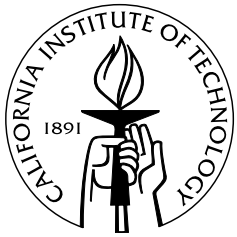
Synthetic Regulation of Eukaryotic Gene Expression by Noncoding RNA

Thesis by

Leopold Daniel d'Espaux

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy



California Institute of Technology
Pasadena, California, USA
2013

(defended May 30, 2013)

Acknowledgements

First, I would like to thank Professor Christina Smolke for having through her patience, inimitable work ethic, and abundant insight helped me develop and mature in the last several years.

I would also like to acknowledge other members of her lab, all of whom have contributed to creating a stimulating and enjoyable environment in which to think about ribozymes and clone them. In particular, Dr. Maung Win first developed the ribozyme switch platform on which this entire thesis is based, and deserves credit—more than he gives himself—for his creation. Andrew Kennedy and Jay Vowles developed the protein-responsive switches with me, and have been a constant source of good ideas, as well as murderers of bad ideas that I have come up with. Joe Liang, Yvonne Chen, Katie Galloway, Andrew Babiskin, and other past and current lab mates have helped me feel at home in the lab, among smells of *E. coli* and the constant hum of machines

Rachel Rose has worked with me these last few years, providing the healthy perspective of an amazingly talented, hardworking, and overall awesome colleague, even at her young age. She's also cloned a lot of stuff. I look forward to seeing what paths this rare combination of gifts takes her on.

To my partner Nick, thank you for being so supportive, kind, and patient, and especially so in these last few months when I've perhaps been too consumed with finishing this work to be properly reciprocal and appreciative of your affection. I've never felt closer to, or more loved by anyone. Or more admiring of anyone's thoughtfulness and *sweetness*.

And lastly, to my family I of course owe everything. I remember when we walked up the stairs to board the plane to leave Cuba, seeing my mom look back, sobbing. While at the time I was puzzled by her reaction, later as I have been reminded of that indelible moment I begin to understand the sacrifice my parents made in leaving friends, careers, language, and family behind, largely for the sake of my brother and me. Although I am sure that to them, as to most people, most of what is written in the following pages may make little sense, I hope that the fact that it is written, bound and all, may serve as a small comfort for their sacrifices.

Abstract

Synthetic biological systems promise to combine the spectacular diversity of biological functionality with engineering principles to design new life to address many pressing needs. As these engineered systems advance in sophistication, there is ever-greater need for customizable, situation-specific expression of desired genes. However, existing gene control platforms are generally not modular, or do not display performance requirements required for robust phenotypic responses to input signals. This work expands the capabilities of eukaryotic gene control in two important directions.

For development of greater modularity, we extend the use of synthetic self-cleaving ribozyme switches to detect changes in input protein levels and convey that information into programmed gene expression in eukaryotic cells. We demonstrate both up- and down-regulation of levels of an output transgene by more than 4 fold in response to rising input protein levels, with maximal output gene expression approaching the highest levels observed in yeast. *In vitro* experiments demonstrate protein-dependent ribozyme activity modulation. We further demonstrate the platform in mammalian cells. Our switch devices do not depend on special input protein activity, and can be tailored to respond to any input protein to which a suitable RNA aptamer can be developed. This platform can potentially be employed to regulate the expression of any transgene or any endogenous gene by 3' UTR replacement, allowing for more complex cell state-specific reprogramming

We also address an important concern with ribozyme switches, and riboswitch performance in general, their dynamic range. While riboswitches have generally allowed for versatile and modular regulation, so far their dynamic ranges of output gene modulation have been modest, generally at most ~10-fold. We address this shortcoming by developing a modular genetic amplifier for near-digital control over eukaryotic gene expression. We combine ribozyme switch-mediated regulation of a synthetic TF with TF-mediated regulation of an output gene. The amplifier platform allows for as much as 20-fold regulation of output gene expression in response to input signal, with maximal expression approaching the highest levels observed in yeast, yet being tunable to intermediate and lower expression levels. EC_{50} values are more than 4 times lower than in previously best-performing, non-amplifier ribozyme switches. The system design retains the modular-input architecture of the ribozyme switch platform, and the near-digital dynamic ranges of TF-based gene control.

Together, these developments promise to allow for wider applicability of these platforms for better-performing eukaryotic gene regulation, and more sophisticated, customizable reprogramming of cellular activity.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Synthetic biology expands possibilities

The work described in this thesis aims to be a very small part in the effort to coopt the myriad functionalities that exist in the natural world and to engineer new life to address some of our most pressing challenges. After billions of years of evolution, we find on our planet an estimated 6 million taxa of prokaryotes, encompassing a perhaps unknowable number of genetically distinct organisms (Curtis et al., 2002), along with some 8.7 million species of eukaryotes (Mora et al., 2011). Many of these organisms have multiple cell types, and many have genetic programs that can be regulated temporally and spatially. The end result is a spectacular diversity of genetically programmed functionality.

Researchers first directly manipulated genes of living organisms more than 40 years ago (Jackson et al., 1972), and have since steadily expanded the toolkit of genetic engineering technologies. Biochemical understanding and sequencing of an ever-increasing number of genomes continues to expand our understanding of gene function. And directed evolution, incorporation of non-natural amino acids, and high-throughput gene modification and activity screening methods allow for development of novel gene function (Bloom et al., 2005; Wang et al., 2009; Isaacs et al., 2011; Johnson et al., 2011; Liang et al., 2012; Michener and Smolke, 2012). In the last decade, the field of synthetic biology has emerged to combine biology with engineering design and construction principles to develop biological systems

with more novel and complex function. Some examples include genetic oscillators (Elowitz and Leibler, 2000), logic gates (Win and Smolke, 2008a; Xie et al., 2011; Moon et al., 2012) and other architectures that can perform computation and allow for programmed gene expression, aiming to begin to recapitulate the sophisticated circuitry of natural systems. This synthetic biology promises to allow for powerful solutions to address many of society's most pressing needs.

1.1.1 Metabolic engineering

Among our most daunting challenges is to be able to supply chemicals needed for modern life in ways that are safe, efficient, and sustainable. Our ancestors first began using the yeast *Saccharomyces cerevisiae* to produce ethanol almost 9,000 years ago (McGovern et al., 2004). Similarly, molds that produce antibiotics have been used to treat infection since at least 2,000 BCE (Forrest, 1982). Even with modern science and chemical synthesis, some 63% of new drugs are still derived from natural products (Newman and Cragg, 2007), and many are produced by methods involving extraction from natural organisms, a testament to the potential of biology for synthesis of chemicals and materials. However, natural microbial organisms did not evolve to produce compounds useful to humans, or do to so in a way that is convenient for large-scale industrial production.

In recent years, the emerging field of metabolic engineering has had large success in engineering microbes to produce valuable compounds not naturally produced in these organisms. Microbial synthesis has several attractive traits when contrasted with traditional manufacturing methods. Because many reactions can be done enzymatically inside cells, microbial production may often

obviate or reduce the need for toxic chemicals. Biosynthetic pathways can be engineered to take as starting material renewable, easily acquired products, such as sugar or biomass. Traditional chemical synthesis methods, by contrast, often require oil and gas products, which are not renewable, often subject to geopolitical and geographic difficulties, and in many instances harmful to people or to the environment. To date, researchers have engineered metabolic pathways to produce the anti-malarial drug precursor artemisinic acid (Ro et al., 2006); benzylisoquinoline alkaloids (Hawkins and Smolke, 2008); biofuels from alkanes (Schirmer et al., 2010), *n*-butanols (Bond-Watts et al., 2011), higher alcohols (Atsumi et al., 2008; Dellomonaco et al., 2011), terpenes (Peralta-Yahya et al., 2011), and fatty acids (Steen et al., 2010; Dellomonaco et al., 2011), as well as many other important compounds. While many of these pathways have yet to be scaled and developed as economically viable industrial production processes, it is clear that as synthetic biology technologies mature further, metabolic engineering is poised to replace traditional manufacturing for the production of many important substances.

1.1.2 Cellular and tissue engineering

Advances in synthetic biology have also been used to reprogram human cells. One promising avenue to combat many diseases is adoptive T-cell therapy, in which immune cells can be reprogrammed to seek out and destroy certain cells, such as cancerous ones (June, 2007). Bacteria, too, have been reprogrammed to specifically invade cancer cells (Anderson et al., 2006). Recent advances also allow for editing genomes to repair damaged genes, or confer novel functionality, by

engineered nucleases. These nucleases generally consist of a DNA-binding domain fused to a DNA-cleaving domain that can perform specific cleavage at a target DNA sequence, allowing for gene deletions, gene corrections, or addition of heterologous genes (Urnov et al., 2010). One recent effort demonstrated using transcription activator-like effector nucleases (TALENs), a type of engineered nucleases, to replace a dysfunctional copy of a gene with a functional one in cells derived from a patient (Osborn et al., 2013).

Another exciting avenue for correcting aberrant cell behavior is developing stem cells to replaced damaged tissues. Embryonic stem cells (ESCs) have been in clinical trials for several years for the treatment of age-related macular degeneration (AMD) (Advanced Cell Technology, Inc.), spinal cord injury (Geron, Inc.), and metastatic melanoma (California stem cell, Inc.), among others. More recently, researchers have been able to create induced pluripotent stem cells (iPSCs) from terminally differentiated cells by forced expression of a set of proteins that confer pluripotency on previously terminally differentiated cells, followed by inducing their differentiation into desired cell types, usually by expression of yet more protein factors (Ferreira and Mostajo-Radji, 2013). Earlier this year, the first clinical trial was approved for using iPSCs to replace damaged epithelium in patients with AMD (Institute for Biomedical Research and Innovation and RIKEN, Japan). There continues to be great excitement about the potential for ESCs and iPSCs for the treatment of many debilitating ailments (Kondo et al., 2012).

1.1.3 Future prospects for synthetic biology

Future applications of synthetic biology will no doubt continue to approach the sophistication and myriad functionality observed in natural systems. Or, perhaps, open the door to new possibilities not found in nature. Researchers have been incorporating non-natural amino acids into proteins for several years, granting properties not possible in nature (Johnson et al., 2011). There has also been much progress in creating an orthogonal genetic code, which would allow greater utilization of these non-natural amino acids to create engineered proteins with novel functionality (Anderson et al., 2004). Other work has generated organisms with *de novo* synthetized DNA (Gibson et al., 2010). Taken together, these developments suggest that the field is poised to coopt the diversity of biological possibility to address many pressing needs.

At present, however, there are several bottlenecks that stall progress. Although the cost of DNA sequencing has been dropping at a rate exceeding Moore's law, from almost $6,000 per megabase in 2000 to less than five cents now, gene synthesis is still relatively expensive, decreasing only 20-fold in the same period (Quail et al., 2012). Implementing new biological designs often requires significant time and resources, while evaluating design performance may be unfeasibly slow, for instance, when no method exists for detecting a desired metabolite. And of course, there are still significant shortcomings in our current understanding of the underlying biology, limiting our ability to design more sophisticated systems.

Many of the current examples in both metabolic engineering and human cell engineering consist of identifying useful genes—whether a working copy of a

mutated gene, genes that direct cellular specialization, or genes encoding enzymes that produce a valuable compound—and moving those genes into the cell of interest where they become constitutively expressed. However, many metabolic pathways involve intermediates that can react to form undesirable products, or enzymes that are expressed when they are either not needed or when they can act on an unintended substrate, limiting yield, and resulting in the production of undesired byproducts. Natural systems have evolved schemes to regulate their genes in complex ways. Although relatively new, synthetic dynamic gene control systems have already greatly increased yields of some engineered metabolic pathways in *E. coli* (Farmer and Liao, 2000; Zhang et al., 2012). These efforts, however, have generally consisted of coopting existing proteins that modulate their activity in response to changing levels of a specific input molecule, and are not generally easy to reengineer to sense and modulate their activity in response to different molecular inputs. Further, there are very few examples of synthetic gene-regulatory systems that can operate in eukaryotes, where many pathways involving enzymes that require glycosylation, membrane localization, and other eukaryotic processing steps are constructed (Nasser et al., 2003).

In cellular and tissue engineering, efforts are similarly limited by the relative dearth of synthetic gene-regulatory technologies that can operate in mammalian cells, in contrast with the complex gene networks that operate in natural systems to allow for proper development and cell function. These natural gene control systems involve regulation of many genes temporally, spatially, and in response to various signals, and can operate at many of the steps involved in gene

expression. A brief survey of the cellular processes involved in the expression of a mature protein is warranted to set in context the players and timescales involved in the life of a transcript, and how these would allow for the development of effective synthetic gene control systems.

1.2 Overview of gene expression in eukaryotes

The process by which genes gain functionality by becoming mature proteins involves myriad steps each of which may provide opportunities for synthetic regulation. Efforts in our laboratory and others have often focused on the yeast *S. cerevisiae*, a model organism whose genetics are thoroughly understood and amenable to manipulation, that recapitulates many of the features of higher eukaryotes, and that is generally nonhazardous to human health, labeled as "Generally Regarded as Safe" (GRAS) by the Food and Drug Administration. These considerations have led to the widespread use of yeast as a microbial production host. For simplicity, I will focus my discussion on this organism, while keeping in mind that many of the processes described below are similar in higher eukaryotes such as humans.

1.2.1 General characteristics of S. cerevisiae

Baker's yeast is a unicellular eukaryotic organism that can exist in haploid or diploid form. Haploids can be of mating type a or α. Two haploids of different mating types can undergo meiosis to generate a diploid cell. Both diploid and haploid organisms can reproduce by mitosis, called budding. Under some conditions, diploids can sporulate and generate four haploid cells. Cells are spherical or ovoid, with an average diameter of 5-10 μm, and an average nuclear

diameter of 1.9 μm (Jorgensen et al., 2007). As with other eukaryotes, yeast cells cycle among G1, S, G2, and M phases. In rich media, doubling time is 1.5 hr, which equates to a growth rate of 0.46/hr.

1.2.2 Genetics of S. cerevisiae

Haploid *S. cerevisiae* contains 16 chromosomes ranging in size from 230 kb to 1,532 kb with some 13 million base pairs encoding more than 6,000 open reading frames (Broach et al., 1991). Its mitochondria contain approximately 78.5 kb of DNA that encode some 15% of the enzymes active in that organelle *(Ibid.)*. Additionally, the circular 2μ plasmid containing 6.3 kb of DNA is present at approximately 20 to 50 copies in most organisms *(Ibid.)*. Some *S. cerevisiae* cells can be infected by any of several dsDNA viruses *(Ibid.)*. Many research laboratories, including ours, primarily derive their yeast strains from the widely studied haploid W303α strain (*MATα; leu2-3, 112; trp1-1; can1-100; ura3-1; ade2-1; his3-11,15*), which lends itself to genetic manipulation through various means. Exogenous genes can be integrated into a desired chromosomal locus by homologous recombination, allowing for different levels of expression depending on the integration site. Genes can also be added in 2μ-like plasmids, which are present at approximately 20 to 50 copies per cell, or in centromeric plasmids, present at 1 copy per cell (Tschumper and Carbon, 1983). Generally, expression level and variability within a population is higher from plasmids than for chromosomal integrations, and particularly so for high-copy plasmids (Romanos et al., 1992). The selection marker used can have a significant effect on expression levels both for chromosomally integrated and plasmid-encoded genes.

1.2.3 Transcription of mRNA

In eukaryotes such as *S. cerevisiae*, messenger RNAs (mRNAs) are transcribed from DNA by RNA Polymerase II (PolII), a holoenzyme composed of 10-12 subunits, which along with accessory proteins binds promoters and proceeds to transcribe the gene until transcription termination. The first step in the process, transcription initiation, is thought to be stochastic (Frieda and Block, 2012) and dependent on the abundance of the specific transcription factors that bind their cognate promoter, with an average occupancy of ~0.078 PolII complexes per kb of gene (Pelechano et al., 2010). Once properly docked and bound by transcription factors, PolII elongation appears to be deterministic and to proceed at ~20-25 nt/sec (Edwards et al., 1991; Pelechano et al., 2010), although this rate can vary throughout the cell cycle. After elongation, termination appears to take 70±41 sec (Frieda and Block, 2012). mRNA steady-state levels average 15,000 mRNA transcripts per cell, or approximately 2.5 mRNAs per each of the 6,000 genes in *S. cerevisiae* (Jelinsky and Samson, 1999), although this number can vary among different genes. Most genes produce 2-30 mRNAs per hour (Pelechano et al., 2010).

1.2.4 mRNA folding and activity

An mRNA can adopt different three-dimensional structures which in many cases can be important in determining how that mRNA acts, binds other factors, and is processed to yield functional proteins (Dethoff et al., 2012a). The problem of predicting which structures a given RNA sequence being transcribed will adopt during its lifetime depends on factors such as transcriptional speed,

thermodynamic and kinetic determinants, the binding of other entities such as ligands or protein chaperones, energy barriers to structural rearrangements, and the timescales and binding factors involved in transcription, RNA processing, translation, and mRNA decay. This problem may be very difficult, if not impossible, to solve with current tools. A simpler, though still not trivial, problem is to predict equilibrium secondary structures for a given sequence, for which methods with relatively high accuracy have been developed (Zuker, 2003; Reuter and Mathews, 2010). However, it remains important to keep in mind how this approximation differs from how RNA folding occurs *in vivo*.

There has long been evidence that certain RNA structures can fold during transcription (Boyle et al., 1980; Kramer and Mills, 1981). Consider, for instance, that RNA domains can form on the timescale of seconds (Bothe et al., 2011; Dethoff et al., 2012a, 2012b). With transcriptional elongation occurring at some 20-25 nt/sec, termination lasting on the order of many tens of seconds, and other processing steps lasting up to several minutes (Oeffinger and Zenklusen, 2012), it is clear that equilibrium folding does not paint a complete picture of *in vivo* folding. Although predictive algorithms for co-transcriptional folding have been developed (Proctor and Meyer, 2013), it remains difficult to evaluate their accuracy. Many of these tools rely on simplifying assumptions such as constant transcriptional speed and absence of RNA-binding factors, which are inaccurate for many important applications. While more research is needed to flush out these details, RNA structure prediction remains much more accurate than protein structure prediction, and continues to be important in determining how RNAs fold and affect gene expression.

1.2.5 mRNA processing and export

Before a transcript can be translated, many processing steps occur during and after transcription, including 5′ capping, 3′ cleavage, polyadenylation, and nuclear export. Splicing, although widespread in higher eukaryotes, is not prevalent in yeast, where only some 5% of the total 6,000 genes contain introns (Lewin, 2008), and thus will be ignored in this discussion.

Many of these processing steps are mediated by modifications to the C-terminal domain of the large subunit of Pol II, the C-terminal domain (CTD), which aids in recruiting other factors yielding a dynamic messenger ribonucleoprotein complex (mRNP) undergoing constant modification and quality control as its being transcribed (Egloff and Murphy, 2008; Buratowski, 2009). During transcription, the cap-binding complex (CBC) attaches to the 5′ end of the mRNP and remains attached until export to the cytoplasm. The 3′ end of the mRNP undergoes 3′ splicing, polyadenylation, and binding by poly(A)-binding proteins such as Pab1p and Nab2p, which also remain attached until export to the cytoplasm.

Nuclear export of mRNP occurs through the nuclear pore complex (NPC). First, the mRNP must diffuse from its location in the nucleus to one of the some 119 NPCs on the nuclear membrane (Maul and Deaven, 1977). Estimates of the diffusion coefficients for mRNPs vary depending on the transcript and research methodology, but many studies list values close to 1 $\mu m^2/sec$ (Grünwald et al., 2011). An average yeast nucleus has a volume of ~2.9 μm^3 (Jorgensen et al., 2007), yielding transit times to the nuclear membrane on the order of seconds. Once at the NPC, transport of most mRNPs is mediated by the Mex67-Mtr2 complex

(Segref et al., 1997). A recent study of mRNA transport in live cells showed that the entire process of binding to the NPC, transport, and release takes approximately 180 msec (Grünwald and Singer, 2010), although again these times may be different for different transcripts.

1.2.6 Translation

Once in the cytoplasm, transcripts can begin to generate proteins. Translation initiation begins by removal of the CBC from the 5′ cap and binding of eukaryotic initiation factors (eIFs) 4G and 4E and associated proteins. eIF4 interacts with Pabp1, generating a circular "activated" mRNP (Hinnebusch and Lorsch, 2012). The activated mRNP is then bound by the 40S ribosomal subunit and associated factors to form a pre-initiation complex (PIP), which scans for a start codon. Initiation efficiency is partly a function of mRNA secondary structure near the start codon (Kozak, 2005), and is thought to be rate-limiting (Preiss and Hentze, 2003). In yeast, most genes contain a relatively unstructured consensus sequence "aAaAaA**ATG**TCt" (with the start codon shown in bold and less-conserved bases in lower case) which increases translation initiation efficiency (Hamilton et al., 1987).

Once the start codon is found, the 60S ribosomal subunit and associated factors bind and translation elongation can begin. Multiple ribosomes can be active on a given transcript, with an average occupancy of 0.64 ribosomes per 100 nt of coding RNA (Arava et al., 2003). On average, translation elongation proceeds at 7-8 amino acids per second per ribosome, although this speed depends on the availability of transfer RNAs (tRNAs) cognate to the codons in the

transcript, with rare codons leading to inefficient translation (Ikemura, 1982). Once the ribosome proceeds to the stop codon, elongation release factors (eRFs) 1 and 3 and associated proteins mediate the release of the nascent peptide from the ribosome (Mugnier and Tuite, 1999).

Protein abundance in yeast ranges from approximately 50-10^6 proteins per cell, with an average of about 2,000 (Ghaemmaghami et al., 2003). Synthesis rates generally fall in the range of 1.8-5.4/hr (Von der Haar, 2008), and protein half-lives average 43 min (Belle et al., 2006).

1.2.7 mRNA decay

For mature transcripts in the cytoplasm, decay can occur either 5′ →3′ or 3′ →5′. For both cases, the process begins with deadenylation until a transcript contains fewer than ten 3′ terminal adenosines. In yeast, the primary decay mechanism involves 5′ decapping by the enzymes Dcp1 and Dcp2 followed by 5′ →3′ degradation by the exoribonuclease Xrn1 (Parker and Song, 2004). Alternatively, following deadenylation, a transcript can be degraded 3′ →5′ by the cytoplasmic exosome. Other degradation pathways clear mRNAs with premature termination codons, ones that lack a stop codon, or ones that stall due to strong RNA secondary structure, termed nonsense mediated decay (NMD), non-stop decay (NSD), and no-go decay (NGD), respectively (Shoemaker and Green, 2012). Degradation can also occur in the nucleus, where mRNAs lacking a 5′ cap are readily degraded 5′ →3′ by the exoribonuclease Rat1, and to some extent by Xrn1, while those lacking polyadenylation can be readily degraded 3′ →5′ by the nuclear exosome (Houseley and Tollervey, 2009). These processes mediate mRNA

half-lives on the order of tens of minutes, with a mean of 11 min (Miller et al., 2011).

1.2.8 Overview of kinetics of gene expression

All in all, a picture emerges of the timescales, concentrations, and organization of the main biological molecules involved in gene expression. Many of these processes are evolutionarily conserved among eukaryotes, if with some important differences. Mammalian cells often require more extensive mRNA processing, including widespread splicing, with 95-100% of genes containing introns and being alternatively spliced (Conze et al., 2010). Additionally, mammalian cells are much larger than yeast, usually with cellular and nuclear volumes 2-10 times larger than yeast's (Rosenbluth et al., 2006). These differences account for longer diffusion times, and often slower processing steps. Mammalian cell doubling times, for example, are typically greater than approximately 16hrs (Kumei et al., 1987), 11 times longer than yeast's. Below, Figure 1.1 summarizes the entire process of gene expression in yeast.

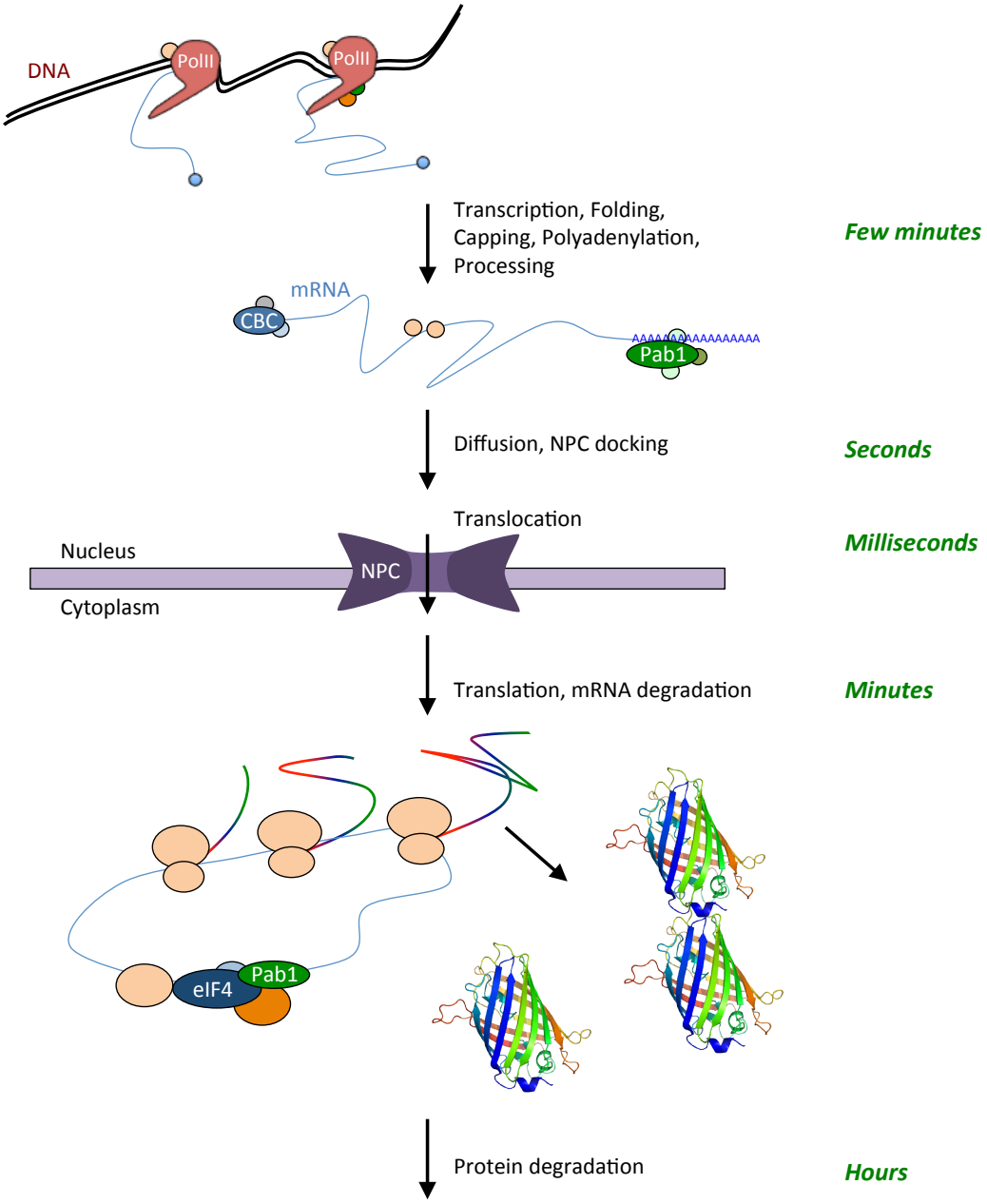


Figure 1.1 Overview of gene expression in yeast

Gene expression in yeast involves several steps with characteristic time scales and component localization-specific processing. The diagram shows DNA being transcribed by PolII and accessory factors, generating a dynamic mRNA-protein complex that translocates to the cytoplasm where further processing and translation yields functional proteins. Average timescales for relevant processes in yeast are shown in green. Molecules are not drawn to scale.

1.3 Synthetic gene-regulatory systems in eukaryotes

For many potential applications of synthetic biology, it would be desirable to tailor gene expression to cellular characteristics. For example, a useful biological device might be instructed with reprogramming cells that display aberrant behavior while not disrupting other cells in a way that would cause pathology. Another useful device might be one that can induce expression of a particular metabolic enzyme only when it is needed, allowing for less waste, and better yields.

Generally, what we would want in a gene control platform is to be able to regulate the expression of a desired output gene, in response to relevant changes of an input signal, to such a degree that it would cause a physiologically useful phenotypic change. For example, if a particular cell contains high levels of a cancer biomarker, an ideal device might increase expression of a protein that reprograms the cell a non-malignant state. Or, if metabolite "K" has reached high levels, begin expressing the perhaps growth-limiting enzyme "Kase," and letting the cell to grow unburdened when "Kase" is not needed. Additionally, depending on the intended application, some important characteristics of such a system are listed in Figure 1.2.

Several kinds of synthetic eukaryotic gene control systems have been developed that operate at different steps of gene expression. Below, I will give a brief description of the more commonly used gene regulation platforms that constitute the current state of the art.

Characteristic	Definition
Modularity	The system could be easily redesigned to respond to different inputs and act on different outputs, often by recombining component "modules" that confer input or output function.
Scalability	The output level of the system could be scaled to cover the range over which a change in output corresponds to a change in phenotype. Similarly, the sensitivity to changes in input would be scaled to cover the range over which the relevant change in input occurs.
Uniformity	In a population of cells, the system would behave similarly enough among all the cells such that perturbations due to genetic noise or other unintended factors would not significantly affect phenotypes.
Orthogonality	The system would not significantly affect the performance of other gene networks, including both other synthetic networks, and the cell's existing networks.
Simplicity	The components of the system would not be too taxing, either by consuming too many cellular resources, or by necessitating too much effort to generate or incorporate genetically into the desired cells.

Figure 1.2 Important ideal characteristics of an ideal synthetic gene-regulatory system

1.3.1. Control over gene insertion

One obvious method of synthetic gene expression control is to introduce desired genes—often constitutively expressed—only into the cells one wishes to reprogram. Such an approach has been employed clinically, but only on certain accessible and somewhat isolated tissues such as the eye and skin (Kondo et al., 2012). For reprogramming cancerous cells distributed deep and throughout the body, such a strategy is not viable. Similarly, there are no methods to constrain presence of a gene only to the subset of cells growing in a fermenter that have at a given time point accumulated sufficient levels of a given intermediate metabolite.

1.3.2 Transcriptional control by protein factors

A commonly used strategy for synthetic gene expression control has been to control transcription. Generally, such systems involve a synthetic transcription factor (TF) that contains a DNA-binding domain, an activation (or repression) domain, and a ligand-binding domain that alters activity of one of the other domains. In one early example, the *E. coli* tetracycline repressor (tetR) was fused to the activation domain of the herpes simplex virus protein 16 (VP16) to generate a tetracycline trans activator (tTA) (Gossen and Bujard, 1992). Binding of the tetR domain to tet operator sequences placed adjacent to a minimal promoter drove expression of a downstream target gene in response to expression of tTA. Addition of tetracycline, which modulates the conformation and activity of tTA, lowered gene expression by up to five orders of magnitude. Variants of tTA have been engineered with various gene expression responses, including turning gene expression "on" in response to tetracycline. Similar systems that respond to several other ligands have also been developed or adopted from natural systems. One group has developed modular binding domains composed of zinc-finger proteins that allow for orthogonal regulation of different target genes by different TF variants with various regulatory dynamics, while still retaining a large dynamic range on the order of approximately 100 fold (Khalil et al., 2012).

