Engineering Transcriptional Control and Synthetic Gene Circuits in Cell Free systems

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Sukanya Iyer
December 2012

Copyright © 2012 by Sukanya Iyer.
All rights reserved.

DEDICATION

This work is dedicated to my family.

Acknowledgements

First and foremost, I would like to extend my gratitude to my advisor, Dr. Mitchel Doktycz for his motivation and guidance that saw me through some difficult phases of my dissertation work. His constant support and freedom that he has given me to explore wild ideas has been instrumental to growth as a scientist. I would further like to thank my committee members Michael Simpson, Albrecht von Arnim, Barry Bruce and Jenny Morrell-Falvey for providing guidance to direct my research plans. I owe a large measure of this work to stimulating conversations and collaborations with David Karig,that not only resulted in development of new projects but also has contributed to my growth as a scientist. Scott Retterer deserves a special mention for his help with the microfluidics experiments and for the frequent laughs that he provided. Ana Kitazono took a great deal of interest in my growth as a scientist and whose advice on all things science served me very well. Liz Norred and Hannah Meredith were instrumental for the construction and testing of plasmids used in chapter 3.

The four years that I have spent in lab would have seemed a lot longer if the lab sisterhood consisting Ninell Mortensen, Elizabeth Vargis, Sandra Davern, Bernadeta Srijanto, Meenaa Kalyanaraman and Sarah Melton weren't around. Dave Allison has been a dear friend and I owe him many thanks. Archana Vijayakumar, Meenakshi Devarajan, Swati Misra, Chandra Ulaganathan, Sangeetha Rajagopalan, Ritin Sharma, Aarthi Sundarrajan and Bijoyita Roy have been great source of strength.

Finally, I would like to thank my parents for their love, their conviction in me, for timely advice and constant motivation that has seen through the ups and downs so far. I would like to thank my sister Sandhya Iyer, who I know is going to make us all very proud. I would like to extend my deepest gratitude Mr. and Mrs Ragunathan for extending a warm welcome into their family. Finally, I would like to thanks Balakumar Ragunathan, with whom I will soon share my address, whose care and affection means to me more than I can put in words.

ABSTRACT

Engineering gene networks offers an opportunity to harness biological function for biotechnological and biomedical applications. In contrast to cell-based systems, cell free extracts offer a flexible and well-characterized context in which to implement predictable gene circuits. Critical to these efforts is the availability of a library of ligand sensitive gene regulatory systems. Here, I describe efforts to develop molecular tools to control gene expression and implement a negative feedback circuit in *E.coli* cell extracts. First, a strategy to regulate T7 RNA polymerase using DNA aptamers is detailed. I test the hypothesis that a DNA aptamer, when placed near the transcription start site, interferes with transcription in the presence of the target molecule. A DNA aptamer that binds thrombin is used as a model system for demonstrating feasibility of the approach. I show that for the hybrid T7-aptamer promoter, thrombin addition results in up to a 5-fold reduction in gene expression. I further demonstrate that gene expression be tuned by altering the position of the aptamer relative to the transcription start site. I then devised a mechanism to engineer dual regulation of T7 promoters using LacI and TetR repressor proteins. To achieve this, a LacI binding site (lacO) was positioned 92bp upstream from a T7lacO promoter, which resulted in an increased repression from T7lacO promoters presumably by a looping based mechanism. TetR binding sites were introduced into this framework to disrupt the DNA looping to create T7 promoters that respond to both LacI and TetR. I show that positioning a tetO operator between the upstream lacO and the T7lacO promoter results in relieving lacO mediated repression by TetR. Finally, a negative feedback circuit was realized using T7lacO promoters. To this end, mono-cistronic and bi-cistronic system assembly approaches for system assembly are examined leading to the realization of an inducible negative feedback circuit in cell free systems. Collectively, the tools developed in this work pave the way for expanding the library of ligands that can be used for regulating gene expression, enabling signal integration at T7 promoters and facilitating engineering of gene networks in cell free systems.

Table of Contents

Chapter 1- Introduction **1**
Background **1**
Synthetic Gene Circuits in Cells: Opportunities, Approach and Challenges **2**
Synthetic Gene Circuits in Cell Free Systems **7**
Gene circuits in well defined buffer systems 7
Gene networks with protein intermediates 10
Challenges to implementing gene networks in cell extracts **16**
Objective **18**
Enabling transcriptional control in cell free system 20
Combinatorial promoters for signal integration in cell free systems 27
Examine modes of system assembly and implement negative feedback motifs in cell-free systems 31
Chapter Outline **35**

Chapter 2- Ligand mediated transcriptional regulation using DNA aptamers in cell free systems **37**
Introduction **37**
Results **42**
Approach for template assembly for achieving aptamer mediated gene regulation 42
Effect of template on aptamer mediated repression 44
Testing the specificity of thrombin mediated repression 46
Discussion **55**
Materials and Methods **59**
Plasmid construction 59
Single stranded DNA template preparation 60
Cell free protein synthesis experiments (CFPS) 61
Fluorescence Anisotropy 61

Chapter 3- Engineering Dual regulation of T7 promoter using LacI and TetR repressors **63**
Introduction **63**
Results **66**
Addition of auxiliary operators upstream to T7lacO promoter results increases repression levels from T7 promoters 66
Engineering dual regulation of T7promoters using LacI and TetR 70
Effect of TetR on T7lacO repression when tetO is positioned downstream from T7lacO promoter 74
Effect of TetR on T7lacO repression when tetO is overlaps with T7lacO promoter 74
TetR bound to tetO placed in between the Lac operators effectively alleviates Lac mediated repression 75
Testing dual input promoters in cell free systems 78
Discussion **80**
Materials and Methods **87**
Plasmids and Bacterial strains 87
Purification of TetR 88
GFP measurements from E. coli experiments 88
Cell-free expression experiments 89

Chapter 4 -Implementing Negative Feedback in Cell free systems using T7lacO promoters **90**
Introduction **90**
Results **94**
System composition 94
Repression of T7lacO promoters by lacI encoded on different plasmids 99
Negative feedback from T7 lacO promoters 103
Discussion **106**
Materials and Methods **112**
Plasmids and strains 112
Cell-free expression experiments 112

Bibliography **120**

Appendices **130**
Appendix A- Methods for construction of double stranded DNA templates with ssDNA regions **131**
Appendix B- Translational Regulation of gene expression using DNA aptamers **135**
Appendix C-Towards continuous exchange cell free protein synthesis reactions in microfluidic devices **143**

VITA **151**

List of figures

FIGURE 1.1 - APPROACH FOR ASSEMBLING SYNTHETIC GENE CIRCUITS 4
FIGURE 1.2- EXAMPLES OF TRANSCRIPTION BASED CELL FREE CIRCUITS 9
FIGURE 1.3 - EXAMPLES OF SYNTHETIC GENE CIRCUITS IMPLEMENTED IN CELL EXTRACTS 13
FIGURE 1.4 - SCHEMATIC SHOWING A HYPOTHETICAL CIRCUIT THAT MIGHT BE ASSEMBLED FROM THE TOOLS DEVELOPED IN THIS WORK. 19
FIGURE 1.5 - SCHEMATIC OF SELEX 25
FIGURE 1.6 CIS REGULATORY SCHEMES FOR IMPLEMENTING LOGIC GATES WITH BACTERIAL PROMOTERS 30
FIGURE 1.7 COMMONLY OCCURRING NETWORK MOTIFS IN BIOLOGICAL CELLS 32
FIGURE 2.1 HYPOTHESIS FOR APTAMER MEDIATED TRANSCRIPTIONAL REGULATION 41
FIGURE 2.2. SCHEMATIC OF STEPS FOR TEMPLATE ASSEMBLY ASSEMBLING 1) BUBBLE DNA TEMPLATES 2) SSDNA WITH DS PROMOTER. 43
FIGURE 2.3 FLUORESCENCE ANISOTROPY MEASUREMENTS FROM BUBBLE TEMPLATES (TOP PANEL) AND SSDNA TEMPLATES(BOTTOM PANEL). THE THROMBIN CONCENTRATIONS BEFORE THE LOGARITHMIC TRANSFORMATION WERE IN NM. 44
FIGURE 2.4 EFFECT OF TEMPLATE ON APTAMER MEDIATED REPRESSION 47
FIGURE 2.5 TIME COURSE FOR RESPONSE OF SSDNA ATKS CONSTRUCT TO THE ADDITION OF 1.8μM THROMBIN. ... 48
FIGURE 2.6 TESTING SPECIFICITY OF APTAMER MEDIATED TRANSCRIPTIONAL REGULATION 50
FIGURE 2.7 DOSE RESPONSE CURVES FOR +2 AND +9 APTAMER CONSTRUCTS 53
FIGURE 2.8 EFFECT OF PLACEMENT OF DIMERIC APTAMERS DOWNSTREAM TO THE TRANSCRIPTIONAL START SITE.... 54
FIGURE 3.1 DESIGN STRATEGY FOR ACHIEVING COMBINATORIAL REGULATION OF EXPRESSION FROM T7PROMOTERS 67
FIGURE 3.2 EFFECT OF PRESENCE OF ADDITIONAL LACI GENE IN THE PLASMID CONTAINING T7LACO PROMOTERS. ... 68
FIGURE 3.3 EFFECT OF AUXILIARY OPERATORS ON LACI MEDIATED REPRESSION OF T7LACO PROMOTERS 69
FIGURE 3.5 EFFECT OF TETO PLACED DOWNSTREAM FROM T7LACO PROMOTERS. 72
FIGURE 3.6 EFFECT OF TETR ON T7LACO REPRESSION WHEN TETO OVERLAPS T7LACO PROMOTERS. 73
FIGURE 3.7 –EFFECT OF TETO ON LACI MEDIATED REPRESSION OF T7LACO WHEN TETO WAS IN BETWEEN THE TWO LAC OPERATORS. 77
FIGURE 3.8 EFFECT OF TETO ON LACI DEPENDENT REPRESSION OF T7LACO WHEN PLACED BETWEEN TWO LAC OPERATORS IN CELL EXTRACTS 79
FIGURE 3.8 TIME COURSE OF EXPRESSION FROM INTERRUPTED T7PROMOTERS. 85
FIGURE 4.1 APPROACH FOR ASSEMBLING SYNTHETIC GENE CIRCUITS IN CELL FREE SYSTEMS. 93
FIGURE 4.2 PLASMIDS AND RESULTS FOR EXPLORING DIFFERENT SYSTEM COMPOSITION APPROACHES. 96
FIGURE 4.3 COMPARISON OF BICISTRONIC AND TWO-PLASMID SYSTEMS. 98
FIGURE 4.4 REPRESSION OF T7LACO BY LACI ENCODED ON A SEPARATE PLASMIDS 101
FIGURE 4.5 EFFECT OF RATIOS OF PT7LACOGFP AND PT7LACI ON EXPRESSION FROM T7LACO PROMOTERS. PT7LACOGFP IN THE INDICATED RATIOS 30 MINUTES AFTER THE ADDITION OF PT7LACI. 102
FIGURE 4.6 DEMONSTRATION OF NEGATIVE FEEDBACK FROM PLACOT7LACOLACIGFP AND PT7LACOLACIGFP. .. 105
FIGURE A1- SCHEMATIC SHOWING LIGATION BASED STRATEGY TO GENERATE TEMPLATES FOR TESTING APTAMER-MEDIATED REGULATION. 132
FIGURE A2 SCHEMATIC FOR APTAMER TEMPLATE ASSEMBLY FROM SSDNA TEMPLATES GENERATED USING STREPTAVIDIN-BIOTIN PURIFICATION 132
FIGURE B2 GRAPH SHOWS RESPONSE FROM PKSGFP PLASMIDS TO LONG AND SHORT VERSIONS OF APTAMER CONTAINING REPRESSOR OLIGONUCLEOTIDES. 139
FIGURE B3 – GRAPH SHOWS CHANGES IN EXPRESSION UPON ADDITION OF REPRESSOR WITH (ANTIL1 AND ANTIL2) AND WITHOUT (ANTIL) THROMBIN APTAMER 141
FIGURE C1- SCHEMATIC OF AN APPLICATION FOR PROPOSED CELL MIMIC DEVICES IN BIOSENSING AND ACTUALTION 145
FIGURE C2- GRAPH SHOWING CHANGE IN FLUORESCENCE AS A FUNCTION OF TIME IN CELL MIMIC DEVICES- DNA ENCODING GFP FROM A T7 PROMOTER WAS INTRODUCED INTO THE DEVICE ALONG WITH *E.COLI* CELL EXTRACT. CHANGE IN FLUORESCENCE WAS MONITORED USING EPIFLUORESCENCE MICROSCOPE. 145
FIGURE C3- NEW MICROFLUIDIC DEVICES WITH EXTENDED CHANNELS FOR ACHIEVING PROTEIN SYNTHESIS ON A MICROFLUIDIC PLATFORM 148

FIGURE C4 COMPARING DIFFUSION OF FLUORESCEIN ACROSS A NANOPOROUS MEMBRANE COATED WITH 4 MINUTES, 6 MINUTES AND 7 MINUTES PECVD. 148

List of tables

TABLE 2.1 –LIST OF PLASMIDS USED IN THIS STUDY. 49
TABLE 2.2 – TABLE DENOTING THE BASAL LEVEL EXPRESSION RELATIVE TO A SSDNA TEMPLATE LACKING AN APTAMER DOWNSTREAM FROM TRANSCRIPTIONAL START SITE. FOLD CHANGE COLUMN INDICATES THE CHANGE IN EXPRESSION FROM THESE TEMPLATES IN RESPONSE TO 1.8 μM THROMBIN. 51
TABLE 2.3 SUMMARY OF FOLD CHANGES IN GENE EXPRESSION OBTAINED FROM SYNTHETIC LIGAND DEPENDENT GENE REGULATION STRATEGIES IN CELL FREE SYSTEMS. 59
TABLE 3.1 LIST OF PLASMIDS USED IN THIS STUDY- THE SUBSCRIPTS NEXT TO THE T7 PROMOTER INDICTES THE NUMBER OF BASES THAT WERE REMOVED FROM T7 PROMOTER 71
TABLE 4.1- MONOCISTRONIC PLASMIDS USED IN THIS STUDY. 94
TABLE 4.2 LIST OF BI-CISTRONIC PLASMIDS USED IN THIS CHAPTER. 95
TABLE B1 LIST OF OLIGONUCLEOTIDES USED FOR TRANSLATIONAL REGULATION IN THIS STUDY 138

Chapter 1- Introduction

Background

A synthetic gene circuit involves "novel regulation of pre-existing or novel cellular function"[1]. This discipline falls under the field of synthetic biology, which encompasses efforts to forward engineer biological function. The discipline entails construction of modules from well characterized components or redesigning an existing module to examine a system in isolation from the rest of the biological circuitry[2]. In contrast to the traditional genetic and biochemistry based approaches, synthetic biology takes an engineering approach to biology- which is to rationally design minimal systems from well characterized parts to test its effect on function[3]. Furthermore, engineered biological cells are being programmed to realize functions that have been traditionally realized from hard materials. The confluence of methods from engineering, biology and mathematics has propelled the development of systems that achieve precision associated with synthetic inorganic materials while utilizing biological systems.

The often-cited inaugural devices namely the toggle switch[4] and the repressilator [5] that showed bi-stability and oscillations paved the way for a construction and implementation of synthetic gene circuits and modules that alter biological function. Since then remarkable improvements in the available parts and development of methodologies to construct circuits of predetermined function have been used to construct "toy systems" to test biological hypotheses and for the construction of application driven systems. Engineered biological systems in contrast to engineered silicon based circuits offer the distinct advantage of being interfaced with the living

world and the ability to link computation and signal output with biologically meaningful responses.

A majority of these synthetic gene circuits are implemented in living cells. These studies rely on the rich diversity of molecular gene regulatory components to assemble gene networks. In contrast to enabling circuits in cells, cell free systems offer a flexible platform in which to enable biological function. However, efforts to realize engineered function in cell free systems have lagged behind those in cell-based systems largely due to the lack of molecular tools to regulate gene expression. This thesis details molecular tools and approaches to realize engineered function in cell free systems. In this introductory chapter, first an overview of the current state of progress in enabling synthetic gene circuits in cells is discussed. This is followed by a discussion of a few examples of *in vitro* gene networks. And finally, a primer for strategies adopted for enabling transcriptional control and network motifs that have been implemented in chapters 2, 3 and 4 is provided.

Synthetic Gene Circuits in Cells: Opportunities, Approach and Challenges

The inherent value of a forward engineering approach to biological investigation lies in the fact that construction of rationally designed genetic circuitry aids in identifying gaps in and assesses the completeness of our understanding of biological systems[6]. Such a synthetic approach has been employed to probe different aspects of cellular function. For instance, Cox *et.al* [7] and Guet *et.al* [8] utilized combinatorial promoter libraries constructed using well-defined parts to correlate promoter architecture with transcriptional regulation. Likewise, synthetic gene circuits have been invaluable for

evaluating and validating theoretical predictions about functional consequences of recurring network motifs[9, 10]. For example, whereas negative feedback circuits accelerate transcriptional response[11] and are resistant to noise[12], positive autoregulation promotes bi-stability in a system[13].

Lessons from synthetic semiconductor based systems have thus far guided the progress made in the field. Given the engineering origins of the field, standards that apply to engineered synthetic devices such as abstraction, modularity, predictability and extensibility are being sought in engineered biological circuits [14]. Parts for constructing synthetic circuits such as transcriptional factors, promoters and other determinants are often derived from well-studied model biological systems. These parts, which have evolved to function within natural biological networks, are adapted for use in synthetic gene circuits. To aid the task of assembling engineered circuits, well-characterized standardized components that function in a variety of cellular contexts have been assembled[14, 15]. These contain well-characterized genetic components such as native and synthetic ligand sensitive promoters, of bacterial and eukaryotic origins, ribosome binding sites and transcriptional terminators that function in predictable manner[16, 17]. Analogous to the manner in which parts such as capacitors and resistors are wired together into circuits that process signals in a desired format, synthetic biologists aim to utilize potentially modular genetic components (that function in a variety of cells), and assemble these components into smaller modules [14] (Figure 1.1).

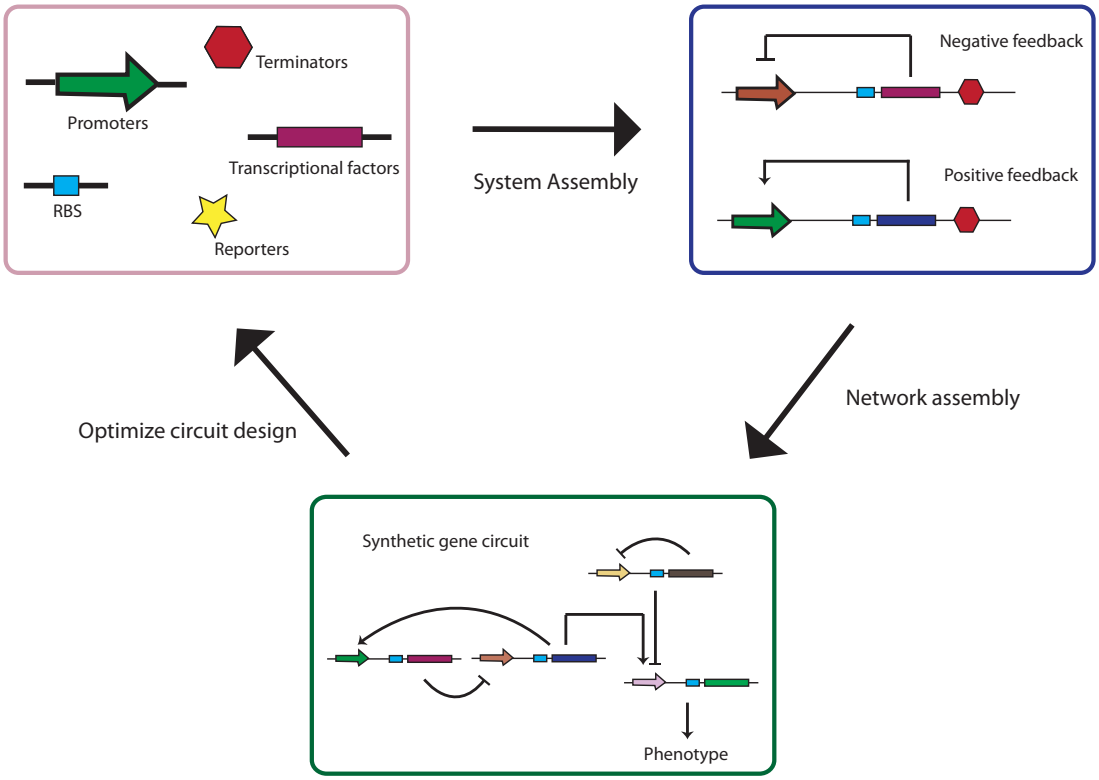


Figure 1.1 - Approach for assembling synthetic gene circuits - Synthetic gene circuits are typically assembled from a library of well-characterized genetic parts. These genetic parts are then assembled into small modules called network motifs, larger sophisticated networks can then be constructed from these modules that can be linked with a biological function. Often several iterations in the assembly process is required before a network that displays a desired phenotype can be assembled

Different strategies have been used for assembling synthetic gene circuits[18, 19]. Iterative rational design, which involves computational modeling of the performance of a gene circuit, construction and evaluation of the circuit and refining the original circuit has yielded synthetic enhancers, and oscillators [20, 21] that are more robust than their original counterparts. Yet another strategy involves constructing circuits comprising different genetic components arranged in a variety of configurations and selecting the circuit variant displaying the desired function. While these two strategies constitute the rational design approach, directed evolution methods are being applied to fine tune DNA, RNA and protein components for use in synthetic gene circuits[1, 22].

The majority of these circuits are embedded within a larger cellular context. The advantages of implementing gene expression in cell-based systems are manifold. Biological cells contain a large repository of ligand sensitive transcription and translation mechanisms and machinery for tuning protein and RNA synthesis and degradation. Consequently, implementation of gene circuits in cells provides access to existing genetic parts in the cell and harnesses the ability of cells to self-sustaining[23]. Cell division provides a convenient method of diluting the built up components. Furthermore, cell based system provides a more biologically relevant context for elucidating network function.

However, exclusive reliance on existing strategies for design and testing biological circuits will ultimately yield systems that remain systems of limited complexity and functionality[18]. The synthetic gene circuits implemented in these systems are extraneous to cell survival and therefore impose fitness effects upon the cells. The evolutionary forces that in nature optimize cellular architecture, function to

deactivate the synthetic system that imposes a heavy burden on cellular machinery. In addition the assembly of biological circuits in a predictable manner remains a non-trivial process because of the lack of predictive power stemming from the use of incompletely understood biological parts[3]. Furthermore, unintended interactions between different components further complicate implementation of reliable computation units. As mentioned earlier, synthetic circuits implemented in cells make use of cellular machinery to achieve a balancing of rate of synthesis and degradation of cellular components. In the case of cell based synthetic gene circuits, the user does not have control over several cellular parameters that play a role in the successful implementation of synthetic gene circuit. These factors include component dilution achieved by the means of cellular replication, relative component concentrations, effects of spatial organization, cellular size and endogenous synthesis and degradation machinery play a role in achieving balance of different RNA and protein components required for proper functioning of a circuit. Incomplete understanding of cellular processes and components further complicate the assembly of predictable systems circuits *in vivo*.

Several approaches are being sought to make biological systems more tractable for implementing predictable engineered systems. Several studies have outlined insulated promoter systems, used orthogonal viral polymerases[24, 25] and orthogonal translation machinery to decouple synthetic systems from endogenous cellular processes[26]. One approach to addressing the task of eliminating or reducing the system complexity is to understand minimal cells[27, 28]. These minimal cells comprise a small genome and contain a fairly limited set of components and yet are self replicating, signal processing mini bioreactors that perform computations that outperform synthetic systems in terms of

the functional density. Not surprisingly then, minimal systems have emerged as the focal point for performing whole cell computation[29], synthesis of a synthetic genome[30] and construction of a bacterial cell whose genome was chemically synthesized[31]. These efforts feed into the goal of uncovering hitherto unknown interactions and design principles that contribute towards cellular functionality.

A complementary approach to constructing engineered gene networks is to reconstitute systems *in vitro*. In contrast to *in vivo* systems, *in vitro* systems that comprise either purified protein components or crude cell extracts for driving gene expression offer a stripped-down "chassis" to realize synthetic gene networks. It is possible to precisely control parts, components and proteins that make up the system. Yet, lack of molecular tools greatly limits the utility of these systems to realize sophisticated engineered functions. The focus of this dissertation is to develop transcriptional tools and employ these tools for enabling simple gene circuits in an *E. coli* cell extract.

In the following section, I describe *in vitro* examples of engineered systems that perform a predetermined function. Herein, I limit the discussion to transcriptional and translational circuits. I then highlight the tools that would be needed to expand the applicability of cell free systems for implementing gene circuits.

Synthetic Gene Circuits in Cell Free Systems

Gene circuits in well defined buffer systems

Reconstituted biochemical systems have long served as a prelude to *in vivo* biochemical investigations. Versatility and the ease of precisely controlling the chemical environment of *in vitro* systems are amenable to precise control and quantitative analysis

of the system under investigation. Since reconstituted *in vitro* systems are devoid of cellular material, they hold great value in defining components that are necessary and sufficient for explaining observed phenomenon. For instance, a reconstituted system that comprises three proteins involved in circadian rhythms –KaiA, KaiB and KaiC proteins purified from cyanobacteria has been shown to exhibit oscillatory phosphorylation behavior in a buffer system with a periodicity of approximately 24 hours in the absence of any other component[32]. Consequently, implementation of simple nucleic acid analogues of natural circuits in well-defined buffer systems provides a great starting point for providing a more reliable quantitative description of circuit behavior. Most prominent examples of *in vitro* transcription based regulatory modules have come from transcriptional circuits built from nucleic acid components[33, 34]. In keeping with the goal of engineering reliable networks, nucleic acid based systems that utilize a small number of proteins that drive transcription and RNA degradation to realize surprisingly sophisticated functionality such as bistability and oscillations have been developed. In both of these nucleic acid based systems, DNA complementation of a promoter sequence is used to mediate transcriptional activation or repression. A DNA oligo that "completes" a T7 promoter was designated as the activator, whereas an RNA oligo that sequesters the activator DNA oligo repressed transcription (Figure 1.2 A). A negative feedback loop between an activator circuit and a repressor circuit resulted in oscillations for about 10 hours[34].

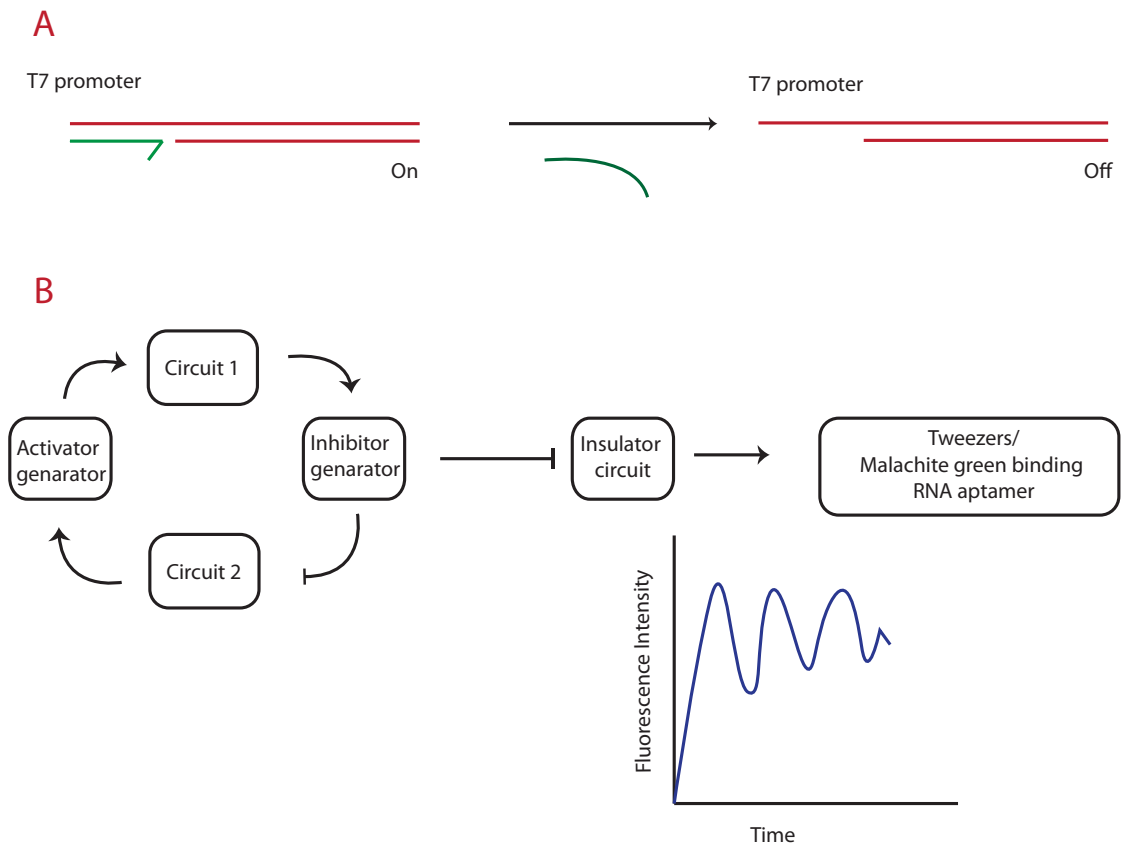


Figure 1.2- Examples of transcription based cell free circuits- A) shows the principle for oligonucleotide mediated regulation of T7 promoter – the red sections of the DNA template represents the non –template strand of T7 promoters. The green oligonucleotide corresponds to a portion of the template strand of T7 promoter, While the green oligonucleotide is hybridized to the promoter region, the template can function as a substrate for transcription, Removal of the template strand results in repression of transcription from the template. B) The transcription scheme described is was utilized for the construction of a transcriptional oscillator consisting a mutually activating and repressing circuit that generates an oligonucleotide that forms a complete promoter and a repressor circuit that generates an RNA oligonucleotide that sequesters the short oligonucleotide away from the template. The insulator circuit was included in the design that siphons partially digested oligonucleotides. Oscillator circuit was linked to DNA based tweezers or an aptamer that binds malachite green and oscillations were monitored by changes in fluorescence intensity.

In a remarkable achievement, in their subsequent work, nucleic acid based oscillators were utilized to drive other nucleic acid based dependent functions namely a nanomechanical DNA tweezer and release of an aptamer that binds a fluorescent dye thereby coupling a timing device with an *in vitro* DNA based function[35] (Figure1.2 B). These systems were able to provide a semi-quantitative model to account for the experimental observations.

Even with these simplified systems, the authors encountered several issues that will prove instructive for future implementation of predictable gene circuits. Build up of short incompletely degraded RNA and DNA fragments was found to interfere with the proper functioning of the circuit. The authors had to build an insulator circuit to eliminate unwanted waste products from interfering with the output circuit. Furthermore, the circuits were found to be sensitive to batch-to-batch variations in the enzymatic machinery, which interfered with the quantitative prediction of circuit behavior. These efforts involve assembling all the components required to enable a dynamic system and therefore represent a bottom up approach to understanding the functioning of a network. Further, they underscore the power of *in vitro* technologies to implement sophisticated functions using relatively simple components and highlight the challenges that would be needed to overcome for realizing larger and diverse systems.

Gene networks with protein intermediates

a) Engineered gene circuits PURE reconstituted systems- *In vitro* biological systems provide a good compromise between using biological material and implementing synthetic protein generating systems without regard for effect on cell viability. In the

following section, I describe features of cell free systems followed by a discussion of the progress made and lessons learnt from implementing synthetic gene circuits in these systems.

PURE (reconstituted translation systems built from bacterial components) offers a great step forward in enabling predictable gene networks in precisely defined chemical environments. PURE systems offer a simplified and precisely defined medium in which to implement these systems[36]. The commercially available reconstituted system comprises T7 RNA polymerase and purified *E.coli* translation proteins. This system has been utilized for synthesis and assembly of protein macromolecular complexes as well as for demonstration of simple regulatory motifs. Asahara and Chong demonstrated the reconstruction of bacterial holoenzyme in a cell free system. The authors showed that at least five different genes encoding the different subunits of *E.coli* RNA polymerase could be simultaneously expressed in these systems to form a functional *E.coli* polymerase holoenzyme [37]. PURE systems have also been used for regulatory motifs that are present in biological system. Karig *et.al* demonstrated that a negative feedback motif could be implemented in these reconstituted translation systems and in traditional cell extracts [38] from T7tetO promoters that drove the expression of a transcriptional fusion of *tetR* and *GFP* genes. Unfortunately, the prohibitively high cost of these reconstituted systems prevents their extensive use towards the implementation of circuits in these systems. Although expensive, these expression systems offer an exciting minimal platform for future implementation of predictable systems.

b) Engineered gene circuits in cell extracts – The development of cell extracts for early biochemical investigations was based on the fact that cellular protein synthesis machinery is functional in the absence of cellular structural material [39]. The ease of carrying out complementation assays combined with the ability to define protein components in environments that simulate cellular conditions long fueled biochemical investigations in cell free systems.

In addition, the capacity of these extracts to achieve high yield protein synthesis has been accomplished. Since their inception in the 1950s, the fundamental technology for generating cell extracts for protein synthesis has not undergone a major change. Extract that furnishes the protein synthesis machinery is supplemented with rNTPs, amino acids, energy sources and stabilizing agents that provide the substrate for protein synthesis[40]. However, in recent years a push in the field of functional genomics has spurred an interest in maximizing yield of protein synthesis in cell free systems mainly by the way of changes in different components of cell free protein synthesis reactions. Extracts derived from different sources have been utilized for producing extracts for high yield protein synthesis and serve as a vehicle for driving biological investigations[41, 42]. Specific components of a cell extract such as composition, bioenergetic components[43, 44] and reaction configuration[45, 46] can be customized for desired applications.

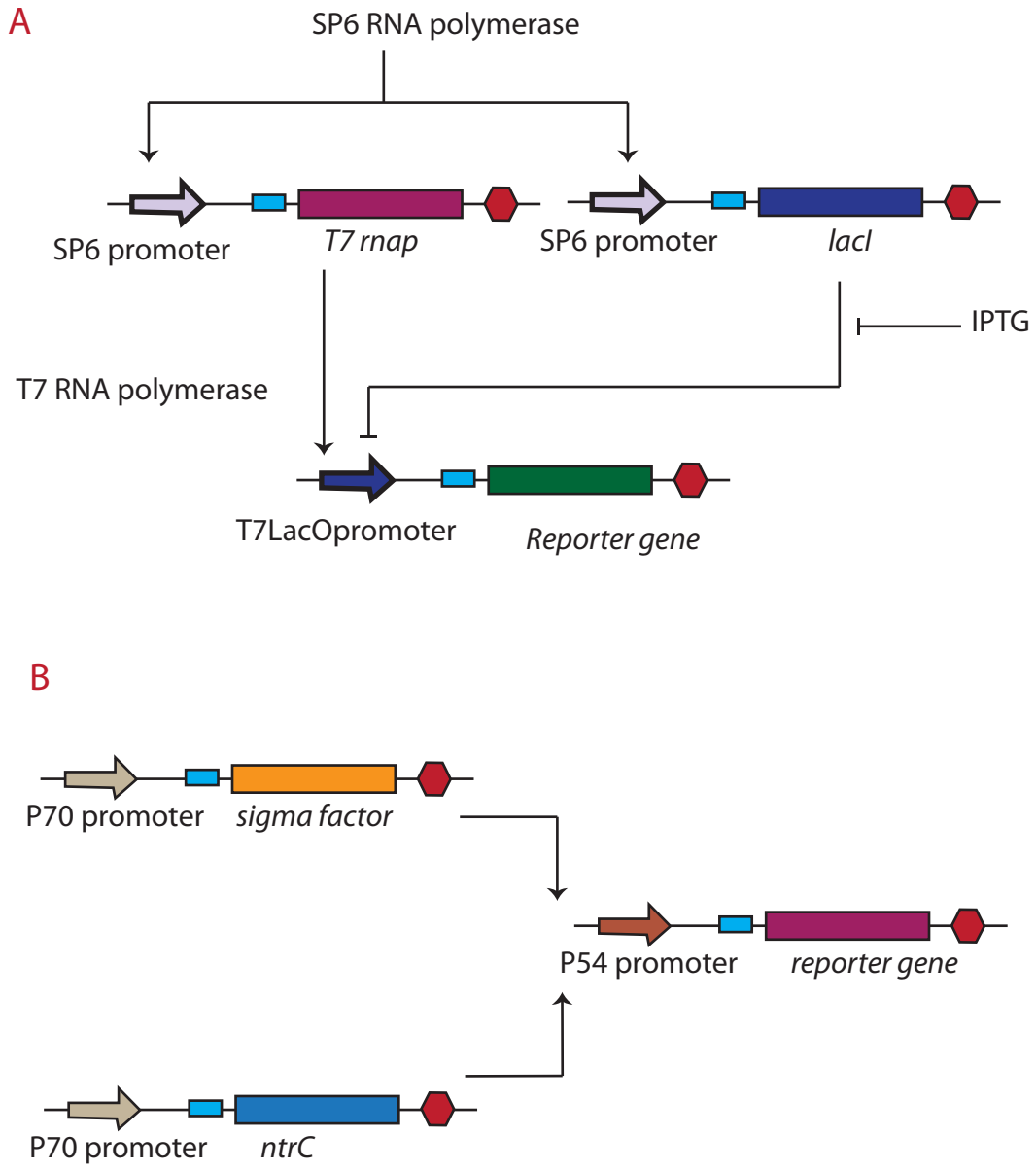


Figure 1.3 - Examples of synthetic gene circuits implemented in cell extracts - A) shows a schematic for cascaded repressor circuit developed by Noireaux et.al . SP6 RNA polymerase was exogenously added to the cell extract, which drives the expression of T7 RNA polymerase and lac repressor from a SP6 promoter. T7 Polymerase in turn drives the expression of the reporter gene from T7lacO promoter which is repressed by lac repressor. Addition of IPTG relieves the repression from T7lacO promoter thereby leading to an induction of expression from T7lacO promoter B) is an example of AND gate using *E.coli* promoters. Wherein the E.coli promoter P70 which drives the expression of a sigma factor and ntrC both of which are necessary to activate expression from P54 promoter, which drives the expression of a reporter gene.

