b>Results and Discussion.....</b>	<b>43</b>
Copy number of plasmid in single plasmid strain .....	43
Copy number of plasmid in multi-plasmid strain .....	44
<b>Methods.....</b>	<b>47</b>
 <b>CHAPTER 5: CONCLUSION .....</b>	 <b>49</b>
<b>Thesis Summary.....</b>	<b>49</b>
<b>Future Directions .....</b>	<b>49</b>
 <b>BIBLIOGRAPHY .....</b>	 <b>51</b>

## Acknowledgements

As is the case with any endeavor, the completion of this project would not have been possible without the support and guidance of others.

I would like to thank my research advisors Professors Jay D. Keasling and Paul R. Gray for the opportunity to work on this project. Both created an environment conducive for growth as an engineer and scientist, allowing me with the ability to explore ideas and research directions within the broader project goals. Indeed, this research project would not have been possible if it was not for my advisors' insights for future directions in science and technology. I thank both of my advisors for always being supportive and enthusiastic mentors, and for finding the time to provide assistance and advice whenever it was needed.

I would also like to thank Professor Bernhard E. Boser for his mentorship during my years at Cal. His lectures and talks were indeed inspiring, framing technical discussions in such a way that the bigger picture was not lost and the important aspects of circuit design not forgotten. His courses keep alive the sense of discovery and enjoyment in circuit design, and his EE 240 course was one of my favorites taken at Cal. It has indeed been a privilege to serve as a teaching assistant for him on a couple of occasions. I also thank Professor Boser for serving as chair of my qualifying exam.

I also thank Professor J. Christopher Anderson for his support and guidance ever since learning of synthetic biology. I learned a great deal by serving as a teaching assistant for his 2007 iGEM team. My contributions to subsequent education efforts in synthetic biology are indeed a result of the skills and tools learned during that experience. I also thank Professor Anderson for serving on my qualifying exam and dissertation committees.

I am also indebted to Dr. Leonard Katz for all of the stimulating and engaging discussions in science and synthetic biology. Advancing this project forward would not have been possible without the generosity shown by Dr. Katz with his time to discuss scientific concepts and issues.

I would also like to take this opportunity to express my gratitude to past and present members of the Keasling lab for all of their time and assistance throughout the challenges encountered in the past few years.

Last but not least, I would like to thank my parents for their commitment to my education and their continual support and encouragement. Their wisdom and guidance has always been a source of strength.

This research was supported by a National Science Foundation Graduate Research Fellowship and the Synthetic Biology Engineering Research Center (SynBERC).



# Chapter 1: Introduction

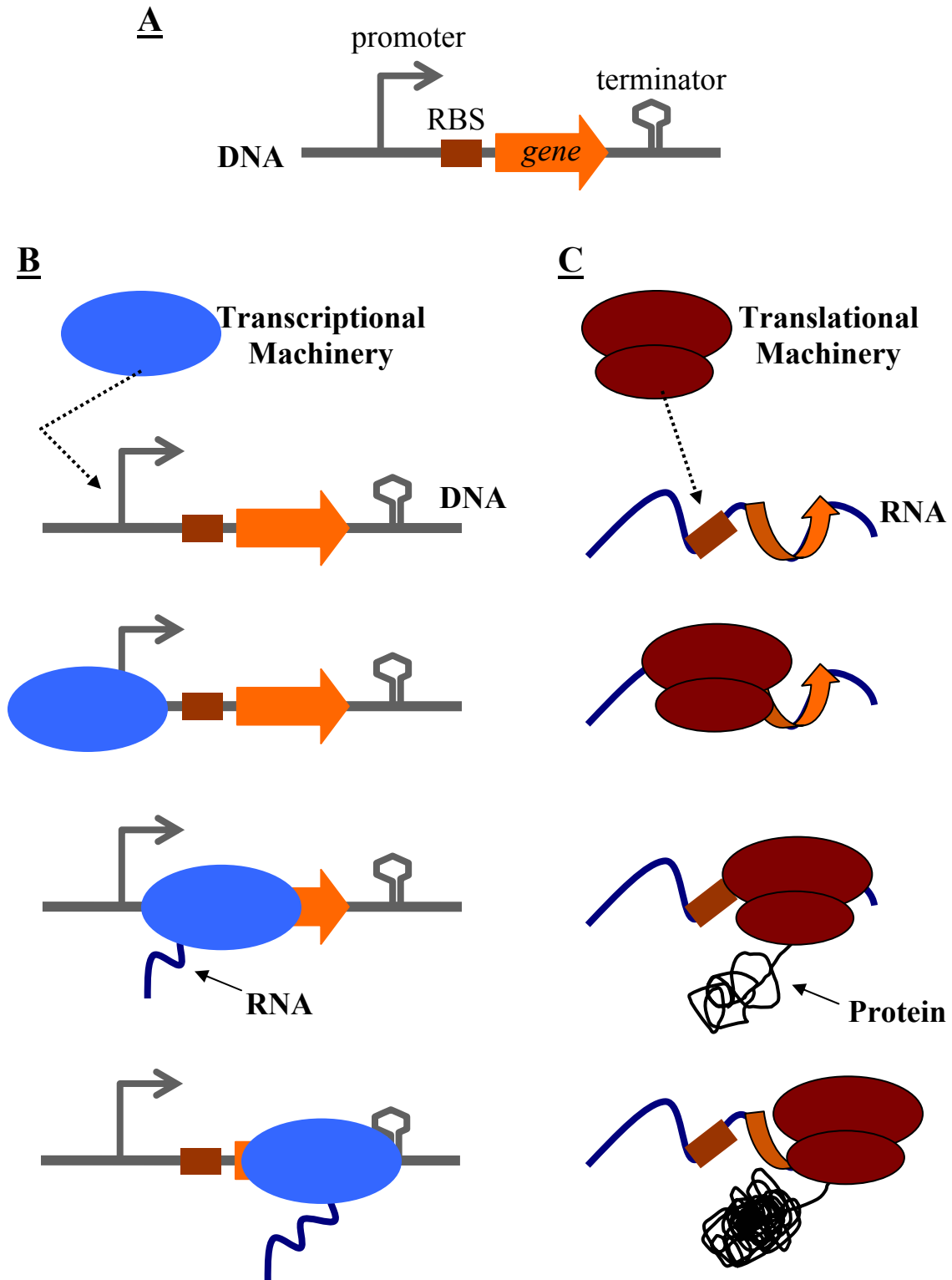
## Background

The genetic engineering of microbial organisms offers benefits to society through biotechnology applications. For instance, engineering microbes to produce industrial products (metabolic engineering) has become increasingly attractive due to the advantages over traditional synthetic methods. Modifications that lead to new biosynthetic capabilities in microbes allow for the production of previously limited products at high levels (such as therapeutic proteins and drugs), while at the same time reducing the use of environmentally destructive chemical processes [1-3]. A brief description regarding the genetic modification of microbes for engineering purposes is appropriate at this time.

RNA and proteins are the molecular players that perform tasks inside of biological cells. They are synthesized by the gene expression machinery of the cell, with the nucleotide sequence of DNA molecules serving as template. The fundamental building block in biological engineering is the single-gene genetic device (Figure 1-1A), which is a sequence of DNA consisting of a promoter, ribosome binding site (RBS), gene of interest, and transcription terminator. A gene is an ordered sequence of nucleotides that encode a product of interest (protein and/or RNA), and it is typically portrayed by a block arrow (Figure 1-1A). Once the DNA sequence of a device has been inserted into an appropriate microbial host, the gene expression machinery of the cell processes its DNA sequence and the encoded gene product is synthesized.

The first step in gene expression is called transcription, and it leads to the synthesis of RNA molecules from the DNA sequence encoded by the device (Figure 1-1B). For a device to undergo proper transcription, the sequences of a promoter and transcription terminator are placed before and after the gene, respectively. The promoter acts as a signal to the host transcriptional machinery to begin transcription prior to the gene nucleotide sequence. The transcription terminator, on the other hand, causes the process of transcription to terminate at a point downstream of the gene nucleotide sequence. In short, the transcriptional machinery of the host cell binds to the promoter, moves along the DNA sequence of the genetic device, and eventually falls off at the terminator. The end result is the synthesis of a RNA transcript that contains the gene sequence plus flanking, so called “non-coding” sequences (Figure 1-1B). An important non-coding sequence on the RNA is the RBS. It is located upstream of the gene, and it controls the process of translation. During translation (the second step in gene expression), the translational machinery of the host cell binds to the RBS, moves along the RNA sequence, and eventually dislodges at the end of the gene sequence. The end result is the synthesis of a protein molecule (Figure 1-1C). The core steps of gene expression have been described at a basic level here. There are a number of books that address this process in greater detail [4-6].

Genetic modifications that lead to the production of desired molecules are made difficult by the fact that many genes (i.e. enzymes) need to be introduced into the microbial host. The situation is further exacerbated by the necessity of having to control

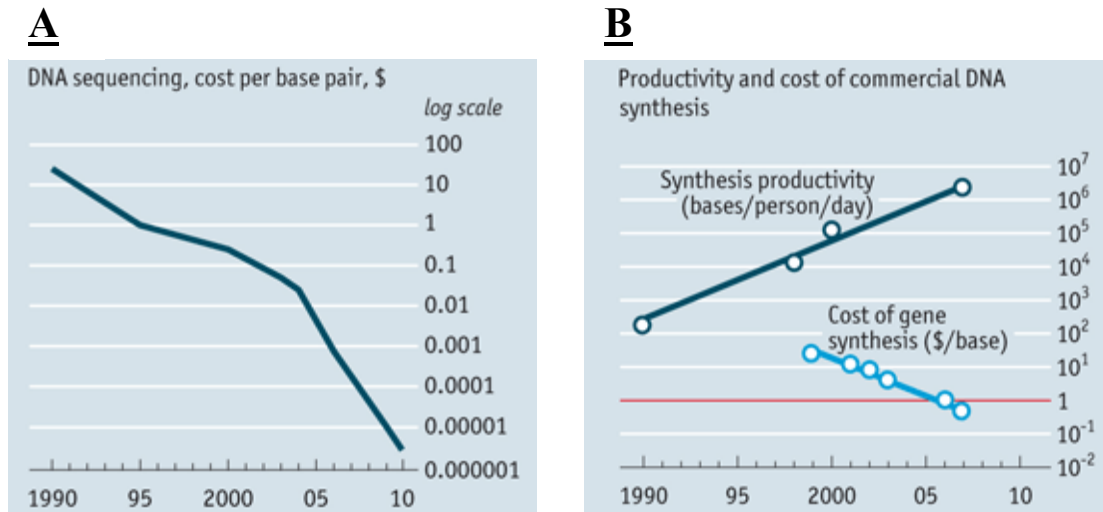


**Figure 1-1** **A.** Schematic representation of the basic components in a genetic device  
**B** and **C.** Schematic representations of the gene expression processes of transcription  
**(B)** and translation **(C)**

and coordinate the activity of each gene to achieve optimal production of the target molecule. This process is generally complex. One of the most cited recent success stories in metabolic engineering has been a system developed to produce a precursor of the antimalarial compound artemisinin in microbes [7]. The final system developed has been estimated to have taken roughly 150 person-years of work, which included the uncovering of genes involved in the pathway, the refinement of components (e.g. promoters and RBSs) to control the expression of the eventual genetic devices, and the testing of many variants to sufficiently increase production of an enzyme necessary to consume a toxic intermediate molecule [8]. Indeed, the “engineering” of microbes to arrive at organisms with predictable, quantitative behavior is not engineering in a strict sense. The process of arriving at functional systems, rather, require months (more often years) of trial-and-error type of experiments, with the undertaking being more akin to art than engineering [9]. The burgeoning area of synthetic biology aims to develop foundational principles and technologies to enable the systematic forward engineering of biological systems [10-11]. In essence, synthetic biology aims to put the engineering “back into” genetic engineering.

Recent technological advances in DNA synthesis and sequencing are important developments that have fueled appeal towards realizing the synthetic biology discipline. In order to engineer microbial organisms, it is necessary to construct assemblies of DNA components and genetic devices (Figure 1-1). Analogous to Moore’s Law in microelectronics, the prices of determining DNA nucleotide sequence (sequencing) and DNA construction (synthesis) have plummeted over the past two decades (Figure 1-2). This not only means that bioinformatics databases are accumulating the sequences of genes from many organisms in nature, but that it is also easier to assemble arbitrary combinations of said genes together in order to engineer a microbial host with new capabilities. One of the earliest contributions of synthetic biology was the application of basic abstraction and physical composition frameworks through BioBricks and the Registry of Standard Biological Parts [14-15], with variations to the original scheme having been subsequently proposed [16]. The Registry lists DNA sequences designed by synthetic biologists that perform specific functions inside of cells, and the BioBricks standard proposes rules that support the physical assembly of these different sequences together. Other early advances in synthetic biology involved the “silicon mimetic” approach, where biological circuits were realized to emulate the functionality of such microelectronic circuits as latches and oscillators [17]. The initial circuits of synthetic biology showed that writeable memory (the toggle switch) and oscillator (the repressilator) elements can be realized in the bacterium *Escherichia coli* (*E. coli*) by using the standard components of biological engineering (Figure 1-1).

The genetic toggle switch [18], having been inspired by the reset-set latch of digital electronics, implements a one-bit memory element using two repressible promoters arranged in a mutually inhibitory network (Figure 1-3). More specifically, it consists of two genetic devices, with the promoter of each being repressed by a regulatory protein encoded by the gene of the other. The transcriptional machinery of the cell (Figure 1-1B) can indeed be repressed (or induced) by certain regulatory proteins and molecules. The fluorescent reporter protein *gfp* (GFP) provided an external readout of the state of the toggle switch, and two inducer species allowed toggling of the switch from one state to the other. In addition to memory elements, clocks are also important



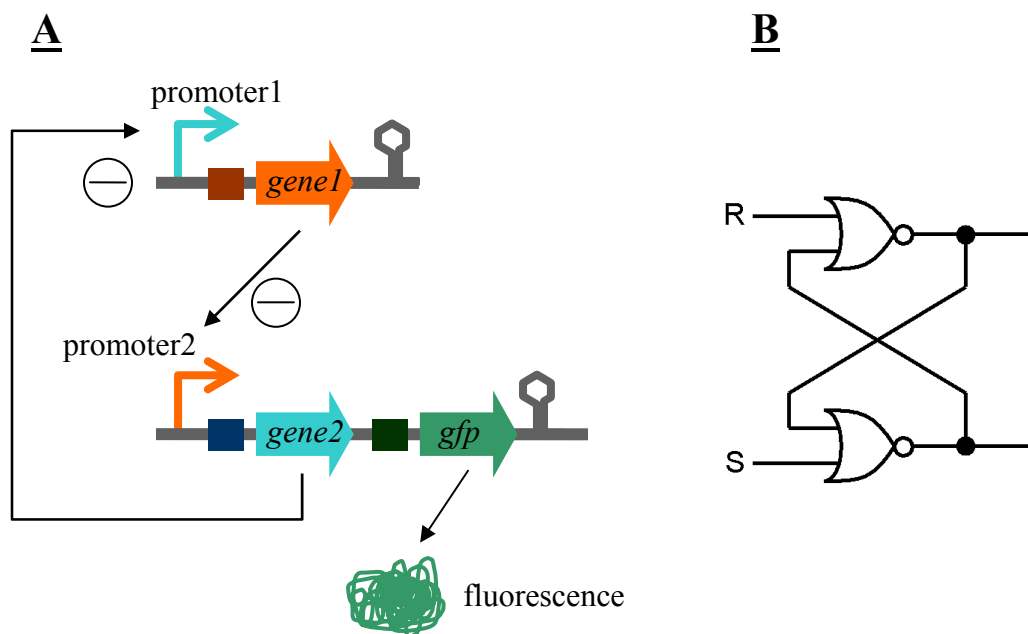
**Figure 1-2** Recent advances in DNA sequencing (A) [12] and synthesis (B) [13]

elements in electronic systems. With the ring oscillator of digital electronics having served as inspiration, the repressilator [19] implements a clock circuit in *E. coli* by using a three transcriptional repressor system (Figure 1-4). This circuit consists of three devices, with the protein of each repressing the promoter of the next device in the system. The addition of the third device leads to the negative feedback loop necessary for oscillation, and the synthesis of GFP (controlled by one of the promoters) provided a readable output of the system.

Such initial advances led to opinions that portrayed synthetic biology as simple design and construction [20]. That is, it inspired the idea that biologists can extend traditional genetic engineering to be more akin to the engineering of a computer system. The methodology would consist of characterizing the genetic sequences that perform useful functions (the “parts”), combine said parts to arrive at genetic devices with more complex functions, and then insert the devices into host cells. As such, synthetic biology could provide a toolbox of reusable genetic components (biological versions of electronic components) to be plugged into biological circuits. This analogy does not accurately capture the current knowledge gap that exists as to how biology works.

## Motivation

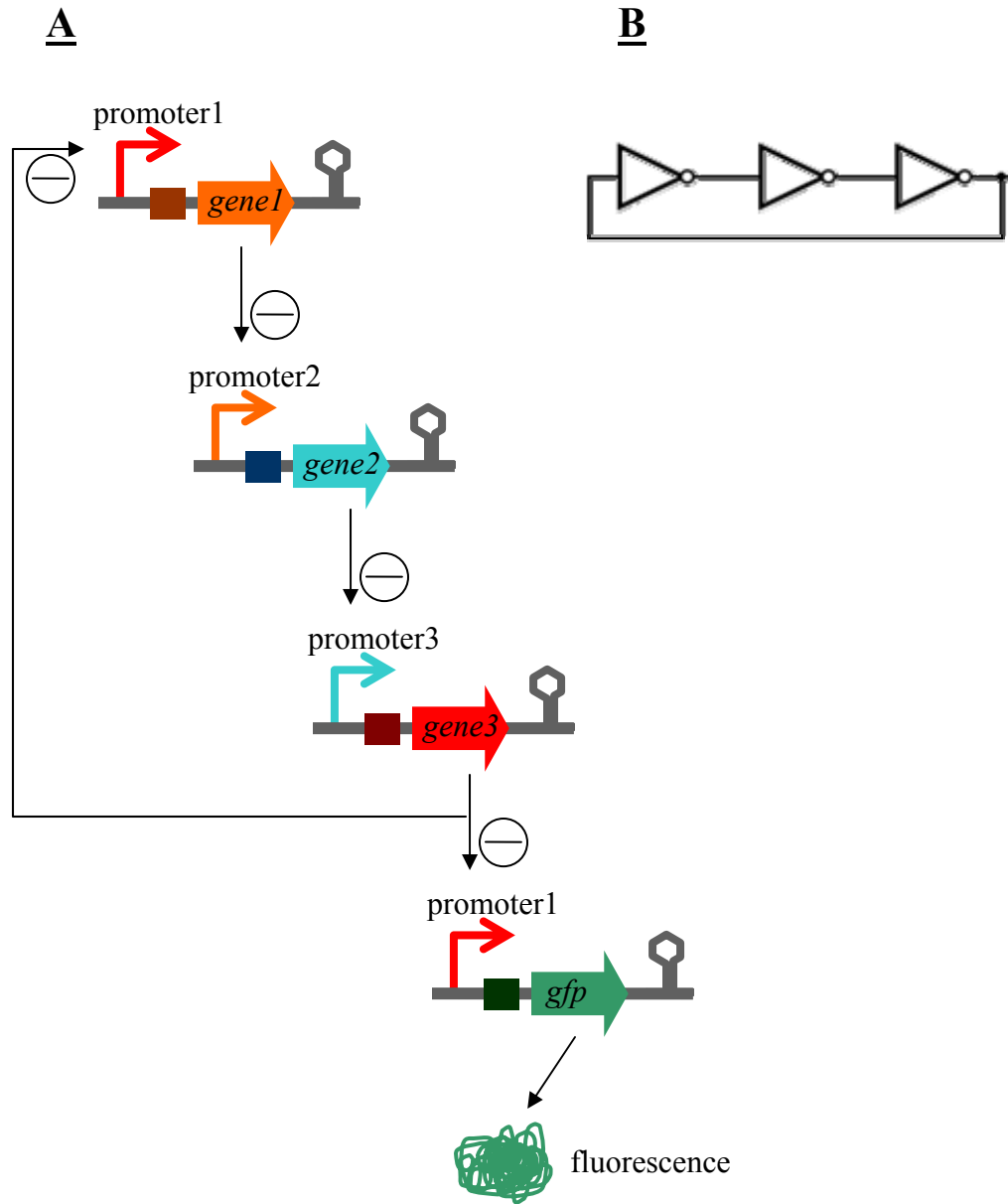
A characteristic feature of established engineering disciplines (such as electronics) is the ability to design and construct systems by way of modularity. The concept of modularity allows engineers to design and build physical systems by bringing together modules that contribute independently to the whole, thereby giving rise to a system whose quantitative behavior can be predicted from its constituent modules [21-23]. A pressing research question is whether the complexity of living organisms allows



**Figure 1-3** Schematic representation of the genetic toggle switch (A) and an electronic circuit analog (B)

engineers with a similar capability to design and construct biological systems from smaller elements characterized in isolation [24-25]. The success of synthetic biology as an engineering discipline will depend, in part, on establishing the conditions necessary for this independence property to hold true in living systems [26]. The research contribution of this study is the application of engineering principles towards realizing modularity and functional composition in biological systems. More specifically, it will be shown that genetic devices can behave in a standardized, quantitatively predictable manner. The ability to view devices as modules may be of benefit in such applications as metabolic pathway construction for the production of natural products and other chemicals (microbial chemical factories).