Although TF-based systems generally display high-fold changes in expression in response to their ligand, they have several disadvantages. First, they generally require the expression of a heterologous transcription factor protein, which may cause immune responses in humans, limiting the application of these systems in clinical settings. Even in yeast, expression of such proteins can often be associated

with toxicity or growth defects. Second, although there are examples of reengineering these synthetic TFs to respond to novel ligands (Collins et al., 2005), these efforts have so far only been successful in evolving sensitivity to close homologues of the original ligands, usually while still maintaining some sensitivity to the original ligand. It remains difficult to engineer TFs that respond specifically to wholly novel entities. And third, often these systems are difficult to tune to different performance parameters, such as different output levels.

1.3.3 Transcriptional control by RNA switches

Transcriptional control can also be achieved by non-protein factors. One group evolved an RNA sequence capable of strongly activating transcription from a specific promoter, and later attached an aptamer to allow allosteric control (Buskirk et al., 2004). Although this work demonstrated up to 10-fold activation upon ligand addition, the riboswitch is specific to the promoter to which it was evolved, and it may be nontrivial to develop riboswitches that bind other promoters. Further, the mechanism through which this design works has not been fully elucidated. And lastly, the design calls for expression of heterologous proteins, which may limit application in human cells due to immunogenicity concerns.

1.3.4 Post-transcriptional control by RNAse III switches

Many groups have developed strategies to control gene expression at processing steps that occur after transcription initiation. For example, a number of regulatory elements have been developed that act through modulating cleavage of a synthetic substrate by different endogenous RNase III enzymes. In yeast, one

strategy involved integrating small-molecule aptamers into synthetic substrates to the *S. cerevisiae* RNase III, Rnt1p (Babiskin and Smolke, 2011b). The study showed up to a 6-fold change in the expression of a reporter gene upon ligand addition. The design is modular in that it can regulate any transgene, and can potentially be generalized to other ligands to which a suitable aptamer exists or can be generated. Of course, the platform is not portable to other eukaryotes, including humans, which do not normally express Rnt1p.

However, similar strategies can be employed in mammalian cells, where the activity of other RNase III enzymes on synthetic substrates can be used for synthetic gene control. One group incorporated an aptamer to a small molecule into a synthetic short hairpin RNA (shRNA) to modulate cleavage by Dicer in response to ligand addition. Dicer cleavage directed subsequent silencing of a target gene, allowing for ~4-fold modulation of target gene expression (An et al., 2006). A similar example involved integrating a protein aptamer into a synthetic shRNA to control expression of anti-apoptotic genes, achieving ~10-fold modulation of cell death (Saito et al., 2011). Independently, our laboratory has demonstrated a similar strategy by integrating small-molecule aptamers into synthetic microRNAs (miRNAs) to modulate Drosha processing and subsequent silencing of a target gene (Beisel et al., 2011). This study was able to achieve ~5-fold modulation of gene expression in response to the presence of the small-molecule ligand. Another group has coupled a small-molecule aptamer to an allosteric ribozyme whose cleavage modulates the release of a primary miRNA (pri-miRNA). Release of the pri-miRNA allows for Drosha processing and gene silencing of a target gene. This study achieved ~3-fold changes in expression of a

reporter gene (Kumar et al., 2009). These designs do not depend on the expression of heterologous proteins, making them attractive candidates to control gene expression in human cells for the treatment of disease. However, so far they exhibit relatively modest dynamic ranges, and there is concern that these small RNA silencers may affect expression of partly homologous off target genes (Birmingham et al., 2006).

The examples just described can detect proteins and small molecules and modulate activity of small RNA silencers. However, levels of endogenous small RNA silencers can also serve as markers of cell state. One group designed a strategy to detect endogenous miRNA expression patterns to regulate expression of pro-apoptotic proteins, achieving ~10-fold modulation of cell death between cell types that express different miRNAs (Xie et al., 2011). However, these designs rely on the expression of heterologous proteins, which may cause immunogenic concerns.

1.3.5 Ribozyme switches

Another strategy developed in our laboratory involves integrating small-molecule aptamers into a self-cleaving hammerhead ribozyme placed in the 3' untranslated region (UTR) of a transgene in yeast (Win and Smolke, 2007), allowing for allosteric regulation of gene expression. The platform has been further developed for complex logical computation (Win and Smolke, 2008a) and high-throughput *in vivo* tuning and *in vitro* characterization (Kennedy et al., 2012; Liang et al., 2012). Further, because the platform does not make use of any proteins that may only be endogenous to specific organisms, it is applicable for

use in human cells, and has been employed to control T-cell proliferation in cultured mammalian cells and live mice (Chen et al., 2010). One drawback is that so far, dynamic ranges have been modest, with a maximum of ~10 fold (Liang et al., 2012). However, the design has been shown to be modular in responding to different small molecule inputs.

1.3.6 Translational control

RNA regulation has also been applied to control translation. One early example involved small-molecule aptamers placed in the 5′ UTR of a gene in mammalian cells that allowed for modulation of gene expression in response to small-molecule binding (Werstuck and Green, 1998). Another group demonstrated a similar approach with protein aptamers, and applied the design to regulate pro-apoptotic genes, achieving ~10-fold modulation of cell death in response to the presence of the protein ligand (Saito et al., 2011). These designs are promising in that they involve no heterologous proteins. However, as with other RNA-based systems described here, the ~10-fold dynamic range is modest. Further, because the system depends on blocking ribosome progress along a transcript, its behavior depends on the characteristics of the aptamer–ligand interaction. With some aptamers, even unbound by ligand, strong secondary structure can significantly lower gene expression, casting doubt on how well such a design can be employed to respond to different inputs.

1.3.7 Evaluation of current platforms

The synthetic gene-regulatory strategies described above can be evaluated by how they approach the desired characteristics described in Figure 1.2, namely,

modularity, scalability, uniformity, orthogonality, and simplicity. Of course, the real evaluation criterion is how well a given synthetic regulatory scheme could be employed for an application of interest, which for different applications might depend on one set of characteristics more than another.

In general, RNA-based platforms have greater modularity. It is generally significantly easier to find an RNA sequence that can bind and switch in response to a ligand of interest than to do the same for a protein. While researchers have long been generating RNA aptamers to ligands of interest *de novo* (Ellington and Szostak, 1990; Tuerk and Gold, 1990), often with high affinity and specificity, such is not the case for generating *de novo* protein domains that bind ligands of interest. This shortcoming is perhaps a technological one. Proteins have 21 naturally occurring amino-acid building blocks whereas RNAs only have four naturally occurring nitrogenous base building blocks. This difference allows for far wider potential functional diversity of proteins. In addition, there are over 500 known amino acids (Wagner and Musso, 1983), which as technologies to incorporate unnatural amino acids develop further could significantly expand the possibilities for protein function. There are, by contrast, only about a dozen known non-natural nitrogenous bases that can be incorporated into RNA, and there are no existing methods to incorporate these into living systems (Yang et al., 2010).

Still, at present, protein engineering as a field is far from being able to design proteins that can regulate gene expression in a modular way. Even if such were the case, for applications in mammalian cells that can address the many significant problems we face in the area of human health, there is still the problem of immunogenicity. Heterologous proteins often cause significant immune

responses. While in the distant future it may be possible to somehow elude this problem, whether by modulating human immune system development or activity, or by developing a method to design proteins that do not elicit such a response, today this remains an important roadblock.

RNA-based gene control systems have their own shortcomings. Broadly generalizing, RNA systems tend to be more modular, owning to the relative ease of developing RNA aptamers to ligands of interest, and to our currently greater ability to predict RNA structure, at least relative to predicting protein structure. Further, most RNAs do not cause an immune response in mammalian cells, making RNA a compelling design substrate for gene control systems designed for applications in human health. However, they tend to exhibit lower dynamic ranges. While different researchers describe dynamic range in different ways—with some using activities of enzymes to describe gene expression levels and some using fluorescent reporter proteins—most RNA-based gene control strategies self-describe at most ~10-fold modulation of gene expression. Is killing 90% of a cancer enough?

While certainly for some applications 10-fold may be sufficient, there is a clear need to develop viable gene control strategies with greater dynamic ranges. It is unclear why most of these RNA-based systems have lower dynamic ranges than most widely used TF-based systems. However, one possibility may have to do with the number of species that are being regulated. As shown in Figure 1.3, there is a large difference in average number of molecules per cell among DNA, RNA, and protein species. This “amplification” occurs in going from mRNAs to proteins, where abundance jumps 1,000-fold. It is reasonable to assume that

control at the "reverse-bottleneck" of transcription may generally more dramatically affect the change in protein expression than control after this "reverse-bottleneck."

	Mean copy number	Mean half-life	Mean synthesis rate	References
DNA	1	--	~0.46/hr	
mRNA	~2.5	~11 min	~2-30/hr	(Jelinsky and Samson, 1999) (Pelechano et al., 2010)
Protein	~2,000	~43 min	~1.8-5.4/hr	(Belle et al., 2006) (Ghaemmaghami et al., 2003) (Von der Haar, 2008)

Figure 1.3 Parameters of various types of biological molecules in yeast

Biological molecule parameters can vary among the more important classes found in cells. DNA, mRNA, and protein average levels and half-lives can differ by orders of magnitude. Values given represent average parameters for a single gene and its mRNA and protein products in a typical haploid yeast cell growing under normal conditions.

Another possibility may involve energy barriers to structural rearrangements. While both proteins and mRNAs exist under a certain structural dynamism owning to thermal noise and kinetic or electrostatic interactions with other molecules, it is important to note that thermodynamic barriers to more dramatic structural changes differ between proteins and RNAs. For one, proteins can contain covalent bonds between distal amino acids, while RNAs only form hydrogen bonds between distal bases. Additionally, certain protein folds such as the alpha helix or beta sheet are extremely stable, while typical RNA folds often consist of relatively short runs of base pairing. In RNA structural rearrangements, step-wise one base-pair rearrangements can result in low energy barriers. A recent study analyzed several riboswitches and proposed folding paths between

their functional conformations, finding most step-wise energy barriers to be relatively small, generally ~1-2 kcal/mol (Quarta et al., 2012). With lower barriers to structural rearrangement, it is reasonable to assume that there may be lower barriers for a given regulatory RNA to switch between an active and an inactive conformation, even when one of those conformations is stabilized by binding to a ligand. Thus, it may be that RNA regulators spend more time sampling both active and inactive conformations, leading to lower differences between their active and inactive states, and thus lower-fold regulation of gene expression. Some of the riboswitches developed in our laboratory, for instance, are designed to switch between active and inactive conformations that differ by energies ~1.2 kcal/mol, only twice the thermal energy available at 30°C, which in an equilibrium Boltzmann distribution would translate to only a ~7-fold difference in the frequency at which these states are populated.

Still another possible factor to explain the modest observed dynamic range for synthetic RNA-based gene control versus the high dynamic range for synthetic protein-based gene control may have to do with cooperativity. It is known that many transcription factors cooperate to generate a more digital response to changing signal levels (Banerjee, 2003; Chang et al., 2006), often resulting in greater sensitivity to signal induction and greater changes in output activity (Bialek and Setayeshgar, 2008). Many other examples of proteins whose activity is cooperative, in that it changes in a more digital manner in response to increasing levels of an input, have been widely studied, such as the classic example of cooperativity of hemoglobin-oxygen binding. While cooperativity has also been observed in RNAs (Vaidya et al., 2012), and it may be possible to design

regulatory RNAs that are also cooperative, the phenomenon remains primarily a characteristic of proteins, and particularly of those involved in the protein-based gene control strategies described above.

Today, it remains of general interest in the field of synthetic biology, and particularly in applications of synthetic biology in eukaryotes such as yeast and mammalian systems, to develop more modular and better-performing gene control strategies. Developing of these technologies would expand our prospects to reengineer life to solve some of society's most onerous challenges.

1.4 Scope of thesis

This thesis describes expanding the capabilities of synthetic gene control in eukaryotes. We take as a starting point a synthetic ribozyme switch platform previously developed in our laboratory (Win and Smolke, 2007, 2008b) which displays many promising attributes. Among these are our ability to employ the platform for complex computation schemes (Win and Smolke, 2008b), established methods to rapidly tune and characterize device properties (Kennedy et al., 2012; Liang et al., 2012), and known design principles stemming from our ability to predict RNA structure-function relationships (Win and Smolke, 2007). The platform has also been demonstrated in mammalian cells (Chen et al., 2010; Wei et al., 2013), where likely absence of immunogenic response stemming from its expression promise powerful applications in gene therapy. In metabolic engineering, too, the platform has proven effective in the development of novel enzymes by correlating enzyme activity to an easily detected signal (Michener and Smolke, 2012). Lastly, the platform is designed to be modular, in that it can

regulate any transgene, or any endogenous gene by 3' UTR replacement (Babiskin and Smolke, 2011b). Moreover, established methods to generate RNA aptamers to molecules of interest suggest the device can be employed to take as input any of a wide array of molecules, in stark comparison to TF-based control platforms, which are not generally generalizable to different inputs.

Thus far, however, the ribozyme switch platform has not been shown to respond to protein inputs. In Chapter 2, we demonstrate the first known instance of ribozyme-mediated, protein input-responsive gene regulation in eukaryotic cells. We describe efforts to integrate protein aptamers directly into a self-cleaving ribozyme at several integration points. Although this strategy does not seem to elicit protein input-mediated gene regulation, the associated studies demonstrate how different modifications of the ribozyme structure affect cleavage activity, informing further development of ribozyme switches. We find success in using strand displacement as a strategy to achieve protein input-responsive ribozyme switches, achieving up to 4-fold modulation of target gene expression in response to the presence of a protein input. *In vitro* characterization shows evidence that the observed gene expression modulation stems from protein input-mediated modulation of ribozyme cleavage activity. We further show that the platform is translatable to mammalian cells. Lastly, we conclude with thoughts about the general applicability of the platform and mechanistic details of its gene-regulatory activity.

Having further demonstrated the modularity of our ribozyme switch platform for controlling gene expression, we next turn to its range of output gene modulation. So far, eukaryotic riboswitch platforms have generally had modest

dynamic ranges of output gene expression modulation, generally on the order of 10 fold, and often in response to induction by high concentrations of input molecules (Liang et al., 2011). TF-based systems, while largely exhibiting modular and high-fold regulation of output gene expression (Khalil et al., 2012), exhibit limitations with respect to their adaptability to respond to different input molecules, limiting their widespread use for synthetic biology. In Chapter 3, we demonstrate a TF-riboswitch gene amplifier in eukaryotes. This platform combines the modular-input characteristics of ribozyme switch-mediated regulation with the high-fold, modular output characteristics of TF-mediated gene expression control. Our amplifier platform allows for as much as 20-fold regulation of output gene expression in response to input signal. EC_{50} values are more than 4 times lower than in our best-performing non-amplifier ribozyme switches, and maximal "ON" state gene expression approaches the highest expression levels commonly observed in yeast.

In Chapter 4, we conclude with future directions for our ribozyme switch and amplifier platforms. We discuss further device development and characterization, as well as applications in gene therapy and metabolic engineering. Collectively, the work included in this thesis extends the capabilities of ribozyme switches, a powerful tool for synthetic gene control, promising more sophisticated customizable applications of synthetic biology.

Chapter 2

Development of protein-responsive ribozyme switches in eukaryotic cells

Abstract

As synthetic biology continues to develop promising solutions to some of society's most pressing needs, the dearth of technologies that can correlate cell state to programmed gene expression in eukaryotic organisms hinders the development of more sophisticated applications. Although some platforms have been developed to detect and program gene expression in response to changing protein levels, current systems either can only operate in certain organisms, risk off-target effects, or are not generally modular or scalable. Here, we demonstrate the use of synthetic self-cleaving riboswitches to detect changes in input protein levels and convey that information into programmed gene expression in eukaryotic cells. We demonstrate both up- and down-regulation of levels of an output transgene by more than 4 fold in response to rising input protein levels, with maximal output gene expression approaching the highest levels observed in yeast. *In vitro* experiments demonstrate protein-dependent riboswitch activity modulation. We further demonstrate the platform in mammalian cells. Our riboswitch devices do not depend on special input protein activity, and can be tailored to respond to any input protein to which a suitable RNA aptamer can be developed. This platform can potentially be employed to regulate expression of any transgene or any endogenous gene by 3' UTR replacement, allowing for more complex cell state-specific reprogramming with applications in gene therapy and cellular engineering.

Partly adopted from d'Espaux, L.D., Kennedy, A.B., Vowles, J.V., Bloom, R.J., and Smolke, C.D. (2013). Development of protein-responsive ribozyme switches in eukaryotic cells. *In preparation*

2.1 Introduction

Synthetic biology continues to develop into a promising avenue for addressing some of our most pressing challenges, including in areas of disease, energy, and the sustainable production of chemicals. For many potential applications, cue-specific control over expression of certain genes would allow greater sophistication and functionality. Constraining gene expression to desired cell states would further enable treatments in which expression of therapeutic genes must be targeted to diseased cells, avoiding undesired side effects (Zhou et al., 2004). It would also further develop our ability to use engineered cells to repair desired tissues. For example, recent efforts have employed embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) to generate new tissues, often by forced expression of transcription factors (TFs) that confer pluripotency, or allow for differentiation into desired cell types (Ferreira and Mostajo-Radji, 2013). However, there remain concerns about unregulated stem cell TF expression potentially leading to cancer (Kondo et al., 2012). For many of these examples, being able to regulate gene expression precisely in response to cues regarding the cellular state would allow for safer and more effective therapeutic strategies.

Cells can differ dramatically in their properties. A one base-pair mutation leading to a small difference in the expression of one protein may differentiate a normal cell from one that would develop into a lethal cancer (Hollstein et al., 1991). Differences among different cell types, or cell states, are often determined

by the levels of a handful of molecules. For example, many cancers can be identified by altered levels of known protein markers (Weigel and Dowsett, 2010; Newton et al., 2012). iPSCs are often created by altering expression of just four transcription factors (Yu et al., 2007). Proteins, generally, are often important identifiers of a cell's state. There are some 50 million total proteins in a typical eukaryotic cell (Futcher et al., 1999), with an average concentration of ~1μM (Ghaemmaghami et al., 2003; Lu et al., 2007), with specific proteins varying by sometimes many orders of magnitude among different cellular states. Being able to detect these differences and convey that information into desirable phenotypic changes in eukaryotic cells is a pressing need in gene therapy and other areas of cellular engineering.

RNA has recently emerged as a powerful platform for the development of synthetic gene control strategies owning partly to its versatility and modularity. Standard protocols allow for the development of RNA sequences that can recognize and tightly bind other molecules, including small molecules and proteins (Ellington and Szostak, 1990; Tuerk and Gold, 1990). RNA structure prediction is often very accurate (Zuker, 2003; Dimitrov and Zuker, 2004; Reuter and Mathews, 2010; Proctor and Meyer, 2013), facilitating the design of molecular switches where a structural change affects function. Additionally, many functional noncoding RNA classes have been identified, including self-cleaving ribozymes present in all domains of life (Perreault et al., 2011), special folds that undergo specific cleavage by RNases, including microRNAs, short hairpin RNAs, and other substrates (Lamontagne and Abou Elela, 2007), and metabolite-responsive riboswitches (Winkler et al., 2004), among many others. Lastly,

synthetic RNAs expressed in living cells do not normally elicit immune responses, while synthetic transcription factors usually do. Inspired by these characteristics, researchers have developed strategies to allosterically regulate gene expression using synthetic functional RNA, although generally these platforms only operate in prokaryotes, and few respond to protein inputs (Liang et al., 2011).

Protein-responsive regulation has been achieved in mammalian cells by placing aptamers within introns to modulate alternative splicing (Culler et al., 2010), although this strategy has not been demonstrated in other eukaryotes, where alternative splicing pathways can be absent or different from mammals. Further, gene expression modulation has been relatively modest, under 4 fold. Another strategy involved placing protein aptamers in the 5′ untranslated region (UTR) of a reporter gene (Saito et al., 2011), although this strategy can lead to low basal gene expression due to strong 5′ UTR secondary structure, and may not be generalizable to other protein inputs. The same group has also demonstrated placing protein aptamers within synthetic short hairpin RNA (shRNA) substrates which can modulate how these shRNAs are processed and lead to gene silencing (Saito et al., 2011; Kashida et al., 2012). However, the strategy is not generalizable to eukaryotes that do not employ shRNA-mediated gene silencing, and reports of shRNA-mediated off-target effects (Birmingham et al., 2006) cast doubt on widespread use of these technologies, even in mammals.

Our laboratory has previously developed a ribozyme switch platform by integrating small-molecule aptamers into a self-cleaving hammerhead ribozyme placed in the 3′ UTR of a heterologous gene in yeast (Win and Smolke, 2007), allowing for allosteric regulation of gene expression in response to changing

levels of small molecules. These switches are composed of a small-molecule aptamer coupled to a hammerhead ribozyme through a transmitter sequence. The transmitter is designed to allow for two primary switch conformations that differ in whether they contain properly folded ribozyme or aptamer domains (Figure 2.1). In switch "ON" designs, aptamer and ribozyme folds are mutually exclusive within each primary conformation. Ligand–aptamer binding stabilizes the aptamer-active conformation, diminishing the fraction of riboswitches in the ribozyme-active conformation, lowering cleavage activity. In the "OFF" design, aptamer and ribozyme folds are mutually inclusive, so that aptamer–ligand binding stabilizes the ribozyme-active conformation, increasing cleavage activity. When these switches are incorporated in the 3′ UTR of a target gene, cleavage activity leads to transcript degradation and loss of expression. Ligand-induced modulation of cleavage activity allows for control over gene expression levels.

The ribozyme switch platform has been further developed for complex logical computation (Win and Smolke, 2008a) and high-throughput *in vivo* tuning and *in vitro* characterization (Kennedy et al., 2012; Liang et al., 2012). It has also been demonstrated in mammalian cells, where it allowed for control of engineered T-cell proliferation in cultured cells and in live mice (Chen et al., 2010). Platform behavior in mammalian cells can be predicted by activity in yeast, allowing for rapid device development and tuning (Wei et al., 2013). Because of its modular construction, the platform can be employed to regulate any transgene, or any endogenous gene by 3′ UTR replacement (Babiskin, 2011). However, so far its inputs have been limited to small-molecule inducers.

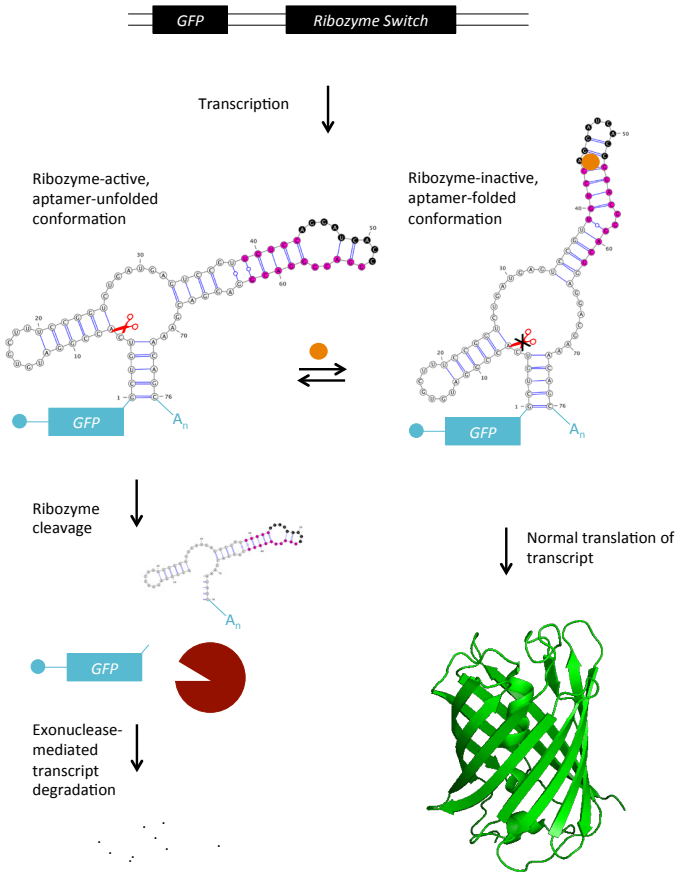


Figure 2.1 Design of *in vivo* hammerhead ribozyme switch platform

A hammerhead ribozyme is a naturally occurring RNA sequence that can undergo rapid, site-specific self-cleavage (site indicated by red scissors). A synthetic ribozyme switch couples this ribozyme to a ligand-binding aptamer sequence (black bases) through engineered transmitter sequences (pink bases). The switch is encoded in the 3' untranslated region of a target gene, in this case GFP, and upon transcription in eukaryotic cells is designed to exist in two primary conformations. In this example of an "ON" switch, conformations are mutually exclusive in whether they contain a ribozyme-active domain or an aptamer-folded domain. Ribozyme-active conformations undergo cleavage, leading to rapid exonuclease-mediated transcript degradation, and loss of GFP expression. Binding of a ligand (orange circle) stabilizes the aptamer-folded conformation, which in this case is diminishes cleavage activity, and allows for greater GFP expression. Thus, increased levels of ligand lead to target gene expression being turned "ON." In "OFF" switch designs (not shown), ribozyme-active and aptamer-folded domains are mutually inclusive. In the absence of ligand, transcripts primarily do not undergo cleavage, but do so when ligand binding stabilizes the aptamer-folded and ribozyme-active conformation.

This ribozyme switch platform has been further developed for complex logical computation (Win and Smolke, 2008a) and high-throughput *in vivo* tuning and *in vitro* characterization (Kennedy et al., 2012; Liang et al., 2012). It has also been demonstrated in mammalian cells, where it allowed for control of engineered T-cell proliferation in cultured cells and in live mice (Chen et al., 2010). Platform behavior in mammalian cells can be predicted by activity in yeast, allowing for rapid device development and tuning (Wei et al., 2013). Because of its modular construction, the platform can be employed to regulate any transgene, or any endogenous gene by 3′ UTR replacement (Babiskin, 2011). However, so far its inputs have been limited to small-molecule inducers.

Here we further develop the platform to be able to detect and modulate gene expression in response to levels of protein inputs. We demonstrate both up- and down-regulation of levels of an output transgene by more than 4 fold in response to rising input protein levels, with maximal output gene expression approaching levels from no-ribozyme controls. *In vitro* experiments demonstrate protein-dependent switch activity modulation. We further demonstrate the platform in mammalian cells. Our ribozyme switch devices do not depend on special input protein activity, and can be generalizable to any input protein to which a suitable RNA aptamer can be developed. This platform can potentially be employed to regulate expression of any transgene, or any endogenous gene by 3′ UTR replacement, allowing for more complex cell state-specific reprogramming with applications in cellular engineering and gene therapy.

2.2 Results and Discussion

2.2.1 Integrating MS2 aptamers into ribozymes by direct coupling

We constructed an initial set of protein-responsive ribozyme switches by integrating known protein aptamers directly into a hammerhead ribozyme. We initially chose a direct-integration strategy because proteins, being much larger than small molecules, might by their sheer size lead to significant disruption of the cleavage-active ribozyme conformation. It has long been known that in addition to conserved sequence and secondary structure elements, hammerhead ribozymes require tertiary interactions between peripheral loops for *in vivo* activity (De la Peña et al., 2003; Khvorova et al., 2003). Thus, we theorized that protein binding near these regions might be sufficient to modulate ribozyme cleavage without the need for a separate transmitter domain that would direct secondary structure changes in the ribozyme. By obviating the need to engineer transmitter sequences, the direct-integration strategy may allow for a more modular plug-and-play architecture that might be generalizable to other protein aptamers.

We designed two classes of switches, ones in which we integrated aptamers directly on loop I of the hammerhead ribozyme, which we call dc1 for "direct coupling into loop I" (Figure 2.2), and ones in which we integrated aptamers into stem III just below the catalytic core, which we call dc3 for "direct coupling into stem III" (Figure 2.3). We decided to integrate aptamers into loop I because with the relatively slow speed of transcription, at ~20 nt/sec (Edwards et al., 1991; Pelechano et al., 2010), loop I would be fully formed more than one full second

before the entire ribozyme is transcribed, allowing extra time for protein binding before the fully-transcribed riboswitch can begin cleaving. We theorized that protein binding during switch transcription might funnel subsequent folding rearrangements to ligand-bound conformations that would not readily form in the absence of protein, serving as a kinetic switch (Quarta et al., 2012). We also decided to try integrating inside stem III itself because this integration point would be closest to the catalytic core, potentially having a greater chance of disrupting cleavage activity after protein binding. For both classes of switches, dc1 and dc3, we integrated aptamers to the bacteriophage MS2 coat protein (MS2) previously shown to bind their cognate protein with high affinity (Parrott et al., 2000; Horn et al., 2004).

We first examined whether these riboswitches can bind MS2 protein via a Biacore surface plasmon resonance (SPR) assay, in which we coupled non-cleaving, mutated-core versions of our riboswitches to a sensor chip, passed various concentrations of MS2 over the chip, and analyzed binding kinetics. We found that the aptamers integrated into the ribozymes generally bound their cognate protein as well as aptamer alone, some with dissociation constants in the tens of nanomolar range (Figure 2.4). Riboswitches with MS2 aptamers integrated into stem III, wherein the aptamer terminal loop is replaced by a stem leading into the ribozyme core, did not exhibit any binding. This is consistent with the previous observation that the bases in this terminal loop of the aptamer participate in MS2 binding (Horn et al., 2004), and changing their configuration from an open loop into a base-paired structure appears to disallow these protein-aptamer interactions.