Alongside these developments, impressive progress has been made in harnessing the protein synthetic capabilities of cell free in vitro systems for enabling synthetic gene networks [47]. Noireaux and colleagues put forth one of the earliest efforts to encode novel functions in cell free systems[48]. In their pioneering study, they described efforts to implement cascading gene networks wherein the authors employed a combination of two viral polymerases and LacI to realize a cascading network[48]. The authors made a modification to the wheat germ extract that allows for long term expression of proteins (6 hours). This modification proved critical to the success of the experiment that involved the expression of T7 RNA polymerase from a SP6 promoter. T7 polymerase then transcribed both a GFP gene and the lac repressor gene. Expression of the GFP gene was, in turn, driven by T7lacO promoters. The authors balanced the DNA template concentrations to yield a system in which expression levels were appropriately balanced to realize a cascading network in the SP6 polymerase expressed T7 RNA polymerase that in turn transcribed the GFP gene. (Figure 1.3 A) Furthermore, mRNA turnover rate was found to be a rate-limiting step for achieving cascaded expression from the three components as saturation of the translational machinery with the products from the first step of the cascade proved detrimental to expression from the subsequent stages [48]. Subsequent studies by Ishikawa *et.al* on cascading networks reported a shortened length of time delay in the onset of expression. Their studies utilized endogenous polymerase to drive the expression of phage polymerase [49]. This resulted in the induction of expression of the reporter gene much more rapidly in the analogous system implemented earlier, because lower mRNA levels generated by endogenous polymerases do not saturate the expression machinery.

This conclusion is further supported by subsequent studies in which expression cascades were implemented exclusively using a host of *E.coli* promoters and a molecular sink for built up proteins in the form of directed protein machinery. Shin *et.al* first developed a bacterial cell extract optimized for the utilization of endogenous polymerases and ClpXP protease to achieve targeted degradation of *ssrA* tagged proteins[50, 51]. Shin *et.al* utilized these extracts to show that complex signal processing networks could be achieved using a broad range of endogenous transcriptional regulation mechanisms and targeted protein degradation[52] (Figure 1.3 B).

Yet another approach to enabling a gene circuit is to tune genetic determinants for gene expression so as to prevent saturation of translation machinery. Accordingly, Karig *et.al* showed that negative feedback motifs could be realized using T7tetO promoters in bacterial cell extracts by modulating other genetic determinants of expression such as altering the tetO operator position relative to transcriptional start site, ribosome binding sites and transcriptional terminators [38].

In addition to encoding function in gene circuits, compartmentalization and temporal separation of processes is key to enabling complex cell like behavior. Use of physical platforms to achieve spatial separation of expression in *in vitro* systems and investigate effects of diffusion on expression dynamics has been demonstrated. For instance, Isalan *et.al* attempted to mimic gap patterning observed in drosophila embryos using a wheat germ cell free system. In this study, DNA was immobilized at specific locations in the reaction chamber and a protein expression gradient was established based on the diffusion of transcriptional activator across the chamber. The study provided insights into some of applicability and limitations of using simple diffusion based models

to explain a complex patterning phenotype in eukaryotes[53]. In addition, spatial patterning of DNA molecules on a chip has been utilized for realizing a transcriptional cascade akin to those implemented in bulk solution[54]. In addition, physical platforms can also enable prolonged expression in cell free systems[55-58] and examine the effects of confinement on expression dynamics[57].

Taken together, these studies represent significant achievement in demonstrating that by tuning parameters of gene expression, a desired network that performs a predetermined function can be implemented.

Challenges to implementing gene networks in cell extracts

Transcriptional regulation – The ability to implement complex function in cell free systems will be greatly enhanced by the availability of modular transcriptional regulation mechanisms that respond to a ligand of choice and achieve signal integration at the transcriptional level. As mentioned earlier, T7 promoters are widely used in cell free systems because of their high processivity and specificity. However, absence of transcriptional regulation strategies with T7 promoters has resulted in cascading networks and negative feedback circuits of limited complexity. Therefore, mechanisms for achieving transcriptional activation and repression akin to those available with bacterial and eukaryotic systems would greatly extend their applicability to cells and cell free systems.

Tuning expression dynamics -As mentioned earlier, cell extracts have traditionally been optimized for achieving protein synthesis, but now are being considered for implementing gene circuits. A common requirement for commercial kits and cell extracts for implementing dynamic behavior is the attainment of high yield protein production.

The goal of commercial cell free protein synthesis extracts is on enabling cost effective mechanisms that can maximize protein production. On the other hand, efficient high yield protein synthesis systems are critical to the success of a network as introduction of several stages of gene expression in cell free systems places a severe burden on expression machinery. However, modifications in commercial cell extracts which include inactivation of RNA and protein degradation mechanisms present a barrier to construction of dynamic gene circuits that require careful balancing of protein and RNA products obtained from different reaction stages. Therefore, cell extracts that have active mechanisms for tuning RNA and protein levels but all the while ensuring high protein productivity would be useful for construction of synthetic gene circuits[59]. In addition, efforts to provide quantitative description of synthetic circuits remain difficult because the cell extracts remain poorly defined and are not robust. Therefore, availability of cost effective cell free protein synthesis systems that contain precisely defined protein and small molecule components would be extremely valuable for enabling robust gene circuits that can be effectively modeled.

To summarize, I hope the examples show the reader that the unprecedented flexibility in terms of the physical components used together with technologies to achieve spatial separation of processes will be incredibly useful for testing biological hypothesis and for creation of metabolic networks for production of industrially relevant products. Therefore, efforts to enable and study complex biochemical reactions would benefit greatly from the following tools-

1) Availability of simple and modular ligand sensitive mechanisms for regulating RNA and protein production levels,

2) Molecular tools to tune RNA and protein amounts in cell free systems
3) A protein synthesis "Chassis" that is comprised of precisely defined components
4) Physical platform that provides a sink for removal of built up RNA and protein products and approaches for maximization of reaction lifetimes

Here, I focus on developing tools that would greatly expand the utility of cell free systems in implementing synthetic gene circuits. As mentioned earlier gene circuits often possess a sensory domain that perceives external signals and an actuator circuit that performs signal integration and mobilizes an appropriate response. Therefore to expand the utility of cell extracts for implementing gene circuits, I set out to develop tools for signal sensing and signal integration. In addition, parameters for assembling gene circuits in a cell free environment that lacks active RNA and protein degradation machinery were examined. Figure 1.4 shows a hypothetical synthetic gene network that might be assembled from tools developed in this dissertation

Objective

The objective of this dissertation is to develop transcriptional switches and gene circuits in cell free systems. In the following sections I describe the progress made in development of molecular tools for following applications:

1) Develop transcriptional control in a cell free system
2) Enable transcriptional signal integration and logic control in cell free systems
3) Examine parameters for system assembly and implement negative feedback systems in cell free systems.

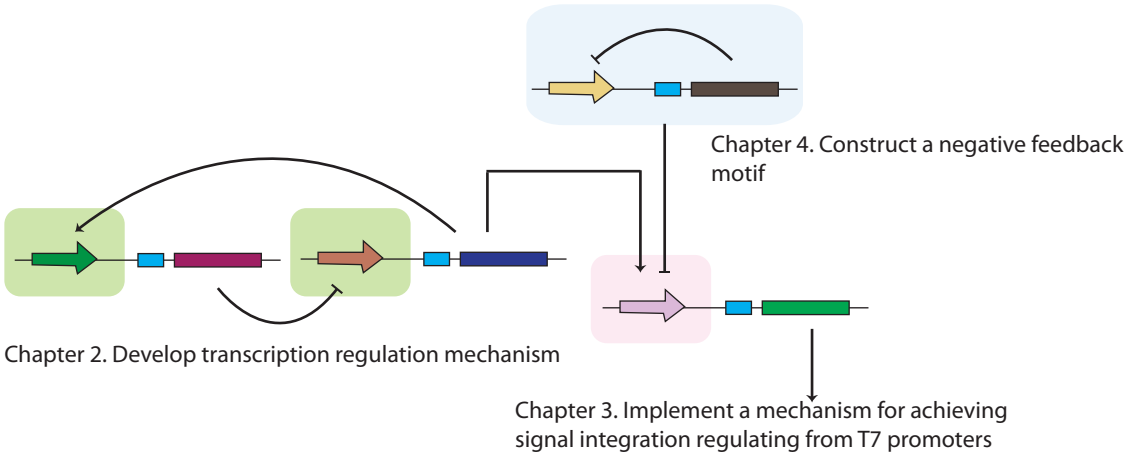


Figure 1.4 - Schematic showing a hypothetical circuit that might be assembled from the tools developed in this work.- Chapter 2 describes a method to achieve ligand dependent transcriptional regulation. Chapter 3 details a strategy to implement a logic gate that is based on T7 promoters. Finally, A negative feedback motif assembled from T7lacO promoter is described in chapter 4.

Enabling transcriptional control in cell free system

Living cells harbor several ligand sensitive gene regulation mechanisms that enable them to recognize environmental and intracellular stimuli and cause the cells to mount an appropriate response. Ligand dependent gene regulation allows the cell to mobilize cellular resources for the expression of gene products at the time of need thereby preventing wasteful expenditure of energy and resources. Biological macromolecules bearing specificity for a type of signal is often coupled to a specific biological pathway that mediates cellular response to the stimuli.

In addition to their role in biological systems, ligand sensitive gene regulation mechanisms are important tools in basic research and in biotechnological applications. In contrast to genetic changes in the system, ligand dependent mechanisms can effect gene expression changes in a spatial and temporal manner[60]. Furthermore, dose dependent responses allow the investigation of effects of intermediate responses of cells in contrast to all or none responses obtained with genetic mutations in the system. However, naturally occurring ligand dependent gene regulation strategies are geared towards molecules and modulate specific cellular processes that are important for cell survival but may not be relevant to biotechnological applications. Therefore, development of modular tools that regulate specific aspects of cellular molecular machinery in response to user defined signal molecules hold the key to harnessing and redesigning biological systems for biotechnological applications[61].

Regulation at different levels of gene expression enables cells to process the information and respond at different rates thereby optimizing response to signals from the

environment. For reasons of economy, transcriptional regulation is a key step in regulating gene expression in bacterial as well as eukaryotic systems[62]. As with biological systems, several efforts to regulate gene expression have centered on modulating transcription in response to user-defined signals. While transcriptional response to an external stimulus is slower than a translational or post-translational response, it offers distinct advantages over other gene regulation mechanisms. Firstly, since transcription is the first level of expression, regulation of transcription achieves a large dynamic range of expression and enables broad regulation of a variety of gene expression targets[61]. Secondly, signal amplification can be achieved, whereby a few signal molecules engender a large change in gene expression[3]. Finally, concatenation of operator subunits enables facile signal integration at the transcriptional level. Therefore, in our quest for building gene networks, here I focus on engineering promoters to respond to ligands that can in turn be used in gene networks[7, 63, 64].

T7 RNA polymerases are extensively used in cell free and cellular environments for achieving high yield protein synthesis[65, 66]. In contrast to the multi-subunit bacterial and eukaryotic polymerases, monosubunit T7 RNA polymerase does not require any co-factors to initiate expression from a specific 17 base pair promoter[67]. While the simplicity of expression is an asset for high yield protein synthesis, availability of ways to achieve transcriptional regulation of T7 RNA polymerase is critical to the ability to create complex gene circuits[18]. As a substrate for engineering transcriptional control, T7 polymerase mediated systems are ideal for rational design because of the simplicity and limited number of parameters that need to be accounted for while designing a novel mechanism for transcriptional regulation. Indeed, Temme *et.al* recently engineered T7

RNA polymerases that recognize different promoters thereby creating four different polymerase promoter pairs which were used for demonstrating an AND logic operation[25]. For these reasons, here we focused on engineering mechanisms for ligand dependent transcriptional regulation that will serve to broaden the applicability of T7 promoters for enabling *in vivo* and *in vitro* synthetic gene circuits.

Present strategies involve repression of T7 RNA polymerase by well-characterized transcriptional repressors. Here we tackle the following aspects of gene regulation in cell free systems –

1) Development of easy to implement and potentially modular strategy for regulating transcription using DNA aptamers.
2) Testing mechanisms for developing multi-input responsive T7 promoters for use in cell free systems.

Nucleic acid aptamers for regulation of gene expression -Cellular mechanisms for signal sensing have served as a template for efforts geared towards rationally designing synthetic sensors and gene switches, which most often is a protein, peptide or a nucleic acid molecule. In theory, transcriptional promoters, transcriptional regulators, location of binding site relative to the promoter region, nascent RNA molecule and the RNA polymerase itself can be engineered for regulation. Most commonly though, transcriptional regulation is achieved by engineering protein based transcriptional repressors and activators. Signal specificity is often encoded onto the sensing molecule, which is transduced to functional domain that operates at the transcriptional, translational or post translational levels[68].

Natural switches contain two components, a signal sensing and an effector domain that allow them to transduce signal changes into a measurable cellular response. Affinity of a sensory domain to a signal often sets the sensory threshold of a system. Ligand binding often induces a conformational change in binding domain, which is transmitted to the effector domain thereby modulating its function[61]. Several proteins and peptides have been utilized for achieving signal specific sensing of molecules. These are based on harnessing the chemical complexity of proteins that enables highly specific and efficient signal transduction. Protein based transcription factors with novel ligand sensitivity have been typically designed by fusing a natural ligand-sensing domain with a DNA binding domain[69-72]. However, rational design of complex tertiary interactions govern the formation of 3D structures of proteins that are responsible for sensing, binding and catalytic functions which makes protein engineering a rather challenging exercise. Therefore, encoding novel signal specificity and rational design of extensible signal transduction mechanisms that are based on protein based transcription factors presents significant challenges.

Nucleic acid (DNA and RNA) aptamers offer a complementary approach for conferring ligand sensitivity onto a recognition module. DNA and RNA aptamers are single stranded nucleic acid molecules that bind their target molecules with high affinity and specificity. Since the breakthrough papers from Szostak and Gold groups demonstrated the selection of RNA ligands that bind T4 DNA polymerase and an organic dye with high affinity[73, 74], aptamers that bind a wide variety of small molecules such as ATP[75], proteins[76, 77] and even whole cells[78] that otherwise do not have nucleic acid recognition properties have been selected. Aptamers that bind to their target

molecule are typically selected from combinatorial library of consisting of 10^{12}-10^{13} oligonucleotides. The subset of oligonucleotides that bind to their target are then subjected to subsequent rounds of selection, until, a nucleic acid ligand that binds its target with high affinity is selected (Figure 1.5). Furthermore, automated strategies for aptamer selection have been developed that permit the rapid selection of RNA aptamers against a target of interest[79].

Naturally occurring RNA structures that bind their target molecules, known as riboswitches, form elaborate structures that regulate gene expression are believed to be amongst the earliest mechanisms for achieving ligand dependent regulation of gene expression[80]. Naturally occurring riboswitches modulate gene expression by coupling a change in nucleic acid conformation that accompanies ligand binding with transcriptional or translational regulation[81-83]. For instance, riboswitches that block ribosome scanning, facilitate or interfere with availability of ribosome binding sites and transcriptional terminators have all served as mechanisms for regulating gene expression[84].

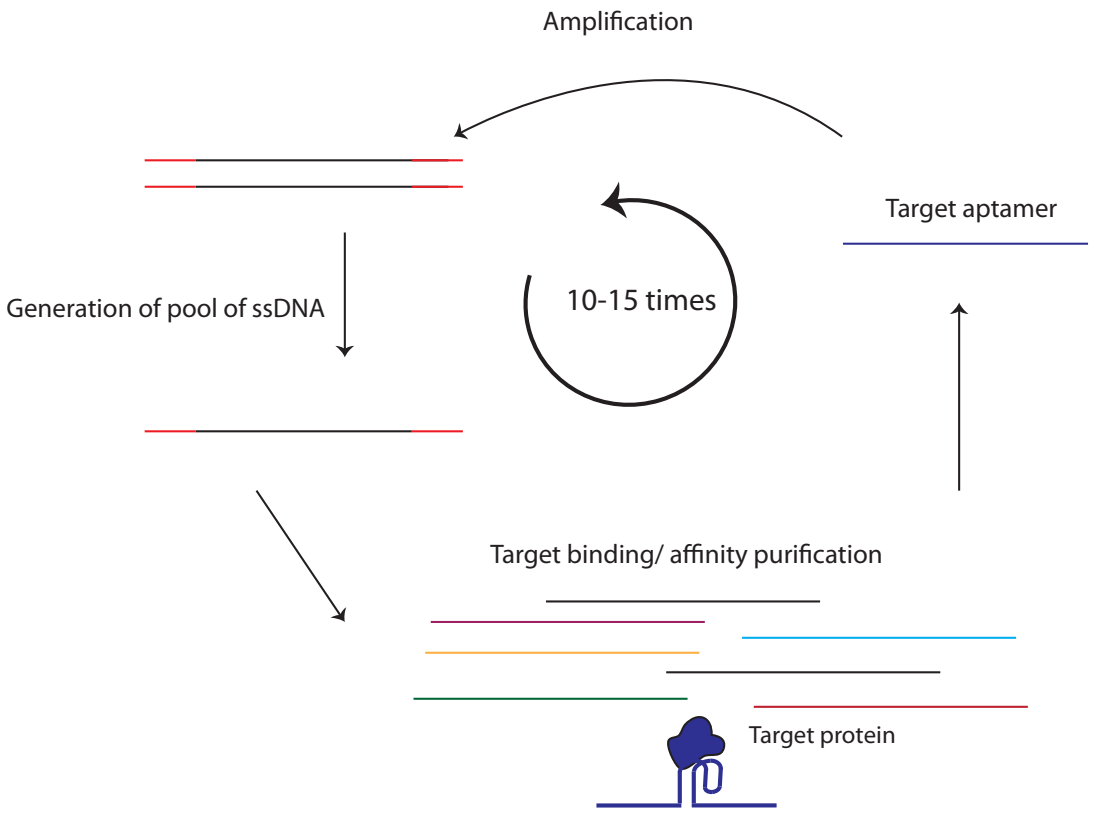


Figure 1.5 - Schematic of SELEX - First a pool of single stranded DNA is generated from a double stranded DNA pool that consists of a randomized pool of DNA molecules . a DNA oligonucleotide that binds the target of interest is selected from a single stranded DNA oligonucleotide library by affinity purification procedure. The DNA oligonucleotide thus selected is used to then generate a library of double stranded and single stranded DNA that can then be used to generate a double stranded pool of double stranded DNA. This selection procedure is repeated several times until a aptamer that binds with very high affinity is selected.

While RNA aptamers have found uses in gene regulation, DNA aptamers have largely been adapted for development of *in vitro* biosensors [91]. Several properties of DNA aptamers make them suitable for *in vitro* applications. Firstly, in contrast to protein-based sensors, nucleic acid based sensors show greater temperature stability and have extended shelf-life. Secondly, since aptamers can be selected in an *in vitro* buffer system, an aptamer that binds a ligand of choice in the system of interest can be selected [79]. And thirdly, ease of chemical modification and immobilization without a significant alteration in the aptamer structure make them amenable to a variety of analytical formats. Hence, aptamer molecules conjugated to fluorophores [92, 93] and nanoparticles[94] to create biosensors harness the conformational change accompanying the target binding that can lead to biochemical, electrochemical and spectroscopic response. In addition, label free methods in which aptamers have additionally been incorporated into longer nucleotide sequences to enable detection via proximity ligation, and ligand dependent amplification of the nucleic acid sequence using RT-PCR[95] and rolling circle amplification[96] have been developed.

However, difficulty in incorporating single stranded DNA aptamers into cells has precluded their use for modulating gene expression. Since we seek to employ transcriptional regulation mechanisms that operate in a cell free context, use of ssDNA does not pose a major obstacle. In Chapter 2, a strategy for harnessing DNA aptamers to regulate transcription in cell free system is described.

Combinatorial promoters for signal integration in cell free systems

Mechanisms to integrate and respond to multiple cellular and environmental cues underlie a biological cell's ability to adapt to its ever-changing environment. Living cells harbor a rich diversity of mechanisms that operate at the level of transcriptional factors, RNA and protein signaling pathways to perform the task of integrating environmental signals and mobilizing a cellular response. Description of these responses as a logic function succinctly captures the nature of the response to different inputs. Transcriptional and translational regulation strategies are often wired to execute certain cellular tasks in response to a combination of internal and external signals.

In addition to their requirement for cell survival, logic control is central to the construction of synthetic gene circuits. Logic gates can be cascaded and integrated and program cellular behavior to respond to multiple signals from the environment. Like their semiconductor-based counterparts, synthetic biological logic gates are critical to our efforts to develop large-scale circuits that perform sophisticated functions.

Accordingly, several designs that rely on nucleic acid substrates and utilize enzyme based transformation for construction of logic gates that function *in vivo* and *in vitro* have been developed. Nucleic acid based logic gates rely on DNA hybridization, nucleic acid catalysis driven by DNA or RNAzymes to perform computation[97-100]. Alternatively, biocatalysts that carry out substrate transformation in the presence of specific input signals have been utilized to implement a single[101] logic operation and a series of concatenated logic gates using a series of enzymes [102]. The output from these two designs is either short DNA or RNA oligonucleotides or transformed small molecules or proteins. In contrast to these strategies, genetic logic gates perform signal

integration at the transcriptional or translational level and can transform a small molecule or protein-based signal to code for an actuator protein. These logic operations that are inspired from natural mechanisms to achieve signal integration can be performed either at the transcriptional level [103, 104], translational level [86] or in a strategy that is a hybrid of the two mechanisms[24].

The simplest and one of the best studied of these mechanisms is combinatorial regulation that occurs at transcriptional promoters. Bacterial cells contain several transcriptional factors that respond to environmental signals and either activate or repress gene expression[64]. Most often global regulators of gene expression such as cAMP and non-specific DNA binding proteins that modulate course in gene expression of several genes, act in concert at a promoter with local activators and repressors thereby effecting stimuli specific changes in the expression of the target genes. For instance, regulation of the lac operon that expresses enzymatic machinery to utilize lactose perhaps is one of the best studied examples of achieving combinatorial regulation of expression wherein, the enzymes are expressed if lactose is the sole carbon source in the media[105].

Combinatorial promoter libraries in which promoters of different strength express a reporter gene such as luciferase and GFP have been valuable for generating synthetic combinations of promoters that differ from each other in terms of promoter strength, and relative location of operators that bind transcriptional repressors and activators[106]. Several studies have taken a synthetic approach to provide quantifiable relationships between promoter architecture and transcriptional activity from a promoter. *E.coli* promoters comprises of a -35 box and a -10 box separated by a core region and regulatory regions that exert an effect on the promoter and span about 100 bases upstream and

downstream to the promoter[62, 107]. While transcriptional activators recruit the polymerase to an otherwise weak promoter, transcriptional repressors block transcription by either sterically occluding transcriptional factor binding to the promoter or preventing promoter clearance. Along with the biochemical role of the transcriptional factor, position of the transcriptional binding is critical to its function. For instance, a transcriptional activator that binds upstream might function to recruit the polymerase to the promoter but when bound at the core region prevents polymerase binding to the promoter and functions as a repressor[108]. Taken together, these results indicate that a detailed understanding of the effects of location of the operators relative to promoter and the proximity of the operator is critical for rational programming of transcriptional logic. These rules have been elucidated for *E.coli* promoters are shown in Figure 1.6.

As mentioned earlier, we set out to engineer transcriptional logic on T7 promoters. A key challenge to regulating T7 promoters is the absence of transcriptional activators and the limited regulatory region flanking the T7 promoter. Unlike multi-subunit polymerases, these polymerases lack activators and co-factors for recruiting the polymerase to the promoter thereby greatly reducing the number of components to regulate expression from T7promoters[67]. Moreover, T7 polymerase responds to repressors that bind at a site proximal to and downstream from the T7 promoter and not to distal sites thereby reducing the region from which to regulate the T7 promoter [109]. In chapter 3, we address the task of constructing multiple input responsive T7 promoters for achieving logical control of gene expression in *E.coli* and in cell free systems by using DNA looping.

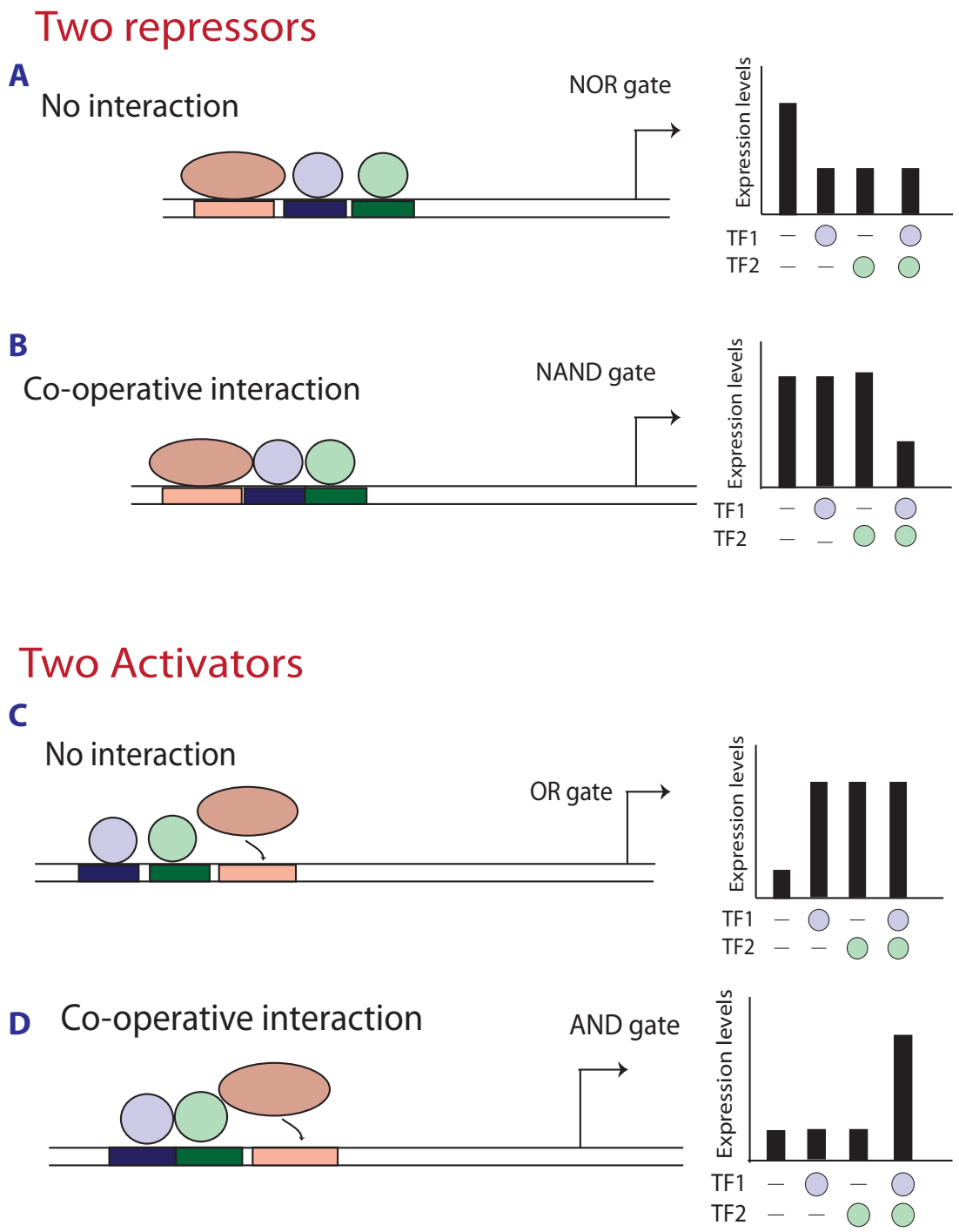


Figure 1.6 Cis regulatory schemes for implementing logic gates with bacterial promoters A) Two repressors bound downstream from transcriptional start site that can repress the promoter individually form a NOR gate , B) whereas if co-operative interaction between two repressors is needed an NAND gate is generated wherein binding of both transcriptional repressors is necessary to repress transcription from the promoter. C) Similarly, two activators that can activate transcription individually form a OR gate where presence of either one of the transcriptional factors is sufficient for activating expression D) a AND gate can be formed when interaction between two activators co-operatively repress expression.

Examine modes of system assembly and implement negative feedback motifs in cell-free systems

Thus far mechanisms and strategies to regulate transcription have been described. These and other ligand sensitive gene regulation mechanisms serve as communication channels from which to communicate with the synthetic device. However the manner in which these signals are processed, is dictated by the information-processing network that exists in cells. Biological systems are characterized by features such as modularity and robustness to fluctuations[110]. Understanding how network connectivity confers these properties onto biological systems is an active area of investigation. The prevailing view is that there may be an underlying simplicity to these complex networks[110, 111]. The simplicity is attributed to the occurrence of network motifs in biological networks that large scale systems biology investigations have revealed [10]. A network motif is a complete sub-network that occurs in biological networks at a frequency higher than that would be expected from a random network built from the same number of modes. Recurrence of sub-networks indicates that network design has functional consequences. Consequently, uncovering and experimentally validating the functional implications of network motifs has garnered considerable interest for understanding biological systems. In addition, the knowledge would be valuable for constructing engineered biological circuits that perform sophisticated tasks.

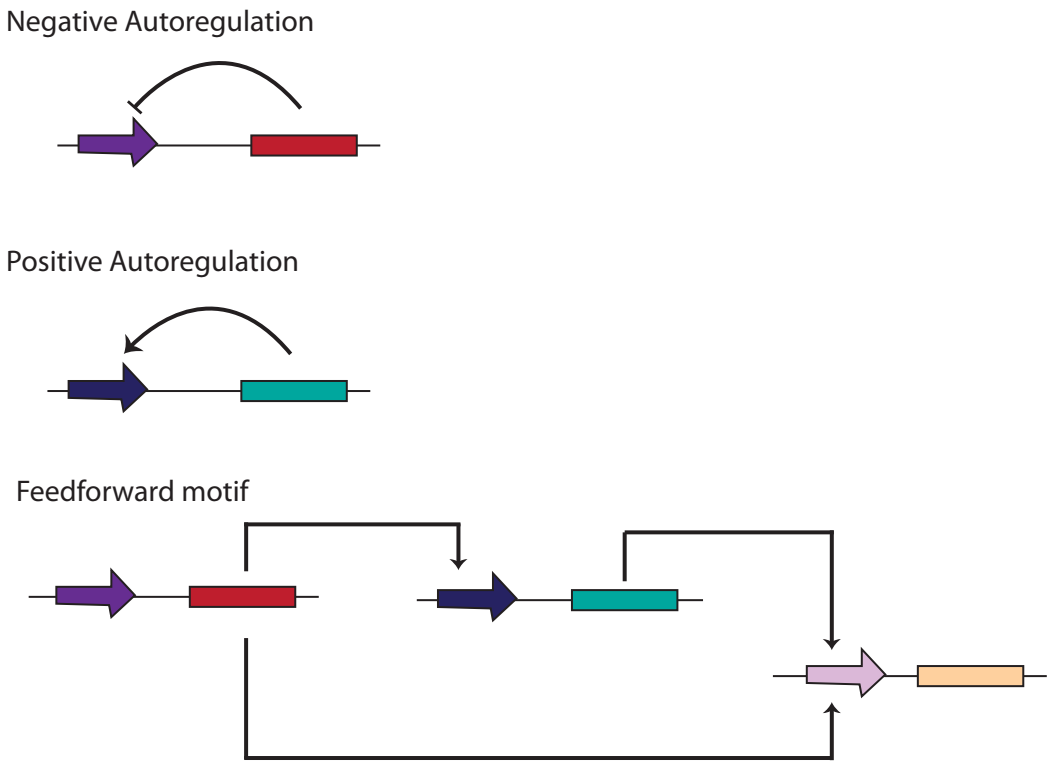


Figure 1.7 Commonly occurring network motifs in biological cells – in a negative feedback circuit, the promoter drives the expression of its own repressor whereas a promoter. The promoter expresses its own activator. In a feedforward loop a transcriptional activator activates the expression of another gene, which in turn activates the expression from a third gene along with its own activator.

Accordingly, several simple network motifs have been built from well-characterized components and their behavior has been evaluated in several biological systems (Figure 1.7). For instance, investigation into transcriptional connectivity in *E.coli* cells revealed that nearly 50 % of transcriptional factors are negatively autoregulated[112]. Experimental evaluation of synthetic negatively autoregulated circuit indicates that the circuit speeds up cellular response to a signal and achieves the same steady state protein levels when compared to a system lacking regulation by a protein whose expression is driven from another promoter[113]. Therefore regulating transcriptional factors that ought to be maintained at a steady state level and be resistant to fluctuations inside a cell, a negative feedback motif is utilized. In contrast, positive autoregulation promotes bistability in a system and helps maintain a mixed population of cells that can respond to changes in the environment[114, 115].

Networks implemented in cells often are encoded onto a single plasmid. However the requirement of transforming plasmids with compatible origins and antibiotic resistance places an upper limit on the number of plasmids that can be transformed into cells. Additionally, manipulating large DNA fragments that span tens of thousands of bases as would be required for assembling large scale networks can be experimentally challenging[116]. In contrast, genetic elements encoded onto different plasmids can be conveniently used in enabling gene networks in cell free systems. Given the limited expression capacity of cell free systems, small changes in RNA expression levels become significant. For instance, transcription termination from even the most efficient mechanisms is known to be only 70% efficient in stopping readthroughs[117]. While this may not have a significant impact on expression from a single gene, these losses become

significant in a regime that involves multiple plasmids and limited energy sources. In chapter 5 therefore, the impact on the efficiency of expression from two assembly schemes was evaluated.

Implementation of analogous networks requires an understanding for tuning expression dynamics in cell free systems. In cells, genetic components such as ribosome binding sites and promoter strength are tuned to generate the required behavior[68]. On the other hand, the parameter space for optimizing expression is greatly expanded in cell free systems. Parameters such as gene dosage, effect of introducing genes in single versus multi- plasmid systems and RNA polymerase concentrations can be precisely controlled in cell free systems and might prove critical in assembly of systems of increasing complexity[117]. In chapter 5, we focus on altering transcriptional rates by modulating polymerase concentrations thereby implementing a negative feedback circuit in cell free systems from well-characterized T7lacO promoters.

The rich diversity of molecular components in biological system provides a valuable arsenal of tools for designing circuits geared towards testing biological hypotheses and for biomedical and biotechnological applications. The availability of well-characterized genetic parts and chassis will be critical to rationally designing predictable biological circuits. To this end, *in vitro* cell extracts provide a flexible platform in which to implement gene circuits. Here, ligand sensitive regulation mechanisms and strategies for signal integration were developed for use in cell free systems. Additionally, methods for system assembly were evaluated and a negative feedback motif was implemented. The tools developed in this study will greatly bolster

efforts to implement gene circuits for biotechnological applications and for testing biological hypotheses.

Chapter Outline

The aim of this dissertation is to develop transcriptional switches and gene circuits in cell free systems.

