Engineering Ligand Control of RNA Interference

Thesis by

Chase Lawrence Beisel

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2009

(Defended May 14th, 2009)

Dedicated in loving memory to my mother

Joanne Hendler Heltzer

ACKNOWLEDGEMENTS

The work described in this thesis represents five years of dedication, sweat, toil, and sacrifice – most of which from those who deserve recognition. First and foremost, I want to thank my thesis advisor, Prof. Christina Smolke. Financial support aside, Christina has patiently shown me the ways of science and offered a consistent source of reason and insight through my own uncertainties and far-fetched project ideas. I will forever benefit from her critical eye and willingness to improve how I communicate my ideas to the world.

Other various professors have made immeasurable contributions to whom I am grateful and now excitedly pursue a career in academia: Prof. Julie Kornfield for her kindness, humility, and willingness to provide aid; Prof. Peter Reilly at Iowa State University for introducing me to scientific research, the joys of a professorial lifestyle, and the global scale of science; Profs. Vassily Hatzimanikatis at the EPFL and Jacqueline Shanks at Iowa State University for their friendship and pointed advice through the highs and lows of my undergraduate and graduate career; Prof. Garrett Soukup at Creighton University for his fascination with RNA and corresponding willingness to fuel my own excitement; and Dr. Michael Vicic (a full-fledged professor in my mind) for personally demonstrating that the Ph.D. is merely a tool in which to engage life's questions and challenges – whether it be topics in fluid mechanics or devising and marketing an online television show database.

This thesis (and my mostly intact sanity) would not have been possible without my treasured friends in the Smolke lab: Dr. Travis Bayer for his willingness to take me

under his wing, his honest excitement about scientific discovery, his positive and outgoing demeanor even in the face of adversity, and his uncanny ability to find an appropriate baseball analogy for any conceivable situation; Dr. Kristy Hawkins for her compassion and unparalleled work ethic; Dr. Kevin Hoff for his limitless academic knowledge and refreshing sense of perspective; and Stephanie Culler for her love of music, wine, movies, alternative splicing, and all things Southern California.

My church community has kept me fully grounded and focused on the most important pursuits in life: Jim, my classmate and former roommate, who first invited me to Warehouse and has been one of my references for a Christ-centered life; my Jr. High group, who taught me about the unlimited potential (and untamed energy) of youth; and my dear friends Mike, Melissa, Todd, Joe, Becca, Katie, Kate, Melinda, and Casey.

Finally, I want to thank my family. My father, Kirk, and stepmother, Maryann, have constantly extended love and encouragement, especially in times when I was filled with self-doubt and despair. My mother, Joanne, instilled in me the importance of contentment and serving others, even when terminally ill. In the same right, my stepfather Dale showed me what unfailing devotion looks like even under extreme duress. Finally I want to thank my wife, Amy. She first entered my life during candidacy and undoubtedly saw the worst of me during those first few months. To my amazement and gratitude, she still pursued the relationship. Since then, she has outpoured love, support, and friendship to which I am forever in her debt. Thankfully she has given me a lifetime to repay and I couldn’t be happier.

ABSTRACT

RNA is a rich and versatile substrate for the construction of information processing devices. These devices detect the levels of specified intracellular biomolecules and control cellular behavior accordingly. With few superficial constraints on the identity of the recognized biomolecule or the targeted gene, RNA-based information processing devices can be rapidly implemented toward various applications in medicine and biotechnology. To advance the design and implementation of RNA-based information processing devices, we delineated general design principles and applied these principles to the construction of devices that operate through RNA interference (RNAi).

RNAi represents an endogenous enzymatic pathway present in humans and other eukaryotes that mediates targeted gene silencing. The pathway has garnered recent interest as a revolutionary biological research tool and as a targeted therapeutic strategy. While RNAi has left an indelible mark on the scientific community, exerting greater control would advance the applicability and safety of this already impressive gene silencing mechanism. Toward this goal, we engineered ligand control of three types of RNAi effectors in mammalian cells: small interfering (si)RNAs, small hairpin (sh)RNAs, and microRNAs (miRNAs). Engineering frameworks enabled facile replacement of the biomolecule sensory and gene targeting domains, thus lending to rapid implementation as biosensors or autonomous control devices. Experimental and computational characterization studies provided a comprehensive understanding of device behavior, thereby facilitating forward design.

Naturally-occurring analogs of RNA-based information processing devices are riboswitches. Riboswitches predominantly mediate dynamic feedback in metabolism and share many traits with current examples of engineered information processing devices. Various experimental characterization studies of riboswitches showed that kinetics underlying events such as conformational switching and ligand binding have a substantial impact on device performance, although these factors remain to be comprehensively evaluated or considered when formulating design principles for synthetic riboswitch construction. We explored the contribution of kinetic factors to riboswitch performance in silico, where model predictions matched experimental observations, including results from our ligand-responsive RNAi effectors. From our modeling results, we developed a general set of design principles that guide riboswitch assembly and performance tuning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Advancing the construction of biological systems

Biological engineering is entering a new, unprecedented era of design and implementation. DNA synthesis technologies currently allow the generation of entire genomes, while continuing insights into the diversity and underlying mechanisms of life are providing the basis to construct autonomous cells and organisms with programmed behaviors and functions. Initial engineered systems resulting from these advances are already addressing global challenges, with prominent examples in sustainable energy production and inexpensive pharmaceutical synthesis (Keasling, 2008; Savage *et al*, 2008). However, these simple systems pale in comparison to the complexity, scale, and diversity of function observed in the natural world. To continue advancing the sophistication and scale of engineered systems, a common need exists to improve the availability and encoded functionalities of biological components that form the basis of these systems. With each development, new tools will emerge that will yield further insights into biological function and address current challenges in our world. While tool development has primarily focused on protein design, a new player – RNA – has gained prominence and is poised to reshape how we approach the construction of biological systems.

The many faces of RNA

RNA is a rich and versatile biomolecule composed of a linear chain of nucleotide units. Each unit includes one of four chemical bases (G, A, U, and C), a ribose sugar, and a negatively-charged phosphodiester backbone. Individual bases can pair through specific hydrogen bonding interactions (G to C, A to U), such that a strand of RNA can identify its sequence complement with high resolution. Base pairing can also occur internally, lending to the formation of complex secondary and tertiary structures. These structures can form unique epitopes for protein recognition, catalyze the cleavage or ligation of nucleic acid molecules, and bind diverse biomolecules with high affinity and specificity. Furthermore, the structure and activity of RNA sequences are amenable to computational prediction based on the hierarchical folding of RNA (Shapiro *et al*, 2007). Finally, the pivotal localization of RNA in gene expression translates the myriad of functionalities associated with RNA into potential gene regulatory strategies in cells.

Within biological systems, RNA is traditionally viewed as a passive information carrier in gene expression. However, the identification of endogenously-expressed RNAs displaying these different functionalities has radically redefined the natural role of RNA. The identified RNAs do not encode proteins – garnering the title noncoding RNAs – yet act as critical regulators in diverse cellular processes. Prominent examples include microRNAs in development (Erson and Petty, 2008), piwi-interacting RNAs in transposon defense and methylation patterning (Klattenhoff and Theurkauf, 2008), bacterial small RNAs in stress responses (Majdalani *et al*, 2005), self-cleaving ribozymes in viral reproduction (Lilley, 1999; Shih and Been, 2002), and riboswitches in dynamic metabolic control (Henkin, 2008). The current list may represent the tip of the iceberg, as

a plethora of recently identified noncoding RNAs still awaits characterization (Guttman *et al*, 2009).

Engineering riboswitches

Inspired by the emerging regulatory role of RNA in nature, biological engineers have begun designing synthetic RNA sequences that serve as genetic regulators in biological systems (Isaacs *et al*, 2006). Design efforts have predominantly focused on riboswitches, naturally occurring noncoding RNAs generally located in the 5' untranslated region (UTR) of genes involved in metabolite biosynthesis and breakdown (Winkler, 2005). Riboswitches dynamically activate (ON behavior) or repress (OFF behavior) the expression of downstream genes in response to changing metabolite concentrations (Figure 1.1). Two domains with prescribed functionalities are responsible for riboswitch activity: an aptamer domain that recognizes and binds a cognate metabolite and a gene regulatory domain that controls gene expression. Metabolites of diverse compositions are specifically recognized, including vitamins, amino acids, nucleic acid bases, metals, and secondary messengers. In addition, natural riboswitches utilize diverse regulatory mechanisms, including alternative splicing, transcriptional termination, translational repression, mRNA destabilization, and combinations thereof (Barrick and Breaker, 2007). In almost all instances of characterized riboswitches, metabolite binding stabilizes an alternative conformation associated with differential expression levels of the downstream gene. Therefore, riboswitches mediate an indirect link between molecular information in the cell and the genetic program primarily through conformational

stabilization, providing a natural RNA-based strategy to coordinate cellular processes in response to environmental perturbations.

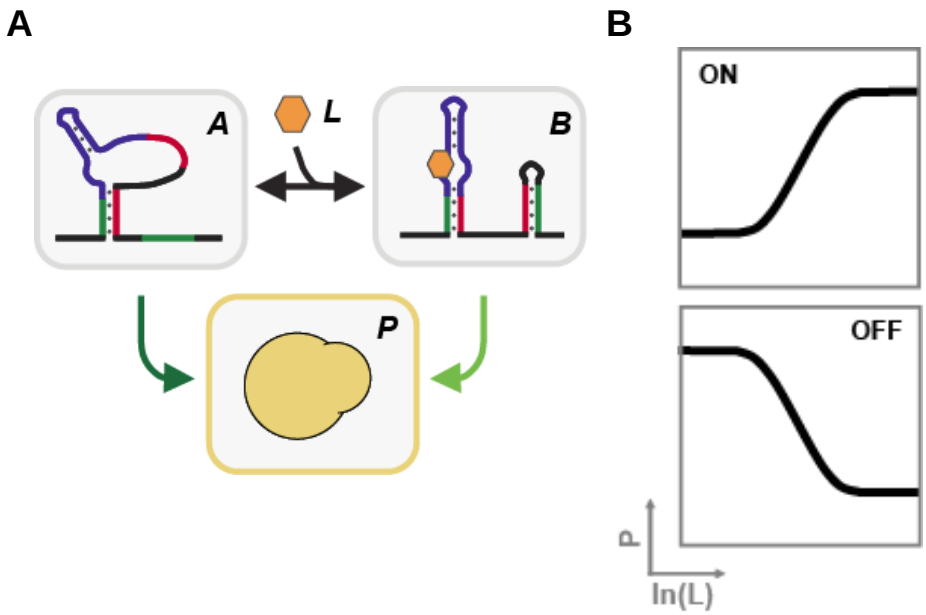


Figure 1.1 General schematic of riboswitch function. Riboswitches are generally composed of two domains that encode a ligand-binding aptamer and a genetic regulatory element. Ligand binding to the aptamer modulates the activity of the regulatory element, thereby affecting the expression levels of genes under the control of the riboswitch. (**A**) Most riboswitches adopt two distinct conformations designated A and B associated with differential expression levels of the encoded protein (P). Only one of the two conformations (conformation B) contains a formed aptamer shown in blue. Interaction between the aptamer and the ligand (L) stabilizes conformation B, thereby biasing conformational partitioning and subsequent protein expression levels. (**B**) Response curves relating ligand concentration (L) and protein levels (P). Ligand addition either increases (ON behavior) or decreases (OFF behavior) protein levels when the conformation with the formed aptamer is associated with higher or lower expression levels, respectively.

Initial efforts to design synthetic riboswitches began after the discovery of aptamers, the RNA elements responsible for molecular recognition. Novel aptamers can be generated in vitro through a procedure termed SELEX (systematic evolution of ligands by exponential enrichment) that isolates RNA sequences with specific ligand binding activity (Ellington and Szostak, 1990; Tuerk and Gold, 1990). Following the development of SELEX, aptamers have been selected against a wide range of molecules – well beyond the known number and types of molecules recognized by natural riboswitches. In an effort to couple genetic control and molecular recognition by aptamers, Werstruck and Green introduced aptamers selected against Hoechst dye into the 5' UTR of a mammalian reporter gene (Werstuck and Green, 1998). Addition of the dye to cells harboring the engineered reporter construct greatly reduced reporter production by perceivably inhibiting ribosomal loading. This seminal demonstration of ligand control of gene expression catalyzed the exploration of synthetic riboswitch design.

Since the communication of this seminal example in 1998, researchers have devised ingenious ways to combine aptamers and regulatory elements in order to introduce ligand control of gene expression. Riboswitch designs now include diverse regulatory mechanisms and operate in bacteria, yeast, and mammalian cells (Suess and Weigand, 2008), all working toward the design of riboswitches to control cellular processes in response to a selected ligand in different applications. Through the design process, a few desirable properties of riboswitches were identified that facilitate the implementation process. The first property is aptamer modularity – namely, the ability to introduce new aptamers in order to change the recognized ligand without riboswitch

redesign. By limiting the necessity for redesign, a single riboswitch design can be widely applied without significant time investment. A second property is tuning of the response curve, the quantitative relationship between ligand concentration and gene expression levels. By modifying the characteristics of riboswitch activity, such as aptamer affinity or the stability of the alternative conformation, the associated response curve could be rationally tuned to meet the regulatory demands of the desired application. A final property is the ability to control endogenous genes. Most riboswitch designs act in cis such that the riboswitch is included in the transcript of the regulated gene. However, directed recombination techniques necessary to couple endogenous genes with synthetic riboswitches are either technically challenging or unavailable in many organisms, limiting most riboswitch designs to the control of heterologous genes. At the start of this thesis work, only one riboswitch design attained regulation in trans, which relied on antisense inhibition as the regulatory mechanism (Bayer and Smolke, 2005). The challenges associated with antisense technology and the advent of a recently discovered trans regulatory mechanism, RNA interference (RNAi), provided the motivation to pursue the design of riboswitches that exploit this powerful mechanism of control (Sahu *et al*, 2007; Scherer and Rossi, 2003).

The emergence and importance of RNA interference

RNAi represents a powerful trans-targeting mechanism originally discovered only ten years ago in the worm *Caenorhabditis elegans* (Figure 1.2). After injecting *C. elegans* with double-stranded (ds)RNA homologous to a target gene, Fire and Mello observed dramatic silencing of the encoded protein from even a few molecules of the applied RNA

(Fire *et al*, 1998). The observed knockdown was sequence-specific, suggesting a targeted approach to silence any individual gene in the genome. Ensuing studies revealed an endogenous enzymatic pathway that conducts gene silencing through a process generally called RNA interference or RNAi. The general pathway was later identified in diverse organisms including insects, fungi, plants, and mammals, and unified seemingly unrelated silencing mechanisms such as quelling in fungi and post-transcriptional gene silencing in plants (Sharp, 1999).

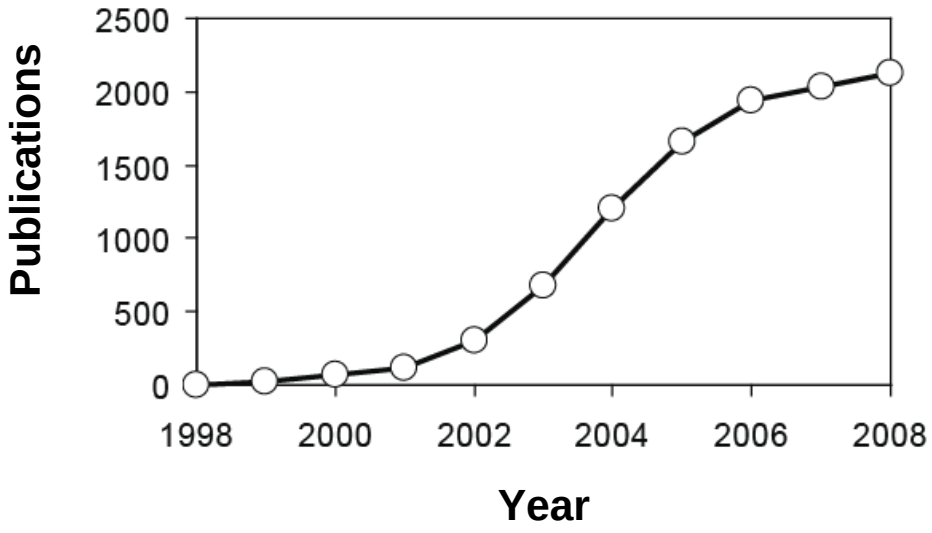


Figure 1.2 Number of RNAi publications each year since the discovery of RNAi in 1998. The search was conducted on the online NCBI PubMed database. Publications with the phrase "RNA interference" in the title or abstract were included in the tally. Notable landmarks in the timeline include the discovery of RNAi in mammalian cells in 2001 and the awarding of the Nobel Prize in Physiology or Medicine for the discovery of RNAi in 2006.

The RNAi pathway encompasses a linear processing cascade that sequentially reduces the applied dsRNA to a single strand of 21 – 24 nucleotides. This strand, termed

the guide strand, is loaded into the RNA induced silencing complex (RISC) and serves as a template to recruit complementary transcripts to RISC through base pairing interactions. Recruited transcripts that are partially or fully complementary to the guide strand are translationally inhibited or cleaved by RISC, thereby suppressing levels of the encoded protein.

RNAi can generally accept four different forms of dsRNA at distinct points in the processing cascade: microRNAs (miRNAs) at the beginning, small hairpin (sh)RNAs and long dsRNA in the middle, and small interfering (si)RNAs at the end (Figure 1.3). miRNAs are large imperfect hairpins present in pol II transcripts and exist as naturally-occurring effectors of RNAi. Processing by the nuclear RNase III Drosha (Lee *et al*, 2003) releases an shRNA with a characteristic 2-nt 3'overhang recognized for nuclear export by the shuttle protein Exportin-5 (Lund *et al*, 2004; Yi *et al*, 2003). Processing of shRNAs and long dsRNA by the cytoplasmic RNase III Dicer produces a duplex siRNA with symmetric 2-nt 3'overhangs. The siRNA is loaded into RISC, where one of the two strands is cleaved and discarded (Matranga *et al*, 2005; Rand *et al*, 2005) based on a collection of factors that include loading orientation and the relative thermodynamic stability on the 3' end of the retained strand (Amarzguioui *et al*, 2006). The remaining piece of the siRNA serves as the guide strand for transcript targeting. Note that the transcribed miRNA, the produced shRNA, and the final guide strand are also called primary, precursor, and mature miRNAs, respectively, within the natural miRNA field due to the difference in nomenclature used for naturally-occurring and synthetic RNAi effectors.

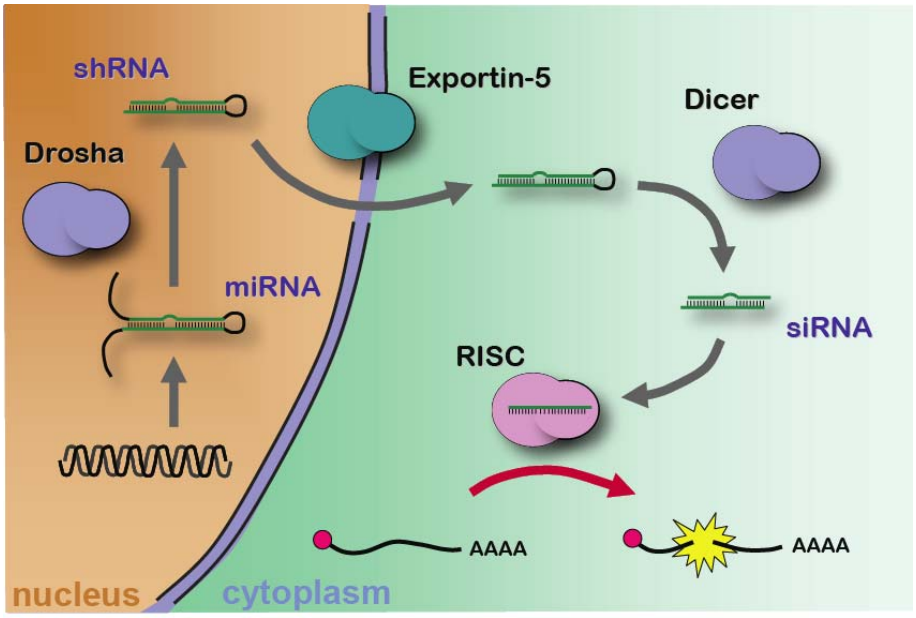


Figure 1.3 RNAi processing pathway and post-transcriptional gene silencing mechanism in mammals. The linear cascade terminates with the loading of RISC with the guide strand responsible for transcript targeting. The guide strand can be generated from three different RNAs: miRNAs, shRNAs, and siRNAs. miRNAs are processed to shRNAs by Drosha located in nucleus. Following export to the cytoplasm by Exportin-5, shRNAs are processed to siRNAs by Dicer. One of the two strands is retained by RISC as the guide strand.

With a greater understanding of the RNAi pathway came the advent of synthetic RNAi effectors for targeted gene silencing. Within mammalian cells, siRNAs were first developed for exogenous delivery in order to circumvent the non-specific interferon response (Elbashir *et al*, 2001). This response is induced by long dsRNA containing at least 30 consecutive base pairs and triggers a global shutdown of protein synthesis. shRNAs were developed after siRNAs as single RNA molecules that induce RNAi, where their use was bolstered by reports that Dicer processing improved RISC loading and subsequent gene silencing (Kim *et al*, 2005; Siolas *et al*, 2005). shRNAs have been

primarily used to endogenously activate RNAi since only one promoter is necessary to transcribe the molecule. With the discovery and characterization of the miRNA-processing enzyme Drosha, synthetic miRNAs dubbed second-generation shRNAs were developed (Silva *et al*, 2005). These RNAi effectors showed improved silencing over shRNAs presumably due to earlier entry into the RNAi processing pathway.

The ability to specifically silence any individual gene through RNAi has had a profound impact on fundamental biological research and holds tremendous potential in the treatment of genetic diseases. Through the facile design and implementation of RNAi effectors targeting individual genes, researchers are capable of rapidly deciphering genetic functions. Moving beyond single-gene studies, comprehensive libraries have been developed to target every expressed coding region in a genome (Haney, 2007). Associated large-scale screens have elucidated entire regulatory networks and identified novel members of critical signaling pathways (Berns *et al*, 2004; Mullenders *et al*, 2009; Tang *et al*, 2008).

The characteristic high specificity and efficiency of silencing has motivated the consideration of RNAi as a therapeutic strategy. RNAi effectors have been tested in model systems for cancer (Takeshita and Ochiya, 2006), HIV infection (Singh, 2008), neurodegenerative diseases (Farah, 2007), and macular degeneration (Whitehead *et al*, 2009) to silence genes associated with the disease state, with promising results that have spurred a flurry of clinical trials (Castanotto and Rossi, 2009). Related research efforts have revealed two major safety and efficacy concerns: off-target effects and cytotoxicity. Off-target effects result from the suppression of non-target genes with partial sequence complementarity to the guide strand, thereby compromising the specificity of targeting

and the potential therapeutic benefit (Kim and Rossi, 2009). Cytotoxicity has been attributed to induction of the interferon response and saturation of the RNAi machinery. Apart from dsRNA with 30 contiguous base pairs, high concentrations of dsRNA, the presence of 5’ phosphates, or some dsRNA sequences can efficiently trigger the innate immune response (Kim and Rossi, 2009). Saturating the RNAi machinery appears to impair the regular function of endogenously-expressed miRNAs, which may be critical for cellular homeostasis. Mouse studies revealed that constitutive over-expression of some shRNAs induced death within a month (Grimm *et al*, 2006). The toxic effect was relieved by co-expressing Exportin-5 or avoided by expressing synthetic miRNAs, suggesting alternative routes to improve the safety of RNAi therapeutics. Through further research efforts to overcome these challenges, the use of RNAi will progress rapidly in the clinic and may revolutionize the treatment of a broad range of genetic diseases.

Design of riboswitches that act through RNA interference

The work described in this thesis exploits the regulatory prowess of RNAi in riboswitch design and evaluates the quantitative relationship between riboswitch function and performance. Chapters 2 through 4 discuss the design of riboswitches that introduce ligand control of siRNA, shRNA, and miRNA processing. Designs rely on known structural requirements for efficient processing by the RNAi machinery elucidated through structural studies. A key pursuit in these chapters is achieving both aptamer modularity and tuning of the response curve to facilitate the widespread implementation of these designs toward different applications. The design of ligand-responsive shRNAs also integrated mathematical modeling and RNA folding algorithms toward the

development of sequence-function relationships for the forward design of riboswitches. Chapter 5 extends on these modeling efforts by investigating how the dynamics of riboswitch function dictate performance for various cis-acting regulatory mechanisms. A collection of design principles were elucidated to guide future design efforts and highlight areas for future experimental studies to better understand the design of synthetic riboswitches and the evolutionary optimization of natural riboswitches. Finally, Chapter 6 provides a perspective on where future research should be focused and the role RNA may play in the construction of biological systems.

REFERENCES

Amarzguioui M, Lundberg P, Cantin E, Hagstrom J, Behlke MA, Rossi JJ (2006) Rational design and in vitro and in vivo delivery of Dicer substrate siRNA. *Nat Protoc* **1:** 508-517.

Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8:** R239.

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23:** 337-343.

Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, Kerkhoven RM, Madiredjo M, Nijkamp W, Weigelt B, Agami R, Ge W, Cavet G, Linsley PS, Beijersbergen RL, Bernards R (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428:** 431-437.

Castanotto D, Rossi JJ (2009) The promises and pitfalls of RNA-interference-based therapeutics. *Nature* **457:** 426-433.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494-498.

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346:** 818-822.

Erson AE, Petty EM (2008) MicroRNAs in development and disease. *Clin Genet* **74:** 296-306.

Farah MH (2007) RNAi silencing in mouse models of neurodegenerative diseases. *Curr Drug Deliv* **4:** 161-167.

Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC (1998) Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391:** 806-811.

Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F, Kay MA (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441:** 537-541.

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223-227.

Haney SA (2007) Increasing the robustness and validity of RNAi screens. *Pharmacogenomics* **8:** 1037-1049.

Henkin TM (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22:** 3383-3390.

Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* **24:** 545-554.

Keasling JD (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol* **3:** 64-76.

Kim DH, Behlke MA, Rose SD, Chang MS, Choi S, Rossi JJ (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol* **23:** 222-226.

Kim DH, Rossi JJ (2009) Overview of gene silencing by RNA interference. *Curr Protoc Nucleic Acid Chem* **Chapter 16:** Unit 16 11.

Klattenhoff C, Theurkauf W (2008) Biogenesis and germline functions of piRNAs. *Development* **135:** 3-9.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415-419.

Lilley DM (1999) Structure, folding and catalysis of the small nucleolytic ribozymes. *Curr Opin Struct Biol* **9:** 330-338.

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U (2004) Nuclear export of microRNA precursors. *Science* **303:** 95-98.

Majdalani N, Vanderpool CK, Gottesman S (2005) Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* **40:** 93-113.

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123:** 607-620.

Mullenders J, Fabius AW, Madiredjo M, Bernards R, Beijersbergen RL (2009) A large scale shRNA barcode screen identifies the circadian clock component ARNTL as putative regulator of the p53 tumor suppressor pathway. *PLoS ONE* **4:** e4798.

Rand TA, Petersen S, Du F, Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123:** 621-629.

Sahu NK, Shilakari G, Nayak A, Kohli DV (2007) Antisense technology: a selective tool for gene expression regulation and gene targeting. *Curr Pharm Biotechnol* **8:** 291-304.

Savage DF, Way J, Silver PA (2008) Defossiling fuel: how synthetic biology can transform biofuel production. *ACS Chem Biol* **3:** 13-16.

Scherer LJ, Rossi JJ (2003) Approaches for the sequence-specific knockdown of mRNA. *Nat Biotechnol* **21:** 1457-1465.

Shapiro BA, Yingling YG, Kasprzak W, Bindewald E (2007) Bridging the gap in RNA structure prediction. *Curr Opin Struct Biol* **17:** 157-165.

Sharp PA (1999) RNAi and double-strand RNA. *Genes Dev* **13:** 139-141.

Shih IH, Been MD (2002) Catalytic strategies of the hepatitis delta virus ribozymes. *Annu Rev Biochem* **71:** 887-917.

Silva JM, Li MZ, Chang K, Ge W, Golding MC, Rickles RJ, Siolas D, Hu G, Paddison PJ, Schlabach MR, Sheth N, Bradshaw J, Burchard J, Kulkarni A, Cavet G, Sachidanandam R, McCombie WR, Cleary MA, Elledge SJ, Hannon GJ (2005) Second-generation shRNA libraries covering the mouse and human genomes. *Nat Genet* **37:** 1281-1288.

Singh SK (2008) RNA interference and its therapeutic potential against HIV infection. *Expert Opin Biol Ther* **8:** 449-461.

Siolas D, Lerner C, Burchard J, Ge W, Linsley PS, Paddison PJ, Hannon GJ, Cleary MA (2005) Synthetic shRNAs as potent RNAi triggers. *Nat Biotechnol* **23:** 227-231.

Suess B, Weigand JE (2008) Engineered riboswitches: overview, problems and trends. *RNA Biol* **5:** 24-29.

Takeshita F, Ochiya T (2006) Therapeutic potential of RNA interference against cancer. *Cancer Sci* **97:** 689-696.

Tang W, Dodge M, Gundapaneni D, Michnoff C, Roth M, Lum L (2008) A genome-wide RNAi screen for Wnt/beta-catenin pathway components identifies unexpected roles for TCF transcription factors in cancer. *Proc Natl Acad Sci U S A* **105:** 9697-9702.

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249:** 505-510.

Werstuck G, Green MR (1998) Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **282:** 296-298.

Whitehead KA, Langer R, Anderson DG (2009) Knocking down barriers: advances in siRNA delivery. *Nat Rev Drug Discov* **8:** 129-138.

Winkler WC (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr Opin Chem Biol* **9:** 594-602.

Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17:** 3011-3016.

Chapter 2

Design of Ligand-Responsive Small Interfering RNAs

ABSTRACT

Small interfering (si)RNAs are duplexed RNAs of 19 base pairs that induce potent and specific gene silencing in metazoans through the RNA interference pathway. We have designed siRNAs, which we call siSwitches, that activate gene silencing only in the presence or absence of a selected ligand. Rational integration of siRNAs and ligand-binding aptamers imparted tunable ligand control of siRNA formation and Dicer processing in vitro. Surprisingly, siSwitches produced by T7 transcription and not chemical synthesis induced potent silencing in mammalian cells even when one of the two siRNA strands were absent, suggesting the presence of spurious T7 transcription products that activate gene silencing. Plasmid-based expression of siRNAs with 25 base pairs from U6 promoters induced greater silencing than the traditional siRNA counterparts, supporting further research in the endogenous expression of siSwitches toward applications in advanced genetic screens and gene-specific therapeutics.

INTRODUCTION

RNA interference (RNAi) is a revolutionary genetic tool that enacts targeted post-transcriptional gene silencing in most organisms including humans. Applied double-stranded or hairpin RNAs are processed by the RNase III enzyme Dicer to 19-base pair duplexes termed small interfering (si)RNAs (Bernstein *et al*, 2001). One of the two siRNA strands termed the guide strand is retained by the RNA induced silencing complex (RISC) and serves as a template for the targeting of transcripts with complementary sequences. Researchers have implemented siRNAs with duplexes of 19 base pairs or 27 base pairs that respectively bypass or undergo Dicer processing to enact gene silencing (Elbashir *et al*, 2001; Kim *et al*, 2005). Simple base pairing interactions between the guide strand and transcript dictate target recognition, lending to the rational design of siRNAs that specifically target any single expressed gene (Tilesi *et al*, 2009). In turn, siRNAs have been used as the basis for high-throughput genetic screens (Echeverri and Perrimon, 2006) and emerging treatment strategies to quell viral infection and genetic diseases such as cancer (Aagaard and Rossi, 2007).

The regulatory capabilities of siRNAs represent only one of many unique qualities afforded to RNA. As a second example, RNAs termed aptamers can bind diverse molecules or ligands with high affinity and specificity. These molecules can be selected through a standard in vitro procedure (Ellington and Szostak, 1990; Tuerk and Gold, 1990), resulting in the identification of aptamer sequences against proteins, metals, metabolites, and small molecule drugs. Utilizing these versatile recognition properties, researchers have integrated aptamers into multifunctional RNAs that activate or repress gene expression in the presence or absence of the recognized ligand (Suess and Weigand,

2008). Researchers have exploited a diversity of regulatory mechanisms in bacteria, yeast, and mammalian cells in the design of these multifunctional RNAs, although none to-date have integrated aptamers and siRNAs for ligand control of RNAi.

In this report, we describe a strategy to engineer ligand control of RNAi through the rational combination of siRNAs and aptamers. The modular and tunable designs are based on a previous framework developed in our group for ligand control of eukaryotic gene expression through antisense inhibition (Bayer and Smolke, 2005). Under our design, ligand binding controls whether or not the siRNA can undergo further processing to activate RNAi. By coupling in vitro aptamer selection technologies and rational siRNA design, virtually any gene can be controlled by a diversity of molecules located in the intracellular environment.

RESULTS AND DISCUSSION

We devised two design schemes that rationally combine aptamers and siRNAs to introduce ligand control of siRNA processing and subsequent RNAi-mediated gene silencing (Figure 2.1). We call the resulting regulatory molecules small interfering switches or siSwitches. In both design schemes, hybridization of the guide strand and the passenger strand form the siRNA that undergoes Dicer processing. Ligand binding to the aptamer stabilizes one of two conformations that either expose or sequester the passenger strand to promote (OFF behavior) or inhibit (ON behavior) siRNA formation, respectively. In the scheme for ON behavior, the aptamer is flanked by the siRNA passenger strand and a competing strand complementary to the passenger strand (Figure 2.1A). The competing strand is intentionally shorter than the passenger strand to avoid

Dicer recognition and processing. In the absence of ligand, the guide strand preferentially displaces the competing strand and hybridizes to the passenger strand, thereby forming an siRNA. Ligand binding stabilizes the aptamer stem, which coincides with base pairing between the passenger strand and competing strand. Thus, the addition of ligand is anticipated to increase target gene levels by inhibiting siRNA formation and subsequent Dicer processing.

In the scheme for OFF behavior, the passenger strand and the aptamer are separated by an intervening sequence complementary to the passenger strand (Figure 2.1B). In the absence of ligand, the intervening sequence base pairs to the passenger strand, sequestering it from guide strand hybridization. Ligand binding to the aptamer displaces the passenger strand, allowing guide strand hybridization, siRNA formation, and subsequent gene silencing. Overall, the two schemes allow either activation or repression of target gene levels in the presence of the cognate ligand by controlling the activity of RNAi.

An important design consideration for both ON and OFF behavior is the number of base pairs in the formed siRNA. In mammals, double-stranded RNAs consisting of more than 30 base pairs efficiently activate the interferon response, which shuts down global protein synthesis. To maintain targeted gene silencing through RNAi, the formed siRNA should be less than the cutoff of 30 base pairs.

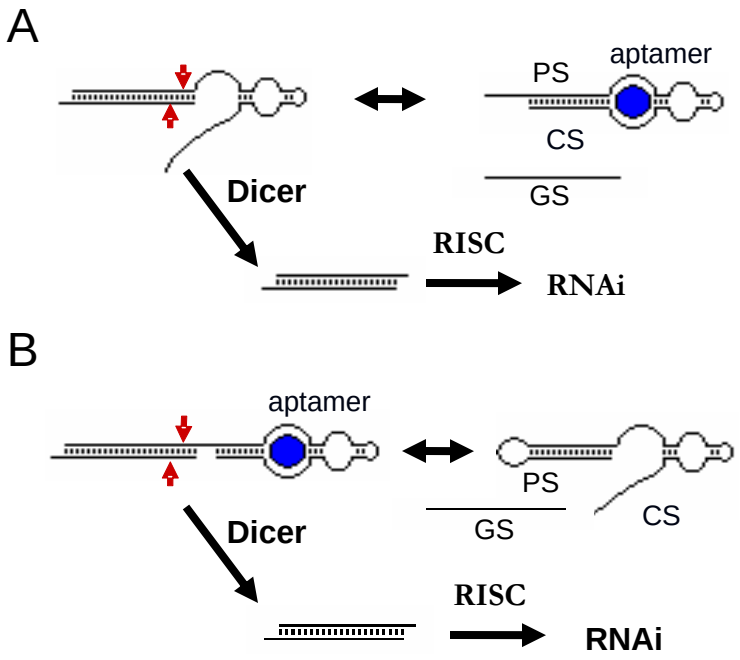


Figure 2.1 Design schemes for ligand-regulated siRNAs or siSwitches that rely upon the necessity of a 2-nt 3'overhange for efficient Dicer processing. Designs combine two strands of RNA. One encodes the guide strand (GS) and the other encodes the passenger strand (PS), a ligand-binding aptamer, and a competing strand (CS) complementary to the passenger strand. Ligand binding (blue circle) stabilizes an aptamer-formed conformation such that the guide strand can or cannot hybridize. Hybridization produces an siRNA with the necessary overhang that can undergo Dicer processing, subsequently activating gene silencing. Designs exhibiting ON behavior (**A**) or OFF behavior (**B**) as designated by the expression level of the target gene in the presence of the ligand. Red arrows designate Dicer cleavage sites.

We first evaluated the behavior of siSwitches in vitro (Figure 2.2). An initial siSwitch (S1) was constructed following the design scheme for ON behavior. S1 contains a passenger strand matching the fluorescent protein encoding gene *gfp*, an in vitro-selected aptamer against the small molecule theophylline (Jenison *et al*, 1994; Zimmermann *et al*, 2000), and a competing strand that forms 19 base pairs with the passenger strand (Figure 2.2A). To assess the impact of ligand binding on siRNA formation and Dicer processing in vitro, T7-transcribed and 5'-radiolabeled S1 was incubated with a 2-fold molar excess of an unlabeled 28-nt guide strand (aS1) and varying concentrations of theophylline in the presence or absence of Dicer. RNAs were then resolved by nondenaturing PAGE (Figure 2.2B). In the absence of Dicer, S1 predominantly hybridized with aS1 to form an siRNA and theophylline addition partially inhibited siRNA formation. The presence of Dicer removed all traces of full-length S1 even in the presence of theophylline. These results suggest that theophylline addition inhibited siRNA formation in line with the predicted behavior for ON-acting siSwitches, although irreversible Dicer processing drove the partitioning toward siRNA formation until the entire S1 pool had been cleaved.

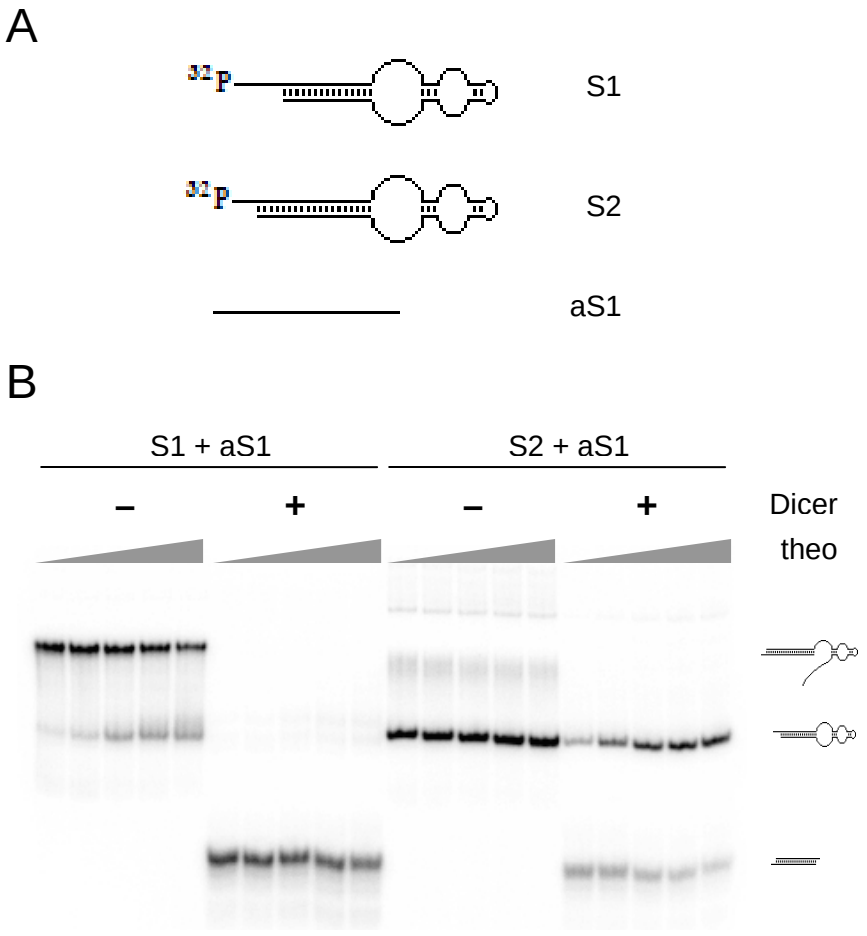


Figure 2.2 Length of the competing stem affects the extent of binding and Dicer processing in vitro. (**A**) Schematic of two ON-behaving long siRNAs with competing strand lengths of 19-nt (S1) and 23-nt (S2) with radiolabeled 5' phosphates. The antisense strand was unlabeled in these experiments. (**B**) Non-denaturing gel results of both sense strands incubated with a 200% molar excess of the antisense strand in the presence (+) or absence (–) of Dicer and varying concentrations of theophylline (theo). Specific theophylline concentrations used (mM): 0.01, 0.1, 1, 10, 25. Gel results are representative of two independent experiments.

The substantial processing even in the presence of theophylline highlights the irreversibility of the Dicer processing step, which may lead to extensive gene silencing in vivo for ON behavior even in the presence of ligand. Therefore, it would be beneficial to tune the extent of processing to manipulate the relationship between ligand concentration and target gene levels. One tuning strategy was suggested by two reports of synthetic RNA-based gene regulation, where the complementarity between hybridizing RNAs was modified to modulate the energetics of complex formation (Bayer and Smolke, 2005; Isaacs *et al*, 2004). We adopted a similar strategy by extending the competing strand of S1 to form 19 base pairs with the passenger strand (Figure 2.2A). The designed siSwitch (S2) was subjected to the in vitro processing assay to compare theophylline control of siRNA formation and Dicer processing to that of its counterpart S1. In vitro results showed that S2 was less prone to form an siRNA and Dicer processing was more sensitive to theophylline addition. By extending the competing strand, siRNA formation was less thermodynamically favorable, providing fewer formed siRNAs for Dicer processing. Modulating the competing strand length can be considered a general design strategy to tailor siSwitch behavior to in vivo silencing requirements.

We next explored the behavior of siSwitches in vivo through two modes of delivery: transfection and endogenous expression. Transfection was initially evaluated based on the predominance of this mode of siRNA delivery in both basic research and disease treatment applications. Transfected siRNAs can be generally produced through chemical synthesis or in vitro transcription, where the latter is more economically viable to mass produce RNAs greater than 50 nts. However, delivery of T7-transcribed RNAs has been shown to induce an innate immune response through PKR recognition of the 5’

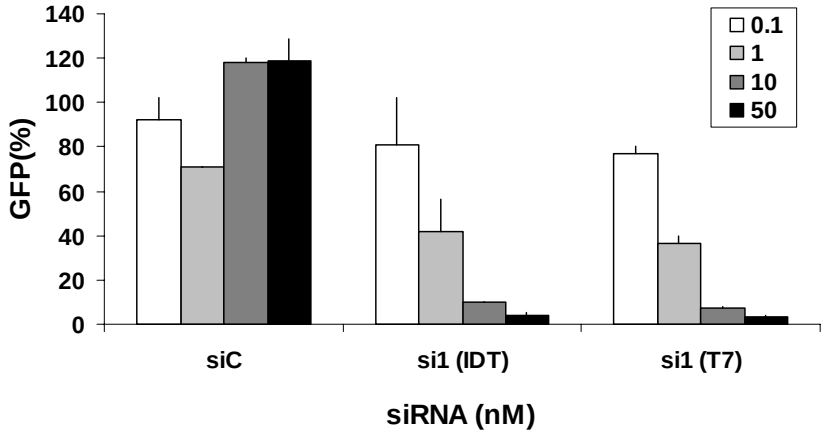
triphosphate (Kim *et al*, 2004). To initially test the silencing capacity of T7-transcribed RNAs through RNAi, we transfected cells with chemically-synthesized or T7-transcribed siRNAs designed to target GFP. 5' phosphates on T7-transcribed siRNAs were removed by phosphatase treatment. Both GFP-targeting siRNAs effectively reduced fluorescence levels, which were insensitive to a T7-transcribed siRNA with a scrambled guide strand (Figure 2.3A).

We anticipated that siSwitches should show the same targeting specificity as the siRNAs when both siRNA strands are present and designed to target GFP. Surprisingly, transfection of the T7-transcribed passenger strands or the guide strands for both ON- and OFF-behaving designs significantly silenced GFP levels (Figure 2.3B). Silencing was not observed for a scrambled guide strand (aSC), suggesting that silencing was sequence-specific. To evaluate whether silencing was linked to T7 transcription, we transfected cells with a T7-transcribed or chemically synthesized guide strand (aS1) targeting GFP and measured relative GFP levels. Only the T7-transcribed guide strand induced GFP silencing, eliminating the possibility of antisense inhibition. One interpretation is that T7 transcription for siSwitches produces spurious products that activate RNAi, either as separate species or extensions of the templated RNAs. These spurious products must be removed to allow ligand control of RNAi-mediated silencing by siSwitches produced through T7 transcription. Alternatively, chemical synthesis may become a viable option with further improvements in RNA synthesis technologies and corresponding cost reductions.

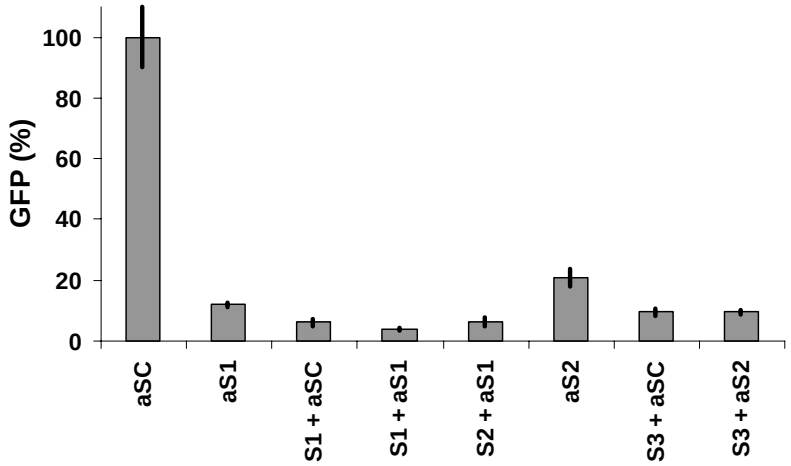
An alternative possibility is endogenously expressing siSwitches in cells to introduce ligand control of gene silencing. Previous work showed that plasmid-based

expression of individual siRNA strands from U6 promoters could significantly silence a target gene through RNAi (Miyagishi and Taira, 2002; Wu *et al*, 2005; Yu *et al*, 2002). Only siRNA duplexes with 19 base pairs were tested, where siRNAs of this length do not undergo Dicer processing. More recent work showed that delivery of longer siRNAs improved silencing efficiency (Kim *et al*, 2005). These longer siRNAs undergo Dicer processing, resulting in rapid loading of the guide strand into RISC (Gregory *et al*, 2005; Maniataki and Mourelatos, 2005). Since these longer siRNAs have not been tested under endogenous expression and form the basis of our siSwitch designs, we evaluated the improved silencing efficiency of longer siRNAs under endogenous expression. The strands of GFP-targeting siRNAs with 19 or 26 base pairs were encoded in plasmids downstream of U6 promoters. The resulting plasmids were cotransfected into HeLa cells with a GFP reporter plasmid. The relative fluorescence showed a much larger decrease in GFP for cells transfected with the longer siRNAs, supporting the improved silencing efficiency of longer siRNAs (Figure 2.4).

A


140
120
100
80
60
40
20
0
GFP(%)
0.1
1
10
50
siC
si1 (IDT)
si1 (T7)
siRNA (nM)

B


100
80
60
40
20
0
GFP (%)
aSC
aS1
S1 + aSC
S1 + aS1
S2 + aS1
aS2
S3 + aSC
S3 + aS2

C

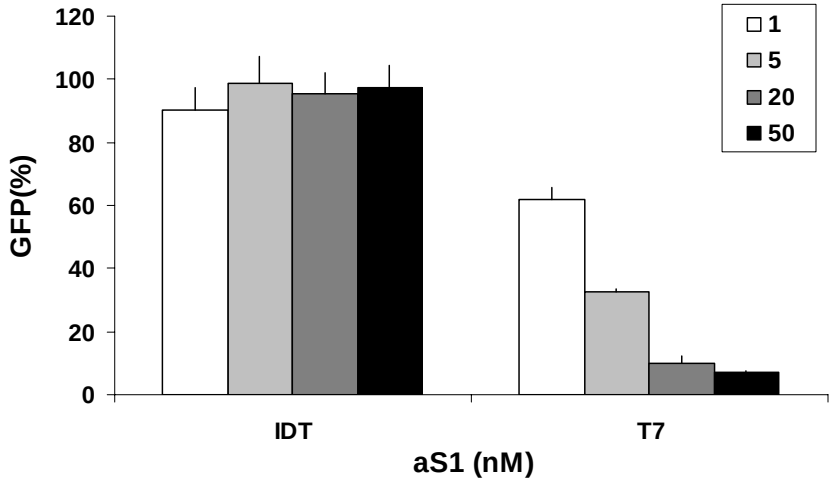

120
100
80
60
40
20
0
GFP(%)
1
5
20
50
IDT
T7
aS1 (nM)

Figure 2.3 Relative EGFP knockdown by T7-transcribed siRNAs in HEK293T tTA-d2EGFP. (**A**) T7-transcribed siRNAs produce sequence-specific knockdown of EGFP. Compared to untransfected cells, a scrambled 25-bp siRNA (siC) produces negligible knockdown, whereas chemically-synthesized (IDT) and T7-transcribed (T7) 25-bp siRNAs (si1) produce dose-dependent knockdown of fluorescence. (**B**) Individual molecules of T7-transcribed ligand-regulated siRNAs produce sequence-specific knockdown of EGFP. Individual molecules were transfected at a concentration of 10 nM, where trials with two RNA strands had a total RNA concentration of 20 nM. Sc, Scrambled RNA strand; aS1, antisense strand for ON behavior; S1, siSwitch for ON behavior containing sense strand and theophylline aptamer; aS2, antisense strand for OFF behavior; S3, sense strand for OFF behavior. (**C**) T7-transcribed antisense RNAs produce sequence-specific knockdown. Chemically-synthesized antisense RNA (IDT) produced no knockdown compared to untransfected cells, while T7-transcribed antisense RNAs produced dose-dependent knockdown of EGFP. Error bars for all three graphs represent three independent measurements.