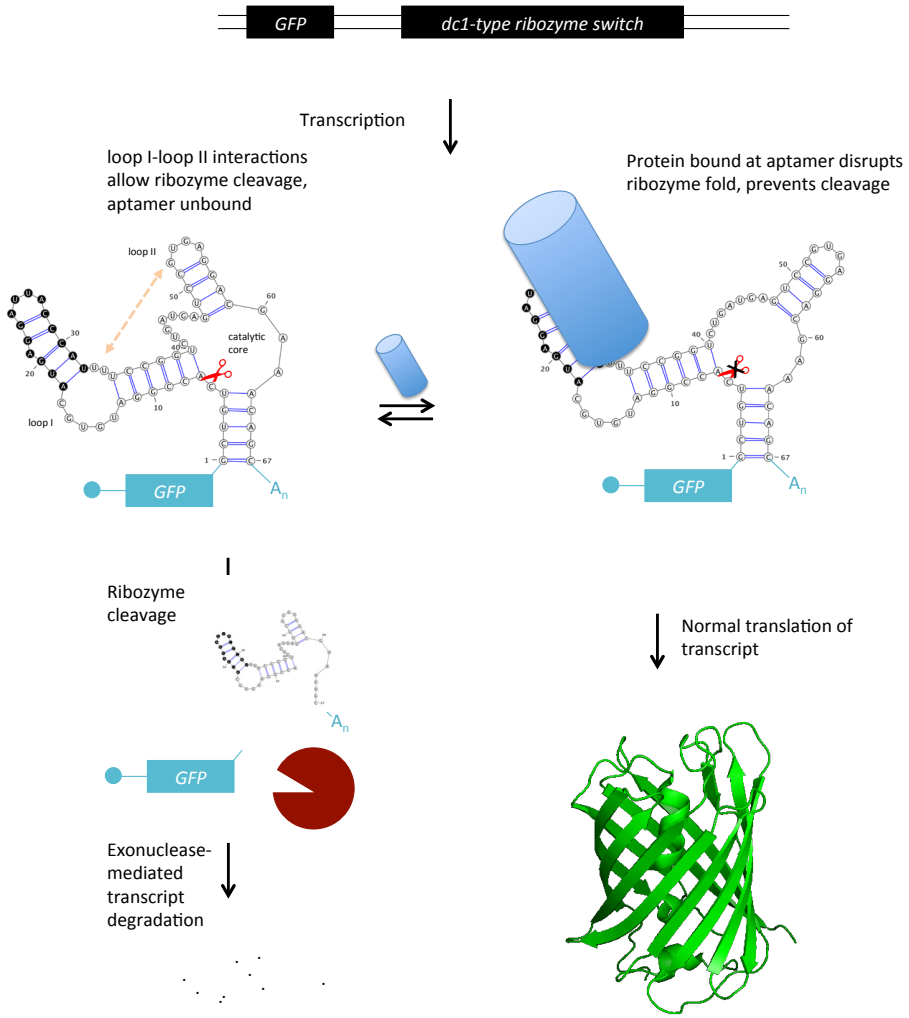


Figure 2.2 Design of protein-responsive ribozyme switch by direct coupling into loop I

dc1-type protein-responsive ribozyme switches were designed by coupling a protein aptamer (black bases) directly into loop I of a self-cleaving hammerhead ribozyme. The switch is encoded in the 3' untranslated region of a target gene, in this case GFP, and upon transcription in eukaryotic cells is designed to exist in two primary conformations. In the ribozyme-active, aptamer-unbound conformation, Loop 1-loop 2 interactions and cleavage-necessary structural elements allow for ribozyme self-cleavage activity. In the aptamer-bound conformation, disruption of cleavage structural elements decreases cleavage. Increased levels of protein ligand (blue barrel) may increase the fraction of transcripts in ribozyme-inactive conformations. Ribozyme cleavage leads to rapid exonuclease-mediated transcript degradation, and lower reporter gene expression. These switches are designed to be "ON"-switches in that increased levels of protein ligand decrease cleavage activity and turn gene expression "ON."

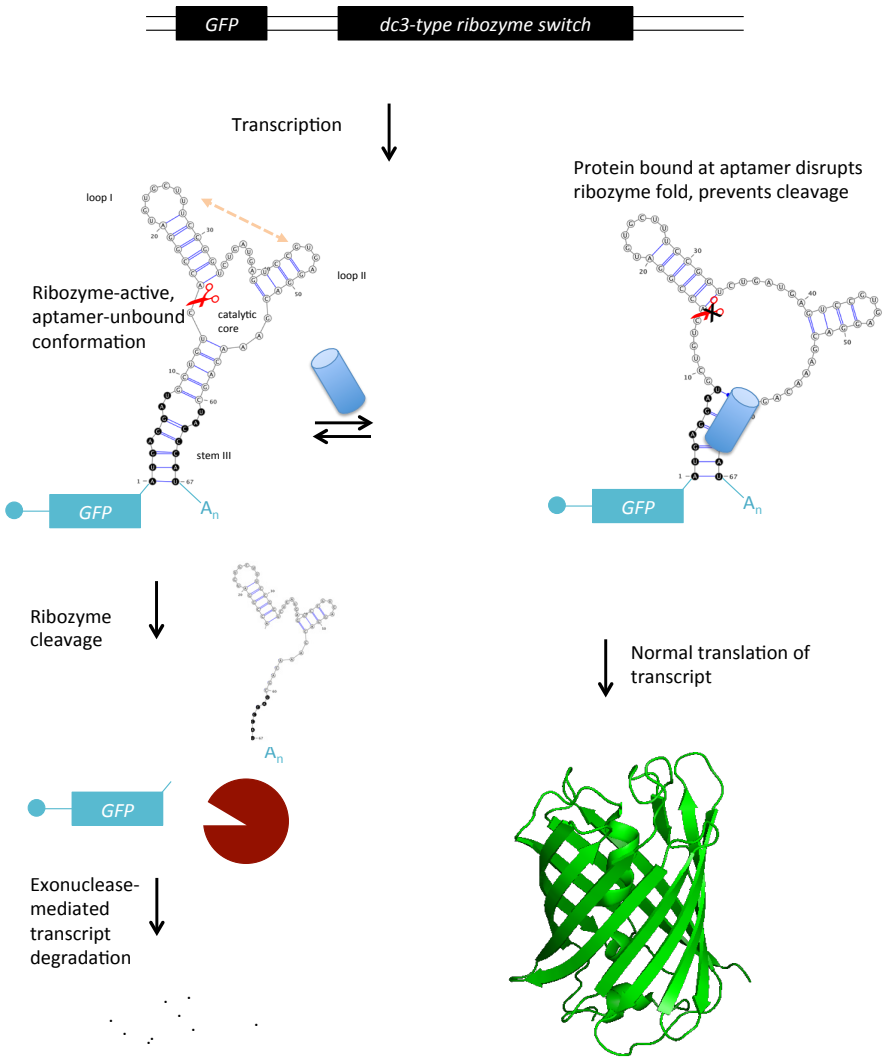


Figure 2.3 Design of protein-responsive ribozyme switch by direct coupling into stem III

dc3-type protein-responsive ribozyme switches were designed by coupling a protein aptamer (black bases) directly in stem III of a self-cleaving hammerhead ribozyme. The switch is encoded in the 3' untranslated region of a target gene, in this case GFP, and upon transcription in eukaryotic cells is designed to exist in two primary conformations. In the ribozyme-active, aptamer-unbound conformation, structural elements allow for ribozyme self-cleavage activity. In the aptamer-bound conformation, disruption of these cleavage-necessary structural elements decreases cleavage. Increased levels of protein ligand (blue barrel) may increase the fraction of transcripts in ribozyme-inactive conformations. Ribozyme cleavage leads to rapid exonuclease-mediated transcript degradation, and lower reporter gene expression. These switches are designed to be "ON"-switches in that increased levels of protein ligand decrease cleavage activity and turn gene expression "ON."

	K_D (nM)	k_a (1/M-s)	k_d (1/s)
M1	9.99	7,338,000	0.07334
M2	0.12	23,890,000*	0.00284*
M3	0.07	27,730,000*	0.00190*
dc1M1	17.10	3,815,000	0.06524
dc1M2	0.02	151,400,000*	0.00230*
dc1M3	0.36	16,100,000*	0.00572*
dc3M1	n.b.	n.b.	n.b.
dc3M2	n.b.	n.b.	n.b.
dc3M3	n.b.	n.b.	n.b.

Figure 2.4 Kinetic parameters of direct-coupling MS2 ribozyme switch *in vitro* protein binding
Dissociation constant (K_D), association (k_a) and dissociation (k_d) rates for selected aptamer and ribozyme switch constructs binding purified MS2 (see Materials and Methods) show that aptamers integrated into ribozymes directly bind cognate protein similarly to aptamer alone. M1–3 are different aptamers to MS2 (Parrott et al., 2000). dc1M1–3 are switches into which corresponding MS2 aptamers were integrated into loop I. dc3M1–3 are switches into which corresponding MS2 aptamers were integrated into stem III. Some constructs showed no binding (n.b.) to MS2. All switches contained a mutated core that abolishes cleavage activity, allowing for clear identification of protein binding and dissociation events. *Some values of k_d and k_a might be masked by mass transfer limitations, which would not affect K_D values (Karlsson, 1999).

Next, we sought to determine whether these switches maintained *in vivo* cleavage activity when modified with these aptamers. We chose to initially develop our devices in yeast, owning to strong observed correlation between ribozyme switch behavior in mammalian cells and *S. cerevisiae* (Wei et al., 2013), allowing for rapid device development and characterization in this latter model organism. We cloned our switches into the 3′ untranslated region (UTR) of a GFP reporter expression construct and expressed these plasmids in yeast. The data indicate that cleavage was generally robust, in some cases as efficient as from wild type ribozyme, achieving as much as 98% reduction in GFP levels (Figure 2.5). In

general, for the same aptamer, integration through stem III elicited greater cleavage activity than through loop I. This observation is perhaps due to the fact that with stem III integration, the generally base-paired structure of the aptamers stabilizes the formation of the stem, helping to bring together the otherwise most distal segments in the ribozyme fold, serving as a molecular clamp. Loop I integration, by contrast, might in some cases interfere with tertiary interactions between loops I and II required for *in vivo* activity (De la Peña et al., 2003; Khvorova et al., 2003).

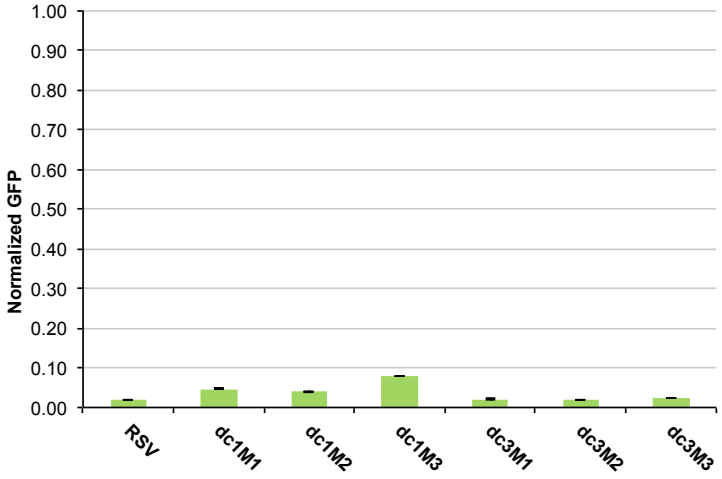


Figure 2.5 Basal expression of transcripts with 3' UTR direct integration MS2 riboswitches
When cloned in the 3' UTR of a reporter GFP gene, direct- coupling, MS2 aptamer- containing ribozyme switches strongly abrogate GFP expression in yeast. RSV is wild type ribozyme; dc1M1–3 are ribozymes into which MS2-binding aptamers M1–3 were cloned at an integration point on loop I; dc3M1–3 contain the same aptamers in stem III. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean.

Knowing that in general these switches cleave efficiently and can bind their cognate proteins relatively tightly, we next examined whether protein binding was able to modulate cleavage activity *in vivo*. We expressed plasmids bearing our switches in the 3′ UTR of GFP alongside either empty control plasmids or plasmids expressing MS2 to observe what effect MS2 expression had on switch activity. Generally, the switches did not show any significant modulation of GFP expression abrogation when co-expressed with MS2, with at most ~0.48-fold change in expression, compared to ~0.20-fold for RSV (Figure 2.7).

We hypothesized that protein localization might play an important role in protein-responsive ribozyme switch regulation. Whereas small molecules such as theophylline can likely diffuse readily through the nuclear pore channel, protein transport between the nucleus and the cytoplasm is protein sequence-dependent, with localization signals dictating protein levels among different cellular compartments (Pemberton et al., 1998). For switches being transcribed in the nucleus and spending up to several minutes undergoing mRNA processing and nuclear transport before reaching the cytoplasm (Oeffinger and Zenklusen, 2012), absence of nuclear-localized ligand may have drastic consequences. For a kinetic switch in which ligand-binding would funnel subsequent folding trajectories into ligand-bound conformations (Quarta et al., 2012), absence of that ligand in the nucleus would result in little to no allosteric regulation. And regardless of switching mechanism, measured ribozyme cleavage kinetics at physiological conditions (Kennedy et al., 2012; Liang et al., 2012) predict a significant fraction of transcripts cleaved even after a few minutes, leaving little chance for cytoplasmic protein-mediated GFP expression modulation (Figure 2.6).

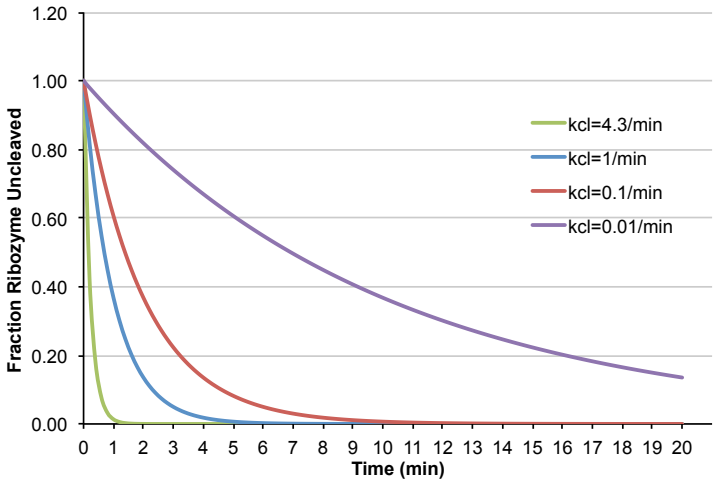


Figure 2.6 Modeled cleavage kinetics at times after ribozyme transcription

Ribozyme cleavage was modeled at different time points by single exponential decay using cleavage constants of ribozyme and ribozyme switch devices determined experimentally at physiologically relevant conditions (Kennedy et al., 2012; Liang et al., 2012). The wild type ribozyme cleaves at 4.3/min. Mean mRNA half-life in yeast is ~11 min (Jelinsky and Samson, 1999).

Thus, we created an MS2 variant containing an N-terminal nuclear localization signal (NLS) and co-expressed it alongside our switches. Generally, NLS-MS2 co-expression did not lead to any evidence of switch cleavage inhibition. If anything, GFP levels decreased (Figure 2.7), although this was also the case for the dc3 constructs previously demonstrated not to bind MS2, suggesting that any GFP expression modulation is not mediated by MS2–aptamer binding.

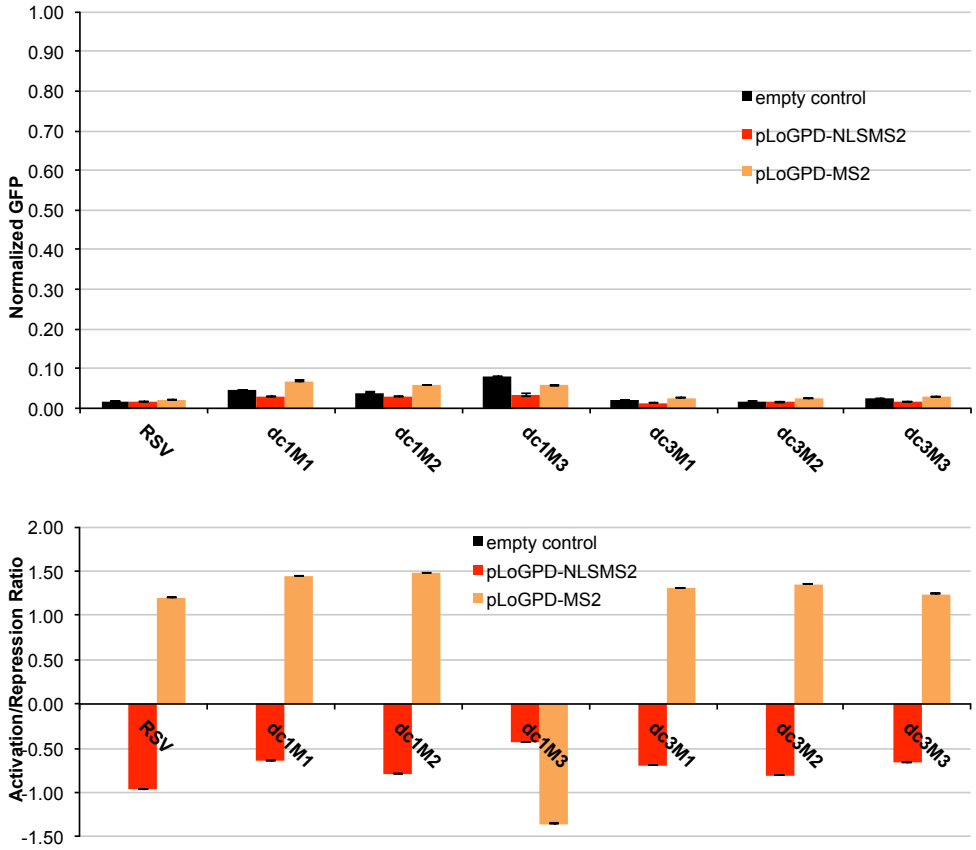


Figure 2.7 MS2 direct-coupling ribozyme switch response to MS2 protein variants

Direct-coupling, MS2 aptamer-containing ribozyme switches appear not to modulate GFP expression in response to cognate MS2 expression in yeast. Selected devices were cloned in the 3' UTR of a reporter GFP plasmid and expressed alongside an empty control or vectors expressing nuclear-localized (NLS-) MS2 or MS2. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean. For samples showing increased GFP levels upon MS2 variant expression, activation ratios (ARs) are reported as ratios of mean GFP levels from device constructs co-expressed with MS2, divided by mean GFP levels from the same device when co-expressed with an empty control. For devices showing decreased GFP levels upon MS2 variant expression, repression ratios (RRs) are reported as the negative ratios of mean GFP levels from device constructs co-expressed with an empty control, divided by mean GFP levels from the same device when co-expressed with MS2. RSV is the wild type ribozyme which contains no MS2 aptamer.

2.2.2 Integrating p50 aptamers into ribozymes by direct coupling

To determine whether the lack of observed switching from direct-coupling MS2 ribozyme switch designs was specific to these particular aptamer–protein pairs being suboptimal, or due to our design strategy not allowing for switching, we next decided to construct switches to a different protein following the direct-coupling strategy. We integrated into the same locations on loop I and stem III an aptamer to Nuclear Factor-κB p50 subunit (p50) previously shown to bind cognate protein with high affinity *in vitro* and *in vivo* (Cassiday and Maher, 2003; Wurster et al., 2009). Again, we used the SPR assay to characterize switch–p50 interactions *in vitro* and observed very tight binding, with dissociation constants in the low nanomolar range, generally 10- to 100-fold lower than for our MS2 aptamers (Figure 2.8).

	K_D (nM)	k_a (1/M-s)	k_d (1/s)
K1	2.57 ± 0.16	339,000 ± 80,050	0.000870 ± 0.000150
dc1K1	2.34 ± 0.19	1,12,000 ± 50,870	0.00262 ± 0.000350
dc3K1	7.37 ± 0.32	310,000 ± 70,100	0.00185 ± 0.000780

Figure 2.8 Kinetic parameters of p50 direct-coupling ribozyme switch *in vitro* protein binding
Dissociation constant (K_D), association (k_a) and dissociation (k_d) rates for selected aptamer and ribozyme switch constructs binding purified p50 (see Materials and Methods) show that aptamers integrated into ribozymes directly bind cognate protein similarly to aptamer alone. K1 is an aptamer to p50. dc1K1 is a ribozyme into which K1 was were integrated into loop I. dc3K1 is a ribozyme into which p50 aptamer K1 was integrated into stem III. All ribozyme switches contained a mutated core that abolishes cleavage activity allowing for clear identification of protein binding and dissociation events.

We then examined whether these integration points maintained *in vivo* cleavage activity by cloning our devices into the 3′ UTR of a reporter GFP

plasmid. We observed higher GFP expression, and presumably lower levels of cleavage, than from the MS2-responsive ribozyme switches (Figure 2.9). For example, dc1K1 stood out in displaying no evidence of cleavage activity. We had chosen our integration point on loop I based on previous experiments with a theophylline aptamer which found that this integration point was the best option for maintaining cleavage, with only four times less cleavage activity than the wild type ribozyme (Kennedy and Smolke, unpublished results). However, seeing that this was far from the case for our p50 aptamer, we investigated other integration points on loop I, and found that many allowed for more efficient cleavage (Figure 2.9). While it is unclear why different aptamers coupled into different loop I integration points disturb ribozyme activity to different extents, it may have to do with the different aptamer sequences interacting differently with the peripheral loop I–loop II tertiary interactions.

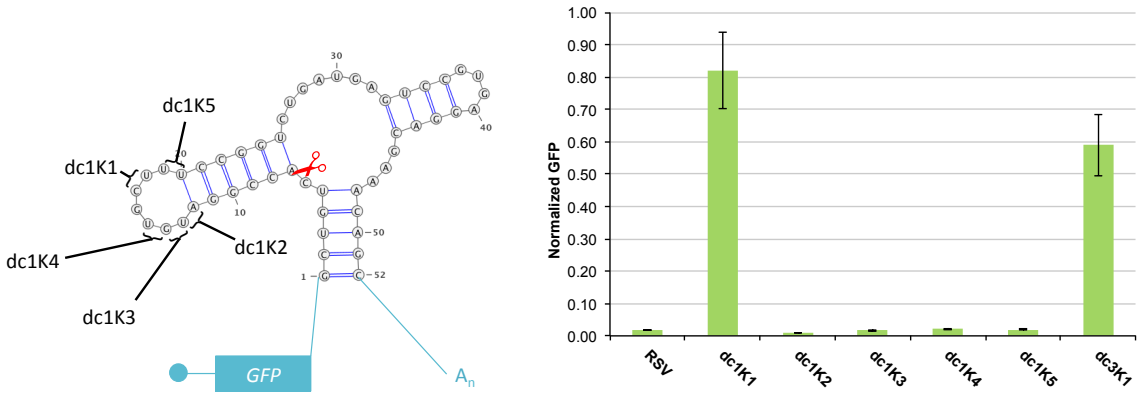


Figure 2.9 p50 direct-coupling ribozyme switch basal *in vivo* expression

When cloned in the 3' UTR of a reporter GFP gene, direct-coupling, p50 aptamer-containing ribozyme switches strongly abrogate GFP expression in yeast. RSV is wild type ribozyme, dc1K1 contains a p50 aptamer integrated into loop I, dc3K1 into stem III. dcK2–5 integration sites shown on left, with cleavage site indicated by a red scissors. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean.

To investigate *in vivo* switching from these devices, we co-expressed our switch-containing GFP plasmids alongside either an empty vector or vectors expressing p50 protein variants. Generally, we did not observe any significant modulation of GFP levels resulting from expression of p50 or NLS-tagged p50 (Figure 2.10), suggesting that protein-switch binding did not modulate cleavage activity. Other work in our group suggests that binding of proteins to cognate aptamers integrated into loop II did not result in cleavage modulation in vitro (Kennedy and Vowles, unpublished work), supporting the in vivo observations.

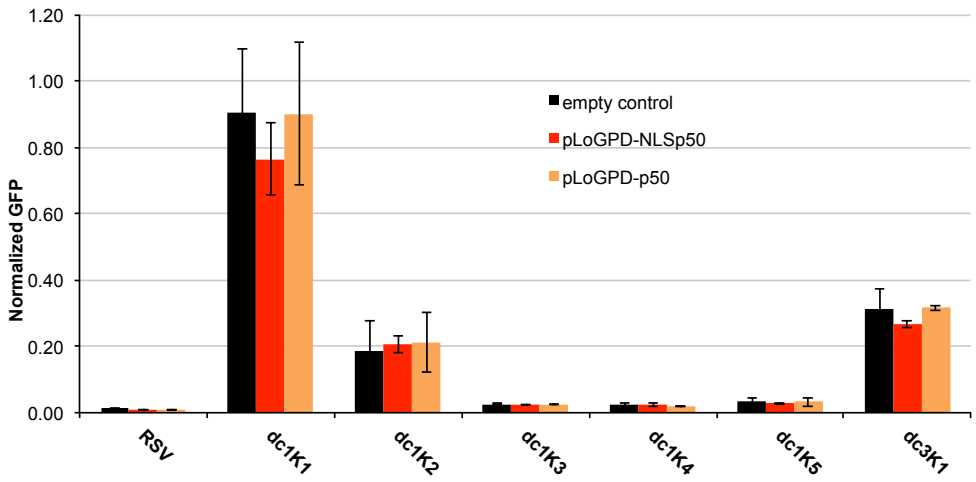


Figure 2.10 p50 direct-coupling ribozyme switch response to *in vivo* p50 expression

Direct-coupling, p50 aptamer-containing ribozyme switches appear not to modulate GFP expression in response to cognate p50 expression in yeast. Selected devices were cloned in the 3' UTR of a reporter GFP plasmid and expressed alongside an empty control or vectors expressing NLS-p50 or p50. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean.

2.2.3 Integrating MS2 aptamers into ribozymes by strand displacement

Seeing that our direct-integration strategy failed to produce functional protein-responsive ribozyme switches, we attempted a new set of switch designs using a strand-displacement design strategy previously developed in our laboratory (Win and Smolke, 2007, 2008a). As with previously described small-molecule aptamers, these switches are composed of an aptamer coupled to a hammerhead ribozyme through a transmitter sequence that allows for two primary switch conformations that differ in whether they contain properly folded ribozyme or aptamer domains. In switch "ON" designs, aptamer and ribozyme folds are mutually exclusive within each primary conformation, whereas in the "OFF" design, they are mutually inclusive. Aptamer–ligand binding stabilizes the aptamer-containing conformation, modulating the fraction of switches containing the ribozyme-active conformation, and consequently gene expression levels (Figure 2.11).

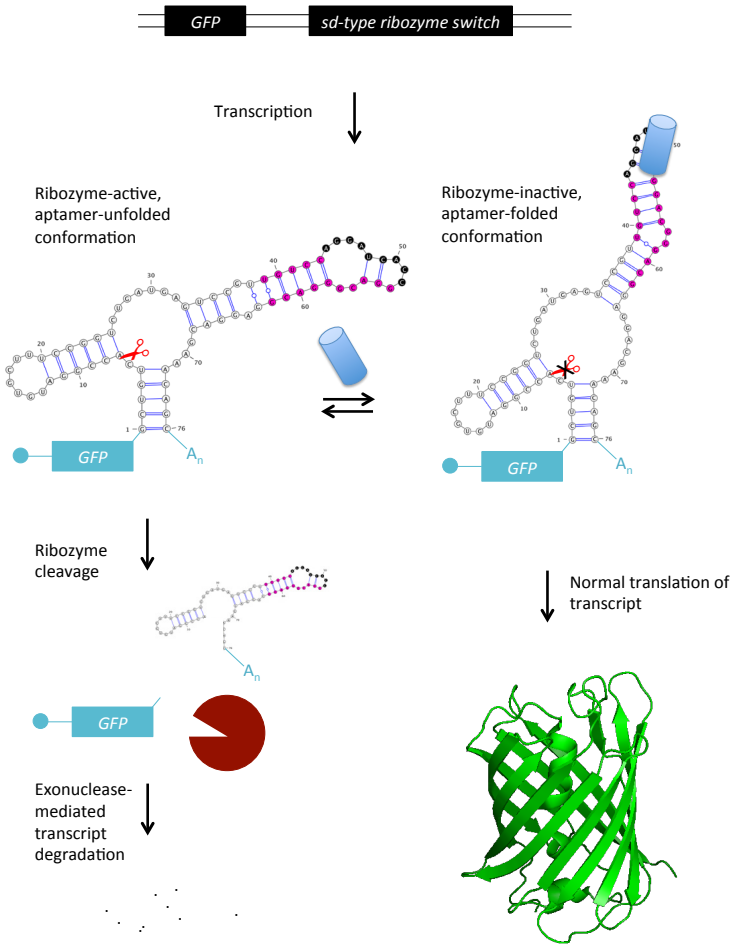


Figure 2.11 Design of strand-displacement ribozyme switches

Strand displacement (sd-type), protein-responsive ribozyme switches were designed by coupling a self-cleaving ribozyme to a protein-binding aptamer sequence (black bases) through engineered transmitter sequences (pink bases). The switch is encoded in the 3' untranslated region of a target gene, in this case GFP, and upon transcription in eukaryotic cells is designed to exist in two primary conformations. In this example of an "ON" switch, conformations are mutually exclusive in whether they contain a ribozyme-active domain or an aptamer-folded domain. Ribozyme-active conformations undergo cleavage, leading to rapid exonuclease-mediated transcript degradation, and loss of GFP expression. Binding of the protein ligand (blue barrel) stabilizes the aptamer-folded conformation, which in this case is diminishes cleavage activity, and allows for greater GFP expression. Thus, increased levels of ligand lead to target gene expression being turned "ON." In "OFF" switch designs (not shown), ribozyme-active and aptamer-folded domains are mutually inclusive. In the absence of protein ligand, transcripts primarily do not undergo cleavage, but do so when protein ligand binding stabilizes the aptamer-folded and ribozyme-active conformation.

We designed switches bearing the MS2 aptamer we found to be tightest binding based on previous characterization, M3, integrated through various transmitter and loop sequences. We refer to these switches by sdMX, where "sd" indicates strand displacement and X refers to loop and transmitter modifications. sdM-1, -2, and -4 are designed to be "ON" switches, in that aptamer- and ribozyme-domain folding are mutually exclusive in the two main conformations, and MS2 binding is expected to increase GFP expression. sdM3 is designed to be an "OFF" switch, in that aptamer- and ribozyme-domain folding are mutually inclusive, and MS2 binding is expected to decrease GFP expression.

First, we examined whether these switches can bind MS2 through the SPR assay. We coupled our switches to a sensor chip under non-cleaving conditions, passed various concentrations of MS2 protein over the chip surface, and analyzed binding kinetics. The data indicate that most of our switches bind MS2 with high affinity, with dissociation constants as low as 6.1 ± 2.4 nM for sdM2 (Figure 2.12).

	K_D (nM)	k_a (1/M-s)	k_d (1/s)
sdM1	24.4 ± 5.7	82,250 ± 29,810	0.00189 ± 0.00028
sdM2	6.1 ± 2.4	254,600 ± 105,800	0.00138 ± 0.00021
sdM3	629.7 ± 152.1	18,800 ± 6,864	0.01115 ± 0.00103
sdM4	253.0 ± 13.9	140,300 ± 59,600	0.03889 ± 0.01901

Figure 2.12 Kinetic parameters of strand-displacement ribozyme switch–protein interaction from SPR
Dissociation constant (K_D), association (k_a) and dissociation (k_d) rates for selected aptamer and ribozyme switch constructs containing MS2 aptamers integrated through strand- displacement transmitter sequences (sdM1–4) bind purified MS2 (see Materials and Methods) with high affinity.

Next, we examined how MS2 aptamer integration via transmitter sequences affected basal cleavage activity. As with the direct-coupling switches, we cloned our switches in the 3′ UTR of a GFP reporter gene and measured fluorescence levels from the cell population. Generally, GFP levels were reduced compared to a non-cleaving control, suggesting robust cleavage activity from our test devices. Basal expression levels from the strand-displacement switches were generally higher than from the direct-coupling switches. This observation may be explained by the fact that strand-displacement switches were designed with transmitter sequences that allow the switch molecule to change conformation between two primary folds, whereas direct-integration switches are designed for one ribozyme-active primary fold. Thus, strand-displacement switches would likely sample both ribozyme-active and -inactive conformations, leading to less overall cleavage. Regardless, many of the strand-displacement switches exhibited robust GFP inhibition in the basal state. For example, sdM2, which was designed by replacing loop I from sdM1 with a sequence previously shown to significantly increase cleavage activity and lower expression in theophylline-responsive ribozyme switches (Kennedy et al., 2012; Liang et al., 2012), showed similarly low GFP expression in the context of the MS2 aptamer, approaching that from wild type ribozyme (RSV).