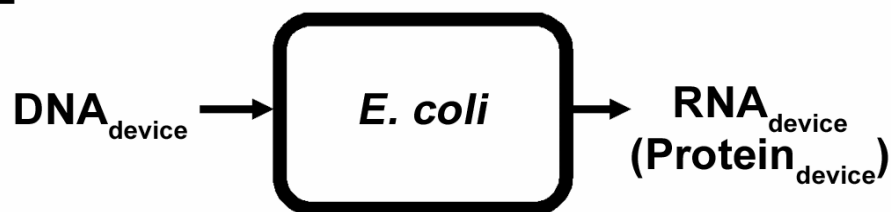
Once introduced into the host cell as DNA, a synthetic device must first be expressed by the native transcriptional and translational machinery in order to give rise to the desired function (e.g. production of transcript, desired protein, or metabolites, transduction of a signal, etc.). Synthetic devices introduced into *E. coli* for engineering purposes are, in essence, additional devices imposed on top of those present in the wild-type (baseline) case. Synthetic devices can, thus, only begin to behave independently if their respective RNA and protein levels are not affected by the addition of other synthetic devices. In order to motivate the experimental approach taken, the gene expression machinery of the *E. coli* host cell was viewed as a system in this study. Devices in the form of DNA are its input, with the resultant RNA (and protein) produced its output (Figure 1-5A). Synthetic devices introduced into *E. coli* for engineering purposes, however, are not the only inputs to the system. Thousands of devices are encoded on the



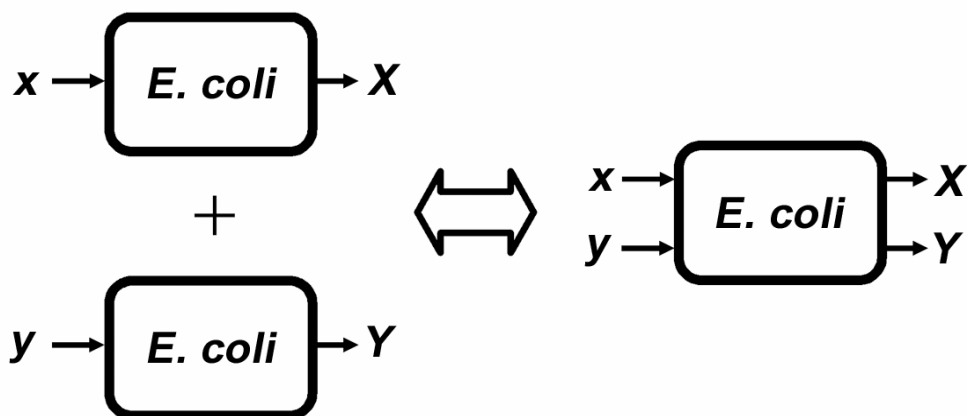
**Figure 1-4** Schematic representation of the repressilator (A) and an electronic circuit analog (B)

*E. coli* genome [27], whose regulated expression allow the organism to survive and grow in a given environment. The same molecular players and building blocks involved in synthesizing RNA (and protein) encoded by the host's native (chromosomally-encoded) devices are involved in the transcription (and translation) of synthetic (heterologous) devices. The system (Figure 1-5A), thus, not only has synthetic devices as inputs, but also the native devices. For independence among devices (synthetic and native) to be possible, the system must be linear, thereby exhibiting the superposition principle by definition [28]. The superposition property of linear systems states that the net response

**A.**



**B.**



**Figure 1-5** Black box representation of the *E. coli* gene expression system.

**A.** Genetic devices (synthetic and native) in the form of DNA act as inputs to the system, with the resultant RNA (and protein) produced the output. **B.** Pictorial depiction of the superposition principle in effect in linear systems. Superposition states that the output from a set of inputs represents the linear sum of the outputs from each individual input.

caused by two or more inputs is the sum of the responses which would have been caused by each stimulus individually. That is, if  $x$  and  $y$  amounts of DNA for a couple of devices alone produce  $X$  and  $Y$  amounts of transcript (protein), respectively, then the concurrent addition of both devices to the system should lead to the formation of  $X$  and  $Y$  amounts of transcript (protein) (Figure 1-5B). Systems in practice, however, are nonlinear and do not abide by the superposition principle. As such, the different synthetic and native inputs to the system cannot be studied in isolation. Design for predictable, quantitative behavior would, thus, not only require an understanding of how synthetic devices couple to one another by way of the nonlinearity present in the *E. coli* system, but also how they couple to the host's native devices. Our current level of understanding of these interactions is limited at best qualitatively, much less so in a quantitative manner. This, in part, may explain the difficulty associated with engineering biological systems with predictable, quantitative behavior. As they are embedded inside complex host cells, many

interactions are possible between the host cell and introduced constructs. And while the implementation of the toggle switch circuit (Figure 1-3) may appear simple on paper, it took years of tweaking the system to make it actually work [8].

This research began with the hypothesis that the introduced synthetic devices can be viewed as perturbations to the *E. coli* system. That is, the amount of DNA acting as input to the gene expression machinery of the host increases with their addition. So long as this increase (i.e. perturbation) is kept “small,” the *E. coli* system may perhaps be approximated as a linear one with respect to the introduced synthetic devices, thereby enabling superposition and the decoupling of synthetic devices from one another. This is the small-signal approximation used in the field of electronic circuit design [29]. There, it is used in the design of analog amplifiers, where voltage and current signals act as inputs to nonlinear, transistor-based systems. The significance and specific contributions of this research are as follows:

1. The current thinking in biology is that organisms are too complex and nonlinear to be predictably engineered. As such, it is thought that the superposition property of linear systems does not apply to the genetic engineering of microbial organisms under conditions relevant in the standard microbiological laboratory. This research demonstrates the contrary. That is, it shows that the nonlinearity in the *E. coli* system is not complex to the point of preventing functional composability (i.e. superposition) with standard elements used to genetically modify the organism.
2. An approach taken in this research to investigate system nonlinearity involved varying the copy number of synthetic genetic devices to generate transfer curves. This research demonstrates that a transfer curve-based framework has application in the engineering of biological systems, with device nonlinearity as gauged by transfer curve response correlating with superposition experimental data. Furthermore, device transfer curves may find application in the general characterization of devices.
3. This research demonstrates the significance of quantitative techniques (such as device transfer curves) to synthetic biology characterization efforts. It will be shown that growth rate alone (a traditional metric used in characterization efforts) is unable to accurately capture changes that take place due to synthetic device addition.
4. This research also presents characterization data for the absolute copy number of a plasmid resulting from a particular replicon, the impact that a change in bacterial cell type has on plasmid copy number, and the effect on a plasmid’s copy number of introducing plasmids with different replicon into the same strain.

## Thesis Organization

In order to demonstrate that biological systems follow principles similar to that of physical ones, Chapter 2 shows that the *E. coli* system response to genetic device additions becomes more linear as the perturbation level is decreased, under conditions relevant in typical molecular biology experiments. This chapter also describes the construction of a plasmid system used to perform the necessary experiments. Chapter 3 proposes and explores the use of device transfer curves to investigate linear system



behavior and superposition in *E. coli*. A plasmid system consisting of three devices was constructed, and it too is described in this chapter. Chapter 4 presents work done to characterize a set of BglBrick-compatible plasmids, with the purpose of having the results documented in a datasheet. Chapter 5 concludes this thesis by presenting a research support in addition to proposing directions for future research.

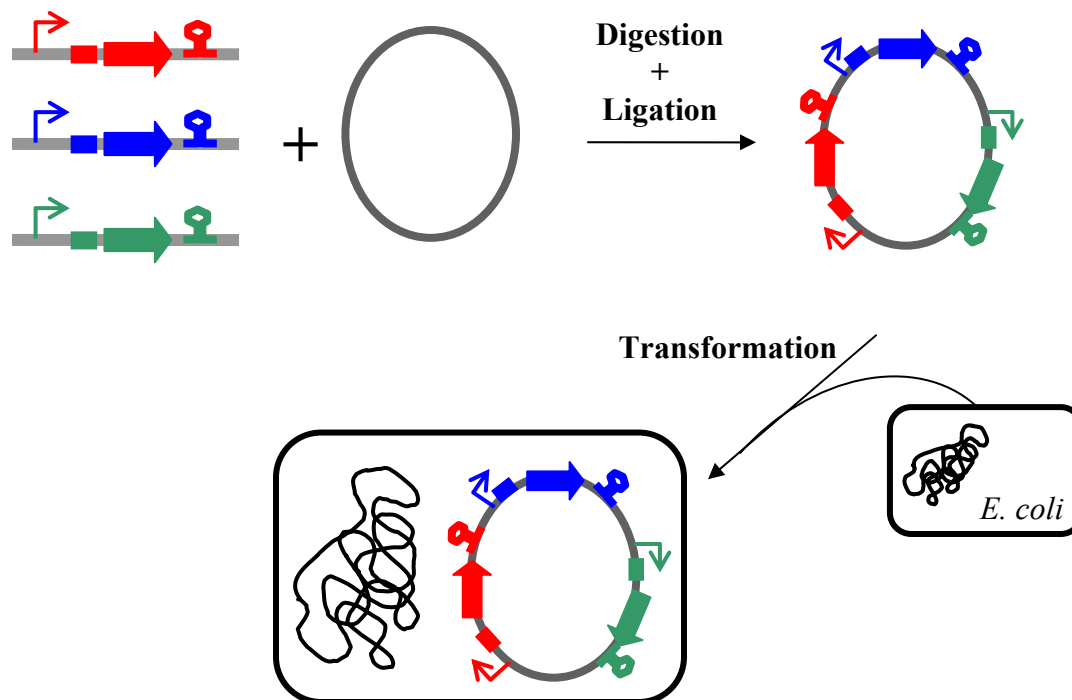
## Chapter 2: Preliminary Investigation of Superposition

### Background

Plasmids were used in this research to construct systems that allow for the introduction of synthetic genetic devices at different copy numbers (i.e. perturbation levels). It should be noted that the copy number of synthetic devices may not be the only factor that perturbs microbial organisms. Promoter strength, ribosome binding site (RBS) strength, gene length, codon usage, and product function are perhaps important factors too. As the intent of this research was to investigate whether *E. coli* can accommodate linear system behavior with standard elements used to genetically modify the organism, copy number was the focus here. Plasmids are circular pieces of DNA that are often used to introduce devices into microbes such as *E. coli* [30-31]. This process entails the use of cloning techniques to first insert the desired device(s) into the plasmid. The plasmid is then transferred into *E. coli* and the construct ends up in the cell's cytoplasm, separate from the host's genomic DNA, expressing its harbored devices. Figure 2-1 depicts this process. An attractive feature of a plasmid is that the degree to which it replicates within the host can be varied. The degree to which a plasmid replicates within the host is governed by a region on the plasmid called the origin of replication (or replicon). By having a different DNA sequence in this region, one can vary the number of copies of plasmid present in the host cell. The number of copies of the synthetic devices introduced into *E. coli* would subsequently be varied as they are harbored on the plasmid.

In addition to the replicon, the other basic component of a plasmid is its selection marker. The maintenance of plasmids is a burden on the cell, and one needs to overcome the cell's desire to lose plasmid. To this effect, *E. coli* is typically grown in growth medium that has been supplemented with an antibiotic. The selection marker is a genetic device harbored on the plasmid and that confers resistance to this added antibiotic, thereby providing an advantage to *E. coli* for maintaining the plasmid in its cytoplasm. That is, the bacterial cell would otherwise be unable to grow and survive in the antibiotic supplemented medium.

This research began by first investigating the applicability of the small-signal approximation in *E. coli* under conditions relevant in the standard microbiological laboratory. In this spirit, plasmids with standard features (i.e. typical replicons, selection marker, etc.) were used to express two devices with different reporter genes. The difference in culture-averaged fluorescence of a device introduced alone and concurrent with the other device was investigated to determine whether superposition improved with a lower copy replicon (i.e. reduced perturbation level). The goal here was to show that microbial organisms follow principles similar to that of physical systems, and that the nonlinearity in *E. coli* is not complex to the point of preventing linearization studies under conditions relevant in the standard microbiological laboratory.



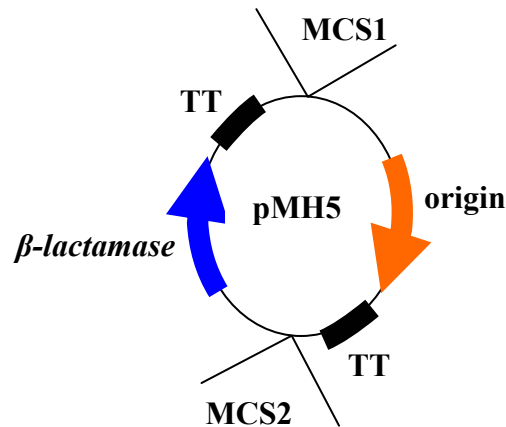
**Figure 2-1** Schematic representation of the cloning process flow

## Results

### Superposition improved as copy number was reduced

A plasmid system was constructed to enable the introduction of genetic devices into *E. coli* at different copy numbers. To this effect, a modular plasmid backbone was constructed as described in the Methods (Figure 2-2). The bacterial *rrnB* terminator was used for the two multi-cloning sites. This strong transcriptional terminator has been widely used, and shown to exhibit bi-directional termination properties [32]. As such, its inclusion in the plasmid minimized possible transcriptional readthrough from the origin of replication, which relies on transcription for functionality [33-34], into the neighboring cloning sites. The same applies with regards to possible transcriptional readthrough originating from the selection marker device  *$\beta$ -lactamase*. This device confers resistance to the antibiotic ampicillin.

In order to investigate the applicability of the small-signal approximation in the model bacterial system *E. coli*, two synthetic devices were cloned into the plasmid

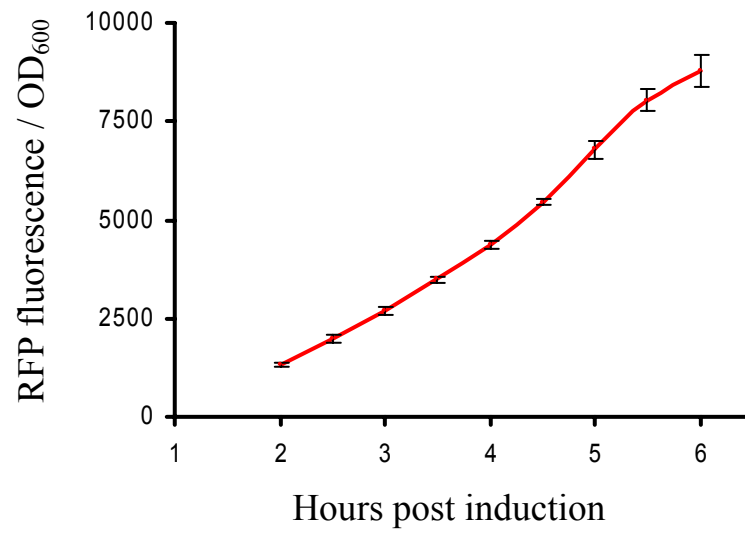
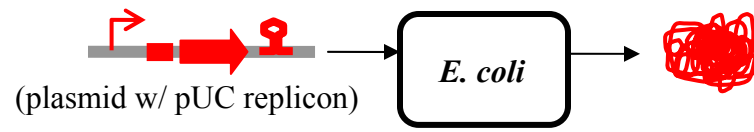


**Figure 2-2** Plasmid backbone with the  $\beta$ -lactamase selection marker. The origins of the plasmids p15A and pUC were cloned into the backbone using the unique restriction site *EcoRI* (not shown). TT denotes *rrnB* transcriptional terminator.

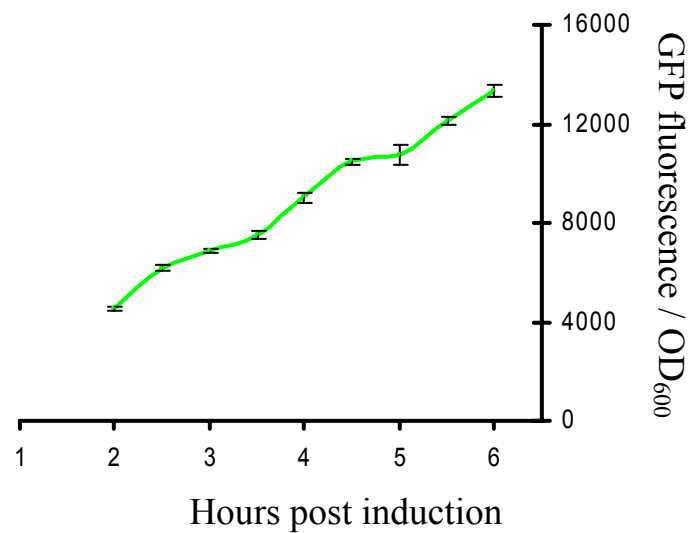
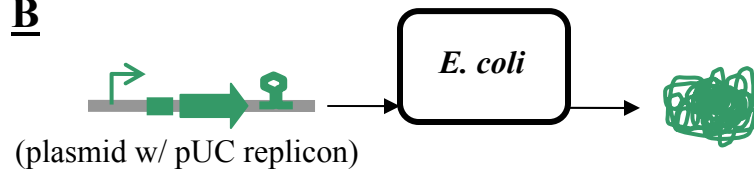
backbone (Figure 2-2). One device had the reporter gene red fluorescent protein (*rfp*) [35], while green fluorescent protein (*gfp*) [36] was the reporter gene of the other device. Both *rfp* and *gfp* are widely used reporter genes in molecular biology and synthetic biology applications. Each device was composed of the IPTG-inducible promoter  $P_{trc}$  and cognate gene encoding the lac repressor (*lacI<sup>Q</sup>*) [37], the strong ribosome binding site AGGAGG [36], and the *rrnB* transcription terminator. The devices were placed on one of two plasmids (Figure 2-2) containing either the high or medium copy number pUC [38] or p15A [39] replicon, respectively. This was to modulate the copy number of the devices and, thus, the degree of perturbation to the *E. coli* system. All other sequences on the plasmids (selection marker, multicloning site, and other sequences) were identical.

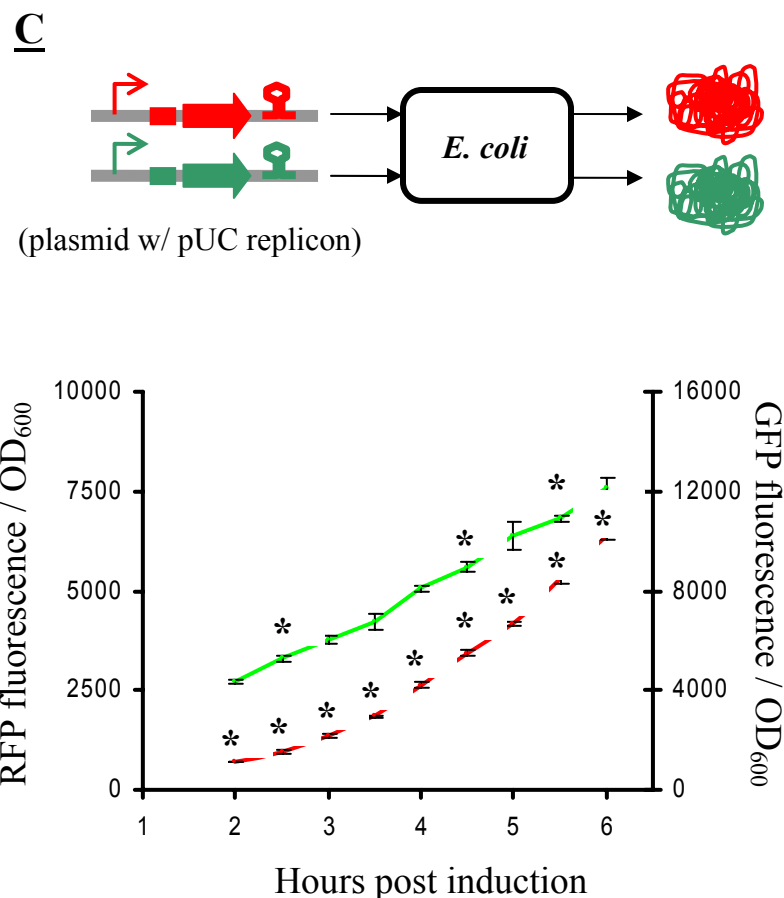
Time course assays were performed to test for superposition in the culture-averaged fluorescence arising from the encoding devices. *E. coli* DH10B cells harboring the pUC- and p15A-borne constructs were grown in LB medium as described in the Methods. Once in mid-exponential growth, cells were induced by adding IPTG to the medium. Samples were taken over the course of time starting two hours post induction in order to measure  $OD_{600nm}$  and fluorescence, with final fluorescence measurements having been normalized for cell density (fluorescence per  $OD_{600nm}$  unit) to arrive at culture-averaged fluorescence. The difference between the culture-averaged fluorescence resulting from the pUC- and p15A-borne constructs was determined statistically ( $P < 0.01$ , t-test) for each time point. Superposition was not observed at the perturbation levels exhibited by the pUC-borne devices. That is, the green and red fluorescence levels resulting from both devices present on the same pUC-based plasmid did not correspond to those levels when either of the devices was present alone on the plasmid (Figure 2-3). The output of the *gfp* device when harbored alone on the high copy number pUC-based plasmid differed significantly at three of nine time points from that when the device was borne on the same high copy number plasmid with the *rfp* device present. The red