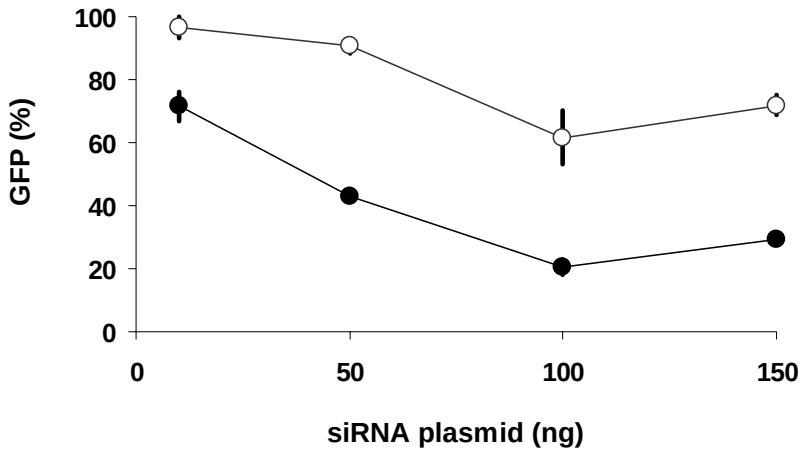


Figure 2.4 Endogenously-expressed long siRNAs induce greater EGFP knockdown than regular siRNAs. Both strands of 19 bp (white) or 25 bp siRNAs (black) targeting EGFP were placed downstream of a U6 promoter encoded in a single plasmid. Different amounts of the siRNA-encoding plasmids were cotransfected with 10 ng of an EGFP-expressing reporter plasmid into HeLa cells. Measured fluorescence was normalized to the fluorescence of cells transfected with the corresponding amount of an empty plasmid. Error bars represent the standard deviation of two independent transfections.

Future work should focus on testing both ON and OFF designs for siSwitches under endogenous expression. Plasmids encoding siSwitches could be transfected in a cell line stably expressing GFP to facilitate the evaluation of siSwitch activity. Once a base assay is developed, aptamer swapping and performance tuning should be evaluated to ascertain how well siSwitches can be tailored to meet any regulatory demands. siSwitch performance should also be compared to the other designs documented in Chapters 3 and 4 to assess which engineered mode of ligand control of RNAi is best for any desired application.

MATERIALS AND METHODS

T7 transcription. All reagents and solutions for T7 transcription were RNase-free. Shorter DNA templates for the T7 transcription reaction were synthesized (IDT) and annealed. For longer templates, the top and bottom strands were ordered truncated at the 3' end. Full-length templates were produced by annealing these strands, adding Klenow (NEB), the supplied buffer, and enough dNTPs for a final concentration of 0.33 mM and incubated at room temperature for 30 mins. Templates were then ethanol-precipitated prior to the T7 transcription reaction. The sequence of the final template contained the T7 promoter sequence (TTCTAATACGACTCACTATAG) at the 5' end. The final sequences of synthesized or transcribed RNAs are reported in Table 2.1. RNA was transcribed using the Ampliscribe T7 transcription kit (Epicentre) according to the manufacturer's instructions. To remove 5' phosphates, 1 U calf intestinal phosphate (NEB) was added to the transcription reaction, followed by incubation at 37 ºC for 1 hr. Unincorporated ribonucleotides were removed using NucAway spin columns (Ambion) and the transcribed RNA was further purified by PAGE. RNA concentration was estimated from the measured absorbance at 260 nm using a DU530 spectrophotometer (Beckman Coulter).

In vitro Dicer assay. Transcribed RNA was 5' labeled by incubating the RNA with γ-[^{32}P]-ATP and T4 polynucleotide kinase (NEB) for 30 min at 37 ºC and purified using the NucAway spin column. Approximately 1 nmol of the labeled RNA was incubated for 16 hrs in 1X Dicer reaction buffer (20 mM Tris-HCl, 150 mM NaCl, 2.5 mM $MgCl_2$, pH 8.0) with varying concentrations of theophylline in the presence or absence of 0.5 U of

Dicer (Stratagene) for a total volume of 10 µl. After the incubation, each sample was combined with 2 µl of 6X loading buffer (40% sucrose), resolved on a 12% non-denaturing polyacrylamide gel, and imaged on an FX phosphorimager (BioRad).

Plasmid construction. The endogenously-expressed siRNAs were encoded in pcDNA3.1(+) (Invitrogen), where the two siRNA strands were separately placed downstream of U6 promoters. The U6 promoters were cloned into BglII/MfeI and SpeI/XhoI of pcDNA3.1(+) after amplification from pSilencer 2.1 puro. Each siRNA strand was encoded in a 5' tail of the reverse oligo primer. sU6.fwd 5'-AATAAGATCTCCCCAGTGGAAAGACGCGCA-3', aU6.fwd 5'- AATAACTAGTCCCCAGTGGAAAGACGCGCA-3', 19nt-sU6.rev 5'-AATACAATTGAAAAAAGATGAACTTCAGGGTCAG**C**GGATCCCGCGTCCTTTCCACA-3', 19nt-aU6.rev 5'-AATACTCGAGAAAAAAGCTGACCCTGAAGTTCAT**C**GGATCCCGCGTCCTTTCCACA-3', 25nt-sU6.rev 5'-AATACAATTGAAAAAATGCAGATGAACTTCAGGGTCAG**C**GGATCCCGCGTCCTTTCCACA-3', 25nt-aU6.rev 5'-AATACTCGAGAAAAAAGCTGACCCTGAAGTTCATCTGCATTT**C**GGATCCCGCGTCCTTTCCACA-3'. Underlined text designates restriction sites while each bolded 'C' marks the start of transcription from the U6 promoter. All restriction enzymes and T4 DNA ligase were purchased from NEB. All constructs were sequence-verified (Laragen).

Cell culture and transfection. HEK293T and HeLa cells were maintained in minimal essential medium alpha media (Invitrogen) with 10% fetal bovine serum (Invitrogen) at 37°C in a 5% CO_2-humidified incubator. Cells were transfected 1 day after seeding with

either plasmid DNA, T7-transcribed RNA, or chemically-synthesized RNA (IDT) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions. pcDNA3.1(+) encoding EGFP was transfected to evaluate siRNA potency. pSilencer 2.1 puro served as a negative control. One day post-transfection, the media was replaced.

Quantification of fluorescence. Three days post-transfection, the media was removed from each well and 100 µl (for 48-well plate) of M-PER Mammalian Protein Extraction Reagent (Pierce) was added to each well. After 5 min of agitation, each well was scraped with a pipette tip and the lysed cells were transferred to a 1.5 ml microcentrifuge tube. Following a 10 min spin at 4 ºC and 14K rpm, the supernatant was transferred to a fresh tube and assayed for total protein using the BCA kit (Sigma). The remaining supernatant was transferred to a 386-well plate and assayed for fluorescence using a Safire (Tecan) fluorescent plate reader set to the appropriate excitation (EGFP, 485 nm) and emission (EGFP, 515 nm) wavelengths. Sample fluorescence was normalized to total protein levels then normalized such that 100% represents relative GFP levels for cells transfected with the negative control plasmid.

One drawback to this procedure is that the measured fluorescence is a population average; hence confounding factors such as variable transfection efficiency bias the fluorescent data. To correct for this, flow cytometry should be used in the future, since the effects of transfection efficiency can be ascertained and removed from the data.

Table 2.1 Sequence of synthesized and T7-transcribed RNAs. Sequences are written 5' to 3'. The theophylline aptamer is indicated in blue text. The guide strand (GS) and passenger strand (PS) are shown for the siRNAs tested in Figure 2.3.

Name	Sequence
S1	GCUGACCCUGAAGUUCAUCUGCGGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCGCAGAUGAACUUCAG
S2	GCUGACCCUGAAGUUCAUCUGCGGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACCGCAGAUGAACUUCAGGGUC
S3	GCUGACCCUGAAGUUCAUCUGCGAACUUCAGGGUCAGCGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCGCUGACCCUU
aS1	GCACCGCAGAUGAACUUCAGGGUCAGCU
aS2	GUUCGCAGAUGAACUUCAGGGUCAGCU
aSC	GUUGCACUGGCUCUACAACUAGACCGCA
si1 (GS)	GGUGCAGAUGAACUUCAGGGUCAGCUU
si1 (PS)	GCUGACCCUGAAGUUCAUCUGCACC
siC (GS)	GUUGCACUGGCUCUACAACUAGACCGC
siC (PS)	GGUCUAGUUGUAGAGCCAGUGCAAC

REFERENCES

Aagaard L, Rossi JJ (2007) RNAi therapeutics: principles, prospects and challenges. *Adv Drug Deliv Rev* **59:** 75-86.

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23:** 337-343.

Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409:** 363-366.

Echeverri CJ, Perrimon N (2006) High-throughput RNAi screening in cultured cells: a user's guide. *Nat Rev Genet* **7:** 373-384.

Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411:** 494-498.

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346:** 818-822.

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123:** 631-640.

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22:** 841-847.

Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-resolution molecular discrimination by RNA. *Science* **263:** 1425-1429.

Kim DH, Behlke MA, Rose SD, Chang MS, Choi S, Rossi JJ (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat Biotechnol* **23:** 222-226.

Kim DH, Longo M, Han Y, Lundberg P, Cantin E, Rossi JJ (2004) Interferon induction by siRNAs and ssRNAs synthesized by phage polymerase. *Nat Biotechnol* **22:** 321-325.

Maniataki E, Mourelatos Z (2005) A human, ATP-independent, RISC assembly machine fueled by pre-miRNA. *Genes Dev* **19:** 2979-2990.

Miyagishi M, Taira K (2002) U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat Biotechnol* **20:** 497-500.

Suess B, Weigand JE (2008) Engineered riboswitches: overview, problems and trends. *RNA Biol* **5:** 24-29.

Tilesi F, Fradiani P, Socci V, Willems D, Ascenzioni F (2009) Design and validation of siRNAs and shRNAs. *Curr Opin Mol Ther* **11:** 156-164.

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249:** 505-510.

Wu MT, Wu RH, Hung CF, Cheng TL, Tsai WH, Chang WT (2005) Simple and efficient DNA vector-based RNAi systems in mammalian cells. *Biochem Biophys Res Commun* **330:** 53-59.

Yu JY, DeRuiter SL, Turner DL (2002) RNA interference by expression of short-interfering RNAs and hairpin RNAs in mammalian cells. *Proc Natl Acad Sci U S A* **99:** 6047-6052.

Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6:** 659-667.

Chapter 3

Design of Ligand-Responsive Small Hairpin RNAs

ABSTRACT

Progress in constructing biological networks will rely on the development of more advanced components that can be predictably modified to yield optimal system performance. We have engineered an RNA-based platform, which we call an shRNA switch, that provides for integrated ligand control of RNA interference (RNAi) by modular coupling of an aptamer, competing strand, and small hairpin (sh)RNA stem into a single component that links ligand concentration and target gene expression levels. A combined experimental and mathematical modeling approach identified multiple tuning strategies and moves towards a predictable framework for the forward design of shRNA switches. The utility of our platform is highlighted by the demonstration of fine-tuning, multi-input control, and model-guided design of shRNA switches with an optimized dynamic range. Thus, shRNA switches can serve as an advanced component for the construction of complex biological systems and offer a controlled means of activating RNAi in disease therapeutics.

INTRODUCTION

To maintain fitness under diverse conditions, biological systems must integrate multiple environmental cues (inputs) to determine appropriate phenotypic outcomes (outputs) over short and long time scales. This relationship, which can be interpreted as an input–output function or transfer function, is specified by the behavior of the individual system components and their network interactions. The complexity of natural biological systems, reflected by the sheer number of associated components and network interactions, can appear intractable to scientists and engineers seeking to understand and reliably construct biological systems.

Synthetic biological systems that perform information processing operations with specified transfer functions can be constructed through the design of either individual complex components encoding multiple integrated functionalities or simpler components assembled into networks with emergent properties. For example, the ultrasensitive switch behavior of the mitogen-activated protein kinase signaling cascade (Huang and Ferrell, 1996) has been replicated with one component (Dueber *et al*, 2007) or a network of components (Hooshangi *et al*, 2005). The majority of previously engineered biological systems have employed design strategies focused on the assembly of simpler components into networks (Elowitz and Leibler, 2000; Gardner *et al*, 2000). However, as biological engineers move towards constructing large-scale systems with more advanced behaviors, integration of complex components into network design will be critical, especially as current network design strategies do not effectively scale with system complexity (Croft *et al*, 2003). Furthermore, engineering of complex components and their integration into networks will facilitate the construction of advanced systems with a reduced number of

constituent parts. Such design strategies will result in a lower energetic load on the cell and will comply with size limitations associated with packaging synthetic systems for delivery applications (Flotte, 2000; Grieger and Samulski, 2005).

RNA is a rich biological substrate for engineering complex components (Isaacs *et al*, 2006), where scalable molecular information processing systems have been built in vitro from nucleic acid components (Stojanovic and Stefanovic, 2003; Seelig *et al*, 2006). Natural biological systems exhibit widespread utilization of regulatory RNAs in larger networks (Shalgi *et al*, 2007) and integrated functionalities by riboswitches (Sudarsan *et al*, 2003; Grundy and Henkin, 2006). The latter is an example of a complex RNA component that converts the intracellular concentration of a molecular signal into levels of a target protein. Building on these natural examples, researchers have integrated synthetic regulatory RNAs into larger networks (Deans *et al*, 2007; Rinaudo *et al*, 2007) and developed synthetic riboswitch elements (Isaacs *et al*, 2006; Suess andWeigand, 2008) to yield desired transfer functions. Synthetic riboswitches have been developed through the coupling of regulatory RNA elements to aptamers, sensory elements that bind specific ligands, to achieve in vivo ligand control of transcription (Buskirk *et al*, 2004), RNA stability (Win and Smolke, 2007), translation (Grate and Wilson, 2001; Suess *et al*, 2003; Bayer and Smolke, 2005; Lynch *et al*, 2007; Ogawa and Maeda, 2008; Wieland and Hartig, 2008), splicing (Thompson *et al*, 2002; Weigand and Suess, 2007), and RNA interference (RNAi) (An *et al*, 2006). As aptamers can be generated de novo to potentially any molecule (Lee *et al*, 2004; Gopinath, 2007) and regulatory RNAs can be designed to target practically any gene of interest, the potential exists to construct complex regulatory RNAs that respond to molecular inputs displaying low cytotoxicity,

tissue targeting, or endogenous production to impose desired phenotypic outcomes. In addition, the ability to alter or tune the component transfer function, which allows manipulation of component performance towards optimal system performance, has been demonstrated for in vivo (Suess *et al*, 2003; Isaacs *et al*, 2004; Bayer and Smolke, 2005; Win and Smolke, 2007) and in vitro systems (Hall *et al*, 2007) by modifying RNA folding energetics. However, broader implementation of these synthetic regulatory components has been limited, as most examples do not support domain swapping of different sensory and regulatory elements. Furthermore, the design of such regulatory systems has lacked predictive tools for the translation of sequence information into component transfer functions to enable in silico optimization of system behavior before construction.

We have developed a framework for the construction of shRNA switches that mediate ligand control of RNAi across diverse mammalian cell types. Our platform utilizes a strand displacement strategy, where the functions of ligand binding, RNAi activation (Kim and Rossi, 2008), and translation of the binding interaction into reduced processing by the RNAi machinery are isolated to individual domains, which increases the generality and ease of successful domain swapping and subsequent broad application. In addition, we systematically investigated tunability of the shRNA switch transfer function through a combined experimental and mathematical modeling approach that resulted in the identification of five tuning strategies. Standard RNA folding algorithms (Mathews *et al*, 2004) were used to establish a quantitative sequence-to-function relationship. Our efforts highlight the current limitations of these broadly used algorithms for the design of RNAs that function in vivo and offer a framework for optimizing

shRNA switch behavior in silico. By demonstrating combinatorial tuning strategies, multi-input control, and model guided forward design of shRNA switches with an optimized dynamic range within a specified context, we show that shRNA switches extend the utility of RNAi as a regulatory tool and are valuable components for the construction of complex biological systems.

RESULTS

Design and characterization of a modular shRNA platform

We engineered a complex RNAi substrate that encodes a ligand-controlled gene regulatory function by replacing the loop of a small hairpin (sh)RNA with two domains: an aptamer and a competing strand (Figure 3.1A). The shRNA switch molecule is designed to adopt distinct 'active' and 'inactive' conformations due to complementarity between the competing strand and the shRNA stem, similar to previously engineered RNA regulatory systems (Bayer and Smolke, 2005; Lynch *et al*, 2007; Win and Smolke, 2007). In the active conformation, irreversible processing by the RNAi machinery of the formed shRNA stem results in small interfering (si)RNA production and subsequent RNAi-mediated silencing of the target gene. In the inactive conformation, base-pairing by the competing strand disrupts the shRNA stem, which is predicted to inhibit processing by the RNAi machinery (Zeng and Cullen, 2004; Macrae *et al*, 2006). This base-pairing coincides with formation of the aptamer domain, such that ligand binding stabilizes the inactive conformation and indirectly reduces siRNA production, thereby linking intracellular ligand concentration to target protein levels through a component transfer function. To decrease the activation energy separating the two conformations, we

removed two nucleotides in the passenger strand, thereby mimicking the bulge from the microRNA (miRNA) mir-30a (Griffiths-Jones, 2004; Griffiths-Jones *et al*, 2006).

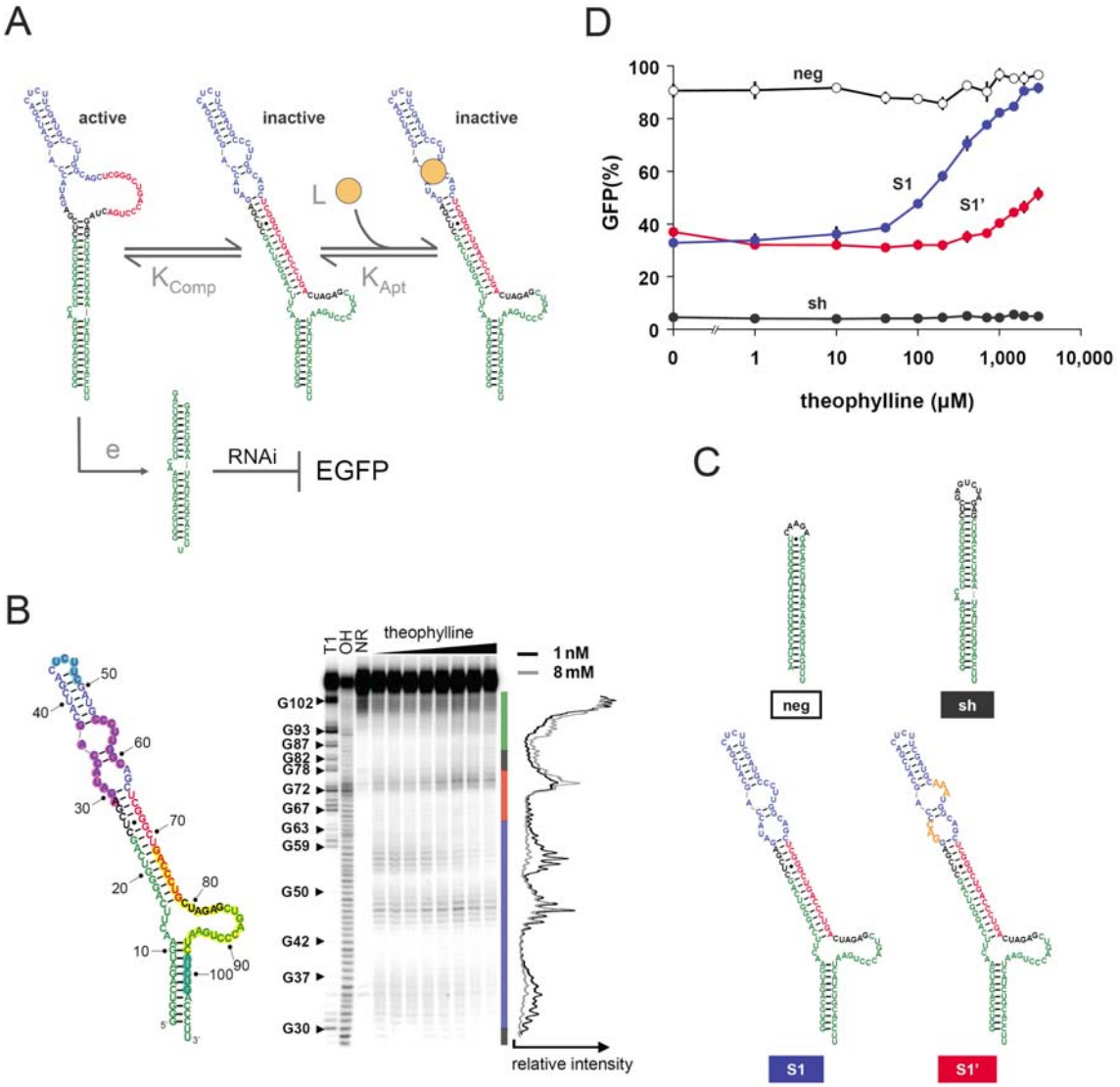


Figure 3.1 Design and characterization of an shRNA switch platform. Color schemes for switches shown in all figures are as follows: shRNA stem, green; aptamer domain, blue; competing strand, red; mutations to aptamer core, orange. (**A**) Sequence and representative structures of shRNA switch S1 and proposed mechanism for ligand control of RNAi-mediated gene silencing. K_{Comp}, K_{Apt}, and e are parameters from the mathematical model; L denotes ligand.

(**B**) In-line probing of S4t under the following theophylline concentrations (μM): 0.001, 0.01, 0.1, 1, 10, 100, 1000, and 8000. S4t was also resolved as unreacted (NR), partially digested with the G-specific RNase T1 (T1), and under basic conditions (OH). The included secondary structure of S4t is representative of the inactive conformation. Band quantification (right) is aligned with the resolved gel image. Nucleotides undergoing constant (●), increased (●), or decreased (●) cleavage in the presence of theophylline are shown. (**C**) Sequence and representative structure of shRNA switch S1 in the inactive conformation and associated controls. (**D**) Component transfer functions of S1 and switch controls. Dependence of GFP levels on theophylline concentration for HEK293T tTA-d2EGFP cells transfected with plasmids harboring the indicated constructs in the presence of varying theophylline concentrations. Median fluorescence values from flow cytometry analysis were normalized to that of untransfected cells in the same well. Error bars represent one standard deviation from duplicate transfected wells.

The three domains that comprise an shRNA switch perform distinct functions: the shRNA stem encodes the guide strand that activates RNAi-mediated silencing of the target gene, the aptamer detects the molecular input concentration through a ligand-binding interaction, and the competing strand translates the binding interaction into a decrease in regulatory activity by affecting processing by the RNAi machinery. On the basis of the action of the competing strand that is complementary to the shRNA stem, the sequences of the shRNA stem and aptamer domains are independent of one another. Therefore, the shRNA stem and aptamer domains can be independently modified without altering the functionality of the opposing domain or requiring sequence reassignment.

We designed an initial shRNA switch (S1) to target EGFP and respond to theophylline by incorporating an EGFP-targeting guide strand and the theophylline

aptamer (Zimmermann *et al*, 2000) into our switch platform (Figure 3.1A). We used in-line probing (Soukup and Breaker, 1999) to assess the structural characteristics of a T7-transcribed variant similar to S1 (S4t; Figure 3.1B). In-line probing provides information on structural changes within the molecule as a result of theophylline binding from the ligand dependence of spontaneous RNA cleavage. Theophylline-dependent changes in cleavage rates at individual nucleotides were observed in the aptamer domain, competing strand, and the downstream shRNA stem sequence. The results suggest that theophylline binding promotes structural changes in the shRNA switch as expected for dynamic RNAs undergoing ligand-dependent conformational switching. The apparent dissociation constant (K_D) of ~5 μM, which was determined by quantifying the cleavage products at two positions (Supplementary Figure S3.1), is an order of magnitude larger than that of the aptamer alone ($K_D \approx 0.29$ μM) (Zimmermann *et al*, 2000). The observed increase in K_D is in agreement with our proposed mechanism (see below), where only the inactive conformation provides a formed aptamer that can bind ligand. As shRNA switches can occupy both conformations, the apparent affinity will be lower because ligand can only bind the inactive conformation that is transiently present in a fraction of the shRNA switch population.