Chapter 1- Introduction

Chapter 2- Ligand mediated transcriptional regulation using DNA aptamers- In the quest for developing a potentially modular strategy for achieving gene regulation a DNA aptamer mediated ligand dependent transcriptional regulation mechanism was developed. In Chapter 2, strategies for assembling DNA templates from phagemids DNA and thrombin binding aptamers to regulate transcription in *E.coli* extracts are described.

Chapter 3- Dual regulation of T7 promoters using lac and tet repressors in cell and cell free systems. Here I turn to engineering a transcriptional dual regulation of T7 promoters using lac and tet repressors. T7lacO promoters that harness a looping based mechanism to repress expression were first developed. Tet repressor was then introduced into this framework and an IMPLIES function was realized in *E.coli* cells as well in cell extracts.

Chapter 4- Implementation of negative feedback using T7lacO promoters in cell free systems- In this chapter, I explored different methodologies to assemble gene networks and realized simple negative feedback in cell extracts using T7 lacO promoters. This chapter contains excerpts from the paper "Expression optimization and synthetic gene circuits in cell free systems".

Chapter 5 – Conclusions and future directions - I summarize the conclusions and indicate future directions for this work.

Appendix A Here, I discuss the DNA assembly techniques that were tested but ultimately not used for the experiment.

Appendix B describes the efforts to repress translation by DNA aptamers.

Appendix C –Finally, I turn to implement cell free reactions in silicon based microfluidic devices.

Chapter 2- Ligand mediated transcriptional regulation using DNA aptamers in cell free systems

Introduction

Synthetic gene circuits comprised of novel genetic regulatory mechanisms have emerged as powerful tools for understanding and harnessing biological function [1]. Engineered arrangements of well characterized genetic components have resulted in systems capable of predetermined functions such as bistability [4], logic control[86] and oscillation[4] of gene expression. Synthetic gene circuits also offer the opportunity to redesign biological systems for the production of biofuels and other chemicals as well as for constructing devices for sensing and responding to biomedical conditions. In practice, the majority of synthetic gene circuits have been implemented in cell-based systems. While these demonstrations benefit from natural mechanisms to sustain a living cell, such as protein synthesis and degradation, creating predictable engineered systems can be complicated by interference from endogenous host machinery and selection pressures that act against unneeded, resource consuming systems[3, 18]. Additionally, conflicts occur when sensing or generating materials that can compromise cell viability and survival. Therefore, alternative strategies to harness and understand biological complexity are a needed complement to existing cell-based approaches. [118].

In this regard, cell free systems provide a versatile platform for understanding and applying the design elements that underlie cellular efficiency [119-121]. Cell free approaches employ select cellular components, produced naturally or synthetically, to carry out defined biological processes. Issues related to plasmid compatibility, protein toxicity or maintenance of a living cell can be mostly ignored, allowing focus on defining

essential system components[122] and implementing predictable dynamic behavior. The flexibility, simplified context and precise specification of system components are distinct advantages of the cell free approach. A number of cell free, *in vitro* gene circuits have been demonstrated. For example, simplified nucleic acid templates in which transcription was regulated by DNA hybridization were co-opted to build bistable switches and oscillators that reasonably agree with quantitative predictions [33, 34]. Additionally, expression cascades[48], negative feedback[38] and logic gates[52] have been realized using circuits involving protein intermediates in cell free protein extracts.

Well-characterized molecular tools for signal sensing and tuning gene expression are essential for the design and construction of synthetic gene circuits in cells and cell free systems [68] [52, 109, 123]. In particular, ligand responsive gene regulation strategies are key. While a myriad of gene regulatory mechanisms are used in natural cells, ligand dependent transcriptional control strategies that function in cell free extracts remain fairly limited. In general, the library of gene promoters available for synthetic constructs is limited. When compared to cell based systems, cell free systems afford an opportunity to expand the repertoire of regulation strategies [28].

Commonly, synthetic gene constructs take advantage viral RNA polymerases and their associated promoter elements for gene expression. For example, T7 RNA polymerase is commonly used in cell extracts for driving transcription due to the enzyme's stability and high processivity [124]. While these characteristics are desirable for achieving high yield protein synthesis, their use in synthetic biology is limited because of the lack of sufficient ligand sensitive T7 promoters[48]. Attempts to engineer ligand regulatable T7 promoters rely on either protein based transcription factors or

modified nucleic acid bases to confer ligand sensitivity [125-127]. For example, existing strategies to regulate transcription by T7 RNA polymerase involve placing a cis acting promoter element, that binds to a repressor, downstream to the transcription start site [128] [109].

While the use of traditional protein-based transcription factors allows for effective and tight transcriptional repression, lack of accurate structure prediction methods makes altering ligand specificities or creating new transcriptional factors with ligand specificity remains non trivial [129]. The use of nucleic acid aptamers presents an alternate approach and can potentially allow regulation of gene expression in response to a wide variety of small molecules and proteins. Aptamers are single stranded DNA and RNA molecules that can be engineered to bind to specific target molecules with high affinity and specificity. RNA aptamers have found extensive application and often couple the binding event and the ensuing conformational change for regulation of transcription or translation [87, 130-132]. Nucleic acid aptamers offer several practical advantages. First, aptamers can potentially be selected against any ligand of interest from a combinatorial library using an iterative affinity selection procedure [73, 74]; second, aptamer target molecules with different affinities can be selected to set different sensory thresholds appropriate for different applications [90, 133]; and finally known hybridization rules facilitate predictive and rational design of DNA domains. The ease and predictability of engineering nucleic acid domains make DNA and RNA molecules particularly useful substrates for engineering flexible platform for achieving tunable sensing and actuation.

Here we describe a new approach to using aptamers to control gene expression at the transcriptional level using viral promoters. The approach involves the

insertion of a DNA aptamer sequence proximal to the T7 promoter such that binding prevents transcription (Figure 2.1). The required single stranded regions are well tolerated by viral polymerases [67, 134] and easily employed in a cell free context. Thrombin binding DNA aptamer (TBA) was selected for demonstrating analyte specific transcriptional control. TBA is well-characterized and is known to bind to human α-thrombin with high affinity (K_d of 10-100nM) and specificity. Presence of additional flanking sequences and aptamer immobilization are not detrimental to thrombin binding, which facilitates the insertion of the aptamer sequence into the DNA template[95, 135]. We show that thrombin can be used to effectively repress expression from single stranded thrombin aptamer containing templates in a cell free context. In addition, exogenous addition of thrombin aptamer oligonucleotides led to the effective reversal of gene expression from these templates.

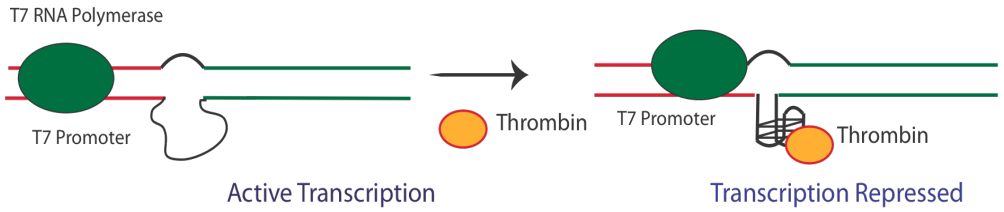


Figure 2.1 Hypothesis for aptamer mediated transcriptional regulation. ssDNA aptamer binding region is placed downstream to the T7 promoter . We tested the hypothesis that thrombin binding to the DNA aptamer represses transcription from T7-aptamer promoters.

Results

Approach for template assembly for achieving aptamer mediated gene regulation

To enable transcriptional regulation using DNA aptamers, double stranded DNA templates were created that contain an unpaired "bubble" DNA region, which contain DNA aptamers on the template and the non-template strands, after the double stranded promoter (Figure 2.2). These structures were created from hybridization of single stranded DNA templates, generated from phagemids, containing complementary and non-complementary regions. ssDNA templates generated from phagemids offer the advantage of producing high yield ssDNA that are long enough to code for a protein[136]. The bubble DNA template was created by placing the thrombin binding aptamer downstream to the transcriptional start site in pBluescript KS II (+) and pBluescript KS II (-) plasmids. Restriction digestion of the double stranded segment of DNA and mung bean nuclease digestion of ssDNA section of the template confirmed formation of bubble regions in double stranded template (Figure 2.3). In addition, to these bubble templates, single stranded templates containing a double stranded T7 promoter were generated by annealing an oligonucleotide to pBluescript KS II (-) phagemid .

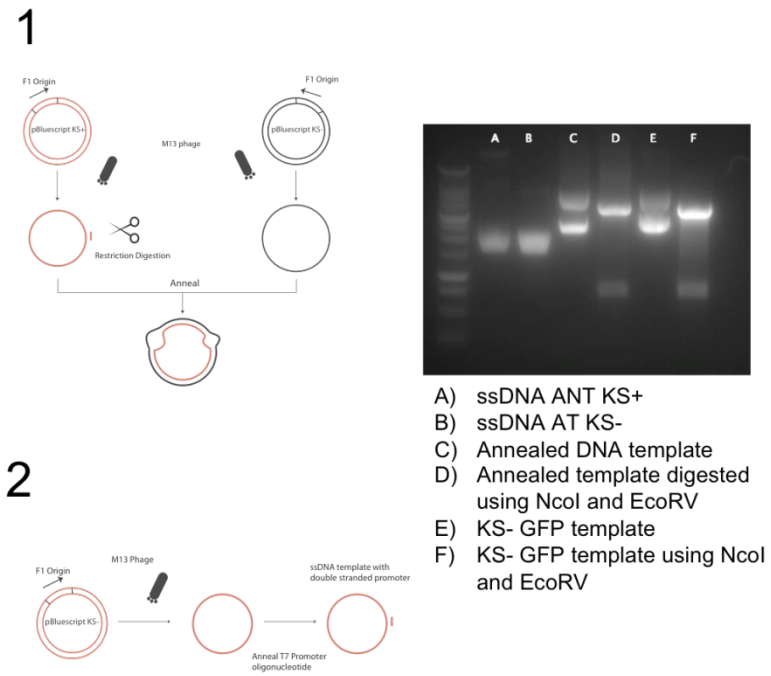
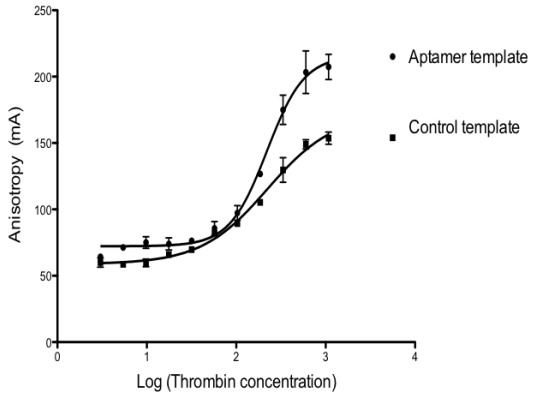


Figure 2.2. Schematic of steps for template assembly assembling 1) bubble DNA templates 2) ssDNA with ds promoter.



Anisotropy from single stranded DNA templates

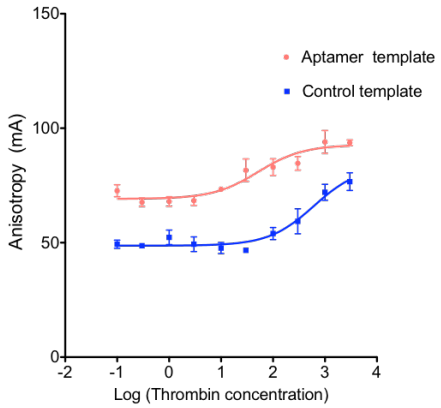


Figure 2.3 Fluorescence anisotropy measurements from bubble templates (top panel) and ssDNA templates(bottom panel). The thrombin concentrations before the logarithmic transformation were in nM.

Effect of template on aptamer mediated repression

Fluorescence anisotropy experiments with ssDNA and bubble DNA confirmed thrombin binding to the aptamer (Figure 2.3). To examine the ability of thrombin to repress transcription from a T7 promoter, an aptamer sequence was placed immediately downstream from the transcription start site. . In addition, a 4bp stem loop structure was added to the thrombin aptamer to facilitate the formation and increase the stability of the thrombin aptamer [137]. GFP coding sequence was inserted in the template and the efficiency of transcriptional repression was monitored by cell free protein synthesis reaction (Figure 2.4). Fluorescence measurements of GFP expression from these templates indicated that placing the aptamer in close proximity to the promoter sequence exhibits increased regulatory efficiency and is accompanied by a lowered basal expression level As expected constitutive expression from plasmid templates was higher when compared to ssDNA templates and the bubble templates. While addition of 1.8 μM thrombin results in modest changes in gene expression from the plasmid template and bubble template, up to a five-fold change in gene expression was observed with the ssDNA aptamer templates. Modest changes in gene expression from double stranded templates indicates that the addition of thrombin does not interfere with transcription and translation in the cell free extract and that thrombin is specific to the single stranded DNA template. Further, rapid repression of expression was observed upon the addition of thrombin (Figure 2.5).

Testing the specificity of thrombin mediated repression

To test the specificity of thrombin dependent repression, the non-specific, single strand DNA binding T4 gene 32 protein [138] was tested for transcription repression on T7 aptamer promoters. The addition of ssDNA binding protein did not have a significant effect on transcription demonstrating that specific protein binding to the aptamer placed proximal to the transcriptional start site is required for effective transcriptional repression (Figure 2.6A).

We tested the specificity of the transcriptional repression further by evaluating competitive inhibition of gene repression by exogenously added thrombin aptamer oligonucleotides to 10nM aptamer template bound to 2μM thrombin (Figure 2.6B). We observed that the addition of thrombin aptamer (12μM), in excess of thrombin protein, completely relieved thrombin mediated gene repression, whereas the addition of a non-specific DNA oligonucleotide did not affect repression. Therefore, the addition of exogenous DNA aptamer allows for "induction" of expression from these promoters.

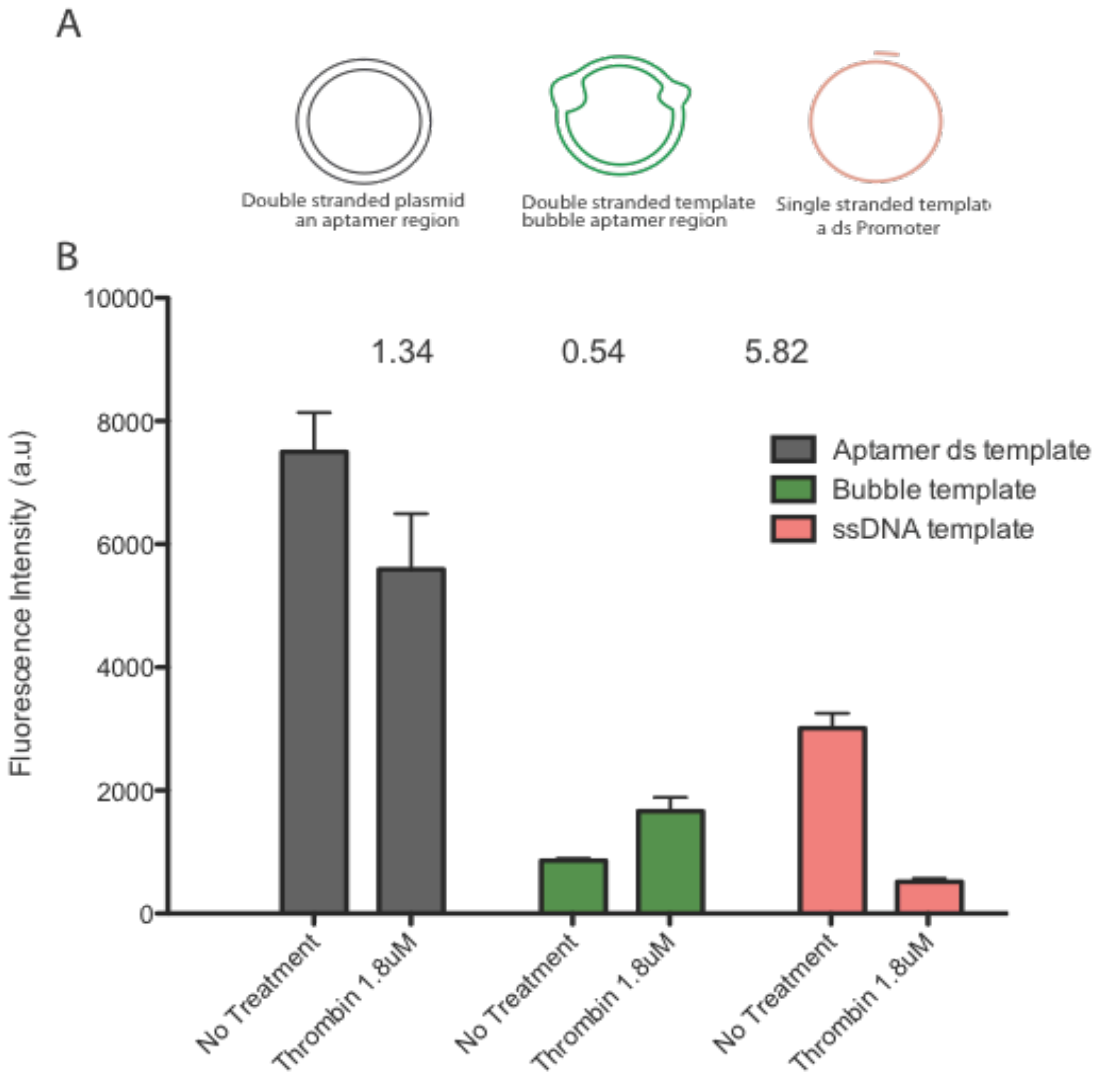


Figure 2.4 Effect of template on aptamer mediated repression A) is a schematic of templates used – The grey cartoon represents the double stranded plasmid with a thrombin aptamer downstream from the promoter, the green template is a "bubble template" that contains an aptamer structure both at B) Flourescence measurements at the end of 6 hours from these templates in absence or presence of thrombin concentrations

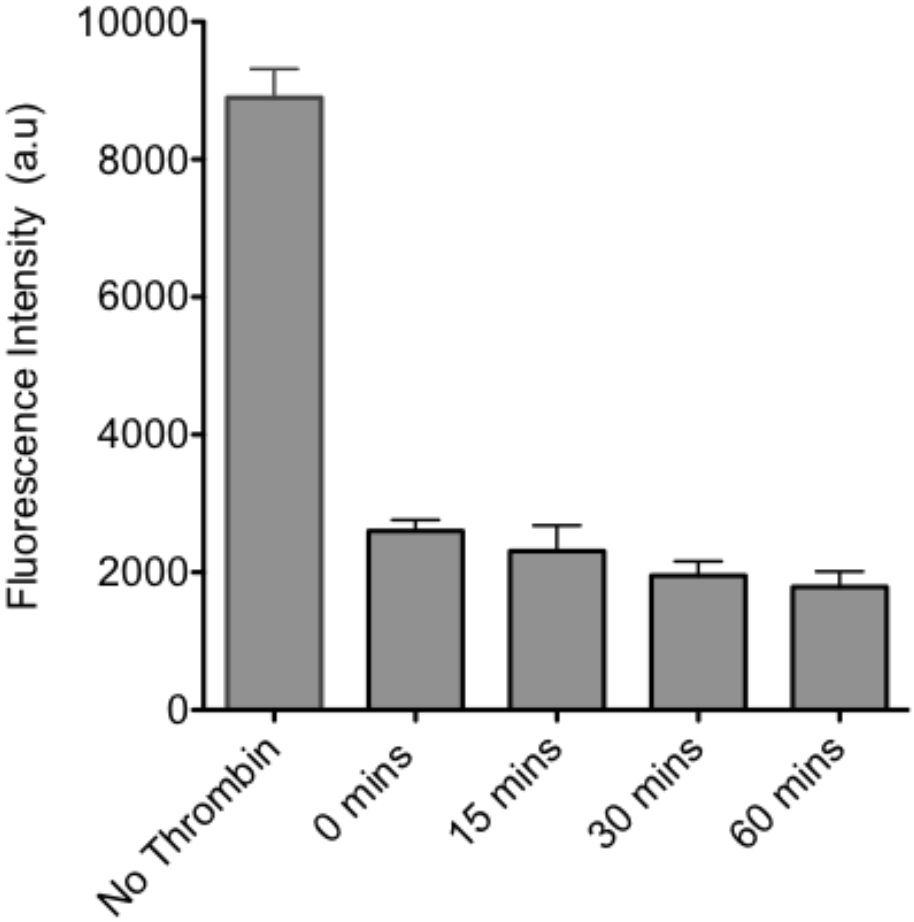


Figure 2.5 Time course for response of ssDNA ATKS construct to the addition of 1.8μM Thrombin –the graph shows fluorescence values after 6 hours. The column indicating 0 mins indicates shows thrombin was added after extract was added to ssDNA ATKS template. Times on the graphs indicate the duration for which thrombin was incubated with ATKS template.

Table 2.1 –List of plasmids used in this study.

Plasmid Name	**TBA position (relative to TSS)**	**Reporter**	**Backbone**
pKSGFP	-	EGFP	pBluescriptKS (-) II
pANTGFP	+2	EGFP	pBluescriptKS (+) II
pATGFP	+2	EGFP	pBluescriptKS (-) II
pNTAGFP	+9	EGFP	pBluescriptKS (+) II
pETAGFP	+9	EGFP	pBluescriptKS (-) II
pETA26GFP	+26	EGFP	pBluescriptKS (-) II
pDETAGFP	+9	EGFP	pBluescriptKS (-) II

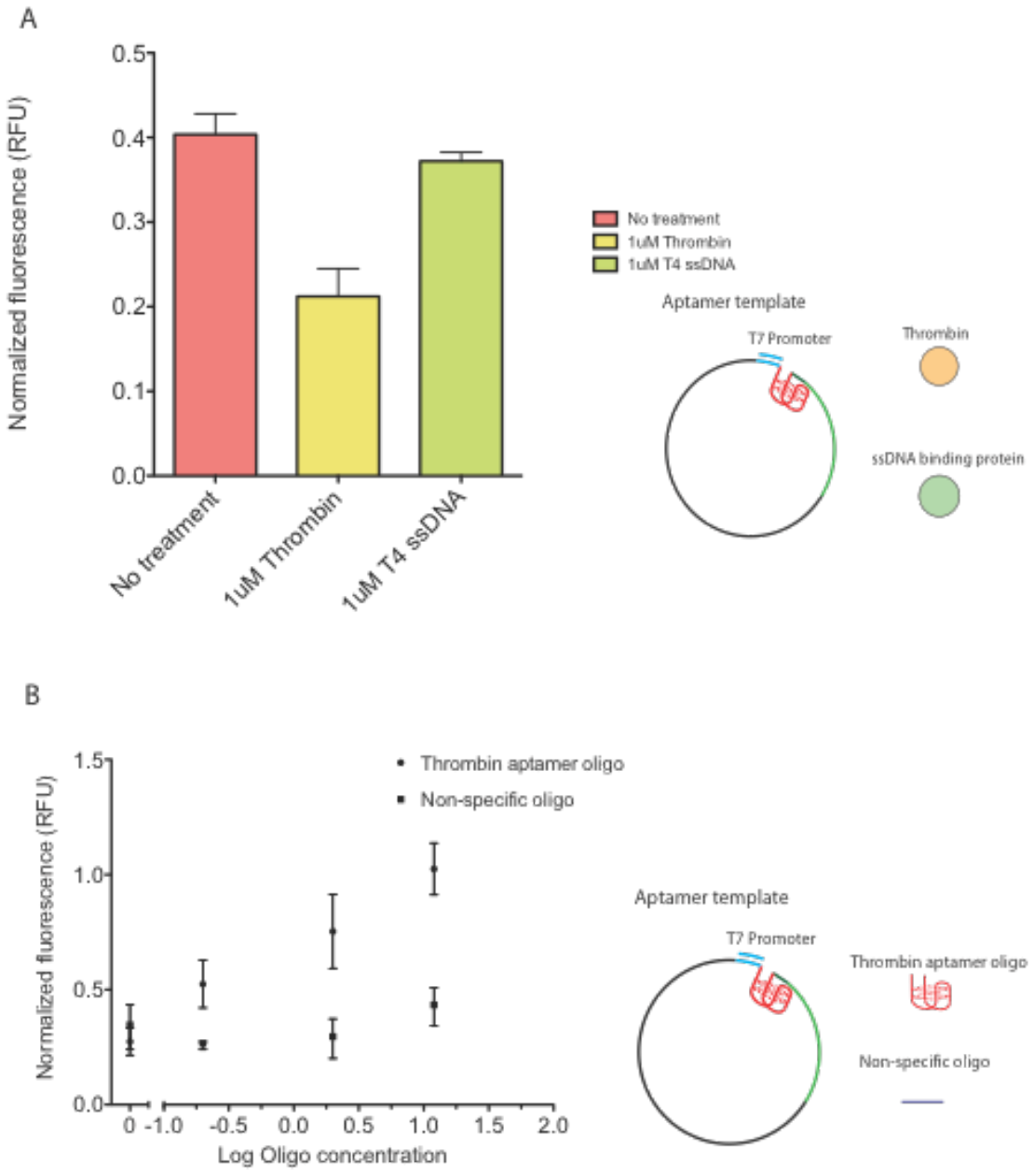
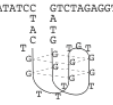
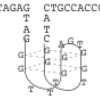
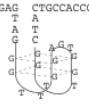


Figure 2.6 Testing specificity of aptamer mediated transcriptional regulation.A)Fluorescence values have been normalized to expression from ssDNA generated from KSGFP. B) Fluorescence values from have been normalized to expression from ssDNA generated from ATKS templates in the absence of thrombin.

Table 2.2 – Table denoting the basal level expression relative to a ssDNA template lacking an aptamer downstream from transcriptional start site. Fold change column indicates the change in expression from these templates in response to 1.8 μM thrombin.

Template	Relative expression	Fold Change
TAATACGACTCACTATAGG ATTATGCTGAGTGATATCC GTCTAGAGGTGGCGCCA 	0.45	4.75
TAATACGACTCACTATAGG ATTATGCTGAGTGATATCCTCTAGAG CTGCCACCGCCGGC 	0.70	2.65
TAATACGACTCACTATAGG ATTATGCTGAGTGATATCCCGCTTAACCTCGAGGTGGCTCTAGAG CTGCCACCGCCGGC 	1.22	1.97

Effects of position and number of the aptamer sequences on transcriptional regulation - Effects of placement of the thrombin aptamer, relative to transcriptional start site on transcriptional efficiency were explored. The thrombin aptamer was placed at three positions relative to the transcription start site- +2, +9 and +26 (Table 2.2). Results from the assay show that placement of the DNA aptamer away from the transcriptional start site resulted in an increase in basal gene expression levels, the magnitude of change in gene expression decreased. Dose response curves with the aptamer at +2 and + 9 positions show a half maximal repressor concentration of 218.00 ± 1.58 nM and 567.90 ± 2.21 nM respectively (Figure 2.7).

To test if the addition of tandem thrombin aptamers results in improved gene repression, we constructed a dimeric DNA aptamer template and tested the template for repression. However, the templates did not express very well (Figure 2.8).

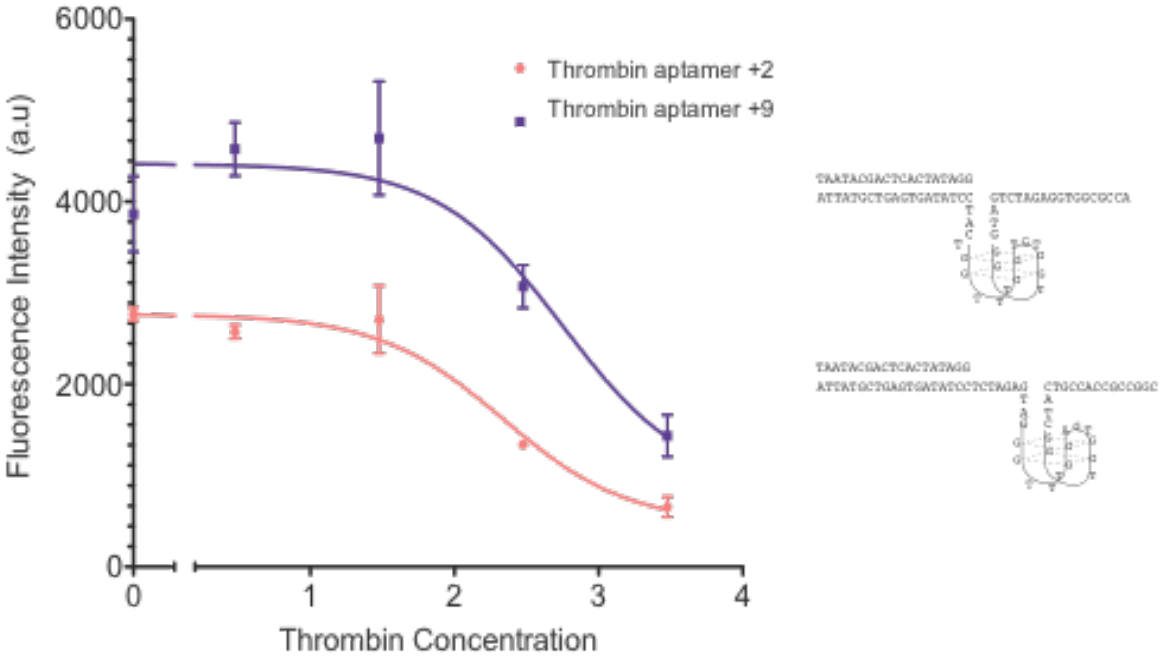


Figure 2.7 Dose response curves for +2 and +9 aptamer constructs- X-axis indicates log of thrombin concentration. The thrombin concentrations were in nM before the logarithmic transformation

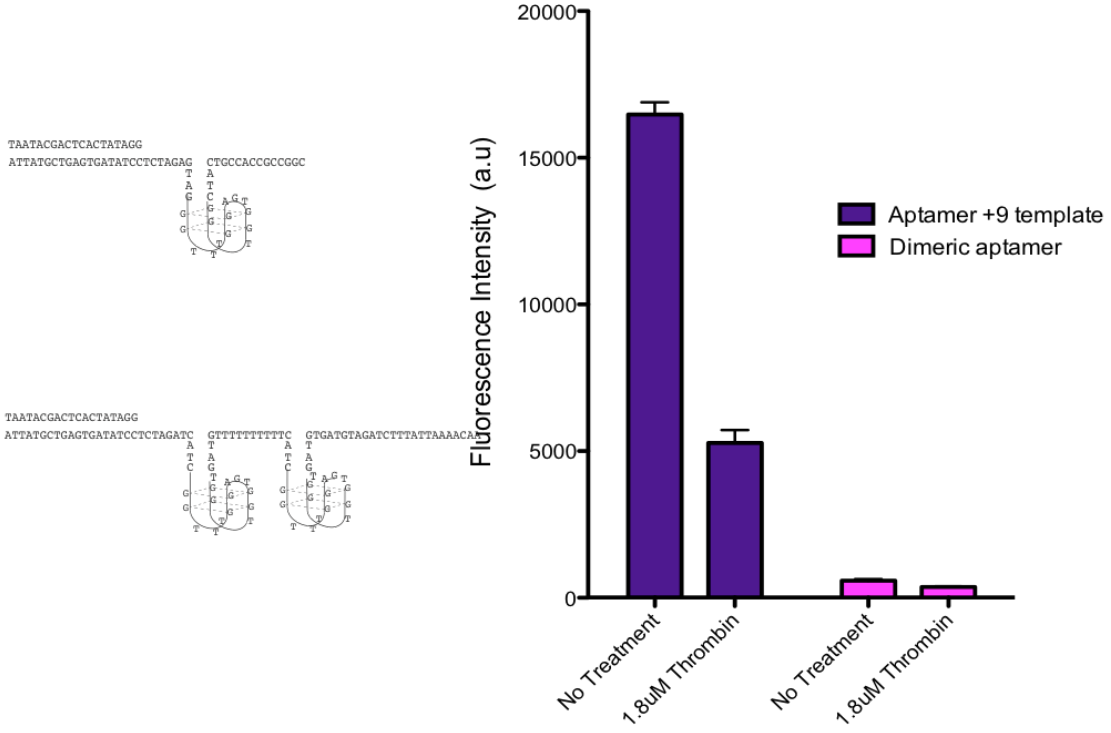


Figure 2.8 Effect of placement of dimeric aptamers downstream to the transcriptional start site. The constructs tested are depicted on the left, whereas the graph indicating the response of the DNA templates to 1.8μM thrombin.

Discussion

Cell free systems are a promising platform for implementing engineered networks from defined components [33, 34]. Effective application of network designs will require the availability of a library of environmentally responsive promoters [27]. Controlling gene expression at the transcriptional level offers several advantages [3]. Being the first level of gene expression, multiple downstream targets can be regulated simultaneously. In addition, signal amplification can be achieved since binding of a single transcriptional factor regulates the expression of several hundred resulting RNA and protein molecules[61].

Aptamers are a promising approach to creating ligand dependent promoters of arbitrary design. The flexibility afforded by RNA aptamers has been utilized to bring about ligand dependent transcription termination, [84, 87, 139, 140]. However, ligand dependent DNA aptamer mediated regulation of transcriptional initiation has not yet been achieved [129]. A first roadblock to employing DNA aptamers for control at the gene transcription level is the requirement for single stranded templates. Several strategies for generating linear ssDNA templates such as affinity purification of biotin labeled ssDNA generated from PCR[141] and rolling circle amplification[142] were evaluated. However, these approaches result in only small amounts of ssDNA, which were insufficient for optimizing protein synthesis reactions (data not shown). The use of ssDNA derived from phagemids allows creation of templates long enough to code for a protein sequence and generates templates in quantities needed for refining transcriptional control of cell free protein synthesis reactions.

Two types of templates for aptamer-mediated regulation were constructed. For preparation of bubble templates, two different, largely complementary phagemid molecules that correspond to the template and non-template strand were annealed. An unpaired, non-complementary region containing the aptamer structure(s) was inserted downstream of the promoter. Templates that contain an aptamer on both template and non-template strands were poor substrates for transcription. DNA topology is known to have a significant effect on transcription both *in vitro* and *in vivo*. Non-canonical DNA structures such as DNA quadruplexes present an obstacle to transcription from T7 promoters, with the effect being more pronounced when they are located proximal to the transcriptional start site [143-145]. The second template design consisted of a single phagemid product, corresponding to the template strand, and hybridization to a short, synthetic oligonucleotide to create a double stranded T7 promoter. Not surprisingly, expression from single stranded DNA templates were found to be lower than expression from corresponding plasmid templates. This is likely a result of non-canonical DNA structures that can form with ssDNA templates. Interestingly, basal transcriptional levels from bubble templates that harbor the thrombin quadruplex aptamer were much lower than corresponding single stranded DNA regions and from annealed templates lacking secondary structures immediately downstream from the transcription start site.

Amongst the different templates designs examined, the single stranded DNA template with a double stranded T7 promoter was the most responsive to thrombin-mediated repression of gene expression. By contrast, the largely double stranded bubble DNA templates, with aptamer DNA on both strands, did not show any thrombin mediated repression. Fluorescence anisotropy data with short, model DNA templates assembled

from oligonucleotides indicates that thrombin binds both the single stranded DNA template and double stranded templates that contain aptamers in the bubble region. The low expression levels from these latter templates may mask any repression that might be occurring. T7 RNA polymerase is known to bypass gaps and discontinuities in the template strand with the aid of the non-template DNA strand [67, 146]. It is therefore also possible that the presence of an alternative path allows the polymerase to bypass the obstacle posed by thrombin binding.

The specificity of repression to thrombin in comparison to T4 ssDNA binding protein, which binds single stranded DNA non-specifically, was tested. Only limited changes in gene expression upon ssDNA binding protein are observed. This is indicative that the position at which the protein binds to the DNA template is critical for effective ligand based gene repression and that this effect was specific to thrombin. Several previous reports have shown that protein binding at a position proximal to the transcription start site is essential for achieving effective repression from T7 promoters[109, 123, 127]. This suggests that aptamer positioning that facilitates the protein binding close to the transcriptional start site is important for specific control of gene expression. Thrombin mediated repression can be reversed by the addition of thrombin aptamer DNA oligonucleotides. This provides additional support that repression is mediated selectively by thrombin and that the DNA aptamer can bind to thrombin in a complex cell extract system. Further, a mechanism for reversing thrombin mediated gene repression from aptamer templates is possible.