**A**



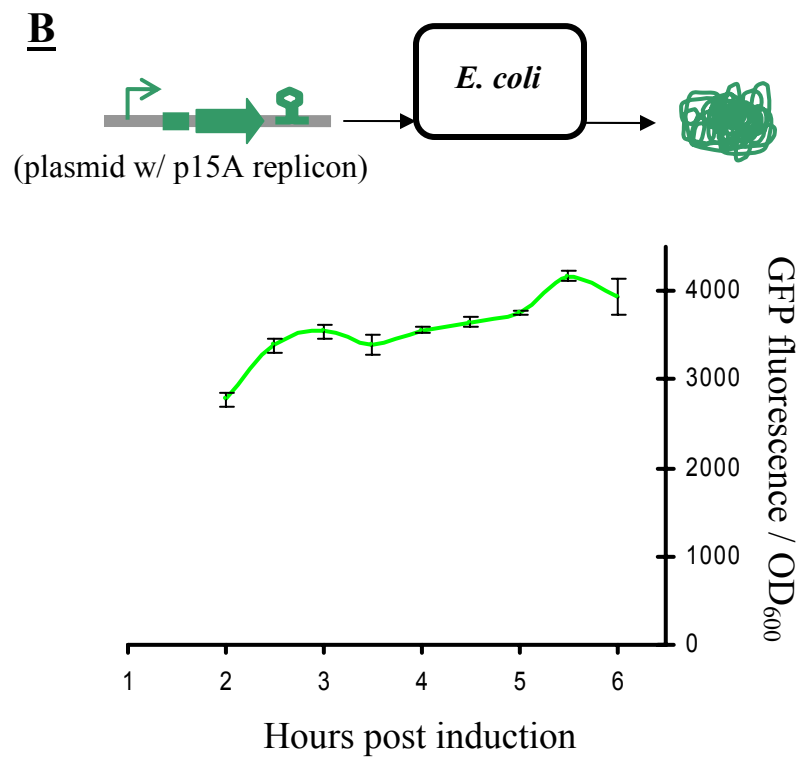
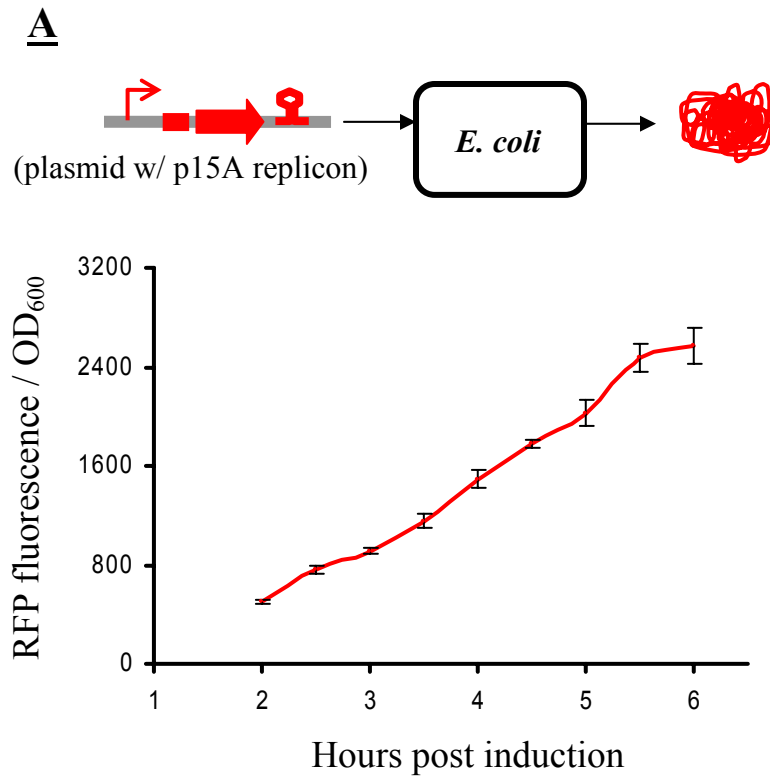
**B**

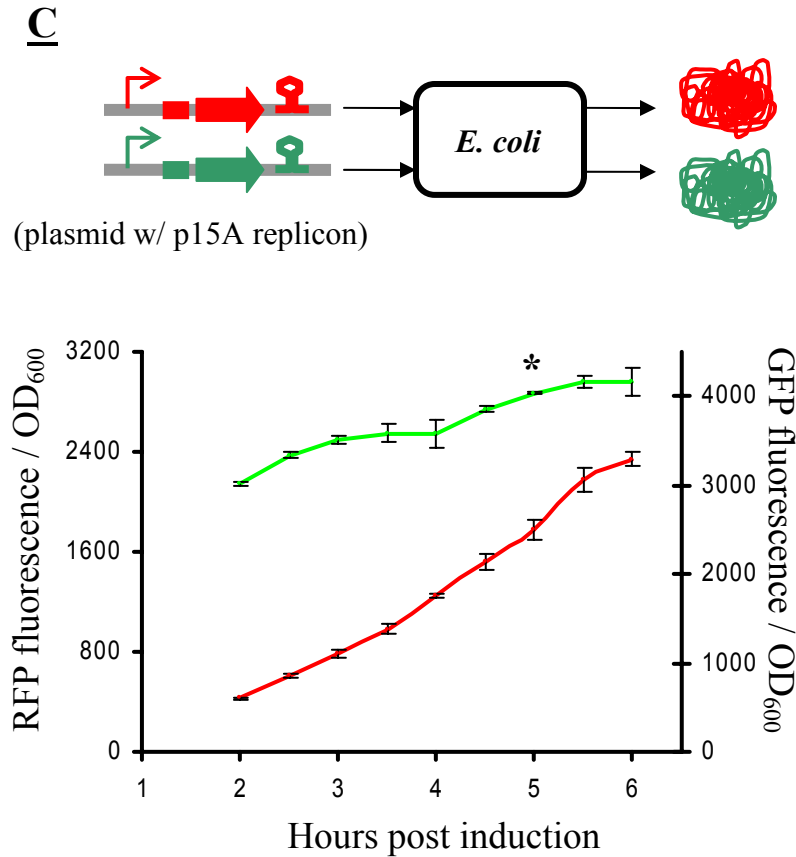




**Figure 2-3** Comparison of superposition in the culture-averaged fluorescence from RFP and GFP encoding devices using plasmids with pUC origins of replication. Data presented are the mean from three replicates +/- standard error. The difference between a device introduced alone (**A**, **B**) and concurrent with the other (**C**) was determined statistically ( $P < 0.01$ , t-test). Asterisks indicate significant difference.

fluorescence resulting from the dual device harboring plasmid differed from that produced by the single *rfp* device harboring plasmid at all nine time points. In contrast, the green and red fluorescence output from the *gfp* and *rfp* devices borne together on the medium copy number p15A-based plasmids differed at only one time point as compared to when the devices were borne singly on the same plasmid backbone (Figure 2-4).





**Figure 2-4** Comparison of superposition in the culture-averaged fluorescence from RFP and GFP encoding devices using plasmids with p15A origins of replication. Data presented are the mean from three replicates  $\pm$  standard error. The difference between a device introduced alone (A, B) and concurrent with the other (C) was determined statistically ( $P < 0.01$ , t-test). Asterisks indicate significant difference.

## Discussion

The experimental results here indicate that the response of *E. coli* becomes more linear as the perturbation incurred by synthetic devices is lowered. This suggests that biological systems follow principles similar to physical ones as it relates to the small-signal approximation. The results also suggest that the nonlinearity in the *E. coli* biological system is not complex to the point of preventing linearization studies under conditions relevant in the standard microbiological laboratory. In the plasmid system presented here, the *gfp* and *rfp* devices each contained the strong  $P_{trc}$  promoter and AGGAGG ribosome binding site. Superposition (linearity) was practically observed for the devices when a plasmid backbone harboring the p15A replicon was used (Figure 2-4). That is, the culture-averaged fluorescence was statistically different at only one time



point. Superposition (linearity) was lost with a pUC origin (Figure 2-3), indicating that the model system presented here enters into the nonlinear regime of operation at this copy number. The results also appear to suggest that the *gfp* device was the more nonlinear device in the model system. At the higher copy number of pUC, *rfp* device addition did not impact *gfp* fluorescence levels to the same extent that *gfp* addition impacted *rfp* fluorescence (three versus nine time points where the difference in culture-averaged fluorescence was statistically significant). This was in spite of the fact that both *gfp* and *rfp* devices were identical except for the sequence of the encoded gene. The larger observed nonlinearity in *gfp* is further investigated in the next chapter.

In a way similar to circuit designers, biological engineers need to be aware of the perturbation levels that synthetic devices incur to observe independence among said devices. Using synthetic devices under small-signal conditions may enable synthetic biologists to engineer microbial organisms with predictable, quantitative behavior. That is, it may enable the realization of functional composition through the assembly of synthetic devices whose quantifiable behavior has been characterized in isolation. The term “small” has yet to be determined. In this spirit, synthetic biologists may be able to use frameworks similar to those developed by electronics engineers to quantify the nonlinearity of synthetic devices. This, in turn, may enable synthetic biologists to determine whether useful biological systems can be engineered by way of superposition (linearity) or if it is necessary to enter into the nonlinear regime of operation where the independence of devices is lost. A possible framework based on the engineering concept of the transfer curve is presented in the next chapter.

## Methods

### Bacterial strains, media, and enzymes

*E. coli* DH10B was used for cloning and time course assays. Luria-Bertani (LB) medium was made as described previously [31]. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs, with digestion and ligation reactions performed as recommended by the enzyme manufacturer. PCR reactions were performed with Phusion polymerase from Finnzymes, and the primers used were synthesized by Integrated DNA Technologies, Inc. The composition of the PCR reactions, cycle times, and temperatures followed those suggested by the enzyme manufacturer. PCR products were sequenced once cloned into the respective plasmids to ensure that no mutations had been introduced during the amplification process. In cases where single digest cloning was performed, sequencing was also used to select for constructs with inserts in the desired orientation.

### Plasmid construction

The modular plasmid backbone (Figure 2-2) was constructed in several cloning steps. DNA encoding the two *rrnB* transcription terminators was amplified from pBAD24-*gfp* [37] using PCR. The primer pairs TT-BamHI-F, TT-BamHI-R and TT-ApaI-F, TT-ApaI-R (Table 2-1) were used for this purpose. PCR products were digested with BamHI (former primer pair) and ApaI (latter primer pair), and ligated into the multi-cloning site of pBlueScript II SK (Stratagene) to yield pMH1 and pMH2, respectively. Subsequently,  $\beta$ -lactamase DNA and the pUC replicon were PCR amplified from pUC19 (New England Biolabs) using the primer pairs Amp-KpnI-F, Amp-KpnI-R and pUCori-EcoRI-F, pUCori-

EcoRI-R (Table 2-1), digested with KpnI and EcoRI, and ligated into the multi-cloning site of pBlueScriptII SK to yield pMH3 and pMH4, respectively. Finally, the modular plasmid backbone (Figure 2-2) was obtained from pMH4 by PCR amplification. Primers circ-AgeI-F, circ-AgeI-R (Table 2-1) were used for this purpose. The obtained PCR product was digested with AgeI, and subsequently ligated to itself to form the circular plasmid backbone pMH5.

The plasmid system used for the fluorescence assays (Figures 2-3 and 2-4) was also constructed in several cloning steps. DNA encoding red fluorescent protein (RFP) was amplified from pZBRG [37] using PCR and the primers RFP-XhoI-F / RFP-XhoI-R, digested with XhoI, and ligated into MCS1 of pMH5 to yield pMH6. Subsequently, DNA encoding the IPTG-inducible promoter  $P_{trc}$  and cognate gene encoding the lac repressor (*lacI<sup>Q</sup>*) were amplified from P<sub>trc</sub>99-*gfp* [37] using PCR and the primers P<sub>trc</sub>-ClaI-F / P<sub>trc</sub>-ClaI-R, digested with ClaI, and ligated into pMH6 to yield pMH7hc. The cloning of the green fluorescent protein into MCS2 proceeded similarly using pBAD24-*gfp* as template and the primers GFP-XbaI-F / GFP-XbaI-R to yield pMH8. DNA encoding the  $P_{trc}$  promoter and *lacI<sup>Q</sup>* were amplified from P<sub>trc</sub>99-*gfp* using PCR and the primers P<sub>trc</sub>-SacII-F / P<sub>trc</sub>-SacII-R, digested with SacII, and ligated into pMH8 to yield pMH9hc. To construct the plasmid with both reporter proteins, RFP<sub>EC</sub> and  $P_{trc}$  promoter (with *lacI<sup>Q</sup>*) were subcloned step-wise from pMH7hc into pMH9hc to eventually yield pMH11hc.

pMH7hc, pMH9hc, and pMH11hc were the pUC-based vectors used in the fluorescence assays. In order to arrive at their p15A-based counterparts, DNA encoding the p15A origin of replication was amplified from pACYC177 using PCR and the primers p15ori-EcoRI-F / p15ori-EcoRI-R, digested with EcoRI, and ligated into pMH7hc, pMH9hc, and pMH11hc to yield pMH7lc, pMH9lc, and pMH11lc, respectively.

### Fluorescence Assays

Cells were grown overnight at 30°C, 200rpm shaking after inoculating 5mL cultures of LB medium (supplemented with 100µg/mL ampicillin) with single colonies from freshly streaked plates. After subculturing (1:25) into shake flasks containing 50mL of LB medium (supplemented with 100µg/mL ampicillin), cells were grown at 30°C, 200rpm shaking until an OD<sub>600nm</sub> of approximately 0.6 was reached. At this time, cells were induced by adding IPTG to a final concentration of 100µM to the medium. Samples were taken every 30 minutes (starting two hours after induction) to make OD<sub>600nm</sub> and fluorescence measurements using a Tecan SpectraFluor Plus plate reader (Tecan-US, Durham, NC). Excitation/emission wavelengths of 558nm/583nm and 400nm/510nm were used for RFP and GFP, respectively. The fluorescence measurements were normalized for cell density (fluorescence per OD<sub>600nm</sub> unit).

**Table 2-1.** List of PCR primers used in the cloning of the plasmids constructed in this chapter

<u>Primer</u>	<u>Sequence</u>
TT-BamHI-F	TATTATGGATCCTTGGCTGTTTTGG
TT-BamHI-R	TATTATGGATCCCCTAGGAGAGTTTGTAGAAAC
TT-ApaI-F	TATTATGGGCCCCTAGCTTGGCTGTTTTGG
TT-ApaI-R	TATTATGGGCCCAGAGTTTGTAGAAACGC
Amp-KpnI-F	TATTATGGTACCGACAGTTACCAATGC
Amp-KpnI-R	TATTATGGTACCGATACGCCTATTTTTATAGG
pUCori-EcoRI-F	TATTATGAATTCTCACTGACTCGCTGC
pUCori-EcoRI-R	TATTATGAATTCAGACCAAGTTTACTCATATATAC
circ-AgeI-F	TATTAGACCGGTTGGGTACCGATAC
circ-AgeI-R	TATTAGACCGGTAACAAAAGCTGGAG
RFP-XhoI-F	TAATAACTCGAGCAGGAGGAATCTAATGGC
RFP-XhoI-R	TAATAACTCGAGAGCATGCCTACAGGAAC
Ptre-ClaI-F	TAATAAATCGATAGGTTTTACCGTCATCAC
Ptre-ClaI-R	TAATAAATCGATTGTGTGAAATTGTTATCCG
GFP-XbaI-F	TAATAATCTAGACAGGAGGAATCTAATGAGTAAAGG
GFP-XbaI-R	TAATAATCTAGAATCATCTTATTTGTAGAGC
Ptre-SacII-F	TAATAACCGCGGAGGTTTTACCGTCATC
Ptre-SacII-R	TAATAACCGCGGTGTGTGAAATTGTTATCCG
p15ori-EcoRI-F	ATAATAGAATTCTTAATAAGATGATCTTCTTG
p15ori-EcoRI-R	ATAATTGAATTCAGTGTATACTGGCTTACTATG

## Chapter 3: Investigation of Linearity via Device Transfer Curves

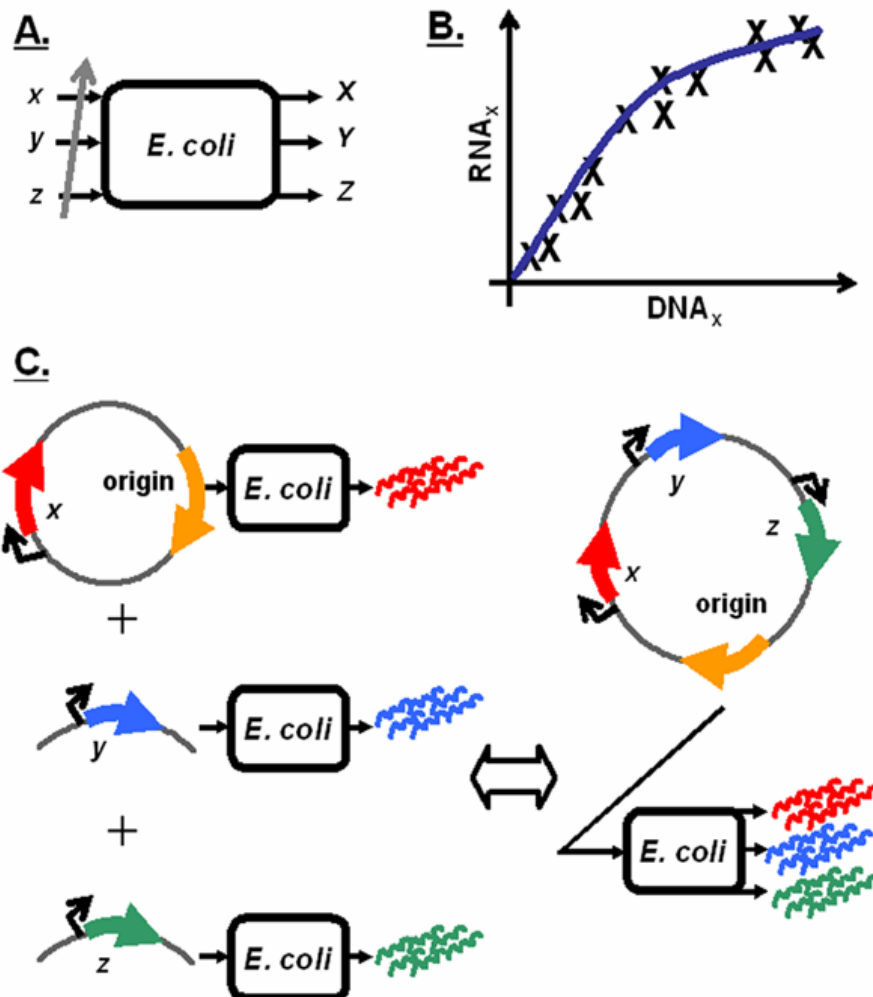
### Background

With the previous chapter suggesting that linearity studies can be done at relevant biological conditions, a more systematic approach was undertaken here. That is, the use of transfer curves was explored in order to investigate linear system behavior (superposition) in the context of synthetic biological systems. A model system was constructed to enable the introduction of synthetic genetic devices into *E. coli* at different copy numbers (Figure 3-1A). The resultant RNA produced was quantified at each copy number to produce RNA versus DNA device transfer curves (DTCs) (Figure 3-1B). DTCs were subsequently analyzed to gauge the nonlinearity of the *E. coli* system response to individual devices, with linear system behavior being verified by showing the presence of superposition (Figure 3-1C). RNA level was the focus here because synthetic devices can only begin to behave independently if their respective transcript levels are not affected by the addition of other synthetic devices. This is because of transcription being the initial process in gene expression.

### Results

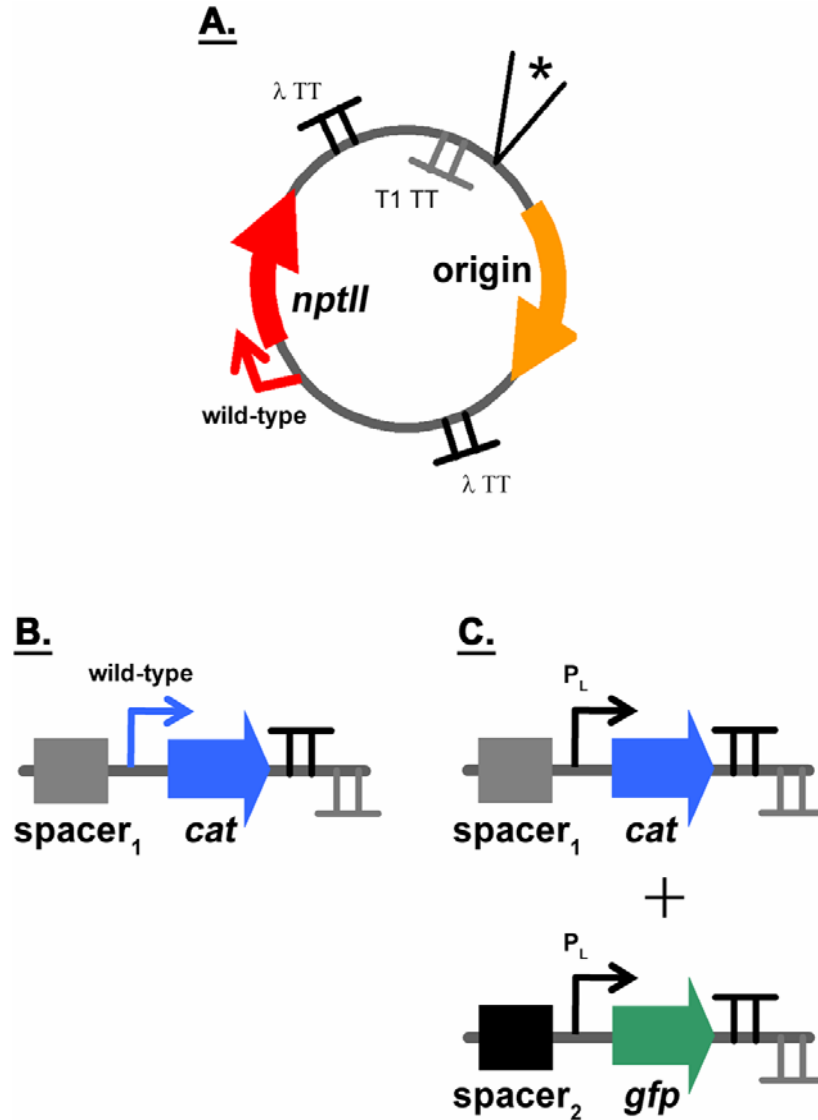
#### Copy number of genetic device varied with plasmid origin of replication

The degree to which a plasmid replicates in *E. coli* is governed by its origin of replication. Using different origins, one can vary the number of copies of plasmid present in the host cell. The number of copies of the synthetic devices introduced into *E. coli* would subsequently be varied as they are harbored on the plasmid. In developing the model system for this work, the plasmid origin of replication was flanked by terminators (Figure 3-2A) to minimize possible transcriptional readthrough from replicons, which rely on transcription for functionality [33-34], into neighboring devices. The two terminators used in the plasmid backbone (and all of the other constructs) were the bacterial *rrnB* T1 and bacteriophage lambda t0. These strong transcriptional terminators have been widely used [32, 40]. It was first verified that the number of copies of a device can be varied in the system by using different replicons. To this effect, the origins of plasmids pMPP6 [41], pSC101 [42], p15A [38], pMB1 [39], and pUC [39] were cloned into the backbone to determine whether the copy number of the neomycin phosphotransferase II (*nptII*) device varied (Figure 3-2A). This device confers resistance to the antibiotic kanamycin. The replicons from pSC101, p15A, and pMB1 are in different incompatibility groups [33-34]. As the origin of pMPP6 (and pUC) is that of pSC101 (and pMB1) with point-mutations [39, 41], these two replicons are not compatible with one another. *E. coli* DH1 cells harboring plasmid backbone constructs (Figure 3-2A) were grown in M9 minimal medium as described in the Methods. The growth rate of cells was comparable among the constructs, with OD<sub>600nm</sub> in the log phase



**Figure 3-1** Experimental approach taken in this chapter. **A.** Copy numbers of the synthetic genetic devices in the model system were varied and the RNA from a particular device was measured as an output. **B.** DTC for a device (device x as an example) was generated by first plotting the RNA produced at each copy number. A regression that minimized the sum of squared residual error was subsequently fitted to the cumulative data. **C.** Linear system behavior was gauged by testing for superposition. Superposition was determined to exist if a correspondence was observed between the left and right hand sides of the depicted figure.

doubling every ~80 minutes (Table 3-1). With the growth rate of cells not varying with replicon (Table 3-1), it appears unlikely that kanamycin (which was used for selection) elicits an effect on *nptII* expression. That is, it appears that the level of *nptII* expressed at the different copy numbers surpasses the minimum threshold necessary for cell survival. Once in mid-exponential growth, cells were harvested and total DNA extracted. Real-time quantitative PCR (qPCR) was subsequently used to determine the copy number of *nptII* for each construct [42-45]. In general, the results were similar to values reported



**Figure 3-2** Diagrams of the plasmids constructed to enable variation of genetic device copy number and to determine DTCs. **A.** Plasmid backbone with the *nptII* selection marker. The origins of the plasmids pMPP6, pSC101, p15A, pMB1, and pUC were cloned into the backbone using the two unique restriction sites *SmaI* and *AvrII* (not shown). TT denotes transcriptional terminator. Two different terminators ( $\lambda$  t0 and *rrnB* T1) had to be used because cloning attempts aimed at having the same terminator present simultaneously in opposing directions met with failure. The cassettes depicted in panels B and C were cloned in the multi-cloning site indicated by an asterisk. **B.** The *cat* device containing cassette cloned into the backbone. **C.** The *cat* and *gfp* device containing cassettes cloned into the backbone to arrive at the model system. Spacers were used to create spatial separation between neighboring devices (see Methods) to minimize the potential of special coupling. The promoter driving *cat* in B is different from that in C. While the native promoter was used in the former,  $P_L$  was used in the latter.