The functionality of shRNA switches was assessed in mammalian cell culture. We transiently transfected plasmids harboring S1 and various switch controls transcribed from a U6 promoter into HEK293T cells stably expressing EGFP (Abbas-Terki *et al*, 2002). Flow cytometry analysis revealed that S1 elicits intermediate knockdown of EGFP as compared to the original shRNA targeting EGFP (sh) and a scrambled shRNA (neg) (Figure 3.1C and D), where the observed silencing by S1 can be attributed to activation

of RNAi on the basis of antisense inhibition of guide strand activity (Hutvagner *et al*, 2004; Meister *et al*, 2004) (Supplementary Figure S3.2). In the presence of theophylline, GFP levels increased in a dose-dependent manner for S1 but not for the control shRNAs. The effective concentration to achieve 50% activity (EC_{50}) for S1 of ~300 mM was much larger than the K_D of 5 μM measured in vitro, which can be primarily attributed to a concentration drop in theophylline across the cellular membrane (Koch, 1956) (J Liang, J Michener, C Smolke, unpublished data, 2008). Mutating the aptamer core of S1 (S10) greatly reduced the observed theophylline dependence without perturbing basal expression levels. We attribute the minor increase in GFP levels at high theophylline concentrations for S10 to pleiotropic effects of the ligand (An *et al*, 2006) and potentially to reduced binding capability of the mutated aptamer. Taken together, S1 links theophylline concentration to GFP levels in vivo through a relationship described by a component transfer function (Figure 3.1D). We obtained qualitatively similar results when shRNA switches targeting EGFP were transiently transfected into other cell lines (Supplementary Figure S3.3), suggesting that shRNA switches can be broadly applied in different cell lines and types.

Mathematical modeling offers tuning parameters to predictively modulate component transfer functions

Previous switch platforms utilizing strand displacement strategies have demonstrated tuning on the basis of aptamer swapping and modulation of folding energetics (Bayer and Smolke, 2005; Win and Smolke, 2007). We systematically evaluated the tuning capabilities of shRNA switches with the aid of a mathematical

model relating ligand concentration and target gene expression levels. Standard model parameters were incorporated to represent each chemical step from our proposed mechanism (Supplementary text S3.1). We assumed that the two adopted conformations are at thermodynamic equilibrium, that ligand only binds the inactive conformation, and that the active conformation is solely processed to an siRNA with a reduced efficiency as compared to the original shRNA. These assumptions yield the following relationship between relative expression levels of the target gene (f; output) and exogenous ligand concentration (L; input):

$$f = 1 - e \cdot f_{shRNA}\left[1 + K_{Comp}\left(1 + K_{Apt} \cdot L\right)\right]^{-h}, \quad (3.1)$$

where e is the processing efficiency, f_{shRNA} is the relative knockdown achieved by the original shRNA (sh), K_{Comp} is the partitioning constant between active and inactive conformations (K_{Comp} = [inactive]/[active]), K_{Apt} is the association constant for binding between ligand and the formed aptamer, and h is the Hill coefficient to account for non-linearity between siRNA concentration and target expression levels. Although mathematical models have been developed for RNAi (Raab and Stephanopoulos, 2004; Bartlett and Davis, 2006; Malphettes and Fussenegger, 2006), our approach utilizes a minimal parameter set that is experimentally tractable, fully represents RNAi in the context of shRNA switches, and captures the steady-state behavior of our system (Supplementary Figure S3.4). For one shRNA stem sequence and input ligand (fixed f_{shRNA}, h), our model provides three tuning parameters that can be varied to tune the component transfer function: K_{Comp}, K_{Apt}, and e (Figure 3.2A–C). Varying K_{Comp} results in a concomitant and opposing variation in EC_{50} and basal expression levels, which are independently tuned by K_{Apt} and e, respectively. In addition, as K_{Comp} approaches zero,

basal expression levels approach a lower limit that is dependent on the value of e and is higher than that of the original shRNA (Figure 3.2D). As each tuning parameter represents individual steps in the proposed mechanism, we examined how modifying the sequence in each domain, specifically the competing strand and aptamer domains, corresponds to parameter variation to identify unique tuning strategies (Figure 3.3A).

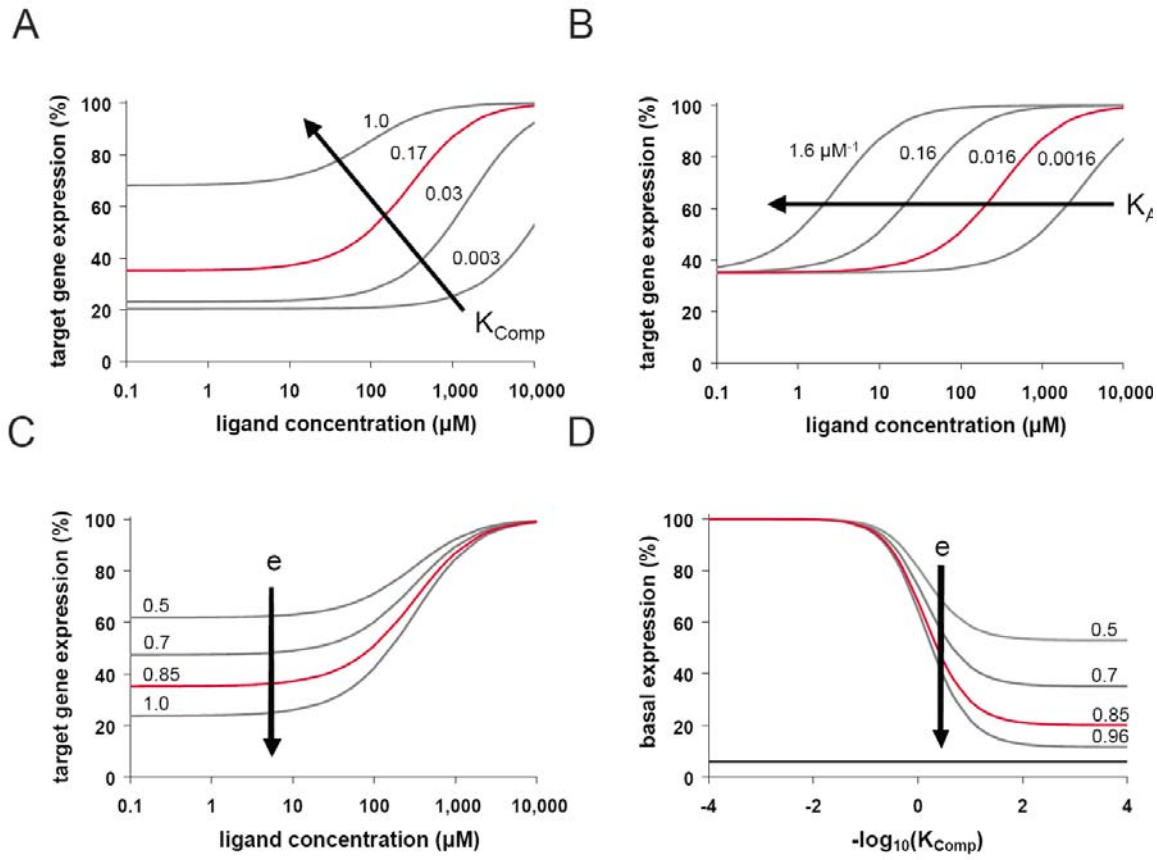


Figure 3.2 Model predicts tuning of the shRNA switch transfer function through variation of identified tuning parameters. Model predictions for the effect on the component transfer function of varying K_{Comp} (**A**), K_{Apt} (**B**), or e (**C**). (**D**) Effect of e on the dependence of basal expression levels on K_{Comp}. Minimal basal expression set by f_{shRNA} (—); the transfer function that fits the S1 theophylline response curve from Supplementary Figure S3.4 (—): K_{Comp} = 0.17, K_{Apt} = 0.016 μM^{-1}, e = 0.85, f_{shRNA} = 0.94, and h = 1.33.

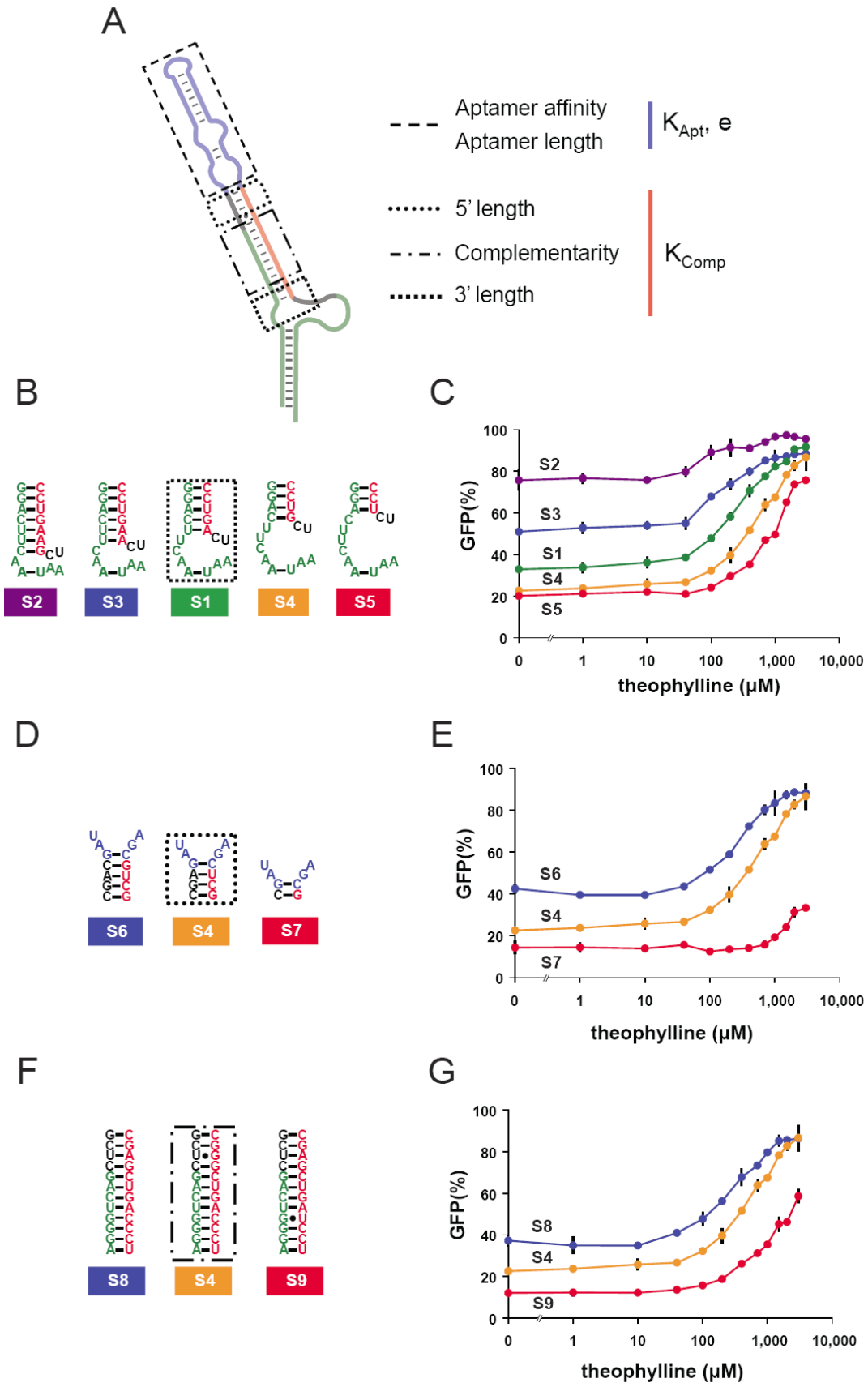

A
Aptamer affinity
Aptamer length
K_{Apt}, e
5' length
Complementarity
3' length
K_{Comp}
B
S2
S3
S1
S4
S5
C
GFP(%)
theophylline (μM)
S2
S3
S1
S4
S5
D
S6
S4
S7
E
GFP(%)
theophylline (μM)
S6
S4
S7
F
S8
S4
S9
G
GFP(%)
theophylline (μM)
S8
S4
S9

Figure 3.3 Experimental validation of competing strand tuning strategies. (**A**) Designated strategies for physical modulation of the tuning parameters. Three strategies pertain to the competing strand (—) and reflect changes in K_{Comp}, and two strategies pertain to the aptamer domain (—) and reflect changes in K_{Apt} and e. (**B–G**) Tuned theophylline response curves as described in Figure 3.1D and associated RNA sequences. Each family of curves represents iterative nucleotide modifications under a single tuning strategy within the competing strand: 3' end (**B**), 5' end (**D**), and complementarity to the shRNA stem (**F**). Indicated sequence variants are swapped into the equivalent box in (A), which designates the applied tuning strategy for each family of curves. Error bars represent one standard deviation from duplicate transfected wells.

Competing strand tuning strategies enable predictive alteration of the component transfer function

Modifying competing strand base-pairing interactions is anticipated to reflect changes in K_{Comp}, as this parameter represents the thermodynamic partitioning between active and inactive conformations. We developed competing strand tuning strategies to target modifications to three regions within the competing strand domain: the length of the competing strand on the 3' end (Figure 3.3B and C) or the 5' end (Figure 3.3D and E), or the base-pairing complementarity (Figure 3.3F and G). We introduced iterative nucleotide changes under each competing strand tuning strategy and generated component transfer functions as before. Regardless of the selected strategy, each nucleotide change resulted in a shift in the response curve in line with the model prediction for variation in K_{Comp}. The results suggest that decreasing the extent of base-pairing interactions between the competing strand and the shRNA stem decreases the stability of, or bias towards, the inactive conformation (lower K_{Comp}), resulting in lower

basal expression levels and a higher EC_{50}. The trend towards higher EC_{50} is consistent with the order-of-magnitude difference between the apparent K_D of S4t observed in the in-line probing experiment and that reported for the aptamer alone (Figure 3.1B). Thus, sequence modifications to the competing strand that affect the extent of base-pairing solely map to variation of K_{Comp}.

Aptamer tuning strategies enable predictive alteration of the component transfer function

Although ligand binding to the formed aptamer directly relates to aptamer affinity, represented by K_{Apt}, sequence changes in the aptamer domain may affect other parameters. To evaluate how sequence modification of the aptamer domain corresponds to parameter variation, we tested two theophylline aptamer variants (S11 and S12) with dissimilar K_D values (Zimmermann *et al*, 2000) and the mutated aptamer (S10) (Figure 3.4A and B). Mutating the aptamer core (S10) without perturbing shRNA switch secondary structure or sequence length resulted in a shift in EC_{50}, whereas decreasing aptamer affinity by decreasing the aptamer stem length (S11 and S12) resulted in a shift in both EC_{50} and basal expression levels. The shifts in EC_{50} for S11 and S12 matched the relative K_D measured in vitro for the aptamer variants alone (Zimmermann *et al*, 2000), suggesting that modulating aptamer affinity is reflected by variation in K_{Apt}. However, K_{Apt} affects only EC_{50}, suggesting that either K_{Comp} or e varies with aptamer size. As the competing strand sequence is preserved for S1, S10, S11, and S12, we hypothesized that the shift in basal expression levels independent of K_{Apt} (most obvious in comparing the

transfer functions of S1 and S11) is solely attributed to the third tuning parameter e (Figure 3.2C).

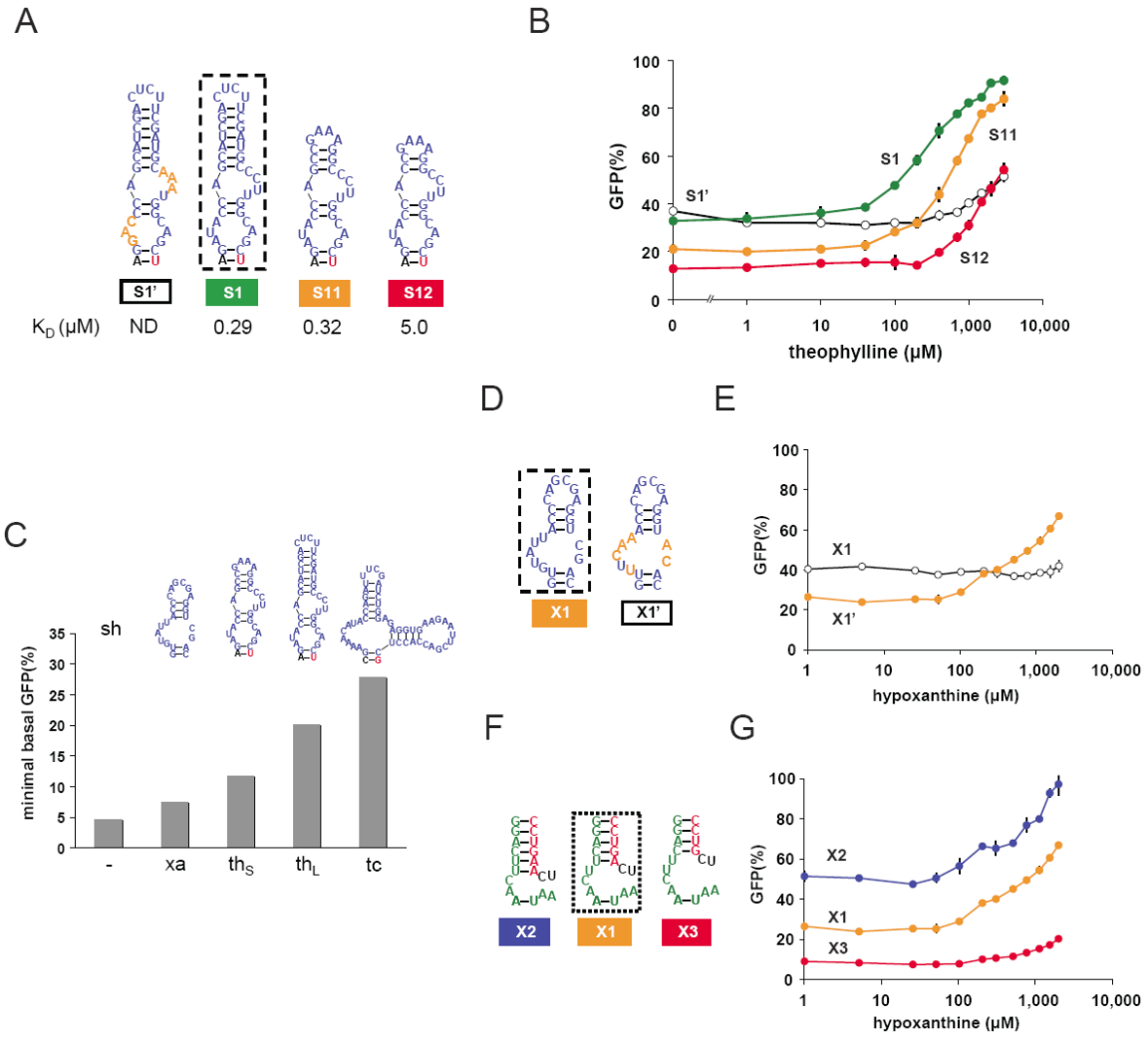


Figure 3.4 Experimental validation of aptamer tuning strategies. (**A**) Theophylline aptamer variants swapped into the equivalent box in Figure 3.3A. Dissociation constants (K_D) as reported previously (Zimmermann et al, 2000) are indicated for each aptamer. (**B**) Tuned theophylline response curves as described in Figure 3.1D for shRNA switches that incorporate aptamers from (A). (**C**) Relationship between aptamer size and the lower limit of basal expression levels estimated from shRNA switches that primarily adopt the active conformation. HEK293T tTA-d2EGFP cells were transfected with shRNA switches containing the following aptamers: none

(— ; sh), xanthine aptamer (xa; X3), smaller theophylline aptamer (th_S; S7, S14, S15), larger theophylline aptamer (th_L; S5, S7, S9, S10), and tetracycline aptamer (tc; T1). Values represent the average of the indicated switches for each aptamer. The original shRNA targeting EGFP (sh) represents the lower theoretical limit in this cellular context. (**D**, **E**) Modular replacement of aptamer imparts new ligand dependence while maintaining switch functionality. Hypoxanthine response curves were generated for shRNA switches incorporating the xanthine aptamer as described in Figure 3.1D, except that cells were grown in the presence of varying concentrations of hypoxanthine. Indicated sequence variants are swapped into the equivalent box in Figure 3.3A. (**F**, **G**) Preservation of competing strand tuning strategies for shRNA switches containing the xanthine aptamer. Variations targeted the length of the 30-end of the competing strand. Error bars represent one standard deviation from duplicate transfected wells.

To evaluate the relationship between aptamer size and the tuning parameter e, we replaced the theophylline aptamer with the smaller xanthine aptamer (Kiga *et al*, 1998) or the larger tetracycline aptamer (Berens *et al*, 2001). As variation of e and K_{Comp} both affect basal expression levels, sole evaluation of e requires estimation of the lower limit of basal expression levels for vanishingly small values of K_{Comp} (Figure 3.2D). To this end, we constructed at least one shRNA switch with each aptamer that strongly prefers the active conformation (low K_{Comp}) and measured GFP basal expression levels of cells transfected with these constructs (Figure 3.4C). Assay results indicated that aptamer size strongly correlated with the lower limit of basal expression levels. The results suggest that the tuning parameter e, which is predicted to have a significant effect on the lower limit of basal expression levels, maps to the size of the aptamer domain. Our observations led to the specification of two aptamer tuning strategies: targeted changes in aptamer

affinity without changing aptamer size alter K_{Apt} and targeted changes to aptamer size to alter the processing efficiency of the switch (e). Taken together, variation of K_{Apt} and e map to the aptamer domain and depend on the nature of the sequence modification.

We examined whether placement of new aptamers into the aptamer domain imparts new ligand dependence while preserving shRNA switch functionality. Previous RNA-based regulatory platforms have demonstrated alteration of ligand dependence by the modular incorporation of new aptamers (Bayer and Smolke, 2005; Win and Smolke, 2007) or minimal mutation of the base aptamer (Thompson *et al*, 2002; Desai and Gallivan, 2004). We evaluated the xanthine aptamer, as it produced low basal expression levels and tightly binds the water-soluble and non-cytotoxic small molecule hypoxanthine. Following construction of shRNA switches that incorporate the xanthine aptamer by direct replacement of the aptamer domain, we generated component transfer functions in HEK293T cells stably expressing EGFP. As observed for S1, intermediate basal expression levels of GFP increased in a dose-dependent manner that was abolished by mutating the aptamer core (Figure 3.4D and E). Furthermore, the competing strand tuning strategies were preserved as evidenced by the effect of changing the competing strand length on the hypoxanthine response curves (Figure 3.4F and G). Contrary to model predictions, mutation of the xanthine aptamer resulted in increased basal expression levels, which may be attributed to base-pairing interactions between the mutated aptamer and the competing strand or shRNA stem sequences, or changes in aptamer folding and stability. However, the shift in basal levels upon mutation of the aptamer sequence is less than that observed for changes in the competing strand, supporting the conclusion that our model serves as a sufficient first approximation. Thus,

our shRNA switch design can accommodate different aptamers to alter the identity of the molecular input that regulates gene expression.

Programming transfer functions by combining competing strand or aptamer tuning strategies

The ligand-regulated behavior of shRNA switches can be programmed through application of the competing strand and aptamer tuning strategies described above. If these programming strategies could be combined, then a collection of shRNA switches could be constructed that display finely tuned transfer functions and respond to a range of molecular inputs. Such capabilities will be integral to the construction of higher order biological networks that display multi-input control over gene expression.

On the basis of the independence of the competing strand tuning strategies, we examined whether the strategies can be combined to fine-tune the component transfer function beyond the capabilities of any single strategy. To generate small deviations in the transfer function of a parent shRNA switch, we added compensatory nucleotide changes under each competing strand tuning strategy in a stepwise manner (Figure 3.5A and B): a point mutation (G68A) within the competing strand to increase complementarity, deletion of two base-pairs to decrease the competing strand length at the 5' end, and a single insertion at the 3' end to increase the competing strand length. Each nucleotide change yielded the expected shift in the transfer function corresponding to the relative stabilization (increased K_{Comp}) or destabilization (decreased K_{Comp}) of the inactive conformation. The final switch, S10, displayed a transfer function slightly shifted from that of the parent switch, S4, demonstrating that nucleotide changes following the

three competing strand tuning strategies can be combined to yield fine-tuning of the component transfer function.