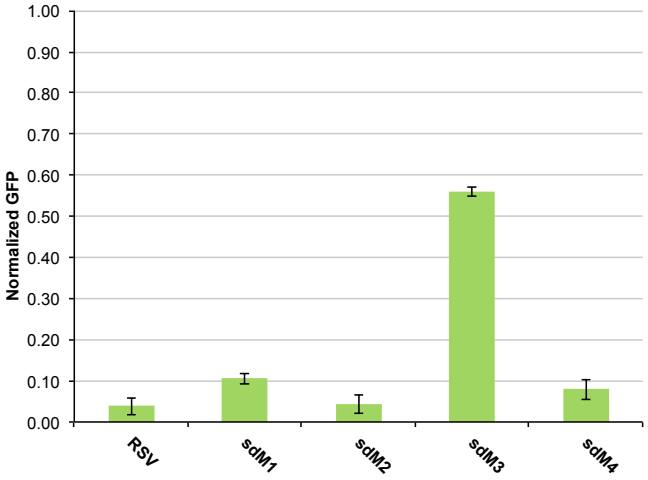


Figure 2.13 Basal expression of transcripts with 3' UTR strand-displacement MS2 switches

When cloned in the 3' UTR of a reporter GFP gene, strand-displacement ribozyme switches containing MS2 aptamers strongly abrogate GFP expression in yeast. RSV is wild type ribozyme, sdM1–4 contain MS2 aptamer integrated through various strand-displacement transmitter sequences. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate standard deviation.

We then examined MS2-mediated regulation of GFP levels by cloning our switches in the GFP expression construct and expressing these plasmids alongside either a control plasmid, or one expressing MS2. We observed more than 4-fold changes in GFP levels upon MS2 expression for both sdM1 and sdM3 (Figure 2.14). Additionally, GFP expression modulation followed the expected pattern, with sdM-1, -2, and -4 having been designed as "ON" switches and increasing GFP levels in response to MS2 expression, and sdM3 designed as an "OFF" switch, decreasing GFP levels in response to MS2 expression.

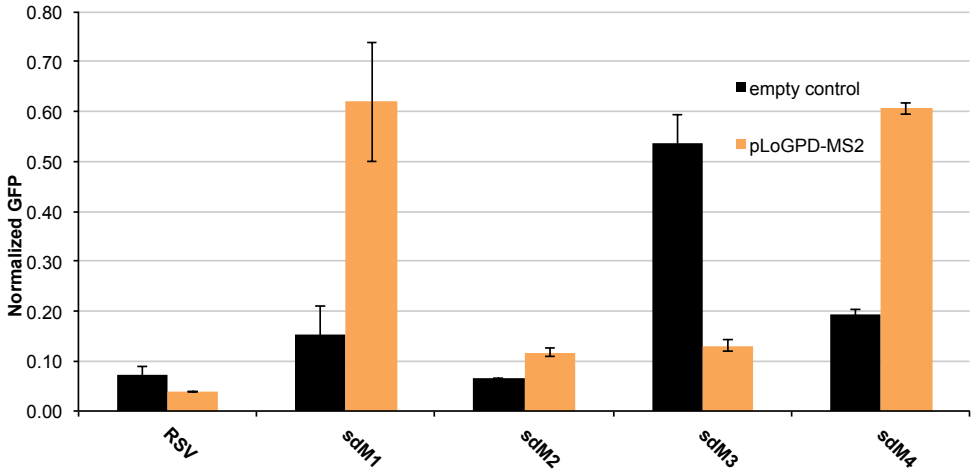


Figure 2.14 MS2 strand-displacement ribozyme switch response to *in vivo* MS2 expression
Strand-displacement, MS2 aptamer-containing ribozyme switches sdM1–4 cloned in the 3' UTR of a reporter GFP gene exhibit MS2 expression-mediated regulation of GFP levels in yeast. RSV is a wild type ribozyme without any MS2 aptamer. In response to increased MS2 levels, sdM-1, -2, and -4 are designed to increase GFP levels; sdM3, to decrease GFP levels. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean.

To confirm that the observed GFP regulation *in vivo* was brought about by MS2-mediated regulation of switch cleavage activity, we performed *in vitro* experiments to characterize switch cleavage modulation in response to varying MS2 levels. We found that the presence of MS2 significantly affected cleavage kinetics for the strand-displacement switches, by up to 13-fold (Figure 2.15). Further, cleavage modulation followed the expected qualitative trends, with "ON" switches exhibiting lower cleavage rates in the presence of MS2, and the "OFF" switch higher. A control switch that cleaves at similar rates as our sdM switches but which contains a theophylline aptamer (sd2T8) and no MS2 aptamer showed no MS2-mediated cleavage modulation, as expected.

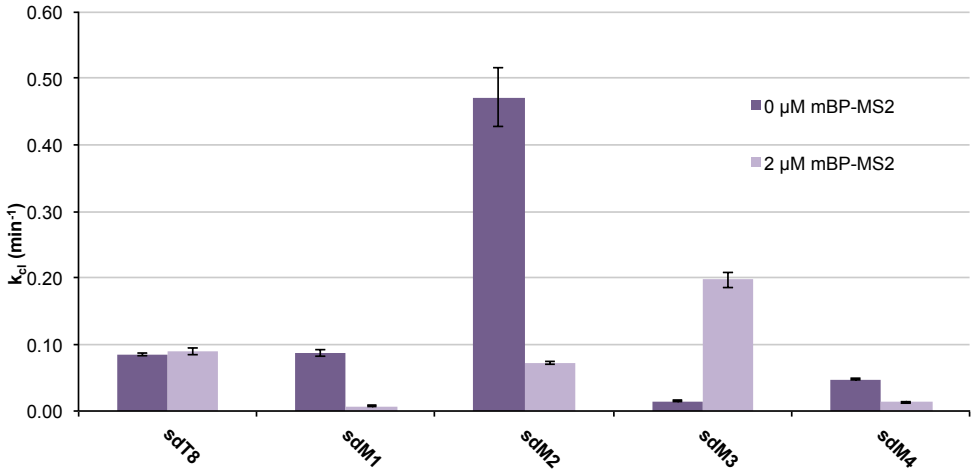


Figure 2.15 *In vitro* cleavage kinetics of selected ribozyme switches with and without MS2 protein
Synthesized ribozyme switches incubated with or without purified MS2 at physiological conditions exhibit MS2-dependent cleavage modulation. sdT8, a ribozyme containing no MS2 aptamer, is unaffected by MS2 levels. k_{cl} is the cleavage rate as predicted single exponential decay fitted from gel-based cleavage assays (See Materials and Methods).

We next sought to explore the mechanism of protein ligand-mediated switching *in vivo*. Riboswitches have been proposed to operate in a spectrum within two extremes—kinetic and thermodynamic switching (Quarta et al., 2012). For kinetic switching, ligand binding during transcription funnels riboswitch folding into aptamer-bound conformations, from which high barriers to structural rearrangements disallow transitions into aptamer-unbound conformations and their corresponding ribozyme states. In thermodynamic switching, energy barriers to structural rearrangements are low, and riboswitches can freely sample aptamer-bound and -unbound conformations, with ligand binding stabilizing aptamer-bound forms and their corresponding ribozyme states. Where on this spectrum our ribozyme switches lie can have important repercussions in how our platform can be modified to respond to different protein ligands. Proteins not

found in the nucleus, for example, would be unlikely to allow kinetic switching. Even for thermodynamic switching, absence of protein ligand during the time required for mRNA processing and nuclear export can result in a significant fraction of transcripts cleaved (Figure 2.15), making it unlikely that protein ligand binding would significantly modulate target gene expression. Regardless of switching mechanism, whether our platform necessitates certain localization of protein inputs for output gene expression modulation has significant ramifications choice of input.

To investigate these localization requirements, we constructed MS2 variants tagged with N-terminal nuclear localization or export signals (NLS or NES, respectively) and co-expressed these along with two of our best-performing strand-displacement switches. We examined what effect input protein localization had on ribozyme switch activity (Figure 2.16). Generally, switching was less significant from both NES- and NLS-tagged MS2 than from untagged protein, the latter expected to localize to both compartments owning to its small size (Singer and Grünwald, 2013). Because transcripts spend time both in the nucleus and in the cytoplasm prior to and during translation, MS2 expression in only one compartment may result in loss of ribozyme cleavage modulation in the other compartment, leading to lower overall GFP switching.

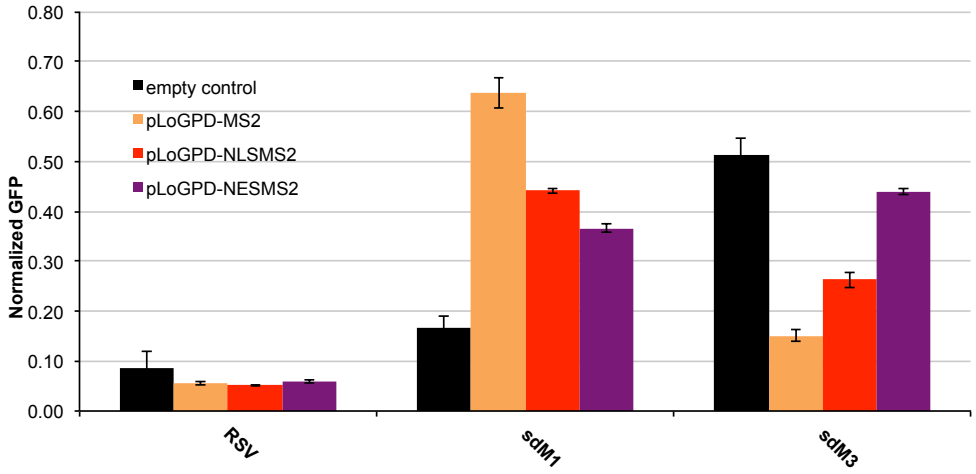


Figure 2.16 MS2 protein localization effect on strand-displacement ribozyme switch response
MS2 aptamer-containing ribozyme switches sdM1 ("ON" switch) and sdM3 ("OFF" switch) cloned in the 3' UTR of a reporter GFP gene exhibit different responses to co-expression of MS2 variants in yeast. MS2 was tagged with nuclear localization signal (NLS), nuclear export signal (NES), or no tag. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean.

Interestingly, GFP expression modulation was greater from NLS-MS2 expression than from NES-MS2 expression. We can interpret this result in various ways. A simple explanation would be that NLS-MS2 is expressed at higher levels than NES-MS2. Another would be that NLS and NES tags affect MS2 variant-aptamer binding to different extents. Or, it may be that nuclear MS2 localization would more significantly affect device switching via a kinetic mechanism alluded to earlier, in which protein binding *during* transcription funnels ribozyme switch folding to aptamer-active conformations and their corresponding ribozyme states. Alternatively, it may be that presence of nuclear ligand may lead to greater overall switching, whether because transcripts spend a significant amount of time in the nucleus, or because the time spent there has a greater effect on ligand-

mediated switching than time spent in the cytoplasm. Of course, the greater switching from NLS-MS2 may have to do with a combination of these factors, and we are currently performing additional experiments to further probe these questions.

2.2.4 Translation of strand-displacement ribozyme switches to mammalian cells

Having demonstrated our platform in yeast, we next explored whether our protein-responsive ribozyme switches can operate directly in mammalian cells without further modification of switch sequences. To be able to induce input MS2 expression, we constructed mammalian expression plasmids encoding MS2 driven by a doxycycline-inducible promoter system. In addition, we placed the strand-displacement ribozyme switch in the 3′ UTR of a blue fluorescent protein (BFP) controlled by a constitutive promoter. The constructs were co-transformed into T-Rex HEK293 cells and MS2 expression induced with doxycycline. The data indicate that the switches function in the mammalian cell host to modulate BFP expression levels in response to MS2 (Figure 2.17).

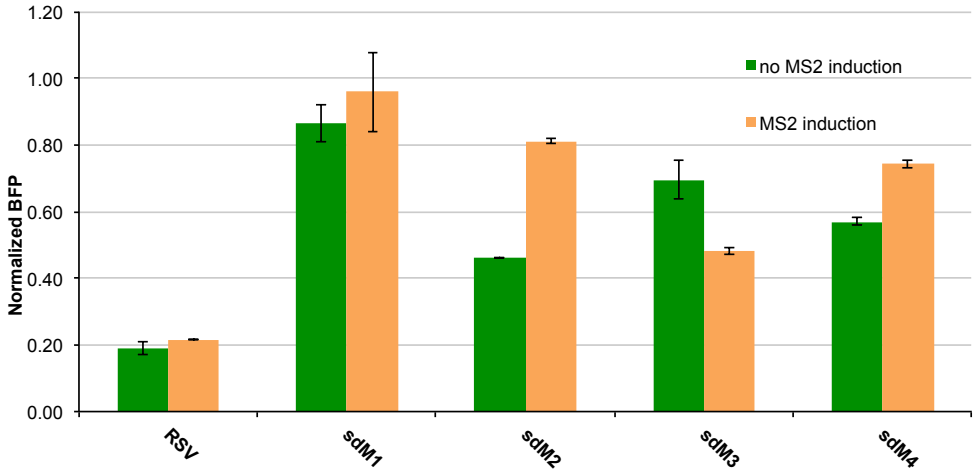


Figure 2.17 MS2-responsive strand-displacement ribozyme switch behavior in mammalian cells
Strand-displacement, MS2 aptamer-containing ribozyme switches sdM1–4 cloned in the 3' UTR of a reporter BFP gene exhibit MS2 expression-mediated regulation of BFP levels in mammalian cells. RSV is a wild type ribozyme without any MS2 aptamer. In response to increased MS2 levels, sdM-1, -2, and -4 are designed to increase BFP levels; sdM3, to decrease BFP levels. Mean BFP levels are shown normalized to a non-aptamer, non-cleaving control. Error bars indicate one standard deviation from the mean.

In general, the gene-regulatory activity of these devices was lower in mammalian cells than in yeast as has been previously observed (Chen et al., 2010; Wei et al., 2013). For example, wild type ribozyme elicited only 80% abrogation of reporter expression compared to ~98% in yeast. As a result, switches with higher basal gene expression tended to exhibit lower dynamic ranges in mammalian cells, as these basal levels were already close to maximal expression. Conversely, sdM2, which in yeast appears to cleave at too high a rate to allow for appreciable gene expression in the "ON" state, exhibited better dynamic range in the human cell line.

2.3 Conclusions

This work is aimed at expanding the capabilities for synthetic gene regulation in eukaryotes. Existing gene control platforms are handicapped owing to possible immunogenicity stemming from expression of platform components, organism-specific functionality, or perhaps most limiting, lack of modularity. These characteristics limit their widespread use for developing more sophisticated synthetic biological systems for treatment of human disease, and for cellular engineering in general. The ribozyme switch platform (Win and Smolke, 2007) addresses many of these concerns. However, it has not yet been shown to respond to changing levels of proteins, which characterize many changes in cellular state.

This study resolves this limitation and demonstrates a synthetic ribozyme-mediated, protein input-responsive gene regulation platform in eukaryotic cells. The switches developed here achieved more than 4-fold modulation of target gene expression in response to the presence of a protein input in yeast. *In vitro* characterization provided evidence that the observed gene expression modulation stems from protein input-mediated modulation of ribozyme cleavage activity. We further demonstrated that the platform retains protein-responsiveness in mammalian cells. And lastly, preliminary experiments indicate that the platform can be used to detect and respond to changes in levels of proteins localized throughout the cell, only in the nucleus, or only in the cytoplasm.

Interestingly, our initial direct-coupling designs failed to elicit protein-responsive gene expression control. Although *in vitro* experiments demonstrated that generally dc1 and dc3 designs bound cognate protein effectively, and maintained *in vivo* cleavage activity, we failed to observe any protein-responsive

cleavage modulation, either *in vitro* or *in vivo*. We demonstrated that certain integration points on loop I modulated basal cleavage dramatically. Ostensibly, this modulation might be due to the addition of aptamer sequences in certain ribozyme locations affecting ribozyme folding, perhaps by disallowing loop I–loop II interactions required for *in vivo* activity. This has been suggested by other work with theophylline-responsive aptamers located on different loop I integration points (Kennedy and Smolke, unpublished results), although different integration points with different aptamers affect cleavage activity differently. Still, it would seem reasonable to assume that for certain integration points, and certain aptamers, addition of aptamer alone to the ribozyme structure might not affect cleavage, whereas protein binding would, either by disallowing loop I–loop II interactions, or by some other structure modulation effect. However, given that different aptamers on the same integration point affect cleavage to different extents, and that we failed to observe protein-responsive cleavage modulation even for dc1K1–5 which had p50 aptamer integrated on various loop I locations, it would seem that this strategy is either flawed, or at the very least not generalizable to a given protein–aptamer pair of interest. Similarly, stem III integration failed to produce any protein input-mediated change in gene expression. And further, the fact that aptamers M1–3 failed to bind cognate protein when integrated in stem III makes it difficult to claim that such a design might be generalizable. Indeed, many aptamers mediate binding to cognate ligands through their unstructured regions—often terminal or internal loops—making it unlikely that ribozyme integration of these aptamers in such a way that

these unstructured regions become part of a stem would maintain binding activity.

Turning back to our strand-displacement (sd-type) switches, this study explored ligand localization-dependence of switch gene-regulatory activity. The question is an important one, potentially limiting the range of ligands that the ribozyme switch platform may be employed to detect. It also brings up certain questions on the molecular design of such switches. In our designs, we have often stated that the switches "are designed" to exist in two primary conformations. Indeed, secondary RNA structure prediction software generally suggests that sd-type switches fold into main conformations with similar energies. Often, two "primary" folds are indeed predicted, which generally follow our designs. For instance, our sdM "ON" switches, as well as sd-type theophylline-responsive "ON" switches described earlier, do generally contain one most-favorable energetic fold (MFEF) and a slightly less-energetically favorable "nearest neighbor." The MFEF contains a properly folded ribozyme domain, but no aptamer domain, whereas the "nearest neighbor" contains a properly folded aptamer, but no ribozyme. In our designs, the free energy of this nearest neighbor often differs from that of the MFEF by ~2 times the ambient thermal energy, RT. In a Boltzmann distribution, this would account for the ribozyme-containing MFEF fold being populated at ~7 times the frequency of its aptamer-containing neighbor. Binding a ligand with, say, K_D~5 nM—similar to the affinity of many of our aptamer-protein pairs—would be expected to lower the energy of the aptamer-bound conformations by approximately twice RT, making this and the ribozyme-containing MFEF equally populated. Such a situation might perhaps

explain why generally these switches modulate gene expression by a maximum of ~10 fold, as do most classes of riboswitches in eukaryotic cells. Although useful for some applications, these dynamic ranges continue to be a limitation of riboswitch-mediated gene regulation in general.

However, there are several ways in which this description differs from the real world. First, RNA structure prediction often fails to recapitulate some characteristics of real-world folding, such as pseudoknots and other interactions including the loop I–loop II tertiary contacts required for *in vivo* ribozyme activity we discussed earlier. Second, it is unclear how these predictions apply to populations of mRNA at the single cell level, where mRNA copy number is generally low, on average ~2.5 (see Figure 1.3), and inherent "noise" in gene expression can lead to stochastic effects (Elowitz et al., 2002; Chang et al., 2006; To and Maheshri, 2010). And third, mRNAs are dynamically processed, with half-lives sometimes on the order of minutes, and it is unclear how folding in such a situation differs from predicted equilibrium folding.

The interplay among rates of cleavage, transcription, domain folding, overall structural rearrangements, and mRNA processing leading to translation are in the end what govern riboswitch function. For example, in the extreme that folding and rearrangement rates are much faster than rates of cleavage and mRNA processing, it would be predicted that RNA structures would reach their thermodynamic equilibrium states quickly. And that ligand addition would allow a fully formed MFEF to rearrange into a new ligand-favored fold, modulating cleavage activity. This situation is what we refer to when we say "thermodynamic switching." With low energy barriers to structural rearrangement, however, the

switches may significantly populate both cleavage-active and cleavage-inactive folds in both ligand-present and ligand-absent environments, potentially leading to low dynamic ranges.

However, with slower rates of structural rearrangements, it may be that certain folds that are not thermodynamically favored at equilibrium nevertheless dominate, owning to "funneling" effects. Conceptually, a "kinetic" switch might function by ligand binding mediating intermediate-energy conformations that funnel the switch into ultimately ligand-bound conformations. These ligand-bound conformations may or may not be thermodynamically favored over the ligand-unbound fold, but may dominate owning to high-energy barriers to structural rearrangement into the ligand-unbound folds (Figure 2.18).

Our direct-coupling switches were designed, at least conceptually, to perhaps follow this kinetic mechanism. It may be the case that owning to these funneling effects, a particular design may lead to a situation where in the presence of ligand, a significant majority of mRNAs are permanently funneled into ligand-binding conformations, and their corresponding ribozyme states. This may lead to higher dynamic ranges compared to thermodynamic switching. In thermodynamic switching, low energy barriers among ligand-bound and -unbound states may allow for rearrangements and thus ligand-mediated cleavage modulation, but perhaps at the expense of high dynamic ranges. With low energy barriers, both ligand-bound and -unbound states will be significantly populated at all times.

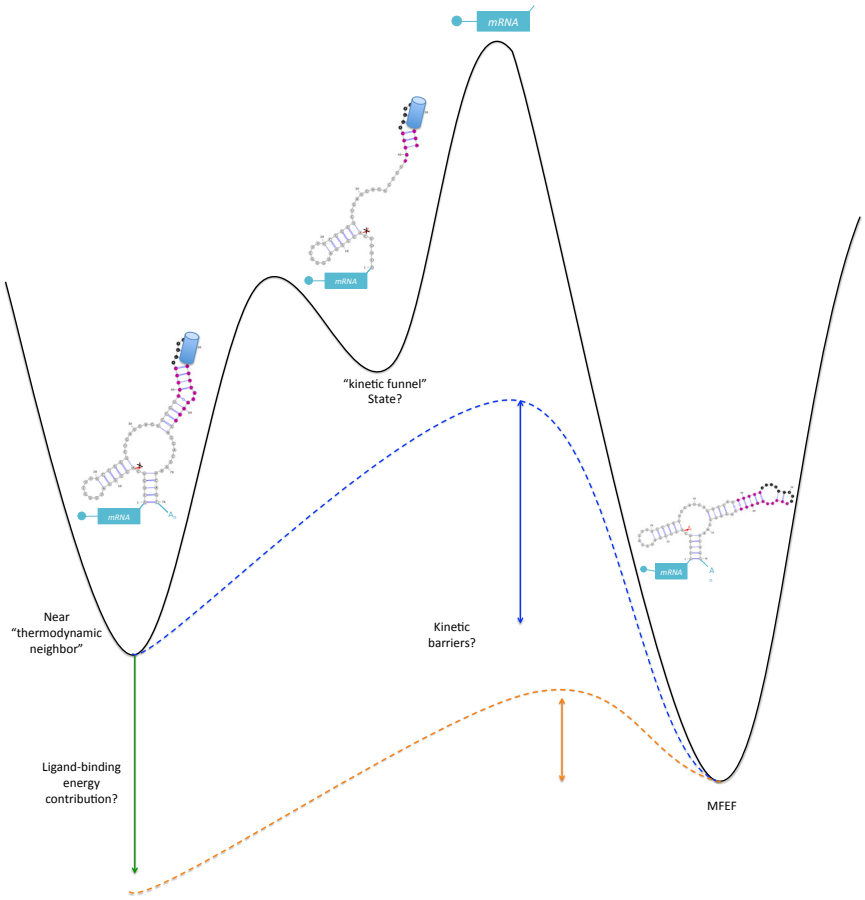


Figure 2.18 Conceptual energy landscape of ribozyme switch folding

Thermodynamic and kinetic considerations, along with rates of transcription, cleavage, and mRNA processing govern ribozyme switch dynamics. An "ON"-acting strand-displacement switch is "designed" to function thermodynamically, wherein the RNA structures generally exist in a most favorable energetic fold (MFEF). Addition of ligand is expected to stabilize ligand-binding, ribozyme-inactive folds, creating a new MFEF (green arrow). Low energy barriers (dashed red line) allow rearrangement of the MFEF into the new ligand-binding MFEF. In contrast, a kinetic switch is "designed" such that ligand-binding before reaching equilibrium "funnels" folding into a ligand-bound conformation (Near "thermodynamic neighbor"). High-energy barriers between this state and the ligand-unbound MFEF state prevent reaching the equilibrium MFEF state (dashed blue line). The interplay among relevant rates of structural rearrangements, mRNA translation, and ribozyme cleavage govern where in the spectrum between "thermodynamic" and "kinetic" switching a given switch operates, with consequences in dynamic ranges. Switch diagrams with incomplete structure represent species currently undergoing transcription.

There is some evidence to suggest that our sd-type switches indeed function thermodynamically. For one, our data suggest that sdM switches respond to MS2 expression both in the nucleus and in the cytoplasm. Because kinetic switches would be expected to "funnel" folding states during or shortly after transcription, this result suggests that the switching mechanism may be more thermodynamic. Second, our *in vitro* experiments consist of taking fully folded switches and exposing them to their ligand in physiological conditions. In these assays, presence of ligand is able to modulate cleavage activity, ostensibly through structural rearrangement between ligand-bound and unbound states, suggesting that these states are indeed accessible from one another. Perhaps also telling, the dynamic ranges of our strand-displacement switches do seem to align with thermodynamic predictions based on binding affinity, as discussed earlier. While not excluding the possibility of developing greater-dynamic range switches, these considerations are important to note as possible limitations of the platform.

Still, the expanded capability of the sd-type ribozyme switch platform to now sense protein inputs is further testament to its versatility and input modularity, a rare characteristic among eukaryotic gene control platforms. This development further expands the potential use of the platform in a variety of situations in which altered levels of protein signals are indicative of changes in cellular state. With increasing biochemical knowledge of protein-mediated phenotypic changes, an ever-expanding set of RNA aptamers (Ellington, 2004), and both rational design (Win and Smolke, 2007, 2008b) and directed evolution methods (Liang et al., 2012) to generate new ligand-responsive ribozyme switches, our platform is poised to allow for greater bespoke cellular reprogramming.

2.4 Materials and Methods

2.4.1 Plasmid and strain construction

Standard molecular biology techniques were used for DNA manipulation and cloning (Sambrook, 2001). Restriction enzymes, T4 DNA ligase, and other cloning enzymes were purchased from New England Biolabs (Ipswich, MA). PfuUltraII (Agilent Technologies, Santa Clara, CA) was used for high-fidelity PCR amplification. LR Clonase II and BP Clonase II (Invitrogen, Carlsbad, CA) were used for recombination into Gateway destination plasmids (Addgene, Cambridge, MA) following manufacturer's instructions. Oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA) or the Stanford Protein and Nucleic Acid Facility (Stanford, CA). Plasmids were transformed into One Shot Top 10 chemically competent E. coli (Invitrogen, Carlsbad, CA; F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galKrpsL (StrR) endA1 nupG) by standard methods. *E. coli* were grown in LB media (BD, Franklin Lakes, NJ) with 100 μg/mL ampicillin (EMD Chemicals, Gibbstown, NJ) or 50 μg/mL kanamycin (EMD Chemicals, Gibbstown, NJ), depending on the plasmid selection gene. Plasmids were prepped from overnight cultures of *E. coli* using Econospin All-in-One Mini Spin Columns (Epoch Biolabs, Missouri City, TX) according to the manufacturers' instructions. Sequencing was performed by Laragen Inc. (Los Angeles, CA) and Elim Biopharmaceuticals (Hayward, CA).

Plasmids for yeast MS2 or p50 protein expression were based on pAG414GPD-ccdB-EGFP (Addgene, Cambridge, MA, Figure 2.20). MS2 was

amplified from pCS1592, a gift of James Vowles (Department of Bioengineering, Stanford University, Stanford, CA) with primers containing the SV40 large T antigen nuclear localization signal, the Protein Kinase Inhibitor nuclear export signal, or no tag, immediately downstream of a Kozak sequence and immediately upstream of the MS2 or p50 start codon. p50 variants were amplified from pJ593, a gift of L. James Maher (Department of Biochemistry and Molecular Biology, College of Medicine, Mayo Clinic, Rochester, MN). Cellular localization tags were added immediately 5′ of start codons. PCR products were cloned into *attL1* and *attR1* sites within pAG414GPD-ccdB-EGFP (Addgene, Cambridge, MA) using BP and LR recombination (Invitrogen, Carlsbad, CA) following manufacturer's instructions.

Plasmids for characterizing switch activity in yeast were based on pCS1748, a centromeric vector encoding an *mCherry* gene under the control of a TEF1 promoter, and a *yEGFP3* gene under the control of TEF1 promoter (Figure 2.19) described previously (Liang et al., 2012). pL203 was constructed by replacing the TEF1 promoter upstream *yEGFP3* with an ADH1 promoter. ADH1 promoter was amplified from plasmid pJZ590d, a gift from Wendell Lim (Department of Biochemistry and Biophysics, University of California, San Francisco), and cloned at unique ClaI and BamHI sites upstream of *yEGFP3* using appropriate restriction endonuclease and ligation-mediated cloning. DNA fragments encoding the ribozyme-based devices were synthesized as 80nt overlapping oligonucleotides, amplified by polymerase chain reaction, and inserted into either pCS1748 or pL203 via the unique restriction sites AvrII and XhoI, which are located 3 nucleotides downstream of the *yEGFP3* stop codon.

Plasmids for characterization of switch activity in mammalian cells were based on pCS2595, a derivative of pcDNA5/FRT (Invitrogen). MS2 was cloned behind a modified CMV promoter containing two 3′ *tetO* binding sites via the unique restriction sites NotI and ApaI. *tetO* sites allow for constitutive suppression of MS2 expression in the host HEK 293 T-Rex line (Invitrogen, Carlsbad, CA), which can be alleviated by doxycycline induction. Blue fluorescent protein (BFP) was cloned downstream of promoter EF1a via unique sites BglII and AvrII containing a 5′ Kozak sequence (CGCCCACC). Riboswitch constructs were cloned immediately downstream of BFP via the unique restriction sites AvrII and AscI. MS2 and BFP sequences used are listed in Figure 2.23.