**Table 3-1 Growth rate of cells harboring empty (Figure 3-2A) or *cat*-device containing (Figure 3-2B) backbone**

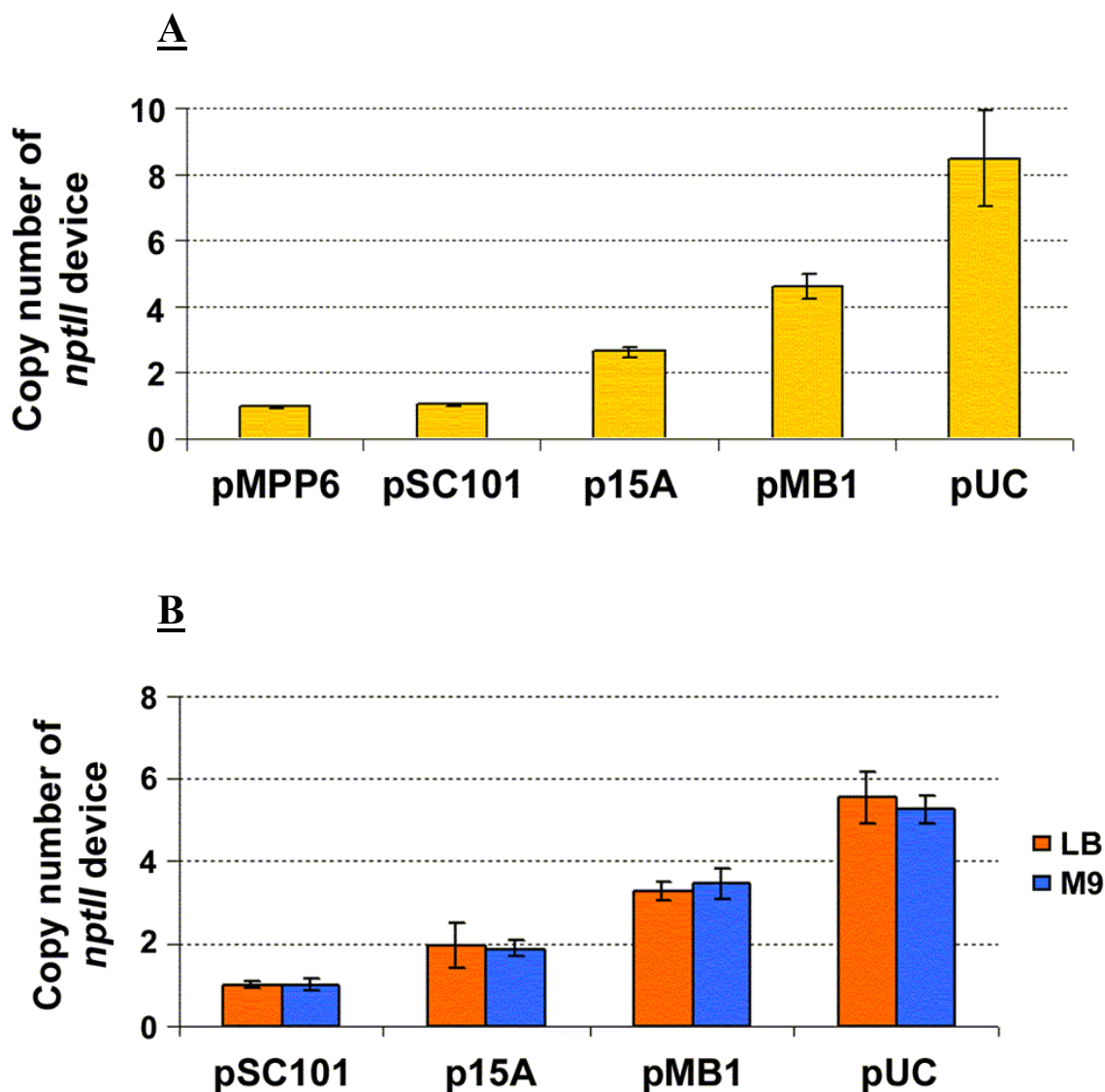
*E. coli* cells were grown in the medium specified in parentheses, with numbers indicating the necessary time in minutes for the optical density at a wavelength of 600nm to double in the log phase. The mean value  $\pm$  95% confidence interval as determined in duplicate has been reported (n.d. = not determined)

	pMPP6	pSC101	p15A	pMB1	pUC
<b>backbone (M9)</b>	80 $\pm$ 3	80 $\pm$ 3	83 $\pm$ 3	83 $\pm$ 5	84 $\pm$ 5
<b>backbone (LB)</b>	n.d.	50 $\pm$ 4	48 $\pm$ 5	48 $\pm$ 1	47
<b>+<i>cat</i><sub>wt</sub> (LB)</b>	n.d.	46 $\pm$ 1	n.d.	n.d.	48 $\pm$ 1

for plasmid copy number (Figure 3-3A) [38-39, 41-42]. While the replicon of pMPP6 has been reported to have a copy number  $\sim$ 1/4 that of wild-type pSC101 [41], the work here indicates that the two origins have comparable copy numbers. Perhaps this difference is due to the experimental measures used for quantification. Real-time qPCR was used here while Southern blot analysis was employed in the work cited. Discrepancies between these two techniques have been observed [46]. It should be mentioned that the pUC origin of replication is temperature sensitive. While higher plasmid copy number values have been reported for this replicon at 37°C and 42°C, the reduced values observed here are consistent with the 30°C growth temperature used in this study [39]. Similar results were obtained when cells were grown in LB medium (Figure 3-3B), with the growth rate of cells again being comparable among the constructs (Table 3-1). This suggests that the copy number resulting from a particular origin (relative to that of pSC101) does not appear to be impacted by the growth medium used. The results here indicate that the copy number of a device can be varied successfully by changing the origin of replication (Figure 3-3), with the range being  $\sim$ 6X for the constructs tested (Figure 3-2A).

### **Linear device transfer curves obtained in *E. coli***

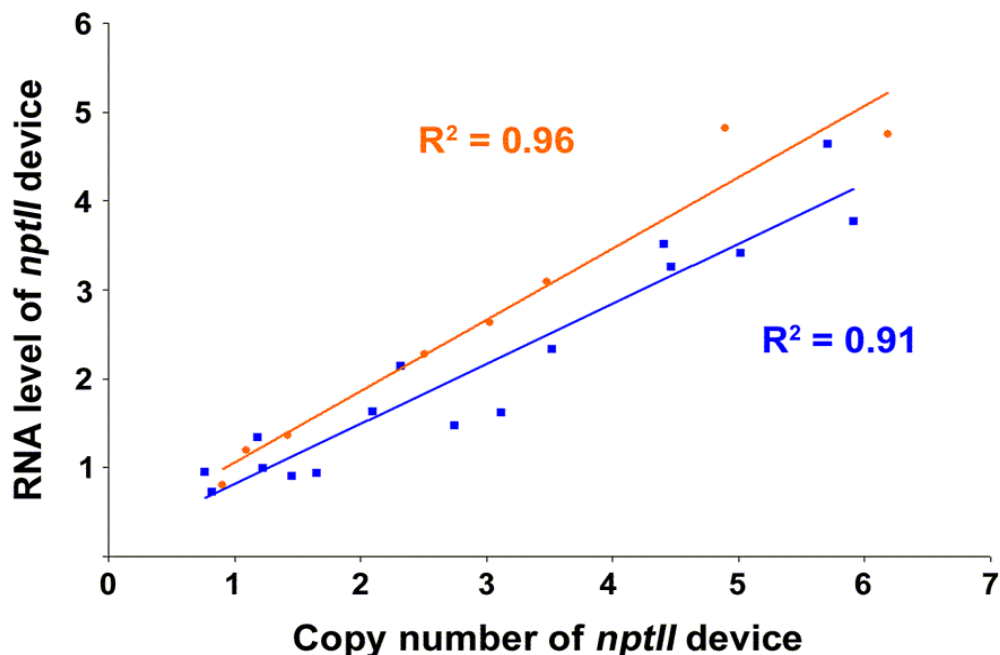
After verifying that the copy number of *nptII* can be varied in the system, experiments were performed to obtain its DTC. Cells that had been harvested prior (Figure 3-3B) were used to quantify the transcript level of the *nptII* device. Total RNA was extracted from cells and transcript level quantified using real-time qPCR [47-49]. The results were plotted against the copy number values determined prior (Figure 3-3B), yielding the DTC (Figure 3-4). RNA transcript level and copy number values have been normalized to that of the pSC101 construct, which were assigned a mean value of one in each case. The y-axis value for each data point indicates how that particular construct's steady-state transcript level compares relative to that of the pSC101 construct. Similarly, the x-axis value for each data point in the plot indicates how that particular construct's



**Figure 3-3** Copy number characterization of backbone (Figure 3-2A). **A.** *nptII* copy numbers obtained from the plasmid backbone harboring the different replicons used. M9 minimal medium was used here. **B.** *nptII* copy numbers obtained from the plasmid backbone harboring the different replicons used. Cells were grown in the indicated growth medium. Each construct was tested at least in duplicate. The mean value has been reported, with the error bars denoting standard error. Values were normalized to that of the pMPP6 (**A**) or pSC101 (**B**) construct for each medium, which was assigned a mean value of one.

steady-state copy number compares relative to that of the pSC101 construct. Also shown in the plot are linear regressions fitted to the data. With  $R^2 > 0.9$ , the data suggest that the *E. coli* system response to the *nptII* device perturbation can be considered linear. The 95% confidence interval for the y-intercepts were also -0.3 – 0.8 and -0.3 – 0.5 for LB and M9 media, respectively. These included the origin, further suggesting that a linear

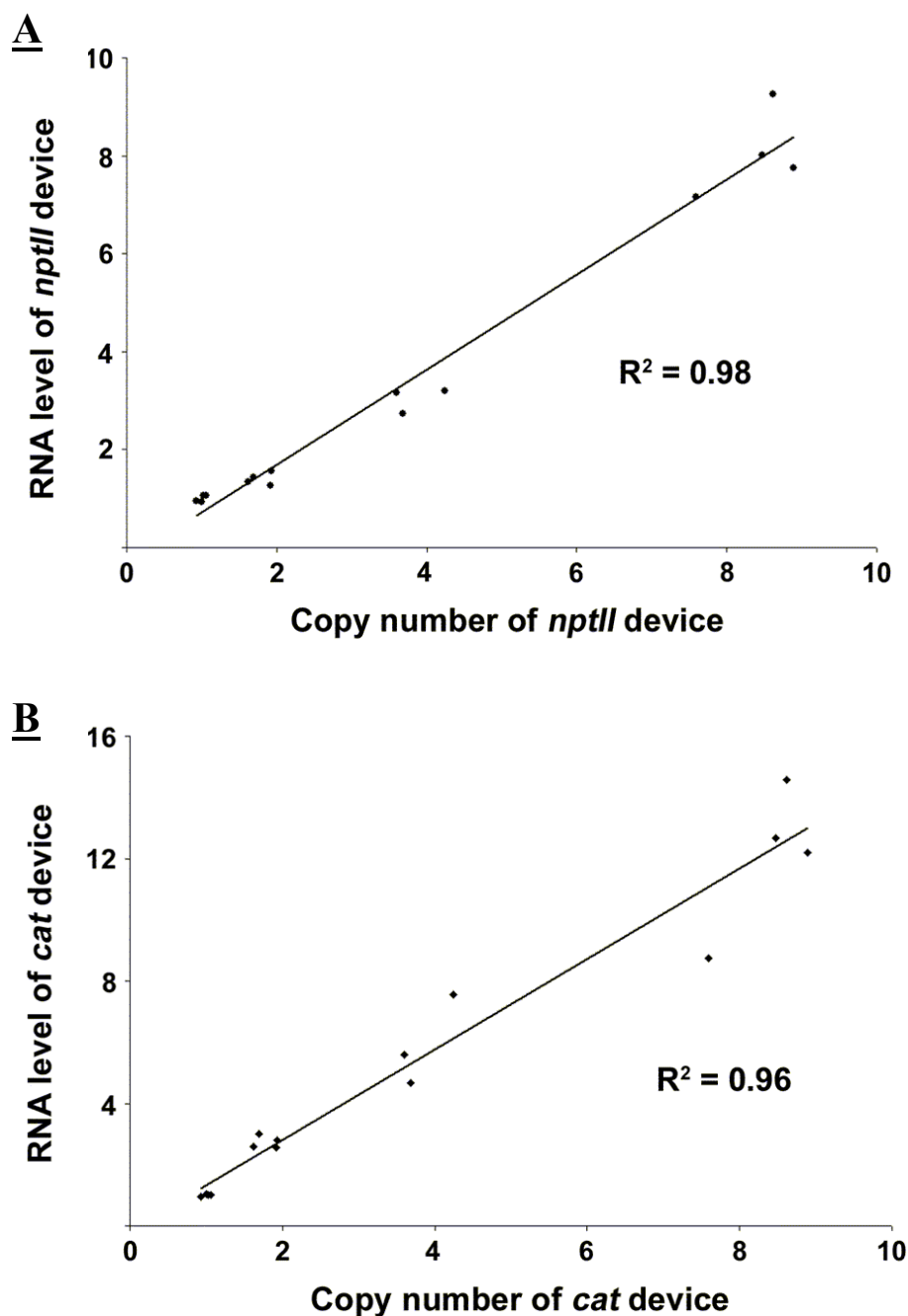




**Figure 3-4** DTC characterization of backbone. *nptII* DTC obtained from the series of constructs (Figure 3-2A) for LB (orange, circle) and M9 (square, blue) media. Each construct was tested at least in duplicate for each medium. RNA and copy number values were normalized to that of the pSC101 construct, with each having been assigned a mean value of one. The regression lines that minimized the sum of squared residual error are also shown, with their corresponding coefficient of determination  $R^2$ .

regression was an appropriate fit for the data. That is, one cannot have *nptII* RNA transcript produced when there is no corresponding DNA present in the cell. The results also suggest that the DTC of *nptII* can be linear under different contexts. That is, the choice of growth medium does not appear to impact system linearity (Figure 3-4).

Next, another device was introduced into the plasmid backbone to increase the perturbation level. The goal was to see whether the presence of an additional device would lead to nonlinear DTCs. This device encoded chloramphenicol acetyl transferase (*cat*), which confers resistance to the antibiotic chloramphenicol. The *cat* device was expressed from its native promoter (Figure 3-2B). *E. coli* DH1 cells harboring the constructs were grown in LB medium as described in the Methods. No chloramphenicol was added to the medium during growth, with only kanamycin having been used for selection purposes. The growth rate of cells was comparable among the constructs, with  $OD_{600nm}$  doubling every ~50 minutes in the log phase (Table 3-1). At mid-exponential growth, cells were harvested and total RNA and DNA extracted. Relative *nptII* and *cat* transcript levels and copy number were subsequently quantified using real-time qPCR (Figures 3-5A and 3-5B). As was done for the backbone (Figure 3-4), linear regressions were fitted to the data. With a  $R^2 > 0.96$  for each device, the data suggest that the combined *nptII* and *cat* device perturbation level appears to be “small” enough to elicit a



**Figure 3-5** DTC characterization of two-device containing constructs. **A,B.** *nptII* and *cat* DTCs obtained from the series of two-device plasmids (Figure 3-2B). The cumulative data presented were the result of two independent experiments. Constructs were tested in duplicate during each independent experiment. Three data points (and not four) are shown for the pMB1 construct because a replicate was lost during sample preparation. RNA and copy number values were normalized to that of the pSC101 construct with each having been assigned a mean value of one. The regression lines that minimized the sum of squared residual error are also shown, with their corresponding coefficient of determination  $R^2$ .

linear response from the *E. coli* system. The 95% confidence intervals were also -0.7 – 0.2 and -1.0 – 0.8 for *nptII* and *cat* y-intercepts, respectively, including the origin for each device once again. The fact that no chloramphenicol was present in the growth medium suggests linear transfer curve response is not due to antibiotic resistance mechanisms.

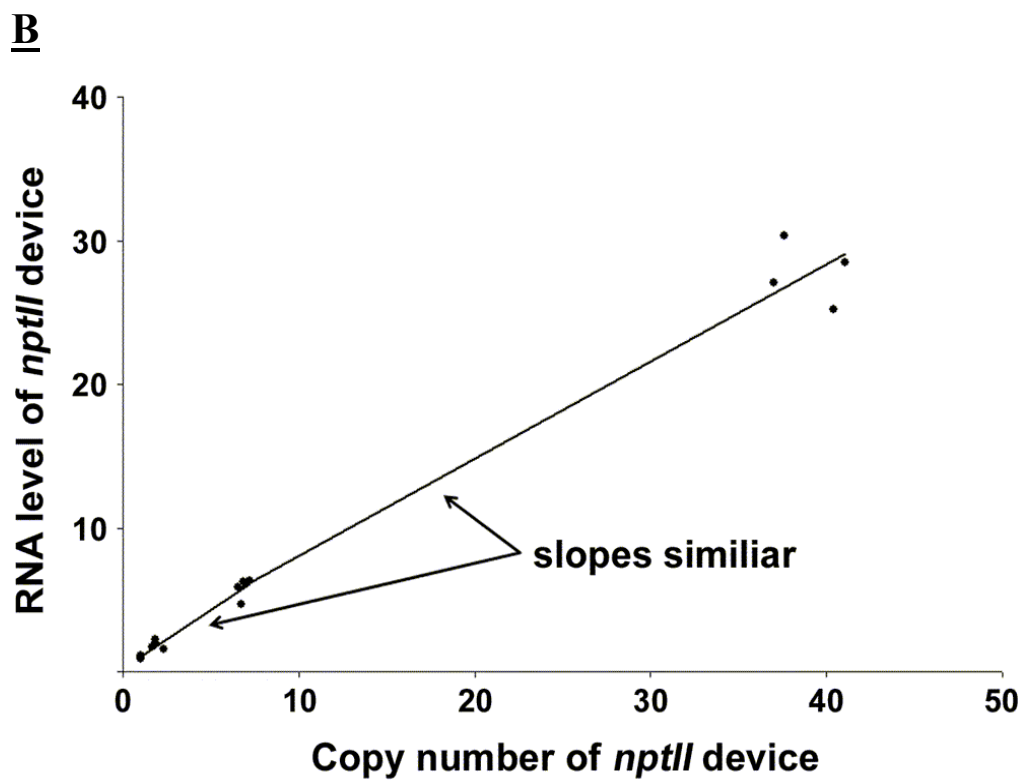
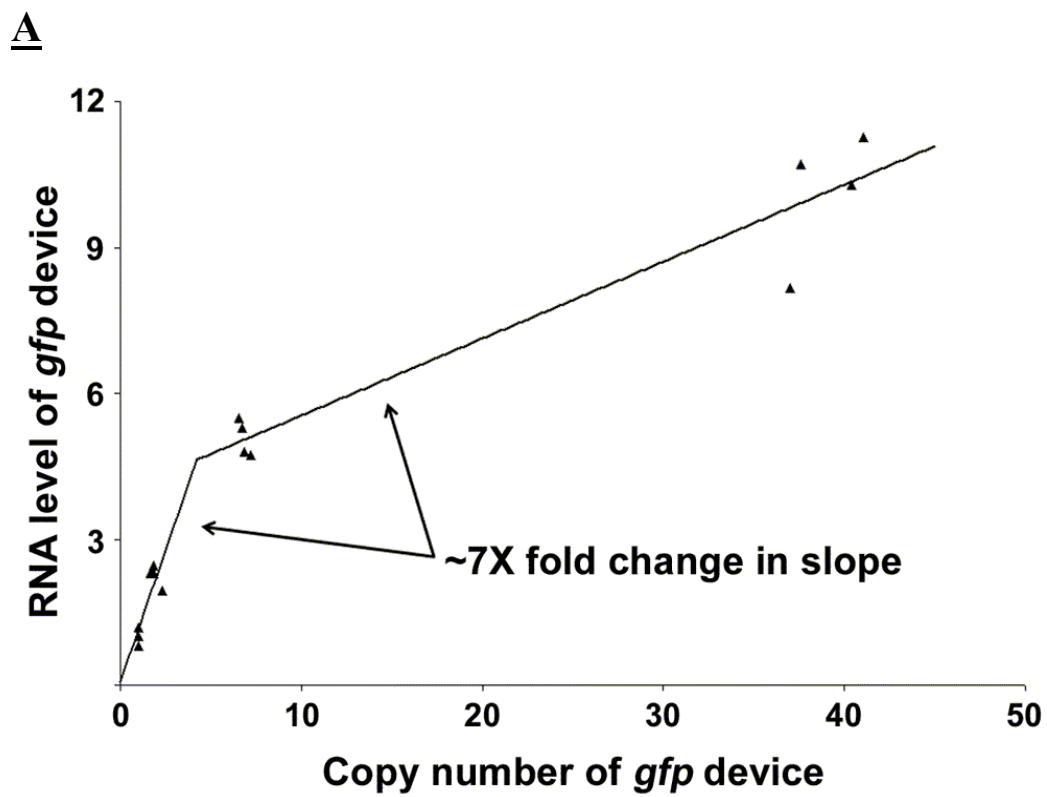
### Introduction of *gfp* genetic device led to nonlinear device transfer curves