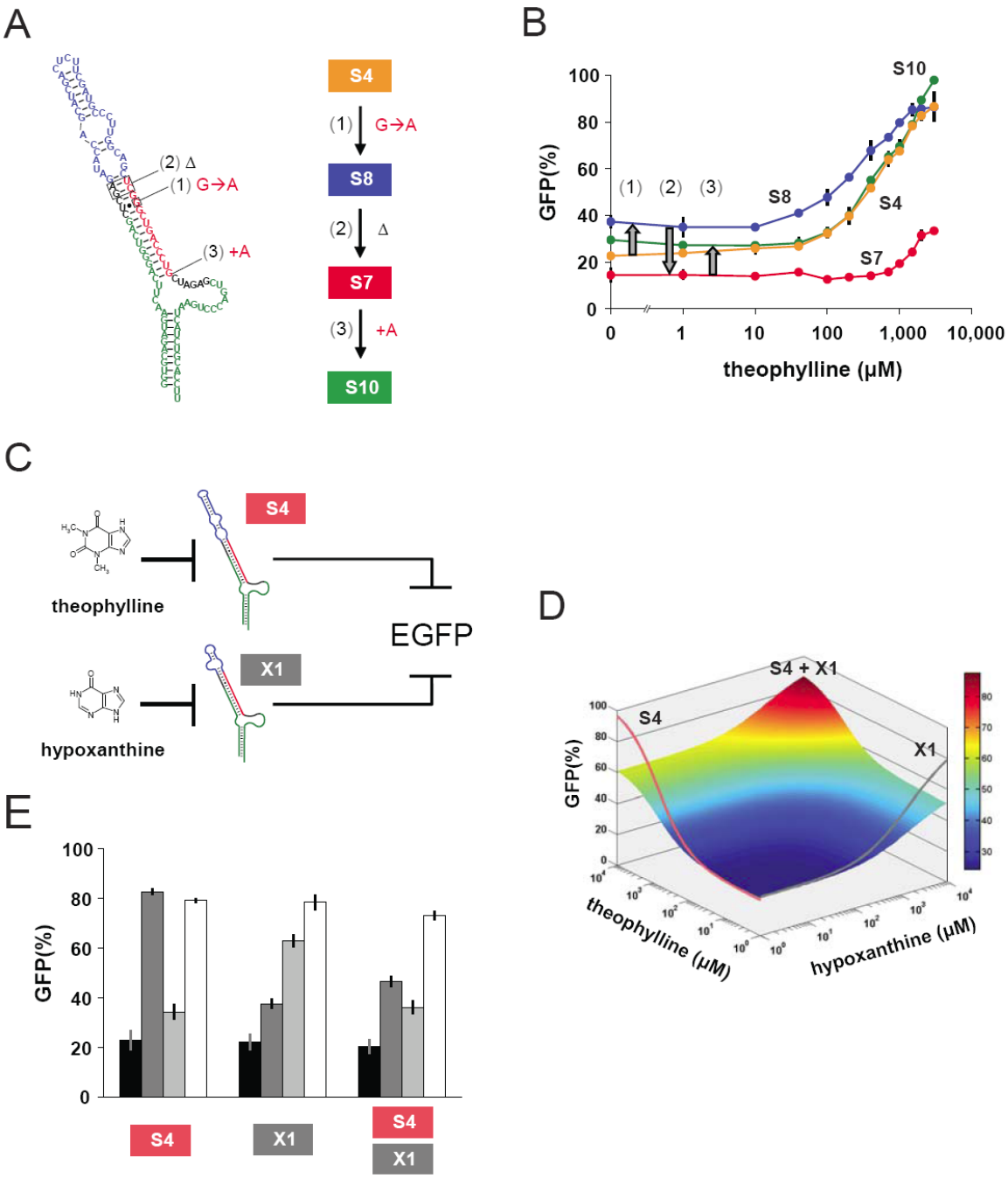


Figure 3.5 Programming transfer functions through combinatorial design strategies. (**A**) Combinatorial tuning strategies enable fine-tuning of the component transfer function. Stepwise nucleotide changes were made to S4, where each change fell under a different competing strand tuning strategy. (**B**) Tuned theophylline response curves as described in Figure 3.1D. Arrows depict the systematic modifications designated in (A). (**C**) Circuit configuration of shRNA switches responsive to theophylline (S4) or hypoxanthine (X1) that both target EGFP. (**D**) Predicted transfer function on the basis of application of the mathematical model to the circuit depicted in (C). Fit curves represent the individual component transfer functions for S4 (—) and X1 (—), respectively. (**E**) Combinatorial implementation of shRNA switches enables construction of networks that process multiple molecular inputs. Results are shown for HEK293T tTA-d2EGFP cells transfected with each shRNA construct (250 ng) or cotransfected with both shRNA constructs (125 ng of each) in the presence of water (■), 3 mM theophylline (■), 2 mM hypoxanthine (■), or both theophylline and hypoxanthine (□). Error bars represent one s.d. from duplicate transfected wells.

In addition to fine-tuning of a single shRNA switch, combining shRNA switches with different ligand dependencies would contribute to the construction of networks that integrate multiple molecular inputs, as suggested from recent work on siRNA-based logic evaluator systems (Rinaudo *et al*, 2007). To evaluate the efficacy of combining shRNA switches, we transfected HEK293T cells stably expressing EGFP with shRNA switches that incorporate the theophylline aptamer (S4) or the xanthine aptamer (X1), where both switches target EGFP and display similar basal expression levels. On the basis of the combined component transfer functions (Materials and methods), we anticipated that the combined regulatory effects of S4 and X1 would require the presence of both hypoxanthine and theophylline to fully inhibit GFP silencing (Figure 3.5C and D).

Relative GFP levels were measured for cells transfected in the presence of theophylline, hypoxanthine, or the combination of the two (Figure 3.5E). GFP levels were high when individual switches were paired with their cognate ligand or both ligands. Some signal cross talk was observed as evidenced by the lower responsiveness of S4 and X1 to their non-cognate ligands, and primarily attributed to low aptamer specificity as observed in previous theophylline aptamer studies (Jenison *et al*, 1994) and not explicitly tested for the xanthine aptamer (Kiga *et al*, 1998). When the switches were cotransfected, high GFP levels coincided only in the presence of both ligands as expected on the basis of the circuit configuration. On the basis of the results, shRNA switches allow the construction of finely tuned genetic networks that can process multiple inputs.

An in silico framework towards component sequence-to-transfer function prediction

The construction of large-scale biological systems will require the simultaneous optimization of the behavior of all system components to yield proper network behavior as suggested for natural (Suel *et al*, 2007) and synthetic (Gardner *et al*, 2000; Yokobayashi *et al*, 2002) systems. Although the transfer functions associated with shRNA switches and other synthetic riboswitches are amenable to physical tuning, a computational framework to effectively navigate through qualitatively functional sequences is necessary for the rapid optimization of switch performance. Folding energetics dictate conformational partitioning and therefore switch performance for a strand displacement mechanism. RNA secondary structure prediction algorithms (Mathews *et al*, 2004) have the potential to perform accurate in silico prediction of in vivo switch performance, although these algorithms have not been sufficiently tested for

in vivo folding dynamics. To investigate the applicability of the secondary structure algorithms to predict in vivo switch behavior, we sought to develop a sequence-to-function relationship for shRNA switches using such algorithms in combination with our model (Figure 3.6A). On the basis of our tuning analysis, we identified K_{Comp} as the sole parameter that reflects partitioning between active and inactive conformations and maps to the competing strand. The free energy difference (ΔG) between conformations is directly related to K_{Comp} such that transfer function prediction is possible by calculating ΔG from sequence information with the aid of structure prediction algorithms, converting this value into K_{Comp}, and inserting K_{Comp} into Equation (3.1) to quantitatively relate ligand concentration and target gene expression levels. A fully determined model requires values for the remaining model parameters; as these parameters are not currently amenable to calculation in silico, experimental estimation can be conducted with a minimal set of experiments on the basis of our model construction (Supplementary text S3.1).

We first determined if ΔG values calculated from the algorithm correlate with the measured basal expression levels for shRNA switches with varying competing strand sequences. The implicit assumption is that competing strand alterations affect only conformational partitioning, which can be calculated with the structure prediction algorithms. We evaluated ΔG (ΔG_{method}) by separating active and inactive conformations on the basis of the minimal free energy (MFE) and the weighted energies from a partition function (PF) calculation (Supplementary text S3.2 and Supplementary Figure S3.5), where both methods are commonly used to evaluate RNA folding in vitro and in vivo.

These methods were employed to calculate ΔG_{method} for shRNA switches S1–10, which differ only in their competing strand sequence.

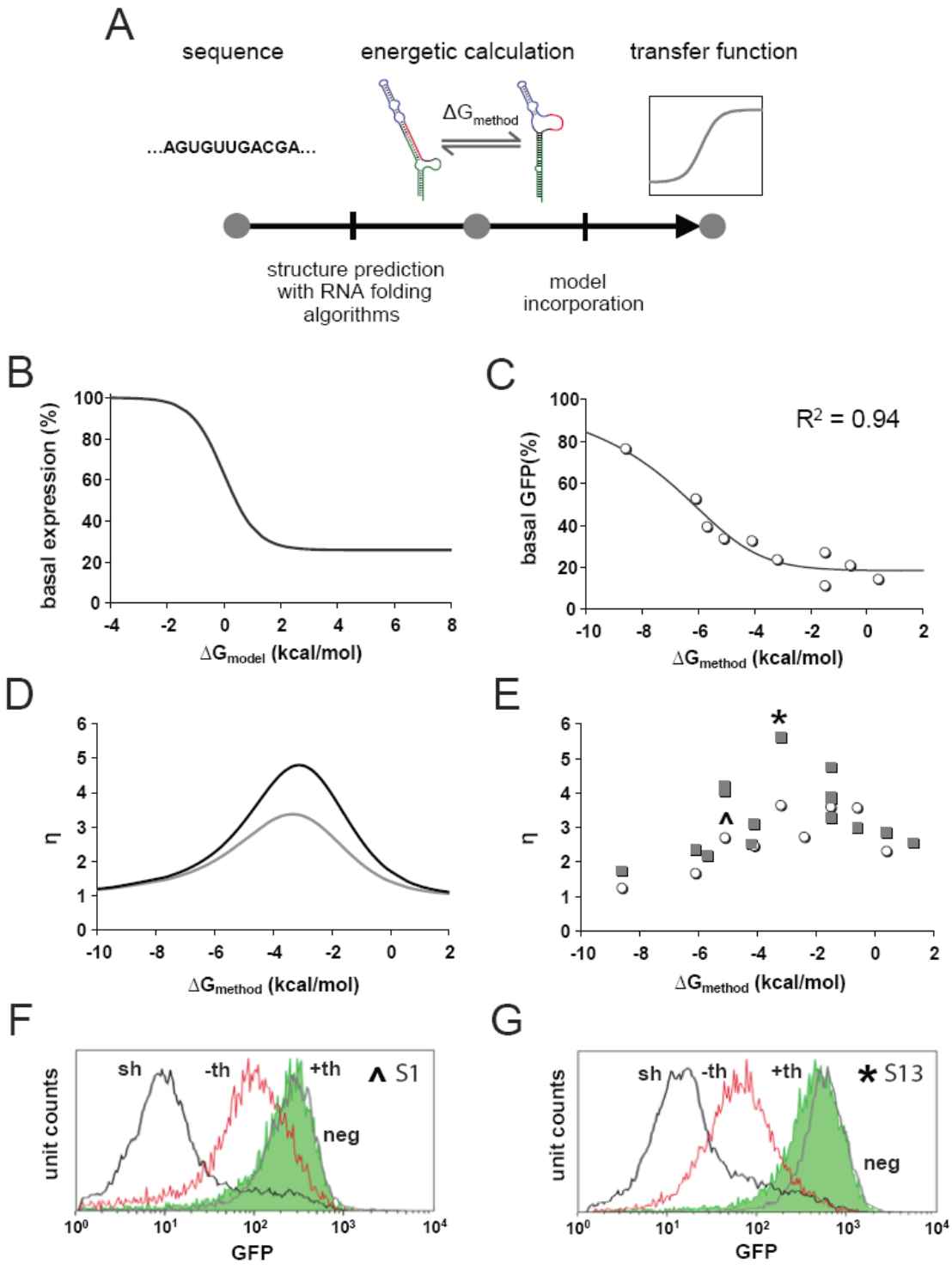


Figure 3.6 Extended model enables sequence-to-transfer function prediction and guides the forward design of optimized shRNA switches. (**A**) General process to convert shRNA switch sequence information into a predicted transfer function. RNA secondary structure algorithms and the method displaying the highest correlation strength ('Stems' method; Supplementary text S3.2 and Supplementary Figure S3.5) were used to calculate the free energy difference between active and inactive conformations (ΔG_{method}). This value is subsequently used to calculate K_{Comp}, which is inserted into the extended model to yield the predicted relationship between ligand concentration and target gene expression levels. (**B**) Predicted relationship between basal expression levels and calculated free energy difference (ΔG_{model}) between active and inactive conformations. (**C**) Sequence–function relationship for shRNA switches under the 'Stems' method. This method links sequence information to basal expression levels with the aid of RNA secondary structure prediction algorithms. ΔG was calculated (ΔG_{method}) according to this method for shRNA switch sequences S1–10 and plotted with the associated measured basal expression levels. The strength of the three-parameter curve fit was evaluated on the basis of the coefficient of determination (R^2). Each data point represents one shRNA switch. (**D**) Extended model predictions for the relationship between ΔG_{method} and dynamic range (η). η is defined as the ratio of GFP (%) at high (3 mM) and low (1 mM) theophylline concentrations. Curves represent shRNA switches containing the smaller theophylline aptamer (━ ; e = 0.94, K_{Apt} = 0.015 μM^{-1}) or the larger theophylline aptamer (— ; e = 0.85, K_{Apt} = 0.016 μM^{-1}), respectively. (**E**) Values of η for shRNA switches containing the larger theophylline aptamer (S1–10; ○) or the smaller theophylline aptamer (S11–25; ■) as a function of ΔG_{method}. Each data point represents one shRNA switch. S13 (the optimized shRNA switch) and S1 (the original shRNA switch) are marked. (**F**, **G**) Flow cytometry data for HEK293T tTA-d2EGFP cells transfected with S1 (F) or S13 (G) in the presence (+th, ━) or absence (-th, ━) of 3 mM theophylline. Histograms are included for the untransfected population of each switch in the absence of theophylline (neg, ━)

or cells transfected with the original shRNA targeting EGFP in the absence of theophylline (sh, —).

To measure the correlation strength between ΔG_{method} and basal expression levels for either method, we performed a least-squares fit using a three parameter equation of the same form as our model with both data sets. Ideally, the fit relationship between ΔG_{method} and measured basal expression levels should align with the same relationship predicted by the model (Figure 3.6B), where ΔG (ΔG_{model}) is related to K_{Comp} according to Equation (3.3). Both MFE and PF calculations failed to provide a significant correlation between ΔG_{method} and basal expression levels (Supplementary Figure S3.6), suggesting that these methods are insufficient for accurate prediction of RNA folding dynamics in vivo.

For all competing strand tuning strategies, increasing the stability of the inactive conformation always resulted in an increase in basal expression (Figure 3.3B–G). The MFE and PF methods did not effectively capture each energetic shift potentially due to the inclusion of binding interactions outside of the major stems. We hypothesized that the interactions outside of the competing strand domain are less prevalent in vivo and are biasing the energetic calculations. To examine this possibility, we devised a third method, the 'Stems' method, that accounts only for the energetics of the major stem in each conformation (Supplementary Figure S3.5). Implementing the 'Stems' method resulted in a strong correlation ($R^2 = 0.94$) between basal expression levels and ΔG_{method} (Figure 3.6C).

Despite the absence of a perfect overlap between the correlation of the 'Stems' method and that predicted by our model (Figure 3.6B and C), the correlation established a

significant empirical link between shRNA switch sequence and behavior in the absence of ligand. This correlation can be assimilated into the model by equating basal expression levels predicted by the fit equation and the model to determine the relationship between ΔG_{method} and K_{Comp} (Supplementary text S3.2). Doing so yields a predictive component transfer function that is now dependent on the calculated value of ΔG_{method}:

$$f_{model} = 1 - e \cdot f_{shRNA}\left[1 + \left[\sqrt[h]{\frac{e \cdot f_{shRNA}}{C_1}\left[C_2 + \exp\left(-\frac{\Delta G_{method}}{k_B N_A T}\right)\right]^{C_3}} - 1\right]\left(1 + K_{Apt} \cdot L\right)\right]^{-h} \qquad (3.2)$$

where $C_{1\text{-}3}$ are empirical constants from the fit correlation. Our extended model provides a general framework for predicting shRNA switch transfer functions from sequence information, where energetic values produced from structure prediction algorithms are inserted into the model for the prediction of switch behavior. Although the extended model currently requires parameter fitting to yield the predicted relationship between ligand concentration and target gene expression levels, the framework establishes a starting point for the development of methods that rely on in silico calculations for transfer function prediction from sequence information.

Model-guided forward design of shRNA switches with optimized transfer functions

To apply the extended model to the forward design of shRNA switches with defined functional properties, we sought to design a theophylline-regulated shRNA switch displaying a maximized dynamic range (η). η is defined as the ratio of GFP levels at high (3 mM) and low (1 mM) theophylline concentrations. We used our extended model to calculate the range of ΔG values where η is maximized. Model predictions suggest that γ is maximized for switches with $\Delta G_{method} \approx$ -3 kcal/mol and that use of the

smaller theophylline aptamer (higher e) yields a higher maximum (Figure 3.6D). To evaluate the predicted landscape, we designed new shRNA switches (S13–25) that include the smaller theophylline aptamer and display ranging ΔG values, generated component transfer functions and calculated η. Plotting ΔG_{method} against the measured value of η for all theophylline-regulated shRNA switches (S1–25; Figure 3.6E) shows a maximum for switches containing the smaller theophylline aptamer that is higher than that for the switches containing the larger aptamer. Furthermore, both maxima existed at $\Delta G_{method} \approx$ -3 kcal/mol as predicted by the extended model supplemented with the empirical parameter values, and the best switch (S13) was approximately equal to the theoretical maximum of η according to model predictions ($\eta_{max,theor} = 5$). Flow cytometry data illustrate the improvement in dynamic range (Supplementary Figure S3.7) for the best shRNA switch (S13; Figure 3.6G) as compared to the original shRNA switch (S1; Figure 3.6F).

To examine the generality of shRNA switch design and functionality, we designed a set of shRNA switches targeting the endogenous La protein. Following selection of an shRNA sequence that yielded moderate knockdown of La as assayed by qRT–PCR, six shRNA switches (L1–6) were constructed with the smaller theophylline aptamer and various competing strand sequences covering a range of ΔG values. Each shRNA switch exhibited variable response to 1.5 mM theophylline that was not observed for the base shRNA (Supplementary Figure S3.8). As observed for the GFP-targeting shRNA switches, use of the ‘Stems’ method provided a suitable correlation between basal levels and ΔG_{method}, whereas the MFE and PF methods did not. Supplying the model with fit values for $C_{1\text{-}3}$ yielded a predicted dynamic range trend that closely

matched the experimental data (Figure 3.7A and B). Interestingly, when the values of f_{shRNA} and e calculated from the La-targeting base shRNA and an shRNA switch preferentially adopting the active conformation (L6) were combined with the remaining parameter values from the GFP experiments, the resulting trend predicted the same maximum dynamic range with a shifted value of ΔG_{method} corresponding to the maximum dynamic range. The results suggest that the shRNA sequence affects calculations with the 'Stems' method such that empirical values are specific to individual sequences and experimental conditions. However, the 'Stems' method produced a strong correlation such that the model may be implemented in future designs by generating a small set of shRNA switches covering ΔG_{method} values of approximately -5 to 0 kcal/mol and measuring basal expression levels. Thus, shRNA switches can be constructed to target different genes, and the model can be used as a tool to guide the forward design of switches displaying optimal behavior.

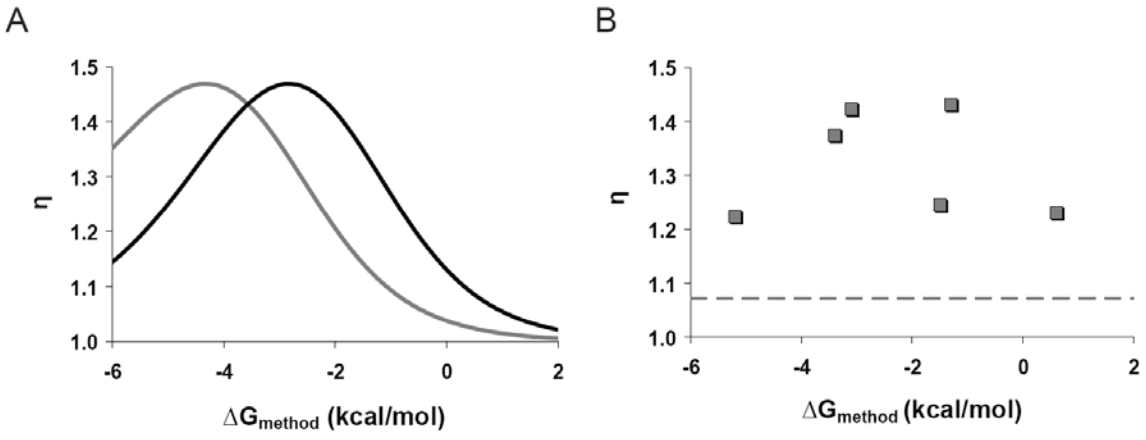


Figure 3.7 Model application guides the forward design of shRNA switches targeting an endogenous gene. (**A**) Extended model predictions for the relationship between ΔG_{method} under the 'Stems' method and dynamic range (η) using empirical parameter values determined from the GFP experiments and experimental parameters determined from the La control experiments (—; f_{shRNA} = 0.6, e = 0.72), or extrapolated empirical parameter values determined from the La switch experiments (**—**) (Supplementary Figure S3.8). (**B**) Relationship between ΔG_{method} and experimental dynamic range (η) for La-targeting shRNA switches L1–6. Plasmids harboring shRNA switches L1–6 displaying a range of ΔG_{method} values were transiently transfected into HEK293T tTA-d2EGFP cells in the presence or absence of 1.5 mM theophylline and La mRNA levels were analyzed by qRT–PCR (Supplementary Figure S3.8). Each data point represents one shRNA switch. The dashed line represents the apparent increase in La mRNA levels upon theophylline addition observed for the original shRNA (shL).

DISCUSSION

A comparison between the framework described here and a recently described ligand-controlled shRNA system (An *et al*, 2006) highlights important design strategies to engineer domain swapping and tuning of the transfer function into synthetic riboswitch systems. In the previous design, ligand control of RNAi was achieved through direct coupling of the theophylline aptamer and an shRNA stem. This design inherently limits aptamer swapping, as the aptamer must perform ligand binding coordinated with modulation of Dicer processing, and prevents tuning of the transfer function, as sequence changes that modulate Dicer processing cannot be implemented without a complete loss of ligand responsiveness. In contrast, we propose here a framework on the basis of the coupling of three distinct domains that carry out separate functions necessary to convert

ligand binding into modulation of RNAi activity. Our system requires that the aptamer performs one function, ligand binding, and the modulation of RNAi processing is performed by a separate domain, the competing strand. The competing strand permits fine-tuning of the transfer function and enables modular coupling of the aptamer and shRNA stem domains, as confirmed by independently replacing each domain and demonstrating preservation of functionality.

We developed a model to enhance our understanding of shRNA switch activity and identified five tuning strategies reflected in three model parameters, K_{Comp}, K_{Apt}, and e, that map specifically to sequence changes in the competing strand or aptamer domains. Our model also established important shRNA switch design guidelines. The first is that basal expression levels are determined by a collection of factors: shRNA potency (f_{shRNA}), shRNA switch processability (e) and prevalence of the active conformation (K_{Comp}). To achieve a desired basal expression level, all factors must be considered in the switch design. Another guideline originates from the observation that larger aptamers coincided with increased basal expression levels, potentially due to sterically hindering processing by the RNAi machinery. The specific contribution of secondary or tertiary structure to the inhibitory effect is unclear, although further understanding of how the RNAi machinery specifically interacts with the shRNA through crystallographic or mutational studies may shed light on this dependence. Our results suggest that shRNA switch sequence length has an upper limit before compromising activity, where future engineering efforts may focus on alleviating or entirely removing this limitation. Furthermore, if achieving low basal expression levels is critical and a set of aptamers against the same ligand are available, use of smaller aptamers may be preferred even at a

cost to aptamer affinity. Such a guideline may even direct library design for the selection of new aptamers by placing an upper limit on the length of the randomized sequence.

We incorporated RNA folding algorithms into our model for in silico prediction of shRNA switch behavior in vivo. The resulting model yielded a framework for the forward design of shRNA switches with specified functional properties. This was achieved by linking RNA secondary structure prediction algorithms, which convert sequence information into energetic values, to our model, which converts energetic values into switch behavior, to provide an empirical sequence–function relationship. The specific method used to calculate the free energy difference (ΔG_{method}) between active and inactive conformations deviated from commonly used methods (MFE and PF calculations) on the basis of observations from the experimental tuning trends. Our alternative method may provide a better correlation with experimental results by focusing the prediction of K_{Comp} to the region of the switch in which the competing strand binding events are occurring, ignoring energetic contributions from other regions of the switch molecule that may not be relevant to the in vivo conformational switching process. Our analysis moves towards direct sequence-to-function relationships and suggests that commonly used methods for predicting RNA structure and behavior should be carefully evaluated when applied to in vivo environments. RNA folding in vivo is a complex process, and algorithms that account for folding kinetics (Danilova *et al*, 2006) and ulterior structural formation (Parisien and Major, 2008), such as pseudoknots or non-canonical base-pairing interactions, may increase the accuracy of the model as well as provide insight into sequences that deviate from model predictions (Figures 3.6E and 7B). For practical application, newer algorithms will need to be more fully developed to

offer the same functionalities as existing algorithms, such as the ability to rapidly scan suboptimal structures, calculate the energetics of multiple RNA strands, and perform a partition function calculation. Although the PF method did not produce a strong correlation using existing algorithms, non-canonical base-pairing interactions may have been an important factor that will be accounted for with newer algorithms.