Effective implementation of cell free circuits will require the ability to tune gene expression in response to a ligand. To investigate ways to alter the fold change in

expression levels from T7-aptamer promoters, the thrombin aptamer was placed at different positions relative to the transcriptional start site. We anticipated that the fold change in gene expression would decrease as the aptamer was moved away from the transcription start site. Accordingly, placement of thrombin binding DNA aptamer 2 bases away from the transcriptional start site resulted in up to a 5-fold change in gene expression from ssDNA templates. While moving the aptamer away from the transcription start site increases basal transcriptional levels, only a 1.8 fold change in gene expression upon thrombin addition was observed. T7 promoters are highly conserved in the region between -17 and +6 region while aptamer placement at +2 TBA position disrupts the original promoter sequence, +9TBA and +26TBA templates do not. In addition, thrombin binding aptamer is known to form a DNA quadruplex structure [76, 147] and alternate DNA secondary structures such as DNA quadruplexes [143, 148], and Z DNA[149] sequences can pose a block to transcription *in vitro*. Taken together, these results indicate the disruption of native promoter sequence and the formation of secondary DNA structures close to the transcriptional site might contribute to lower basal expression levels from the +2TBA construct. The finding that operator placement proximal to the transcriptional start site achieves effective repression from T7 aptamer promoters at the expense of lower basal expression mirrors results obtained with T7lacO and T7tetO promoters[109, 123].

Table 2.3 Summary of fold changes in gene expression obtained from synthetic ligand dependent gene regulation strategies in cell free systems.

Mechanism for gene regulation	Fold change	Reference
Triplex DNA	3	[126]
DNA aptamers	5	Present study
Azobenzene mediated photo-regulation	7.6	[150]
T7tetO	10	[123]

Several *in vitro* small molecule/signal sensitive T7 gene regulation systems have been developed [123, 126, 150, 151]. These previous gene regulation systems resulted in 2 to 10 fold changes in gene expression upon the addition of the ligand (Table 2.3). The repression values obtained in this study compare favorably with the most utilized gene regulation systems while offering a strategy for extending the range of signals that can be used to control gene expression. The use of DNA aptamers for transcriptional repression paves the way for creation of cell free feedback circuits with novel sensory capabilities. Selection and use of DNA aptamers that work with known promoter elements will result in new approaches to regulating gene expression in response to a wide range of molecules.

Materials and Methods

Plasmid construction

All the plasmid constructions were carried out using standard techniques [152]. GFP was cloned into pBluescript KS (+) II and pBluescript KS (-) II vector

backbones and the aptamer sequences were then inserted downstream to T7 promoters using inverse PCR. The aptamer constructs are listed in Table 2.1.

Single stranded DNA template preparation

ssDNA templates were assembled by annealing template strands generated from pBluescript KS (-) II variants with T7 promoter oligo in 10mM Tris-HCl pH 7.5 50mM KCl and 1mM $MgCl_2$.

Thrombin aptamer sequence was cloned into the phagemid vectors pBluescript KS II (+) and pBluescript KS II (-) at different locations downstream to transcription start site. These two backbones differ from each other only with respect to the orientation of the F1 origin. ssDNA molecules were then generated using a standard procedure[152]. For the preparation of double stranded templates with a bubble regions, ssDNA generated from ANTKS + DNA was annealed to complementary oligonucleotide that is complementary to a HindIII site on the KS+ DNA backbone prior to digestion with HindIII restriction endonuclease. The resulting template was purified and annealed to the single stranded DNA template derived from pBluescript KS II (-) by slow cooling from 95° C to room temperature in a thermocycler in the presence of 10mM Tris HCl, 50mM KCl and 1 mM $MgCl_2$. The resulting construct contains a mismatch bubble region corresponding to the f1 origin region. The efficiency of annealing and dsDNA generation was verified by digesting the DNA template using restriction endonucleases.

Cell free protein synthesis experiments (CFPS)

The Promega S30 T7 High-Yield Expression System kit (Promega TM306) was used for the CFPS experiments. The S30 premix and the cell extract were mixed in proportions recommended by the manufacturer and 300ng of the template was used per reaction. Reactions were set up following manufacturer's instructions except that the final reaction volume was 15 μL. Reactions were set up in Corning CLS3820 plates. Samples were incubated at 30°C with shaking and measured every 7 minutes in a Biotek Synergy 2 plate reader. Error bars on the florescence measurements represent standard deviation of three replicates.

Thrombin dependent gene repression was tested by incubating the DNA templates with thrombin (diluted into 10mM Tris-HCl pH 7.5 and 50mM KCl) along with 0.01%Tween-20 for one hour at room temperature, followed by the addition of the cell extract. Human α-Thrombin was purchased from Haematologic technologies,VA. To test the effect of different thrombin aptamer and non-specific oligonucleotide concentrations on thrombin mediated repression, different oligonucleotide amounts were heat denatured and slowly cooled to room temperature in the presence of 10mM Tris-HCl pH 7.5, 5mM KCl and 1mM $MgCl_2$ before they were added to cell free protein synthesis reaction. The data in the figure 2.6b is normalized to expression from ATKS templates in the absence of thrombin.

Fluorescence Anisotropy

Fluorescently labeled, gel purified DNA oligonucleotides were purchased from IDTDNA (Corville,IA) The Oligonucleotides were annealed in the presence of

binding buffer (100mM Tris-HCl pH 7.5, 200mM NaCl, 2 mM $MgCl_2$ Fluorescence anisotropy measurements were made in the presence of 10mM Tris HCl pH 7.5, 5mM KCl, 1 mM $MgCl_2$ and 0.02% Tween 20. 10nM DNA templates were incubated with different Thrombin concentrations and anisotropy measurements were made on Biotek Synergy 2 plate reader.

Chapter 3- Engineering Dual regulation of T7 promoter using LacI and TetR repressors

Introduction

Synthetic gene circuits entail redesigning of existing or creating novel genetic function to perform a predetermined task[1]. Construction of these circuits has been invaluable in attaining a bottom up understanding of biological systems and offers potential for harnessing biological function for biotechnology and biomedicine. A library of well-characterized genetic components have been integrated into circuit components that function as logic gates[153], memory elements[4], clocks[5] and counters[154]. Ultimately, like their electronic analogues, components that could be assembled into larger circuits that might find applications in medicine, bioremediation[90] and production of synthetic compounds[155] that are of commercial interest.

These circuits are embedded within cellular systems that comprise of well-characterized components often utilize endogenous promoters and translational machinery to drive circuit function. However, unintended interactions with endogenous processes make implementation of predictable and rationally designed circuits rather difficult. Consequently, several orthogonal expression systems are being sought to insulate the expression of the synthetic gene circuits from biological networks.

One such orthogonal expression systems is the mono-subunit T7 polymerase. T7 RNA polymerase is commonly utilized to drive the expression of genes in cell free systems because of its stability, simplicity and processivity. Furthermore, since T7 polymerase is highly specific for their promoters, its use permits exclusive expression from user-defined genes in a variety of cell cellular backgrounds. In contrast to multi-subunit bacterial RNA polymerases, T7 polymerase recognizes a specific 17 base pair

promoter sequence and does not require any co-factors to activate transcription [67]. For these reasons, T7 promoters have been utilized for achieving expression in cells but are also favored for achieving high yield protein synthesis in cell free systems. Therefore, circuit elements assembled using T7 promoters will be valuable not just in cell based systems for achieving orthogonal expression but can also be used in cell free systems to assemble sophisticated circuits.

Here we focused on constructing logic gates based on T7 promoters. Like their electronic counterparts, logic gates are key circuit components that enable integration of multiple signals and are critical for implementation of sophisticated gene circuits. Consequently, several strategies have been described for achieving logic control of gene expression using bacterial promoters. Most commonly logic control of gene expression is achieved at the transcriptional level wherein a combination of activators and repressors that function individually or act in concert to generate responses from promoters. These promoters facilitate complex gene responses by the action of a multiple transcriptional factors temper transcriptional output to a combination of environmental signals. Indeed, combinatorial promoters have been critical for enabling synthetic devices and motifs such as feed-forward motifs[156] and logic gates[104].

However, unlike bacterial promoters, viral promoters have very few mechanisms for activating or repressing gene expression from T7 promoter. Moreover, T7 promoters are repressed by transcriptional factors such as LacI and TetR that bind a relatively short regulatory region that is proximal to and downstream from the transcription start site [38, 109, 128, 151]. An alternative to cis-regulation is to harness DNA looping to enable regulation of T7 promoters from locations distal to the T7 promoter. DNA looping

mediated by protein dimerization is commonly used to enable regulation of transcription by the synergistic action of repressors bound at two different locations [157]. For instance, Lac repressor proteins (LacI) bind their operators as a tetramer or a dimer of dimers [158]. Native *E.coli* lac promoters contain auxiliary Lac operators (LacO) upstream and downstream to the *E.coli* lac promoter. At low LacI concentrations, presence of additional Lac operators induces a DNA loop formation in the intervening DNA thereby increasing the probability of LacI occupancy of *E.coli* lac promoter [159, 160], which results in enhanced repression from *E.coli* promoters.

Here, a rational design approach to engineer logic control of T7 promoters by harnessing a DNA looping mechanism is described. Existing versions of lac repressible promoters contain LacO downstream from T7 promoters. We hypothesized that as with *E.coli* lac promoters, an improvement in repression of T7lacO promoters could be attained by appropriately a spaced lac operator upstream to a T7lacO promoter. We then examined the effect of placement of tetO, which is the binding site for the TetR protein into this framework to generate T7 promoters that respond to both TetR and LacI (Figure 1). These TetR and LacI repressible T7 promoters were tested in both *E.coli* cells and in cell extracts and constitute the first demonstration of logic control of T7 promoters using two different transcription factors.

Results

Addition of auxiliary operators upstream to T7lacO promoter results increases repression levels from T7 promoters

Existing T7lacO promoters contain the lacO1 operator 4 bases from the transcriptional start site. While presence of this operator represses transcription, it is effective in reducing the basal level expression only upon substantial accumulation of LacI in the cell (Figure 3.2). T7lacO promoters were constructed that utilize an auxiliary lac operator to enable effective repression.

We set out to increase LacI dependent repression of T7lacO promoters by harnessing DNA looping mechanism. A LacO1 operator was placed proximal to and downstream from the transcriptional start site of a T7 promoter. An auxiliary Lac operator LacO1 was placed 92 bases upstream to the primary LacO1 operator (Figure 3.1a) These plasmids were co-transformed into BL21-AI *E.coli* cells along with a pTetRLacI plasmid that expresses tet (TetR) and lac (LacI) repressors from an *E.coli* promoter. The cells were induced with L-arabinose and response of the promoters to lac repressors was measured by monitoring fluorescence changes in response to Isopropyl β-D-1-thiogalactopyranoside (IPTG), which is a negative regulator of Lac repressor.

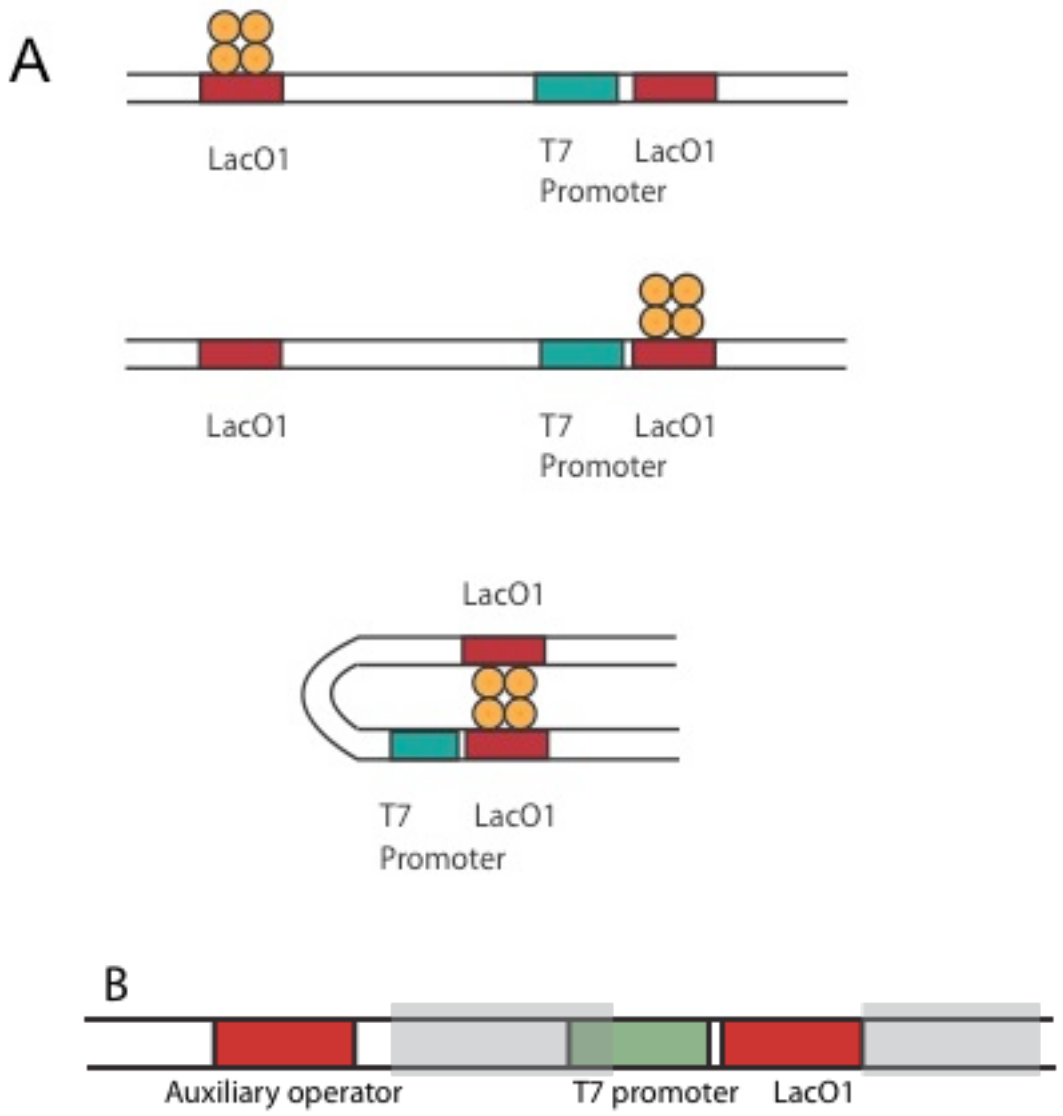


Figure 3.1 Design strategy for achieving combinatorial regulation of expression from T7promoters.A) Auxiliary lacO are placed 92 bases upstream to T7lacO promoters to create strong LacI repressible T7 promoters. The upstream lacO increases repression from T7lacO promoters due to DNA looping. B) TetR binding regions (tetO) are placed within this framework at regions indicated by grey boxes.

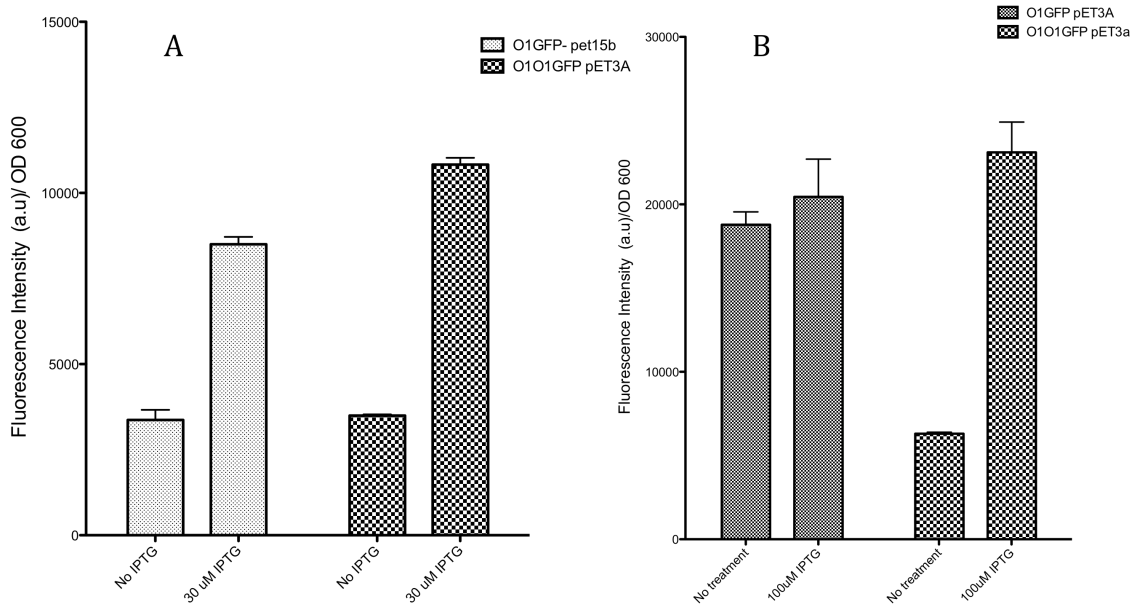


Figure 3.2 Effect of presence of additional lacI gene in the plasmid containing T7lacO promoters A) compares expression from T7lacO promoters in pET15b backbone with O1O1GFP pET3a in the presence and absence of 30μM IPTG. pET15b vector has an additional copy of *lacI* B) Shows expression responses to addition of IPTG from O1GFP and O1O1GFP encoded on pET3a backbones. The graphs indicate fluorescence values normalized to optical density readings.

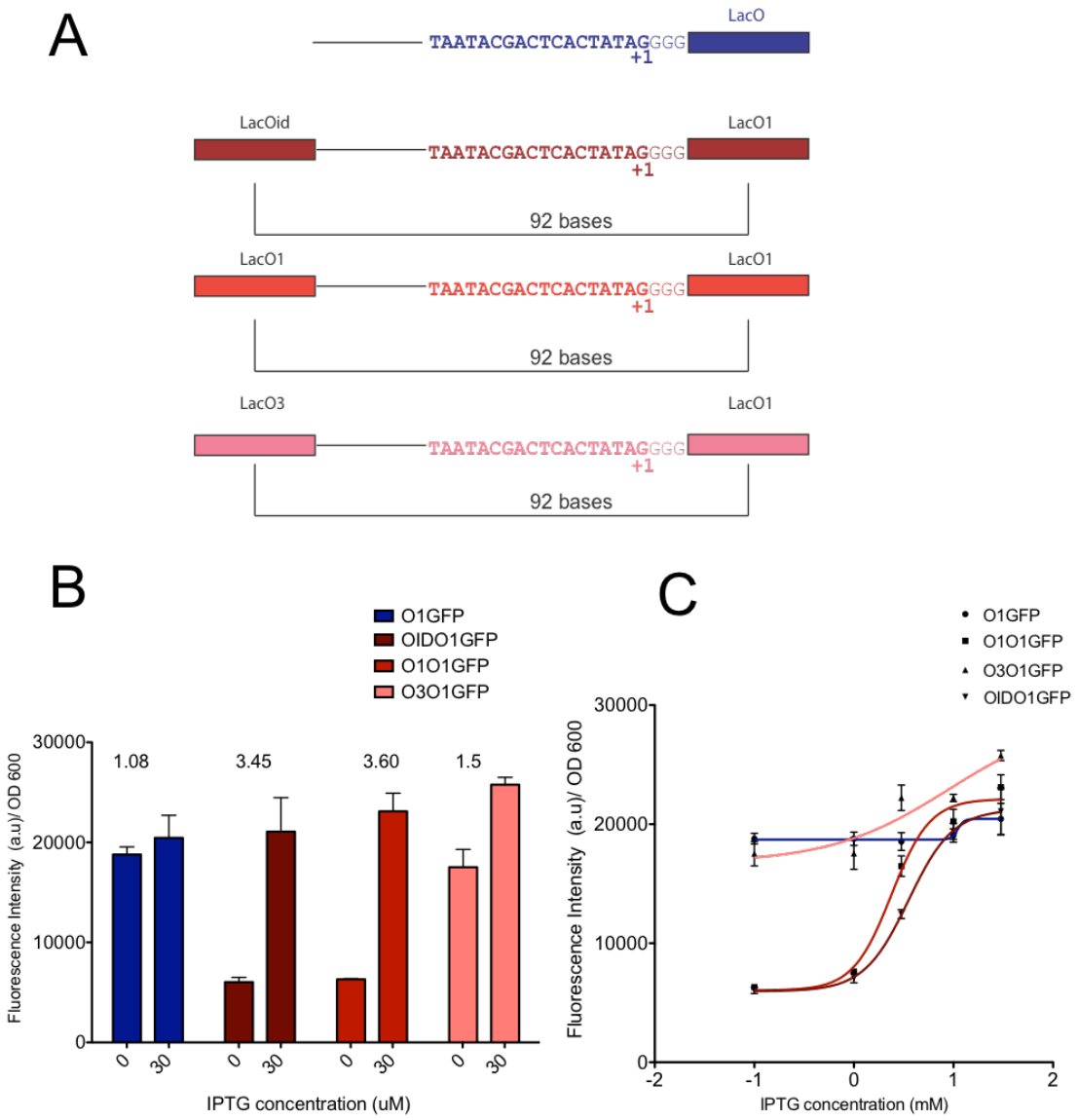


Figure 3.3 Effect of auxiliary operators on LacI mediated repression of T7lacO promoters A) indicates promoter sequences contain T7lacO promoters with auxiliary operator sequences of different strengths. B) Graph depicts responses to 30µM IPTG from the constructs depicted in A). C) Dose response to IPTG from the different constructs. Fluorescence response values are normalized to Optical density values.

The presence of the auxiliary operator was found to increase repression from T7lacO1 promoters (Figure 3.2). In addition, reducing the arabinose concentrations from 0.2% as recommended by the manufacturer, to 0.02% resulted in a greater fold change in gene expression upon L-arabinose addition (Figure 3.4). In the analogous *E.coli* lac regulation scheme, auxiliary operators upstream increase the local concentration of lac repressor around the T7lacO operators and increase the probability of LacI binding to the Lac operator resulting in an increase in repression levels from *E.coli*, lac promoters[159]. Consistent with this hypothesis, weakening the upstream operator by introducing lac operators (LacO3) that bind with lower affinity than lacO1 did not result in improved repression from T7lacO promoters over the control constructs without the auxiliary operators. Overall, these results show that presence of an auxiliary operator that binds with high affinity to the lac repressor 92 bases upstream to the primary operator results in improved repression from T7lacO promoters. (Figure 3.3)

Engineering dual regulation of T7promoters using LacI and TetR

A tet operator site (tetO) that binds tet repressor was inserted into the T7lacO framework. TetO was positioned such that it could potentially interfere or alternatively co-operatively repress along with the LacI dependent repression and therefore result in a multi-input responsive T7 promoter. We tested the effect of placing tet operators at 3 different locations- downstream from the primary lac repressor, at positions interfering with T7 promoters and in between the two lac operators (Figure 3.1b). To test the responses of these dual input promoters, the plasmids were co-transformed with pTetRLacI into BL21-AI cells and fluorescence response to the addition of anhydrotetracycline (aTc) and IPTG was measured.

Table 3.1 List of plasmids used in this study- the subscripts next to the T7 promoter indictes the number of bases that were removed from T7 promoter

Plasmid Name	Promoter	Upstream Operator	Downstream operators	Gene	Backbone
pT7lacOGFP	T7lacO			*EGFP*	pET3a
pLacO1T7O1GFP	T7lacO	LacO1		*EGFP*	pET3a
pLacO3T7O1GFP	T7lacO	LacO3		*EGFP*	pET3a
pLacOIDT7O1GFP	T7lacO	LacOID		*EGFP*	pET3a
pT7lacOtetOGFP	T7lacO		LacO, tetO	*EGFP*	pET3a
pLacO1T7lacOtetOGFP	T7lacO		LacO, tetO	*EGFP*	pET3a
placO1tet21T7lacOGFP	$T7_{-6}$lacO1	LacO1,TetO	LacO	*EGFP*	pET3a
pLacO1tet23T7lacOGFP	$T7_{-4}$lacO1	LacO1,TetO	LacO	*EGFP*	pET3a
pLacO1tet25T7lacOGFP	$T7_{-2}$lacO1	LacO1,TetO	LacO	*EGFP*	pET3a
pLacO1tet27T7lacOGFP	T7lacO1	LacO1,TetO	LacO	*EGFP*	pET3a
placO159tetT7lacOGFP	T7lacO1	LacO1,TetO	LacO	*EGFP*	pET3a
p59tetT7lacOGFP	T7lacO1	TetO	LacO	*EGFP*	pET3a
pTetRLacI	*E.coli promoter*			*TetR ,lacI*	pPROLAR

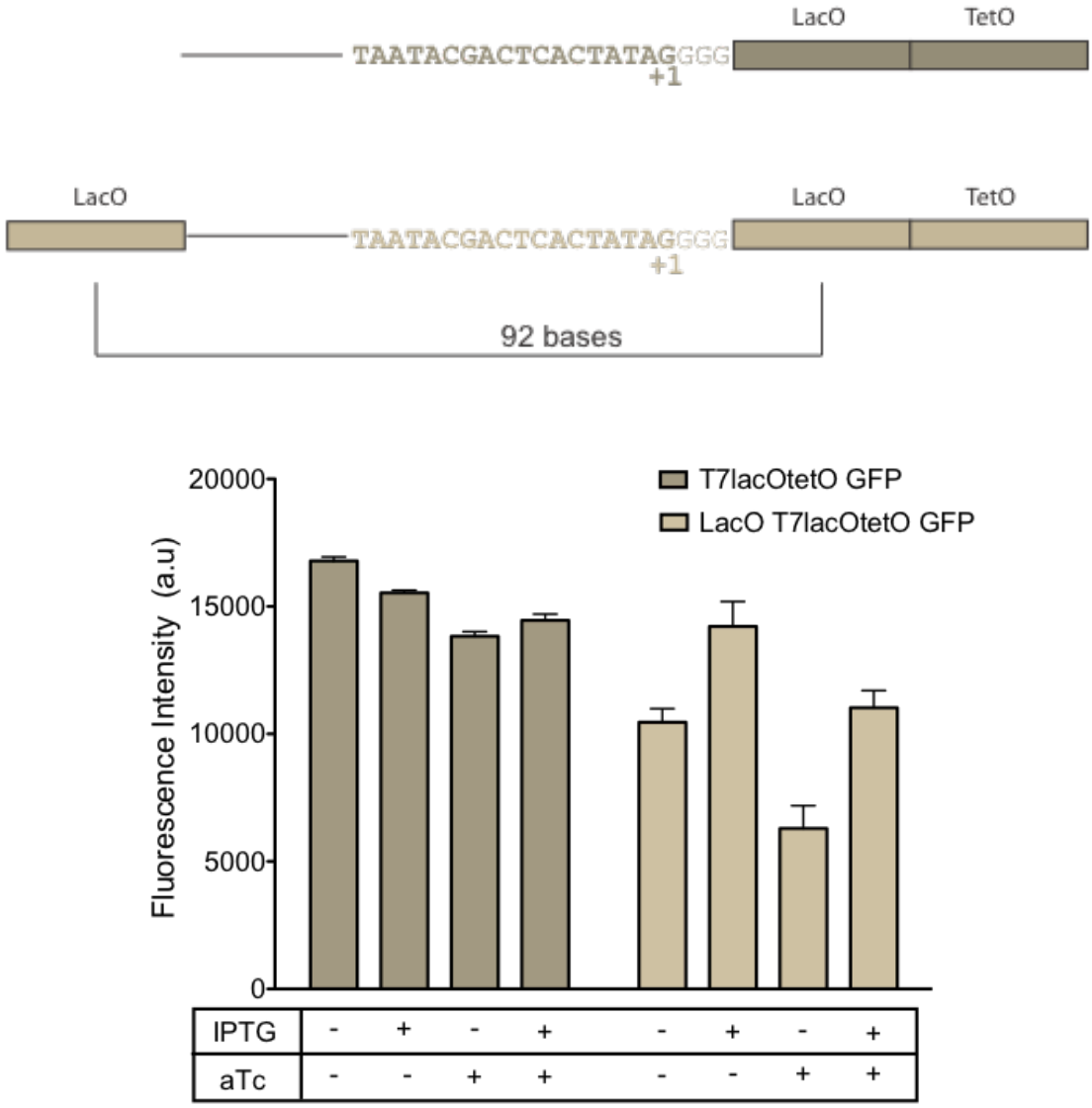


Figure 3.5 Effect of tetO placed downstream from T7lacO promoters. Upper panel depicts the schematic of the pT7lacOtetOGFP and pLacOT7lacOtetOGFP. The graphs indicate the response of pT7lacOtetOGFP and pLacOT7lacOtetOGFP TO 30 μM IPTG and 200ng/ml aTc.

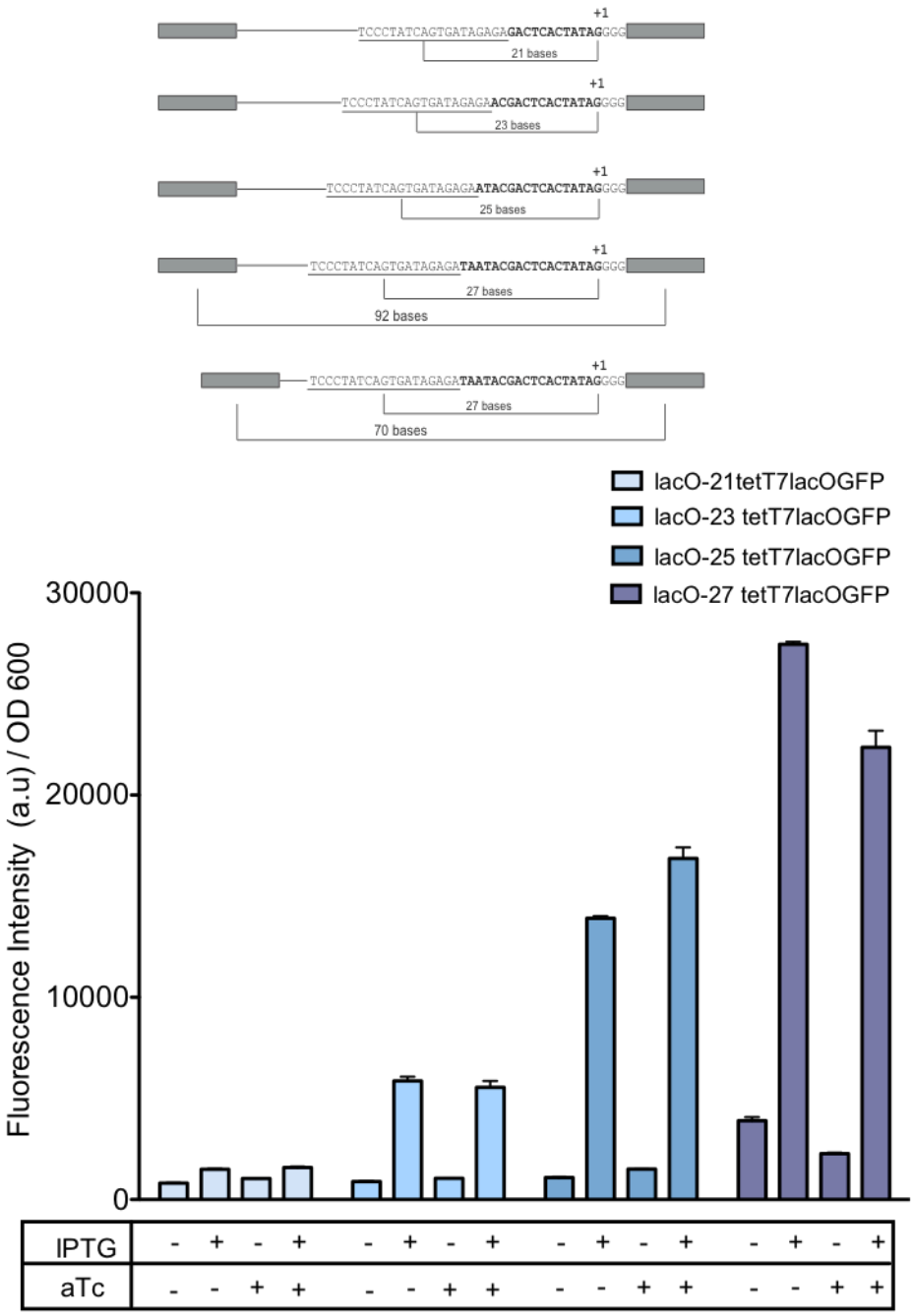


Figure 3.6 Effect of TetR on T7lacO repression when tetO overlaps T7lacO promoters. Upper panel depicts the schematic of the placO1tet21T7lacOGFP, placO1tet23T7lacOGFP, placO1tet25T7lacOGFP and placO1tet27T7lacOGFP. The graphs indicate the response of these plasmids to 30 μM IPTG and 200ng/ml aTc.

Effect of TetR on T7lacO repression when tetO is positioned downstream from T7lacO promoter

The placement of operators that bind to different repressor can elicit a NOR response from the promoter wherein, both repressors independently repress from the promoter[64]. Therefore hypothesized that placing a tetO downstream from T7lacO promoter would lead to LacI and TetR repressible T7 promoters. So to test this hypothesis, the tet operator was positioned 34 bases (distance between +1 base and the center of tetO) and downstream from the transcriptional start site to create placT7lactet. Contrary to our hypothesis, however, we observed that while LacI repressed expression from T7 promoters, the TetR appeared to counter the effect of lac repression instead of repressing expression from T7 promoters (Figure 3.5). By comparison, the control template (pT7lacOtetOGFP) without the auxiliary LacO operator did not show any significant LacI or TetR dependent change in gene expression at 200 minutes. This result indicates that although TetR bound downstream to a lac operator interfered with LacI mediated repression of T7 promoters, it has little to no effect on repression from T7 promoters on its own.

Effect of TetR on T7lacO repression when tetO is overlaps with T7lacO promoter

Several studies show that barring a few key positions on the T7 promoter, several bases can be mutated at the expense of reduced transcriptional output [161]. We therefore examined if replacing a portion of the T7 promoter with tetO would repress transcription. Therefore, tetO were centered at 21, 23, 25 and 27 bases upstream from transcriptional

start site of T7lacO promoters to create placO21tetT7lacOGFP, placO23tetT7lacOGFP, placO25tetT7lacOGFP, and placO27tetT7lacOGFP respectively. While the construct placO27tetT7lacOGFP contains tet operator 27 bases upstream from the transcriptional start site contains an intact T7 promoter, placO21tetT7lacOGFP, placO23tetT7lacOGFP and placO21tetT7lacOGFP contain truncated versions of the T7 promoters with 15, 13 and 11 bases of the wild type T7 promoter remaining, respectively. As expected, we observed that progressively decreasing six bases from the T7 promoter had a significant effect on basal expression levels from T7 promoters. (Figure 3.6) Furthermore, LacI mediated repression was observed for the placO21tetT7lacOGFP, placO23tetT7lacOGFP and placO25tetT7lacOGFP constructs. Intriguingly, however, the tet repressor did not repress expression at -21, -23, -25 and -27 positions. Instead, the tet repressor bound at -27 position, which abuts the T7 promoter, was found to interfere with LacI repression from the T7lacO repressors but did not completely relieve LacI based repression. An additional construct was created from placO27tetT7lacOGFP in which the distance between the lacO was shortened to 70 bases while retaining the tetO at -27 position to yield p70lac27tetT7lacOGFP. Although LacI mediated repression from this construct was found to be stronger, as with p92lac27tetT7lacOGFP TetR negatively regulated LacI mediated repression.

TetR bound to tetO placed in between the Lac operators effectively alleviates Lac mediated repression

We asked if placing a tet operator between the two lac operators and in phase with the lacO loop that might form relieve LacI mediated transcriptional repression. Therefore,

the tetO was placed 59 bases upstream to the primary lac operator in placO1T7lacGFP to create pLacO159tetT7lacOGFP. As a control, the tet operator was introduced 59 bases upstream to T7lacO operator in pT7lacO which does not have the auxiliary lac operator. Protein expression as measured by fluorescence at 200 minutes showed that the tet repressor at this position effectively interferes with LacI repression and presence of TetR repressor at the operator site relieved LacI dependent repression (Figure 3.7) We further tested the response of this construct in response to a wide range of aTc and IPTG concentrations. The plasmid exhibited an IMPLIES gate, wherein the tet repressor hinders LacI loop mediated repression.