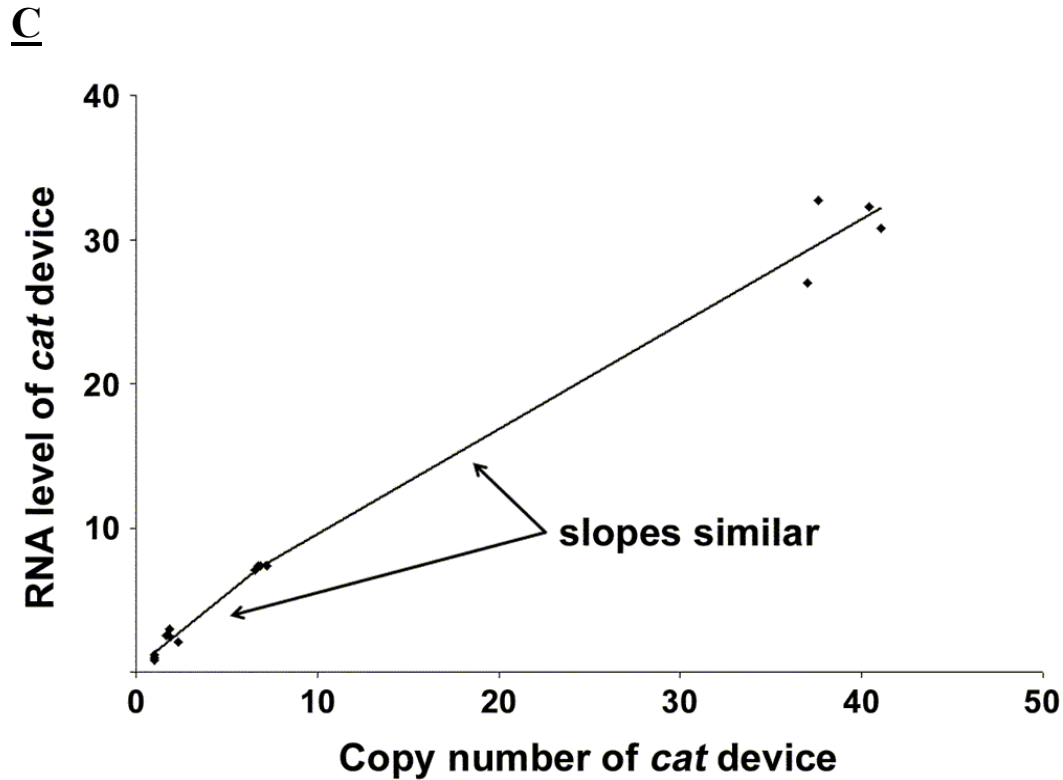
Considering the results described above, a model system consisting of three genetic devices was constructed to determine if nonlinear DTCs would be obtained (Figure 3-2C). The model system is the plasmid backbone analyzed previously (Figure 3-2A) with two additional devices added. One genetic device encodes *cat*, while the other encodes green fluorescent protein (*gfp*). In order to investigate whether other factors besides the identity of a device's promoter impacts linear system behavior, a derivative of the constitutive bacteriophage P<sub>L</sub>  $\lambda$  promoter was used for both *cat* and *gfp* [50]. *E. coli* DH1 cells harboring the model system constructs (Figure 3-2C) were grown in LB medium as described in the Methods. At mid-exponential growth, cells were harvested and total RNA and DNA extracted. Relative *nptII*, *cat*, and *gfp* transcript levels and copy number were subsequently quantified using real-time qPCR to obtain DTCs (Figure 3-6). Note that the copy number of the pUC construct relative to its pSC101 counterpart was ~6X greater compared to the similar constructs in the previously studied series of plasmids (Figures 3-3, 3-4, and 3-5). Unlike the latter series of plasmids, the growth rate of cells was not comparable among constructs harboring the three devices of the model system. While cells harboring the pSC101, p15A, and pMB1 constructs had doubling times comparable to one another (and similar to the ~50 minute doubling time found for the constructs used in the experiments of Figures 3-3, 3-4, and 3-5, see Table 3-1), the doubling time of cells harboring the pUC construct was ~2X greater (Table 3-2). It has been observed that the plasmid copy number of constructs with ColE1-derived origins (e.g. pUC) increases under slow-growth conditions [51]. The results here are consistent with these findings.

**Table 3-2 Growth rate of cells grown in LB medium harboring the different constructs in model system**

Numbers indicate the time in minutes necessary for the OD<sub>600nm</sub> to double in the log phase. The mean value  $\pm$  95% confidence interval as determined in duplicate has been reported.

origin	pSC101	p15A	pMB1	pUC
backbone	45 $\pm$ 1	50 $\pm$ 2	54 $\pm$ 3	49
backbone with <i>cat</i> device	45	51 $\pm$ 2	58 $\pm$ 4	50
backbone with <i>gfp</i> device	46 $\pm$ 1	53 $\pm$ 2	56 $\pm$ 1	80 $\pm$ 2
backbone with <i>cat</i> and <i>gfp</i> devices	46 $\pm$ 1	53	62 $\pm$ 11	103 $\pm$ 2





**Figure 3-6** DTC characterization of the model system. *gfp* (A), *nptII* (B), and *cat* (C) DTCs obtained from the series of model system plasmids (Figure 3-2C). The cumulative data presented were the result of two independent experiments. Constructs were tested in duplicate during each independent experiment. Three data points (and not four) are shown for the pSC101 construct in all panels because a replicate was lost during sample preparation. RNA and copy number values were normalized to that of the pSC101 construct, with each having been assigned a mean value of one. The piecewise-linear fits that minimized the sum of squared residual error are also shown.

As done for the other systems, a linear regression was fitted to the *gfp* data (Figure 3-6A). The results suggested that the *E. coli* system response to the *gfp* device was not linear (not shown). This was due to the 95% confidence interval for the y-intercept not including the origin (i.e. 1.2 – 3). The 95% confidence interval for the y-intercept still did not include the origin if only the first three constructs (i.e. the ones with the pSC101, p15A, and pMB1 replicons) were considered. This suggested a piecewise-linear model for the data, with the first segment consisting of data points for the pSC101, p15A constructs (where the 95% interval for the y-intercept included the origin) and the second segment data points for the pMB1, pUC constructs. The piecewise-linear approximation is used in electrical engineering to model nonlinear transfer curves [52]. To arrive at a piecewise-linear model in a systematic manner, the problem was approached as a nonlinear least squares optimization [53]. The NLIN Gauss-Newton

procedure in SAS was used to fit the data to a piecewise-linear model consisting of two segments with unknown slopes and an unknown breakpoint (Figure 3-6A). The algorithm was not forced to go through the origin. This way, the appropriateness of the fit could later be verified by noting the y-intercept obtained from the slopes and breakpoint numerically computed by the NLIN procedure. A y-intercept of 0.07 was obtained, which is approximately equal to the origin. The change in slope between the two segments (~7X fold) was taken as a means to report the nonlinearity observed in the *gfp* DTC.

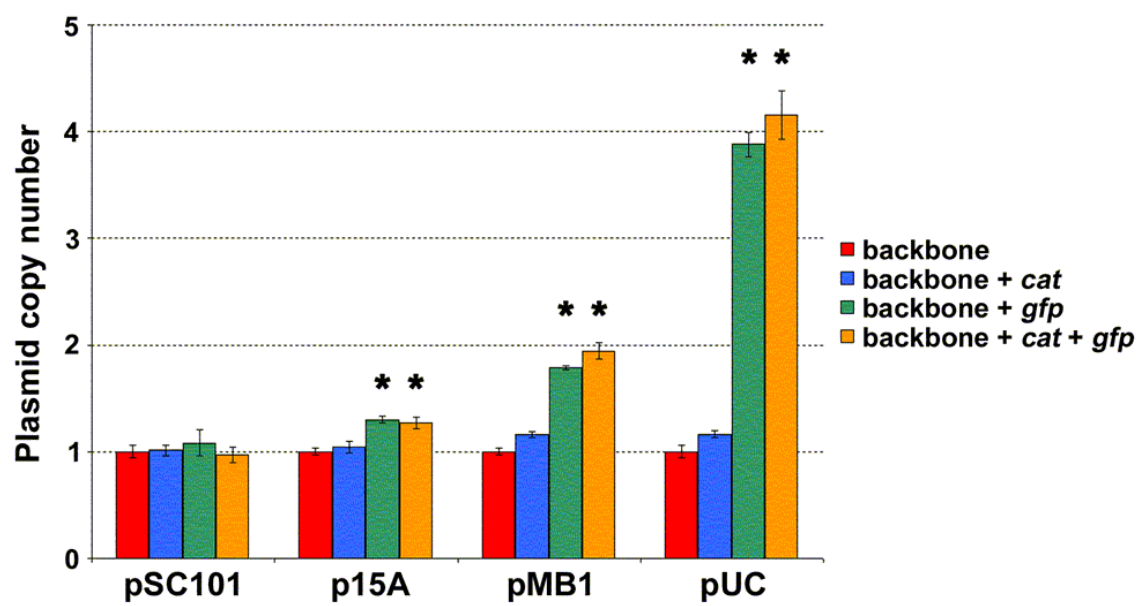
As was done for *gfp*, SAS was used to fit piecewise-linear models to the data for the *nptII* and *cat* devices (Figures 3-6B and 3-6C, respectively). Unlike the former, however, the fits that minimized the sum of squared residuals had the first segment consisting of data points for the pSC101, p15A, and pMB1 constructs (not shown). The second segment could, thus, not be determined because the pUC construct remained as the only available point (i.e. one needs two points to fit a line). As an approximate solution to this problem, a piecewise-linear model was determined for each device by fitting a linear regression to the pSC101, p15A, pMB1 and pMB1, pUC constructs for the first and second segments, respectively (Figures 3-6B and 3-6C). The two segments for the *nptII* and *cat* devices had similar slopes. This was noticeably smaller than the ~7X fold change observed for *gfp* (Figure 3-6A).

### **Superposition lost at higher expression levels with the addition of *gfp***

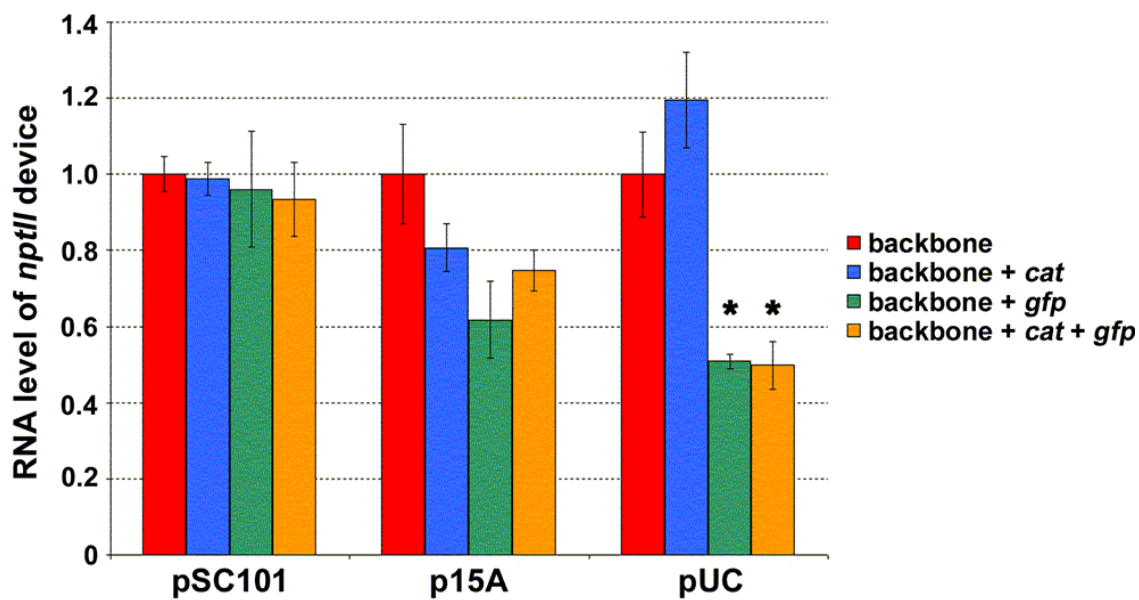
Next, superposition experiments were performed to verify the DTC results of the previous section and to determine whether the *E. coli* system can indeed behave as a linear system under “small” perturbation conditions. If the nonlinear *E. coli* system can be approximated as a linear one, the perturbing devices may be studied independent of one another (Figure 3-1C). This, in turn, would allow one to predict the response of *E. coli* to the complete system (backbone with both *cat* and *gfp* devices) from characterization data of the individual devices. In other words, the addition of new devices would not impact the expression levels of the devices present prior. *E. coli* DH1 cells harboring either the empty plasmid backbones or the various backbones containing *cat* or/and *gfp* devices were grown in LB medium. At mid-exponential growth, cells were harvested and total DNA extracted. Plasmid copy number was subsequently quantified using real-time qPCR (Figure 3-7A). The results indicated that the plasmid copy number was unaffected by the addition of *cat* and/or *gfp* devices to the backbone with a pSC101 origin. Device addition, however, began to have an impact at higher copy numbers. The change in plasmid copy number was most pronounced by the addition of the *gfp* device, with those resulting from *cat* not being statistically significant even with the pUC replicon. While the growth rate data (Table 3-2) also support this finding, the numbers suggest that a change in the plasmid copy is not necessarily reflected by a corresponding change in the cell doubling time.

The larger nonlinearity observed in the *gfp* DTC (Figure 3-6A) interestingly also manifested itself in superposition experiments involving device transcript levels. Cells containing the pSC101, p15A, and pUC origins that had been harvested above were used to quantify the transcript levels of the different devices. The latter two replicons were chosen so as to have data points on either side of the DTC breakpoints (Figure 3-6). The

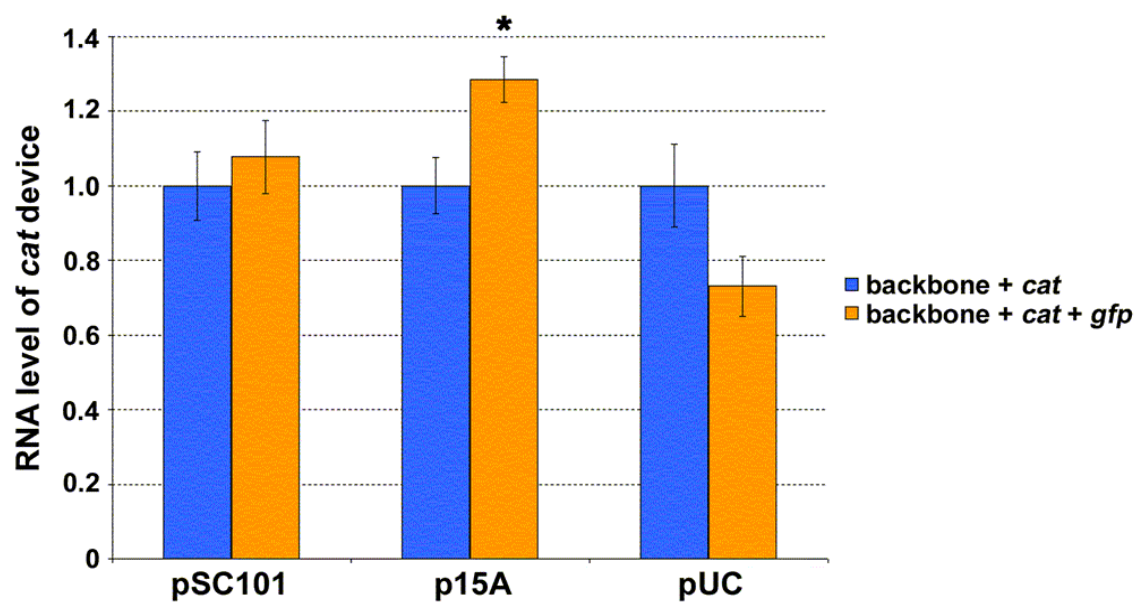
**A**



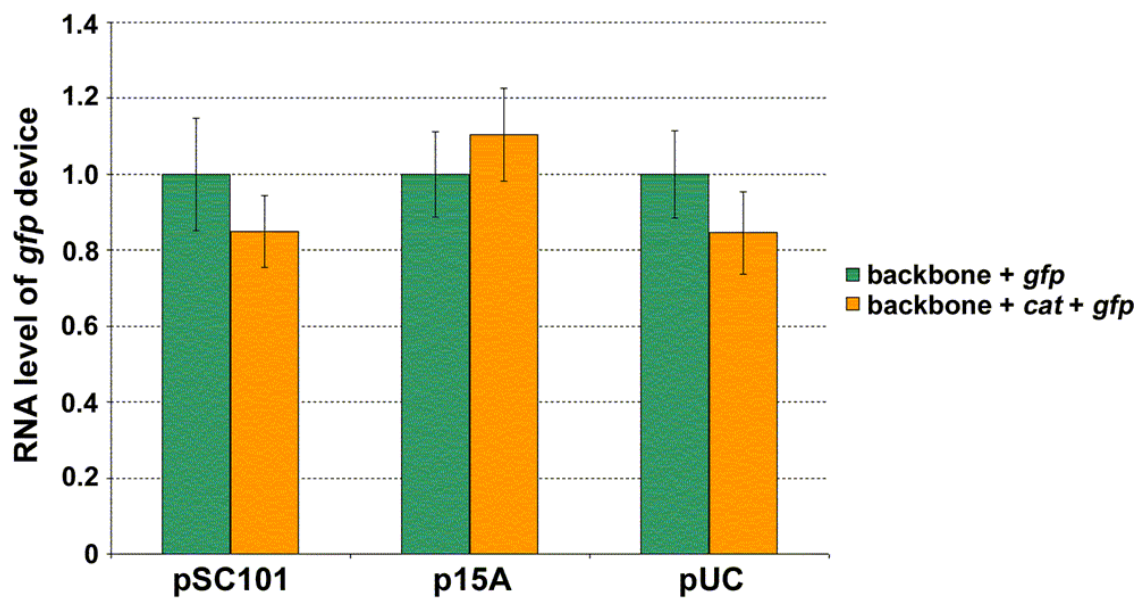
**B**



**C**



**D**





**Figure 3-7** Testing for the presence of superposition in the model system. For each of the indicated replicons, plasmid copy number (**A**) and *nptII* transcript level per plasmid copy (**B**) were determined for empty backbone and backbone harboring *cat* or/and *gfp* devices. Values for each replicon were normalized to that of the empty backbone construct, which was assigned a mean value of one. The difference in value between empty and device harboring backbone was determined statistically for each replicon ( $P < 0.05$ , t-test). Asterisks indicate significant difference. **C**. For each of the indicated replicons, *cat* transcript level per plasmid copy was determined for backbone containing either *cat* device alone or both *cat* and *gfp* devices. Values for each replicon were normalized to that of the backbone construct with only *cat* device present, which was assigned a mean value of one. The difference in value between backbone containing either *cat* device alone or both *cat* and *gfp* devices was determined statistically for each replicon ( $P < 0.05$ , t-test). Asterisks indicate significant difference. **D**. For each of the indicated replicons, *gfp* transcript level per plasmid copy was determined for backbone containing either *gfp* device alone or both *cat* and *gfp* devices. Values for each replicon were normalized to that of the backbone construct with only *gfp* device present, which was assigned a mean value of one. The difference in value between backbone containing either *gfp* device alone or both *cat* and *gfp* devices was determined statistically for each replicon ( $P < 0.05$ , t-test). No significant differences were observed. For all panels, the mean value as determined from two independent experiments (constructs were tested in duplicate during each independent experiment) has been reported. Error bars denote standard error.

pSC101 origin was selected to investigate whether superposition observed at the plasmid copy level (Figure 3-7A) also applied to device transcript levels. Total RNA was extracted from cells and transcript levels quantified using real-time qPCR (Figures 3-7B, 3-7C, and 3-7D). As the plasmid copy varied among the constructs for a particular replicon (Figure 3-7A), transcript levels were not only normalized to the endogenous 16S but also to the plasmid copy number. That is, values reported are RNA produced per unit plasmid. Superposition would be in effect if the amount of RNA produced by the *nptII* device (Figure 3-7B) did not change after additional devices had been introduced into the plasmid. That is with superposition, if the plasmid backbone harboring genetic device *nptII* led to the production of a certain amount of that device's RNA, one would obtain the same amount upon addition of *cat* and/or *gfp* devices. Similar arguments apply to superposition for the *cat* and *gfp* devices (Figures 3-7C and 3-7D, respectively). As was the case for plasmid copy (Figure 3-7A), the results indicated that *nptII* transcript level is unaffected by the addition of *cat* and/or *gfp* genetic devices to a plasmid with a pSC101 replicon (Figure 3-7B). Device addition began to have an impact at higher copy numbers. The changes in *nptII* transcript level were, once again, the most pronounced by the addition of the *gfp* device, with those resulting from *cat* not being statistically significant even with the pUC origin. The data for *cat* and *gfp* RNA exhibited a similar pattern. Once again, *cat* or *gfp* transcript level was unaffected by the addition of the other device to a plasmid with a pSC101 replicon (Figures 3-7C and 3-7D). The addition of *gfp*,

however, affected *cat* RNA at the higher copy numbers (Figure 3-7C). This was not the case in the reverse direction. That is, *cat* device addition did not impact *gfp* RNA at the higher copy numbers of p15A and pUC (Figure 3-7D). These results suggest that the extent of the changes brought about by *gfp* is large enough to mask those caused by the addition of *cat*.