On the basis of the demonstrated modularity and tunability of our platform, shRNA switches can be implemented towards various applications. The required dynamic regulatory range of a given application will be one of the main considerations in utilizing shRNA switches, as the switches are practically limited to an ~10-fold induction ratio on the basis of the maximum achievable knockdown with an endogenously expressed shRNA. However, many endogenous non-coding RNAs, including miRNAs, exhibit similar restrictions on dynamic regulatory range and have important functions in diverse biological processes, suggesting that this limited dynamic range is not absolutely restrictive to the utility of shRNA switches as dynamic gene regulatory components.

As one potential application, shRNA switches can be applied to disease therapy by sensing intracellular disease markers and inducing apoptosis or cell cycle arrest only in the affected cells as suggested previously (Rinaudo *et al*, 2007). When a context-dependent concentration threshold divides diseased and normal cells, tunability is essential to reduce the likelihood of false positives or negatives. In addition, shRNA switches can be integrated into synthetic genetic circuits to generate advanced control schemes in biological systems. Such systems often exhibit complex dependencies on the dynamics of component interactions, and tuning of component behavior is often necessary to achieve optimal system performance. Through the fine-tuning strategies and

model-guided forward design tools described here, shRNA switches may be used to address challenges faced in biological network design and serve as complex regulatory components in synthetic biology (Endy, 2005; Keasling, 2008; Savage *et al*, 2008).

MATERIALS AND METHODS

Plasmid construction. All shRNAs were cloned into pSilencer 2.1-U6 puro (Ambion). The original shRNA present in pSilencer was used as a scrambled shRNA control. The pSilencer backbone was modified to co-express DsRed-Express in 293T cells by cloning the SV40 origin of replication, CMV IE promoter, and DsRed-Express into the NsiI/MfeI restriction sites. The original XhoI site present in the backbone was removed by XhoI cleavage, extension with the large Klenow fragment (New England Biolabs), and ligation. To clone the shRNA switches, the original shRNA followed by a 6-nucleotide (nt) string of T's was cloned into BamHI/HindIII directly downstream of the U6 promoter. The original shRNA was converted into an shRNA switch by cloning the remaining sequence (Supplementary Table S3.1) into XhoI/XbaI contained within the shRNA loop region. This allowed cloning in parallel of multiple shRNA switches that are comprised of the same shRNA region. All cloning steps involved annealing of 5'-phosphorylated synthetic oligonucleotides (Integrated DNA Technologies) and ligation into the backbone vector. All restriction enzymes and T4 DNA ligase were purchased from NEB. All constructs were sequence-verified (Laragen Inc.), where sequences are provided in Supplementary Table S3.1.

Preparation of RNAs. S4t was transcribed in vitro from an annealed template containing the T7 promoter (5'-TTCTAATACGACTCACTATA**G**GG-3', where **G** is the first transcribed nucleotide) using the Ampliscribe T7 transcription kit (Epicentre) according to the manufacturer's instructions. Following transcription and DNase treatment, unincorporated NTPs were removed using a NucAway clean-up column (Ambion). 5'-phosphates were subsequently removed using Antarctic phosphatase (NEB). Dephosphorylated RNA was then gel-purified on a 6% denaturing polyacrylamide gel and quantified using an ND-1000 spectrophotometer (NanoDrop). RNAs were 5'-radiolabeled using T4 PNK (NEB) and [γ-^{32}P]-ATP, purified using a NucAway clean-up column, and gel-extracted on a 6% denaturing polyacrylamide gel.

In-line probing. In-line probing was conducted as described previously (Soukup and Breaker, 1999). After heating at 70°C for 2 min followed by slow cooling to room temperature, 5'-radiolabeled RNAs (0.2 pmol) were incubated for 40 h at 25°C in varying amounts of theophylline with 50 mM Tris–HCl pH 8.5 and 20 mM $MgCl_2$. Reactions were terminated by adding an equal volume of loading buffer (10 M urea, 1.5 mM EDTA). The alkaline hydrolysis ladder was generated by incubating RNA in 50 mM $NaHCO_3/Na_2CO_3$ pH 9.2 and 1 mM EDTA for 6 min at 95°C. The G-specific cleavage ladder was generated by incubating RNA in 1U RNase T1 (Ambion) with 20 mM sodium citrate pH 5.0, 1 mM EDTA, 7 M urea, and 3 mg yeast RNA for 25 min at 25°C. RNAs were resolved on an 8% denaturing polyacrylamide gel, dried for 90 min at 70°C, then visualized on an FX phosphorimager (Bio-Rad). Band quantification was performed using the Quantity One software package (Bio-Rad). To account for well-loading

variability, quantified band intensities were normalized to an adjacent band of similar intensity showing negligible theophylline dependence.

Cell culture and transfection. All cells were maintained at 37°C in a 5% CO_2-humidified incubator. HEK293T, HEK293, HeLa, and HEK293T tTA-d2EGFP cells were maintained in minimal essential medium alpha media (Invitrogen) supplemented with 10% fetal bovine serum (FBS) (Invitrogen), whereas MDA-MB-231 cells were maintained in RPMI 1640 with glutamine (Invitrogen) supplemented with 10% FBS. Cells were transfected 1 day after seeding using Fugene 6 (Roche) according to the manufacturer's instructions, followed by the immediate addition of ligand. HEK293T tTA-d2EGFP were transfected with shRNA vector (250 ng), whereas cells lacking endogenous GFP were cotransfected with shRNA vector (250 ng) and pcDNA3.1(+) (Invitrogen) harboring the d2EGFP gene (25 ng) (Clontech). One day post-transfection, the media and ligand were replaced. Transfected cells were collected 3 days post-transfection for flow cytometry analysis.

Cell fluorescence analysis. Three days post-transfection, cells were trypsinized and subjected to flow cytometry analysis using the Cell Lab Quanta SC MPL (Beckman Coulter). Cells were first gated twice for (1) viability as assessed by electronic volume versus side scatter and (2) green fluorescence above autofluorescence to remove a non-fluorescent subpopulation. Cells were then gated for either low or high DsRed-Express fluorescence, representing untransfected or transfected cells, respectively. To minimize well-to-well variability, the median green fluorescence value of transfected cells were

divided by that of untransfected cells in the same well and reported as GFP(%). For cells cotransfected with shRNA and GFP plasmids, GFP(%) is the relative GFP levels when normalized to mean red fluorescence followed by normalization to cells transfected with the scrambled shRNA. See Supplementary Figure S3.9 for representative plots and the corresponding gates for transfected and untransfected cells.

Modelling and RNA energetic calculations. Calculation of RNA free energy and partition functions were performed using RNAStructure (Mathews *et al*, 2004). K_{Comp} and the energy difference between inactive and active conformations are related by the following expression:

$$\Delta G_{model} = E\left(\right) - E\left(\right) = -N_A k_B T \cdot \ln\left(K_{Comp}\right) \quad , \tag{3.3}$$

where N_A is Avogadro's number, k_B is the Boltzmann constant, and T is temperature (K). See Supplementary texts S3.1 and S3.2 for a full description of the model derivation, methods for calculating folding energetics, and prediction of the transfer function for a given shRNA switch sequence. Equation fits to measure the correlation strength between ΔG_{method} and basal expression levels were performed by least squares analysis using the following expression that has the same mathematical form as Equation (3.1):

$$f_{fit} = 1 - C_1\left[C_2 + \exp\left(-\frac{\Delta G_{method}}{k_B N_A T} \right) \right]^{-C_3} \quad , \tag{3.4}$$

where $C_{1\text{-}3}$ are the fit constants. Supplementary Table S3.2 contains energetic values calculated under each method along with experimentally determined expression levels.

To model the multi-input system attained by cotransfecting equimolar concentrations of plasmids harboring S4 and X1, the knockdown achieved by the original

shRNA was halved to reflect the 50% decrease in delivered plasmid DNA. The ligand-dependent contributions to decreased expression levels were combined into a single expression to reflect the additive nature of shRNA levels mediating knockdown of the target gene:

$$f = 1 - \frac{f_{shRNA}}{2} \cdot \left[\frac{e_{S4}}{\left[1 + K_{Comp,S4}\left(1 + K_{Apt,S4} \cdot L_{th}\right)\right]^{h_{S4}}} + \frac{e_{X1}}{\left[1 + K_{Comp,X1}\left(1 + K_{Apt,X1} \cdot L_{xan}\right)\right]^{h_{X1}}} \right], \quad (3.5)$$

where the subscripts S4 and X1 designate parameter values generated by a fit to the corresponding individual component transfer functions, and L_{th} and L_{xan} represent the exogenous theophylline or hypoxanthine concentration, respectively.

ACKNOWLEDGEMENTS

We thank G Soukup and D Endy for critical reading and comments on this manuscript, P Aebischer for providing the HEK293T tTA-d2EGFP cells, and S Culler for providing assistance with qRT–PCR. This work was supported by the Caltech Joseph Jacobs Institute for Molecular Engineering for Medicine (grant to CDS), the Defense Advanced Research Projects Agency (grant to CDS), the National Institutes of Health (training grant to TSB; fellowship to KGH), and the Department of Defense (fellowship to CLB).

REFERENCES

Abbas-Terki T, Blanco-Bose W, Deglon N, Pralong W, Aebischer P (2002) Lentiviral-mediated RNA interference. *Hum Gene Ther* **13**:2197–2201.

An CI, Trinh VB, Yokobayashi Y (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer–small molecule interaction. *RNA* **12**: 710–716.

Bartlett DW, Davis ME (2006) Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. *Nucleic Acids Res* **34**: 322–333.

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23**:337–343.

Berens C, Thain A, Schroeder R (2001) A tetracycline-binding RNA aptamer. *Bioorg Med Chem* **9**: 2549–2556.

Buskirk AR, Landrigan A, Liu DR (2004) Engineering a ligand-dependent RNA transcriptional activator. *Chem Biol* **11**: 1157–1163.

Croft LJ, Lercher MJ, Gagen MJ, Mattick JS (2003) Is prokaryotic complexity limited by accelerated growth in regulatory overhead? *Genome Biol* **5**: P2.

Danilova LV, Pervouchine DD, Favorov AV, Mironov AA (2006) RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J Bioinform Comput Biol* **4**: 589–596.

Deans TL, Cantor CR, Collins JJ (2007) A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**: 363–372.

Desai SK, Gallivan JP (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* **126**: 13247–13254.

Dueber JE, Mirsky EA, LimWA (2007) Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol* **25**: 660–662.

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

Endy D (2005) Foundations for engineering biology. *Nature* **438**: 449–453.

Flotte TR (2000) Size does matter: overcoming the adeno-associated virus packaging limit. *Respir Res* **1**: 16–18.

Gopinath SC (2007) Methods developed for SELEX. *Anal Bioanal Chem* **387**: 171–182.

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**: 339–342.

Grate D, Wilson C (2001) Inducible regulation of the S. cerevisiae cell cycle mediated by an RNA aptamer-ligand complex. *Bioorg Med Chem* **9**: 2565–2570.

Grieger JC, Samulski RJ (2005) Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. *J Virol* **79**: 9933–9944.

Griffiths-Jones S (2004) The microRNA registry. *Nucleic Acids Res* **32**: D109–D111.

Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140–D144.

Grundy FJ, Henkin TM (2006) From ribosome to riboswitch: control of gene expression in bacteria by RNA structural rearrangements. *Crit Rev Biochem Mol Biol* **41**: 329–338.

Hall B, Hesselberth JR, Ellington AD (2007) Computational selection of nucleic acid biosensors via a slip structure model. *Biosens Bioelectron* **22**: 1939–1947.

Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci USA* **102**: 3581–3586.

Huang CY, Ferrell Jr JE (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci USA* **93**: 10078–10083.

Hutvagner G, Simard MJ, Mello CC, Zamore PD (2004) Sequence-specific inhibition of small RNA function. *PLoS Biol* **2**: E98.

Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* **24**: 545–554.

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22**: 841–847.

Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-resolution molecular discrimination by RNA. *Science* **263**: 1425–1429.

Keasling JD (2008) Synthetic biology for synthetic chemistry. *ACS Chem Biol* **3**: 64–76.

Kiga D, Futamura Y, Sakamoto K, Yokoyama S (1998) An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. *Nucleic Acids Res* **26**: 1755–1760.

Kim D, Rossi J (2008) RNAi mechanisms and applications. *Biotechniques* **44**: 613–616.

Koch AL (1956) The metabolism of methylpurines by Escherichia coli. Tracer studies. J Biol Chem 219: 181–188 Lee JF, Hesselberth JR, Meyers LA, Ellington AD (2004) Aptamer database. *Nucleic Acids Res* **32**: D95–100.

Lynch SA, Desai SK, Sajja HK, Gallivan JP (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem Biol* **14**: 173–184.

Macrae IJ, Zhou K, Li F, Repic A, Brooks AN, Cande WZ, Adams PD, Doudna JA (2006) Structural basis for double-stranded RNA processing by Dicer. *Science* **311**: 195–198.

Malphettes L, Fussenegger M (2006) Impact of RNA interference on gene networks. *Metab Eng* **8**: 672–683.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci USA* **101**: 7287–7292.

Meister G, Landthaler M, Dorsett Y, Tuschl T (2004) Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. *RNA* **10**: 544–550.

Ogawa A, Maeda M (2008) An artificial aptazyme-based riboswitch and its cascading system in E. coli. Chembiochem 9: 206–209 Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51–55.

Raab RM, Stephanopoulos G (2004) Dynamics of gene silencing by RNA interference. *Biotechnol Bioeng* **88**: 121–132.

Rinaudo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol* **25**: 795–801.

Savage DF, Way J, Silver PA (2008) Defossiling fuel: how synthetic biology can transform biofuel production. *ACS Chem Biol* **3**: 13–16.

Seelig G, Soloveichik D, Zhang DY, Winfree E (2006) Enzyme-free nucleic acid logic circuits. *Science* **314**: 1585–1588.

Shalgi R, Lieber D, Oren M, Pilpel Y (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput Biol* **3**: e131.

Soukup GA, Breaker RR (1999) Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**: 1308–1325.

Stojanovic MN, Stefanovic D (2003) A deoxyribozyme-based molecular automaton. *Nat Biotechnol* **21**: 1069–1074.

Sudarsan N, Barrick JE, Breaker RR (2003) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9**:644–647.

Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* **315**: 1716–1719.

Suess B, Hanson S, Berens C, Fink B, Schroeder R, Hillen W (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res* **31**: 1853–1858.

Suess B, Weigand JE (2008) Engineered riboswitches: overview, problems and trends. *RNA Biol* **5**: 24–29.

Thompson KM, Syrett HA, Knudsen SM, Ellington AD (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol* **2**: 21.

Weigand JE, Suess B (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res* **35**: 4179–4185.

Wieland M, Hartig JS (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl* **47**: 2604–2607.

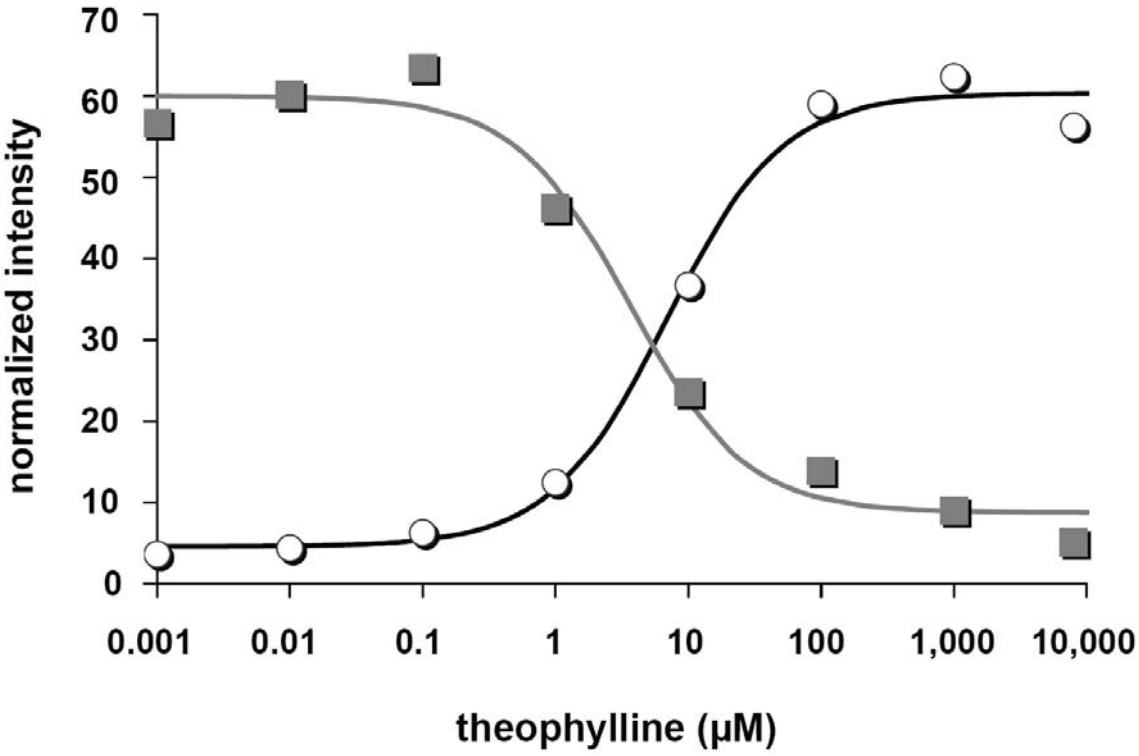
Win MN, Smolke CD (2007) From the cover: a modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci USA* **104**: 14283–14288.

Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci USA* **99**: 16587–16591.

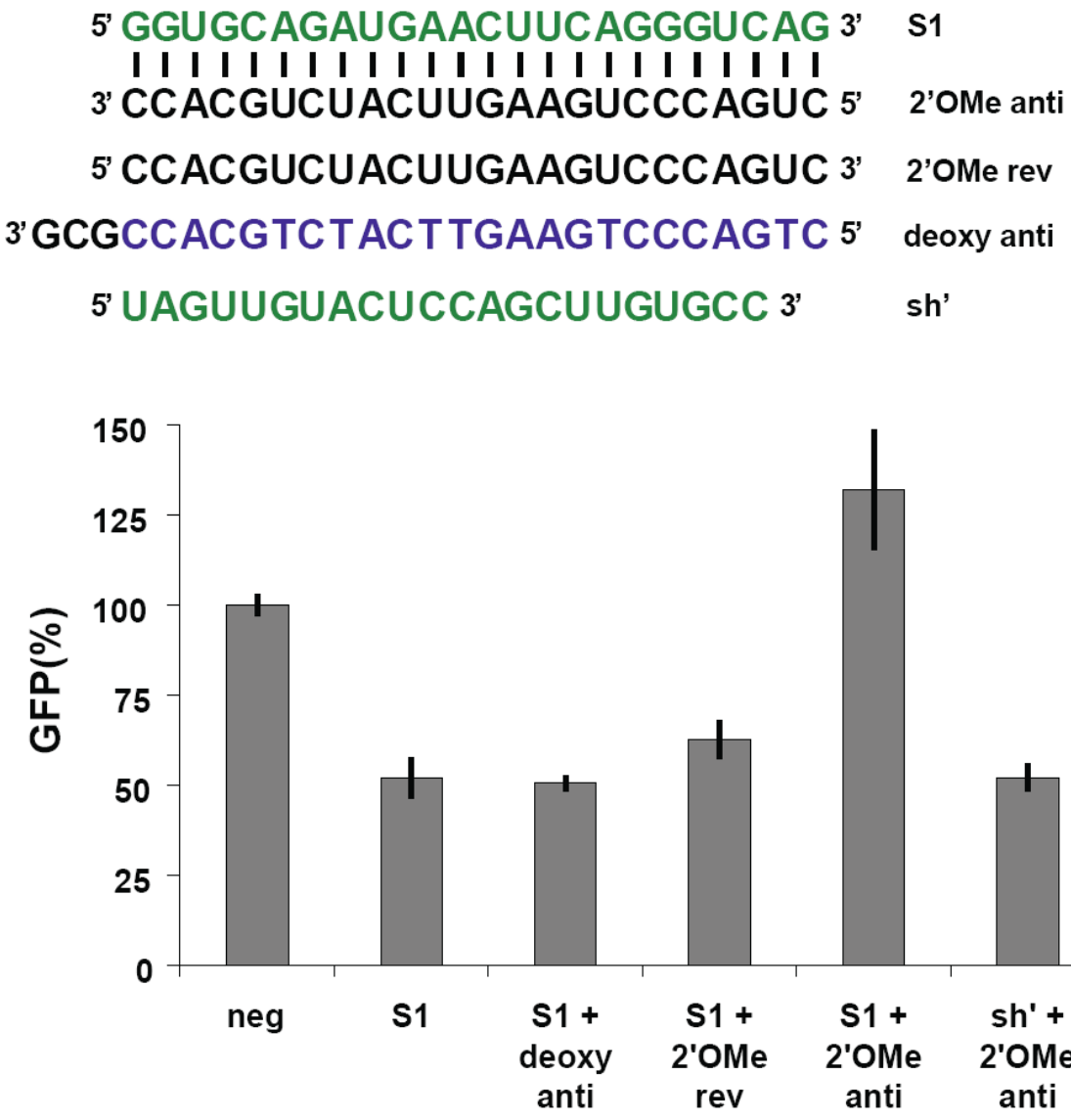
Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* **32**: 4776–4785.

Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6**: 659–667.

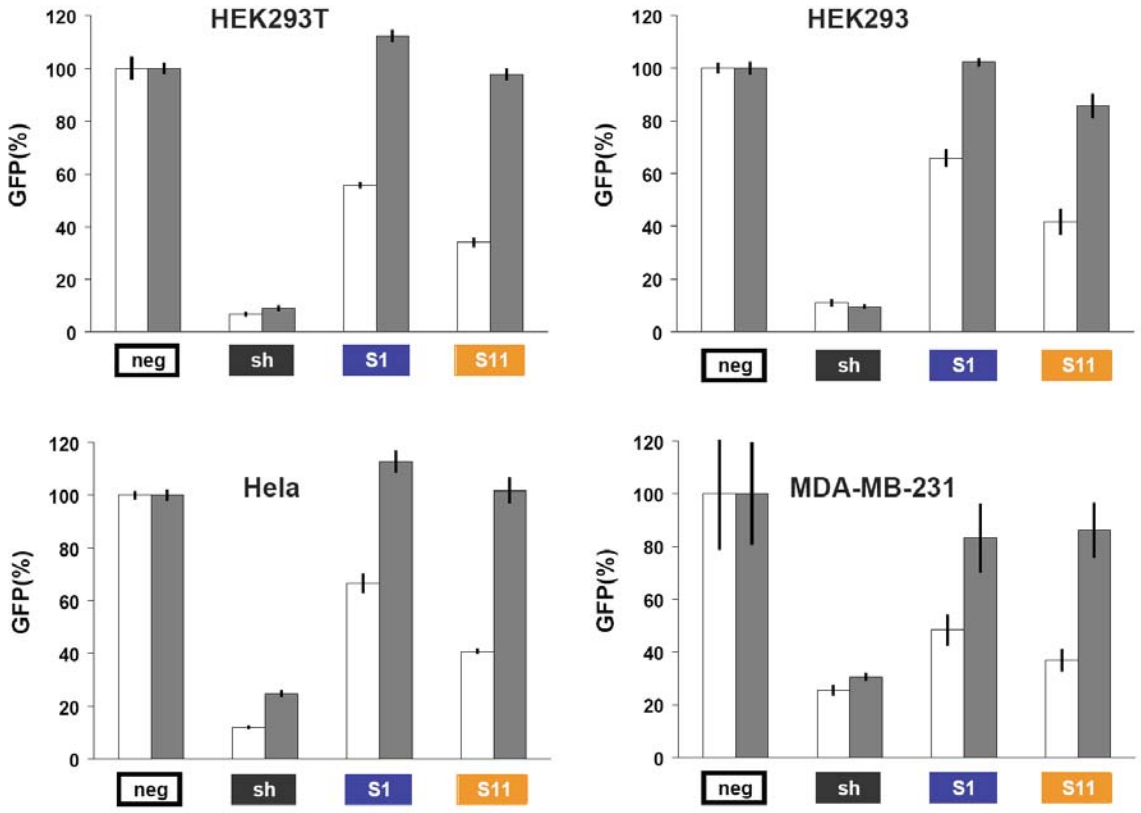
SUPPLEMENTARY INFORMATION



Supplementary Figure S3.1 Theophylline dependence of in-line cleavage at C56 (■) and C74 (○) of S4t. Individual bands from the in-line gel (Figure 3.1B) were quantified and normalized to an adjacent constant region to remove bias from inconsistent well loading. C56 bands were normalized to U47 bands, while C74 bands were normalized to U100-G102 bands. Curves were fit using a standard Michaelis-Menten model, with determined K_D values of 3.6 μM and 6.8 μM for C56 and C74, respectively.