2.4.2 Characterization of ribozyme switches in yeast

Individual plasmids were transformed into yeast strain CSY22, a *gal2Δ* mutant of W303α (*MATα; leu2-3,112; trp1-1; can1-100; ura3-1; ade2-1; his3-11,15*), described earlier (Hawkins and Smolke, 2006) using standard lithium-acetate methods (Gietz and Woods, 2002). Yeast were grown in YPD or appropriate dropout media (BD, Franklin Lakes, NJ) with 2% w/v glucose. Yeast strains were inoculated in 500 μl appropriate liquid drop out media in 96-well plates and grown in a Lab-Therm HT-X (Kühner, Basel, Switzerland) at 480 rpm, 30°C, and 80% relative humidity overnight. The next day, samples were back-diluted to OD_{600} ~0.1, grown 3–6 hours in the same conditions, and analyzed for fluorescence

2.4.3 Characterization of ribozyme switches in mammalian cells

Individual plasmids were transfected into HEK 293 T-Rex line (Invitrogen, Carlsbad, CA) using FuGENE HD (Promega, Fitchburg, WI) approximately 24

hours after seeding cells on 24-well plates following manufacturer's instructions. Cells were grown in DMEM supplemented with 10% Fetal Bovine Serum (Gibco, Carlsbad, CA). Immediately following transfection, 1 mg/L doxycycline was added to wells to relieve repression of MS2. Approximately 48 hours after transfection, cells were dislodged from adherence to the 24-well plates using standard trypsin methods, and analyzed for fluorescence.

2.4.4 Flow cytometry

Fluorescence was measured using a MACSQuant VYB (Miltenyi Biotec, Cologne, Germany). For GFP characterization, cells were excited with a 488 nm laser and signal measured after passing through a 525/50 nm filter. For BFP characterization, cells were excited with a 405 nm laser and signal measured after passing through a 450/50 nm filter. Viable cells were gated by electronic volume and side scatter. Approximately 10,000 cells were analyzed for each culture, and the arithmetic mean fluorescence of replicate or triplicate biological samples reported. Generally, mean GFP or BFP measurements were normalized by dividing by the mean value of a non-cleaving, non-aptamer control ribozyme, Ctl (Figure 2.22).

2.4.5 SPR-based ribozyme binding assays

SPR experiments were performed on a Biacore X100 (Biacore, Uppsala, Sweden) on a CM5 sensor chip (Biacore) modified with a DNA activator strand (5′-AAACAACTTTGTTTGTTTCCCCC-/AmMO/) as described previously (Kennedy et al., 2012). Full-length RNA was prepared as previously described for the cis-blocking strategy without the addition of the radiolabeled nucleotide. The

Biacore X100 instrument was equilibrated with the physiologically relevant reaction buffer at 37°C unless otherwise specified before all ribozyme binding assays. The SPR baseline was stabilized by performing 2–5 startup cycles, where each cycle includes a capture and a regeneration step. The capture step was performed by an injection of a total of 10–25 ng transcribed cis-blocked RNA diluted in HBS EP+ Buffer (Biacore) over the reaction flow cell (FC2) for 1 min at a flow rate of 10 µl/min. The capture step typically yielded ~50–300 RU of the SPR signal for the described constructs. The regeneration step was performed by an injection of 25 mM NaOH over both flow cells for 30 sec at a flow rate of 30 µl/min.

Following the startup cycles, assay cycles were performed. Each assay cycle includes a capture, a reaction and a regeneration step. The capture and regeneration steps in an assay cycle were performed as described for those in the startup cycle. The reaction step was performed by an injection of the running buffer containing 500 µM $MgCl_2$ over both FCs for 5 min at a flow rate of 10 µl/min. When noted, appropriate concentrations of purified mBP-MS2 (gift of Rachel Green, Department of Chemical Engineering, Johns Hopkins University), were added to FC1. Biacore sensorgram processing and analysis were performed using Biacore X100 Evaluation Software v2.0 (Biacore). The processed sensorgram (R) was fit to a simple exponential equation $R = (R_0 – R_\infty) \times (e^{-kdt}) + R_\infty$, where R_0 (fit globally for a given replicate) is the initial SPR signal before the cleavage reaction, R_∞ (fit locally for a given replicate) is the residual response at the end of the cleavage reaction and kd is the first-order RNA dissociation rate constant.

2.4.6 Gel-based ribozyme cleavage assays

Generation of radiolabeled, full-length MS2-responsive RNA devices was carried out as previously described for natural hammerhead ribozymes and theophylline-responsive RNA devices (Kennedy et al., 2012). Cleavage assays to determine cleavage kinetics were also adapted from the previous work. Briefly, gel-based ribozyme cleavage assays were performed in a physiologically relevant reaction buffer (40 μl) composed of 500 μM $MgCl_2$, 150 mM NaCl, 1 mM DTT and 10 mM HEPES (pH 7.4) at 37°C. In the reaction volume, 200–400 nM of radiolabeled, full-length RNA generated from the cis-blocking strategy was first incubated with 2.5 μM DNA activator strand (5′-AAACAACTTTGTTTGTTTCCCCC), for 2 min to activate the blocked RNA and specified amount of mBP-MS2 protein (Courtesy Rachel Green, Department of Chemical Engineering, Johns Hopkins University). The zero time-point aliquot was taken before initiating the self-cleavage reaction with the addition of $MgCl_2$. Reactions were quenched at specified time points with addition of 3 volumes of RNA stop/load buffer (95% formamide, 30 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol) on ice. Samples were size-fractionated on a denaturing (8.3 M Urea) 8% polyacrylamide gel at 55 W for 35–45 min. Gels were exposed overnight on a phosphor screen and imaged on a FX Molecular Imager (Bio-Rad, Hercules, CA). The relative levels of the full-length transcript and cleaved products were determined by phosphorimaging analysis. To determine k, the first-order rate constant of self-cleavage, the cleaved product fraction at each time point (F_t) was fit to the single exponential equation $F_t = F_0 + (F_\infty - F_0) \times (1 - e^{-kt})$ using Prism 5 (GraphPad, La Jolla, CA), where F_0 and F_∞ are the fractions cleaved

before the start of the reaction and at the reaction endpoint, respectively. All reported cleavage rate constants are the mean of at least three independent experiments.

Author Contributions

The chapter is partly adopted from

d'Espaux, L.D., Kennedy, A.B., Vowles, J.V., Bloom, R.J., and Smolke, C.D. (2013). Development of protein-responsive ribozyme switches in eukaryotic cells. *In preparation.*

LDD designed, performed, and analyzed the *in vivo* yeast experiments and the *in vitro* experiments on direct-coupling riboswitches. ABK designed, performed, and analyzed the *in vitro* experiments on strand-displacement riboswitches. JVV designed, performed, and analyzed the *in vivo* experiments in mammalian cells. RJB, as well as LDD, ABK, and JVV designed ribozyme switches. CDS designed and analyzed experiments.

Supplementary Information

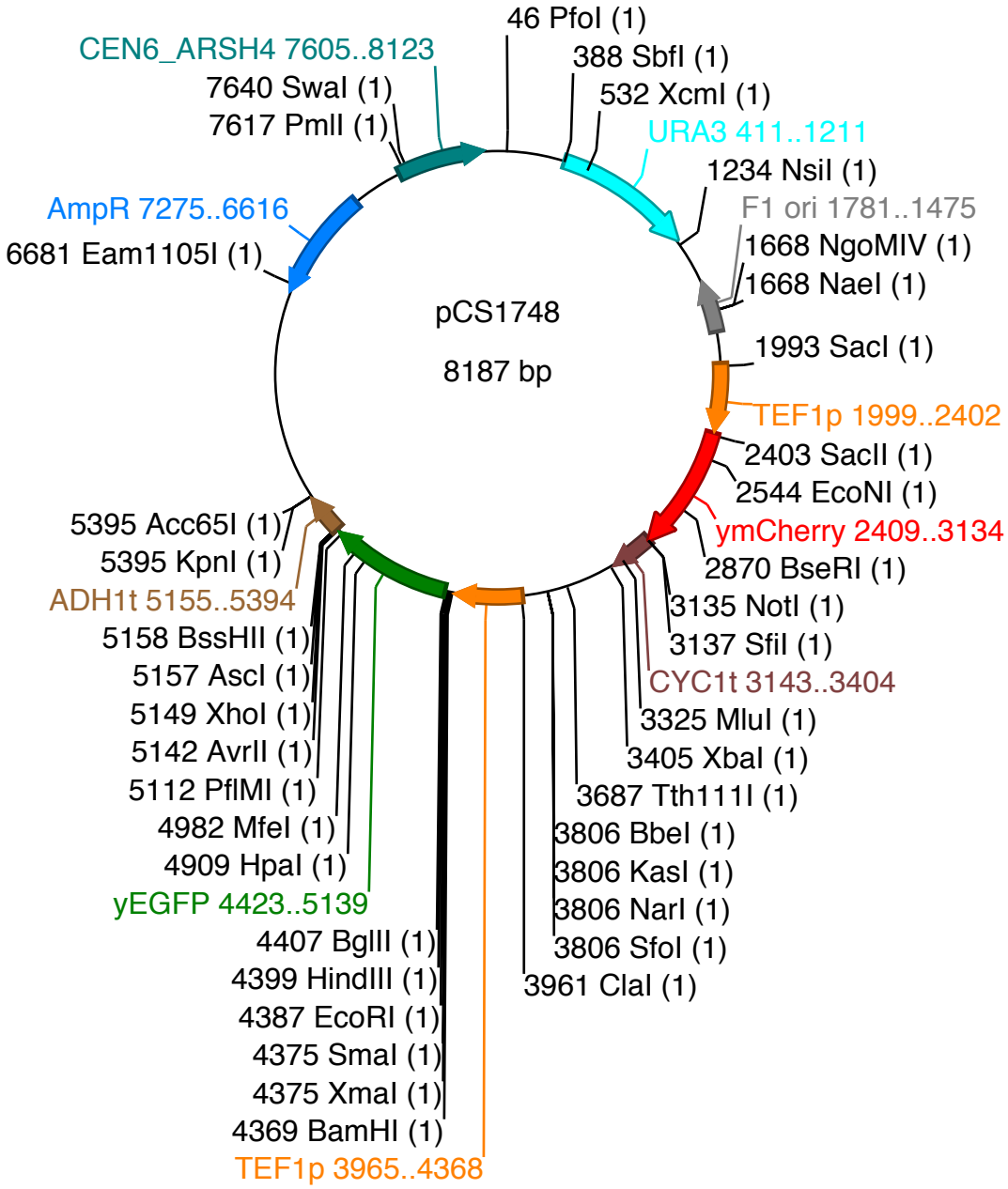


Figure 2.19 Plasmid pCS1748 used for device characterization in yeast

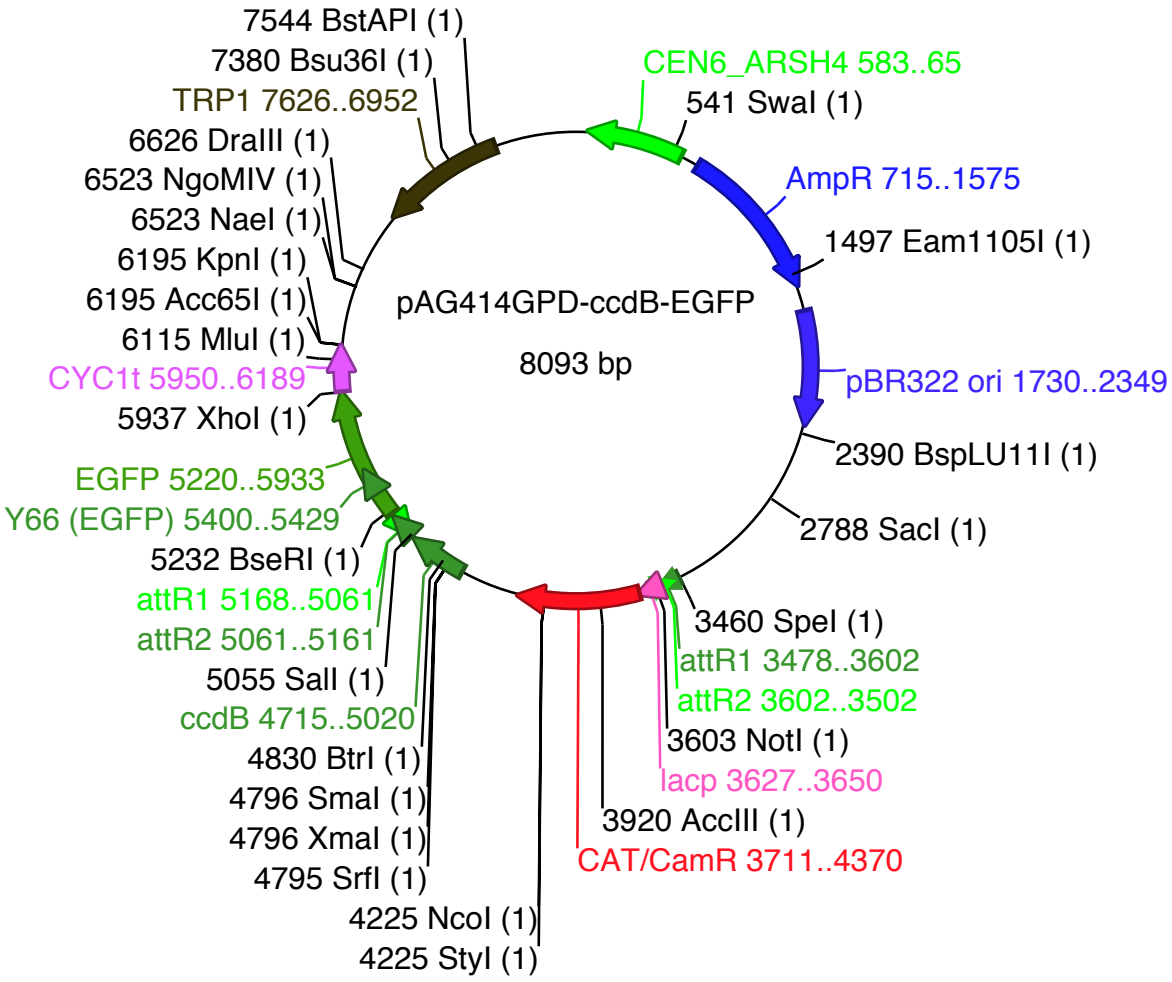


Figure 2.20 Plasmid used for MS2 and p50 expression in yeast

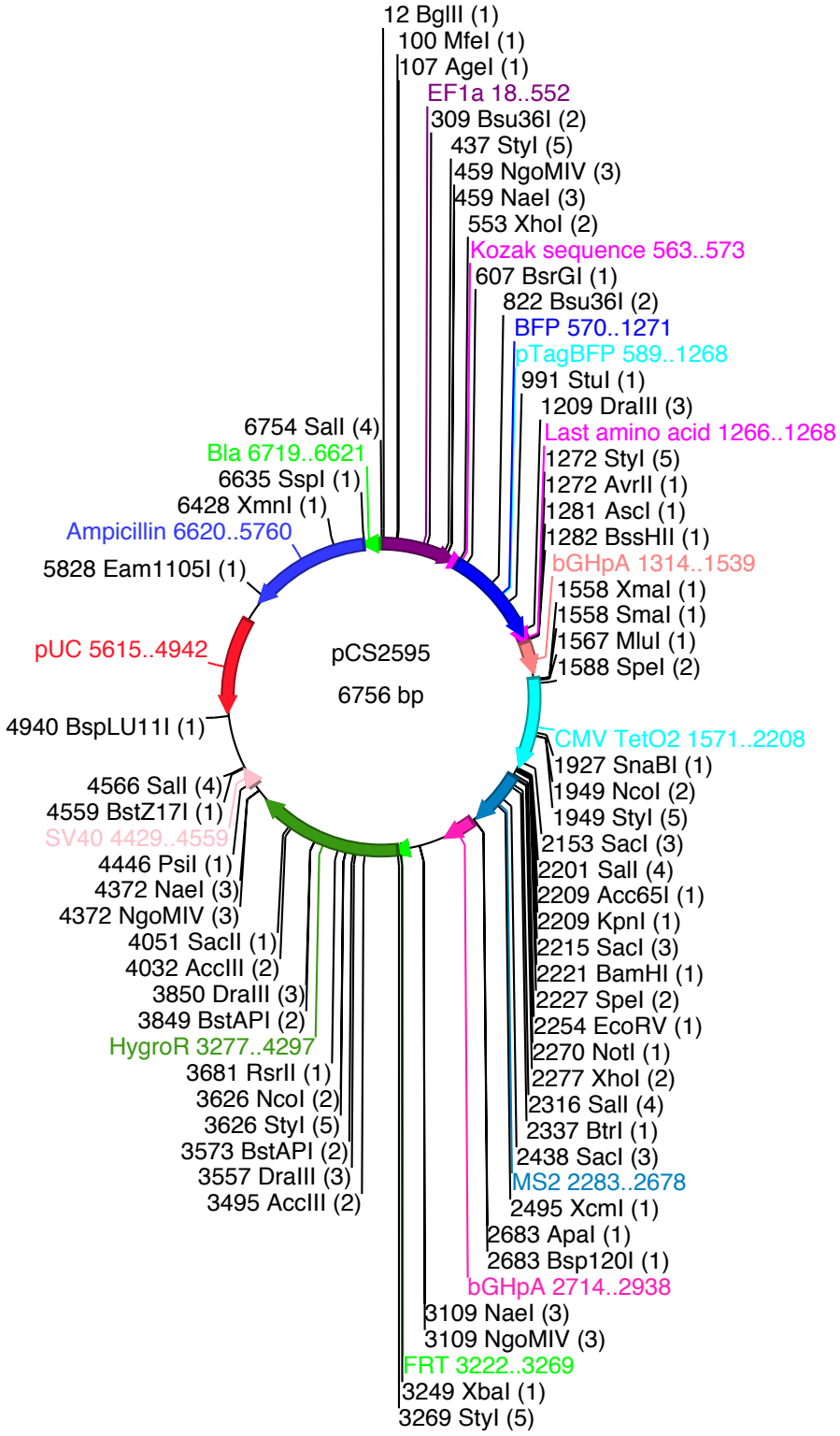


Figure 2.21 Plasmid use for device characterization in mammalian cells

Figure 2.22 Ribozyme switch sequences used in the study

Name in text	Laboratory name	Plasmid	Sequence
Ctl	Ctl	pCS1751	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTACGTGAGGTCCGT GAGGACAGAACAGCAAAAAGAAAAATAAAAA
RSV	RSV	pCS1750	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT GAGGACGAAACAGCAAAAAGAAAAATAAAAA
sdM1	D1	pL401	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT TGTCCAGGATCACCGGACGGGACGGAGGACGAAACAGCAAAAAGAAAAA TAAAAA
sdM2	D2	pL402	AAACAAACAAAGCTGTCACCGGAATCAAGGTCCGGTCTGATGAGTCCGT TGTCCACCATCAGGGGACGGGACGGAGGACGAAACAGCAAAAAGAAAAA TAAAAA
sdM3	D3	pL403	AAACAAACAAAGCTGTCACCGGATGTGCTGCAGGATCACCGCATTTCCG GTCTGATGAGTCCGTGAGGACGAAACAGCAAAAAGAAAAATAAAAA
sdM4	D7	pL404	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT GGTCCACCATCAGGGGACTGGACTGAGGACGAAACAGCAAAAAGAAAAA TAAAAA
dc1M1	dc1M1	pL254	AAACAAACAAAGCTGTCACCGGATGTGCATGAGGATTACCCATTTTCCG GTCTGATGAGTCCGTGAGGACGAAACAGCAAAAAGAAAAATAAAAA
dc1M2	dc1M2	pL259	AAACAAACAAAGCTGTCACCGGATGTGCATGAGGATCACCCATTTTCCG GTCTGATGAGTCCGTGAGGACGAAACAGCAAAAAGAAAAATAAAAA
dc1M3	dc1M3	pL255	AAACAAACAAAGCTGTCACCGGATGTGCCCCAGGATCACCGGGTTTCCG GTCTGATGAGTCCGTGAGGACGAAACAGCAAAAAGAAAAATAAAAA
dc3M1	dc3M1	pL256	AAACAAACAAAATGAGGATGCTGTCACCGGATGTGCTTTCCGGTCTGAT GAGTCCGTGAGGACGAAACAGCCACCCATAAAAAGAAAAATAAAAA
dc3M2	dc3M2	pL257	AAACAAACAAAATGAGGATGCTGTCACCGGATGTGCTTTCCGGTCTGAT GAGTCCGTGAGGACGAAACAGCCACCCATAAAAAGAAAAATAAAAA
dc3M3	dc3M3	pL258	AAACAAACAAACCCAGGATGCTGTCACCGGATGTGCTTTCCGGTCTGAT GAGTCCGTGAGGACGAAACAGCCACCGGGAAAAAGAAAAATAAAAA
dc1K1	dc1K1	pL220	AAACAAACAAAGCTGTCACCGGATGTGCGGATCCTGAAACTGTTTTAAG GTTGGCCGATCCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCAA AAAGAAAAATAAAAA
dc3K1	dc3K1	pL222	AAACAAACAAAGGATCCTGAAACTGTTTCTGTCACCGGATGTGCTTTCC GGTCTGATGAGTCCGTGAGGACGAAACAGTAAGGTTGGCCGATCCAAAA AGAAAAATAAAAA
dc1K2	Kb45	pL250	AAACAAACAAAGCTGTCACCGGAGGATCCTGAAACTGTTTTAAGGTTGG CCGATCCTGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCAA AAAGAAAAATAAAAA
dc1K3	Kb55	pL251	AAACAAACAAAGCTGTCACCGGATGGATCCTGAAACTGTTTTAAGGTTG GCCGATCCGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCAA AAAGAAAAATAAAAA
dc1K4	Kb65	pL252	AAACAAACAAAGCTGTCACCGGATGGGATCCTGAAACTGTTTTAAGGTT GGCCGATCCTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCAA AAAGAAAAATAAAAA
dc1K5	Kb75	pL253	AAACAAACAAAGCTGTCACCGGATGTGCTGGATCCTGAAACTGTTTTAA GGTTGGCCGATCCTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCAA AAAGAAAAATAAAAA

Figure 2.23 Genetic element sequences used in the study

Name in text	Plasmid	Sequence
NES		TTGGCTTTGAAATTGGCTGGTTTGGATATT
NLS		CCAAAAAAGAAGAGAAAGGTC
KZ		TAAAAAAAATG
P50	pL195	ATGGCAGAAGATGATCCATATTTGGGAAGGCCTGAACAAATGTTTCATTTGGA TCCTTCTTTGACTCATACAATATTTAATCCAGAAGTATTTCAACCACAGATGG CACTGCCAACAGATGGCCCATACCTTCAAATATTAGAGCAACCTAAACAGAGA GGATTTCGTTTCCGTTATGTATGTGAAGGCCCATCCCATGGTGGACTACCTGG TGCCTCTAGTGAAAAGAACAAGAAGTCTTACCCTCAGGTCAAAATCTGCAACT ATGTGGGACCAGCAAAGGTTATTGTTCAGTTGGTCACAAATGGAAAAAATATC CACCTGCATGCCCACAGCCTGGTGGGAAAACACTGTGAGGATGGGATCTGCAC TGTAACTGCTGGACCCAAGGACATGGTGGTCGGCTTCGCAAACCTGGGTATAC TTCATGTGACAAAGAAAAAAGTATTTGAAACACTGGAAGCACGAATGACAGAG GCGTGTATAAGGGGCTATAATCCTGGACTCTTGGTGCACCCTGACCTTGCCTA TTTGCAAGCAGAAGGTGGAGGGGACCGGCAGCTGGGAGATCGGGAAAAAGAGC TAATCCGCCAAGCAGCTCTGCAGCAGACCAAGGAGATGGACCTCAGCGTGGTG CGGCTCATGTTTACAGCTTTTCTTCCGGATAGCACTGGCAGCTTCACAAGGCG CCTGGAACCCGTGGTATCAGACGCCATCTATGACAGTAAAGCCCCCAATGCAT CCAACTTGAAAATTGTAAGAATGGACAGGACAGCTGGATGTGTGACTGGAGGG GAGGAAATTTATCTTCTTTGTGACAAAGTTCAGAAAGATGACATCCAGATTCG ATTTTATGAAGAGGAAGAAAATGGTGGAGTCTGGGAAGGATTTGGAGATTTTT CCCCCACAGATGTTCATAGACAATTTGCCATTGTCTTCAAAACTCCAAAGTAT AAAGATATTAATATTACAAAACCAGCCTCTGTGTTTGTCCAGCTTCGGAGGAA ATCTGACTTGGAAACTAGTGAACCAAAACCTTTCCTCTACTATCCTGAAATCA AAGATAAAGAAGAAGTGCAGAGGAAACGTCAGAAGCTCATGCCCAATTTTTCG GATAGTTTCGGCGGTGGTAGTGGTGCCGGAGCTGGAGGCGGAGGCATGTTTGG TAGTGGCGGTGGAGGAGGGGGCACTGGAAGTACAGGTCCAGGGTATAGCTTCC CACACTATGGATTTCCTACTTATGGTGGGATTACTTTCCATCCTGGAACTACT AAATCTAATGCTGGGATGAAGCATGGAACCATGGACACTGAATCTAAAAAGGA CCCTGAAGGTTGTGACAAAAGTGATGACAAAAACACTGTAAACCTCTTTGGGA AAGTTATTTGA
MS2	pCS1592	ATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGA CGTGACTGTCGCCCCAAGCAACTTCGCTAACGGGGTCGCTGAATGGATCAGCT CTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGTTCGTCAGAGCTCT GCGCAGAATCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAACCCA GACTGTTGGTGGTGTAGAGCTTCCTGTAGCCGCATGGCGTTCGTACTTAAATA TGGAACTAACCATTCCAATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTT AAGGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCGGCCATCGC GGCAAACTCCGGCATCTACTAG
ADH1p	pL203	TAAAACAAGAAGAGGGTTGACTACATCACGATGAGGGGGATCGAAGAAATGAT GGTAAATGAAATAGGAAATCAAGGAGCATGAAGGCAAAAGACAAATATAAGGG TCGAACGAAAAATAAAGTGAAAAGTGTTGATATGATGTATTTGGCTTTGCGGC GCCGAAAAAACGAGTTTACGCAATTGCACAATCATGCTGACTCTGTGGCGGAC CCGCGCTCTTGCCGGCCCGGCGATAACGCTGGGCGTGAGGCTGTGCCCGGCGG AGTTTTTTGCGCCTGCATTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATT GGAGAAGCAATAAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGACAAC TGGTGTCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAACATGAGT

		ATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGTGGTGCGA ACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATAACCGCTAGAGTAC TTTGAAGAGGAAACAGCAATAGGGTTGCTACCAGTATAAATAGACAGGTACAT ACAACACTGGAAATGGTTGTCTGTTTGAGTACGCTTTCAATTCATTTGGGTGT GCACTTTATTATGTTACAATATGGAAGGGAACTTTACACTTCTCCTATGCACA TATATTAATTAAAGTCCAATGCTAGTAGAGAAGGGGGGTAACACCCCTCCGCG CTCTTTTCCGATTTTTTTCTAAACCGTGGAATATTTCGGATATCCTTTTGTTG TTTCCGGGTGTACAATATGGACTTCCTCTTTTCTGGCAACCAAACCCATACAT CGGGATTCCTATAATACCTTCGTTGGTCTCCCTAACATGTAGGTGGCGGAGGG GAGATATACAATAGAACAGATACCAGACAAGACATAATGGGCTAAACAAGACT ACACCAATTACACTGCCTCATTGATGGTGGTACATAACGAACTAATACTGTAG CCCTAGACTTGATAGCCATCATCATATCGAAGTTTCACTACCCTTTTTCCATT TGCCATCTATTGAAGTAATAATAGGCGCATGCAACTTCTTTTCTTTTTTTTTC TTTTCTCTCTCCCCCGTTGTTGTCTCACCATATCCGCAATGACAAAAAAATGA TGGAAGACACTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCG TTGTTCCAGAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCCTTCC TTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCCTTCCTTCCA GTTACTTGAATTTGAAATAAAAAAAAGTTTGCTGTCTTGCTATCAAGTATAAA TAGACCTGCAATTATTAATCTTTTGTTTCCTCGTCATTGTTCTCGTTCCCTTT CTTCCTTGTTTCTTTTTCTGCACAATATTTCAAGCTATACCAAGCATACAATC AACTATCTCATATACA
yEGFP3	pCS321	ATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGC TGCCAAGCTATTTAATATCATGCACGAAAAGCAAACAAACTTGTGTGCTTCAT TGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAGCATTAGGTCCC AAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGA GGGCACAGTTAAGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCT TCGAAGACAGAAAATTTGCTGACATTGGTAATACAGTCAAATTGCAGTACTCT GCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGT GGTGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCAGAAGAAGTAACAA AGGAACCTAGAGGCCTTTTGATGTTAGCAGAATTGTCATGCAAGGGCTCCCTA TCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCGACAAAGA TTTTGTTATCGGCTTTATTGCTCAAAGAGACATGGGTGGAAGAGATGAAGGTT ACGATTGGTTGATTATGACACCCGGTGTGGGTTTAGATGACAAGGGAGACGCA TTGGGTCAACAGTATAGAACCGTGGATGATGTGGTCTCTACAGGATCTGACAT TATTATTGTTGGAAGAGGACTATTTGCAAAGGGAAGGGATGCTAAGGTAGAGG GTGAACGTTACAGAAAAGCAGGCTGGGAAGCATATTTGAGAAGATGCGGCCAG CAAAACTAA
mBFP	pCS2595	ATGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCAC CGTGGACAACCATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACG AGGGCACCCAGACCATGAGAATCAAGGTGGTCGAGGGCGGCCCTCTCCCCTTC GCCTTCGACATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCTTCATCAA CCACACCCAGGGCATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCA CATGGGAGAGAGTCACCACATACGAAGACGGGGGCGTGCTGACCGCTACCCAG GACACCAGCCTCCAGGACGGCTGCCTCATCTACAACGTCAAGATCAGAGGGGT GAACTTCACATCCAACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGG CCTTCACCGAGACGCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAACGAC ATGGCCCTGAAGCTCGTGGGCGGGAGCCATCTGATCGCAAACATCAAGACCAC ATATAGATCCAAGAAACCCGCTAAGAACCTCAAGATGCCTGGCGTCTACTATG TGGACTACAGACTGGAAAGAATCAAGGAGGCCAACAACGAGACCTACGTCGAG CAGCACGAGGTGGCAGTGGCCAGATACTGCGACCTCCCTAGCAAACTGGGGCA CAAGCTTAATTGA

Chapter 3

Development of a modular genetic amplifier for near-digital control over eukaryotic gene expression

Abstract

In synthetic biology, engineered gene regulation is often achieved using synthetic transcription factors (TFs). Although allowing for high-fold changes in output gene expression in response to various inputs, it is generally exceedingly difficult to generate TFs that respond to a new input of interest. RNA-based gene control, on the other extreme, has proved versatile and modular, although generally regulatory dynamic ranges of output gene expression have been modest. The limitations of both kinds of regulation hamper efforts in developing more sophisticated synthetic biological systems. Here we develop a modular genetic amplifier for near-digital control over eukaryotic gene expression. We combine ribozyme switch-mediated regulation of a synthetic TF with TF-mediated regulation of an output gene. The amplifier platform allows for as much as 20-fold regulation of output gene expression in response to input signal, with maximal expression approaching the highest levels observed in yeast, yet tunable to intermediate and lower levels. EC_{50} values are more than 4 times lower than in previously best-performing, non-amplifier ribozyme switches. The system design retains the modular-input architecture of the ribozyme switch platform, and the near-digital dynamic ranges of TF-based gene control. This combination of characteristics suggest great potential for the wide applicability of this amplifier platform for more sophisticated, customizable regulation of cellular activity.