## Discussion

The results indicate that the *E. coli* biological system can exhibit linear system behavior (Figure 3-7). In the model system presented in this chapter, the necessary condition with all three genetic devices present was to use a plasmid backbone harboring the pSC101 replicon. That is, the experimental results showed superposition to be present at this copy number. The presence of superposition, however, was not only a consequence of having used the pSC101 origin. In the absence of the *gfp* device, superposition was found even with a pUC origin (Figure 3-7). The finding that superposition is possible under different contexts is important. It suggests that the nonlinearity in the *E. coli* system is not complex to the point of preventing design efforts to elicit a linear system response.

A simple mathematical model that captures the DTC results can be derived by noting the rate of change of a molecular entity is a synthesis term minus a degradation term. Assuming that the synthesis of RNA is proportional to the amount of DNA present and that its degradation is proportional to the amount of RNA, the equation for RNA becomes [54]

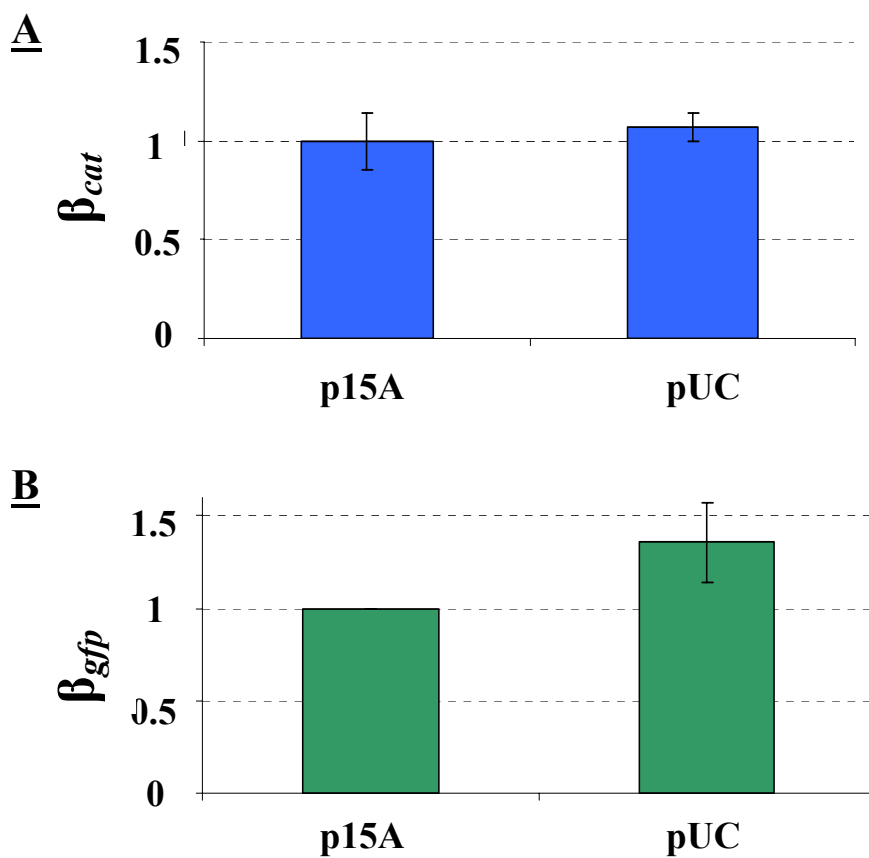
$$\frac{d}{dt} RNA = \alpha \cdot DNA - \beta \cdot RNA, \quad (1)$$

where  $\alpha$  and  $\beta$  are proportionality constants that capture the synthesis and degradation rates, respectively. Under steady-state conditions, the left hand side of equation (1) becomes zero and one arrives at the following equation

$$RNA_{SS} = \frac{\alpha}{\beta} \cdot DNA_{SS}, \quad (2)$$

where  $RNA_{SS}$  and  $DNA_{SS}$  are the steady-state RNA and copy number of the encoding DNA, respectively. The nonlinearity observed in the piecewise-linear DTCs (Figure 3-6) may, thus, be modeled by a change in the  $\frac{\alpha}{\beta}$  slope term of equation (2). That is, the

synthesis and/or degradation rate varies for the devices at higher copy numbers. Analysis of RNA degradation after cells had been treated with rifampicin did not reveal a noticeable change in the decay rates of *gfp* and *cat* transcripts at the higher copy number of pUC relative to p15A (Figure 3-8). This suggests that the larger nonlinearity observed in *gfp* is due to a modulation in the synthesis rate. The fact that *gfp* and *cat* have identical  $P_L$  promoters in our model system further suggests that the mechanism involved does not affect the initiation of transcription. Perhaps, the stringent response is implicated in this matter. Previous work has indicated that the *E. coli* stringent response can differentially impact the elongation rate of transcripts [55]. Cells harboring the model system plasmids (Figure 3-2C) with the pMB1 and pUC replicons did not grow in M9 minimal medium, suggesting that cells might be in a starvation like condition at the higher copy numbers. This is also supported by the growth rate data in LB medium, with the doubling time



**Figure 3-8** Effect of copy number variation on RNA decay rate  $\beta$ . The fold change in the decay rate  $\beta$  of *cat* (A) and *gfp* (B) transcripts as copy number increases from using origin p15A to pUC. The cumulative data presented were the result of two independent experiments. The mean value has been reported, with the error bars denoting standard error. Values have been normalized to that of the p15A construct, which was assigned a mean value of one.

increasing at the higher copy numbers (Table 3-2). Gene sequence has been shown to influence transcriptional pausing of RNA polymerase in the presence of guanosine tetraphosphate (ppGpp) [56]. Perhaps, the coupling present between transcription and translation in *E. coli* facilitates this effect, with ribosomal ppGpp synthesis affecting upstream RNA polymerase that is in the process of transcript elongation. Cooperative activity between RNA polymerase and ribosomes has been shown to modulate the elongation rate of transcripts in *E. coli* [57], with this linkage involving the NusE-NusG complex [58]. It also needs to be acknowledged that copy number was the only perturbing factor considered in our study. The *cat* and *gfp* genes were expressed as done previously [50] (see below). As such, the devices have different RBS sequences (strengths). Promoter strength, RBS strength, and codon usage may be coupled perturbing factors because of the cooperative activity between RNA polymerase and ribosomes. Indeed, a range of promoters, RBS strengths, and codon usage need to be

used to better elucidate the mechanism underlying the large observed nonlinearity in *gfp*'s DTC (Figure 3-6).

The results do, however, suggest that the transfer curve-based framework has application in the engineering of biological systems. A correspondence between DTC nonlinearity and the break down of linear system behavior was observed. That is, the *gfp* device was found to elicit a more nonlinear DTC response from the *E. coli* system than the other tested devices (~7X change in slope as compared to no change for *nptII* and *cat* devices, Figure 3-6), which was reflected in superposition being lost when *gfp* was present at copy numbers above the pSC101 level. While a change in growth rate offers an alternate gauge for nonlinearity, it does not appear to provide one with the same level of accuracy. The doubling time as monitored by OD<sub>600nm</sub> only began to change noticeably with the pUC replicon (Table 3-2), failing to indicate changes to copy number (Figure 3-7A) and transcript (Figures 3-7B, 3-7C, and 3-7D) due to *gfp* at the other origins. This indicates the significance of quantitative techniques (such as DTCs) to synthetic biology characterization efforts because growth rate alone is unable to accurately capture changes that take place due to device addition. DTCs may have application in the general characterization of devices. A device could be characterized by cloning it into a standard plasmid and its copy number varied by way of different replicons. Based on the nonlinearity gauged from its resulting DTC, one may subsequently be able to determine whether the device is well suited for eliciting a predictable, linear response from *E. coli* when used in combination with other devices.

Determining factors that impact linear system behavior in *E. coli* would also be of benefit to synthetic biology. Such knowledge may enable the construction of biological systems using superposition because guidelines for the conditions necessary that ensure linear system behavior would be available. In this research, the focus was on device copy number as the perturbing factor. Promoter strength is another important factor (as are RBS strength, gene length, codon usage, and product function). A library of constitutive promoters has been characterized using the *cat* and *gfp* genes [50]. By expressing *cat* and *gfp* in the manner done in that study, the model system constructed in this work can be used to investigate the effect promoter strength has on linear system behavior. The results here appear to suggest that the identity of a device's promoter is not the only factor that impacts linearity in its RNA expression profile. While *gfp* had a promoter identical to that of *cat*, the former was the most nonlinear in our three-device model system (Figure 3-6). In fact, *cat* and *nptII* had similar DTCs in the three-device model system, but yet had different promoters. Comparison of the *cat* DTCs in the two-device (Figure 3-5B) and three-device (Figure 3-6C) constructs also supports this premise. The *cat* DTC was primarily linear for both cases. The *cat* device in one experiment, however, had its native promoter (Figure 3-5B), while the P<sub>L</sub> promoter was used in the other (Figure 3-6C). Comparison of the DTC results of the two-device (Figure 3-5) and three-device (Figures 3-6B and 3-6C) constructs also suggest that the amount of nonlinearity in the *E. coli* system response to devices harboring *nptII* and *cat* genes is not impacted greatly by slow-growth conditions. The results from the backbone, two-device, and three-device constructs also suggest that the DTC slope (which is the transfer curve gain) may act as a useful metric for characterizing promoter strength of a gene. The gain of the native *nptII* promoter was found to be ~1 in the various constructs tested (Figures 3-4, 3-

5A, and 3-6B), with similar values having been found irrespective of the choice of growth medium (Figure 3-4) or a change in growth rate (Figures 3-5A and 3-6B).

Small-signal linearization techniques may also have application to other aspects of biological system behavior. Input-output relationships can be defined and experimentally measured to generate transfer curves, where piecewise linear models may subsequently be employed to determine the linear range of the system. Examples could include inducer concentration to activated transcription factor, activated transcription factor to RNA, and RNA to protein transfer curves. Measuring input-output characteristics and applying small-signal linearization techniques have the potential of reducing the complex mathematical equations used to model biological interactions to their simplest form; thereby, permitting predictable, quantitative behavior predictions. The limitation of small-signal linearization techniques is that the linearity property needs to be checked. As was observed in the model system, however, some systems can have a linear regime. So long as experiments are performed within this regime, one can avoid nonlinear effects and apply the simplifications associated with a small-signal linear model. And even if the system is to be operated in the nonlinear regime, it may be possible to introduce nonlinear correction factors to the obtained small-signal linear model. In Equation (2), for instance, this can be modeled by having the magnitudes of  $\alpha$  and/or  $\beta$  dependent on the copy number (as opposed to constant values). With the results suggesting that the transfer curve and small-signal concepts used in electrical engineering can likewise be employed towards biological systems, the application of other concepts may also be of benefit to synthetic biology. The transfer curve concept is primarily of use in studying the steady-state behavior of a system. Linear systems can also be studied in the frequency domain by using Fourier techniques, which enable engineers to predict time-domain system response. These techniques have been used previously to study the yeast osmo-adaptation system [59]. Indeed, the application of analysis and design techniques of other established engineering disciplines may enable the systematic forward engineering of biological systems for improved biotechnology applications.

## Methods

### Bacterial strains, media, and enzymes

*E. coli* DH10B and DH5 $\alpha$  were used for cloning. *E. coli* DH1 was used for expression work. Luria-Bertani (LB) media was made as described in [31]. M9 minimal media + 0.5% glucose supplemented with micronutrients was made as described in [60]. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs, with digestion and ligation reactions performed as recommended by the enzyme manufacturer. PCR reactions were performed with Phusion polymerase from Finnzymes, and the primers used were synthesized by Integrated DNA Technologies, Inc. The composition of the PCR reactions, cycle times, and temperatures followed those suggested by the enzyme manufacturer. PCR products were sequenced once cloned into the respective plasmids to ensure that no mutations had been introduced during the amplification process. In cases where single digest cloning was performed, sequencing was also used to select for constructs with inserts in the desired orientation.

## Plasmid construction

The plasmid backbones with the replicons of pMPP6, pSC101, p15A, pMB1, and pUC (Figure 3-2A) were constructed using standard molecular biology techniques [31]. These plasmids were named pAmin 82 [GenBank:HQ283397], pAmin81 [GenBank:HQ283398], pAmin78 [GenBank:HQ283399], pAmin79 [GenBank:HQ283400], and pAmin80 [GenBank:HQ283401], respectively.

The construction of the two-device series of plasmids (Figure 3-2B) proceeded as follows. PCR was used to obtain a spacer and the *cat* gene (with its corresponding native promoter) as inserts. Primers lacZ\_1\_F and lacZ\_1\_R (Table 3-3) were used to obtain the spacer, with p50gl [35] having been used as template. A spacer was used to create spatial separation between the neighboring *nptII* and *cat* devices, and to not have the devices right next to each other. The spacer sequence consisted of a ~600bp fragment taken from within the coding region of the bacterial *lacZ* gene. The *cat* gene was obtained by using the primers cat\_wt\_F and cat\_wt\_R (Table 3-3), with pACYC184 having served as template. The spacer and *cat* inserts were digested with *AvrII*, *XbaI* and *XbaI*, *SacI*, respectively, and ligated into a *AvrII*, *SacI* digested pAmin81 in a three-fragment ligation reaction. The *cat* genetic device (complete with spacer and terminator) was subsequently transferred into pAmin78, pAmin79, pAmin80, and pAmin81 using *MluI* single digest to arrive at the desired series of plasmids (Figure 3-2B).

The construction of the model system series of plasmids (Figure 3-2C) proceeded by first creating a series of *cat* device (with the P<sub>L</sub> promoter) containing plasmids. PCR was used to obtain the *cat* open-reading frame as insert. The primer pairs used were cat\_orf\_F and cat\_orf\_R (Table 3-3), with pACYC184 having served as template. This insert was digested with *KpnI*, *MluI*, and ligated into pZE21 [61] to create pAmin92. PCR was subsequently used to obtain the *cat* gene (with the P<sub>L</sub> promoter) as insert using the primer pairs PL\_F and cat\_wt\_R (Table 3-3) and pAmin92 as template. This insert was digested with *AatII*, *StuI*, and ligated into the two-device (Figure 3-2B) series of plasmids to yield a series of *cat* device (with the P<sub>L</sub> promoter) containing constructs. Next, work began on constructing a *gfp* device (with the P<sub>L</sub> promoter) containing construct with pSC101 as the origin. This plasmid was called pAmin81+*gfp*<sub>PL</sub>. PCR was used to obtain a spacer (different in sequence from that above) and the *cat* gene as inserts. Primer pairs lacZ\_2\_F and lacZ\_2\_R (Table 3-3) were used for the spacer, and p50gl served as the template. The *cat* gene was obtained by using the primers cat\_wt\_F and cat\_wt\_R (Table 3-3), with pACYC184 having served as template. The spacer here was to create spatial separation between the neighboring *cat* and *gfp* devices, and consisted of a ~600bp fragment taken from within the coding region of the bacterial *lacZ* gene. The spacer and *cat* inserts were digested with *AvrII*, *AatII* and *AatII*, *SacI*, respectively, and ligated into a *AvrII*, *SacI* digested pAmin81 in a three-fragment ligation reaction to create pAmin93. The *cat* genetic device (complete with spacer and terminator) was subsequently transferred into pAmin81 to create pAmin99. PCR was next used to obtain the *gfp* open-reading frame as insert. Primer pairs gfp\_F and gfp\_R (Table 3-3) were used, with BBa\_E0044 [14] serving as the template. This insert was digested with *KpnI*, *HindIII*, and ligated into pZE21 to create pAmin100. PCR was then used to obtain the *gfp* gene (with P<sub>L</sub> promoter) using the primer pairs PL\_F and gfp\_2\_R (Table 3-3) and pAmin100

as template. The creation of pAmin81+*gfp*<sub>PL</sub> subsequently proceeded by performing a three-fragment ligation reaction of this fragment digested with *AatII*, *SacI*, the ~2.5kb fragment released from a *AvrII*, *SacI* digested pAmin93, and the ~2kb fragment released from a *AvrII*, *AatII* digested pAmin99. The *gfp* genetic device (complete with spacer and terminator) was finally transferred from pAmin81+*gfp*<sub>PL</sub> into the *cat* device (with the P<sub>L</sub> promoter) containing series of plasmids described prior using *BamHI*, creating the desired series of plasmids (Figure 3-2C). Sub-cloning was used in order to arrive at the *gfp* device (with the P<sub>L</sub> promoter) containing constructs with the other three origins of replication. More specifically, the origins released from a *AvrII*, *SmaI* digested pAmin78, pAmin79, and pAmin80 were ligated into pAmin81+*gfp*<sub>PL</sub> to yield pAmin78+*gfp*<sub>PL</sub>, pAmin79+*gfp*<sub>PL</sub>, and pAmin80+*gfp*<sub>PL</sub>.

### **Bacterial growth conditions**

*E. coli* DH1 cells were grown overnight at 30°C, 200 rpm shaking after inoculating 5 mL cultures of LB media (supplemented with 50 µg/mL kanamycin) with single colonies from freshly streaked plates. After sub-culturing (1:50) into shake flasks containing 50 mL of either M9 minimal or LB media (supplemented with 50 µg/mL kanamycin), cells were grown at 30°C, 200 rpm shaking until an OD<sub>600nm</sub> of 0.3–0.4 was reached to approximate steady-state conditions. At this time, 1 mL of cells were added to ice chilled tubes with 100 µL of 10% phenol:90% EtOH stop solution [62], mixed, spun down, supernatant removed, and total RNA isolation proceeded immediately thereafter. Another 1 mL of cells were spun down, supernatant removed, and cell pellets subsequently frozen for total DNA isolation at a future date.

### **Collection of bacterial cells for RNA decay measurements**

*E. coli* DH1 cells were grown overnight at 30°C, 200rpm shaking after inoculating 5mL cultures of LB media (supplemented with 50µg/mL kanamycin) with single colonies from freshly streaked plates. After sub-culturing (1:50) into shake flasks containing 50mL of LB media (supplemented with 50µg/mL kanamycin), cells were grown at 30°C, 200rpm shaking until an OD<sub>600nm</sub> of 0.3–0.4 was reached to approximate steady-state conditions. Rifampicin dissolved in methanol was added at this time to arrest transcription (final concentration = 250µg/mL). Cells were collected at serial time points for analysis of RNA decay after rifampicin treatment. At each time point, 1mL of cells were added to ice chilled tubes with 100µL of 10% phenol:90% EtOH stop solution [62], mixed, spun down, supernatant removed, and frozen for total RNA isolation at a future date.

### **Bacterial total RNA isolation to quantify *nptII*, *cat*, and *gfp* expression levels**

Bacterial cell pellets were resuspended in 700 µL buffer RLT (Qiagen), to which beta-mercaptoethanol had been added according to the manufacturer's instructions. Cells were subsequently lysed using 0.1 mm diameter glass beads in the Mini-Beadbeater-8 (Biospec). Following lysis, tube contents were spun down and 500 µL of lysate was transferred to new tubes. Total RNA extraction then proceeded by adding 500 µL of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexing vigorously for ~1 min, allowing the tubes to sit at bench for a few minutes subsequent, and centrifugation for 15 min at

12000 x g, 4°C. Next, 300 µL of the upper aqueous phase was transferred to a new tube containing 300 µL nuclease free water. RNA extraction continued by adding 600 µL of chloroform to each tube, vigorous vortexing for ~1 min, allowing the tubes to sit at bench for a few minutes subsequent, and centrifugation for 15 min at 12000 x g, 4°C. Next, 300 µL of the upper aqueous phase was transferred to a new tube. Following chloroform extraction, total RNA was ethanol precipitated overnight, washed with 70% ethanol, and finally resuspended in 30 µL of nuclease free water. RNA concentration and purity were assayed using a Nanodrop spectrophotometer, and integrity examined on 2% agarose gels.

### **cDNA synthesis and real-time qPCR quantification of cellular *nptII*, *cat*, and *gfp* transcript levels**

Total RNA extracted was treated with Turbo DNase (Ambion) to reduce DNA contamination. First-strand cDNA was synthesized by using reverse gene-specific primers (Table 3-4) and SuperScript III Reverse Transcriptase (Invitrogen) following the manufacturer's instructions. Transcript levels were normalized to that of endogenous 16S rRNA. The primer sets specific to *nptII*, *cat*, *gfp*, and 16S rRNA (Table 3-4) amplified a single product of the expected size as confirmed by the melting temperatures of the amplicons. Real-time qPCR was conducted on a BioRad iCycler with 96-well reaction blocks in the presence of SYBR Green under the following conditions: 1X iQ SYBR Green Supermix (BioRad), 150 nM *nptII*, 300 nM *cat*, 100 nM *gfp*, or 500 nM 16S primers in a 25 µL reaction. Real-time qPCR cycling was 95°C for 3 min, followed by 40 cycles of 30 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C. Threshold cycles (Ct) were determined with iCycler (BioRad) software for all samples. A standard curve was prepared for quantification. For this purpose, a fourfold dilution series of a total of seven dilutions was prepared from a digested total DNA sample, and each dilution was subjected to qPCR analysis in triplicate with either the *nptII*-, *cat*-, *gfp*-, or 16S-specific primers. Obtained Ct values were used by the iCycler software package to plot a standard curve that allowed quantification of *nptII*, *cat*, *gfp*, or 16S in the total RNA samples (i.e. unknowns) relative to the RNA sample used to prepare the standard curve.