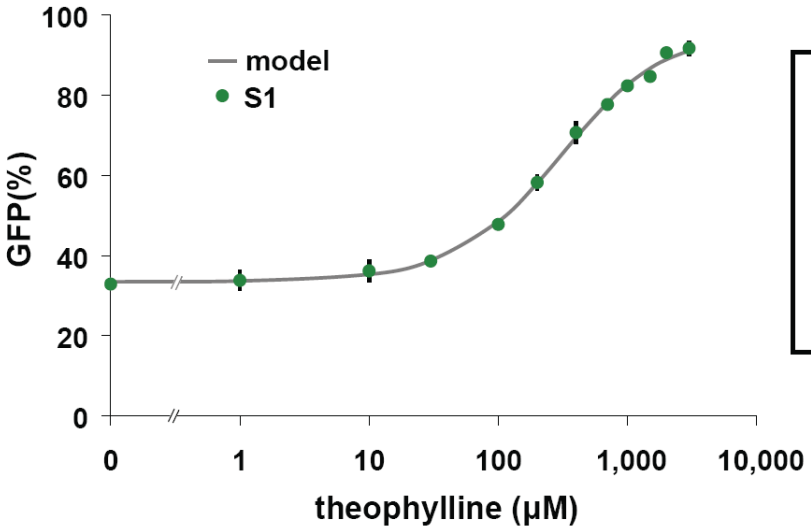


Supplementary Figure S3.2 Antisense inhibition of guide strand activity represses RNAi-mediated silencing of GFP. Cells stably expressing EGFP were cotransfected with the designated oligo and a plasmid that expresses a scrambled shRNA (neg), shRNA switch S1, or an shRNA targeting a different region of the EGFP mRNA (sh'). Mean fluorescence relative to neg are based on flow cytometry measurements of transfected cells. The guide strands are shown in green, 2'-O-Methyl nucleotides in black, and deoxy nucleotides in blue. Error bars represent one standard deviation from duplicate transfected wells.



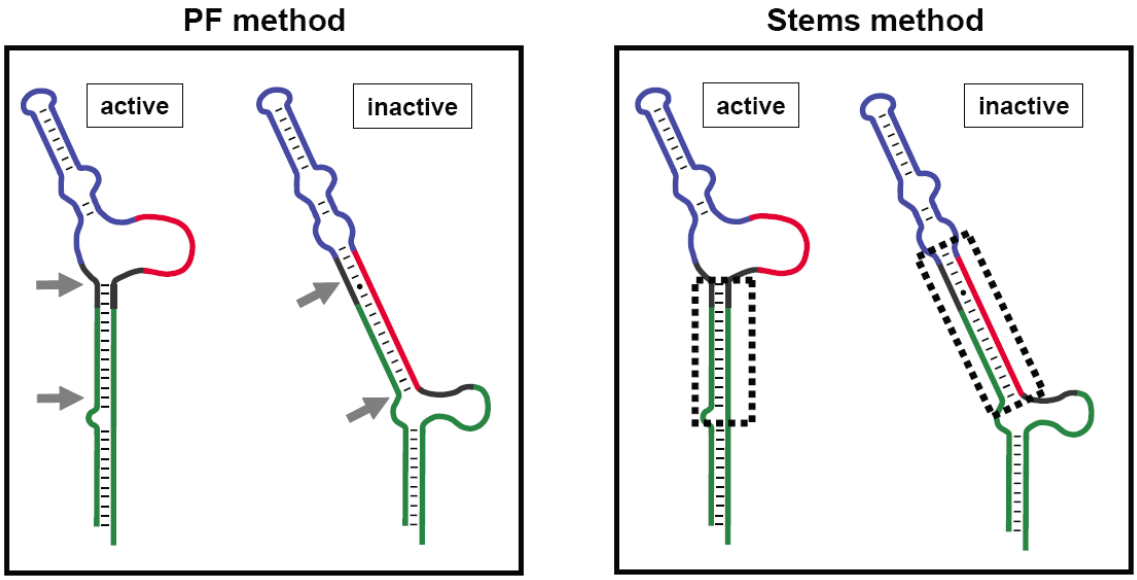
Supplementary Figure S3.3 Functionality of shRNA switches in different cell lines. Cells lacking endogenous EGFP expression were cotransfected with the shRNA construct and an EGFP expression plasmid in the presence (■) or absence (□) of 3 mM theophylline. Mean fluorescence values were normalized to red fluorescence values (DsRed-Express) contributed by the shRNA construct. These values were then normalized to those cells transfected with a scrambled shRNA (neg). Error bars represent one standard deviation from triplicate transfected wells.

$$f(L) = 1 - e \cdot f_{shRNA}\left[1 + K_{Comp}\left(1 + K_{Apt} \cdot L\right)\right]^{-h}$$

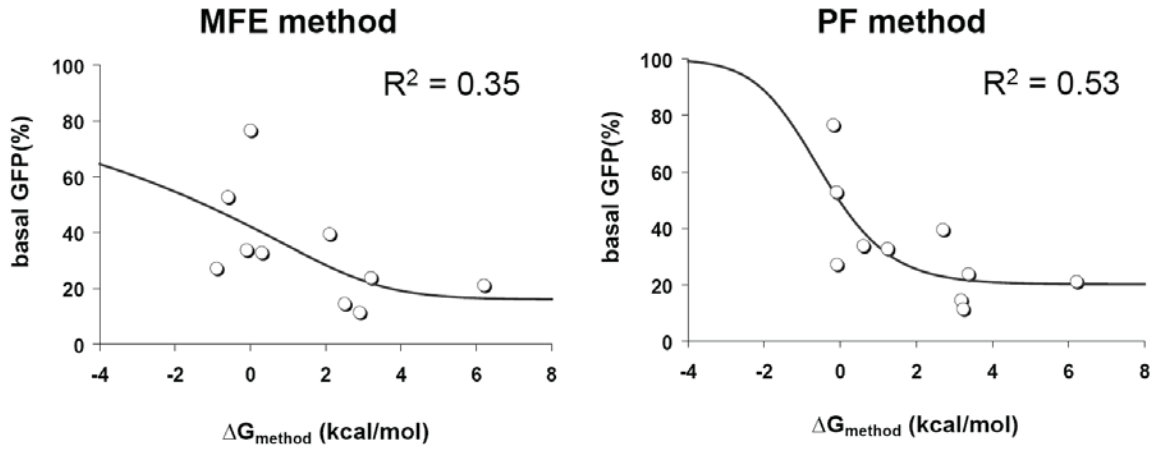


K_{Comp}	0.17
K_{Apt} (μM^{-1})	0.016
e	0.85
f_{shRNA}	0.94
h	1.33

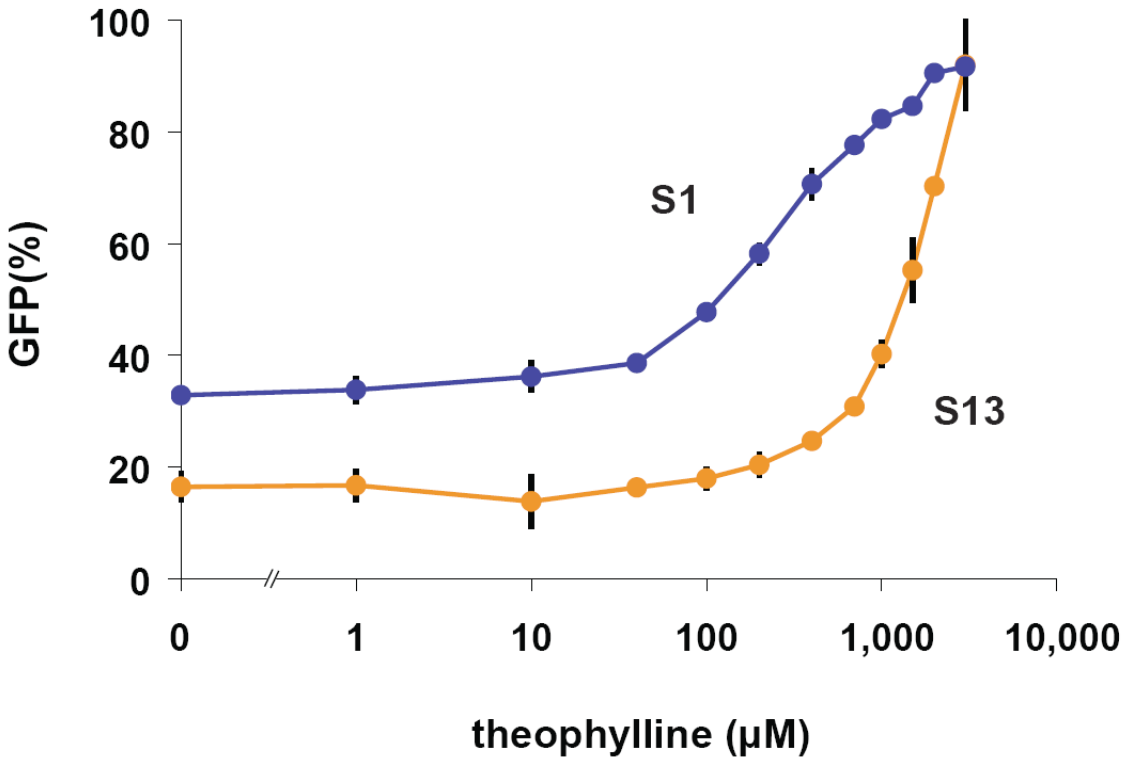
Supplementary Figure S3.4 Derived model equation and model fit of theophylline response curve for S1 data from Figure 3.1D. See Supplementary Text S3.1 for model derivation. The value of f_{shRNA} was set by a separate transfection experiment with the original shRNA targeting EGFP (sh) under the same conditions. The value of e was determined from the average basal expression levels of shRNA switches that highly favor the active conformation (S5, S7, S9, S10). Parameters K_{Comp}, K_{Apt}, and h were produced by a least-squares fit (—) to the S1 data (●). Parameter values are reported to the right of the plot.



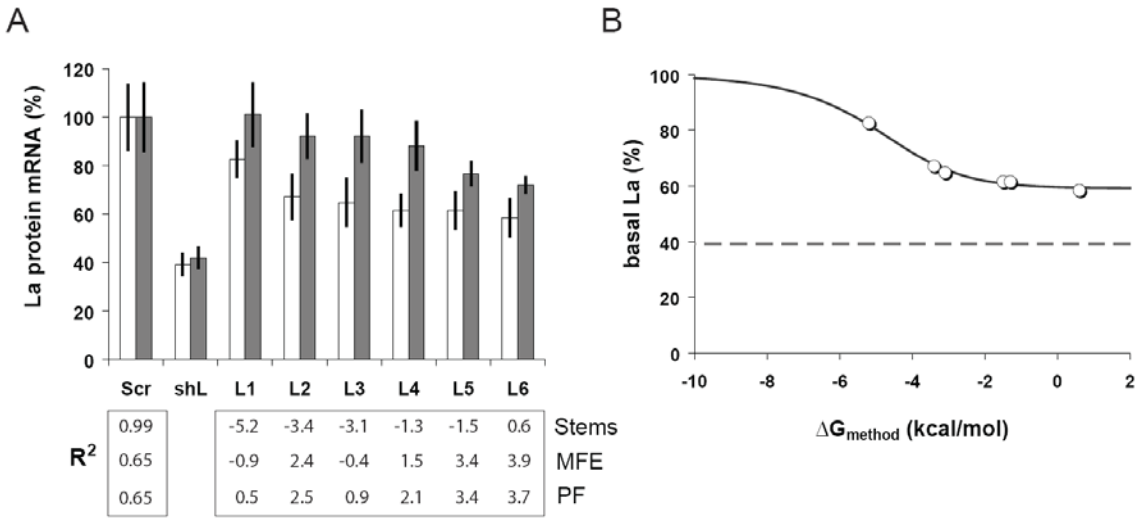
Supplementary Figure S3.5 Depiction of methods to calculate ΔG from shRNA switch sequence information. Base-pairing probabilities of base-pairs designated by arrows were used for the PF method. The boxed sections designate the major stem in the active and inactive conformations used under the Stems method.



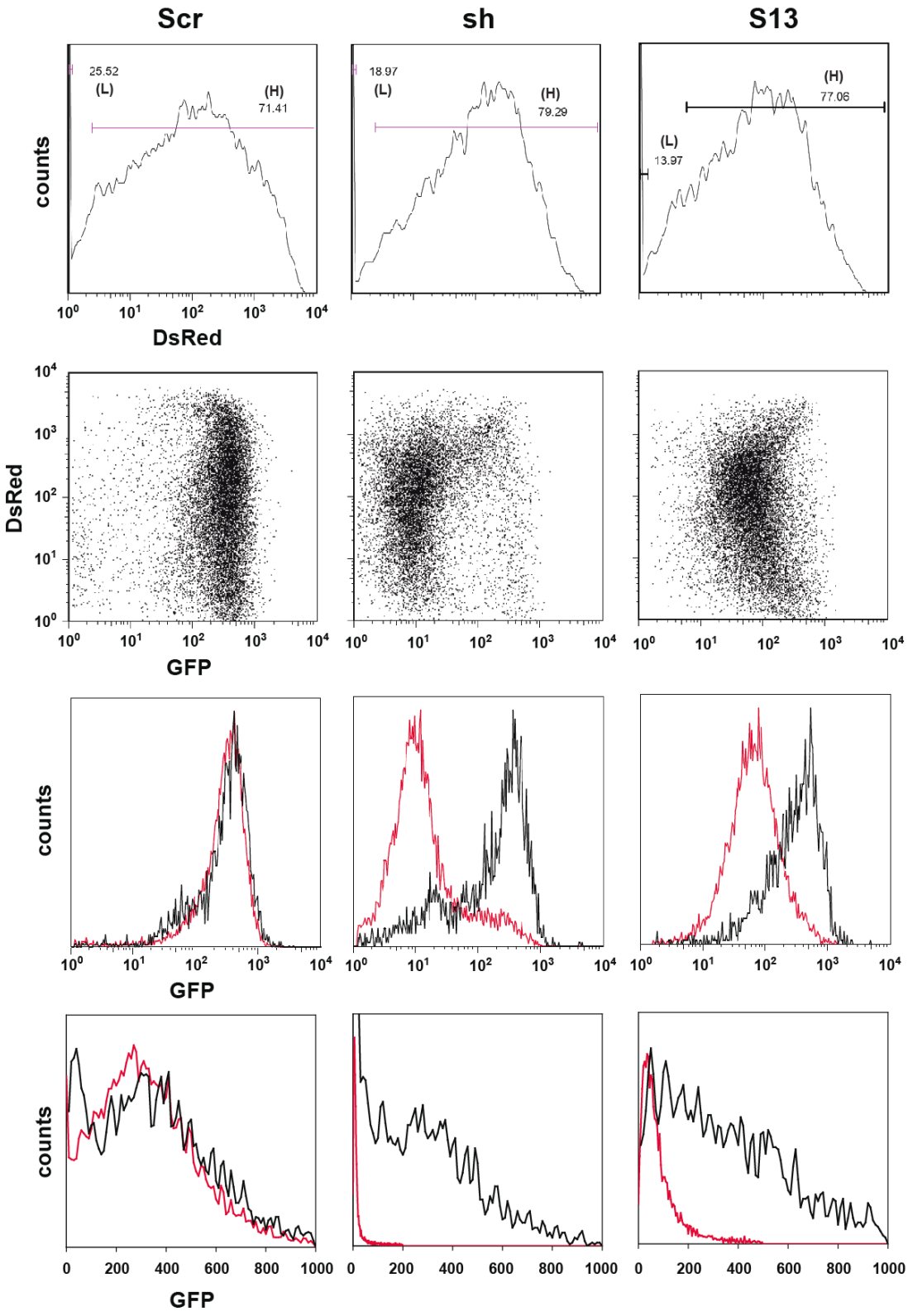
Supplementary Figure S3.6 Alternative methods to relate shRNA switch sequence and in vivo basal expression levels. ΔG was calculated (ΔG_{method}) for shRNA switches S1-10 using RNA secondary structure prediction algorithms. Plots relating ΔG_{method} and measured basal expression level for shRNA switches S1-10, where ΔG was calculated using the MFE method or the PF method. A three-parameter equation with the same mathematical form as the model was fit by least-squares analysis to each data set.



Supplementary Figure S3.7 Theophylline response curve for the initial (S1, ●) and optimized (S13, ●) shRNA switches. Median fluorescence values from flow cytometry analysis were normalized to that of untransfected cells in the same well. Error bars represent one standard deviation from duplicate transfected wells.



Supplementary Figure S3.8 Theophylline-mediated gene regulation of endogenous La protein with shRNA switches. (**A**) qRT-PCR of La protein mRNA from HEK293T tTA-d2EGFP cells transfected in the presence (■) or absence (□) of 1.5 mM theophylline. Calculated free energy differences from the 'Stems,' MFE, and PF methods are displayed below each shRNA switch. Our model predicts decreasing basal levels for increasing bias toward the active conformation (higher values of ΔG_{method}). Coefficient of determination (R^2) for each method is included. Error bars represent the standard deviation for quadruplicate qRT-PCR measurements. (**B**) Curve fit to extrapolate empirical parameters $C_{1\text{-}3}$ using the 'Stems' method using a least-squares fit. Dashed line marks knockdown achieved by base shRNA shL.



$$\mathrm{GFP}(\%) = 100\%\left(\frac{\mathrm{GFP_H}}{\mathrm{GFP_L}}\right)$$

Supplementary Figure S3.9 Representative histograms and dot plots for HEK293T tTA-d2EGFP cells transiently transfected with a plasmid expressing DsRed-Express and either a scrambled shRNA (Scr), the base shRNA targeting EGFP (sh), or shRNA switch S13. Gates in the histogram capture the transfected (H) and untransfected (L) populations in each well, where the untransfected gate was set based on a mock-transfected control that was below the limit of detection (data not shown). The two bottom rows show the resulting GFP histograms on linear or logarithmic axes for the transfected (—) and untransfected (—) gates. The calculation of GFP(%) used in the main text is shown on the right. Median GFP levels of each transfected population were normalized to that of untransfected cells in the same well, thereby reducing well-to-well variability. As compared to the mean, the median of each histogram gave more consistent results for the transient assays.

Supplementary Table S3.1 shRNA and shRNA switch sequences. Color schemes correspond to Figure 1A. Oligos are written from 5' to 3' and reflect the insert sequence cloned into the base plasmid. L1-6 were cloned into shL, while all other switch sequences were cloned into sh.

Name	Aptamer	Sequence	Cloning sites (5'/3')	Database #
neg		GGATCCACTACCGTTGTTATAGGTGTTCAAG AGACACCTATAACAACGGTAGTTTTTTGGAA AAGCTT		pCS626
sh	N/A	GGATCCGGTGCAGATGAACTTCAGGGTCAG CTCGAGTCTAGAGCTGACCCTGAATCATCT GCACCTTTTTTGGAAGCTT	BamHI/ HindIII	pCS741
shL		GGATCCGGCTTCCCAACGATGATGCAACTC CTCGAGTCTAGAGGAGTTGCATCAGTTGGG AAGCCTTTTTTGGAAGCTT		pCS1457
S1	theophylline	CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGGGCTGACCCTGACTAGA	XhoI/ XbaI	pCS630
S1'		CTCGAGGACCCAGCATCGACTCTTCGATGC AAATGGCAGCTCGGGCTGACCCTGACTAG A		pCS847
S2		CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGGGCTGACCCTGAAGCTA GA		pCS633
S3		CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGGGCTGACCCTGAACTAG A		pCS631
S4		CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGGGCTGACCCTGCTAGA		pCS628
S5		CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGGGCTGACCCTCTAGA		pCS632
S6		CTCGACGATACCAGCATCGACTCTTCGATG CCCTTGGCAGCGTCGGGCTGACCCTGCTA GA		pCS848
S7		CTCGATACCAGCATCGACTCTTCGATGCCC TTGGCAGCGAGCTGACCCTGCTAGA		pCS807
S8		CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGAGCTGACCCTGCTAGA		pCS629
S9		CTCGAGATACCAGCATCGACTCTTCGATGC CCTTGGCAGCTCGAGCTGATCCTGCTAGA		pCS1005
S10		CTCGATACCAGCATCGACTCTTCGATGCCC TTGGCAGCGAGCTGACCCTGACTAGA		pCS808
S11		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGACCCTGACTAGA		pCS634
S12		CTCGAGATACCACCGAAAGGCCTTGGCAGC TCGGGCTGACCCTGACTAGA		pCS635
S13		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGACCCTGCTAGA		pCS911

Supplementary Table S3.1 cont'd.

Name	Aptamer	Sequence	Cloning sites (5'/3')	Database #
S14	theophylline	CTCGATACCAGCCGAAAGGCCCTTGGCAGC GAGCTGACCCTGCTAGA	XhoI/ XbaI	pCS908
S15		CTCGATACCAGCCGAAAGGCCCTTGGCAGC GGGCTGACCCTGCTAGA		pCS909
S16		CTCGATACCAGCCGAAAGGCCCTTGGCAGC GAGCTGACCCTGACTAGA		pCS910
S17		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGAGCTGACCCTGCTAGA		pCS941
S18		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGAGCTGACCCTACTAGA		pCS942
S19		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGACCCTGAACTAGA		pCS1001
S20		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGACCCTGAAGCTAGA		pCS1002
S21		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGACCCTGGCTAGA		pCS1003
S22		CTCGATACCAGCCGAAAGGCCCTTGGCAGC GAGCTGACCCTGAACTAGA		pCS1004
S23		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGAGCTGATCCTGCTAGA		pCS1061
S24		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGATCCTGCTAGA		pCS1062
S25		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGGGCTGATCCTGACTAGA		pCS1063
X1	xanthine/guanine	CTCGAGTGTATTACCCAGCGAGGTCGACTC GAGCTGACCCTGACTAGA	XhoI/ XbaI	pCS870
X1'		CTCGAGTTTCAAACCCAGCGAGGTACACTC GAGCTGACCCTGACTAGA		pCS913
X2		CTCGAGTGTATTACCCAGCGAGGTCGACTC GAGCTGACCCTGAACTAGA		pCS972
X3		CTCGAGTGTATTACCCAGCGAGGTCGACTC GAGCTGACCCTGCTAGA		pCS869
T1	tetracycline	CTCGAAAACATACCAGAGAAATCTGGAGAG GTGAAGAATACGACCACCTCGAGCTGACCC TGCTAGA	XhoI/ XbaI	pCS895

Supplementary Table S3.1 cont'd.

Name	Aptamer	Sequence	Cloning sites (5'/3')	Database #
L1	theophylline	CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGAGGAGTTGCATCCTAGA	XhoI/ XbaI	pCS1458
L2		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGAGGAGTTGCATTCTAGA		pCS1459
L3		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTTGAGGAGTTGCATCCTAGA		pCS1460
L4		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTTGAGGAGTTGCATACTAGA		pCS1462
L5		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTCGAGGAGTTGCACTAGA		pCS1463
L6		CTCGAGATACCAGCCGAAAGGCCCTTGGCA GCTTGAGGAGTTGCACTAGA		pCS1464

Supplementary Table S3.2 Calculated free energies and corresponding expression levels of theophylline-regulated shRNA switches. η: ratio of relative GFP levels at exogenous theophylline concentrations of 3 mM and 1 μM for each shRNA switch.

Name	Aptamer	Basal expression levels (%)	η	ΔG (method)		
				MFE	PF	Stems
neg	N/A	90.7	1.07		N/A	
sh		4.6	1.19		N/A	
S1	theophylline	33.8	2.71	-0.1	0.6	-5.1
S1'		37.0	1.39	-0.3	0.8	-5.1
S2		76.7	1.25	0.0	-0.2	-8.6
S3		52.8	1.68	-0.6	-0.1	-6.1
S4		23.8	3.65	3.2	3.4	-3.2
S5		21.1	3.58	6.2	6.2	-0.6
S6		42.5	2.20	2.1	2.7	-5.7
S7		14.5	2.32	2.5	3.2	0.4
S8		37.3	2.47	0.3	1.2	-4.1
S9		12.1	4.75	2.9	3.2	-1.5
S10		29.4	3.60	-0.9	-0.1	-1.5
S11		20.0	4.20	0.6	0.9	-5.1
S12		13.4	4.06	-0.1	0.8	-5.1
S13		16.7	5.61	4.0	3.8	-3.2
S14		14.3	2.86	2.6	3.3	0.4
S15		11.2	2.56	9.9	9.1	1.3

Supplementary Table S3.2 cont'd.

Name	Aptamer	Basal expression levels (%)	η	ΔG (method)		
				MFE	PF	Stems
S16	theophylline	21.7	3.85	0.1	0.1	-1.5
S17		17.5	3.12	2.5	2.0	-4.1
S18		19.5	3.89	5.7	4.5	-1.5
S19		38.3	2.38	-0.6	0.0	-6.1
S20		49.8	1.75	0.0	-0.2	-8.6
S21		10.8	2.52	4.6	4.6	-4.2
S22		35.0	2.73	-0.9	-0.9	-2.4
S23		16.6	3.30	5.1	4.4	-1.5
S24		10.8	3.00	6.6	6.1	-0.6
S25		11.8	2.63	3.2	3.3	-2.5

Supplementary Text S3.1

DERIVATION OF MATHEMATICAL MODEL

We initially developed a mathematical model to examine the mechanism through which shRNA switches mediate ligand control of RNA interference (RNAi). Instead of drawing from existing models (Bartlett and Davis, 2006; Malphettes and Fussenegger, 2006; Raab and Stephanopoulos, 2004) that take into account the mechanistic steps and kinetics of RNAi that are well characterized, we chose to derive a simplified model that captures the steady-state behavior of shRNA switches and the fundamental mechanism that provides for ligand regulation of gene expression. The goal was to develop a model that predicts the relative steady-state expression levels of the target gene (f; output) as a function of exogenous ligand concentration (L; input) and can be easily adapted to predict shRNA switch activity in different cellular environments.