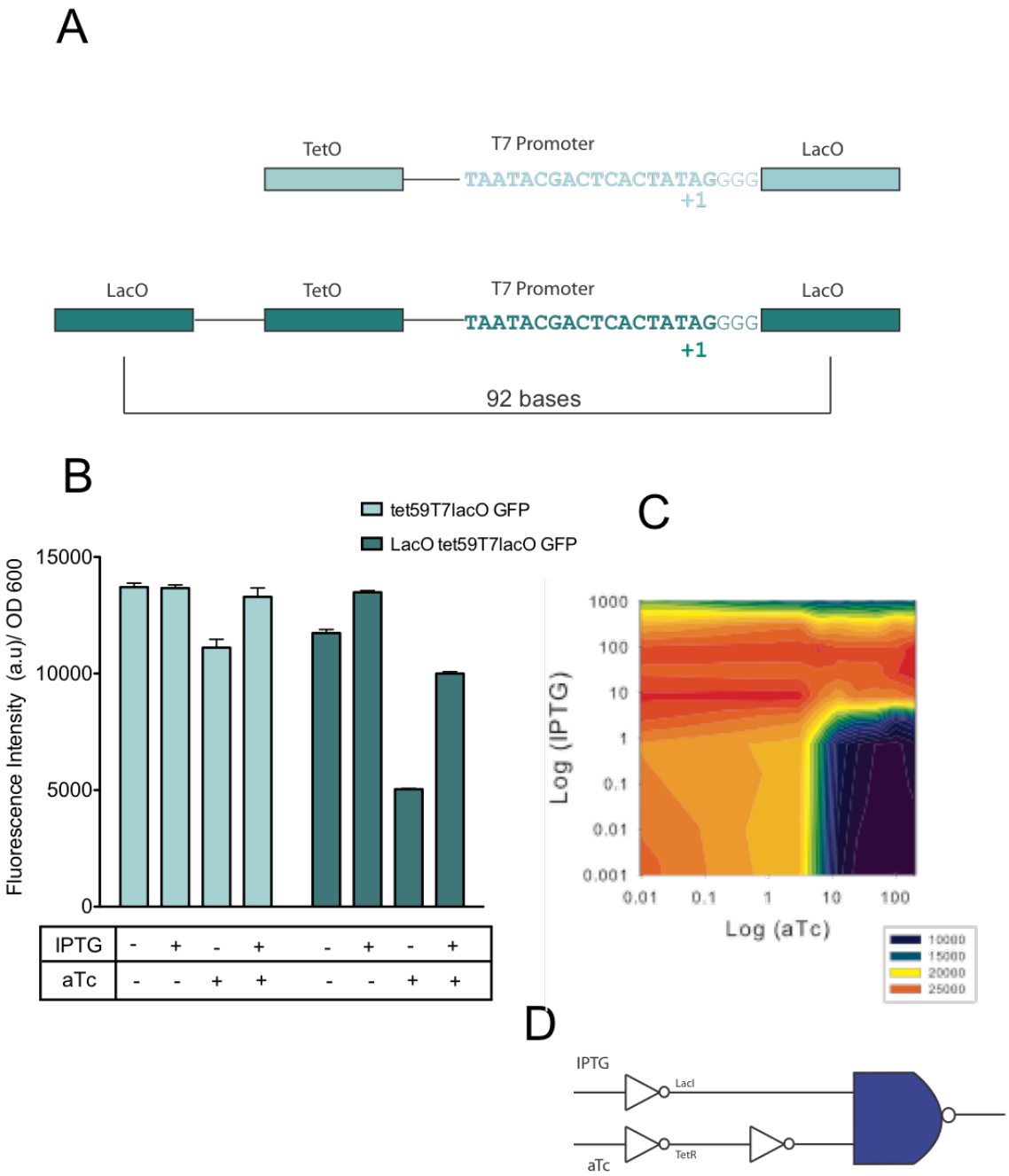


Figure 3.7 –Effect of tetO on LacI mediated repression of T7lacO when tetO was in between the two lac operators- A) Depicts the schematic of p59tetOT7lacOGFP and pLacO59tetOT7lacOGFP B) Graph indicates the responses of these plasmids to IPTG and aTc C) Contour plot indicates the Normalized Fluorescence responses to a range of IPTG and aTc concentrations. The X axis indicates the logarithm of aTc concentrations (ng/ML) and whereas Y-axis contains the Log of IPTG concentrations (µM) Fluorescence measurements in B and C were normalized to the optical density measurements at 600 nm D) is a schematic of the IMPLIES logic gate realized using the pLacO59tetOT7lacOGFP plasmid.

Testing dual input promoters in cell free systems

The plac59tetT7lacOGFP constructs were tested in commercial extracts to test their response to LacI and TetR repressors. The extracts were chosen since they contained a cache of LacI protein. To enable tetR-mediated repression, purified TetR protein was added to the reaction and the response of the multi-input promoters was measured by monitoring fluorescence changes upon the addition of IPTG and aTc.

The responses of the pLacO159tetT7lacOGFP and p59tetT7lacOGFP (control without the auxiliary operator) to IPTG indicate that auxiliary operators improve LacI dependent repression by about 8 fold at the concentrations of LacI and template concentrations tested here (Figure3.8). Furthermore, TetR was found to relieve LacI dependent repression and expression levels were found to be similar to those from the constructs lacking the auxiliary operators. This provides support to the fact that the presence of an upstream operator improves LacI dependent repression presumably through a looping based mechanism. In addition, similar expression levels from pLacO159tetT7lacOGFP and p59tetT7lacOGFP in the absence of inducers implies that TetR bound in between the lac operators eliminates the advantage conferred by the upstream operator.

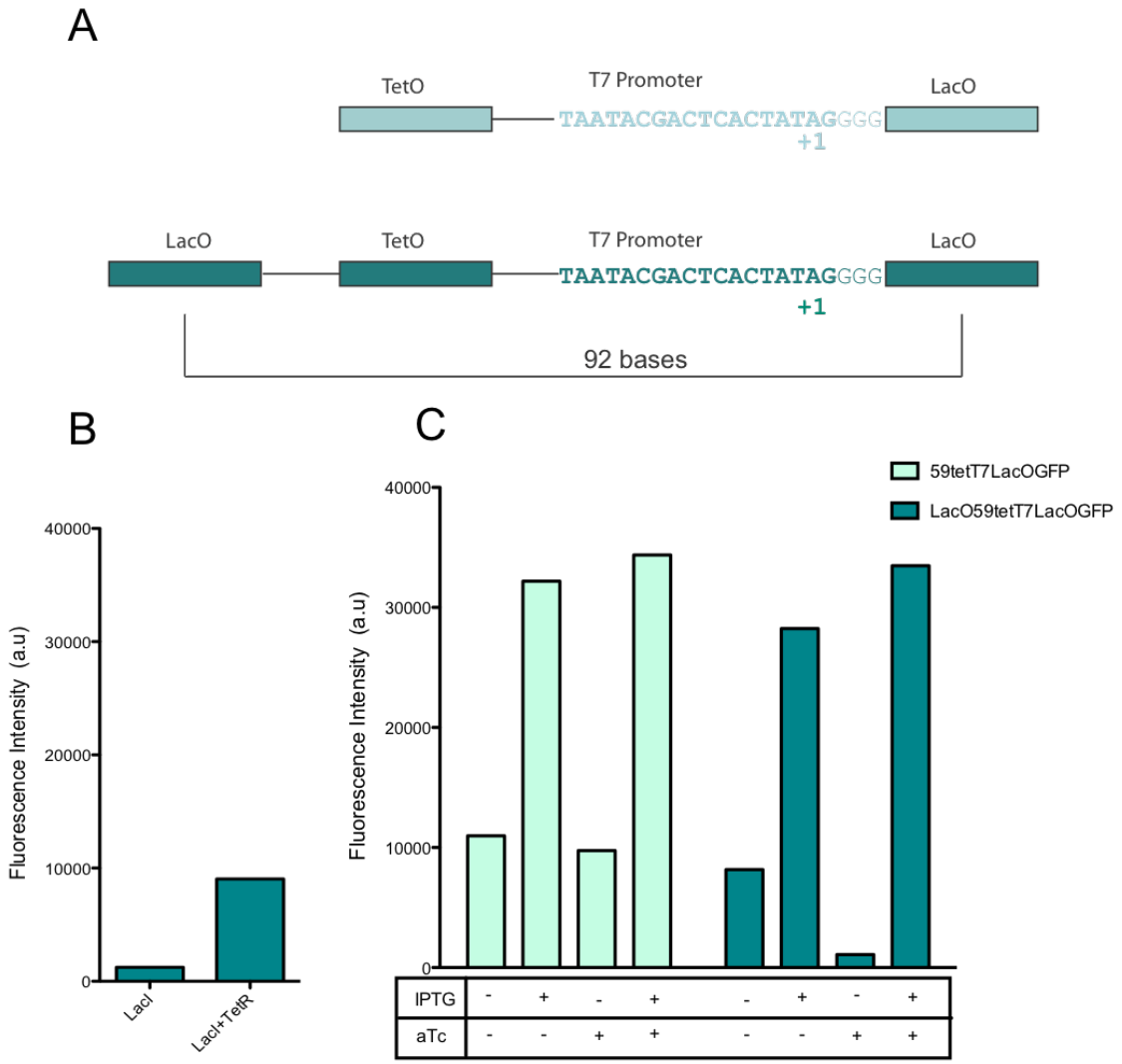


Figure 3.8 Effect of tetO on LacI dependent repression of T7lacO when placed between two lac operators in cell extracts- A) Depicts the schematic of p59tetOT7lacOGFP and pLacO59tetOT7lacOGFP plasmids B) Graph showing fluorescence response from pLacO59tetOT7lacOGFP LacI and TetR proteins. C) Shows fluorescence response from p59tetOT7lacOGFP and pLacO59tetOT7lacOGFP plasmids to addition of 300μM IPTG and 200ng/ml aTc

Discussion

Several studies have outlined the relationship between promoter architecture and function [7, 8, 162] using a combination of well-characterized activators and repressors with bacterial promoters [7, 8]. These studies rely on the fact that bacterial promoters respond to transcriptional activators and repressors that bind cis-regulatory regions that span at least 100 bp upstream and downstream to the promoter [62, 107]. Consequently, a broad array of regulation mechanisms can be incorporated that operate either individually or in concert to create programmable multiple input responsive promoters. Varying the location of the transcriptional factor binding sites relative to the promoter, tuning the affinity of transcriptional factors to their respective operator sites and controlling the interaction between the transcriptional factors results in a wide variety of combinatorial responses with bacterial systems[7, 8]. As an illustration of the power of this approach to realize diverse functions, Hunziker *et.al* designed 12 different types of logic functions using a combination of cAMP-CRP activator protein, GalR repressor protein and promoters of diverse strengths [163]. Whereas several reports have attempted to define rules for achieving logic control by modulating cis-regulatory regions around bacterial promoters, similar systems have not yet been established for use with viral promoters such as T7 promoters. This paper described efforts to engineer an IMPLIES gate in live cells and in cell free systems using T7 promoters.

A difficulty with engineering T7 promoters that respond to multiple transcription factors arises from the fact that T7 polymerase has no known recruiters or transcriptional activators and are repressed only by proteins that bind at a location close to the

transcriptional start site. To circumvent this limitation, a DNA looping based mechanism was harnessed in order to achieve gene regulation from distal locations. Specifically, a LacO1 auxiliary operator was placed 92 bases upstream from the primary LacO1 operator to enable strong repression from the T7lacO promoter. Within this framework the effects of placing a TetR binding site at three sets of locations relative to the T7 promoters - downstream from T7lacO, region overlapping the T7lacO promoter and in between the two lac operators, were examined.

Protein mediated DNA looping has been utilized for achieving transcriptional control in both natural and synthetic promoters. In particular, the role of DNA looping in the regulation of promoters such as *lac*[159, 164], *araBAD*[165] *and gal*[166] has been extensively studied in prokaryotes[167]. Zhan *et.al* utilized two different LacO binding sites to introduce looping thereby enabling regulation by the concerted action of two different repressors bound at different sites[103].

Previous demonstrations of LacI dependent repression of T7 promoters have relied on LacI binding to a location downstream to the transcriptional start site[109, 151]. While LacI enhances repression from T7 lacO promoters, efficient repression requires sufficient build up of lac repressor protein in the cell. In addition, pTetRLacI is expressed from an *E.coli* promoter that is weaker than the T7 promoter that drives the expression of the reporter gene. The high processivity of T7 polymerase generates a large amount of RNA transcript that saturates the translational machinery[66] and therefore masks repression that might occur at later time points.

To develop lac repressible T7 promoters that are strongly repressed at low LacI concentration, a DNA looping based mechanism was utilized. In native *E.coli* lac

promoter systems, looping facilitates tight repression of *E.coli* lac promoters even at low LacI repressor concentrations[168]. Our results mirror the effect of auxiliary operators on expression from *E.coli* lac promoters whereby, as shown in Figure 3.2, introduction of an auxiliary lac mediated operator 92 bases upstream to the primary lacI binding site resulted in a 4 fold increase in LacI repression in comparison to T7lacO promoters lacking the auxiliary distal operator. Additionally, changing the auxiliary operator strength facilitated the tuning of repression levels from T7lacO promoters. In contrast to these results, Dubendorff and Studier observed that an auxiliary operator placed 238 bp away from T7lacO promoter resulted only a modest increase in repression from T7lacO promoter[151]. The difference between their study and findings described here can be reconciled from that fact that the distance and the phase difference between lac operators is critical for achieving LacI mediated repression [168]. Muller *et.al* observed that a substantial decrease in repression levels accompanied an increase in operator distance. A 50-fold change in gene expression was observed at an inter-operator distance of 70.5, while only a 15-fold change in expression was found where the distance was 150 bp. Thus an auxiliary operator located a distance of 238bp might be too far to increase repression from T7lacO promoters. Moreover, at shorter distances, the energy required for formation of a LacI dependent loop between operators located on the same side of the DNA helix is lower than the energy required for formation of a loop between operators that are on opposite phases[160, 168]. Therefore, the interoperator distances of 92 and 72 were chosen based on *in vivo* data for lac mediated transcriptional control with *E.coli* promoters and lead to a significant increase in repression levels from T7lacO promoters.

These redesigned operators were then utilized in conjunction with a tet operator to realize T7 promoters that can be regulated by multiple signals. As has been postulated in several studies, the nature of the promoter, placement of repressors and activators relative to the transcriptional start site and the nature of interaction between the different proteins affect the response from promoters [64, 169]. A simple design to achieve repression from two transcriptional factors involves placement of two transcriptional repressors downstream from the transcriptional start site. Binding of two strong repressors downstream from the transcriptional start site of bacterial promoters is known to result in repression from both repressors. In fact, a similar design with T7 promoters regulated by two distinct zinc finger proteins that bind contiguous binding sites has been utilized to realize a NOR gate with T7 promoters[53]. In a sharp contrast to these studies, however, upon simultaneous binding of LacI and TetR repressors downstream from the initiation, tet repressor bound downstream from lac operator interfered with LacI dependent repression instead of repressing expression from T7 promoters. These observations can be ascribed to the higher processivity of T7 promoters in comparison to *E.coli* promoters. Repressors placed several bases downstream from transcriptional start site of a T7 promoter are known to be weaker repressors of transcription[38] thereby placing an upper limit on the region downstream from the T7promoter that regulates expression from it. In contrast to short 9 bases zinc finger operators, the 19 and 21 base length tetO and lacO pushes the secondary operator further downstream from the transcriptional start site thereby reducing the efficiency of repression from the secondary operator[53]. In fact, binding of a strong protein such as TetR downstream from lacO affects the ability of LacI to bind to its cognate operator.

Our finding that interruption of the T7 promoters with tetO did not yield tet regulatable promoters was surprising. Only in the case of positioning tetO centered at -25 site did we observe a slight effect of TetR. Interruption of the binding region of the T7 promoter is thought to facilitate promoter release during transcriptional initiation thereby increasing transcriptional efficiency. Therefore, any tet repression that may have occurred might be offset by the higher transcriptional output from the truncated promoters.

The high processivity of T7 RNA polymerase makes it difficult to maintain tight repression. Data from the constructs containing interrupted promoters suggest that mutations in the promoter region result in a tighter LacI dependent repression of expression from T7lacO promoters. At later time points, expression from placO1tet21T7lacOGFP, placO1tet23T7lacOGFP and placO1tet25T7lacOGFP, is effectively repressed in the absence of IPTG, whereas expression from placO1tet27T7lacOGFP with the intact promoter sequence shows a steady rise in fluorescence even in the absence of IPTG at later time points (Figure 3.8). Therefore, it is likely that the apparent lack of repression observed with pT7lacOGFP plasmids is a reflection of weak repression of T7lacO promoters and the resulting accumulation of mRNA at low LacI concentrations.

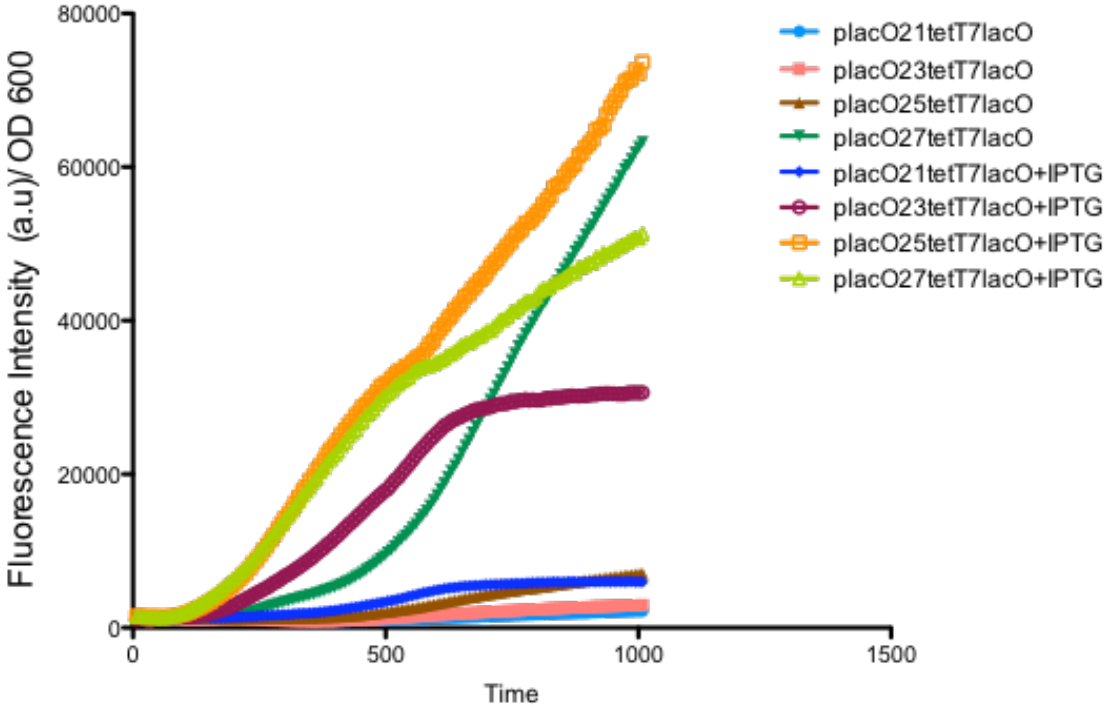


Figure 3.8 Time course of expression from interrupted T7promoters. The graph depicts change in fluorescence from truncated T7lacO promoters in response to 30 μM IPTG.

Finally, we were able to demonstrate that placement of a TetR repressor at a region between the two lac operators interferes with LacI dependent repression of the T7lacO promoter. Participation of protein tetramers for achieving repression contributes to the steepness of the response to IPTG. This resulted in a logic behavior characterized by sharp transition between the on and the off states. At short distances, stable formation of lac repressor mediated looping requires that the operators be in phase as significant energy is expended for the formation of the loop. TetR was placed in between lac operators and found to effectively interfere with the mediated repression of T7 promoters. One scenario that would explain this observation is that binding of the tet repressor may stiffen the DNA template thereby increasing the persistence length of the DNA and making the formation of the DNA loop energetically unfavorable. Alternatively, it is possible that the tet repressor sterically hinders the formation of the LacI based loop. In the case of *E.coli* promoters, especially at short distances, periodic dependence on intervening distances between the lac operators has been observed with a repression maxima for distances that place the operators in phase and minima for those on the opposite sides of DNA. Repression maxima have been shown to occur with a periodicity of 11.5 bp with *E.coli* promoters of 59 bases, 70.5, 81.5 and 92 bases. Based on this model tetO centered 59 base upstream to the primary lac operator would be in phase with the primary lac operator and may sterically hinder lac mediated loop formation. In any case, binding to the tet operator at a distal site interfered with LacI dependent repression and enabled an IMPLIES function.

The T7 promoter regulated by both lac and tet repressor were finally tested in cell extracts and constitutes the first demonstration of an IMPLIES gate in *E.coli* cell free

systems using T7 promoters. Cell free systems are ideal test-beds for implementing simple regulatory circuits and as sensors of environmental signals[122]. Cell free extract combine the simplicity of an *in vitro* system along with the remarkable capability for continuous protein production from DNA encoded instructions for enabling synthetic gene circuits[47]. As mentioned earlier, the high processivity of T7 polymerase is ideal for cell free applications and has been used in conjunction with *E.coli* promoters and the SP6 promoter to realize cascading networks in cell free systems[48, 52]. The lac and tet reguatable promoters described here would be valuable tool for assembling cell free gene circuits[51].

In conclusion, LacI mediated looping has been shown to increase repression from T7lacO promoters. In addition, binding of TetR protein in between the lac operators interferes with LacI dependent looping to realize an IMPLIES function for T7 dependent protein expression. This study paves the way for introducing modular transactivating domains with the help of additional transcription factors. Temme *et.al* recently reported the development of set of orthogonal a set of T7 RNA polymerase- promoter pairs[25]. The strategy put forth here combined with the availability of promoters of varied strength, provides the opportunity to harness the portability of T7 promoters for realizing networks in cells and in cell free systems.

Materials and Methods

Plasmids and Bacterial strains

All plasmids used in this study were constructed using standard molecular biology techniques. The plasmids constructed and used in this study are listed in Table 3.1 .DNA

used in cell free experiments was prepared using Qiagen Plasmid Midi prep kits or Biorad midi prep kits. *E. coli* strain BL21-AI (Invitrogen Inc, WI) was used for protein purification and for live cell expression experiments. LB media with 100 μg/mL ampicillin was used to culture cells for protein purification and preparation of starter cultures for live cell experiments. Minimal media for testing plasmids has the following composition- M9 salts with Casamino acids (Amresco), 2 mM $MgSO_4$, 0.5% glycerol, 300 μM thiamine, and 100 μg/mL ampicillin.

Purification of TetR

TetR was purified as previously described. Briefly, BL21-AI *E.coli* strains (Invitrogen Inc, WI) harboring pET-TetRHis[38] was grown in LB media with 100 μg/mL ampicillin at 37°C and were induced using 0.2% L Arabinose. The cells were resuspended in binding buffer (50 mM Sodium Phosphate buffer pH 8.0 300, mM NaCl, 10 mM Imidazole) and lysed by sonication. The supernatant obtained after centrifugation of the samples was applied to a Ni-NTA column. The column was subsequently washed with buffer (50mM Sodium Phosphate buffer pH 8.0, 300mM NaCl, 50 mM imidazole). TetR-His6 was then eluted with elution buffer (50mM Sodium Phosphate buffer pH 8.0, 300 mM NaCl, 500mM imidazole). Finally, the purified protein was concentrated and dialysed into the final storage buffer (20mM sodium phosphate pH 7.2, 50 mM NaCl).

GFP measurements from E. coli experiments

The GFP expressing plasmids bearing different T7lacO and tet operator regions were co-transformed along with pTetRLacI plasmids into BL21-AI cells. A single colony from

the transformation plate was used to initiate an overnight culture in LB media. A small aliquot of overnight culture was then transferred into M9 minimal media (M9 salts with Casamino acids (Amresco), 2 mM $MgSO_4$, 0.5% glycerol, 300 μM thiamine) supplemented with the 100ug/mlAmpicillin and 50ug/ml Kanamycin. The culture was incubated at 37°C for about 5 hours before this starter culture was again diluted in M9 media to a final Optical Density of 0.01. 0.02% L-arabinose was added to the culture to induce the expression of T7 RNA polymerase. 100 ul aliquots of culture were dispensed into a 96-well plate (Corning 3370). Subsequently, IPTG and aTc were added to the wells as indicated. 50 μl of mineral oil was added to each of these wells to prevent drying of the samples. Absorbance (at 600 nm wavelength) and fluorescence measurements (485/20 nm, emission was 528/20 nm) were made at intervals of 7 minutes. Fluorescence values were corrected for background fluorescence of the media, and absorbance readings at 600 nm were used to normalize for cell density.

Cell-free expression experiments

Qiagen cell free protein synthesis kits were used to carry out cell free protein synthesis reactions. Reactions were set up following the manufacturer's instructions, but the final reaction volume was reduced to 15 μL .The reactions were overlaid with 10 μL mineral oil was added to prevent drying. 300μM IPTG and 200ng/ul aTc were added to the reactions to induce expression. Reactions were set up in Corning CLS3820 plates. Fluorescence measurements were made at an interval of 7 minutes in a Biotek Synergy 2 plate reader. The fluorescence units shown the Figure 3.7 represent values obtained after 6 hours.

Chapter 4 -Implementing Negative Feedback in Cell free systems using T7lacO promoters

(This chapter contains excerpts from the paper "Expression optimization and synthetic gene networks in cell-free systems" by David K. Karig, **Sukanya Iyer**, Michael L.Simpson and Mitchel J.Doktycz)

Introduction

The field of synthetic biology, which aims to forward engineer biological systems, offers a tremendous opportunity to harness biological function for biomedical and biotechnological applications[68]. Synthetic gene circuits that perform a predetermined operation in response to external input have been successfully enabled in cells. These circuits have been used to program cells to integrate multiple input using logic gates [153] , encode memory using bi-stable switches[4] and program periodic behavior by incorporating oscillators[5]. Consequently, engineered gene circuits have found extensive applications in bio-sensing, bioremediation and biomedicine[68]. However, undesired crosstalk between the synthetic and the host networks make engineering gene circuits in live cells very challenging. Furthermore, use of living cells precludes sensing and production of compounds that might be toxic to living cells.

Cell free systems offer a unique opportunity to tap transcriptional and translational capabilities for applied uses. Logical behavior and the controlled synthesis of commercially important bio-molecules can be enabled [120]. The open nature of cell free systems and the ability to synthesize and process bio-molecules, without concern for cell viability, present significant advantages over natural cells for the synthesis of potentially toxic bio-molecules. Furthermore, the ability to exclusively direct the biosynthesis machinery of cell free systems for the desired application, without interference from the

host machinery, can facilitate implementation of predictable, engineered systems. Furthermore, the user has precise control over the components that comprise a system thus making *in vitro* systems ideal for implementing predictable networks for diagnostics and for testing hypotheses about phenomena in isolation from the host system. Cell free systems have been utilized for realizing cascading systems [48, 49], for demonstrating pattern formation in cell extracts [170] and for creating logic gates[52]. In addition, flexibility and the open nature of cell free systems can fast track testing of genetic components before being transferred to cell based systems. Implementation of networks in cell free systems serves as a stepping-stone for realizing complex functions.

Gene circuits programmed to elicit a particular response impinges on attaining a proper balance between the RNA and protein components[3, 171]. As a starting point for construction of gene circuits, we selected the simple yet important negative autoregulatory circuit as a starting point for our investigations[2]. In a negative autoregulatory motif, transcriptional factor represses its own expression. The ability of these motifs to accelerate circuit response and reduce gene expression noise makes them important components of natural and synthetic gene circuits[5, 11, 12]. Given the importance and the simplicity of this system, construction of a negative autoregulatory circuit would be useful not only for assembling increasingly complex gene circuits but also highlight the parameters that can be tuned for encoding functionality in cell extracts.

Karig *et.al* have implemented a negative autoregulatory circuit that utilizes TetR repressible T7 promoters for enabling a negative feedback circuit[38]. The authors showed that tuning genetic determinants for expression such as promoter strength, ribosome binding site and terminator strength that can be used to program negative

feedback response that demonstrates a desired level of expression. Here we use a complementary approach for enabling negative autoregulatory circuits using a LacI repressible T7 promoter (Figure 4.1). Because LacI is a weaker repressor of the T7 promoter, different determinants for system assembly that would facilitate the realization of negative feedback in cell free systems were examined. Two different variants of T7lacO promoters with and without auxiliary operators (that have been described and characterized in detail in chapter 3), were utilized to demonstrate negative feedback in cell free systems. As a step towards assembling circuits in cell free systems, the expression outputs between multi-cistronic systems and multi-plasmid systems were compared. We then examined repression of T7lacO promoters by LacI encoded on separate plasmids. Finally, a negative feedback circuit expressing lacI and GFP from T7lacO promoters were implemented.

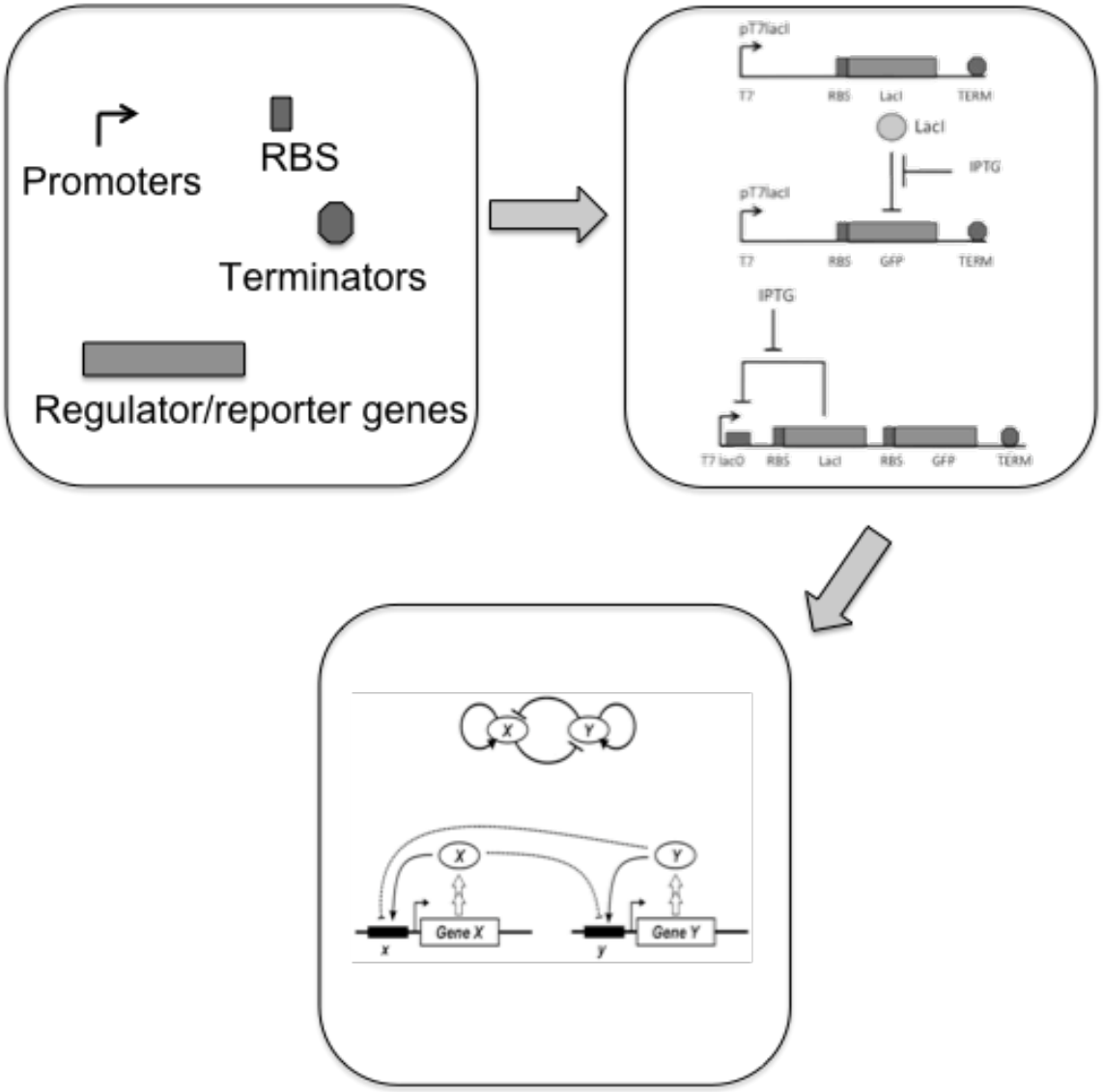


Figure 4.1 Approach for assembling synthetic gene circuits in cell free systems- Construction of gene circuits requires the availability of well-characterized genetic parts such as ligand sensitive promoters, ribosome binding sites. These components are then assembled into small recurring network motifs such as negative feedback circuits and positive feedback circuits. An understanding of how to assemble these networks would be critical to construction of more sophisticated circuits such as bi-stable switches in cell free systems.

Results

System composition

As a step towards creating larger and more complex synthetic systems, we explored the effect of different system composition approaches. One system composition approach, which is commonly employed in living cells, involves the use of multi-cistronic sequences for co-regulating subsets of genes in the system. As a simple investigation into expression efficiency in multicistronic sequences, two bicistronic sequences were constructed. The first, placI-GFP consists of *lacI* inserted upstream of GFP and the second pGFP-lacI consists of *lacI* inserted downstream of GFP (Figure 4.2a). As expected, fluorescence measurements for all of the bicistronic sequences were significantly lower than for the pKSGFP control. (Figure 4.2b) While fluorescence of the lacI bi-cistronic constructs were approximately 25% lower than pKSGFP, no significant difference was observed between the insertions of *lacI* upstream vs. downstream of GFP.

Table 4.1- Monocistronic plasmids used in this study.

Plasmid Name	Promoter	Gene1	Backbone
pKSGFP	T7	*EGFP*	pET3a
pT7LacI	T7	*LacI*	pET3a
pT7LacOGFP	T7LacO	*EGFP*	pET3a

An alternative to multi-cistronic sequences is to use a separate plasmid for each gene in the system. As later discussed, this approach is particularly amenable to cell-free systems. However, it is important to understand the effect of altering the concentrations of the different plasmids on expression. To this end, we combined a plasmid expressing only GFP (pKSGFP) and a plasmid expressing only LacI (pT7lacI) in different molar ratios, while keeping the sum of the molar concentrations fixed at 8 nM. The resulting GFP fluorescence exhibited a non-linear increase as a function of the percentage of the GFP plasmid. Specifically, the effect of increasing the percentage of pKSGFP grew more pronounced at higher pKSGFP percentages. (Figure 4.2c)

Table 4.2 List of bi-cistronic plasmids used in this chapter.

Plasmid Name	**Promoter**	**Gene 1**	**Gene2**	**Backbone**
placIGFP	T7	*LacI*	*EGFP*	pBluescriptKS(+) II
pGFPLacI	T7	*GFP*	*LacI*	pET3a
pLacOLacIGFP	T7LacO1	*LacI*	*EGFP*	pET3a
pLacOT7LacOLacIGFP	92LacO1T7LacO1	*LacI*	*EGFP*	pET3a

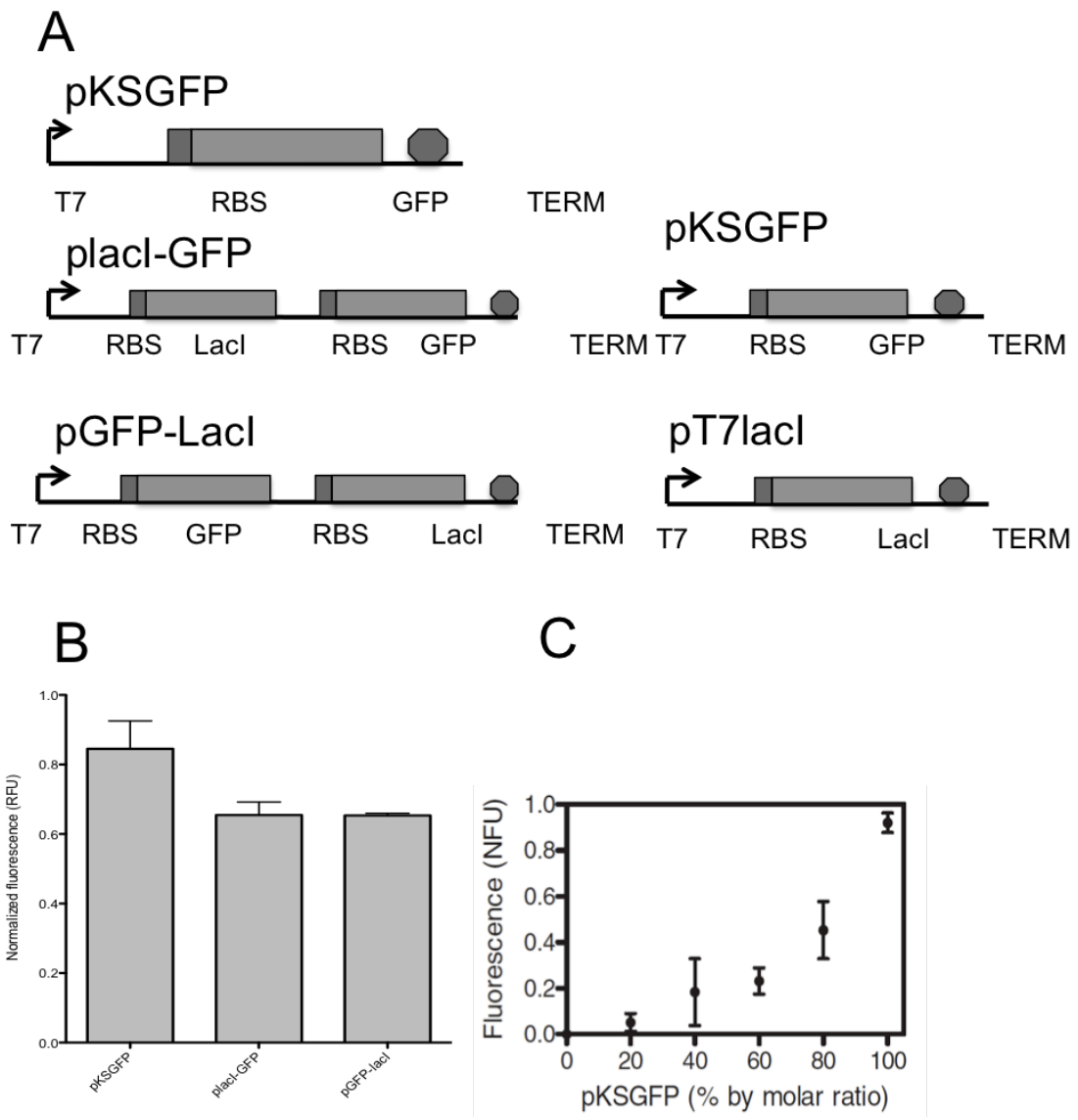


Figure 4.2 Plasmids and results for exploring different system composition approaches. a) Constitutive T7 construct pKSGFP, pT7lacI and bicistronic constructs placI-GFP, and pGFP-lacI. b) Fluorescence after 10 hours of EGFP expression from these constructs. c Results for co-expression of pKSGFP and pT7lacI for different percentages of pKSGFP by molar concentration.