### **Bacterial total DNA isolation to quantify plasmid copy number**

The DNA isolation method reported in the previous publications [43, 63] was adopted. Bacterial cell pellets were resuspended in 400 µL of 50 mM Tris/50 mM EDTA, pH 8, by vortex. Cell membranes were permeabilized by the addition of 8 µL of 50 mg/mL lysozyme (Sigma) in 10 mM Tris/1 mM EDTA, pH 8, followed by incubation at 37°C for 30 min. To complete cell lysis, 4 µL of 10% SDS and 8 µL of 20 mg/mL Proteinase K solution (Invitrogen) were added to each tube, mixed with a syringe with 21 gauge 1.5 inch needle, and incubated at 50°C for 30 min. Proteinase K was subsequently heat inactivated at 75°C for 10 min, and RNA was digested with the addition of 2 µL of 100 mg/mL RNase A solution (Qiagen) followed by incubation at 37°C for 30 min. Total DNA extraction then proceeded by adding 425 µL of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexing vigorously for ~1 min, allowing the tubes to sit at bench for a few minutes subsequent, and centrifugation for 5 min at 12000 x g, 4°C. Next, 300 µL of the upper aqueous phase was transferred to a new tube using a wide-opening pipet tip. DNA extraction continued by adding 400 µL of chloroform to each tube, vigorous vortexing for ~1 min, allowing the tubes to sit at bench for a few minutes subsequent, and



centrifugation for 5 min at 12000 x g, 4°C. Next, 200 µL of the upper aqueous phase was transferred to a new tube using a wide-opening pipet tip. Following chloroform extraction, total DNA was ethanol precipitated overnight, washed with 70% ethanol, and finally resuspended in 40 µL of nuclease free water. DNA concentration and purity were assayed using a Nanodrop spectrophotometer, and integrity examined on 1% agarose gels.

### **Real-time qPCR quantification of plasmid copy number**

Primer sets specific to the *nptII* and 16S rDNA genes were used (Table 3-4). These primers amplified a single product of the expected size as confirmed by the melting temperatures of the amplicons. The *nptII* gene is a single-copy gene of the plasmids characterized in this study, while 16S is a multi-copy gene of *E. coli* chromosomal DNA [64] and was used for normalization purposes [43, 45]. Total DNA isolated from each strain was first digested overnight using EcoRI (New England Biolabs) at 37°C. Real-time qPCR was conducted on a BioRad iCycler with 96-well reaction blocks in the presence of SYBR Green under the following conditions: 1X iQ SYBR Green Supermix (BioRad), 150 nM *nptII*, or 500 nM 16S primers in a 25 µL reaction. Real-time qPCR cycling was 95°C for 3 min, followed by 40 cycles of 30 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C. Threshold cycles (Ct) were determined with iCycler (BioRad) software for all samples. A standard curve was prepared for quantification. For this purpose, a fourfold dilution series of a total of seven dilutions was prepared from a digested total DNA sample, and each dilution was subjected to qPCR analysis in triplicate with either the *nptII*- or 16S-specific primers. Obtained Ct values were used by the iCycler software package to plot a standard curve that allowed quantification of *nptII* or 16S in the digested total DNA samples (i.e. unknowns) relative to the DNA sample used to prepare the standard curve.

**Table 3-3** List of PCR primers used in the cloning of the plasmids constructed in this chapter

<u>Primer</u>	<u>Sequence</u>
lacZ_1_F	tattatctcgagtacctaagggtaacagtttctttatgg
lacZ_1_R	tattattctagattcgctggcacttcgatggtttg
cat_wt_F	tattattctagagacgtcgaataaatacctgtgacggaag
cat_wt_R	tattaagagctcaggcctaataactgccttaaaaaaattacg
cat_orf_F	tattatggtacctttcaggagctaaggaagctaaaatg
cat_orf_R	taataaacgcgtccaataactgccttaaaaaaattacg
PL_F	tattatgacgtctccctatcagtgatagagattgacatc
lacZ_2_F	tattaacctaaggaggatccatgttgccactcgc
lacZ_2_R	taataagacgtcatcggtcagacgattcattg
gfp_F	tattatggtaccgcatgcgtaaaggagaagaacttttcactggagttgtcc
gfp_R	taataaaagcttattaaactgatgcagcgtagtttcgctgcttgctgcaggccttttg
gfp_2_R	tattaagagctcgaagtgcctcaagcttattaaactgatgcagcgtag

**Table 3-4** List of real-time qPCR primers used in this study

<u>Primer</u>	<u>Sequence</u>	<u>Reference</u>
qpcr_nptII_F	gcgttggtaccggtgatat	[65]
qpcr_nptII_R	aggaagcggtcagcccat	
qpcr_cat_F	cgcaaggcgacaagggtg	[66]
qpcr_cat_R	ccatcacaacggcatgatg	
qpcr_gfp_F	aagcgttcaactagcagacc	[50]
qpcr_gfp_R	aaagggcagattgtgtggac	
qpcr_16S_F	ccggattggagtctgcaact	[43]
qpcr_16S_R	gtggcattctgatccacgattac	

## Chapter 4: Plasmid Copy Number Characterization of a Set of BglBrick-Compatible Vectors

### Background

As stated in Chapter 1, one of the earliest contributions of synthetic biology was the creation of the Registry of Standard Biological Parts [14]. A biological part is a DNA sequence that encodes a definable biological function (e.g. a gene, promoter, terminator, RBS, origin). Parts can also be composed of multiple functions (e.g. a gene and its promoter, RBS, and terminator). The registry serves as a repository of biological parts (mostly sent in by undergraduates participating in the International Genetically Engineered Machine competition), while also proposing standard rules that support the physical assembly of the different parts together. The goal of the registry is to promote sharing of parts among members of the synthetic biology community, thereby allowing biological engineers to design and assemble engineered organisms in a more systematic and rapid manner [67]. One problem is that many parts have not been well characterized. Experimental tests have not always been performed that characterize what a particular part does and whether the performance changes with different cell types. In response, the registry has been stepping up efforts to encourage contributors to include documentation on part function and performance [8]. In parallel, a new effort named BIOFAB has received funding from the National Science Foundation to professionally develop and characterize existing and new parts [8]. In the spirit of these efforts, parts conferring origin of replication functionality were further characterized in this chapter. The work here extends that done in the previous chapter by determining the absolute copy number of plasmid that results from a particular replicon (as opposed to the relative quantification done in Chapter 3), the impact (if any) that a change in bacterial cell type has on plasmid copy number, and the effect (if any) on a plasmid's copy number of introducing plasmids with different replicon into the same strain. The characterization data reported here may serve as useful information for future assays that examine linearity by way of transfer curves and/or metabolic pathway design efforts towards producing valuable chemicals in *E. coli*. It should be noted that the experiments conducted in this chapter were done in the context of a larger team effort to design, construct, and characterize 96 BglBrick-compatible plasmids harboring different origins of replication, selection markers, and promoters [68]. As the manuscript reporting this work is still in progress, details have been omitted here.

### Results and Discussion

#### Copy number of plasmid in single plasmid strain

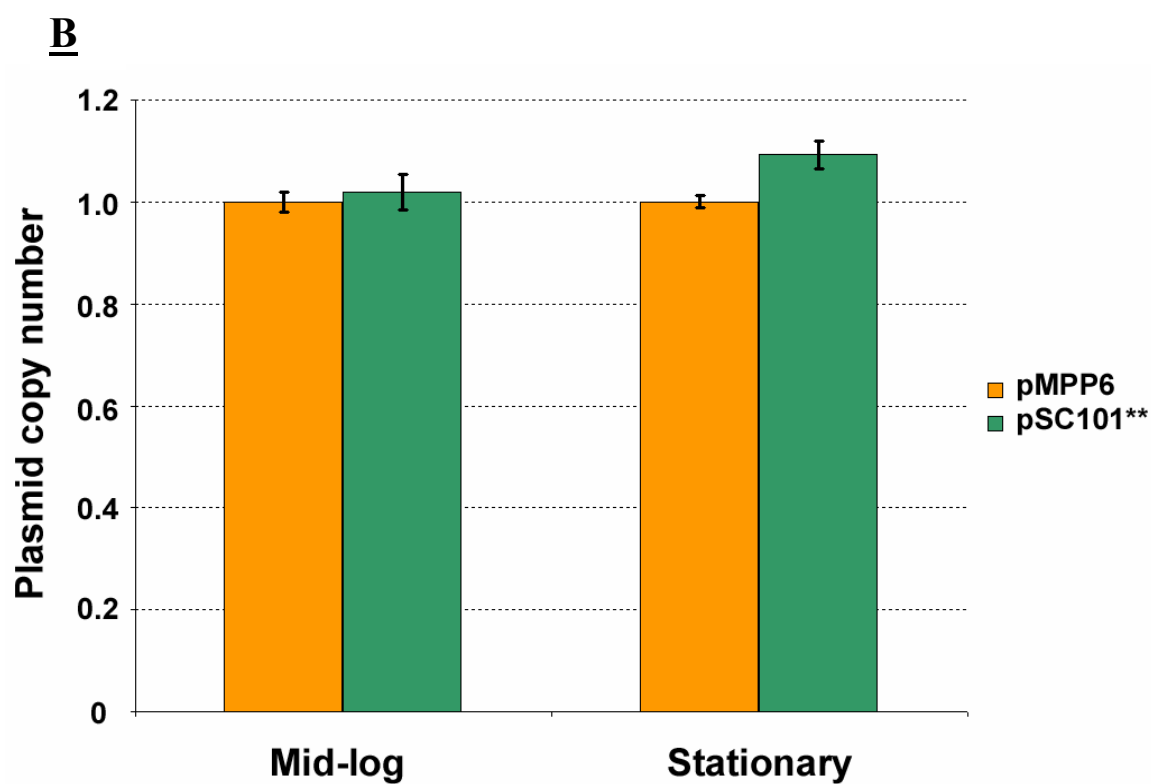
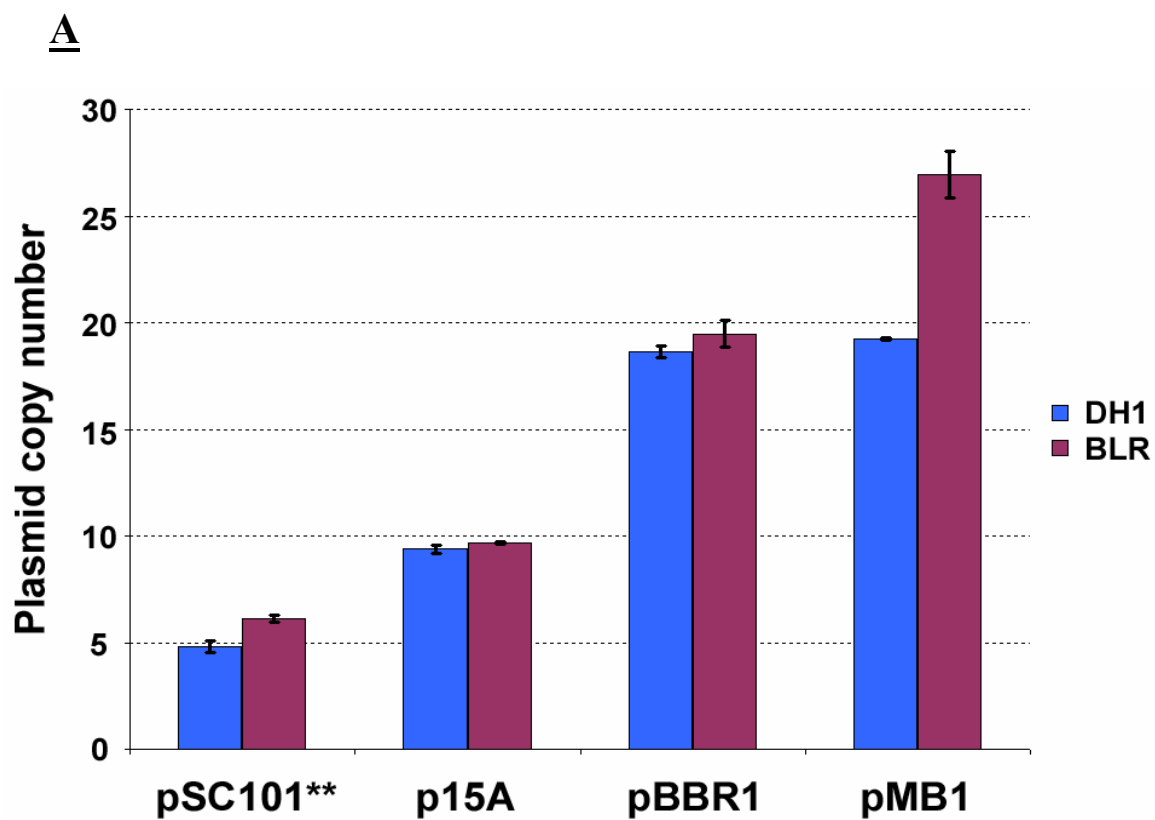
A set of plasmids compatible with the *BglII* and *BamHI* restriction sites (i.e. BglBrick-compatible) was constructed [68]. Four replicons (belonging to different incompatibility groups) were chosen in order to vary the copy number of the constructed

plasmids. A couple of these origins of replication were also used in the previous chapter, namely p15A [38] and pMB1 [39]. One of the additional replicons utilized here is a mutated version originating from pMPP6 [41], which will be referred to as pSC101\*\* [68]. To arrive at the pSC101\*\* origin, site-directed mutagenesis was used to introduce a single nucleotide substitution in the pMPP6 origin [68]. This was to eliminate a *Bgl*III restriction site present in the latter. The other additional replicon is a derivative of the broad host origin pBBR1. The pBBR1 origin has been widely used in metabolic engineering because of its compatibility with most other replicons, thereby allowing the simultaneous use of plasmid harboring this origin with additional plasmids harboring other replicons [45, 69-70].

Real-time quantitative PCR was used to determine the copy numbers of plasmid [43-45, 71-72] harboring the aforementioned replicons. Plasmid copy numbers were determined for plasmids transformed into both BLR and DH1 *E. coli* strains. The single copy *nptII* gene harbored on plasmid served as the target, with the multi-copy 16S rDNA gene harbored on host chromosome [64] having been used for normalization purposes [43, 45]. Here, plasmid copy number is defined as the number of copies of plasmid present per chromosomal equivalent in *E. coli* [73], and absolute plasmid copy numbers were obtained by using BLR and DH1 transgenic strains containing single *nptII* integration in the chromosome as the reference sample. The plasmid copy number values obtained for strains with a single plasmid of a particular replicon are comparable to those found in literature (Figure 4-1A) [42, 61, 74]. The results suggest that the DH1 and BLR strains do not differentially impact the regulation of plasmid copy number for the pSC101\*\*, p15A, and pBBR1 replicons. This is interesting as the replication mechanism of plasmids harboring these origins is different. While the mechanism of pBBR1 has not yet been elucidated, p15A-based plasmids replicate under relaxed control while pSC101 (the parent of pSC101\*\*) is stringent [33]. The results, however, suggest that the choice of strain does impact the copy number obtained from a pMB1-based plasmid (Figure 4-1A). Interestingly, similar copy numbers were obtained for plasmid harboring the pBBR1 and pMB1 origins in strain DH1 and pBBR1 origin in strain BLR (Figure 4-1A). Perhaps the discrepancy observed in pMB1-based plasmids is due to differences in the concentrations and/or activity levels of the molecular players involved in its replication control. While the specific mechanism underlying this discrepancy is not known, the results here suggest that the pBBR1 origin may be preferable in applications necessitating a higher plasmid copy number. That is, the plasmid copy number appears to be less sensitive to the choice of strain as compared to the pMB1 replicon. The pSC101\*\* origin is a derivative of that of pMPP6, and comparable copy numbers for plasmids harboring either one of these replicons was also observed (Figure 4-1B). This was examined for control purposes, and suggests that the point-mutation introduced in pMPP6 to eliminate its *Bgl*III restriction site does not greatly affect the plasmid copy number resulting from the replicon.

### **Copy number of plasmid in multi-plasmid strain**

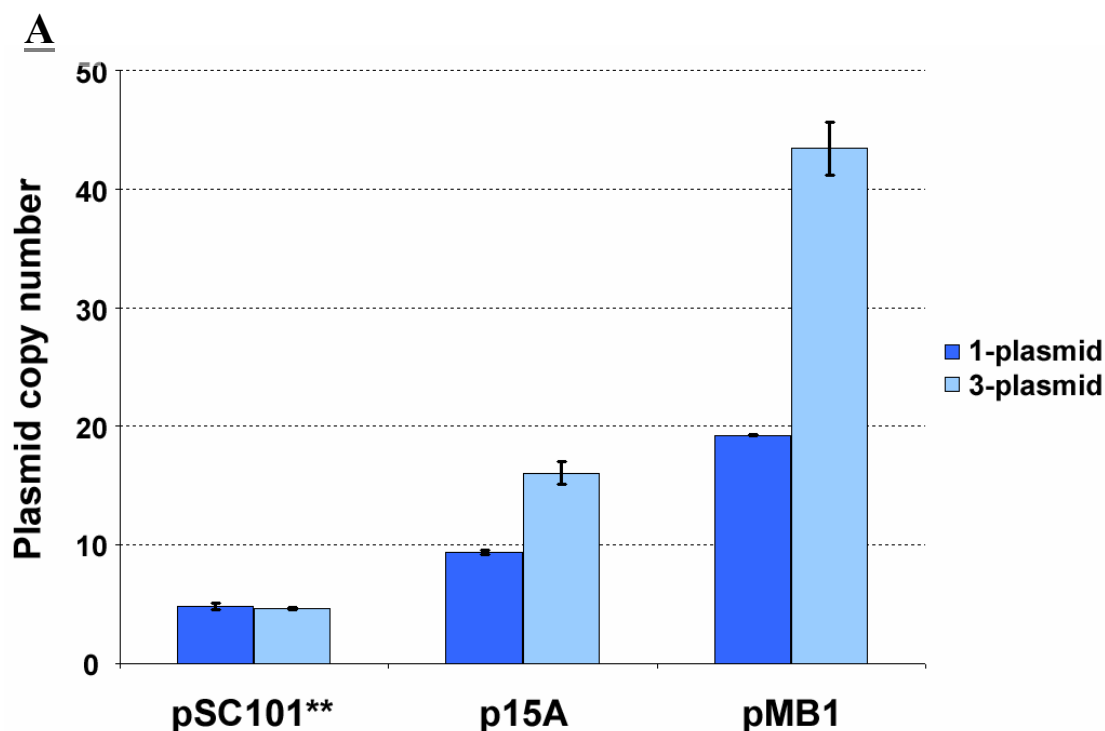
In order to assess the impact of metabolic burden on plasmid copy number, the copy number of individual plasmids in the context of a three-plasmid system was determined. Here, DH1 and BLR *E. coli* strains were co-transformed with pSC101\*\*, p15A, and pBBR1 plasmids.

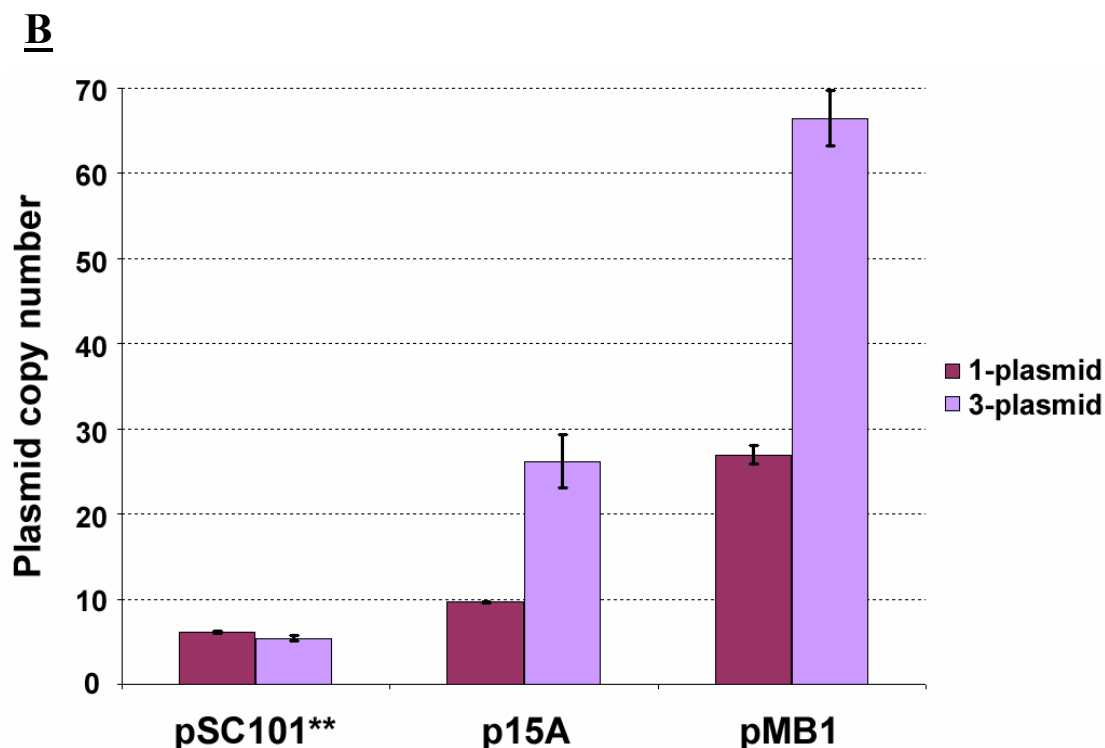


**Figure 4-1** BglBrick plasmid copy numbers in strains with single plasmid. **A.** For each of the indicated replicons, absolute plasmid copy number was determined in *E. coli* DH1 and BLR strains. **B.** For each of the indicated replicons, plasmid copy number was determined in *E. coli* DH1 when cell pellets were retrieved either in mid-log or stationary phase of growth. Values for each time point were normalized to that of the pMPP6 construct, which was assigned a mean value of one. For all panels, the mean value as determined in duplicate has been reported. Error bars denote standard error.

p15A, and pMB1 origin of replication containing plasmids. Each plasmid contained a different selection marker (resistant to ampicillin, chloramphenicol, or kanamycin), and cultures were grown in medium supplemented with all three antibiotics. Using real-time quantitative PCR, absolute copy number was determined for the plasmid harboring the kanamycin selection marker *nptII*. The results (Figure 4-2) suggest that p15A and pMB1-based plasmids lead to a higher copy number for plasmid under increased metabolic load conditions. For pSC101\*\*, on the other hand, the plasmid copy number was unchanged from that observed under single plasmid conditions.