3.1 Introduction

Cells have evolved to survive and make efficient use of often scant and changing resources by tailoring gene expression to environmental conditions. Gene control circuits must often convey input signals—usually changing levels of relevant molecules—into homogeneous phenotypic outputs—usually by altering expression of phenotype-setting genes. This must be achieved even when circuit component levels can be "noisy" due to inherent properties of biological processes (Elowitz et al., 2002; Munsky et al., 2012). A commonly evolved strategy has been to institute near-digital control over expression of phenotype-setting genes by amplification. In this manner, an input signal causes an output gene to be unambiguously turned "ON" or "OFF," allowing a biological system architecture to elicit homogenous changes in phenotype in a population of cells even when input signal and cell-to-cell variability may be significant. For example, whereas yeast preferentially consume glucose as a food source, in conditions of glucose depravation and galactose abundance, complex circuits containing various feedback loops allow for high-fold changes in expression of galactose metabolism genes (Ramsey et al., 2006; Sellick et al., 2008). This amplified high-fold, or "digital," regulation of galactose metabolism genes allows the organism to thrive in the presence of galactose, while not

wasting resources synthesizing unneeded genes when preferred glucose is abundant.

For many components of the galactose system, as with many other natural gene circuits, this digital control over output gene expression is achieved by feedback loops, wherein a transcription factor (TF) can activate (or repress) both the output gene and itself. This allows a small input-mediated change in TF activity to self-amplify for maximal output gene modulation. Examples of both positive and negative feedback amplification abound in natural systems (Bateman, 1998).

Another phenomenon that can allow for amplification, and digital control over gene output, is cooperativity, wherein a small change in effector protein activity amplifies not the level of the effector itself, but further changes in effector activity. For example, the *E. coli Tn*10 tetracycline repressor (tetR) is a homodimer protein that can tightly bind tetracycline operator (*tetO*) DNA sites, and block transcription of neighboring genes. Tetracycline binding to tetR mediates cooperative regulation of tetR–*tetO* binding activity (Reichheld et al., 2009). tetR family proteins have been shown to display cooperative binding to adjacent operator sites by modulating conversion of neighboring target operators from B-form DNA to the under-twisted form that facilitates further tetR binding (Ramos et al., 2005). It is interesting to note that TFs, whose evolved function is often to convey an input signal into a digital response, in many examples consist of multimeric proteins that bind repetitive target sequences, e.g., Lambda repressor cI, tetR, Nuclear Factor-κB, and many other classic examples. Further, there are many examples of cooperative binding between different transcription

factors (Banerjee, 2003; Chang et al., 2006). Nature, it may be argued, has found cooperativity a useful tool for achieving digital control over gene expression.

As synthetic biology continues to mature and promise solutions to many pressing needs, burgeoning potential applications necessitate more digital control over output gene expression. For some applications, this has already been successfully demonstrated. Synthetic TFs, for example, have been widely used in synthetic biology applications to date to achieve digital control over target gene expression. In one such example, tetR was fused to the activation domain of the herpes simplex virus protein 16 (VP16) to generate a tetracycline trans activator (tTA) (Gossen and Bujard, 1992). Binding of the tetR domain to *tetO* sites placed adjacent to a minimal promoter drove expression of a downstream target gene in response to expression of tTA. Addition of tetracycline, which modulates the conformation and activity of tTA, lowered gene expression by up to five orders of magnitude. Similar systems that respond differently to tetracycline, or to several other ligands, have also been developed or adopted from natural architectures. In yeast, one group has developed modular binding domains composed of zinc-finger proteins that allow for orthogonal regulation of different target genes by different TF variants with various regulatory dynamics, while still retaining large dynamic ranges on the order of ~100 fold (Khalil et al., 2012). Synthetic TF-mediated gene control has been applied to metabolic pathways in *E. coli*, sensing levels of key intermediates and accordingly regulating enzyme expression, greatly increasing end product yields (Farmer and Liao, 2000; Zhang et al., 2012). This suggests that employing this strategy in other pathways may similarly result in better system performance. Synthetic TFs have also been

constructed with feedback architectures, leading to gene regulation at impressive dynamic ranges (Nistala et al., 2010). Other TF-based strategies, such as TF-regulated integrases, have allowed for the construction of complex biological circuits with greatly amplified digital outputs (Bonnet et al., 2013).

Impressive as they are, these efforts largely make use of TF elements that respond to specific inputs, and except for rare examples that modify an existing TF sensor element to respond to an input molecule very similar to the original (Collins et al., 2005), it remains very difficult to engineer a TF to respond to a novel input. Another approach to gene regulation has been to use synthetic RNA (Liang et al., 2011). These platforms enjoy potentially greater input modularity owing to the wide availability of RNA sequences that sense various molecules of interest and established methods to develop new sensors (Ellington and Szostak, 1990; Tuerk and Gold, 1990; Ellington, 2004), combined with relative accuracy of RNA structure-function prediction (Mathews et al., 1999; Zuker, 2003; Reuter and Mathews, 2010),

One such platform is the ribozyme switch (Win and Smolke, 2008b), in which small-molecule aptamers integrated into a self-cleaving hammerhead ribozyme placed in the 3′ untranslated region (UTR) of a transgene can allow for allosteric regulation of transgene expression in both yeast and mammalian cells (Win and Smolke, 2007, 2008b; Chen et al., 2010; Wei et al., 2013). The platform has been further developed for complex logical computation (Win and Smolke, 2008a) and high-throughput *in vivo* tuning and *in vitro* characterization (Kennedy et al., 2012; Liang et al., 2012). Because of its modular construction, it can be used to regulate any transgene, and to respond to any input to which a suitable aptamer can be

developed, including small molecules (Win and Smolke, 2008a, 2008b) and proteins (d'Espaux et al., 2013). Unfortunately, as with RNA-based gene control systems in general, dynamic ranges have so far been modest compared to TF-based control, with a maximum of ~10-fold gene expression modulation even after engineered tuning and directed evolution (Liang et al., 2012). For many applications, however, the importance of the modular-input architecture of RNA-based gene control cannot be overstated. It thus has remained an important goal in the field to develop ways of amplifying riboswitch-mediated gene control, empowering applications that necessitate sensing novel inputs.

Here we report the development of modular genetic amplifier platform for near-digital control over gene expression in eukaryotic cells. The amplifier platform combines a synthetic TF with ribozyme switch-mediated allosteric regulation. It exhibits high-dynamic range modulation of output gene expression, up to 20 fold, stemming from cooperative activity of the TF component. The platform's output is amplified over output from a ribozyme switch-only platform, allowing for lower input EC_{50} values and maximal output gene expression in the "ON" state. The amplifier retains the modular-input architecture of the ribozyme switch platform. Further, availability of TF components that can activate or repress with various regulatory dynamics (Witzgall et al., 1994; Lamartina, 2003), and that can bind orthogonal modular output operator sequences with digital responses (Khalil et al., 2012) suggest that the amplifier may generally be employed to sense varied and novel modular inputs to control modular outputs, with various and often digital regulatory profiles.

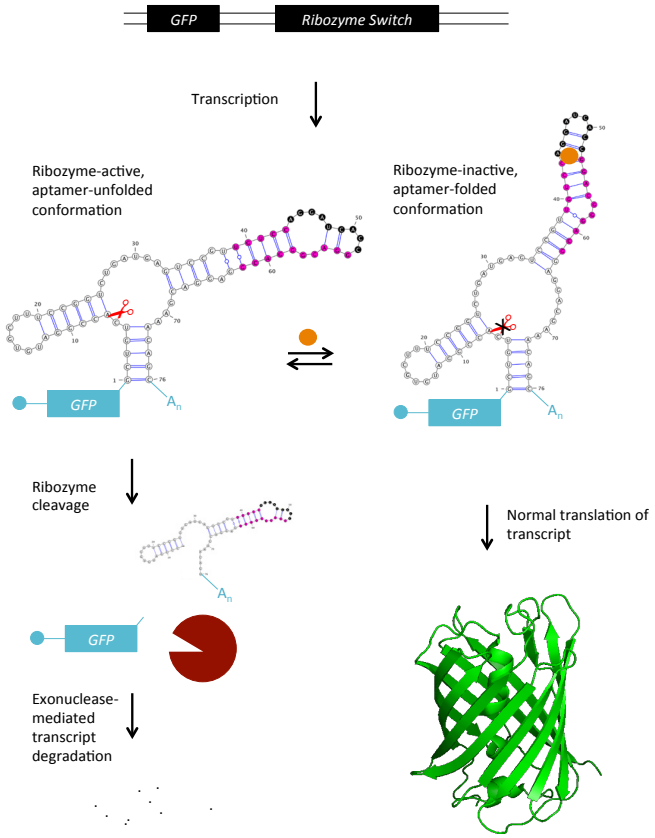


Figure 3.1 Design of ribozyme switch platform

A hammerhead ribozyme is a naturally occurring RNA sequence that can undergo rapid, site-specific self-cleavage (site indicated by red scissors). A synthetic ribozyme switch couples this ribozyme to a ligand-binding aptamer sequence (black bases) through engineered transmitter sequences (pink bases). The switch is encoded in the 3' untranslated region of a target gene, in this case GFP, and upon transcription in eukaryotic cells is designed to exist in two primary conformations. In this example of an "ON" switch, conformations are mutually exclusive in whether they contain a ribozyme-active domain or an aptamer-folded domain. Ribozyme-active conformations undergo cleavage, leading to rapid exonuclease-mediated transcript degradation, and loss of GFP expression. Binding of a ligand (yellow circle) stabilizes the aptamer-folded conformation, which in this case is diminishes cleavage activity, and allows for greater GFP expression. Thus, increased levels of ligand lead to target gene expression being turned "ON." In "OFF" switch designs (not shown), ribozyme-active and aptamer-folded domains are mutually inclusive. In the absence of ligand, transcripts primarily do not undergo cleavage, but do so when ligand binding stabilizes the aptamer-folded and ribozyme-active conformation.

3.2 Results and Discussion

3.2.1 Design of a modular genetic amplifier

We developed a genetic amplifier platform composed of two genetic cassettes: a synthetic transcription factor (TF) whose levels are regulated by a ribozyme switch, and a target gene whose levels are regulated by a TF-responsive promoter. For our TF, we chose the well-characterized tetracycline trans activator (tTA), which consists of the *E. coli* tetracycline repressor (tetR) fused to the activation domain of the herpes simplex virus protein 16 (VP16) (Gossen and Bujard, 1992). We placed a ligand-responsive riboswitch in the 3' UTR of the tTA gene to allow for ribozyme switch-mediated modulation of tTA expression in response to ligand levels. For our target, we cloned a reporter GFP gene under the control of tTA-inducible synthetic promoters, which consist of tetracycline operator (*tetO*) sequences adjacent to a minimal promoter containing a "TATA" box and a leader sequence from the yeast CYC1 promoter. The minimal promoter does not yield appreciable GFP expression alone. However, binding of tTA at *tetO* sites mediate recruitment of other TFs to induce expression of the target downstream gene (Garí et al., 1997) (Figure 3.2).

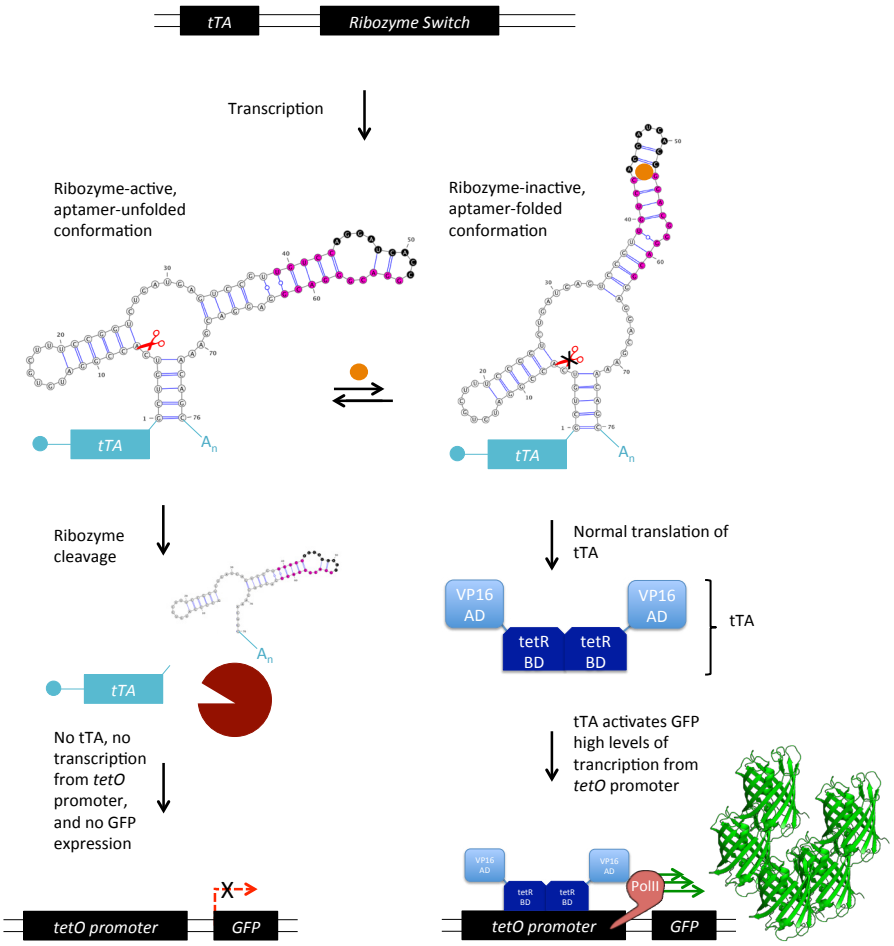


Figure 3.2 Design of modular genetic amplifier platform

The amplifier platform consists of the ribozyme switch platform layered with transcription factor (TF)-mediated regulation of the output gene. Ribozyme switch-mediated, ligand-dependent regulation occurs as in the normal ribozyme switch architecture, but in the amplifier platform is employed to regulate expression of TF tTA. In this "ON" switch design, increased levels of ligand modulate ribozyme cleavage activity to increase levels of tTA. The amplifier architecture adds another layer of regulation, in that tTA regulates expression of a reporter GFP gene by transcriptional control. The tTA synthetic TF is composed of a transcriptional activator VP16 domain (AD) fused to a tetR DNA binding domain (BD). In the absence of tTA, the tetO promoter does not allow for appreciable GFP expression. Increased levels of tTA lead to BD binding at *tetOp*, recruiting other transcription factors, and leading to high levels of GFP. Because of the tTA–*tetO* binding dynamics, the amplifier architecture allows for an amplified GFP output in response to addition of ligand over the ribozyme switch-only architecture, while retaining the modular-input characteristics of the latter.

3.2.2 Input–output profiles of tTA expression–tetO target gene levels

We first sought to examine the properties of GFP activation by tTA. We cloned two, four, or seven copies of *tetO* upstream the minimal CYC1 promoter driving GFP expression and chromosomally integrated these target constructs into yeast at the *trp1* locus (*trp::tetO$_2$-GFP*, *trp::tetO$_4$-GFP*, *trp::tetO$_7$-GFP*, respectively). We then transformed into these strains plasmids bearing tTA behind a GAL1 promoter (*GAL1p*) and titrated galactose to examine the relationship between tTA expression and output GFP levels allowed by our design architecture. To allow for tight control o tTA expression levels, the assays were performed in a parent strain (CSY22) in which the galactose permease *GAL2* was knocked out to allow for linear induction of gene expression from *GAL1p* (Hawkins and Smolke, 2006). A control plasmid containing a *GAL1p-GFP* expression cassette was transformed into the *trp::tetO7-GFP* strain, showing that *GAL1p* induction by galactose was indeed linear and relatively weak. In contrast, *GAL1p* induction of tTA, expected to exhibit a similar expression profile as the *GAL1p-GFP* construct, led to dramatically increased GFP levels from all *tetO*-containing promoters. The amplifier architecture allowed for saturating output GFP levels even at low induction, with EC_{50} values ~ 0.04% galactose, whereas *GAL1p-GFP* did not reach maximal expression in the galactose range tested (Figure 3.3A). The data also show that maximal expression increased with *tetO* copy number, allowing for tuning of output level, and compared favorably with even the highest-strength promoters commonly used in yeast (Figure 3.3C).

Importantly, the data indicate that GFP expression levels rose logarithmically with tTA induction. As discussed previously, prior work has shown that tetR family proteins can display cooperative binding to adjacent operator sites by modulating conversion of neighboring target operators from B-form DNA to the under-twisted form that facilitates further tetR binding (Ramos et al., 2005). We used GFP expression from *GAL1p* as a measure of tTA induction from the same promoter, and $tetO_7$-*GFP* activation as a measure of output, to calculate a hill coefficient for tTA–$tetO_7$ binding of $n_{H,tTA}$ ~2.47 (for comparison, the classic example of hemoglobin–O_2 binding displays $n_{H,\ Hg}$~2.7 in adult humans (Pittman, 2011)), suggesting that our system is cooperative (Figure 3.3B). The logarithmically increasing GFP output levels from linearly increasing tTA input levels lend support to the hypothesis that in the regime of low tTA expression, a small increase in tTA could lead to a dramatic increase in target GFP levels. We hoped that such a situation could allow a ligand-responsive switch effecting a relatively modest change in tTA expression to dramatically increase GFP output levels.

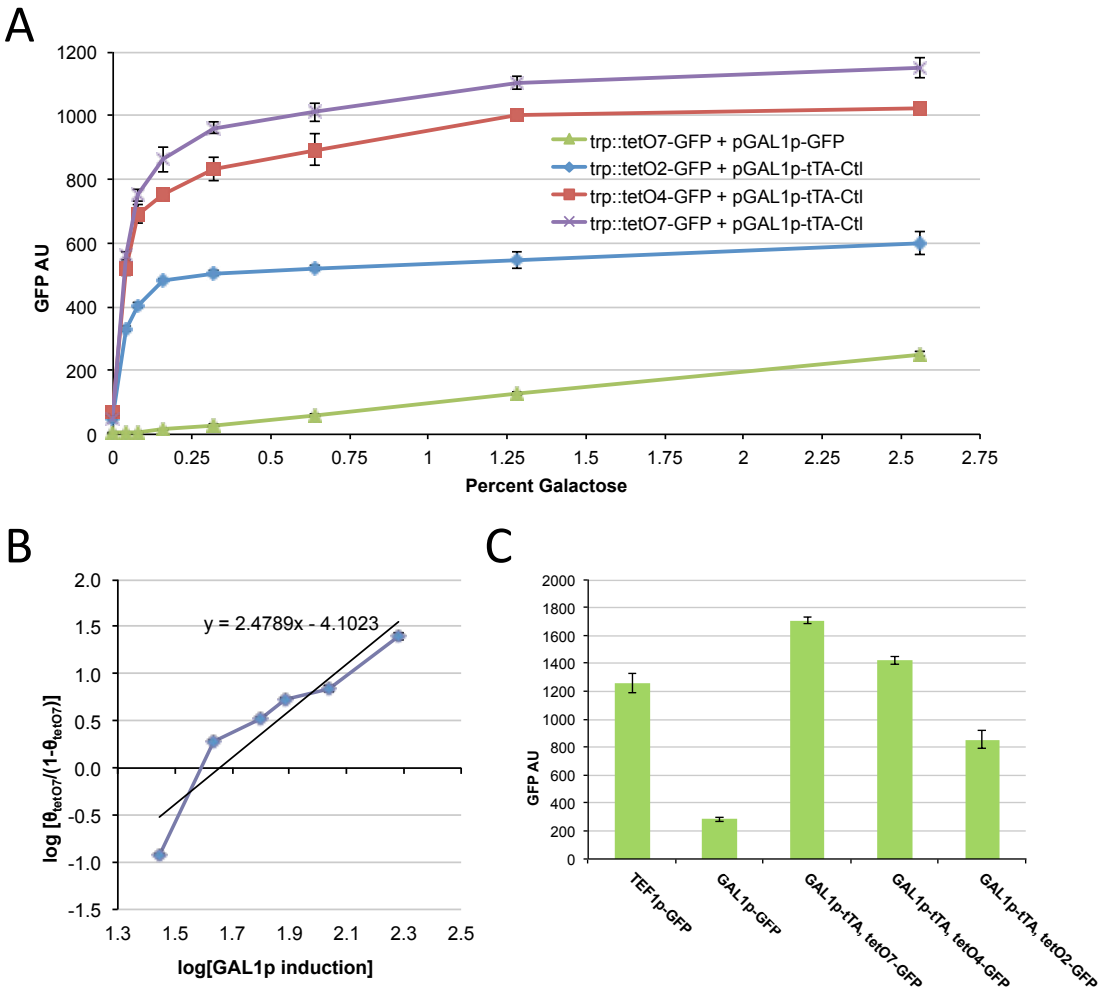


Figure 3.3 Profiles of GFP target activation following linear induction of tTA

Linearly increasing levels of tTA in the amplifier architecture mediate robust and logarithmically increasing levels of output GFP expression. **(A)** Yeast strains contain target cassettes composed of tTA-responsive promoters *tetO*$_{-2,-4,\ and\ -7}$ driving GFP expression, and a separate actuator cassette composed of galactose-responsive *GAL1p* driving expression of the transcription factor tTA. Galactose induction of *GAL1p-GFP* (green) elicits linearly increasing and low-level expression of GFP, whereas galactose induction *GAL1p-tTA* results in robust and digital expression of GFP output from the various *tetO*-containing promoters (purple, red, blue). **(B)** Hill plot of tTA induction-GFP output performance from tTA-*tetO*$_7$-GFP indicate cooperative binding, with n_H~2.47. **(C)** Maximal allowed GFP expression from the amplifier designs compares favorably to expression from the constitutive high-strength yeast promoter *TEF1p*, and can be tuned by *tetO* copy number. All samples are at 2% galactose induction. Error bars indicate one standard deviation from mean GFP levels.

3.2.3 Ribozyme switch-mediated regulation of tTA for target expression control

We next examined the ability of tTA regulation by small molecule-responsive ribozyme switches to increase the regulatory range of these RNA-based control elements. Several theophylline-responsive ribozyme switches previously developed in our laboratory (Win and Smolke, 2007) were placed in the 3′ UTR of the tTA gene driven by a constitutively-active TEF1 promoter (*TEF1p*). All *TEF1p-tTA-riboswitch* constructs led to near-saturating levels of GFP expression with or without theophylline, with only wild type ribozyme (RSV) abolishing GFP expression (Figure 3.4).

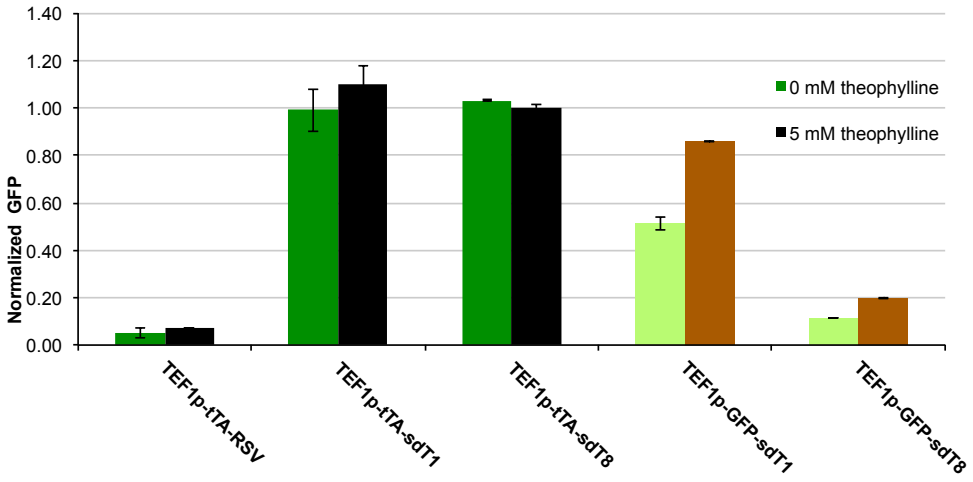


Figure 3.4 Switches in the amplifier platform driven by high-strength promoter TEF1

For ribozyme switches in the amplifier platform behind high-strength promoter *TEF1p*, amplifier platform fails to confer ligand-mediated regulation. Theophylline-responsive switches sdT1 and sdT8 placed in the no-amplifier *GFP-switch*-only platform (light green and brown) regulate GFP expression in response to increased theophylline levels. The same switches behind *TEF1p-tTA* show saturating levels of GFP expression under all conditions (dark green, black). RSV is a wild type ribozyme. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control cloned behind *TEF1p-tTA*. Error bars indicate one standard deviation from the mean. All experiments were performed in a yeast strain containing *trp1::tetO$_7$-GFP*.

In the switch-only, no-amplifier architecture (containing only *TEF1p-GFP-switch*) basal expression levels from these switches were all greater than 10% of that from a non-cleaving control ribozyme (Win and Smolke, 2007). These expression levels would be similar to induction of *GAL1p* at more than 1% galactose induction. At these expression levels, tTA already begins to yield maximal GFP expression (Figure 3.3A), explaining why any possible ligand-induced increase in tTA protein levels would not lead to a significant increase in GFP expression. These results suggest that perhaps lowering overall tTA expression might lead to the desired high-fold modulation of target GFP levels. To achieve that, we tested various low-strength promoters driving a reporter GFP and based on the results (Figure 3.5) redesigned our tTA-switch constructs to be driven by *STE5p*, which drives expression at just 1% of that from *TEF1p*.

We cloned *STE5p* upstream of our various tTA-switch constructs and tested whether these low-expression plasmids can modulate GFP levels in response to theophylline. A non-cleaving ribozyme control (Ctl) resulted in GFP expression as high as from *TEF1p-tTA-Ctl*, even though tTA levels would be expected to be just 1% of the latter. Although we observed some theophylline-responsive control over GFP expression by some of our tTA-switch constructs, dynamic ranges were lower than in the no-amplifier *GFP-switch*-only context. These results can again be explained by our tTA input–output profile (Figure 3.3A). Since *STE5p-tTA-Ctl* already yields close to maximal GFP expression, riboswitches with basal expression of 10% of Ctl or higher would be expected to yield GFP levels at ~60% or higher in the amplifier architecture, close to the

values we observe. Thus, the data indicate that we need to lower tTA levels even further to achieve high-fold modulation of GFP expression.

We next obtained second-generation, low-basal level, theophylline-responsive ribozyme switches developed in our laboratory using a high-throughput screening method (Liang et al., 2012). These newly-developed switches have been shown to allow for basal levels as low as ~3% of that from Ctl, and dynamic ranges of up to 10 fold in response to theophylline. To explore a wide array of basal states, we cloned one, two, and three copies of these second-generation switches (sdT8a and sdT8t) behind *STE5-tTA*. Additional ribozyme switch copies are expected to increase the likelihood of any given transcript being cleaved, leading to lower overall gene expression levels (Chen et al., 2010). In our experiments, the second-generation switches led to dramatic modulation of GFP expression in response to theophylline (Figure 3.7A). In addition, gene expression "ON" states were very high, similar to GFP levels from plasmids bearing non-cleaving ribozyme controls. These switches regulating tTA in the amplifier architecture exhibited higher-fold GFP modulation and higher "ON" states than the same switches directly regulating GFP (Figure 3.7B). The amplifier switches exhibited dynamic ranges as high as 20 fold, with maximal GFP expression approaching the highest levels commonly achieved in yeast. GFP expression in the absence of theophylline was almost indistinguishable from output from wild type ribozyme. In addition, GFP expression profiles were unimodal, with populations of cells in the presence and absence of theophylline showing little overlap (Figure 3.7B).

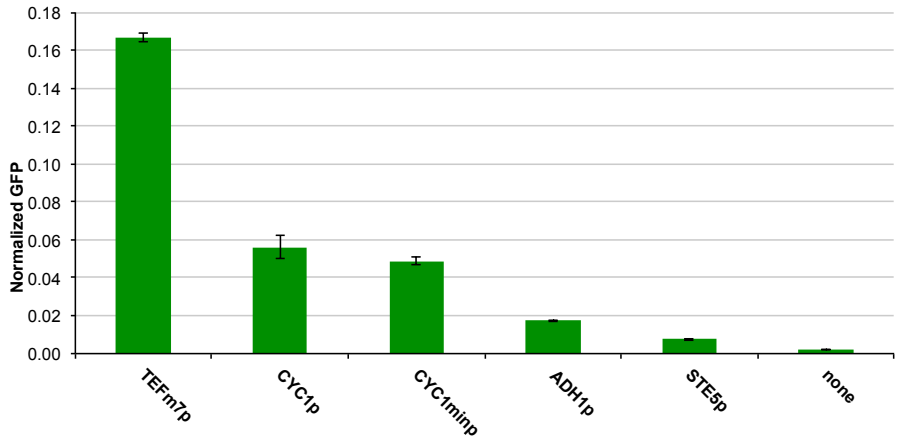


Figure 3.5 Expression levels of various promoters driving GFP

Various promoters upstream of *GFP* show different expression levels, shown here normalized to expression from *TEF1p-GFP*. Cells harboring no GFP gene are labeled as "none."

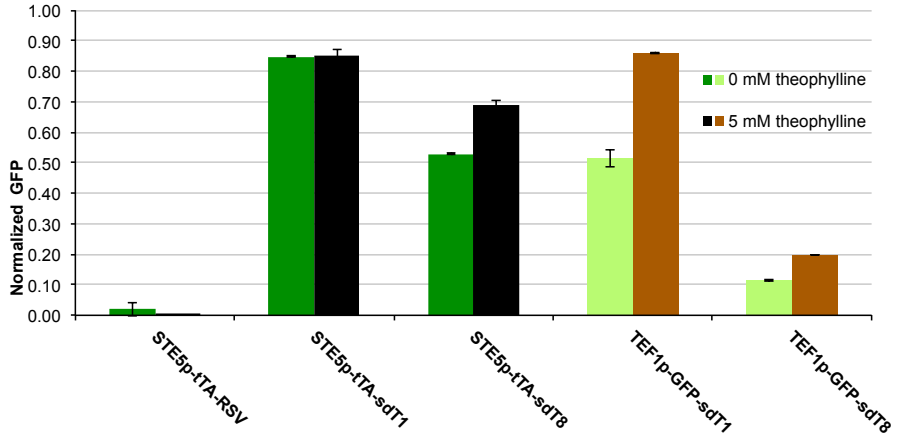


Figure 3.6 Switches in the amplifier platform driven by low-strength promoter STE5

For ribozyme switches in the amplifier platform behind low-strength promoter *STE5p*, amplifier platform confers modest ligand-mediated regulation. Theophylline-responsive switches sdT1 and sdT8 placed in the no-amplifier *GFP-switch*-only platform (light green and brown) regulate GFP expression in response to increased theophylline levels, but show modest modulation behind *STE5p-tTA* (dark green, black). RSV is a wild type ribozyme. Mean GFP levels are shown normalized to a non-aptamer, non-cleaving control cloned behind *STE5p-tTA*. Error bars indicate one standard deviation from the mean. All experiments were performed in a yeast strain containing *trp1::tetO$_7$-GFP*. Expression from *STE5p* is ~1% of that from *TEF1p*.