To compare GFP expression for the two different system composition approaches, we expressed the bicistronic constructs pLacI-GFP and pGFP-LacI at the concentrations shown in figure 4.3, and we also co-expressed pKSGFP and pLacI such that the molar concentrations of each plasmid were also as shown in the figure. Thus each concentration on the x-axis corresponds to the same number of copies of the *EGFP* and *lacI* genes for each approach. Lower EGFP expression, as measured by fluorescence after 10 hours of expression, was realized with the two-plasmid approach.

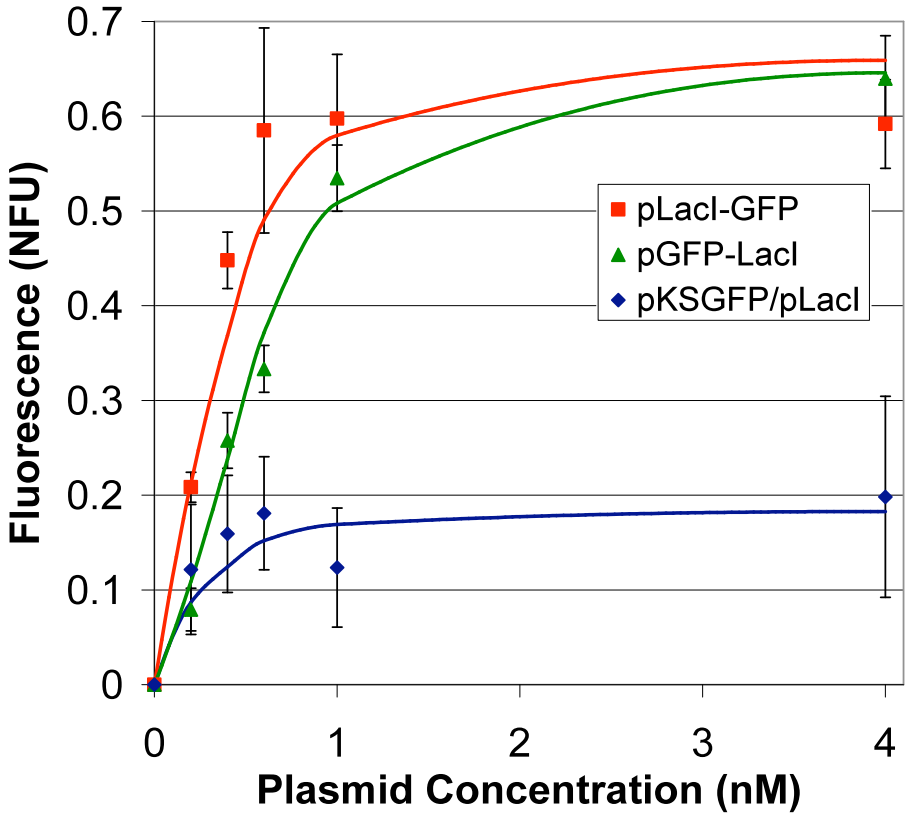


Figure 4.3 Comparison of bicistronic and two-plasmid systems. Normalized fluorescence after 10 hours of expression is shown for the bicistronic constructs pLacI-GFP and pGFP-lacI, along with the two-plasmid system pKSGFP/pLacI.

Repression of T7lacO promoters by lacI encoded on different plasmids

Effective negative regulation of a promoter by a repressor expressed in the same milieu depends on factors such as relative transcription and translation efficiency of the repressor and reporter gene, affinity of the promoter for the repressor and efficiency of repression by the bound repressor[25, 172]. Here, we adopted a multi-plasmid approach to vary relative gene copy number instead of varying genetic determinants of expression efficiency such as ribosome binding site [173] and terminators to control protein amounts in the extract. Accordingly, a repressor cascade was implemented using two separate plasmids –pT7lacI that expresses LacI from a T7 promoter and pT7lacOGFP in which T7lacO promoters drive the reporter gene GFP expression (Figure 4.4a) . We hypothesized that addition of plasmid expressing LacI in excess of T7lacOGFP might ensure that lac repressor was present in sufficient amounts in the extract. Accordingly, 200ng pT7lacI along with 80 ng of pT7lacOGFP was added to the cell extract and response to 1mM IPTG that binds and relieves lac repression was measured. Intriguingly, only a 1.2 fold increase in fluorescence was observed upon IPTG addition when the pT7lacI and pT7lacOGFP were expressed simultaneously, even though pT7lacI was present in greater amounts (Figure 4.4b).

To investigate whether build up of sufficient lac repressor in the extract would bring about repression, 100ng of pT7lacI was incubated in the cell extract for 10 minutes prior to the addition of 200ng pT7lacOGFP. In this case, higher amounts of pT7lacOGFP plasmid were added in comparison of pT7lacI to compensate for the loss of expression capacity. Presence of lac repressor cache prior to the addition of pT7lacOGFP resulted in

a 3.5 fold increase in expression from pT7lacOGFP in response to the addition of 1mM IPTG (Figure 4.4c). Increasing the time of incubation of T7lacI to 30 minutes severely affected the expression capacity from the extract and high GFP expression could not be achieved even when 800 ng GFP was used in the reaction (Figure 4.5). These results suggest that large quantities of GFP transcripts that saturate the translation machinery were synthesized before lac repressor reaches a concentration that can effectively repress T7lacO promoters. In addition, while prior accumulation of lacI addresses this issue, effective repression of T7lacO promoter is attained at the expense of expression capacity.

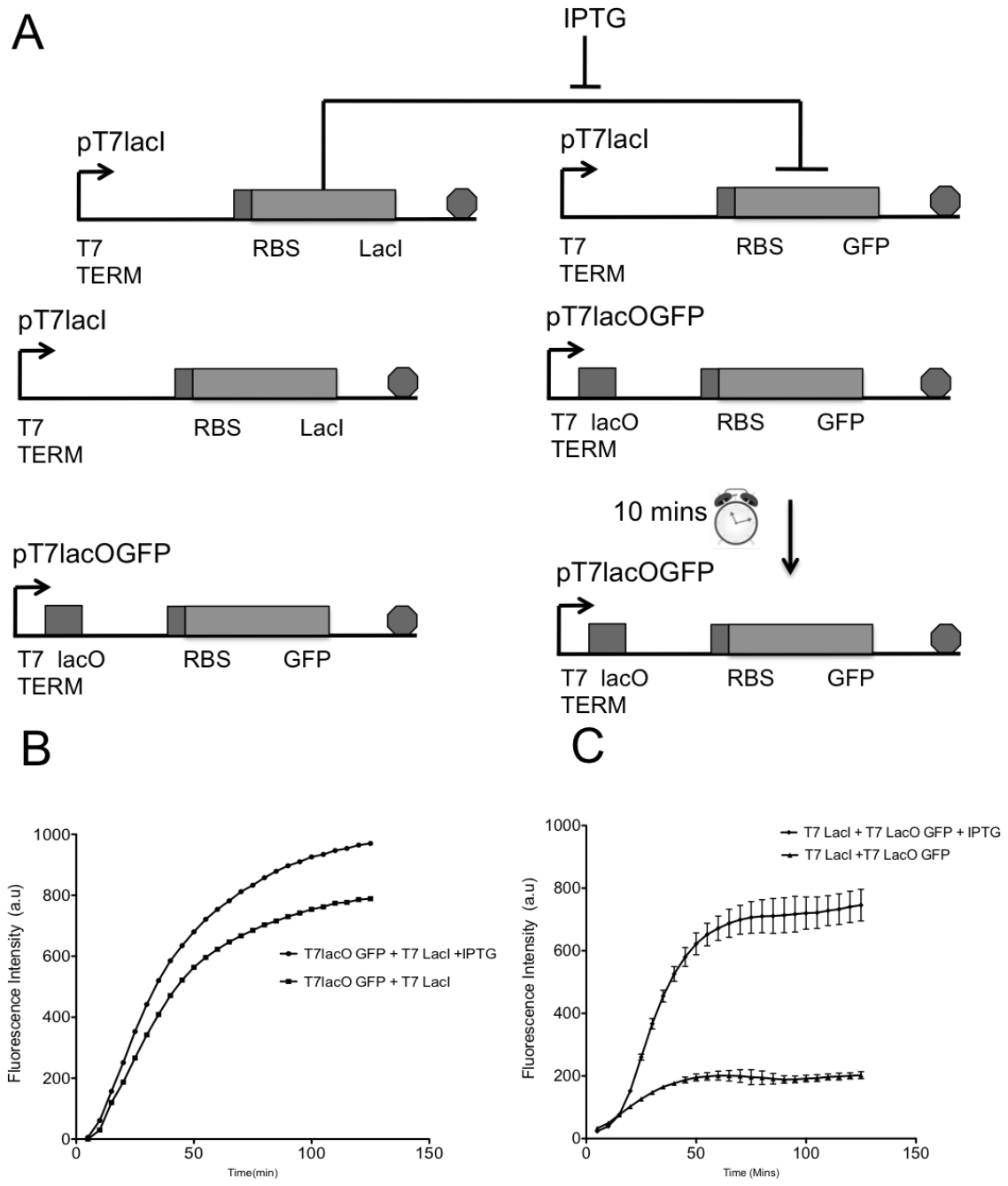


Figure 4.4 Repression of T7lacO by LacI encoded on a separate plasmid A) schematic of repression strategy B) shows expression from T7lacOGFP plasmids upon incubation with pT7lacI C) Repression from pT7lacOGFP added 10 mins after incubating pT7lacI in the extract.

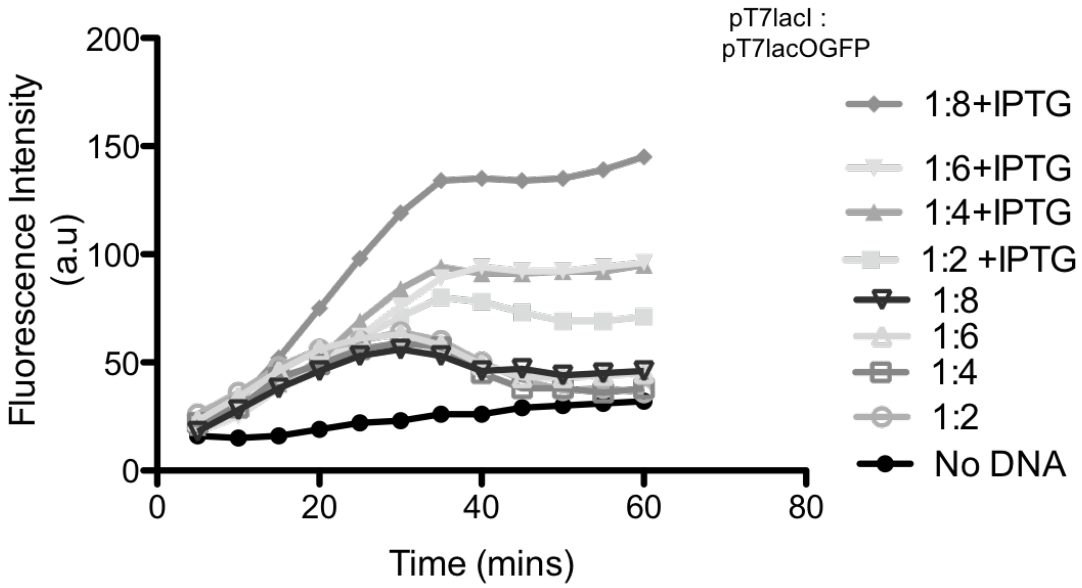


Figure 4.5 Effect of ratios of pT7lacOGFP and pT7lacI on expression from T7lacO promoters. pT7lacOGFP in the indicated ratios 30 minutes after the addition of pT7lacI.

Negative feedback from T7 lacO promoters

A circuit in which a transcription factor negatively regulates its own expression is called a negative autoregulation motif. Typically a transcriptional fusion of this motif with a reporter gene is constructed to provide a read-out of repression dynamics. [11, 12]. Here a T7lacO promoter that drives the expression of lacI and GFP was constructed to demonstrate negative autoregulation. Repression cascade experiments suggested that for effective repression from T7lacO required accumulation of lac repressor in the extract. Since temporal separation of lacI and GFP expression from a bicistronic construct is difficult to achieve, it was hypothesized that a variant of T7lacO promoters that tightly represses expression from T7lacO promoters at low lac repressor concentration will enable the construction of negative feedback circuit. To this end, auxiliary lac operators upstream to the T7lacO promoters were incorporated into the design of negative feedback circuits. As shown in chapter 3, the presence of an auxiliary lacO operator located 92 base upstream to the T7lacO promoter enables stronger repression from T7lacO promoters, presumably by increasing the probability of lacI binding the T7lacO promoter. To examine the effects of auxiliary operators to realize a negative feedback motif , two different negative feedback bicistronic circuits expressing lacI and GFP that differed in the promoters from which they were expressed were constructed- while pT7lacOlacIGFP contained T7lacO promoter, placOT7lacOlacIGFP contained the lacO operator 92 bases upstream to the T7lacO promoter. (Figure 4.6a) To test negative feedback response, 1mM IPTG, the negative regulator of lacI, was added to the cell free reactions and fluorescence response was measured. As a control, placIGFP was constructed, in which the lacI and

GFP genes are expressed from a T7 promoter. Interestingly, addition of IPTG resulted in a little to no change in expression from any of these constructs (Figure 4.6b).

It is known that the high processivity of T7 polymerase results in a rapid saturation of translation machinery thereby masking the repression occurring in cell free systems. We hypothesized therefore, that rapid accumulation of RNA transcripts from T7 promoters might result in the saturation of translational machinery. Hence, to reduce the mRNA load, T7 RNA polymerase amounts were reduced to 20% of the amount prescribed by the manufacturer for the reaction. Under this new regime, whereas expression from pT7lacIGFP and pT7lacOlacIGFP remained unchanged, LacOT7lacOlacIGFP showed a 1.6 fold change in gene expression between the induced and uninduced states thereby showing the benefits of using a stronger repression mechanism and optimization of transcription for achieving negative feedback in cell free extracts (Figure 4.6c)

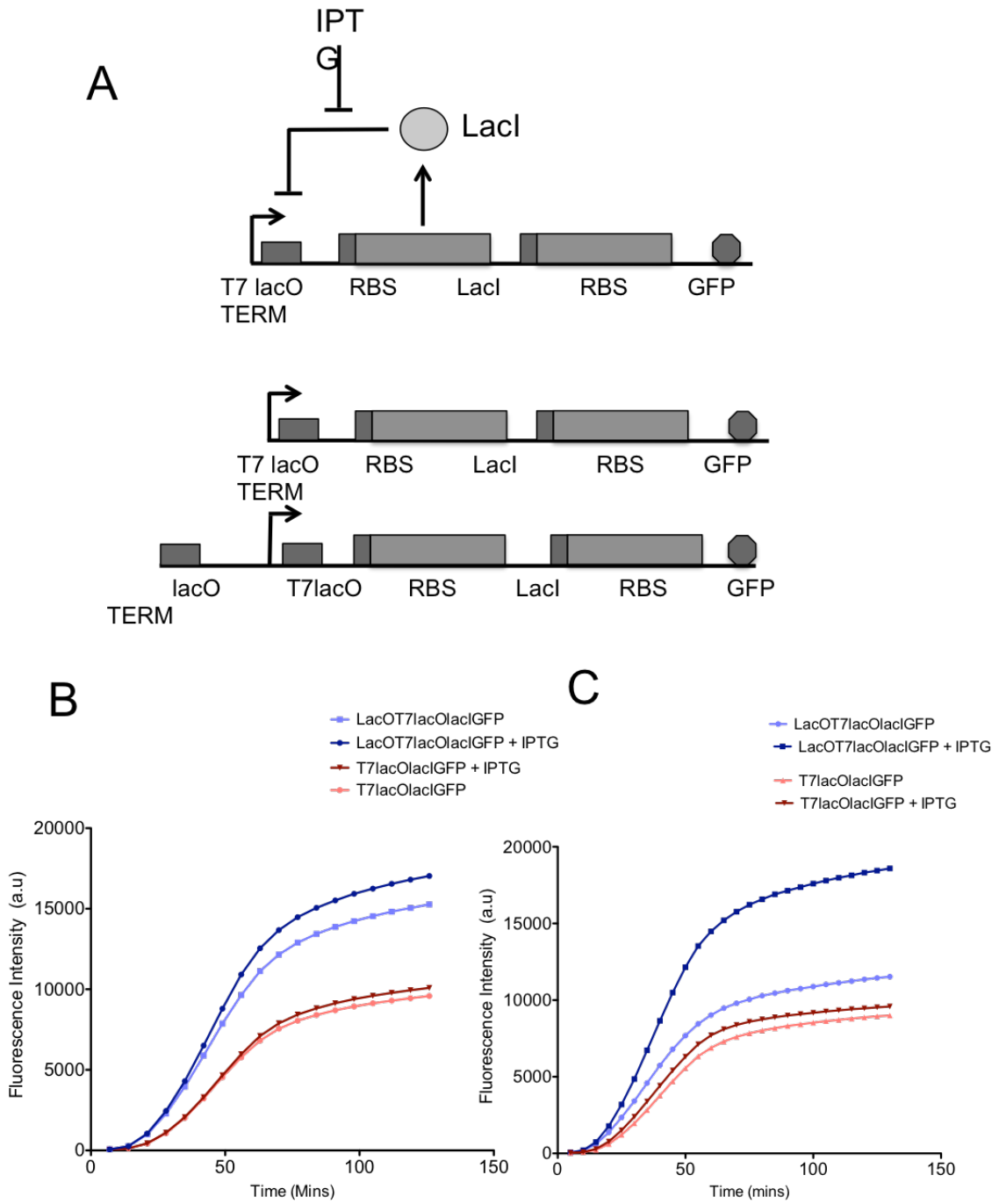


Figure 4.6 Demonstration of negative feedback from pLacOT7lacOlacIGFP and pT7lacOlacIGFP.

Discussion

Cell extracts provide a simplified context for implementing cell free gene circuits. Here, a cell free negative feedback motif and repression cascades using T7lacO promoters are described. First, we investigated monocistronic and bicistronic strategies to assemble cell free circuits. This was followed by a demonstration of a repression cascade built from T7lacO promoters and lac repressors. Finally,a negative feedback system using the lacO – lac repressor system was constructed.

Beyond offering a simplified context, several features of cell-free synthetic biology are appealing for direct applications and also for the initial prototyping of both genetic devices and assembled systems in live cells. Our characterization of assembly methods exemplifies benefits of forward engineering biological functionality in cell-free contexts. Specifically, direct quantitative characterization of transcriptional regulation is greatly facilitated. By contrast, in live cells, such dosage responses must be inferred indirectly through careful analysis of single cell responses with consideration of the significant amounts of noise in gene expression [174]. Another advantage of cell-free systems is that they enable fast screening of construct libraries, as transformation is not required. Also, as shown in Figure 4.4, multi-plasmid systems may be implemented without regard to backbone compatibility, and DNA concentration is easily tunable, unlike in cells. Finally, testing regulatory systems in a context free of mutation and recombination can simplify initial system development can aid troubleshooting efforts in live cells systems.

First, we sought to examine different approaches for assembling regulatory networks in cell-free systems. When implementing large gene networks in live *E. coli*,

one option is to integrate the genetic components into the chromosome. However, with this approach, it is cumbersome to explore a large number of combinations of different network variants, in terms of different ribosome sites, different promoter variants, different protease tags, etc. In addition, location in the genome can impact expression. For these reasons, most live *E. coli* systems in synthetic biology have relied on the use of plasmids. Still, at most two or three different plasmid types can be used, and plasmid compatibility must be carefully considered in terms of the origins of replication and the antibiotic resistances. On the other hand, with cell-free systems, the same backbone can be used for different system components. This enables an approach for constructing large synthetic gene networks, whereby each component is encoded on a separate plasmid, and different plasmids are combined in cell-free extract to form the final system. The DNA copy number of each network component can be easily and precisely tuned, whereas in live cells, copy number can only be coarsely tuned for each plasmid by using different origins of replication. Thus, with this multi-plasmid approach, a large number of network variants can be quantified without the need for chromosomal integrations or transformations.

As expected, when either *tetR* or *lacI* was inserted in a bicistronic sequence with *GFP*, fluorescence decreased due to sharing of expression capacity between the repressor and GFP (Figure 4.2b). Insertion of *tetR* reduced expression by approximately half, while insertion of *lacI* reduced expression by approximately a quarter. This difference in the effects of inserting *tetR* vs. *lacI* implies that inserting a gene in a bicistronic sequence impacts relative expression in a manner that is dependent on the particular gene inserted. Trading the order of *GFP* and *lacI* had no impact on fluorescence.

By comparison, when each gene is expressed on a separate plasmid, normalized fluorescence is reduced by half even when the EGFP expressing plasmid comprises 80% of the total plasmid concentration (Figure 4.2c). In general, one tradeoff with this approach is that the flexibility of easily tuning relative gene copy numbers can potentially come at the cost of weaker expression, as shown in Figure 4.3. For example, due to the higher ratio of promoters to genes, the effects of inefficiency in transcriptional termination may be more pronounced. Nonetheless, it has previously been shown that, with the multiple plasmid approach, properly tuning the ratio of plasmids, along with the use of common downstream box sequences, can help to achieve efficient expression of all genes in the system [175]. Interestingly, we observed a nonlinear relationship between relative plasmid ratio and expression (Figure 4.2c). This is potentially due to a competition between the constructs for translational resources [175], and future experiments to quantify the yield of both proteins will help to further elucidate the cause of this nonlinearity.

As a step toward implementing negative feedback circuits, we set out to investigate the gene dosing effects on repression from T7lacO promoters. Implementation of dynamic gene circuits in cell free systems entails attaining a proper balance between the RNA and protein components[1, 176]. Typically, different genetic elements specifying protein expression efficiency are explored to provide appropriate dosing of RNA and protein components to enable a particular function[38, 117]. Here, a multi-plasmid approach to control the gene copy number in cell extracts to generate LacI proteins in different proportions for negatively regulating T7lacO promoters. Use of a cell free platform offers unprecedented opportunity to control the DNA components and

achieve temporal separation of expression from pT7lacI and PT7lacOGFP plasmids to demonstrate effective repression. In fact, such a multiplasmid approach has been recently used in assembling a variety of circuits such as multistep cascades, AND logic gates and feedback loops in cell free systems[52]. Oue results show that whereas simultaneous addition of pT7lacI and pT7lacOGFP did not have a significant effect on GFP expression, incubating pT7lacI for 10 minutes before the addition of pT7lacOGFP plasmid resulted in an effective repression that could be reversed by the addition of 1mM IPTG. This result mirrors a similar study done in wheat germ extracts where the authors observed no repression upon simultaneous expression from similar constructs[48]. This could be because a critical concentration of lac repressor needs to be attained before it can effectively repress T7lacO promoters. In case of simultaneous expression of both pT7lacOGFP and pT7lacI, GFP transcripts are rapidly generated before lac repressor concentration that can repress T7lacO promoters can be reached. Accumulation of lac repressor prior to GFP expression addresses this need effectively. Increases in repressor plasmid concentration result in an increase in the repression levels from T7lacO come at the cost of expression levels. In addition, the increase in GFP expression with increasing T7lacOGFP concentration indicates that translation machinery was not saturated for the tested plasmid concentrations and that this parameter can be tuned to achieve the desired level of expression from a multiplasmid system. Tuning operator –repressor interactions and parameters to achieve efficient gene expression in cell free systems might be used to mitigate this issue.

Efficient genetic components in our negative feedback circuit such as g10 RBS and T7 terminators were used to ensure high expression levels[38]. To demonstrate

simple cell-free gene networks, negative feedback circuits based on two different lac repressible promoters were constructed. Presence of auxiliary lac operators 92 bases upstream to T7lacO promoters facilitates effective repression even in the presence low repressor concentration (Chapter 3). The two negative feedback designs tested relied two different lac repressible promoter systems, T7lacO promoter and T7lacO promoters with auxiliary operators to drive the expression of lac repressor and the reporter gene GFP. Surprisingly, addition of IPTG, which is expected to induce expression from T7lacO promoters, did not result in any observable increase in expression from either of these templates. In cell extracts using highly processive T7 RNA polymerases, this balance is skewed in favor of product accumulation due to rapid synthesis of RNA and protein products and absence of mechanisms to dilute RNA and protein[48, 50]. This combined with the time lag that occurs due to synthesis of repressor protein, operator binding and repression of expression from the promoter contribute to lower change in expression realized in cell free systems in contrast to a cell based system. This effect is particularly pronounced with bacterial cell extracts that have a limited reaction lifetime that prevents the manifestation of repression that occurs at later time points.

Reduction in T7 RNA polymerase amounts addressed the overshoot issue and resulted in a 1.5 fold increase in fluorescence levels from pLacOT7lacOlacIGFP whereas no observable change was observed with pT7lacOlacIGFP plasmids showing the benefits of using a tightly repressible T7 lacO promoter and reducing the transcriptional output by lowering polymerase concentrations in the extract. As with the repressor cascades, repression of T7lacO promoters in pT7lacOlacIGFP plasmids can occur only after the formation of functional lac repressor tetramers. The initial burst of transcripts before the

LacI mediated repression tends to saturate the translation machinery thereby masking the effects of repression that occurs at later time points. In fact, similar saturation of translational machinery has been observed even in cells[66]. To address the issues arising from saturation of translational machinery, Karig *et.al* showed tuning of genetic determinants of gene expression such as strength of ribosome binding site and terminators resulted in different magnitudes of fold change in expression between the uninduced and induced state[38]. They observed a 1.8 fold induction with a tet repressible negative feedback circuit. Lowering polymerase concentrations, weakening the expression determinants and use of directed RNA and protein degradation mechanisms together might yield greater fold changes in expression [38, 50].

All cell-free reactions were performed in batch mode, which prevents the influx of nutrients and the efflux of waste products and consequently limits the reaction dynamics [43, 45, 177-180]. Extension to a continuous flow system would help to harness the advantages of repressible T7 promoter variants [39, 181]. The flow of fresh nutrients and removal of waste products would help to preserve the initially strong rate of expression. The additional incorporation of mechanisms to actively degrade mRNA and EGFP [50] would mitigate the previously described overshoot problem with cell free negative feedback circuits and would reduce the high yield in the absence of inducer.

Alternative approaches to engineering regulation in cell-free contexts avoid the use of translational machinery, thus further simplifying the engineering of fast, complex systems [182, 183]. At the same time, protein expression is clearly useful for a number of applications such as production of protein-based therapeutics and chemical sensors. Ultimately, the expression components and simple feedback systems that we present can

be interfaced to more complex regulatory networks based on a simplified set of mechanisms [184].

Our results contribute to bottom-up approaches to engineering biological function [185]. The simplified context and the facilitation of direct, quantitative component characterization offered by cell-free systems will aid efforts to transcend the complexity of systems currently engineered in living cells [18]. For assembling regulatory systems in cell-free contexts, both the traditional multicistronic approach and an approach whereby a separate plasmid is used for each gene appear to be viable. Furthermore, the successful demonstration of inducible negative feedback embodies an initial step towards more complex regulatory systems. In the future, coupling the ability to forward engineer cell-free genetic regulation with efforts to compartmentalize reaction components in small liposomes [186-188] or nanofabricated wells [189-192] will help to close the gap between harnessing the unique capabilities of living cells and capitalizing on the comparative ease of engineering in simpler contexts.

Materials and Methods

Plasmids and strains

All plasmids used in this study were constructed using standard methods. These plasmids are described in Table 4.1 and 4.2. DNA used in cell free experiments was prepared using Qiagen Plasmid Maxi prep kits.

Cell-free expression experiments

The Promega S30 T7 High-Yield Expression System kit (Promega TM306) was used for the experiments depicted in Figures 4.2 and 4.3 and Expressway high yield protein

synthesis kit (Invitrogen) was used for carrying out expression in depicted in rest of the figures. For experiments in Figure 4.2 and 4.3 samples consisting of reaction mix with no DNA were assayed to quantify background fluorescence of the reaction mix, and samples with pDEST17-EGFP [191, 192] were assayed for the purpose of normalizing fluorescence values. "Normalized fluorescence" (NFU) for a given sample was calculated by subtracting background fluorescence of the reaction mix from that sample's fluorescence value and then dividing by the background corrected fluorescence of the benchmark construct pDEST17-EGFP. Results depicting final yield (Figure 4.2 and 4.3) are the normalized fluorescence values after 10 hours of expression. In case of figure 3 and 4, reactions were set up following manufacturer's instructions except that the final reaction volume was 15 μL and the reaction was treated as a batch reaction instead of a fed batch reaction format with the feed buffer containing both 2.5X IPVS and 2XIVPS buffers provided in the kit. 15 μL mineral oil was added to each of the reactions to prevent drying. For induction experiments, IPTG [193] was added in the denoted concentrations. Reactions were set up in Corning CLS3820 plates. Samples were incubated at 30°C with shaking and measured every 6 minutes in a Biotek Synergy 2 plate reader. For the measurements, excitation was 485/20 nm, emission was 528/20 nm, the optics position was set at "Top 510," and the sensitivity was set at 40. Error bars in all figures represent standard deviation of at least three replicates.

Chapter 5 Conclusions and future directions

Advances in molecular biology and genetic engineering have offered unprecedented insights into the molecular workings of a cell. Decades of biological research are providing glimpses into the molecular underpinnings of biological systems. Omics technologies are transforming the field from defining the cell as a sum of molecular parts to describing biological systems in terms of the nature of interactions. One approach to understanding interactions is to computationally identify recurring patterns of network connectivity. These motifs are embedded in networks along with other components, which makes the investigation of its individual contribution to the properties of the system rather difficult. Synthetic construction of these motifs provides an alternative approach to investigation of the role of these components in a larger network.

Synthetic biology is aimed at forward engineering biological function using well-characterized molecular components. The driving philosophy for synthetic biology is the iterative design and understanding of a system entwined. Deliberate design of biological systems serves as a reductionist approach to understanding the design principles that govern biological processes. Biological systems offer advantages of miniaturization, self-replication and obvious biocompatibility for realizing engineered functions that can be harnessed for biotechnological and biomedical applications.

Analogies with traditional engineering disciplines are being evoked for describing and designing engineered circuits using biological parts. Yet, there are important distinctions between traditional engineered circuits and biological circuits that require a fundamentally different approach to designing cellular systems that perform desired tasks. Unlike semiconductor based circuits, biological systems are not inherently

engineered and therefore the components are not standardized and are difficult to interface with cells.

Therefore, there is a growing interest in implementing information processing gene circuits in cell free systems akin to those that have been implemented in cells[34]. In addition to offering flexible platforms for implementing logical behavior, as discussed earlier, cell free systems allow for quantitative component characterizations that may be extended to cell based systems[123].

Accordingly, the approach taken in this work was to construct systems from the bottom up to facilitate the implementation of dynamic cell like behavior in a cell free system. A large portion of this dissertation work was devoted to the development of tools to achieve gene regulation. As mentioned earlier in this document, the strength of cell free systems lies in the opportunity to decouple viability of the system and utilize biological machinery for translating DNA based instructions. The field would benefit from advancements in the production of extracts for achieving robust and high-yield expression, tools to manipulate gene expression and a platform in which to implement cell free reactions. The approaches outlined here would greatly bolster efforts to implement network motifs in cell free systems.

In this thesis, I focused on developing mechanisms for gene regulation and engineering T7 promoters for use in synthetic gene circuits. Noireaux *et.al* have asserted that ligand responsive expression and implemention transcriptional logic is difficult to achieve with T7 promoters[51, 118]. Contrary to this perception, here I have shown that the T7 transcriptional system is tractable. Innovative designs for regulating T7 promoters have

been harnessed to demonstrate that the simplicity of the polymerase can be asset to signal sensing and integration.

T7 RNA polymerase can transcribe through several DNA structures, DNA lesions and even gaps in template that might otherwise block other polymerases. These properties along with the use of cell free systems provide the flexibility to utilize novel gene regulation mechanisms that might be difficult to implement in living cells. At the same time, for the regulation strategy to be widely applicable, it must be specific, sensitive, tunable, modular and achieve a large dynamic range of gene expression upon target addition. To this end, in chapter 2, a DNA aptamer based strategy to achieve gene regulation was utilized. An aptamer that binds thrombin with high specificity was selected to demonstrate proof of principle. Firstly, ssDNA generated from phagemids was utilized to assemble the appropriate DNA structures. This technique allows for easy and large-scale preparation of an aptamer harboring single stranded DNA strands that can be used to direct protein synthesis in cell free systems. Up to a 5-fold change in gene expression could be attained from these systems. Furthermore, repression levels could be easily tuned by moving the aptamer away from the transcriptional start site. This work sets the stage for testing other aptamers for regulating gene expression. Aptamers selected to bind their target in cell free systems might be still more effective in regulating expression from these templates. It also remains to be seen whether the nature of the target molecule (small molecule or protein), and aptamer structure has an effect on the efficacy of this gene regulation strategy.

In addition to their use in transcriptional regulation, DNA aptamers were also adapted for translational regulation as described in Appendix B. The mechanism harnessed the ability

of RNaseH enzyme to digest an RNA transcript hybridized to a DNA oligo. The repressor oligonucleotides were designed to contain a sensing domain comprised of the thrombin DNA aptamer oligonucleotide and a repressor domain with a sequence complementary to the RNA transcript. The repressor oligo was designed such that, in the absence of thrombin, the repressor secondary structure sequesters the binding domain and prevents it from hybridizing to the RNA transcript. Addition of thrombin exposed the repressor domain thereby allowing it to hybridize to the RNA transcript and repressing expression by 10 fold. This result, together with transcriptional repression, shows that the modularity of DNA aptamers can be effectively used to regulate both transcription and translation in cell free systems.

Along with mechanisms to engineer ligand sensitivity, it is beneficial to develop strategies to achieve signal integration with T7 promoters. To this end, chapter 3 described two different transcriptional factors TetR and LacI for regulating expression from T7 promoters in concert. To achieve this, first lac operators that utilize DNA looping to achieve tight repression from T7lacO promoters were used. Next tetO sites were placed in this framework to achieve combinatorial regulation of gene expression from these promoters. Placement of tetO operators downstream from lacO and at locations overlapping the T7 promoter did not alter response to lac repressors. However, TetR binding to a site upstream from T7lacO promoters resulted in interference with lac dependent looping and therefore lac repression. This resulted in the first demonstration of an IMPLIES gate implemented *in vivo* and *in vitro* with T7 promoters. This gate was also realized in cell free systems using TetR and LacI. It might be interesting to investigate the effects of combining the operator arrangement with different T7 promoter strengths to

investigate the range of phenotypes that can be achieved. This study provides a way forward for using the gate developed here for enabling synthetic gene circuits.