The results obtained (Figure 4-2) are consistent with those found in the literature [51], and may be explained by differences in the replication control mechanism of the replicons. As a broad generalization, plasmid replication control is either relaxed or stringent, with plasmid replication being uncoupled from host chromosomal DNA





**Figure 4-2** BglBrick plasmid copy numbers in strains with three plasmids. For each of the indicated replicons, absolute plasmid copy number was determined in *E. coli* DH1 (A) and BLR (B) strains. In the case of the 3-plasmid system, strains were co-transformed with pSC101\*\*, p15A, and pMB1 origin of replication containing plasmids (with each containing a different selection marker). The absolute plasmid copy number reported here was for the one harboring the indicated replicon. For all panels, the mean value as determined in duplicate (for 1-plasmid system) and triplicate (for 3-plasmid system) has been reported. Error bars denote standard error.

synthesis in the former case [75]. Plasmids harboring pMB1 and p15A replicons replicate under relaxed control while pSC101 (the parent of pSC101\*\*) is stringent [33]. Consistent with the results (Figure 4-2), variation in pSC101 copy number is not to be expected under stress conditions because plasmid replication is tightly coupled to the bacterial cell cycle.

## Methods

### Bacterial total DNA isolation to quantify plasmid copy number

The DNA isolation method reported in the previous publications [43, 63] was adopted. Bacterial cell pellets were resuspended in 400  $\mu$ L of 50 mM Tris/50 mM EDTA, pH 8, by vortex. Cell membranes were permealized by the addition of 8  $\mu$ L of 50 mg/mL

lysozyme (Sigma) in 10 mM Tris/1 mM EDTA, pH 8, followed by incubation at 37°C for 30 min. To complete cell lysis, 4 µL of 10% SDS and 8 µL of 20 mg/mL Proteinase K solution (Invitrogen) were added to each tube, mixed with a syringe with 21 gauge 1.5 inch needle, and incubated at 50°C for 30 min. Proteinase K was subsequently heat inactivated at 75°C for 10 min, and RNA was digested with the addition of 2 µL of 100 mg/mL RNase A solution (Qiagen) followed by incubation at 37°C for 30 min. Total DNA extraction then proceeded by adding 425 µL of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexing vigorously for ~1 min, allowing the tubes to sit at bench for a few minutes subsequent, and centrifugation for 5 min at 12000 x g, 4°C. Next, 300 µL of the upper aqueous phase was transferred to a new tube using a wide-opening pipet tip. DNA extraction continued by adding 400 µL of chloroform to each tube, vigorous vortexing for ~1 min, allowing the tubes to sit at bench for a few minutes subsequent, and centrifugation for 5 min at 12000 x g, 4°C. Next, 200 µL of the upper aqueous phase was transferred to a new tube using a wide-opening pipet tip. Following chloroform extraction, total DNA was ethanol precipitated overnight, washed with 70% ethanol, and finally resuspended in 40 µL of nuclease free water. DNA concentration and purity were assayed using a Nanodrop spectrophotometer, and integrity examined on 1% agarose gels.

### **Real-time qPCR quantification of plasmid copy number**

Primer sets specific to the *nptII* and 16S rDNA genes were used (Table 4-1). These primers amplified a single product of the expected size as confirmed by the melting temperatures of the amplicons. The *nptII* gene is a single-copy gene of the plasmids characterized in this study, while 16S is a multi-copy gene of *E. coli* chromosomal DNA [64] and was used for normalization purposes [43, 45]. Total DNA isolated from each strain was first digested overnight using EcoRI (New England Biolabs) at 37°C. Real-time qPCR was conducted on a BioRad iCycler with 96-well reaction blocks in the presence of SYBR Green under the following conditions: 1X iQ SYBR Green Supermix (BioRad), 150 nM *nptII*, or 500 nM 16S primers in a 25 µL reaction. Real-time qPCR cycling was 95°C for 3 min, followed by 40 cycles of 30 sec at 95°C, 30 sec at 60°C, and 30 sec at 72°C. Threshold cycles (Ct) were determined with iCycler (BioRad) software for all samples. A standard curve was prepared for quantification. For this purpose, a fourfold dilution series of a total of seven dilutions was prepared from a digested total DNA sample, and each dilution was subjected to qPCR analysis in triplicate with either the *nptII*- or 16S-specific primers. Obtained Ct values were used by the iCycler software package to plot a standard curve that allowed quantification of *nptII* or 16S in the digested total DNA samples (i.e. unknowns) relative to the DNA sample used to prepare the standard curve.

**Table 4-1** List of real-time qPCR primers used in this study

<u>Primer</u>	<u>Sequence</u>	<u>Reference</u>
qpcr_nptII_F	gcgttggtaccggtgat	[65]
qpcr_nptII_R	aggaagcggcagcccat	
qpcr_16S_F	ccggattggagtctgcaact	
qpcr_16S_R	gtggcattctgatccacgattac	
		[43]



## Chapter 5: Conclusion

### Thesis Summary

This research has demonstrated that the nonlinearity in the *E. coli* gene expression system is not complex to the point of preventing functional composability (i.e. superposition) with standard elements used to genetically modify the organism. A model system and framework has been presented to investigate linear system behavior in *E. coli*. The proposed framework involves varying the copy number of synthetic devices to generate transfer curves. The research results obtained suggest that a transfer curve-based framework has application in the engineering of biological systems, with device nonlinearity as gauged by transfer curve response correlating with superposition experimental data.

In the model system presented in this research, superposition was shown to exist at the pSC101 copy number level with all three genetic devices present. In the absence of the *gfp* device, linear system behavior was present even with a pUC replicon. The amount of nonlinearity in the model system appears to be biased towards the *gfp* device. This was also the case in preliminary superposition experiments performed in the context of another plasmid system (Chapter 2). The bias towards *gfp* was in spite of the fact that the device had an identical promoter (and RBS sequence in Chapter 2) to other devices present. Such a finding suggests additional factors besides promoter strength impact the amount of nonlinearity in a device's steady-state RNA expression profile. The developed device transfer curve (DTC) method may find application in the systematic testing of device nonlinearity to determine whether a device will give a predictable output when used in combination with other devices. This, in turn, may enable the design and construction of biological systems with predictable, quantitative behavior from smaller elements characterized in isolation.

### Future Directions

Research ideas that extend the work presented include:

#### 1. Elucidation of the mechanism underlying the observed non-linearity in *gfp*

The copy number of synthetic devices may not be the only factor that perturbs microbial organisms. Promoter strength, RBS strength, codon usage, and product function are perhaps important factors too. Transcription and translation are coupled in *E. coli*, and there is cooperative activity between RNA polymerase and ribosomes [57-58]. As such, it would be interesting to use a range of promoters, RBS strengths, and rare codons to investigate the mechanism underlying the large observed non-linearity in *gfp*'s DTC. And while codon usage may contribute to the large observed non-linearity, it should be noted that the number of rare codons in the *nptII* and *gfp* genes are comparable (nine versus twelve, respectively). As such, it would be

meaningful to investigate the additional factors above to determine whether they manifest due to transcriptional and translational coupling.

## **2. Investigation of the impact of gene length and position**

The genes investigated in this research are all less than 1000 bp in length. It would be interesting to investigate what role (if any) gene length has on DTC non-linearity and superposition. This may be investigated by adding nonsense sequences before and/or after the coding sequence in order to vary the overall sequence length of the device. Another interesting issue to investigate involves placing a gene inside of an operon to determine whether its DTC is impacted with position.

## **3. Investigation of RNA superposition *in trans***

The genetic devices investigated in this research were not engineered to realize useful functions in the RNA domain. RNA molecules, however, can be engineered to realize interesting biological functions in the cell (e.g. antisense RNAs and riboregulators) [76-78]. It would be interesting to investigate whether RNA superposition extends into the functional domain (i.e. *in trans*). Suppose, for instance, one has a couple of RNA devices, each engineered for desired antisense and/or riboregulator behavior. Do the devices preserve their quantitative behavior when expressed concurrently, assuming that superposition is enforced in the RNA expression profile (say by using a low copy number plasmid)? Identification of conditions where superposition is present (both *in cis* and *trans*) would benefit future synthetic biology efforts aimed at realizing systems with predictable, quantitative behavior.

## **4. Mechanistic computer aided design (CAD) tools**

CAD programs are an indispensable tool to analyze and design present day electronic circuits. Recently, progress has been made towards realizing software tools that describe the components and topology of synthetic gene networks [79-80]. Even though synthetic genetic devices are embedded inside of complex microbial hosts, the results of this research suggest that devices in a synthetic system may be modeled by a lumped circuit model. That is, devices behave independent of one another unless they are engineered to interact. As such, it would be of benefit if future research efforts began to quantitatively model the mechanistic details of gene expression. This may, in turn, lead to the creation of computational models for synthetic genetic devices and a SPICE-like simulation environment for synthetic biology.

## Bibliography

1. Khosla C, Keasling JD: **Metabolic engineering for drug discovery and development.** *Nat Rev Drug Discov.* 2003, **2**:1019-1025
2. Raab RM, Tyo K, Stephanopoulos G: **Metabolic engineering.** *Adv Biochem Eng Biotechnol.* 2005, **100**:1-17
3. Nowroozi FF-B: **Construction and optimization of mevalonate pathway for production of isoprenoids in *Escherichia coli*.** *Ph.D. thesis.* University of California, Berkeley, Bioengineering Department; 2009
4. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P: *Molecular biology of the cell.* 4<sup>th</sup> ed. New York: Garland; 2002
5. Watson JD, Baker TA, Bell SP, Gann A, Levine M, Losick R: *Molecular biology of the gene.* 5<sup>th</sup> ed. Redwood City: Benjamin Cummings; 2003
6. Ptashne M, Gann A: *Genes and signals.* New York: Cold Spring Harbor Lab.; 2001
7. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, Keasling JD: **Production of the antimalarial drug precursor artemisinic acid in engineered yeast.** *Nature.* 2006, **440**:940-943
8. Kwok R: **Five hard truths for synthetic biology.** *Nature* 2010, **463**:288-290
9. Keasling JD: **Synthetic biology for synthetic chemistry.** *ACS Chem Biol* 2008, **3**:64-76
10. Andrianantoandro E, Basu S, Karig DK, Weiss R: **Synthetic biology: new engineering rules for an emerging discipline.** *Mol Syst Biol* 2006, **2**:E1-E14
11. Endy D: **Foundations for engineering biology.** *Nature* 2005, **438**:449-453
12. Carlson R: **What lies within.** *The Economist* (2010, Aug 14)
13. Carlson R: **Genesis redux.** *The Economist* (2010, May 22)
14. **Registry of Standard Biological Parts** [<http://partsregistry.org>]
15. Shetty RP, Endy D, Knight TF Jr: **Engineering BioBrick vectors from BioBrick parts.** *J Biol Eng* 2008, **2**:5

16. Ellis T, Adie T, Baldwin GS: **DNA assembly for synthetic biology: from parts to pathways and beyond.** *Integr Biol* 2011, Jan 19 [Epub ahead of print]
17. Simpson ML, Saylor GS, Fleming JT, Applegate B: **Whole-cell biocomputing.** *Trends Biotechnol.* 2001, **19**:317-323
18. Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in *Escherichia coli*.** *Nature.* 2000, **403**:339-342
19. Elowitz MB, Leibler S: **A synthetic oscillatory network of transcriptional regulators.** *Nature.* 2000, **403**:335-338
20. **Life, Reinvented** [<http://www.wired.com/wired/archive/13.01/mit.html>]
21. Sauro HM: **Modularity defined.** *Mol Syst Biol* 2008, **4**:166
22. Del Vecchio D, Ninfa AJ, Sontag ED: **Modular cell biology: retroactivity and insulation.** *Mol Syst Biol* 2008, **4**:161
23. Agapakis CM, Silver PA: **Modular electron transfer circuits for synthetic biology.** *Bioengineered Bugs* 2010, **1**:6
24. Arkin A: **Setting the standard in synthetic biology.** *Nat Biotechnol* 2008, **26**:771-774
25. Canton B, Labno A, Endy D: **Refinement and standardization of synthetic biological parts and devices.** *Nat Biotechnol* 2008, **26**:787-793
26. Benner SA, Sismour AM: **Synthetic biology.** *Nat Genetics* 2005, **6**:533-543
27. Blattner FR, Plunkett G 3<sup>rd</sup>, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277**:1453-1462
28. Lathi BP: *Linear systems and signals.* Carmichael: Berkeley-Cambridge Press; 1992
29. Gray PR, Hurst PJ, Lewis SH, Meyer RG: *Analysis and design of analog integrated circuits.* 4<sup>th</sup> ed. New York: John Wiley & Sons, Inc.; 2001
30. Ingraham JL, Neidhart FC: *Escherichia coli and Salmonella typhimurium: cellular and molecular biology.* Washington, DC: ASM Press; 1987
31. Sambrook J, Fritsch EF: *Molecular cloning: a laboratory manual.* 2<sup>nd</sup> ed. New York: Cold Spring Harbor Laboratory Press; 1989

32. Orosz A, Boros I, Venetianer P: **Analysis of the complex transcription termination region of the *Escherichia coli* *rrnB* gene.** *Eur J Biochem* 1991, **201**:653-659
33. Kües U, Stahl U: **Replication of plasmids in gram-negative bacteria.** *Microbiol Rev* 1989, **53**:491-516
34. Selzer G, Som T, Itoh T, Tomizawa J: **The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids.** *Cell* 1983, **32**:119-129
35. Smolke CD, Carrier TA, Keasling JD: **Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures.** *Appl Environ Microbiol* 2000, **66**:5399-5405
36. Pfleger BF, Fawzi NJ, Keasling JD: **Optimization of DsRed production in *Escherichia coli*: effect of ribosome binding site sequestration on translation efficiency.** *Biotechnol Bioeng* 2005, **92**:553-558
37. Lee SK, Keasling JD: **A propionate-inducible expression system for enteric bacteria.** *Appl Environ Microbiol* 2005, **71**:6856-6862
38. Lin-Chao S, Chen WT, Wong TT: **High copy number of the pUC plasmid results from a *Rom/Rop*-suppressible point mutation in RNA II.** *Mol Microbiol* 1992, **6**:3385-3393
39. Chang AC, Cohen SN: **Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid.** *J Bacteriol* 1978, **134**:1141-1156
40. Scholtissek S, Grosse F: **A cloning cartridge of lambda t(o) terminator.** *Nucleic Acids Res* 1987, **15**:3185
41. Manen D, Xia G, Caro L: **A locus involved in the regulation of replication in plasmid pSC101.** *Mol Microbiol* 1994, **11**:875-884
42. Peterson J, Phillips GJ: **New pSC101-derivative cloning vectors with elevated copy numbers.** *Plasmid* 2008, **59**:193-201
43. Lee CL, Ow DS, Oh SK: **Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria.** *J Microbiol Methods* 2006, **65**:258-267
44. Lee C, Kim J, Shin SG, Hwang S: **Absolute and relative QPCR quantification of plasmid copy number in *Escherichia coli*.** *J Biotechnol* 2006, **123**:273-280

45. Tao L, Jackson RE, Cheng Q: **Directed evolution of copy number of a broad host range plasmid for metabolic engineering.** *Metab Eng* 2005, **7**:10-17
46. Ballester M, Castelló A, Ibáñez E, Sánchez A, Folch JM: **Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals.** *Biotechniques* 2004, **37**:610-613
47. Bustin SA: **Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays.** *J Mol Endocrinol* 2000, **25**:169-193
48. Giulietti A, Overbergh L, Valckx D, Decallonne B, Bouillon R, Mathieu C: **An overview of real-time quantitative PCR: applications to quantify cytokine gene expression.** *Methods* 2001, **25**:386-401
49. Wong ML, Medrano JF: **Real-time PCR for mRNA quantification.** *Biotechniques* 2005, **39**:75-85
50. Alper H, Fischer C, Nevoigt E, Stephanopoulos G: **Tuning genetic control through promoter engineering.** *Proc Natl Acad Sci USA* 2005, **102**:12678-12683
51. Grabherr R, Bayer K: **Impact of targeted vector design on ColE1 plasmid replication.** *Trends Biotechnol* 2002, **20**:257-260
52. Weigandt TC: **Low-phase-noise, low-timing-jitter design techniques for delay cell based VCOs and frequency synthesizers.** *Ph.D. thesis.* University of California, Berkeley, Electrical Engineering Department; 1998.
53. Robbins KR, Saxton AM, Southern LL: **Estimation of nutrient requirements using broken-line regression analysis.** *J Anim Sci* 2006, **84**:E155-E165
54. Bailey JE, Ollis DF: *Biochemical engineering fundamentals.* 2<sup>nd</sup> ed. New York: McGraw-Hill; 1986
55. Vogel U, Jensen KF: **Effects of guanosine 3',5'-bisdiphosphate (ppGpp) on rate of transcription elongation in isoleucine-starved Escherichia coli.** *J Biol Chem* 1994, **269**:16236-16241
56. Krohn M, Wagner R: **Transcriptional pausing of RNA polymerase in the presence of guanosine tetraphosphate depends on the promoter and gene sequence.** *J Biol Chem* 1996, **271**:23884-23894
57. Proshkin S, Rahmouni AR, Mironov A, Nudler E: **Cooperation between translating ribosomes and RNA polymerase in transcription elongation.** *Science* 2010, **328**:504-508

58. Burmann BM, Schweimer K, Luo X, Wahl MC, Stitt BL, Gottesman ME, Rösch P: **A NusE:NusG complex links transcription and translation.** *Science* 2010, **328**:501-504
59. Mettetal JT, Muzzey K, Gómez-Urbe C, van Oudenaarden A: **The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*.** *Science* 2008, **319**:482-484
60. Neidhardt FC, Bloch PL, Smith DF: **Culture medium for enterobacteria.** *J Bacteriol* 1974, **119**:736-747
61. Lutz R, Bujard H: **Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements.** *Nucleic Acids Res* 1997, **25**:1203-1210
62. Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN: **Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays.** *Proc Natl Acad Sci USA* 2002, **99**:9697-9702
63. Pushnova EA, Geier M, Zhu YS: **An easy and accurate agarose gel assay for quantification of bacterial plasmid copy numbers.** *Anal Biochem* 2000, **284**:70-76
64. Lee C, Lee S, Shin SG, Hwang S: **Real-time PCR determination of rRNA gene copy number: absolute and relative quantification assays with *Escherichia coli*.** *Appl Microbiol Biotechnol* 2008, **78**:371-376
65. Mason G, Provero P, Vaira AM, Accotto GP: **Estimating the number of integrations in transformed plants by quantitative real-time PCR.** *BMC Biotechnol* 2002, **2**:20
66. Olivares EC, Hollis RP, Chalberg TW, Meuse L, Kay MA, Calos MP: **Site-specific genomic integration produces therapeutic Factor IX levels in mice.** *Nat Biotechnol* 2002, **20**:1124-1128
67. Voigt CA: **Genetic parts to program bacteria.** *Curr Opin Biotechnol* 2006, **17**:548-557
68. Lee TS, Krupa RA, Hajimorad M, Prasad N, Lee SK, Keasling JD: **Biobrick vectors and datasheets: a synthetic biology platform for metabolic engineering.** *Manuscript in preparation*
69. Kovach ME, Phillips RW, Elzer PH, Roop RM 2<sup>nd</sup>, Peterson KM: **pBBR1MCS: a broad-host range cloning vector.** *Biotechniques* 1994, **16**:800-802

70. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: **Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids.** *Nat Biotechnol* 2003, **21**:796-802
71. Carapuca E, Azzoni AR, Prazeres DM, Monteiro GA, Mergulhao FJ: **Time-course determination of plasmid content in eukaryotic and prokaryotic cells using real-time PCR.** *Mol Biotechnol* 2007, **37**:120-126
72. Skulj M, Okrslar V, Jalen S, Jevsevar S, Slanc P, Strukelj B, Menart V: **Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes.** *Microb Cell Fact* 2008, **7**:6
73. Friehs K: **Plasmid copy number and plasmid stability.** *Adv Biochem Eng Biotechnol* 2004, **86**:47-82
74. Antoine R, Loch C: **Isolation and molecular characterization of a novel broad-host-range plasmid from *Bordetella bronchiseptica* with sequence similarities to plasmids from gram-positive organisms.** *Mol Microbiol* 1992, **6**:1785-1799
75. Glass RE: *Gene function: E. coli and its heritable elements.* University of California Press; 1982
76. Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ: **Engineered riboregulators enable post-transcriptional control of gene expression.** *Nat Biotechnol* 2004, **22**:841-847
77. Carothers JM, Goler JA, Kapoor Y, Lara L, Keasling JD: **Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity.** *Nucleic Acids Res* 2010, **38**:2736-2747
78. Isaacs FJ, Dwyer DJ, Collins JJ: **RNA synthetic biology.** *Nat Biotechnol* 2006, **24**:545-554
79. Densmore D, Van Devender A, Johnson M, Sritanyaratana N: **A platform-based design environment for synthetic biological systems.** In *Proceedings of the 2009 Conference on Diversity in Computing, ACM*: April 2009; Portland
80. Densmore D, Kittleson JT, Bilitchenko L, Liu A, Anderson JC: **Rule based constraints for the construction of genetic devices.** In *Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS*: May 2010; Paris