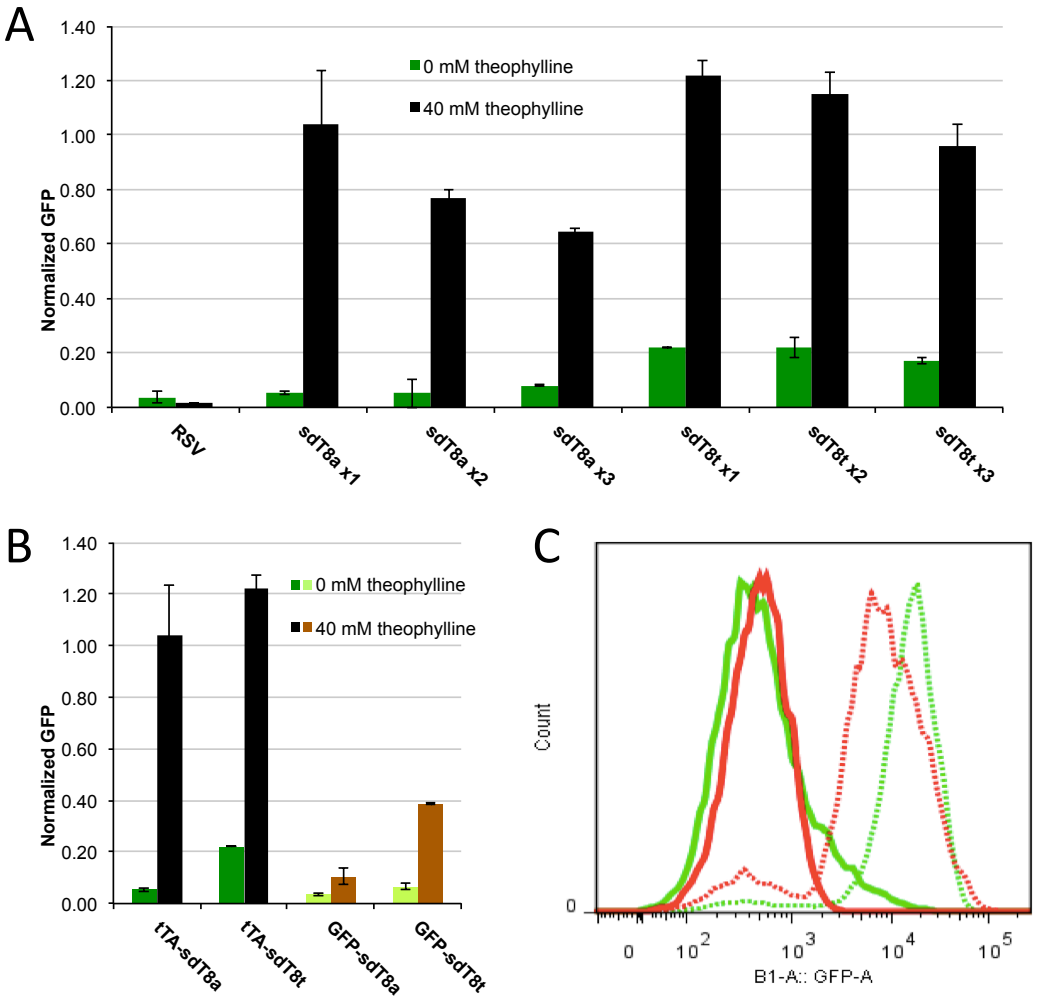


Figure 3.7 Low-basal level switches allow high dynamic ranges and "ON" states

(A) One, two, and three copies of theophylline-responsive riboswitches sdT8a and sdT8t in the 3' UTR of tTA are able to significantly modulate GFP expression in response to theophylline addition. **(B)** The same riboswitches in the 3' UTR of GFP exhibit much lower-fold dynamic range than behind tTA, and also much lower GFP "ON" states. **(C)** Histograms show unimodal distributions with clear separation between GFP "OFF" and "ON" states. Red, *TEF1p-GFP-sdT8a*, Green, *STE5p-tTA-sdT8a*; solid, 0 mM theophylline, dotted, 40 mM theophylline. All plasmids are expressed in a CSY22 *trp1::tetO$_7$-GFP* strain. GFP values are normalized to the average from a plasmid bearing *STE5p-tTA-Ctl*. Maximal expression from *STE5p-tTA-Ctl* is similar than from *TEF1p-GFP-Ctl*. All experiments were performed in a yeast strain containing *trp1::tetO$_7$-GFP*. All *GFP-switch* constructs are expressed from *TEF1p*, all *tTA-switch* constructs from *STE5p*. Expression from *STE5p* is ~1% of that from *TEF1p*.

We then explored the regulatory dynamics of ligand-mediated gene expression. We titrated the theophylline inducer for switches controlling GFP expression directly or through the amplifier architecture. Results showed increased sensitivity to theophylline addition, with amplifier switches having IC50 values ~2–4 mM theophylline, versus >10 mM for direct GFP control switches (Figure 3.8). Further, for tTA-sdT8t, near-saturating GFP levels were observed at as low as 10 mM theophylline, whereas the same switch in the direct-GFP architecture did not appear to reach saturation.

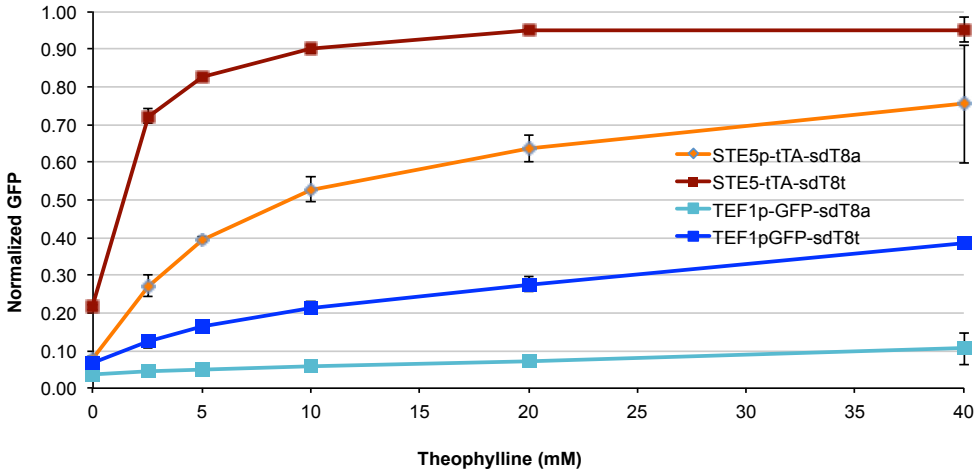


Figure 3.8 Regulatory dynamics of theophylline-induced riboswitch control

Ligand-induced modulation of ribozyme switch activity is conveyed into an amplified response in the amplifier platform. Switches sdT8a and sdT8t placed behind *TEF1p-GFP* mediate higher GFP expression upon theophylline induction, although maximal expression levels and dynamic range are modest. The same switches in the amplifier context behind *STE5p-tTA* mediate rapid saturation of GFP output even at low theophylline concentration. Expression from *STE5p* is ~1% of that from *TEF1p*.

3.3 Conclusions

We set out to develop a mechanism to amplify output gene regulation by ribozyme switches. Although the ribozyme switch platform exhibits a highly modular design, as with other classes of synthetic riboswitches, its output dynamic ranges have so far been modest, at most ~10 fold, and often in response to induction by high concentrations of input molecules (Liang et al., 2011). As discussed earlier in Chapter 2, this limitation may be due to the thermodynamic nature of their switching mechanism, making it difficult to exceed this dynamic range. TF-based systems, while largely exhibiting modular and high-fold regulation of output gene expression (Khalil et al., 2012), exhibit greater limitations in terms of their adaptation to different input molecules, limiting their widespread use for synthetic biology.

This work resolves this problem by developing a TF-riboswitch gene amplifier in eukaryotes combining the modular-input characteristics of ribozyme switch-mediated regulation with the high-fold, modular output characteristics of TF-mediated gene control. The amplifier platform allows for as much as 20-fold regulation of output gene expression in response to input signal. EC_{50} values are more than 4 times lower than in our best-performing non-amplifier ribozyme switches, and maximal "ON" state gene expression approaches the highest expression levels commonly observed in yeast, whereas "OFF" states approach those from a wild type, maximally-cleaving ribozyme.

The amplifier platform architecture allows for modularity of input stemming from the versatility of the riboswitch platform. The theophylline-responsive ribozyme switches used here can be replaced with riboswitches to novel ligands,

including already-developed protein-responsive ribozyme switches described in Chapter 2, or with more complex computational logic gates as previously described (Win and Smolke, 2008a). Other classes of riboswitches can also be potentially employed to regulate tTA levels (Babiskin and Smolke, 2011b; Liang et al., 2011). Further, with an ever-expanding set of RNA aptamers (Ellington, 2004), and both rational design (Win and Smolke, 2007, 2008b) and directed evolution methods (Liang et al., 2012) to generate new riboswitches, there do not appear to be any foreseen limitations regarding choice of input molecule. While in this study we needed to tune promoter strength and riboswitch basal state to achieve high-fold modulation of output gene expression, there are myriad tools with which to tune expression levels (Alper et al., 2005; Babiskin and Smolke, 2011a). The riboswitches we used also did not originally exhibit exceedingly large dynamic ranges, being similar to other riboswitch classes previously developed, indicating that the amplifier platform can likely be employed successfully with new inputs. Immediate future work will be aimed at demonstrating this explicitly.

In addition to riboswitch-mediated, ligand-responsive regulation, TFs that respond to various ligands have been widely used. For example, tTA is known to lose its gene activation characteristics upon binding doxycycline. Similar transcription factors bind other inducer molecules and modulate output gene expression accordingly. While alone these TF-altering inducers may comprise a small and not generally expansible set of inputs, layered in combination with riboswitch-mediated control, they can provide for more sophisticated regulation over gene expression.

The amplifier platform is also designed to be modular with regards to output protein. tTA-responsive promoters can be placed upstream of any transgene, or any endogenous gene by 3′ UTR replacement (Babiskin and Smolke, 2011b). There is no reason to believe that the range of expression regulation of a different output protein would be significantly different than for our reporter GFP.

For some applications, it may be desirable to simultaneously regulate different genes in response to different outputs. Previous work in yeast has demonstrated using modular binding domains fused to various activation domains to allow for orthogonal gene regulation by synthetic TFs (Khalil et al., 2012). The limitation of this platform is lack of input modularity, as these synthetic TFs cannot be generally modified to respond to a new ligand. With the amplifier architecture, however, tTA can be modified by replacing tetR with these orthogonal DNA-binding domains to allow for simultaneous control over many genes in response to any ligand of interest to which a proper riboswitch can be developed.

Another desirable characteristic of a gene control platform is to be able to set output gene expression at a specified level. Here we demonstrated that the number of *tetO* sites upstream our output gene can tune its expression. We show expression levels as high as generally observed in yeast, as well as at intermediate levels from lower *tetO*-copy promoters. It is likely that more step-wise changes in *tetO* copy-number can provide even more precise control over output gene expression. Further, several strategies have shown that mutating promoter sequence can affect output strength (Alper et al., 2005), a strategy which we may also employ. Our platform thus seems to allow precise control

over “ON” state expression level, while still maintaining ability to maintain tight expression in the “OFF” state.

Lastly, while we demonstrate turning gene expression “ON” in response to our ligand, our laboratory has developed several examples of “OFF” switches which we may likely employ to invert the response to ligand induction. Alternatively, synthetic TFs have also been developed that contain silencing domains (Witzgall et al., 1994), potentially allowing “ON” switches to turn gene expression “OFF” in response to increased ligand levels..

While we continue to demonstrate and further develop these characteristics, the amplifier platform appears poised to resolve important limitations in current eukaryotic gene control strategies. As synthetic biology designs continue to evolve in sophistication, we hope that our genetic amplifier may generally be employed to sense varied and novel molecular inputs, to control modular outputs, with orthogonal, tailored, and digital regulatory responses, allowing for the development of powerful solutions to many pressing needs.

3.4 Materials and Methods

3.4.1 Plasmid and strain construction

Standard molecular biology techniques were used for DNA manipulation and cloning (Sambrook, 2001). Restriction enzymes, T4 DNA ligase, and other cloning enzymes were purchased from New England Biolabs (Ipswich, MA). PfuUltraII (Agilent Technologies, Santa Clara, CA) was used for high-fidelity PCR amplification. LR Clonase II and BP Clonase II (Invitrogen, Carlsbad, CA) were used for recombination into Gateway destination plasmids (Addgene,

Cambridge, MA) following manufacturer's instructions. Oligonucleotides were synthesized by Integrated DNA Technologies (Coralville, IA) or the Stanford Protein and Nucleic Acid Facility (Stanford, CA). Plasmids were transformed into One Shot Top 10 chemically competent E. coli (Invitrogen, Carlsbad, CA; F- mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 Δ lacX74 recA1 araD139 Δ(araleu)7697 galU galKrpsL (StrR) endA1 nupG) by standard methods. *E. coli* were grown in LB media (BD, Franklin Lakes, NJ) with 100 μg/mL ampicillin (EMD Chemicals, Gibbstown, NJ) or 50 μg/mL kanamycin (EMD Chemicals, Gibbstown, NJ), depending on the plasmid selection gene. Plasmids were prepped from overnight cultures of *E. coli* using Econospin All-in-One Mini Spin Columns (Epoch Biolabs, Missouri City, TX) according to the manufacturers' instructions. Sequencing was performed by Laragen Inc. (Los Angeles, CA) and Elim Biopharmaceuticals (Hayward, CA).

Plasmids for tetOn-GFP target integration were based on pAG304GAL-ccdB (Addgene, Cambridge, MA), where a stop codon was added 5' of the constructs to prevent any induction from GAL1p, followed by tetOx-CYC1mp-GFP. PCR constructs were cloned into *attL1* and *attR1* sites within pAG304GPD-ccdB-EGFP (Addgene, Cambridge, MA) using BP and LR recombination (Invitrogen, Carlsbad, CA) following manufacturer's instructions. Sequences of relevant promoters and proteins are listed in Figure 3.11.

Plasmids for riboswitch GFP expression characterization were based on pCS1748, a centromeric vector encoding an *mCherry* gene under the control of a TEF1 promoter, and a *yEGFP3* gene under the control of TEF1 promoter (Figure 2.19) described previously (Liang et al., 2012). pL192 was constructed by

replacing the TEF1 promoter upstream *yEGFP3* by a STE5 promoter. STE5 promoter was amplified from plasmid pJZ525, a gift from Wendell Lim (Department of Biochemistry and Biophysics, University of California, San Francisco) using primers Lcla_Ste5 (ATATatcgatATCAAGTTTCCTTTAAAGGG) and Rbam_Ste5* (ATATggatccCATTTAAAAGTTGTTTCCGCTGtat) and cloned at unique ClaI and BamHI sites upstream of *yEGFP3* using appropriate restriction endonuclease and ligation-mediated cloning. DNA fragments encoding the ribozyme-based devices were synthesized as 80nt overlapping oligonucleotides, amplified by polymerase chain reaction, and inserted into either pL192 via the unique restriction sites AvrII and XhoI, which are located 3 nucleotides downstream of the yEGFP3 stop codon. For construction of plasmids bearing multiple copies of a riboswitch, riboswitches were amplified by Rxhoxba_s8X (ATATctcgagtctagaTTTTTATTTTTCTTTTTGCTGTTTCGTCCTCC) and Lavr_s8X (ATATCCTAGGAAACAAACAAAGCTGTCACCGGA), the first copy cloned into the unique restriction sites AvrII and XhoI, and subsequent copies cloned in compatible enzyme scar method, in which further copies of riboswitches were digested with AvrII and XhoI, the vector by XbaI and XhoI, and the two ligated, resulting in a configuration of AvrII site—riboswitch—scar—riboswitch—Xba I site—XhoI site. All switch, protein, and promoter sequences used are listed in Figure 3.10 and Figure 3.11.

Individual plasmids were transformed into yeast strain CSY22, a *gal2Δ* mutant of W303α (*MATα; leu2-3,112; trp1-1; can1-100; ura3-1; ade2-1; his3-11,15*) , described earlier (Hawkins and Smolke, 2006) using standard lithium-acetate

methods (Gietz and Woods, 2002). Yeast were grown in YPD or appropriate dropout media (BD, Franklin Lakes, NJ) with 2% w/v glucose. For galactose induction experiments, yeast were instead grown in appropriate dropout media with 1% w/v sucrose and the appropriate galactose concentration.

3.3.2 Flow cytometry

For fluorescence characterization, yeast strains were inoculated in 500 μl appropriate liquid drop out media in 96-well plates and grown in a Lab-Therm HT-X (Kühner, Basel, Switzerland) at 480 rpm, 30°C, and 80% relative humidity overnight with the appropriate concentration of theophylline (Sigma Aldrich, St. Louis, MO). The next day, samples were back-diluted to OD_{600} ~0.1, grown 3–6 hours in the same conditions, and analyzed for fluorescence using a MACSQuant VYB (Miltenyi Biotec, Cologne, Germany). Cells were excited with a 488 nm laser and signal measured after passing through a 525/50 nm filter. Viable cells were gated by electronic volume and side scatter. Approximately 10,000 cells were analyzed for each culture, and the arithmetic mean fluorescence of replicate biological samples reported. Generally, mean GFP values were normalized by the mean GFP value of a non-cleaving control ribozyme. Theophylline was prepared as a 40 mM stock in appropriate media, and diluted when appropriate.

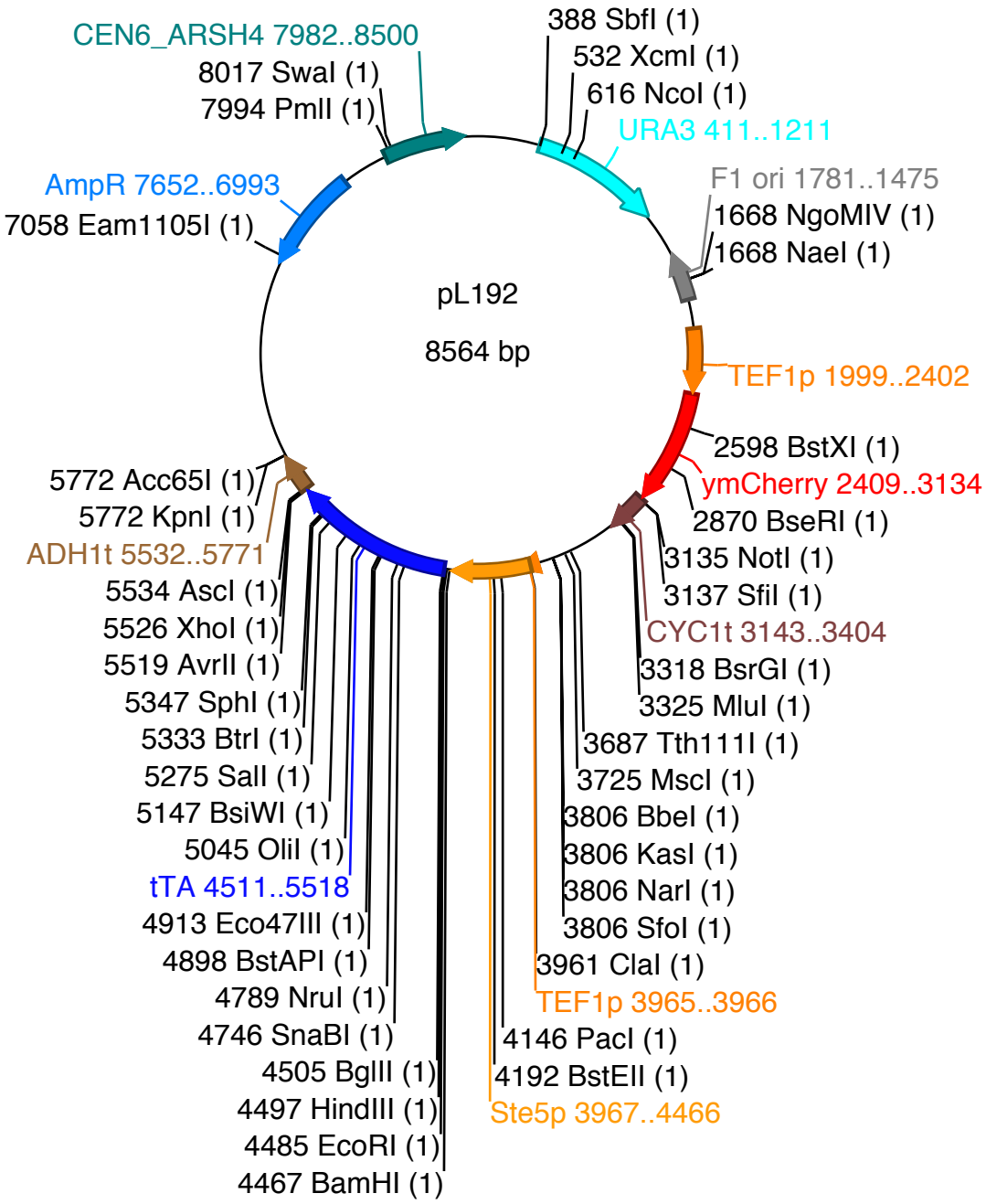


Figure 3.9 Plasmid pL192 for characterization of amplifier output

Name in text	Laboratory name	Plasmid	Sequence
Ctl	sTRSV-Ctl	pCS1751	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTACGTGAGGTCCGTGAGGACAGAACAGCAAAAAGAAAAATAAAAA
RSV	sTRSV	pCS1750	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGCAAAAAGAAAAATAAAAA
sdT8	L2b8	pL121	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTTGTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGGAGGACGAAACAGCAAAAAGAAAAATAAAAA
sdT1	L2b1	pL71	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGCAAAAAGAAAAATAAAAA
sdT8a	L2b8-a1	pCS2261	AAACAAACAAAGCTGTCACCGGAATCAAGGTCCGGTCTGATGAGTCCGTTGTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGGAGGACGAAACAGCAAAAAGAAAAATAAAAA
sdT8t	L2b8-t241	pCS2273	AAACAAACAAAGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTTGCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTTCAGCGGAGGACGAAACAGCAAAAAGAAAAATAAAAA

Figure 3.10 Ribozyme switch device sequences used in experiments

Name in text	Plasmid	Sequence
KZ		TAAAAAAAATG
STE5p	pL192	ATCAAGTTTCCTTTAAAGGGATATATAACAGATTCTAAAACTGACAG AAATATTTCGAGTGAAGAAGAAGCGTTAAATATTGGATCTTTCCGCA GTTCTACTCTGATACATTTTTGAAGTAGGAGAGTCATTTAGAAGGCG TATTGCTCAATAGTAGAAAGCAGGCCTGTGCACATGAATTAATTAAA AAATATAAAGGTAGTGATTAGACGACACATGTCCATAGGTAACCTGT CATAATTTTGAACAATTTCCCTTCTTTTCTTTTTTTTTTTTGGGTGC GGCGATATGTAGCTTGTTAATTTACACATCATGTACTTTTCTGCATC AAAATATGAAAGGCGATAGTAGCTAAAGAAAATACCGAGAATTTCCT CGAAAAGTTGACGACAAAAGAAAGGCATAAAAAAGTAATTTGAAAAT ATTTTAAAACTGTTTTAACCCATCTAGCATCCGCGCTAAAAAAGGAA GATACAGGATACAGCGGAAACAACTTTTAA
TEF1p	pCS1748	ATAGCTTCAAAATGTTTCTACTCCTTTTTTACTCTTCCAGATTTTCT CGGACTCCGCGCATCGCCGTACCACTTCAAAACACCCAAGCACAGCA TACTAAATTTCCCCTCTTTCTTCCTCTAGGGTGTCGTTAATTACCCG TACTAAAGGTTTGGAAAAGAAAAAAGAGACCGCCTCGTTTCTTTTTC TTCGTCGAAAAAGGCAATAAAAATTTTTATCACGTTTCTTTTTCTTG AAAATTTTTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATTGA TATTTAAGTTAATAAACGGTCTTCAATTTCTCAAGTTTCAGTTTCAT TTTTCTTGTTCTATTACAACTTTTTTTACTTCTTGCTCATTAGAAAG AAAGCATAGCAATCTAATCTAAGTTTTG
ADH1p	pL203	TAAAACAAGAAGAGGGTTGACTACATCACGATGAGGGGGATCGAAGA AATGATGGTAAATGAAATAGGAAATCAAGGAGCATGAAGGCAAAAGA CAAATATAAGGGTCGAACGAAAAATAAAGTGAAAAGTGTTGATATGA TGTATTTGGCTTTGCGGCGCCGAAAAAACGAGTTTACGCAATTGCAC AATCATGCTGACTCTGTGGCGGACCCGCGCTCTTGCCGGCCCGGCGA TAACGCTGGGCGTGAGGCTGTGCCCGGCGGAGTTTTTTGCGCCTGCA TTTTCCAAGGTTTACCCTGCGCTAAGGGGCGAGATTGGAGAAGCAAT AAGAATGCCGGTTGGGGTTGCGATGATGACGACCACGACAACTGGTG TCATTATTTAAGTTGCCGAAAGAACCTGAGTGCATTTGCAACATGAG TATACTAGAAGAATGAGCCAAGACTTGCGAGACGCGAGTTTGCCGGT GGTGCGAACAATAGAGCGACCATGACCTTGAAGGTGAGACGCGCATA ACCGCTAGAGTACTTTGAAGAGGAAACAGCAATAGGGTTGCTACCAG TATAAATAGACAGGTACATACAACACTGGAAATGGTTGTCTGTTTGA GTACGCTTTCAATTCATTTGGGTGTGCACTTTATTATGTTACAATAT GGAAGGGAACTTTACACTTCTCCTATGCACATATATTAATTAAAGTC CAATGCTAGTAGAGAAGGGGGGTAACACCCCTCCGCGCTCTTTTCCG ATTTTTTTCTAAACCGTGGAATATTTCGGATATCCTTTTGTTGTTTC CGGGTGTACAATATGGACTTCCTCTTTTCTGGCAACCAAACCCATAC ATCGGGATTCCTATAATACCTTCGTTGGTCTCCCTAACATGTAGGTG GCGGAGGGGAGATATACAATAGAACAGATACCAGACAAGACATAATG GGCTAAACAAGACTACACCAATTACACTGCCTCATTGATGGTGGTAC ATAACGAACTAATACTGTAGCCCTAGACTTGATAGCCATCATCATAT CGAAGTTTCACTACCCTTTTTCCATTTGCCATCTATTGAAGTAATAA TAGGCGCATGCAACTTCTTTTCTTTTTTTTTCTTTTCTCTCTCCCCC GTTGTTGTCTCACCATATCCGCAATGACAAAAAAATGATGGAAGACA

		CTAAAGGAAAAAATTAACGACAAAGACAGCACCAACAGATGTCGTTG TTCCAGAGCTGATGAGGGGTATCTCGAAGCACACGAAACTTTTTCCT TCCTTCATTCACGCACACTACTCTCTAATGAGCAACGGTATACGGCC TTCCTTCCAGTTACTTGAATTTGAAATAAAAAAAAGTTTGCTGTCTT GCTATCAAGTATAAATAGACCTGCAATTATTAATCTTTTGTTTCCTC GTCATTGTTCTCGTTCCCTTTCTTCCTTGTTTCTTTTTCTGCACAAT ATTTCAAGCTATACCAAGCATACAATCAACTATCTCATATACA
CYC1p	pL108	GAGCAGATCCGCCAGGCGTGTATATATAGCGTGGATGGCCAGGCAAC TTTAGTGCTGACACATACAGGCATATATATATGTGTGCGACGACACA TGATCATATGGCATGCATGTGCTCTGTATGTATATAAAACTCTTGTT TTCTTCTTTTCTCTAAATATTCTTTCCTTATACATTAGGACCTTTGC AGCATAAATTACTATACTTCTATAGACCCACAAACACAAATACCCCC CCTAAATTAATA
CYC1mp	pL74	TATGGCATGCATGTGCTCTGTATGTATATAAAACTCTTGTTTTCTTC TTTTCTCTAAATATTCTTTCCTTATACATTAGGTCCTTTGTAGCATA AATTACTATACTTCTATAGACACGCAAACACAAATACACACACTAAA TTACCG
tetO7-CYC1m	pL74	CTTATTGACCACACCTCTACCGGCAGATCCGCTAGGGATAACAGGGT AATATAGATCAATTCCTCGA TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG TCGAGCTCGGTACCC TATGGCATGCATGTGCTCTGTATGTATATAAAACTCTTGTTTTCTTC TTTTCTCTAAATATTCTTTCCTTATACATTAGGTCCTTTGTAGCATA AATTACTATACTTCTATAGACACGCAAACACAAATACACACACTAAA TTACCG GATCAATTCG
tetO	pL74	TCGAGTTTACCACTCCCTATCAGTGATAGAGAAAAGTGAAAG
tTA	pL192	ATGTCTAGATTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCT GCTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCCC AGAAGCTAGGTGTAGAGCAGCCTACATTGTATTGGCATGTAAAAAAT AAGCGGGCTTTGCTCGACGCCTTAGCCATTGAGATGTTAGATAGGCA CCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTT TACGTAATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGC GATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGAAAAACAGTA TGAAACTCTCGAAAATCAATTAGCCTTTTTATGCCAACAAGGTTTTT CACTAGAGAATGCATTATATGCACTCAGCGCTGTGGGGCATTTTACT TTAGGTTGCGTATTGGAAGATCAAGAGCATCAAGTCGCTAAAGAAGA AAGGGAAACACCTACTACTGATAGTATGCCGCCATTATTACGACAAG CTATCGAATTATTTGATCACCAAGGTGCAGAGCCAGCCTTCTTATTC GGCCTTGAATTGATCATATGCGGATTAGAAAAACAACTTAAATGTGA AAGTGGGTCCGCGTACAGCCGCGCGCGTACGAAAAACAATTACGGGT CTACCATCGAGGGCCTGCTCGATCTCCCGGACGACGACGCCCCCGAA GAGGCGGGGCTGGCGGCTCCGCGCCTGTCCTTTCTCCCCGCGGGACA CACGCGCAGACTGTCGACGGCCCCCCCGACCGATGTCAGCCTGGGGG ACGAGCTCCACTTAGACGGCGAGGACGTGGCGATGGCGCATGCCGAC GCGCTAGACGATTTCGATCTGGACATGTTGGGGGACGGGGATTCCCC GGGTCCGGGATTTACCCCCCACGACTCCGCCCCCTACGGCGCTCTGG ATATGGCCGACTTCGAGTTTGAGCAGATGTTTACCGATGCCCTTGGA ATTGACGAGTACGGTGGGTAG

yEGFP3	pCS321	ATGTCTAAAGGTGAAGAATTATTCACTGGTGTTGTCCCAATTTTGGT TGAATTAGATGGTGATGTTAATGGTCACAAATTTTCTGTCTCCGGTG AAGGTGAAGGTGATGCTACTTACGGTAAATTGACCTTAAAATTTATT TGTACTACTGGTAAATTGCCAGTTCCATGGCCAACCTTAGTCACTAC TTTCGGTTATGGTGTTCAATGTTTTGCGAGATACCCAGATCATATGA AACAACATGACTTTTTCAAGTCTGCCATGCCAGAAGGTTATGTTCAA GAAAGAACTATTTTTTTCAAAGATGACGGTAACTACAAGACCAGAGC TGAAGTCAAGTTTGAAGGTGATACCTTAGTTAATAGAATCGAATTAA AAGGTATTGATTTTAAAGAAGATGGTAACATTTTAGGTCACAAATTG GAATACAACTATAACTCTCACAATGTTTACATCATGGCTGACAAACA AAAGAATGGTATCAAAGTTAACTTCAAAATTAGACACAACATTGAAG ATGGTTCTGTTCAATTAGCTGACCATTATCAACAAAATACTCCAATT GGTGATGGTCCAGTCTTGTTACCAGACAACCATTACTTATCCACTCA ATCTGCCTTATCCAAAGATCCAAACGAAAAGAGAGACCACATGGTCT TGTTAGAATTTGTTACTGCTGCTGGTATTACCCATGGTATGGATGAA TTGTACAAATAA

Figure 3.11 Coding and promoter sequences used for plasmid construction

Chapter 4

Conclusions and future directions

Natural biological diversity promises spectacular possibilities for synthetic biology. Advances in the understanding of gene function and biological system design have allowed for the development of novel life with characteristics that are already beginning to address important societal needs. By reengineering metabolic pathways, researchers have created organisms that produce a wide array of important compounds, including medicines (Ro et al., 2006; Hawkins and Smolke, 2008) and biofuels (Atsumi et al., 2008; Dellomonaco et al., 2011). And by engineering human cells, other research has allowed for the reprogramming of cells to many desired tissues, for tailored immune cell-mediated targeting of cancers, and for rectification of aberrant gene expression (Yu et al., 2007; Santiago et al., 2008; Chen et al., 2010).