Finally, in chapter 5 a negative feedback circuit was implemented in cell free systems. We have shown that successful design of a functional gene circuits requires attention to component details, quantitative component characterizations and tradeoffs associated with achieving high signal levels and gene circuit function [122]. In addition, synthetic DNA aptamer based systems for achieving modular gene regulation in cell free systems have been developed. Insights garnered from the implementation of the simple negative feedback circuits described here will enable the realization of predictable information processing systems of higher complexity that will find use in a variety of biotechnological applications and for bottom up understanding of biological network organization.

To be a valuable tool for scientific investigation, cell free systems need to be easy to implement, cost effective and robust. For cell free systems to be recognized as a viable complement to existing technologies, applications that take advantage of the flexibility of this system would be necessary. Physical platforms such as those described in Appendix C would be beneficial for enabling long-term gene expression and for implementing more sophisticated functions. Moreover, spatial separation of biochemical reactions on a microfluidic platform would facilitate the construction of evermore sophisticated synthetic gene circuits[194, 195].

One could argue that operation and investigation within biological complexity might be necessary to replicate biological function. Indeed, redundancy of function in almost all biological systems is thought to confer robustness to biological systems and a reductionist

approach might be stripping the system out of its inherent advantages. Yet, incomplete knowledge of components even in model systems such as *E.coli* cells, present overwhelming challenges in defining all interactions. Therefore, a minimal model that comprises a minimal set of genes to enable cell like functionality such as replication and protein synthesis would be valuable for implementing predictable networks. Indeed, *Mycoplasma genitalium, a* model minimal system has been a subject of several recent efforts to construct and understand the functions of a minimal cell[196]. As a demonstration of the power of a pared down system to understand primary workings of the cell, researchers from JCVI (J Craig Venter Institute) and Stanford University put forth the computational simulation of a whole working cell[29]. Although the investigations might be far from complete, efforts such as these represent a step forward in providing a complete description of biological systems.

These studies along with the approach taken in this dissertation, highlight the potential to understand biological networks from the bottom up. As the costs associated with gene synthesis reduce and we have precise control over protein components present in the system, these bottom-up strategies can be expected to yield more tractable and cost efficient approaches for engineering biology. The tools developed here would expand the applicability of cell free systems for implementing predictable gene circuits that would ultimately facilitate bottom-up understanding of biological function. Ultimately, implementation of predictable synthetic biological networks would enable us to harness the power of biological circuitry for biotechnological applications.

Bibliography

1. Haseltine, E.L. and F.H. Arnold, *Synthetic gene circuits: design with directed evolution.* Annu Rev Biophys Biomol Struct, 2007. **36**: p. 1-19.
2. Hasty, J., D. McMillen, and J.J. Collins, *Engineered gene circuits.* Nature, 2002. **420**(6912): p. 224-30.
3. Andrianantoandro, E., et al., *Synthetic biology: new engineering rules for an emerging discipline.* Mol Syst Biol, 2006. **2**: p. 2006 0028.
4. Gardner, T.S., C.R. Cantor, and J.J. Collins, *Construction of a genetic toggle switch in Escherichia coli.* Nature, 2000. **403**(6767): p. 339-42.
5. Elowitz, M.B. and S. Leibler, *A synthetic oscillatory network of transcriptional regulators.* Nature, 2000. **403**(6767): p. 335-8.
6. Elowitz, M. and W.A. Lim, *Build life to understand it.* Nature, 2010. **468**(7326): p. 889-90.
7. Cox, R.S., 3rd, M.G. Surette, and M.B. Elowitz, *Programming gene expression with combinatorial promoters.* Mol Syst Biol, 2007. **3**: p. 145.
8. Guet, C.C., et al., *Combinatorial synthesis of genetic networks.* Science, 2002. **296**(5572): p. 1466-70.
9. Milo, R., et al., Network motifs: simple building blocks of complex networks. Science, 2002. **298**(5594): p. 824-7.
10. Alon, U., *Network motifs: theory and experimental approaches.* Nat Rev Genet, 2007. **8**(6): p. 450-61.
11. Rosenfeld, N., M.B. Elowitz, and U. Alon, *Negative autoregulation speeds the response times of transcription networks.* J Mol Biol, 2002. **323**(5): p. 785-93.
12. Austin, D.W., et al., *Gene network shaping of inherent noise spectra.* Nature, 2006. **439**(7076): p. 608-11.
13. Maeda, Y.T. and M. Sano, Regulatory dynamics of synthetic gene networks with positive feedback. J Mol Biol, 2006. **359**(4): p. 1107-24.
14. Endy, D., *Foundations for engineering biology.* Nature, 2005. **438**(7067): p. 449-53.
15. Bio, F.A.B.G., et al., *Engineering life: building a fab for biology.* Sci Am, 2006. **294**(6): p. 44-51.
16. Kelly, J.R., et al., Measuring the activity of BioBrick promoters using an in vivo reference standard. J Biol Eng, 2009. **3**: p. 4.
17. Canton, B., A. Labno, and D. Endy, *Refinement and standardization of synthetic biological parts and devices.* Nat Biotechnol, 2008. **26**(7): p. 787-93.
18. Purnick, P.E. and R. Weiss, *The second wave of synthetic biology: from modules to systems.* Nat Rev Mol Cell Biol, 2009. **10**(6): p. 410-22.
19. Tsvetanova, B., et al., *Genetic assembly tools for synthetic biology.* Methods Enzymol, 2011. **498**: p. 327-48.
20. Marucci, L., et al., How to Turn a Genetic Circuit into a Synthetic Tunable Oscillator, or a Bistable Switch. PLoS One, 2009. **4**(12).
21. Rosenfeld, N., et al., Accurate prediction of gene feedback circuit behavior from component properties. Mol Syst Biol, 2007. **3**: p. 143.
22. Slusarczyk, A.L., A. Lin, and R. Weiss, *Foundations for the design and implementation of synthetic genetic circuits.* Nat Rev Genet, 2012. **13**(6): p. 406-20.
23. Shetty, R.P., D. Endy, and T.F. Knight, Jr., *Engineering BioBrick vectors from BioBrick parts.* J Biol Eng, 2008. **2**: p. 5.
24. Anderson, J.C., C.A. Voigt, and A.P. Arkin, *Environmental signal integration by a modular AND gate.* Mol Syst Biol, 2007. **3**: p. 133.

25. Temme, K., et al., Modular control of multiple pathways using engineered orthogonal T7 polymerases. Nucleic Acids Res, 2012.
26. An, W. and J.W. Chin, *Synthesis of orthogonal transcription-translation networks.* Proc Natl Acad Sci U S A, 2009. **106**(21): p. 8477-82.
27. Jewett, M.C. and A.C. Forster, *Update on designing and building minimal cells.* Curr Opin Biotechnol, 2010. **21**(5): p. 697-703.
28. Forster, A.C. and G.M. Church, *Synthetic biology projects in vitro.* Genome Res, 2007. **17**(1): p. 1-6.
29. Karr, J.R., et al., A whole-cell computational model predicts phenotype from genotype. Cell, 2012. **150**(2): p. 389-401.
30. Gibson, D.G., et al., One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20404-9.
31. Gibson, D.G., et al., Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 2010. **329**(5987): p. 52-6.
32. Nakajima, M., et al., Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. Science, 2005. **308**(5720): p. 414-5.
33. Kim, J., K.S. White, and E. Winfree, Construction of an in vitro bistable circuit from synthetic transcriptional switches. Mol Syst Biol, 2006. **2**: p. 68.
34. Kim, J. and E. Winfree, *Synthetic in vitro transcriptional oscillators.* Mol Syst Biol, 2011. **7**: p. 465.
35. Franco, E., et al., Timing molecular motion and production with a synthetic transcriptional clock. Proc Natl Acad Sci U S A, 2011. **108**(40): p. E784-93.
36. Ueda, T., *[PURE system: a reconstituted cell-free protein synthesis system].* Tanpakushitsu Kakusan Koso, 2009. **54**(12 Suppl): p. 1448-51.
37. Asahara, H. and S. Chong, In vitro genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. Nucleic Acids Res, 2010. **38**(13): p. e141.
38. Karig, D.K., et al., Expression optimization and synthetic gene networks in cell-free systems. Nucleic Acids Res, 2012. **40**(8): p. 3763-74.
39. Katzen, F., G. Chang, and W. Kudlicki, *The past, present and future of cell-free protein synthesis.* Trends Biotechnol, 2005. **23**(3): p. 150-6.
40. Swartz, J.R., M.C. Jewett, and K.A. Woodrow, *Cell-free protein synthesis with prokaryotic combined transcription-translation.* Methods Mol Biol, 2004. **267**: p. 169-82.
41. Hino, M., et al., Efficiency of cell-free protein synthesis based on a crude cell extract from Escherichia coli, wheat germ, and rabbit reticulocytes. J Biotechnol, 2008. **133**(2): p. 183-9.
42. Tsuboi, T., et al., The wheat germ cell-free protein synthesis system: a key tool for novel malaria vaccine candidate discovery. Acta Trop, 2010. **114**(3): p. 171-6.
43. Calhoun, K.A. and J.R. Swartz, *Energizing cell-free protein synthesis with glucose metabolism.* Biotechnol Bioeng, 2005. **90**(5): p. 606-13.
44. Calhoun, K.A. and J.R. Swartz, Energy systems for ATP regeneration in cell-free protein synthesis reactions. Methods Mol Biol, 2007. **375**: p. 3-17.
45. Kim, D.M. and J.R. Swartz, *Prolonging cell-free protein synthesis by selective reagent additions.* Biotechnol Prog, 2000. **16**(3): p. 385-90.
46. Spirin, A.S., et al., A continuous cell-free translation system capable of producing polypeptides in high yield. Science, 1988. **242**(4882): p. 1162-4.
47. Hockenberry, A.J. and M.C. Jewett, *Synthetic in vitro circuits.* Curr Opin Chem Biol, 2012. **16**(3-4): p. 253-9.

48. Noireaux, V., R. Bar-Ziv, and A. Libchaber, *Principles of cell-free genetic circuit assembly.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12672-7.
49. Ishikawa, K., et al., *Expression of a cascading genetic network within liposomes.* FEBS Lett, 2004. **576**(3): p. 387-90.
50. Shin, J. and V. Noireaux, Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. J Biol Eng, 2010. **4**: p. 9.
51. Shin, J. and V. Noireaux, Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. J Biol Eng, 2010. **4**: p. 8.
52. Shin, J. and V. Noireaux, An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. Acs Synthetic Biology, 2012. **1**(1): p. 29-41.
53. Isalan, M., C. Lemerle, and L. Serrano, Engineering gene networks to emulate Drosophila embryonic pattern formation. PLoS Biol, 2005. **3**(3): p. e64.
54. Buxboim, A., et al., A single-step photolithographic interface for cell-free gene expression and active biochips. Small, 2007. **3**(3): p. 500-510.
55. Mei, Q., et al., Protein synthesis in a device with nanoporous membranes and microchannels. Lab Chip, 2010. **10**(19): p. 2541-5.
56. Siuti, P., S.T. Retterer, and M.J. Doktycz, *Continuous protein production in nanoporous, picolitre volume containers.* Lab Chip, 2011. **11**(20): p. 3523-9.
57. Collier, C.P. and M.L. Simpson, *Micro/nanofabricated environments for synthetic biology.* Current Opinion in Biotechnology, 2011. **22**(4): p. 516-526.
58. Karzbrun, E., et al., Coarse-grained dynamics of protein synthesis in a cell-free system. Phys Rev Lett, 2011. **106**(4): p. 048104.
59. Jewett, M.C., et al., An integrated cell-free metabolic platform for protein production and synthetic biology. Molecular systems biology, 2008. **4**: p. 220.
60. Chang, A.L., J.J. Wolf, and C.D. Smolke, *Synthetic RNA switches as a tool for temporal and spatial control over gene expression.* Curr Opin Biotechnol, 2012. **23**(5): p. 679-88.
61. Buskirk, A.R. and D.R. Liu, Creating small-molecule-dependent switches to modulate biological functions. Chem Biol, 2005. **12**(2): p. 151-61.
62. Browning, D.F. and S.J. Busby, *The regulation of bacterial transcription initiation.* Nat Rev Microbiol, 2004. **2**(1): p. 57-65.
63. Schilstra, M.J. and C.L. Nehaniv, *Bio-logic: gene expression and the laws of combinatorial logic.* Artif Life, 2008. **14**(1): p. 121-33.
64. Buchler, N.E., U. Gerland, and T. Hwa, *On schemes of combinatorial transcription logic.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5136-41.
65. Kim, D.M., et al., A highly efficient cell-free protein synthesis system from Escherichia coli. Eur J Biochem, 1996. **239**(3): p. 881-6.
66. Studier, F.W. and B.A. Moffatt, Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J Mol Biol, 1986. **189**(1): p. 113-30.
67. Rong, M., R.K. Durbin, and W.T. McAllister, *Template strand switching by T7 RNA polymerase.* J Biol Chem, 1998. **273**(17): p. 10253-60.
68. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nat Rev Genet, 2010. **11**(5): p. 367-79.
69. Sakaguchi-Mikami, A., et al., Construction of a novel glucose-sensing molecule based on a substrate-binding protein for intracellular sensing. Biotechnol Bioeng, 2011. **108**(4): p. 725-33.
70. Wang, X., X. Chen, and Y. Yang, Spatiotemporal control of gene expression by a light-switchable transgene system. Nat Methods, 2012. **9**(3): p. 266-9.

71. Beerli, R.R., et al., *Chemically regulated zinc finger transcription factors.* J Biol Chem, 2000. **275**(42): p. 32617-27.
72. Lohmueller, J.J., T.Z. Armel, and P.A. Silver, *A tunable zinc finger-based framework for Boolean logic computation in mammalian cells.* Nucleic Acids Research, 2012. **40**(11): p. 5180-5187.
73. Tuerk, C. and L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science, 1990. **249**(4968): p. 505-10.
74. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* Nature, 1990. **346**(6287): p. 818-22.
75. Huizenga, D.E. and J.W. Szostak, *A DNA Aptamer That Binds Adenosine and Atp.* Biochemistry, 1995. **34**(2): p. 656-665.
76. Macaya, R.F., et al., Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. Proc Natl Acad Sci U S A, 1993. **90**(8): p. 3745-9.
77. Green, L.S., et al., *Inhibitory DNA ligands to platelet-derived growth factor B-chain.* Biochemistry, 1996. **35**(45): p. 14413-14424.
78. Hamula, C.L.A., et al., *Selection of aptamers against live bacterial cells.* Analytical Chemistry, 2008. **80**(20): p. 7812-7819.
79. Cox, J.C., et al., Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. Nucleic Acids Res, 2002. **30**(20): p. e108.
80. Vitreschak, A.G., et al., Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet, 2004. **20**(1): p. 44-50.
81. Mandal, M. and R.R. Breaker, *Gene regulation by riboswitches.* Nat Rev Mol Cell Biol, 2004. **5**(6): p. 451-63.
82. Serganov, A. and D.J. Patel, Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. Nat Rev Genet, 2007. **8**(10): p. 776-90.
83. Brantl, S., Bacterial gene regulation: from transcription attenuation to riboswitches and ribozymes. Trends Microbiol, 2004. **12**(11): p. 473-5.
84. Link, K.H. and R.R. Breaker, Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches. Gene Ther, 2009. **16**(10): p. 1189-201.
85. Liang, J.C. and C.D. Smolke, *Rational design and tuning of ribozyme-based devices.* Methods Mol Biol, 2012. **848**: p. 439-54.
86. Win, M.N. and C.D. Smolke, Higher-order cellular information processing with synthetic RNA devices. Science, 2008. **322**(5900): p. 456-60.
87. Bayer, T.S. and C.D. Smolke, Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat Biotechnol, 2005. **23**(3): p. 337-43.
88. Jin, Y., et al., Use of a Riboswitch-controlled Conditional Hypomorphic Mutation to Uncover a Role for the Essential csrA Gene in Bacterial Autoaggregation. Journal of Biological Chemistry, 2009. **284**(42): p. 28738-28745.
89. Topp, S. and J.P. Gallivan, *Guiding bacteria with small molecules and RNA.* J Am Chem Soc, 2007. **129**(21): p. 6807-11.
90. Sinha, J., S.J. Reyes, and J.P. Gallivan, *Reprogramming bacteria to seek and destroy an herbicide.* Nat Chem Biol, 2010. **6**(6): p. 464-70.
91. Cho, E.J., J.W. Lee, and A.D. Ellington, *Applications of Aptamers as Sensors.* Annual Review of Analytical Chemistry, 2009. **2**: p. 241-264.
92. Rajendran, M. and A.D. Ellington, *Selection of fluorescent aptamer beacons that light up in the presence of zinc.* Anal Bioanal Chem, 2008. **390**(4): p. 1067-75.

93. Hall, B., et al., *Kinetic optimization of a protein-responsive aptamer beacon.* Biotechnol Bioeng, 2009. **103**(6): p. 1049-59.
94. Javier, D.J., et al., Aptamer-targeted gold nanoparticles as molecular-specific contrast agents for reflectance imaging. Bioconjug Chem, 2008. **19**(6): p. 1309-12.
95. Yang, L. and A.D. Ellington, Real-time PCR detection of protein analytes with conformation-switching aptamers. Anal Biochem, 2008. **380**(2): p. 164-73.
96. Yang, L., et al., Real-time rolling circle amplification for protein detection. Anal Chem, 2007. **79**(9): p. 3320-9.
97. Seelig, G., et al., *Enzyme-free nucleic acid logic circuits.* Science, 2006. **314**(5805): p. 1585-8.
98. Benenson, Y., et al., Programmable and autonomous computing machine made of biomolecules. Nature, 2001. **414**(6862): p. 430-434.
99. Benenson, Y., et al., An autonomous molecular computer for logical control of gene expression. Nature, 2004. **429**(6990): p. 423-429.
100. Bi, S., et al., *Colorimetric Logic Gates Based on Supramolecular DNAzyme Structures.* Angewandte Chemie-International Edition, 2010. **49**(26): p. 4438-4442.
101. Baron, R., et al., *Logic gates and elementary computing by enzymes.* Journal of Physical Chemistry A, 2006. **110**(27): p. 8548-8553.
102. Niazov, T., et al., *Concatenated logic gates using four coupled biocatalysts operating in series.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(46): p. 17160-17163.
103. Zhan, J., et al., Develop reusable and combinable designs for transcriptional logic gates. Mol Syst Biol, 2010. **6**: p. 388.
104. Kramer, B.P., C. Fischer, and M. Fussenegger, *BioLogic gates enable logical transcription control in mammalian cells.* Biotechnol Bioeng, 2004. **87**(4): p. 478-84.
105. Kuhlman, T., et al., *Combinatorial transcriptional control of the lactose operon of Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(14): p. 6043-6048.
106. Hammer, K., I. Mijakovic, and P.R. Jensen, *Synthetic promoter libraries - tuning of gene expression.* Trends in Biotechnology, 2006. **24**(2): p. 53-55.
107. Busby, S. and R.H. Ebright, Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell, 1994. **79**(5): p. 743-6.
108. Hunziker, A., et al., *Genetic flexibility of regulatory networks.* Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(29): p. 12998-13003.
109. Lopez, P.J., et al., On the mechanism of inhibition of phage T7 RNA polymerase by lac repressor. J Mol Biol, 1998. **276**(5): p. 861-75.
110. Alon, U., *Simplicity in biology.* Nature, 2007. **446**(7135): p. 497.
111. Alon, U., *Biological networks: the tinkerer as an engineer.* Science, 2003. **301**(5641): p. 1866-7.
112. Thieffry, D., et al., From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. Bioessays, 1998. **20**(5): p. 433-40.
113. Madar, D., et al., Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli. BMC Syst Biol, 2011. **5**: p. 111.
114. Becskei, A., B. Seraphin, and L. Serrano, Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J, 2001. **20**(10): p. 2528-35.

115. Wolf, D.M. and A.P. Arkin, *Motifs, modules and games in bacteria.* Curr Opin Microbiol, 2003. **6**(2): p. 125-34.
116. Ellis, T., T. Adie, and G.S. Baldwin, *DNA assembly for synthetic biology: from parts to pathways and beyond.* Integr Biol (Camb), 2011. **3**(2): p. 109-18.
117. Du, L., R. Gao, and A.C. Forster, Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase. Biotechnol Bioeng, 2009. **104**(6): p. 1189-96.
118. Noireaux, V., Y.T. Maeda, and A. Libchaber, *Development of an artificial cell, from self-organization to computation and self-reproduction.* Proc Natl Acad Sci U S A, 2011. **108**(9): p. 3473-80.
119. Jewett, M.C., et al., An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol, 2008. **4**: p. 220.
120. Eric Hodgman, C. and M.C. Jewett, *Cell-free synthetic biology: Thinking outside the cell.* Metab Eng, 2011.
121. Simpson, M.L., Cell-free synthetic biology: a bottom-up approach to discovery by design. Mol Syst Biol, 2006. **2**: p. 69.
122. Doktycz, M.J. and M.L. Simpson, *Nano-enabled synthetic biology.* Mol Syst Biol, 2007. **3**: p. 125.
123. Karig, D.K., et al., Expression optimization and synthetic gene networks in cell-free systems. Nucleic Acids Res, 2011.
124. Tabor, S. and C.C. Richardson, A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc Natl Acad Sci U S A, 1985. **82**(4): p. 1074-8.
125. Sakaguchi-Mikami, A., et al., Construction of a novel glucose-sensing molecule based on a substrate-binding protein for intracellular sensing. Biotechnol Bioeng, 2010.
126. Bello-Roufai, M., T. Roulon, and C. Escude, Ligand-mediated transcription elongation control using triplex-based padlock oligonucleotides. Chem Biol, 2004. **11**(4): p. 509-16.
127. Skoog, J.U. and L.J. Maher, 3rd, *Repression of bacteriophage promoters by DNA and RNA oligonucleotides.* Nucleic Acids Res, 1993. **21**(9): p. 2131-8.
128. Maher, L.J., 3rd, Inhibition of T7 RNA polymerase initiation by triple-helical DNA complexes: a model for artificial gene repression. Biochemistry, 1992. **31**(33): p. 7587-94.
129. Buskirk, A.R., A. Landrigan, and D.R. Liu, *Engineering a ligand-dependent RNA transcriptional activator.* Chem Biol, 2004. **11**(8): p. 1157-63.
130. An, C.I., V.B. Trinh, and Y. Yokobayashi, Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. RNA, 2006. **12**(5): p. 710-6.
131. Suess, B., et al., Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res, 2003. **31**(7): p. 1853-8.
132. Bunka, D.H. and P.G. Stockley, *Aptamers come of age - at last.* Nat Rev Microbiol, 2006. **4**(8): p. 588-96.
133. Topp, S. and J.P. Gallivan, *Emerging applications of riboswitches in chemical biology.* ACS Chem Biol, 2010. **5**(1): p. 139-48.
134. Milligan, J.F., et al., Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. Nucleic Acids Res, 1987. **15**(21): p. 8783-98.
135. Baldrich, E., A. Restrepo, and C.K. O'Sullivan, Aptasensor development: elucidation of critical parameters for optimal aptamer performance. Anal Chem, 2004. **76**(23): p. 7053-63.

136. Harrington, C. and F.W. Perrino, *Initiation of RNA-primed DNA synthesis in vitro by DNA polymerase alpha-primase.* Nucleic Acids Res, 1995. **23**(6): p. 1003-9.
137. Macaya, R.F., et al., Structural and functional characterization of potent antithrombotic oligonucleotides possessing both quadruplex and duplex motifs. Biochemistry, 1995. **34**(13): p. 4478-92.
138. Alberts, B.M., Function of gene 32-protein, a new protein essential for the genetic recombination and replication of T4 bacteriophage DNA. Fed Proc, 1970. **29**(3): p. 1154-63.
139. Chang, A.L., J.J. Wolf, and C.D. Smolke, Synthetic RNA switches as a tool for temporal and spatial control over gene expression. Curr Opin Biotechnol, 2012.
140. Win, M.N. and C.D. Smolke, A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A, 2007. **104**(36): p. 14283-8.
141. Mitchell, L.G. and C.R. Merril, Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction. Anal Biochem, 1989. **178**(2): p. 239-42.
142. Lin, C., et al., Rolling circle enzymatic replication of a complex multi-crossover DNA nanostructure. J Am Chem Soc, 2007. **129**(46): p. 14475-81.
143. Broxson, C., J. Beckett, and S. Tornaletti, Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene. Biochemistry, 2011. **50**(19): p. 4162-72.
144. Wolffe, A.P. and H.R. Drew, *Initiation of transcription on nucleosomal templates.* Proc Natl Acad Sci U S A, 1989. **86**(24): p. 9817-21.
145. Portugal, J. and A. Rodriguez-Campos, *T7 RNA polymerase cannot transcribe through a highly knotted DNA template.* Nucleic Acids Res, 1996. **24**(24): p. 4890-4.
146. Zhou, W., D. Reines, and P.W. Doetsch, T7 RNA polymerase bypass of large gaps on the template strand reveals a critical role of the nontemplate strand in elongation. Cell, 1995. **82**(4): p. 577-85.
147. Schultze, P., R.F. Macaya, and J. Feigon, Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). J Mol Biol, 1994. **235**(5): p. 1532-47.
148. Neaves, K.J., et al., Direct visualization of G-quadruplexes in DNA using atomic force microscopy. Nucleic Acids Res, 2009. **37**(18): p. 6269-75.
149. Droge, P. and F.M. Pohl, The influence of an alternate template conformation on elongating phage T7 RNA polymerase. Nucleic Acids Res, 1991. **19**(19): p. 5301-6.
150. Liu, M., H. Asanuma, and M. Komiyama, *Azobenzene-tethered T7 promoter for efficient photoregulation of transcription.* J Am Chem Soc, 2006. **128**(3): p. 1009-15.
151. Dubendorff, J.W. and F.W. Studier, Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. J Mol Biol, 1991. **219**(1): p. 45-59.
152. Maniatis, T., E.F. Fritsch, and J. Sambrook, *Molecular cloning : a laboratory manual*. 1982, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory. x, 545 p.
153. Wang, B., et al., Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat Commun, 2011. **2**: p. 508.
154. Friedland, A.E., et al., *Synthetic gene networks that count.* Science (New York, N Y), 2009. **324**(5931): p. 1199-202.
155. Ro, D.K., et al., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature, 2006. **440**(7086): p. 940-3.

156. Mangan, S., A. Zaslaver, and U. Alon, The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. J Mol Biol, 2003. **334**(2): p. 197-204.
157. Vilar, J.M. and S. Leibler, *DNA looping and physical constraints on transcription regulation.* J Mol Biol, 2003. **331**(5): p. 981-9.
158. Oehler, S., S. Alberti, and B. Muller-Hill, Induction of the lac promoter in the absence of DNA loops and the stoichiometry of induction. Nucleic Acids Res, 2006. **34**(2): p. 606-12.
159. Oehler, S., et al., Quality and position of the three lac operators of E. coli define efficiency of repression. EMBO J, 1994. **13**(14): p. 3348-55.
160. Bond, L.M., et al., *Gene repression by minimal lac loops in vivo.* Nucleic Acids Res, 2010. **38**(22): p. 8072-82.
161. Ikeda, R.A., C.M. Ligman, and S. Warshamana, *T7 promoter contacts essential for promoter activity in vivo.* Nucleic Acids Res, 1992. **20**(10): p. 2517-24.
162. Kinkhabwala, A. and C.C. Guet, *Uncovering cis regulatory codes using synthetic promoter shuffling.* PLoS One, 2008. **3**(4): p. e2030.
163. Hunziker, A., et al., *Genetic flexibility of regulatory networks.* Proc Natl Acad Sci U S A, 2010. **107**(29): p. 12998-3003.
164. Muller, J., S. Oehler, and B. Muller-Hill, Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. J Mol Biol, 1996. **257**(1): p. 21-9.
165. Carra, J.H. and R.F. Schleif, Variation of half-site organization and DNA looping by AraC protein. EMBO J, 1993. **12**(1): p. 35-44.
166. Weickert, M.J. and S. Adhya, Control of transcription of gal repressor and isorepressor genes in Escherichia coli. J Bacteriol, 1993. **175**(1): p. 251-8.
167. Schleif, R., *DNA looping.* Annu Rev Biochem, 1992. **61**: p. 199-223.
168. Han, L., et al., Concentration and length dependence of DNA looping in transcriptional regulation. PLoS One, 2009. **4**(5): p. e5621.
169. Lanzer, M. and H. Bujard, *Promoters largely determine the efficiency of repressor action.* Proc Natl Acad Sci U S A, 1988. **85**(23): p. 8973-7.
170. Isalan, M. and Y. Choo, Engineered zinc finger proteins that respond to DNA modification by HaeIII and HhaI methyltransferase enzymes. J Mol Biol, 2000. **295**(3): p. 471-7.
171. Babiskin, A.H. and C.D. Smolke, Synthetic RNA modules for fine-tuning gene expression levels in yeast by modulating RNase III activity. Nucleic Acids Res, 2011. **39**(19): p. 8651-64.
172. Win, M.N., J.C. Liang, and C.D. Smolke, *Frameworks for programming biological function through RNA parts and devices.* Chem Biol, 2009. **16**(3): p. 298-310.
173. Voigt, S., et al., Evaluation of the Jena skin flap model for the training of the closure of facial skin defects. Biomed Tech (Berl), 2012.
174. Rosenfeld, N., et al., Accurate prediction of gene feedback circuit behavior from component properties. Mol Syst Biol, 2007. **3**.
175. Keum, J.W., et al., The presence of a common downstream box enables the simultaneous expression of multiple proteins in an E. coli extract. Biochem Biophys Res Commun, 2006. **350**(3): p. 562-7.
176. Babiskin, A.H. and C.D. Smolke, A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. Mol Syst Biol, 2011. **7**: p. 471.

177. Jewett, M.C. and J.R. Swartz, Substrate replenishment extends protein synthesis with an in vitro translation system designed to mimic the cytoplasm. Biotechnol Bioeng, 2004. **87**(4): p. 465-72.
178. Jewett, M.C. and J.R. Swartz, Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol Bioeng, 2004. **86**(1): p. 19-26.
179. Michel-Reydellet, N., K. Calhoun, and J. Swartz, Amino acid stabilization for cell-free protein synthesis by modification of the Escherichia coli genome. Metab Eng, 2004. **6**(3): p. 197-203.
180. Nakano, H., et al., Efficient coupled transcription/translation from PCR template by a hollow-fiber membrane bioreactor. Biotechnol Bioeng, 1999. **64**(2): p. 194-9.
181. Spirin, A.S., High-throughput cell-free systems for synthesis of functionally active proteins. Trends Biotechnol, 2004. **22**(10): p. 538-45.
182. Kim, J., K.S. White, and E. Winfree, Construction of an in vitro bistable circuit from synthetic transcriptional switches. Mol Syst Biol, 2006. **2**.
183. Kim, J. and E. Winfree, *Synthetic in vitro transcriptional oscillators.* Mol Syst Biol, 2011. **7**.
184. Karig, D.K. and M.L. Simpson, *Tying new knots in synthetic biology.* HFSP J, 2008. **2**(3): p. 124-8.
185. Guido, N.J., et al., *A bottom-up approach to gene regulation.* Nature, 2006. **439**(7078): p. 856-860.
186. Monnard, P.A., Liposome-entrapped Polymerases as Models for Microscale/Nanoscale Bioreactors. Journal of Membrane Biology, 2003. **191**(2): p. 87-97.
187. Noireaux, V. and A. Libchaber, *A vesicle bioreactor as a step toward an artificial cell assembly.* Proc Natl Acad Sci U S A, 2004. **101**(51): p. 17669-74.
188. Pereira de Souza, T., P. Stano, and P.L. Luisi, The minimal size of liposome-based model cells brings about a remarkably enhanced entrapment and protein synthesis. Chembiochem, 2009. **10**(6): p. 1056-63.
189. Fletcher, B.L., et al., Microarrays of Biomimetic Cells Formed by the Controlled Synthesis of Carbon Nanofiber Membranes. Nano Letters, 2004. **4**(10): p. 1809-1814.
190. Karig, D.K., et al., Model for biological communication in a nanofabricated cell-mimic driven by stochastic resonance. Nano Communication Networks, 2011. **In Press, Corrected Proof**.
191. Retterer, S.T., et al., Development and fabrication of nanoporous silicon-based bioreactors within a microfluidic chip. Lab on a Chip, 2010. **10**(9): p. 1174-1181.
192. Siuti, P., et al. Cell free translation in engineered picoliter volume containers. in Biomedical Science & Engineering Conference, 2009. BSEC 2009. First Annual ORNL. 2009.
193. Milne, L., et al., An approach to gene-specific transcription inhibition using oligonucleotides complementary to the template strand of the open complex. Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3136-41.
194. Szita, N., et al., *Microfluidic approaches for systems and synthetic biology.* Curr Opin Biotechnol, 2010. **21**(4): p. 517-23.
195. Retterer, S.T., et al., Development and fabrication of nanoporous silicon-based bioreactors within a microfluidic chip. Lab Chip, 2010. **10**(9): p. 1174-81.
196. Glass, J.I., et al., *Essential genes of a minimal bacterium.* Proc Natl Acad Sci U S A, 2006. **103**(2): p. 425-30.

Appendices

Appendix A- Methods for construction of double stranded DNA templates with ssDNA regions

Several strategies were tested to construct double stranded DNA templates with melted or ssDNA regions inserted downstream to transcriptional start site- Here, I summarize the strategies that were tested but eventually not adopted for assembling dsDNA templates with ssDNA regions. The pros and cons of each of these techniques are also described.

1. Ligation of PCR products for generation of linear templates with gap regions

The initial design of the aptamer templates was generate to linear templates with a gap in the template strand so as to accommodate the aptamer on the non-template strand. 3 different fragments of DNA templates corresponding to promoter, aptamer and reporter DNA were ligated together to generate the desired template. The long linear template that encoded GFP was generated using PCR. The PCR primer that annealed on the 5' end contained a Hpy99I restriction site and was subsequently digested with Hpy99I restriction enzyme to generate a template with a 5 base long 3' overhang. A short oligo corresponding to the IgE aptamer with ends complementary to the GFP fragment and T7 promoter region were added to a 20μl ligation reaction. The gapped template output was difficult to evaluate as the length of the fragment attached was ~50 bases long which does not yield an observable shift on an agarose gel. The strategy was discarded because of the low yield of the reaction (Figure A1).

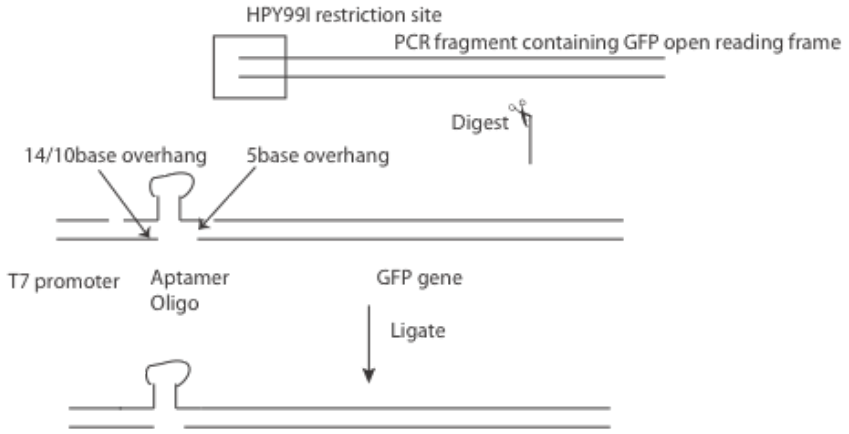


Figure A1- Schematic showing ligation based strategy to generate templates for testing aptamer-mediated regulation.

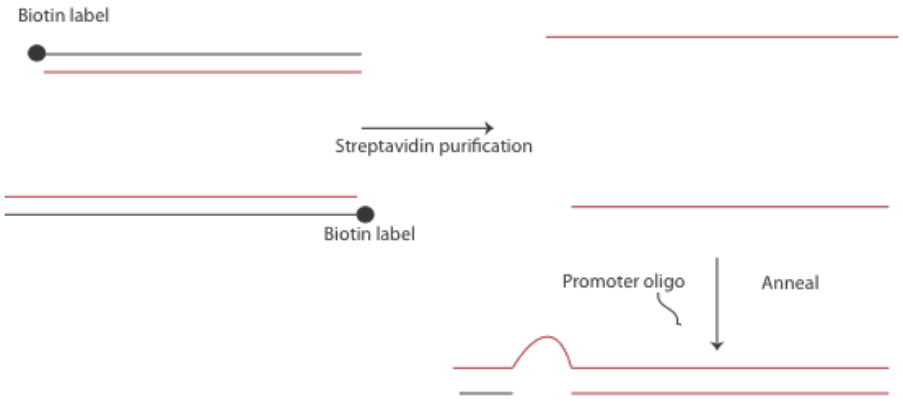


Figure A2 Schematic for aptamer template assembly from ssDNA templates generated using Streptavidin-Biotin purification

2.Generation of linear ssDNA templates using biotinylated PCR

As an alternative strategy to assembling gapped linear DNA templates, instead of ligating three DNA products, ssDNA strands corresponding to the template and the non-template strands were generated. Single strands for assembling the DNA templates were generated using a streptavidin-biotin mediated affinity purification strategy[1]. Briefly, a biotin tag on the strand complementary to the strand needed for template generation was incorporated on a PCR primer. PCR product was amplified using this primer and purified using Qiagen PCR purification kit. The product was resuspended in Biotin binding buffer (10mM Tris-HCl pH 7.5, 1M NaCl, 0.01% Triton X-100) and added to 100 μl streptavidin agarose beads (Thermoscientific) and incubated overnight at 4 °C. The beads were subsequently washed with the binding buffer 3 times and finally incubated with 150μl of 0.2M NaOH for 6 minutes to denature dsDNA. The beads were added to a column and the ssDNA eluted and neutralized using acetic acid. The resulting DNA was purified once using the Qiagen PCR purification kit and subsequently annealed with the complementary ssDNA to yield a double stranded DNA template. The individual products were analyzed using agarose gel electrophoresis.

We were able to generate ssDNA using this procedure, although some dsDNA products were also present in the final DNA preparation. The annealed products were found to be 1kb, which corresponds to the size of correctly annealed DNA products. However, the yield was around 50ng/μl in 30μl volume, which is not sufficient for driving high yield expression in cell extracts (Figure A2).

3.Generation of ssDNA templates using rolling circle amplification

Finally we tested a strategy to generate long ssDNA templates using phi29 DNA polymerase. Phi29 DNA polymerase is commonly used to achieve repeated extension of a primer annealed to a circular DNA template thereby generating tandem copies of the DNA template[2]. Commonly, a primer is annealed to the template and the resulting double stranded DNA is digested using a restriction enzyme to resolve the amplified product into shorter DNA fragments that correspond to a complement of the original DNA template. To generate ssDNA, T7 terminator primer was annealed to ssDNA template generated from pBluescript KS II (+) and amplified by phi29 DNA polymerase (NEB) for 16 hours at room temperature. The template thus obtained was annealed to cT7asymterm primer that contained an EcoRV restriction site and subsequently digested using EcoRV restriction enzyme. While the yields of amplified DNA were fairly high (5ug of DNA from 100ng of ssDNA), the products were insufficient for generating DNA for cell free protein synthesis.

1. Mitchell, L.G. and C.R. Merril, *Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction.* Anal Biochem, 1989. **178**(2): p. 239-42.
2. Lin, C.X., et al., *Rolling circle enzymatic replication of a complex multi-crossover DNA nanostructure.* Journal of the American Chemical Society, 2007. **129**(46): p. 14475-14481.

Appendix B- Translational Regulation of gene expression using DNA aptamers

Introduction

Ligand sensitive mechanisms for gene regulation find extensive use in biotechnology and in applied sciences. They allow for temporal regulation of expression in response to environmental signals and facilitate communication between the external environment and the components of a gene circuit[1].

In this regard, nucleic acid aptamers have been extensively used for achieving expression regulation. The conformational change that ensues ligand binding is coupled to a functional domain that mediates the change in expression. These responses include change in access of the ribosome to the ribosome-binding site, blocking ribosome scanning or by transcription attenuation[2]. In addition, RNA aptamers coupled to an antisense RNA molecule added in trans enables modular control of expression. While RNA aptamers have been used to achieve gene regulation, the potential of DNA aptamers to regulate translation has not yet been realized.

DNA has several advantages over RNA as a regulation molecule. DNA oligonucleotides have longer lifetimes in cells and cell free systems. In addition, end protected DNA protected by phosphorothioate bonds prevent digestion of DNA in cells and cell free systems. As there are no membrane barriers, cell free systems are particularly amenable for utilizing large amounts of exogenous oligonucleotides.

Here, translational regulation of gene expression using DNA aptamers in cell free systems is described. DNA oligonucleotides have been harnessed for regulating gene expression in cells and *in vitro* systems and have been promising candidates for gene therapy. DNA oligonucleotides have been utilized for regulating both transcription and

translation in cell free systems. For instance, DNA oligonucleotides have been targeted to polypyrimidine tracts, which are capable of forming triplex structures and have been employed to repress transcription from viral, bacterial and eukaryotic promoters[3, 4]. These strategies suffer from the disadvantage that DNA triplexes can be formed by specialized sequences and under specific ionic conditions. Alternatively, chemically modified oligonucleotides may be used that allow for triplex formation under a variety of conditions.

Perhaps the most extensively used DNA based gene regulation strategy is the antisense regulation of expression. A DNA oligonucleotide is hybridized to a complementary RNA sequence, which is subsequently cleaved by RNaseH thereby knocking down gene expression[5]. This strategy has been widely adopted for use in both cells and cell free systems. In contrast to the triplex DNA strategy, antisense DNA technology does not have special sequence requirements or ionic conditions for enabling DNA hybridization.

While these strategies describe constitutive repression from these templates, an approach to enable ligand responsive antisense DNA aptamers would be very useful.

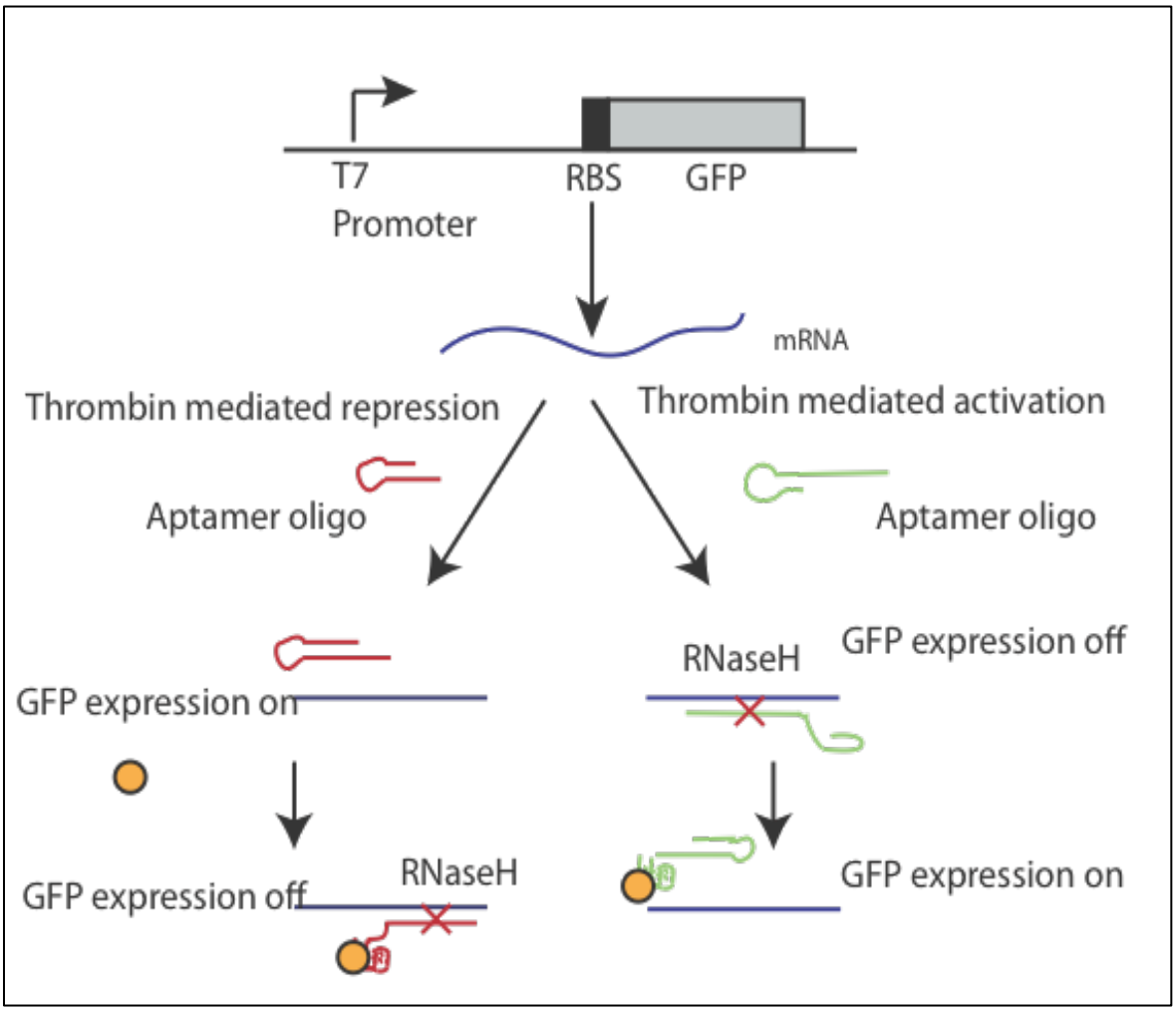


Figure B1- Schematic of translation activation and repression using DNA aptamers.

Results and Discussion

Here we describe preliminary of efforts to enable ligand sensitivity onto DNA aptamers for achieving ligand dependent translational regulation (Figure B1). Published studies on translational repression from trans-acting DNA oligonucleotides utilized a 20 bases DNA oligonucleotide to knock down gene expression in *E.coli* cell extracts[6].

Table B1 List of oligonucleotides used for translational regulation in this study

	Antisense DNA length	Sequence
Antil1thr	22	CGTACGGTTGGTGTGGTTGGTGTACGTTTTTTTATCTTGAAAAGCATTGAACACCA
Antil2thr	22	CGTAGGGTTGGTGTGGTTGGTCTACGTTTTTTTATCTTGAAAAGCATTGAACACCA
Antis1thr	18	CGTACGGTTGGTGTGGTTGGTGTACGTTTTTTGAAAAGCATTGAACACCA
Antis2thr	18	CGTAGGGTTGGTGTGGTTGGTCTACGTTTTTTGAAAAGCATTGAACACCA
Antis	18	GAAAAGCATTGAACACCA
Antil	22	TATCTTGAAAAGCATTGAACACCA

A 22 base and an 18 base long antisense DNA oligonucleotide were utilized to repress gene expression. The 5' end of antisense DNA oligonucleotide was modified to contain a thrombin aptamer region such that, in the absence of thrombin, the thrombin aptamer region would hybridize with the antisense region thereby preventing its interaction with the cognate RNA molecule. It was hypothesized that upon binding to thrombin, however, the antisense region on the DNA would be exposed thereby repressing expression. An RNA sequence bearing partial complementarity to the thrombin aptamer region was selected to facilitate the design. The oligonucleotides used in this study are listed in Table B1. The red sequence corresponds to the thrombin aptamer region and blue sequence refers to the antisense sequence. The program MFOLD was used to predict the secondary structure under the salt conditions used[7].

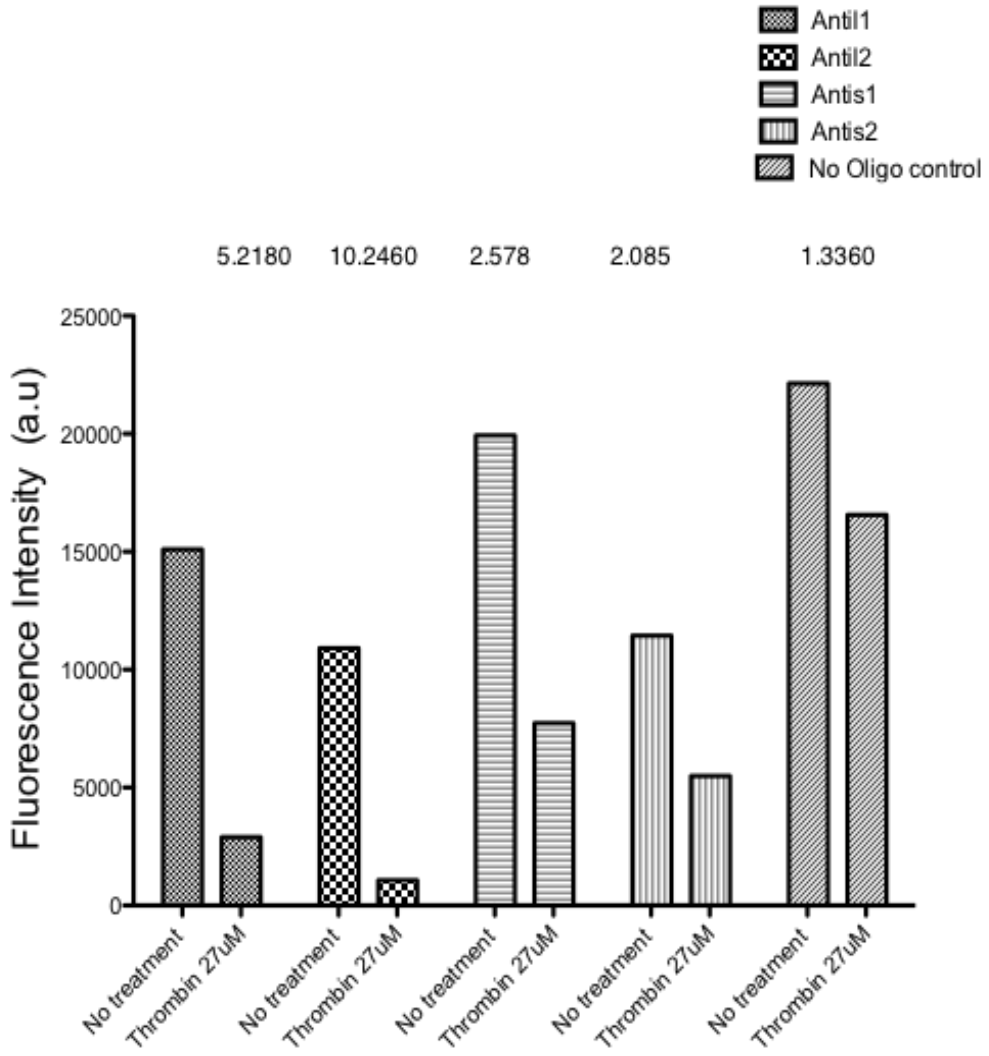


Figure B2 Graph shows response from pKSGFP plasmids to long and short versions of aptamer containing repressor oligonucleotides.

Before using them in an experiment, 9µl of 100µM oligonucleides in a volume of 10 µl of solution were heat denatured in the presence of 10mM Tris-HCl pH 7.5 and 50mM NaCl. 2µl of the heat-denatured oligonucleotide was incubated with 1ul of 270µM thrombin solution and 0.1% Tween-20 for 10 minutes. To test the effect of oligonucleotides on gene expression in cell extracts, these oligonucleotides in the presence and absence of thrombin were added to *E.coli* cell extract along with 180µg of GFP expressing plasmid (pKSGFP described in chapter 2). To carry out protein synthesis, the High yield protein synthesis kit from Promega corporation was used and the reactions were assembled as recommended by the manufacturer except that the final reaction volume was 15µl and the reaction was assembled in a Corning 384 well plate. To prevent drying 10ul of mineral oil was overlaid on the reaction and fluorescence measurements were made every 7 minutes for 6 hours.

As shown in Figure B2, oligonucleotides Antil1 and Antil2 that have longer antisense regions resulted in an increased thrombin dependent repression of gene expression over oligonucleotides with shorter antisense region. Furthermore, as shown in Figure B3, thrombin dependent repression was observed only with the aptamer containing repression oligonucleotide AntiL1 and AntilL2. Extract that contained just the antisense DNA region showed the same level of reduction in the presence of thrombin as the control reaction without any antisense oligonucleotides added.

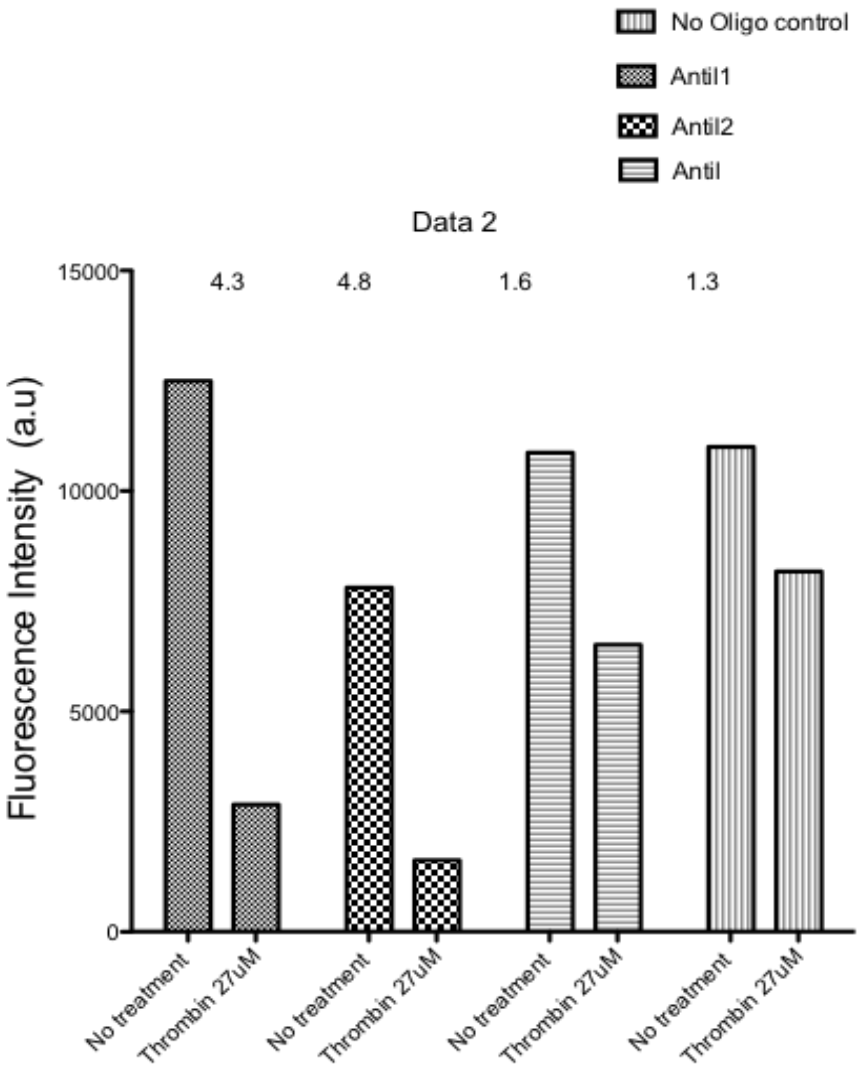


Figure B3 – Graph shows changes in expression upon addition of repressor with (Antil1 and Antil2) and without (Antil) thrombin aptamer

Future work

Different concentrations of thrombin and oligonucleotides would need to be tested to determine the critical concentrations of these agents for achieving effective repression. Furthermore, designs for achieving thrombin dependent activation would need to be explored to examine if DNA aptamer mediated activation of expression can be achieved.

1. Buskirk, A.R., A. Landrigan, and D.R. Liu, *Engineering a ligand-dependent RNA transcriptional activator.* Chem Biol, 2004. **11**(8): p. 1157-63.
2. Link, K.H. and R.R. Breaker, *Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches.* Gene Ther, 2009. **16**(10): p. 1189-201.
3. Maher, L.J., 3rd, *Inhibition of T7 RNA polymerase initiation by triple-helical DNA complexes: a model for artificial gene repression.* Biochemistry, 1992. **31**(33): p. 7587-94.
4. Skoog, J.U. and L.J. Maher, 3rd, *Repression of bacteriophage promoters by DNA and RNA oligonucleotides.* Nucleic Acids Res, 1993. **21**(9): p. 2131-8.
5. Keum, J.W., et al., *Combinatorial, selective and reversible control of gene expression using oligodeoxynucleotides in a cell-free protein synthesis system.* Biotechnol Bioeng, 2009. **102**(2): p. 577-82.
6. Ahn, J.H., T.J. Kang, and D.M. Kim, *Tuning the expression level of recombinant proteins by modulating mRNA stability in a cell-free protein synthesis system.* Biotechnol Bioeng, 2008. **101**(2): p. 422-7.
7. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

Appendix C-Towards continuous exchange cell free protein synthesis reactions in microfluidic devices

Synthetic gene circuits consist of a network of sensory elements that are integrated to perform a logical operation [1]. Successful implementation of gene circuits requires proper balancing of RNA and protein products in the reaction. Accordingly, gene circuits have been implemented *in vitro* in a batch reaction format. Prominent examples of these include cascaded gene circuits[2, 3] and a bi-stable stable switch that does not rely on protein intermediates[4]. Yet, these circuits lag behind analogous circuits implemented in cells. In particular, two features of commercially available *E.coli* cell extracts for cell free protein synthesis pose a challenge toward enabling elaborate *in vitro* gene circuits akin to those that have been built *in vivo*-

a. **Absence of mechanisms to specifically degrade RNA and proteins built up in the cell extracts-** Higher order networks can be created in environments that allow for continuous turnover of accumulated RNA and protein products. Commercially available cell extract contain proprietary optimizations that increase the mRNA and protein lifetimes. While this ensures high protein yields, this poses a significant challenge with regard to enabling cell free gene circuits in a batch reaction.
b. **Limited lifetime of cell free protein synthesis** - The cell free protein synthesis reactions in a batch mode typically last for ~2 hours. The transcription and translation reactions are inhibited by inhibitory by-products of protein synthesis and depletion of energy resources. Hence, to achieve sustained cell free protein synthesis for long periods of time require a mechanism for continuous supply of precursors and energy sources and removal of inhibitory by products.

To address these issues, I worked on developing physical platforms that extend cell free protein synthesis reaction life times and allows for removal of built up RNA and protein products. Large-scale continuous cell free protein synthesis reactions have been implemented by enclosing the cell extracts in a dialysis chamber that is suspended in the feed buffer solution. Continuous influx of fresh nutrients and efflux of inhibitory by products from the dialysis chamber extends the reaction life-time to about 24 hours[5, 6]. In addition to large-scale studies, continuous exchange reactions have also been enabled in nanoporous microfluidic platforms. These devices have been built to facilitate rapid evaluation of gene products using small amounts of extracts[7] and to mimic the size and scale of biological cells and evaluate its effect on reaction rates[8].

As a step towards achieving greater functionality in cell free systems, cell free protein synthesis reactions were conducted in picoliter scale devices. These devices have been described by Retterer *et.al*[9] and Siuti *et.al*[8]. The "cell mimic" devices are nanoporous reaction containers that are integrated with a microfluidic system. These reaction vessels have 40 micron inner diameter and the walls of the container are 15 micron high with a working volume of 18 picoliters. These vessels are embedded in a microfluidic channel to facilitate addition of fresh substrate and removal of by products[9]. The pores on the walls of the device are designed to facilitate substrate exchange between the channel and the reaction vessel. Siuti *et.al* have shown constitutive cell free protein synthesis using *E.coli* cell extracts for 24 hours using these devices[9]. The devices might ultimately be used as platforms for biosensing and for testing biological hypotheses in a cell free context (Figure C1).

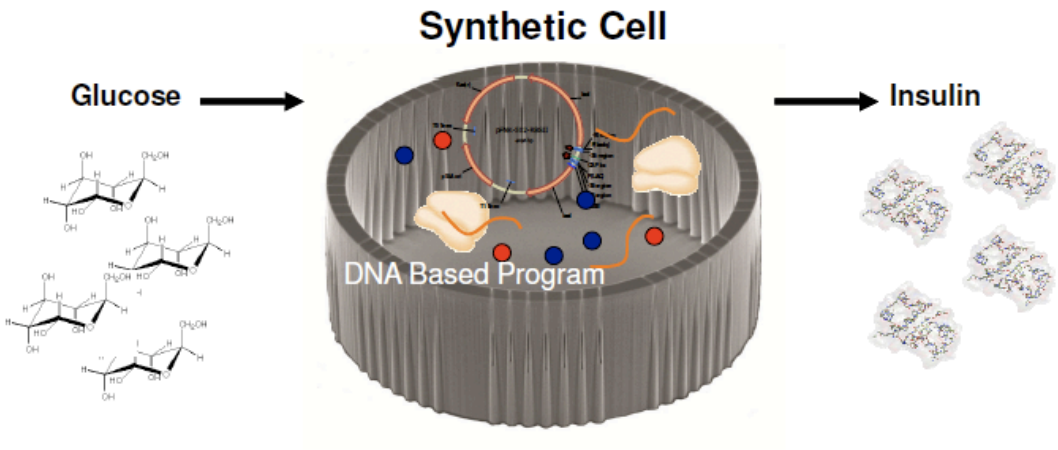


Figure C1- Schematic of an application for proposed cell mimic devices in biosensing and actualtion- A DNA program encoding genes of interest is introduced into the cell mimic device along with the cell extract containing the protein synthesis machinery. The analyte such as glucose traverses the device enclosure through the nanoporous membrane. The analyte activates expression of a gene such as insulin and the protein is then released from the device.

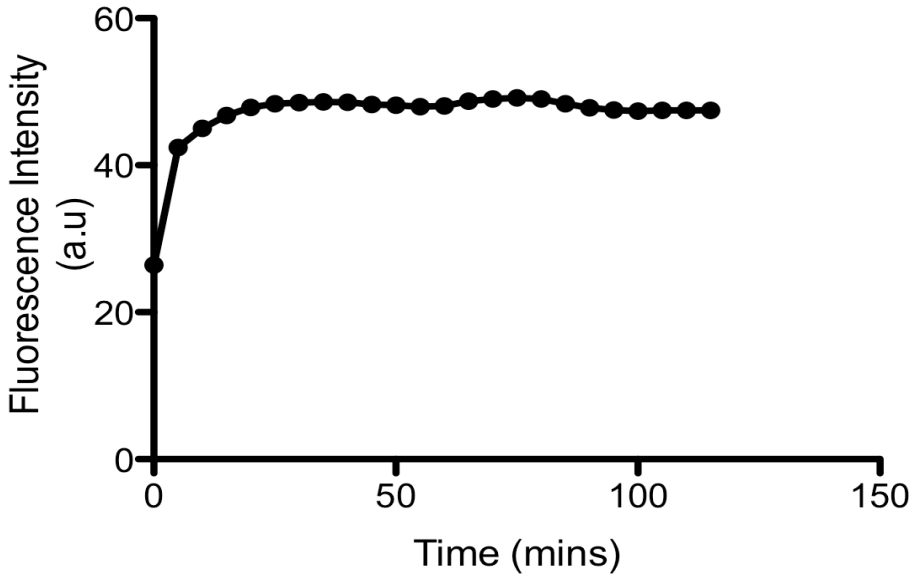


Figure C2- Graph showing change in fluorescence as a function of time in cell mimic devices- DNA encoding GFP from a T7 promoter was introduced into the device along with *E.coli* cell extract. Change in fluorescence was monitored using epifluorescence microscope.

In an effort to extend the functionality of these devices and enable negative feedback circuits in these devices, I attempted to express bi-cistronic constructs from pT7lacIGFP that expresses both the lac repressor LacI and GFP from a T7 promoter. In addition, a cell extract preparation from a different manufacturer was employed to carry out protein synthesis (High yield protein synthesis kit from Invitrogen.inc). The extracts, premixed with the buffer that supplies small molecule substrates for protein synthesis, were loaded onto the mimic device using a micromanipulator, as has been described earlier.

The device was then covered with a thin PDMS layer, which contained bores that aligned with the inlet and outlets of the microfluidic channel. The feed buffer comprising 2.5X IVPS buffer, 2X IVPS buffer and the amino acid mixture provided by the manufacturer was injected into the channel by means of a syringe pump at a rate of 5μl/hr. Change in fluorescence was monitored using an epifluorescence microscope while the silicon device rested on a heated stage set at a temperature of 37°C. As shown in Figure C2, increase in fluorescence (which is indicative of GFP expression) was observed for ~ 20 minutes and fluorescence was retained in the devices for 2 hours, as is the case of a batch reaction on a conventional scale. A decline in fluorescence was also observed at the end of the two hours, which might be due to bleaching of GFP and/or diffusion of GFP out through the pores of the device.

Several difficulties were encountered with implementing the reaction in these mimic devices

1) Large surface area to volume ratio hastened the evaporation of extract material from the device

2) Reproducible loading is rather difficult to achieve when devices are loaded using a micromanipulator. The devices are loaded with the cell extract by tapping the surface of the device with a fine loading needle attached to the micromanipulator. This prevents the loading a predetermined volume of the cell extract into the device.
3) Tight sealing of the silicon device with PDMS is needed to ensure continuous flow of liquid in the device. To accomplish this, PDMS and the silicon devices were plasma cleaned before the PDMS slab was placed on the silicon device. The extract loaded in the device appeared to be washed away upon flowing the buffer through the microfluidic channel. It is possible that there may be a gap between the top surface of the mimic device and the PDMS slab through which the extract was washed out.

To address these issues, a new device was tested which consisted of two long microfluidic channels, each of which contained the cell extract and the buffer that was separated by a porous membrane. (Figure C3) The extract was fed into the channel using a syringe pump, which eliminates the problem of uneven device loading and drying of the device.

Preliminary device characterization involved flowing the fluorescein through one channel and examining its diffusion into the other channel (Figure C4).

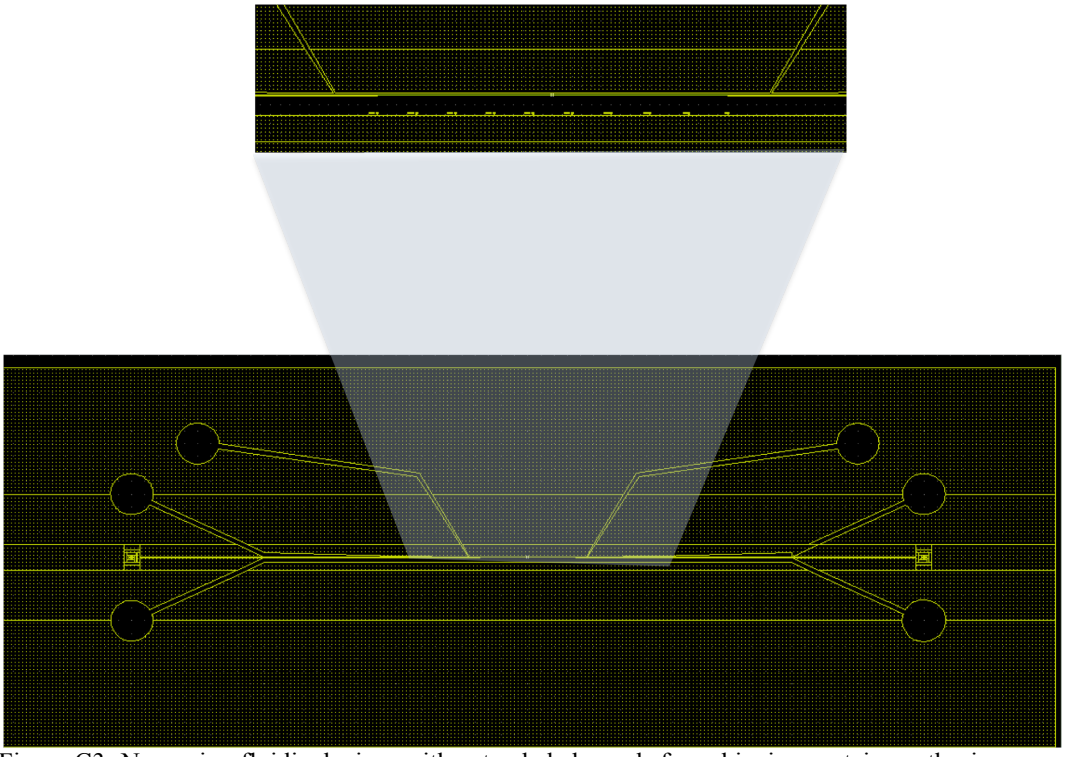


Figure C3- New microfluidic devices with extended channels for achieving protein synthesis on a microfluidic platform

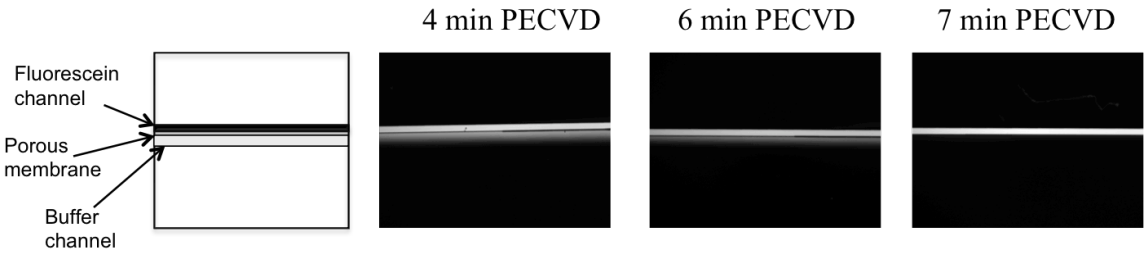


Figure C4 Comparing diffusion of fluorescein across a nanoporous membrane coated with 4 minutes, 6 minutes and 7 minutes PECVD.

We observed that devices with 4 minutes and 6 minutes PECVD permitted the diffusion of the small molecule across the membrane, whereas devices with 7 minutes PECVD did not allow for fluorescein diffusion from the fluorescein channel. Further device characterizations will be required for creating a new platform for achieving long-term protein synthesis.

Conclusions

The primary requirement for enabling long term gene expression in cell extracts is the continuous exchange of metabolites across a porous membrane that facilitates the influx of nutrients and the efflux of inhibitory by-products into the cell extract containing enclosure. To this end, picoliter cell mimic devices were employed to enable protein synthesis in a nanoporous microfluidic platform. Several issues such rapid drying of extract and incomplete sealing of the device were encountered with the picoliter devices. To address these issues a new device for enabling protein continuous protein synthesis was tested. As a prelude to enabling protein synthesis, and to model diffusion of small molecules across the nanoporous membrane, the effect of PECVD times on fluorescein diffusion across a nanoporous membrane was examined. Among the devices tested, 4 and 6 minutes PECVD devices were found to permit the diffusion of fluorescein across the channel. These devices may be used in the future to enable long-term protein synthesis in these devices.

1. Khalil, A.S. and J.J. Collins, *Synthetic biology: applications come of age.* Nat Rev Genet, 2010. **11**(5): p. 367-79.
2. Noireaux, V., R. Bar-Ziv, and A. Libchaber, *Principles of cell-free genetic circuit assembly.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12672-7.

3. Ishikawa, K., et al., *Expression of a cascading genetic network within liposomes.* FEBS Lett, 2004. **576**(3): p. 387-90.
4. Kim, J., K.S. White, and E. Winfree, *Construction of an in vitro bistable circuit from synthetic transcriptional switches.* Mol Syst Biol, 2006. **2**: p. 68.
5. Spirin, A.S., et al., *A continuous cell-free translation system capable of producing polypeptides in high yield.* Science, 1988. **242**(4882): p. 1162-4.
6. Kigawa, T. and S. Yokoyama, *A continuous cell-free protein synthesis system for coupled transcription-translation.* J Biochem, 1991. **110**(2): p. 166-8.
7. Mei, Q., et al., *Protein synthesis in a device with nanoporous membranes and microchannels.* Lab Chip, 2010. **10**(19): p. 2541-5.
8. Siuti, P., S.T. Retterer, and M.J. Doktycz, *Continuous protein production in nanoporous, picolitre volume containers.* Lab Chip, 2011. **11**(20): p. 3523-9.
9. Retterer, S.T., et al., *Development and fabrication of nanoporous silicon-based bioreactors within a microfluidic chip.* Lab Chip, 2010. **10**(9): p. 1174-81.

VITA

Sukanya Iyer was born to Mrs Lalitha Ravikumar and Mr.V.S.Ravikumar in Gujarat, India. She went to C.M.Academy where her teachers encouraged her to pursue a higher degree in Biotechnology. She went on to pursue a Bachelor of Technology Degree from SRM university, Chennai,India,and received her degree in 2007. Here, she completed a thesis under the guidance of Dr. S.Meignanlakshmi who motivated her to pursue a doctoral degree. She then accepted a graduate research assistantship at the University of Tennessee, Knoxville in 2007. During her time here, she joined the Biological and Nanoscale Systems Group at Oak Ridge National Laboratory,Oak Ridge TN and completed her dissertation under the mentorship of Dr. Mitchel J.Doktycz.