For many potential applications, however, existing gene-regulatory schemes hinder the development of more sophisticated control over gene expression. In eukaryotic cells, in particular, few engineered platforms allow for the customizable reprogramming of cells. Such platforms are generally either not modular, not applicable in organisms of interest, or not able to meet requirements needed for robust and reliable modulation of target gene expression. This work expanded our current capabilities in two directions: by expanding input modularity to be able to sense protein inputs, and by amplifying the range of output gene modulation.

As a result, it may be possible to build on these developments and create better-performing applications. For example, our ability to use the ribozyme switch platform to sense protein inputs may allow us to employ this platform for safer cellular reprogramming in human cells. This platform, containing no component heterologous proteins, may be free of the immunogenicity concerns that generally prevent the use of TF-based gene control in humans. With our now-expanded ability to detect protein inputs, we extend our ability to detect cellular states. Many characteristics of proteins make them excellent candidates as input molecules. For instance, high intracellular concentrations—average of ~1μM (Ghaemmaghami et al., 2003; Lu et al., 2007)—and wide availability of RNA aptamers that bind cognate proteins with often nM affinity appear promising parameters for design of gene control that responds to relevant changes of an input molecule. Additionally, current biochemical understanding often provides ample knowledge about what changes in protein expression are associated with aberrant cell function. It may be possible to use this information to design ribozyme switches that can detect these protein markers, and effect necessary cellular reprogramming by modulating expression of target genes.

Perhaps one shortcoming of this platform is a modest dynamic range, which may be a problem for some applications. We discussed earlier that this might be an inherent problem of the designs. Thermodynamic switches, which our data seem to insinuate our strand-displacements switches are examples of, may sacrifice dynamic range for accessibility of the different regulatory folds. The modest dynamic ranges may thus be a characteristic constraint of the platform.

Our amplifier platform largely addresses this shortcoming, as we move closer to the type of regulation observed in natural systems, and learn important lessons about possible reasons why nature evolved such architectures. Although there are many examples of naturally occurring riboswitches (Serganoc and Nudler, 2013), generally gene control is more often mediated by proteins, and largely by transcription factors. This may stem from the greater diversity of structure, and probably function, allowed from greater number of amino-acid building blocks compared to nucleic acid building blocks. Today, with the guidance of evolution, the species that more often mediate ligand-dependent control over gene expression are TFs. It may be that these species are the best options for mediating gene control. Earlier, we discussed characteristics of TF structure, including an often-repeated architecture of multimeric proteins binding repetitive DNA sites, sometimes cooperatively. Such was indeed the case for our amplifier platform, which allowed for high-fold regulation of gene expression levels. It may thus be that in the distant future, when we may be able to engineer proteins with customizable input domains, that we will be using synthetic custom TFs to regulate gene expression.

However, one general limitation of TF systems is immunogenicity associated with expression of heterologous proteins in vertebrates, limiting their use for many important applications. Perhaps in the distant future, such an immunogenic response may be prevented, either by somehow modulating the immune system, or by constructing synthetic TFs in a way that doesn't cause an immune response. In the near future, however, this limitation will continue to limit the use of TF-based strategies in humans. The amplifier platform, perhaps,

may be an exception. While general expression of a heterologous protein may be immunogenic, it may be that in our amplifier, the low-level expression of the TF needed to elicit high-level expression of an output gene may not cause a problem. Many TF-based systems function by constitutive high-expression of a TF, where ligand binding alters not the expression level of the TF, but its conformation, modulating output. Our amplifier system, by contrast, is basally "OFF." We only need very low-level expression of the transgenic tTA to elicit output protein expression. In yeast, basal expression of tTA was set to levels allowed by STE5 promoter, and further abrogated by our switches. For sdT8a, for example, basal expression would be expected to be at 3% of 1% of that allowed by, say, TEF1 promoter, a commonly used constitutive promoter—a very low value. And "ON" state expression was seen at very low induction of tTA by a GAL1 promoter in our cell strain, which again we chose because it allowed for low-level, titratable galactose induction. Potentially, the fact that we need very low levels of the synthetic TF inducer, if also a characteristic of a working human cell-version of the amplifier platform, may lower concerns for immunogenicity.

Even in applications in lower eukaryotes, such as for the construction of synthetic metabolic pathways in yeast, there are still limitations for widespread use of TF-based systems. For example, it remains exceedingly difficult to predict what protein sequence might fold into a given input domain. Indeed, it is difficult in general to generate a novel input domain, whether via rational design or directed evolution strategies. Thus, it may prove difficult for some time to develop TFs that respond to novel inputs. Our amplifier then remains a very good option. It's modular. And it has near-digital output. It may even help with

the development of new protein function. Even with many methods to generate novel biological functionality, whether by directed evolution of proteins, genome-wide genetic manipulation, or other technologies, measuring the effect of such manipulation on a given performance parameter may be difficult for many applications. Recently, a strategy was demonstrated for using theophylline-binding ribozyme switches to evolve enzymes with increased theophylline-producing properties (Michener and Smolke, 2012). Because RNA aptamers, and ribozyme switches, may perhaps generally be employed to detect a metabolite of interest, this strategy may be widely helpful. One hindrance, however, was the low dynamic range of the switch employed. Low-fold changes in reporter expression may be difficult to distinguish from signal noise. The amplifier perhaps might improve signal to noise ratios, simplifying this evolutionary method.

The work described here constitutes what may be important improvements over current synthetic gene control schemes in eukaryotes, combining greater modularity with near-digital control over output protein expression. In the near future, it will be important to further demonstrate the modularity and versatility of these platforms to allow for, perhaps, when combined with other tools such as complex genetic circuits, gene evolution strategies, and improved understanding of biological design principles, greater capabilities for synthetic biology to address many pressing societal needs.

References

Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. Proceedings of the National Academy of Sciences of the United States of America *102*, 12678–12683.

An, C., Trinh, V.U.B., and Yokobayashi, Y. (2006). Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer – small molecule interaction. RNA 710–716.

Anderson, J.C., Clarke, E.J., Arkin, A.P., and Voigt, C. a (2006). Environmentally controlled invasion of cancer cells by engineered bacteria. Journal of Molecular Biology *355*, 619–627.

Anderson, J.C., Wu, N., Santoro, S.W., Lakshman, V., King, D.S., and Schultz, P.G. (2004). An expanded genetic code with a functional quadruplet codon. Proceedings of the National Academy of Sciences of the United States of America *101*, 7566–7571.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America *100*, 3889–3894.

Atsumi, S., Hanai, T., and Liao, J.C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature *451*, 86–89.

Babiskin, A.H. (2011). Development of RNA-based Genetic Control Elements for Predictable Tuning of Protein Expression in Yeast Thesis by. California Institute of Technology.

Babiskin, A.H., and Smolke, C.D. (2011a). A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. Molecular Systems Biology *7*, 471.

Babiskin, A.H., and Smolke, C.D. (2011b). Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules. Nucleic Acids Research *39*, 5299–5311.

Banerjee, N. (2003). Identifying cooperativity among transcription factors controlling the cell cycle in yeast. Nucleic Acids Research *31*, 7024–7031.

Bateman, E. (1998). Autoregulation of eukaryotic transcription factors. Progress in Nucleic Acid Research and Molecular Biology *60*, 133–168.

Beisel, C.L., Chen, Y.Y., Culler, S.J., Hoff, K.G., and Smolke, C.D. (2011). Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. Nucleic Acids Research *39*, 2981–2994.

Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E.K. (2006). Quantification of protein half-lives in the budding yeast proteome. Proceedings of the National Academy of Sciences of the United States of America *103*, 13004–13009.

Bialek, W., and Setayeshgar, S. (2008). Cooperativity, Sensitivity, and Noise in Biochemical Signaling. Physical Review Letters *100*, 258101.

Birmingham, A., Anderson, E.M., Reynolds, A., Ilsley-tyree, D., Leake, D., Fedorov, Y., Baskerville, S., Maksimova, E., Robinson, K., Karpilow, J., et al. (2006). 3′ UTR seed matches , but not overall identity , are associated with RNAi off-targets. Nature *3*, 199–204.

Bloom, J.D., Meyer, M.M., Meinhold, P., Otey, C.R., MacMillan, D., and Arnold, F.H. (2005). Evolving strategies for enzyme engineering. Current Opinion in Structural Biology *15*, 447–452.

Bond-Watts, B.B., Bellerose, R.J., and Chang, M.C.Y. (2011). Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nature Chemical Biology *7*, 222–227.

Bonnet, J., Yin, P., Ortiz, M.E., Subsoontorn, P., and Endy, D. (2013). Amplifying Genetic Logic Gates. Science *599*,.

Bothe, J.R., Nikolova, E.N., Eichhorn, C.D., Chugh, J., Hansen, A.L., and Al-hashimi, H.M. (2011). Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *8*,.

Boyle, J., Robillard, G.T., and Kim, S.-H. (1980). Sequential folding of transfer RNA. Journal of Molecular Biology *139*, 601–625.

Broach, J.R., Jones, E.W., and Pringle, J.R. (1991). The Molecular and Cellular Biology of the Yeast Saccharomyces: Genome Dynamics, Protein Synthesis, and Energetics (Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press).

Buratowski, S. (2009). Progression through the RNA polymerase II CTD cycle. Molecular Cell *36*, 541–546.

Buskirk, A.R., Landrigan, A., and Liu, D.R. (2004). Engineering a Ligand-Dependent RNA Transcriptional Activator. *11*, 1157–1163.

Cassiday, L. a, and Maher, L.J. (2003). Yeast genetic selections to optimize RNA decoys for transcription factor NF-kappa B. Proceedings of the National Academy of Sciences of the United States of America *100*, 3930–3935.

Chang, Y.-H., Wang, Y.-C., and Chen, B.-S. (2006). Identification of transcription factor cooperativity via stochastic system model. Bioinformatics (Oxford, England) *22*, 2276–2282.

Chen, Y.Y., Jensen, M.C., and Smolke, C.D. (2010). Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. Proceedings of the National Academy of Sciences of the United States of America *107*, 8531–8536.

Collins, C.H., Arnold, F.H., and Leadbetter, J.R. (2005). Directed evolution of Vibrio fischeri LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. Molecular Microbiology *55*, 712–723.

Conze, T., Göransson, J., Razzaghian, H.R., Ericsson, O., Oberg, D., Akusjärvi, G., Landegren, U., and Nilsson, M. (2010). Single molecule analysis of combinatorial splicing. Nucleic Acids Research *38*, e163.

Culler, S.J., Hoff, K.G., and Smolke, C.D. (2010). Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. Science (New York, N.Y.) *330*, 1251–1255.

Curtis, T.P., Sloan, W.T., and Scannell, J.W. (2002). Estimating prokaryotic diversity and its limits. Proceedings of the National Academy of Sciences of the United States of America *99*, 10494–10499.

D'Espaux, L.D., Kennedy, A.B., Vowles, J.V., and Bloom, R.J. (2013). Development of protein-responsive ribozyme switches. In Preparation.

Dellomonaco, C., Clomburg, J.M., Miller, E.N., and Gonzalez, R. (2011). Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. Nature *476*, 355–359.

Dethoff, E. a, Chugh, J., Mustoe, A.M., and Al-Hashimi, H.M. (2012a). Functional complexity and regulation through RNA dynamics. Nature *482*, 322–330.

Dethoff, E. a, Petzold, K., Chugh, J., Casiano-Negroni, A., and Al-Hashimi, H.M. (2012b). Visualizing transient low-populated structures of RNA. Nature *491*, 724–728.

Dimitrov, R. a, and Zuker, M. (2004). Prediction of hybridization and melting for double-stranded nucleic acids. Biophysical Journal *87*, 215–226.

Edwards, A.M., Kanel, M., Young, R.A., and Kornberg, R.D. (1991). Two Dissociable Subunits of Yeast RNA Polymerase I1 Stimulate the. *266*, 71–75.

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends in Genetics : TIG *24*, 280–288.

Ellington, A.D. (2004). The Aptamer Database. http://aptamer.icmb.utexas.edu/

Ellington, A.D., and Szostak, J.W. (1990). In vitro selection of RNA molecules that bind specific ligands. Nature *346*, 818–822.

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

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science (New York, N.Y.) *297*, 1183–1186.

Farmer, W.R., and Liao, J.C. (2000). Improving lycopene production in Escherichia coli by engineering metabolic control. Nature Biotechnology *18*, 533–537.

Ferreira, L.M.R., and Mostajo-Radji, M. a (2013). How induced pluripotent stem cells are redefining personalized medicine. Gene *520*, 1–6.

Forrest, R.D. (1982). Early history of wound treatment. Journal of the Royal Society of Medicine *75*, 198–205.

Frieda, K.L., and Block, S.M. (2012). Direct observation of cotranscriptional folding in an adenine riboswitch. Science (New York, N.Y.) *338*, 397–400.

Futcher, B., Latter, G.I., Monardo, P., McLaughlin, C.S., and Garrels, J.I. (1999). A sampling of the yeast proteome. Molecular and Cellular Biology *19*, 7357–7368.

Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. Yeast (Chichester, England) *13*, 837–848.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Gibson, D.G., Glass, J.I., Lartigue, C., Noskov, V.N., Chuang, R.-Y., Algire, M. a, Benders, G. a, Montague, M.G., Ma, L., Moodie, M.M., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science (New York, N.Y.) *329*, 52–56.

Gietz, R.D., and Woods, R. a (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods in Enzymology *350*, 87–96.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proceedings of the National Academy of Sciences of the United States of America *89*, 5547–5551.

Grünwald, D., and Singer, R.H. (2010). In vivo imaging of labelled endogenous β-actin mRNA during nucleocytoplasmic transport. Nature *467*, 604–607.

Grünwald, D., Singer, R.H., and Rout, M. (2011). Nuclear export dynamics of RNA-protein complexes. Nature *475*, 333–341.

Von der Haar, T. (2008). A quantitative estimation of the global translational activity in logarithmically growing yeast cells. BMC Systems Biology *2*, 87.

Hamilton, R., Watanabe, C.K., and Boer, H.A. De (1987). Compilation and comparison of the sequence context around the AUG start codons in Saccharomyces cerevisiae mRNAs. Nucleic Acids Research *15*, 3581–3593.

Hawkins, K.M., and Smolke, C.D. (2006). The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae. The Journal of Biological Chemistry *281*, 13485–13492.

Hawkins, K.M., and Smolke, C.D. (2008). Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. Nature Chemical Biology *4*, 564–573.

Hinnebusch, A.G., and Lorsch, J.R. (2012). The mechanism of eukaryotic translation initiation: new insights and challenges. Cold Spring Harbor Perspectives in Biology *4*,.

Hollstein, M., Sidransky, D., Vogelstein, B., and Curtis, C. (1991). p53 Mutations in Human Cancers. Science (New York, N.Y.) *5015*, 49–53.

Horn, W.T., Convery, M.A., Stonehouse, N.J., Adams, C.J., Liljas, L., Phillips, S.E. V, and Stockley, P.G. (2004). The crystal structure of a high affinity RNA stem-loop complexed with the

bacteriophage MS2 capsid : Further challenges in the modeling of ligand – RNA interactions. 1776–1782.

Houseley, J., and Tollervey, D. (2009). The many pathways of RNA degradation. Cell *136*, 763–776.

Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer R. Journal of Molecular Biology *158*, 573–597.

Isaacs, F.J., Carr, P. a, Wang, H.H., Lajoie, M.J., Sterling, B., Kraal, L., Tolonen, A.C., Gianoulis, T. a, Goodman, D.B., Reppas, N.B., et al. (2011). Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science (New York, N.Y.) *333*, 348–353.

Jackson, D.A., Symonst, R.H., and Berg, P. (1972). Biochemical Method for Inserting New Genetic Information into DNA of. *69*, 2904–2909.

Jelinsky, S. a, and Samson, L.D. (1999). Global response of Saccharomyces cerevisiae to an alkylating agent. Proceedings of the National Academy of Sciences of the United States of America *96*, 1486–1491.

Johnson, J.A., Lu, Y.Y., Deventer, J.A. Van, and Tirrell, D.A. (2011). into proteins : recent developments and applications. *14*, 774–780.

Jorgensen, P., Edgington, N.P., Schneider, B.L., Rupes, I., Tyers, M., and Futcher, B. (2007). The Size of the Nucleus Increases as Yeast Cells Grow. *18*, 3523–3532.

June, C.H. (2007). Science in medicine Adoptive T cell therapy for cancer in the clinic. *117*,.

Karlsson, R. (1999). Affinity analysis of non-steady-state data obtained under mass transport limited conditions using BIAcore technology. 285–292.

Kashida, S., Inoue, T., and Saito, H. (2012). Three-dimensionally designed protein-responsive RNA devices for cell signaling regulation. Nucleic Acids Research *40*, 9369–9378.

Kennedy, A.B., Liang, J.C., and Smolke, C.D. (2012). A versatile cis-blocking and trans-activation strategy for ribozyme characterization. Nucleic Acids Research 1–13.

Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K., and Collins, J.J. (2012). Resource A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions. 647–658.

Khvorova, A., Lescoute, A., Westhof, E., and Jayasena, S.D. (2003). Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nature Structural Biology *10*, 708–712.

Kondo, T., Takahashi, R., and Inoue, H. (2012). Stem Cells and Cancer Stem Cells, Volume 2. *2*, 241–247.

Kozak, M. (2005). Regulation of translation via mRNA structure in prokaryotes and eukaryotes. Gene *361*, 13–37.

Kramer, F.R., and Mills, D.R. (1981). Secondary structure formation during RNA synthesis. Nucleic Acids Research *9*, 5109–5124.

Kumar, D., An, C.-I., and Yokobayashi, Y. (2009). Conditional RNA interference mediated by allosteric ribozyme. Journal of the American Chemical Society *131*, 13906–13907.

Kumei, Y., Nakajima, T., Sato, A., and Kamata, N. (1987). Reduction of G ^ phase duration and enhancement of c-myc gene expression in HeLa cells at hypergravity. 221–226.

De la Peña, M., Gago, S., and Flores, R. (2003). Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. The EMBO Journal *22*, 5561–5570.

Lamartina, S. (2003). Construction of an rtTA2s-m2/ttskid-Based transcription regulatory switch that displays no basal activity, good inducibility, and high responsiveness to doxycycline in mice and Non-Human primates. Molecular Therapy *7*, 271–280.

Lamontagne, B., and Abou Elela, S. (2007). Short RNA guides cleavage by eukaryotic RNase III. PloS One *2*, e472.

Lewin, B. (2008). Genes (Oxford University Press).

Liang, J.C., Bloom, R.J., and Smolke, C.D. (2011). Engineering biological systems with synthetic RNA molecules. Molecular Cell *43*, 915–926.

Liang, J.C., Chang, A.L., Kennedy, A.B., and Smolke, C.D. (2012). A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. Nucleic Acids Research *40*, 1–14.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E.M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nature Biotechnology *25*, 117–124.

Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. Journal of Molecular Biology *288*, 911–940.

Maul, G.G., and Deaven, L. (1977). Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. Journal of Cell Biology *73*, 748–760.

McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nuñez, A., Butrym, E.D., Richards, M.P., Wang, C.-S., et al. (2004). Fermented beverages of pre- and proto-historic China. Proceedings of the National Academy of Sciences of the United States of America *101*, 17593–17598.

Michener, J.K., and Smolke, C.D. (2012). High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. Metabolic Engineering *14*, 306–316.

Miller, C., Schwalb, B., Maier, K., Schulz, D., Dümcke, S., Zacher, B., Mayer, A., Sydow, J., Marcinowski, L., Dölken, L., et al. (2011). Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Molecular Systems Biology *7*, 458.

Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C., and Voigt, C. a (2012). Genetic programs constructed from layered logic gates in single cells. Nature *491*, 249–253.

Mora, C., Tittensor, D.P., Adl, S., Simpson, A.G.B., and Worm, B. (2011). How many species are there on Earth and in the ocean? PLoS Biology *9*, e1001127.

Mugnier, P., and Tuite, M.F. (1999). Translation termination and its regulation in eukaryotes: recent insights provided by studies in yeast. Biochemistry. Biokhimiia̋ *64*, 1360–1366.

Munsky, B., Neuert, G., and Van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. Science (New York, N.Y.) *336*, 183–187.

Nasser, M.W., Pooja, V., Abdin, M.Z., and Jain, S.K. (2003). Evaluation of Yeast as an Expression System. *2*, 477–493.

Newman, D.J., and Cragg, G.M. (2007). Natural products as sources of new drugs over the last 25 years. Journal of Natural Products *70*, 461–477.

Newton, K.F., Newman, W., and Hill, J. (2012). Review of biomarkers in colorectal cancer. Colorectal Disease : the Official Journal of the Association of Coloproctology of Great Britain and Ireland *14*, 3–17.

Nistala, G.J., Wu, K., Rao, C. V, and Bhalerao, K.D. (2010). A modular positive feedback-based gene amplifier. Journal of Biological Engineering *4*, 4.

Oeffinger, M., and Zenklusen, D. (2012). To the pore and through the pore: A story of mRNA export kinetics. Biochimica Et Biophysica Acta *1819*, 494–506.

Osborn, M.J., Starker, C.G., McElroy, A.N., Webber, B.R., Riddle, M.J., Xia, L., DeFeo, A.P., Gabriel, R., Schmidt, M., Von Kalle, C., et al. (2013). TALEN-based Gene Correction for Epidermolysis Bullosa. Molecular Therapy 1–9.

Parker, R., and Song, H. (2004). The enzymes and control of eukaryotic mRNA turnover. Nature Structural & Molecular Biology *11*, 121–127.

Parrott, a M., Lago, H., Adams, C.J., Ashcroft, a E., Stonehouse, N.J., and Stockley, P.G. (2000). RNA aptamers for the MS2 bacteriophage coat protein and the wild-type RNA operator have similar solution behaviour. Nucleic Acids Research *28*, 489–497.

Pelechano, V., Chavez, S., and Perez-Ortin, J. (2010). A Complete Set of Nascent Transcription Rates for Yeast Genes. PloS One *5*,.

Pemberton, L.F., Blobel, G., and Rosenblum, J.S. (1998). Transport routes through the nuclear pore complex. Current Opinion in Cell Biology *10*, 392–399.

Peralta-Yahya, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D., and Lee, T.S. (2011). Identification and microbial production of a terpene-based advanced biofuel. Nature Communications *2*, 483.

Perreault, J., Weinberg, Z., Roth, A., Popescu, O., Chartrand, P., Ferbeyre, G., and Breaker, R.R. (2011). Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. PLoS Computational Biology *7*, e1002031.

Pittman, R. (2011). Regulation of Tissue Oxygenation (San Rafael, CA).

Preiss, T., and W Hentze, M. (2003). Starting the protein synthesis machine: eukaryotic translation initiation. BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology *25*, 1201–1211.

Proctor, J.R., and Meyer, I.M. (2013). COFOLD: an RNA secondary structure prediction method that takes co-transcriptional folding into account. Nucleic Acids Research 1–11.

Quail, M. a, Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., and Gu, Y. (2012). A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics *13*, 341.

Quarta, G., Sin, K., and Schlick, T. (2012). Dynamic energy landscapes of riboswitches help interpret conformational rearrangements and function. PLoS Computational Biology *8*, e1002368.

Ramos, J.L., Martı, M., Molina-henares, A.J., Tera, W., Brennan, R., and Tobes, R. (2005). The TetR Family of Transcriptional Repressors. Microbiology and Molecular Biology Reviews *69*, 326–356.

Ramsey, S. a, Smith, J.J., Orrell, D., Marelli, M., Petersen, T.W., De Atauri, P., Bolouri, H., and Aitchison, J.D. (2006). Dual feedback loops in the GAL regulon suppress cellular heterogeneity in yeast. Nature Genetics *38*, 1082–1087.

Reichheld, S.E., Yu, Z., and Davidson, A.R. (2009). The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. Proceedings of the National Academy of Sciences of the United States of America *106*, 22263–22268.

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

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K. a, Eachus, R. a, Ham, T.S., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940–943.

Romanos, M. a, Scorer, C. a, and Clare, J.J. (1992). Foreign gene expression in yeast: a review. Yeast (Chichester, England) *8*, 423–488.

Rosenbluth, M.J., Lam, W. a, and Fletcher, D. a (2006). Force microscopy of nonadherent cells: a comparison of leukemia cell deformability. Biophysical Journal *90*, 2994–3003.

Saito, H., Fujita, Y., Kashida, S., Hayashi, K., and Inoue, T. (2011). Synthetic human cell fate regulation by protein-driven RNA switches. Nature Communications *2*, 160.

Sambrook, J. (2001). Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell (Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory).

Santiago, Y., Chan, E., Liu, P.-Q., Orlando, S., Zhang, L., Urnov, F.D., Holmes, M.C., Guschin, D., Waite, A., Miller, J.C., et al. (2008). Targeted gene knockout in mammalian cells by using engineered zinc-finger nucleases. Proceedings of the National Academy of Sciences of the United States of America *105*, 5809–5814.

Schirmer, A., Rude, M. a, Li, X., Popova, E., and Del Cardayre, S.B. (2010). Microbial biosynthesis of alkanes. Science (New York, N.Y.) *329*, 559–562.

Segref, a, Sharma, K., Doye, V., Hellwig, a, Huber, J., Lührmann, R., and Hurt, E. (1997). Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. The EMBO Journal *16*, 3256–3271.

Sellick, C. a, Campbell, R.N., and Reece, R.J. (2008). Galactose metabolism in yeast-structure and regulation of the leloir pathway enzymes and the genes encoding them. International Review of Cell and Molecular Biology *269*, 111–150.

Serganoc, A., and Nudler, E. (2013). A Decade of Riboswitches. Cell *152*, 17–24.

Shoemaker, C.J., and Green, R. (2012). Translation drives mRNA quality control. Nature Structural & Molecular Biology *19*, 594–601.

Singer, R.H., and Grünwald, D. (2013). Multiscale dynamics in nucleocytoplasmic transport. *24*, 100–106.

Steen, E.J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S.B., and Keasling, J.D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature *463*, 559–562.

To, T.-L., and Maheshri, N. (2010). Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science (New York, N.Y.) *327*, 1142–1145.

Tschumper, G., and Carbon, J. (1983). Copy number control by a yeast centromere. Gene *23*, 221–232.

Tuerk, C., and Gold, L. (1990). Systematic Evolution of Ligands by Exponential Enrichment : RNA Ligands to Bacteriophage T4 DNA Polymerase. Science (New York, N.Y.) *249*, 506–510.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., and Gregory, P.D. (2010). Genome editing with engineered zinc finger nucleases. Nature Reviews. Genetics *11*, 636–646.

Vaidya, N., Manapat, M.L., Chen, I. a, Xulvi-Brunet, R., Hayden, E.J., and Lehman, N. (2012). Spontaneous network formation among cooperative RNA replicators. Nature *491*, 72–77.

Wagner, I., and Musso, H. (1983). New Naturally Occurring Amino Acids. Angewandte Chemie International Edition *22*, 816–828.

Wang, H.H., Isaacs, F.J., Carr, P. a, Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898.

Wei, K.Y., Chen, Y.Y., and Smolke, C.D. (2013). A yeast-based rapid prototype platform for gene control elements in mammalian cells. Biotechnology and Bioengineering *110*, 1201–1210.

Weigel, M.T., and Dowsett, M. (2010). Current and emerging biomarkers in breast cancer: prognosis and prediction. Endocrine-related Cancer *17*, R245–62.

Werstuck, G., and Green, M.R. (1998). Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions. Science *282*, 296–298.

Win, M.N., and Smolke, C.D. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proceedings of the National Academy of Sciences of the United States of America *104*, 14283–14288.

Win, M.N., and Smolke, C.D. (2008a). Higher-order cellular information processing with synthetic RNA devices. Science (New York, N.Y.) *322*, 456–460.

Win, M.N., and Smolke, C.D. (2008b). Higher-order cellular information processing with synthetic RNA devices. Science (New York, N.Y.) *322*, 456–460.

Winkler, W.C., Nahvi, A., Roth, A., Collins, J. a, and Breaker, R.R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. Nature *428*, 281–286.

Witzgall, R., O'Leary, E., Leaf, a, Onaldi, D., and Bonventre, J. V (1994). The Krüppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. Proceedings of the National Academy of Sciences of the United States of America *91*, 4514–4518.

Wurster, S.E., Bida, J.P., Her, Y.F., and Maher, L.J. (2009). Characterization of anti-NF-kappaB RNA aptamer-binding specificity in vitro and in the yeast three-hybrid system. Nucleic Acids Research *37*, 6214–6224.

Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011). Multi-input RNAi-based logic circuit for identification of specific cancer cells. Science (New York, N.Y.) *333*, 1307–1311.

Yang, Z., Chen, F., Chamberlin, S.G., and Benner, S. a (2010). Expanded genetic alphabets in the polymerase chain reaction. Angewandte Chemie (International Ed. in English) *49*, 177–180.

Yu, J., Vodyanik, M. a, Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G. a, Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science (New York, N.Y.) *318*, 1917–1920.

Zhang, F., Carothers, J.M., and Keasling, J.D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nature Biotechnology *30*, 354–359.

Zhou, H., Liu, D., and Liang, C. (2004). Challenges and strategies: the immune responses in gene therapy. Medicinal Research Reviews *24*, 748–761.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research *31*, 3406–3415.