To accomplish this we first began with the proposed mechanism for shRNA switch functionality (Figure 3.1A). This mechanism asserts that a single shRNA switch can adopt two conformations due to distinct base-pairing interactions. The active conformation (left) is processed by the RNAi machinery to an siRNA that initiates RNAi-mediated silencing of target transcripts. Processing includes nuclear export by Exportin-5 (Yi *et al*, 2003) and cleavage by the RNase III-like enzyme Dicer (Ketting *et al*, 2001). Conversely, the inactive conformation (middle) is not processed by the RNAi machinery. Ligand binding to the formed aptamer domain in the inactive conformation stabilizes this

conformation (right), thereby reducing overall processing of the shRNA switch to an siRNA.

Model derivation began by assuming that the three conformations (active, inactive, and inactive bound to ligand) are at thermodynamic equilibrium as determined by K_{Comp} and K_{Apt}. K_{Comp} is the equilibrium partitioning constant between active and inactive conformations, while K_{Apt} is the association constant for binding between ligand and the inactive conformation. When normalized to the total shRNA switch concentration, the fraction of shRNA switches in the active conformation is

$$\left[\text{  }\right] = \frac{1}{1 + K_{comp}\left(1 + K_{apt} \cdot L\right)} \quad . \tag{3.6}$$

The next step was correlating the fraction of shRNA switches in the active conformation to relative expression levels of the target gene. Previous models have highlighted the importance of absolute expression levels of the RNAi substrate, target gene transcripts, and the RNA-induced silencing complex (RISC), as well as the rate of cell division (Bartlett and Davis, 2006). Recent work has elaborated on the mechanism of RNAi, including the emerging role of Dicer binding partners TRBP and PACT (Gregory *et al*, 2005; Kok *et al*, 2007; Lee *et al*, 2006), association of RISC and Dicer (Gregory *et al*, 2005), shuttling of the cleaved siRNA from Dicer to RISC (Gregory *et al*, 2005), cleavage and release of the passenger strand (Matranga *et al*, 2005; Rand *et al*, 2005), target site availability for efficient degradation of the target transcript (Westerhout and Berkhout, 2007), and the potential for saturation of Exportin-5 (Grimm *et al*, 2006; Yi *et al*, 2005). Rather than offer a descriptive model of RNAi that incorporates all of these

mechanisms that are still under investigation, we chose an empirical route that requires minimal experimental data.

Excluding nuclear export by Exportin-5, the mechanistic steps described above apply to the linear cascade downstream of and including Dicer recognition and processing. Incorporation of three parameters, f_{shRNA}, e, and h, can account for the dynamics of these steps. f_{shRNA} is the relative knockdown achieved by the original shRNA – an RNA molecule comprised of a loop region and the shRNA stem sequence, e is the efficiency of shRNA switch processing by the RNAi machinery, and h is the hill coefficient that accounts for the nonlinearity between the concentration of Dicer-cleaved siRNAs and relative expression levels of the target gene. To capture the correlation between the prevalence of the active conformation and target gene expression levels, we used the following relationship:

$$f = 1 - e \cdot f_{shRNA} \left[\text{  } \right]^{h} . \tag{3.7}$$

Introducing equation (3.6) into equation (3.7) yields the final form of the model:

$$f = 1 - e \cdot f_{shRNA} \left[1 + K_{Comp} \left(1 + K_{Apt} \cdot L \right) \right]^{-h} , \tag{3.8}$$

where the relative expression levels of the target gene (f) are a function of exogenous ligand concentration (L).

The power of our model lies in the ability to calculate realistic parameter values from a minimal set of experiments: f_{shRNA} can be found in one experiment by measuring the relative knockdown of the target gene induced by an shRNA that contributes the shRNA stem, e can be calculated from basal expression levels from a few shRNA

switches that strongly prefer the active conformation, and h can be calculated by generating a ligand response curve with one shRNA switch – as long as administration of the highest ligand concentration results in negligible knockdown of the target gene. The remaining model parameters, K_{Comp} and K_{Apt}, can be found from the same response curve used to calculate e, since varying K_{Apt} only changes the EC_{50} while varying K_{Comp} changes both EC_{50} and basal expression levels. A summary of the model parameters and how values are experimentally obtained are included in Supplementary Table S3.3 below.

Supplementary Table S3.3 Description of model parameters.

parameter	initial determination	description
K_{Comp}	fit to data	Equilibrium partitioning constant between conformations equal to [inactive]/[active] (-)
K_{Apt}	fit to data	Association constant between ligand and formed aptamer (1/μM)
e	extrapolated from data	RNAi processing efficiency (-)
f_{shRNA}	from shRNA data	Relative knockdown by original shRNA (-)
h	fit to data	Hill coefficient (-)

To investigate the validity of the model, we experimentally determined model parameter values as described above: f_{shRNA} was equated to the knockdown achieved with the original shRNA targeting EGFP (sh); e was calculated from the average basal expression levels produced by shRNA switches S5, S7, S9, and S10; and K_{Comp}, K_{Apt}, and h were determined by a model fit of the theophylline response curve for S1. The resulting parameter values are shown in Supplementary Figure S3.4. The fit curve aligns with the

response curve for S1, and the fit parameter values are realistic as described below for K_{Apt} and K_{Comp}. The EC_{50} is related to K_{Apt} and K_{Comp} according to the following:

$$EC_{50} = \left(1 + K_{Comp}^{-1}\right)K_{Apt}^{-1} \quad , \tag{3.9}$$

From the in-line assay results, the ratio of the apparent K_D of S4t (5 μM) to the K_D of the aptamer alone (0.29 μM (Zimmermann *et al*, 2000)) was ~ 17. Solving for K_{Comp} in equation (4) yields a value of 0.06. While this is below the fit value from the S1 data of 0.17, S1 has one less base pair than S4 contributed by the competing strand. Thus, the value from S4t is anticipated to be closer to 0.17 if the extra base pair is included. The fit value for K_{Apt} (0.016 μM^{-1}) from the S1 data was lower than that for the aptamer alone (3.4 μM), which can be attributed to a theophylline concentration drop across the cellular membrane as observed in *E. coli* (Koch, 1956) and *S. cerevisiae* (J Liang, J Michener, C Smolke, unpublished data, 2008). Hence the model faithfully follows the underlying mechanism of ligand regulation of gene expression mediated by shRNA switches and can capture *in vivo* behavior by utilizing a minimal set of experiments.

REFERENCES

Bartlett DW, Davis ME (2006) Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. *Nucleic Acids Res* **34:** 322-333.

Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R (2005) Human RISC couples microRNA biogenesis and posttranscriptional gene silencing. *Cell* **123:** 631-640.

Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F, Kay MA (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441:** 537-541.

Ketting RF, Fischer SE, Bernstein E, Sijen T, Hannon GJ, Plasterk RH (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes Dev* **15:** 2654-2659.

Koch AL (1956) The metabolism of methylpurines by Escherichia coli. I. Tracer studies. *J Biol Chem* **219:** 181-188.

Kok KH, Ng MH, Ching YP, Jin DY (2007) Human TRBP and PACT directly interact with each other and associate with dicer to facilitate the production of small interfering RNA. *J Biol Chem* **282:** 17649-17657.

Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN (2006) The role of PACT in the RNA silencing pathway. *EMBO J* **25:** 522-532.

Malphettes L, Fussenegger M (2006) Impact of RNA interference on gene networks. *Metab Eng* **8:** 672-683.

Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD (2005) Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. *Cell* **123:** 607-620.

Raab RM, Stephanopoulos G (2004) Dynamics of gene silencing by RNA interference. *Biotechnol Bioeng* **88:** 121-132.

Rand TA, Petersen S, Du F, Wang X (2005) Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. *Cell* **123:** 621-629.

Westerhout EM, Berkhout B (2007) A systematic analysis of the effect of target RNA structure on RNA interference. *Nucleic Acids Res* **35:** 4322-4330.

Yi R, Doehle BP, Qin Y, Macara IG, Cullen BR (2005) Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *RNA* **11:** 220-226.

Yi R, Qin Y, Macara IG, Cullen BR (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17:** 3011-3016.

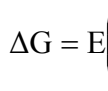
Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6:** 659-667.

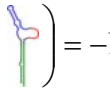
Supplementary Text S3.2

FREE ENERGY CALCULATIONS AND MODEL EXTENSION

The model derived in Supplementary Text S3.1 identified different tuning trends that were observed in our experimental analysis, although this form of the model only predicts qualitative shifts in the transfer function based upon nucleotide changes to a parent shRNA switch. We sought to augment the model with predictive capabilities for the forward design of shRNA switch sequences that yield desired transfer functions. We initially focused on K_{Comp}, the partitioning coefficient between active and inactive conformations, since it solely captured the effect of multiple changes to the competing strand and has a thermodynamic basis. Under basic thermodynamic assumptions, K_{Comp} is related to the free energy difference (ΔG) between the active and inactive conformations according to

$$\Delta G = E\left(\right) - E\left(\right) = -N_A k_B T \cdot \ln\left(K_{Comp}\right), \qquad (3.10)$$





where N_A is Avogadro's number, k_B is the Boltzmann constant, and T is temperature (K). If ΔG can be calculated for a given shRNA switch sequence, then the corresponding value of K_{Comp} can be calculated. When paired with the other experimentally-determined parameter values (Supplementary Text S3.1), this value of K_{Comp} can then be used in the model to predict the transfer function relating ligand concentration and relative gene expression levels. The initial challenge is calculating an experimentally valid ΔG from a given shRNA switch sequence.

Free energy calculation

To calculate ΔG, we employed the RNA secondary structure prediction program RNAStructure 4.5 (Mathews *et al*, 2004) to output structural and energetic information for a given sequence. The program's dynamic folding algorithm utilizes empirical energy values measured *in vitro* (Mathews *et al*, 2004) to predict RNA conformations and their relative free energy. Since application of the program to *in vivo* folding has rarely been addressed (Mathews *et al*, 2004), we first asked if ΔG values calculated from the program (ΔG_{method}) correlated with measured basal expression levels for each shRNA switch. Two commonly used methods were initially employed to calculate ΔG_{method} for S1-10 (switches with the same aptamer domain and shRNA stem): minimal free energy of the active and inactive conformation (MFE method) and partition function calculation to find the relative probability of either general conformation (PF method). ΔG_{method} values were then plotted with the associated basal expression levels measured *in vivo* (Supplementary Table S3.2) and compared to the expected trend from the model (ΔG_{model}; Figure 3.6B). A three-parameter equation with the same mathematical form as the model was then fit to each data set using a least-squares analysis to evaluate the correlation strength, since a strong correlation is necessary for accurate prediction of the transfer function. The mathematical form used to fit the data was

$$f_{fit} = 1 - C_1\left[C_2 + \exp\left(-\frac{\Delta G_{method}}{k_B N_A T} \right) \right]^{-C_3}, \qquad (3.11)$$

where C_1 , C_2 , and C_3 are fit constants and f_{fit} is the basal expression of the target gene for the fit curve.

MFE method

The minimal free energy conformation – the most stable conformation – has been considered to be representative of the actual tertiary structure, and the free energy of this conformation is often considered to represent overall energetics of the RNA sequence. Under the MFE method, the free energy is recorded for the most stable active and inactive conformation. The difference in these free energy values is then reported as ΔG_{method}. The resulting plot (Supplementary Figure S3.6) shows no significant correlation and an associated weak fit ($R^2 = 0.35$), suggesting that this method is insufficient for predicting transfer functions.

PF method

Calculation of the partition function is a more advanced and considered to be a more accurate method for the approximation of RNA energetics. All possible secondary structure conformations and their energies are calculated in order to identify the most prevalent conformation, which often deviates from the minimal free energy conformation. Under the PF method, the program outputs the probability of a given base-pair based on the partition function calculation. To convert these probabilities into a value of ΔG, we first found the smaller value of the base-pair probabilities near the top and bottom of the upper shRNA stem (starting at the stem bulge) in the active conformation and the stem formed by the competing strand and the shRNA stem in the inactive conformation (Supplementary Figure S3.5). Base-pairs were chosen such that the same nucleotide in the shRNA stem was part of the selected base-pair in both conformations. This ensures that a base-pair probability only applies to one of the two conformations. In

other words the sum of the base-pair probabilities that include the same nucleotide for both conformations should always be less than one. Ideally, the sum should equal one, where all calculated sums for S1-10 were between 85% and 99% (data not shown). The value of ΔG_{method} can be calculated from the base-pair probabilities according to the following:

$$\Delta G_{method} = -k_B N_A T \cdot \ln\left(\frac{P_I}{P_A}\right) \quad , \tag{3.12}$$

where P_A and P_I are the base-pair probabilities representing the active or inactive conformations, respectively. ΔG_{method} values were calculated using the PF method and plotted in the same way as above (Supplementary Figure S3.6). The PF method provided a better fit ($R^2 = 0.53$) when compared to the MFE method that qualitatively matched the model trend, although the fit is not suitable for predictive purposes.

Stems method

While increasing the extent of base-pairing between the competing strand and shRNA stem always resulted in an increase in basal expression levels (Figure 3.3B-G), the MFE and PF calculations output predicted an increase or decrease in free energy changes based on binding interactions outside of the major stems. We attributed the inaccuracy of the MFE and PF methods to the equal weight placed on these binding interactions. To remove these contributions to the energetic calculation, we devised a third method we term the ‘Stems’ method. This method only accounts for the energetic contributions from the major stems in the active and inactive conformations. The major stem for the active conformation spans from the shRNA stem bulge to the top of the

shRNA stem, while the major stem for the inactive conformation includes base-pairs formed between the shRNA stem and the competing strand (Supplementary Figure S3.5). The lower portion of the shRNA stem is ignored since it is present in both conformations. As before, we calculated ΔG_{method} for S1-10 and plotted these values against the basal expression levels. The resulting plot (Figure 3.6C) shows a strong correlation ($R^2 = 0.94$), a significant improvement over the other methods.

It is surprising yet insightful that the most accurate method only accounts for energetic contributions from regions that interact with the competing strand, which is precisely and solely where K_{Comp} maps. An inequality does exist between the fit curve from the 'Stems' method and model predictions in terms of the abscissa values and curve slope, which suggests that sequences outside of the major stem contribute to folding energetics *in vivo* in a way that is improperly treated by the MFE or PF method.

Model extension

Based on the strong correlation between ΔG_{method} calculated from shRNA switch sequence and *in vivo* basal expression levels, the fit curve from the 'Stems' method can be incorporated into our model for the forward design of shRNA switches. This is accomplished by converting the value of ΔG_{method} calculated from the 'Stems' method into K_{Comp} that can be used in the model to predict the transfer function. To perform this conversion, f from the model equation and f_{fit} from the curve fit are set equal to each other. For successful conversion, the dynamic range (the range of f) of the model and fit curves must match exactly. This can be done by ensuring that

$$1 - e \cdot f_{shRNA} = f_{fit}\left(\Delta G_{method} \to \infty\right) , \tag{3.13}$$

where e and f_{shRNA} are model parameters. Once set equal to each other, K_{Comp} can be found in terms of ΔG_{method}:

$$K_{Comp} = \sqrt[h]{\frac{e \cdot f_{shRNA}}{C_1}\left[C_2 + \exp\left(-\frac{\Delta G_{method}}{k_B N_A T}\right)\right]^{C_3}} - 1 \quad . \tag{3.14}$$

Replacing K_{Comp} in the model with equation (5) yields the extended model:

$$f_{model} = 1 - e \cdot f_{shRNA}\left[1 + \left[\sqrt[h]{\frac{e \cdot f_{shRNA}}{C_1}\left[C_2 + \exp\left(-\frac{\Delta G_{method}}{k_B N_A T}\right)\right]^{C_3}} - 1\right]\left(1 + K_{Apt} \cdot L\right)\right]^{-h} \quad . \tag{3.15}$$

Following experimental determination of the remaining model parameter values (Supplementary Text S3.1), this equation can be used to predict relative expression levels of the target gene (f_{model}) as a function of ligand concentration (L) by calculating ΔG_{method} under the 'Stems' method using RNAStructure.

Since the obtained fit parameter values are specific to shRNA switches S1-10, there is a question as to how parameter values and model accuracy will change for a new aptamer, target sequence, or cellular context. To address the generality of the model, we constructed six shRNA switches (L1-6) targeting the endogenous La protein that covered a range of values for ΔG_{method}. Cells were transiently transfected with plasmids harboring each shRNA in the presence or absence of theophylline and assayed for La levels by qRT-PCR (Figure 3.7, Supplementary Figure S3.8). Parameter values $C_{1\text{-}3}$ were again extrapolated through a least-squares fit of the basal expression data for L1-6. Similar to the GFP-targeting shRNA switches, a strong correlation was determined using the 'Stems' method, which was not observed with the MFE and PF methods, that yielded a relationship between dynamic range and ΔG_{method} closely matching experimental results.

However, the predicted relationship between ΔG_{method} and dynamic range was slightly shifted as compared to that generated from the fit parameter values for S1-10 supplemented with the shRNA potency (f_{shRNA}) and aptamer inhibitory effect (e) determined for the La-targeting shRNA switches. The parameters f_{shRNA} and e were calculated using basal levels from the base shRNA (shL) and an shRNA switch (L6) strongly biased toward the active conformation, respectively. The results suggest that the 'Stems' method is suitable for the prediction of shRNA switch behavior *in vivo* when supplemented with empirical parameter values specific to each shRNA stem sequence. As demonstrated here the parameter values can be determined from the basal expression levels of only a few switches. As our understanding of dynamic RNA behavior *in vivo* progresses, future modeling efforts may provide more accurate methods that move toward *de novo* sequence-function prediction.

REFERENCES

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101:** 7287-7292.

Supplementary Methods

Antisense inhibition of RNAi. Oligonucleotides were chemically synthesized (Integrated DNA Technologies) with 3' amino linkers (L). The following sequences were used, where 2'-O-methyl nucleotides are underlined and all other nucleotides are 2'-deoxy: 2'OMe anti, 5'-CUGACCCUGAAGUUCAUCUGCACCL-3'; 2'OMe rev, 5'-CCACGUCUACUUGAAGUCCCAGUCL-3'; deoxy anti, 5'-CTGACCCTGAAGTTCA TCTGCACCGCGL-3'. Oligonucleotides were cotransfected with the designated shRNA plasmid (500 ng) into HEK293T tTA-d2EGFP cells seeded in a 12-well plate using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions with a final oligo concentration of 50 nM. The media was replaced one day post-transfection and the cells were trypsinized and assayed by flow cytometry three days post-transfection. Only transfected cells were included in the analysis based upon high DsRed-Express expression levels. A separate shRNA (sh') that targets the EGFP mRNA in a different location was included in the analysis.

qRT-PCR. The following oligos were used for qRT-PCR against La protein (Acc # X13697) and the loading control GAPDH (Acc # NM_002046): La_fwd, 5'-GGTTGAACCGTCTAACAACAG-3'; La_rev, 5'-ATGTCATCAAGAGTTGCATCAG-3'; GAPDH_fwd, 5'-GAAGGTGAAGGTCGGAGTC-3'; GAPDH_rev, 5'-GAAGATG GTGATGGGATTTC-3'. HEK293T tTA-d2EGFP cells were transfected in a 12-well plate in the presence or absence of 1.5 mM theophylline with plasmids harboring the hygromycin B resistance gene and an shRNA targeting La protein mRNA (Acc #

X13697). shRNA sequences are contained in Supplementary Table S3.1. One day post-transfection, cells were subcultured and seeded into a 6-well plate. One day later, the media was replaced and supplemented with 300 μg/ml puromycin. Four days after adding hygromycin, dead cells were removed with a 1X PBS wash and total RNA was extracted using the RNeasy Protect Mini kit (Qiagen) according to the manufacturer's instructions and DNase I-treated for 20 minutes at 37°C. Following purification using a NucAway column (Ambion), total RNA (up to 5 μg) was reverse-transcribed using Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions using the gene-specific reverse primers for La protein and GAPDH followed by the recommended incubation with RNase H. qRT-PCR was conducted with this cDNA on the iCycler iQ system (BioRAD) according to the manufacturer's instructions. Samples were prepared in quadruplicate using the iQ SYBR green supermix and data were analyzed using the iCycler iQ software.

Chapter 4

Design of Ligand-Responsive MicroRNAs

This chapter was adapted with permission from Beisel CL, Culler SJ, Hoff KG, and Smolke CD. Design of ligand-responsive microRNAs based on structural insights into Drosha processing. The manuscript was in submission at the time of the thesis defense.

ABSTRACT

MicroRNAs (miRNAs) are endogenous small RNAs that mediate gene silencing in higher eukaryotes and play critical roles in regulating diverse cellular processes. Here, we describe a strategy for engineering ligand-responsive miRNAs in mammalian cells based on structural requirements for efficient Drosha processing – namely, the single-stranded nature of the miRNA basal segments. We utilized a unique property of aptamer binding, termed adaptive recognition, to introduce ligand control over structure within the basal segments, rendering Drosha processing and subsequent gene silencing sensitive to ligand concentrations. The generality of this strategy as a means of introducing ligand control over gene silencing was demonstrated for three different aptamer-small molecule ligand pairs. Furthermore, we explored the versatility of ligand-responsive miRNAs through the engineering of miRNA clusters, cis-acting miRNAs, and self-targeting miRNAs that combine cis and trans regulation. This versatility is an important property for tailoring genetic regulatory systems to applications in biotechnology and medicine that require finely tuned, combinatorial control of transgene or endogenous gene expression. Finally, our synthetic ligand-responsive miRNAs suggests an as yet undiscovered regulatory mechanism for small molecule control of natural miRNAs.

INTRODUCTION

miRNAs comprise a conserved class of small noncoding RNAs that direct targeted gene silencing through the RNA interference (RNAi) pathway in humans and other eukaryotes (Bartel, 2004). Most miRNAs are encoded within long transcripts transcribed from Pol II promoters (Cai *et al*, 2004; Lee *et al*, 2002). The primary (pri-) miRNA is initially processed to an ~65 nucleotide (nt) precursor (pre-)miRNA by the Microprocessor (Gregory *et al*, 2004; Han *et al*, 2004; Lee *et al*, 2003) composed of the cleaving enzyme Drosha and the RNA binding protein DGCR8. Following pri-miRNA cleavage and export from the nucleus, the pre-miRNA is processed by Dicer to a 22-25 nt miRNA duplex. One of the duplex strands termed the mature miRNA is incorporated into the RNA-induced silencing complex (RISC), which subsequently cleaves or translationally represses the target transcript depending on the degree of complementarity between the guide sequence of the miRNA and the target. miRNA-mediated gene regulation has been implicated in diverse biological processes ranging from development to angiogenesis and may be involved in the regulation of a majority of the human genome (Friedman *et al*, 2009).

Engineered genetic systems that display ligand control of gene silencing mediated by miRNAs will provide a powerful and versatile means to control transgene and endogenous gene expression. In addition, the application of such control systems to natural miRNAs will allow coordinated regulation of potentially hundreds of genes and diverse cellular functions (Baek *et al*, 2008; Selbach *et al*, 2008). Recently, researchers have designed synthetic RNA-based regulatory systems that integrate sensing and gene-regulatory functions, where the former are encoded in RNA aptamer sequences that

recognize small molecule ligands (Suess and Weigand, 2008). Such integrated ligand-responsive RNA-based control systems offer several advantages over more traditional protein-based regulatory systems in avoiding potential immunogenicity of heterologous protein components and providing a more tunable control system. In addition, as aptamers can be selected against a wide range of biomolecules (Ellington and Szostak, 1990; Osborne and Ellington, 1997; Tuerk and Gold, 1990), such integrated RNA systems provide platforms for gene expression control in response to potentially any molecular input.

Recently, integrated RNA-based control systems that mediate gene silencing through the RNAi pathway in response to small molecule ligands were built from gene regulatory functions encoded by intermediate substrates in the processing pathway, small hairpin RNAs (shRNAs) (An *et al*, 2006; Beisel *et al*, 2008). The ligand-responsive shRNA system designs linked small molecule RNA aptamers to shRNA elements to modulate the extent of Dicer processing and subsequent gene silencing through ligand binding events based on known structural requirements for efficient Dicer processing (An *et al*, 2006; Beisel *et al*, 2008). However, sequence restrictions and in vivo toxicity of shRNAs (Boudreau *et al*, 2009; Grimm *et al*, 2006; McBride *et al*, 2008) establish significant hurdles toward broader implementation of these control systems. Recent research indicates that such practical application issues may be avoided through the implementation of integrated ligand-responsive miRNAs (Bauer *et al*, 2009; Boudreau *et al*, 2009; McBride *et al*, 2008).

We have designed ligand-responsive miRNAs and demonstrated their implementation in regulatory circuits that modulate and tune the resulting regulatory

response. Our design is based on elucidated structural requirements for miRNA processing, specifically that the bulge size in the miRNA basal segments dictates the extent of Drosha processing and in vivo silencing. By integrating an aptamer into the miRNA basal segments, we utilized aptamer-ligand binding interactions to increase the local structure in that region, such that Drosha processing and subsequent gene silencing were inhibited with increasing ligand concentration. The sequence flexibility of the basal segments allows for the introduction of different aptamer sequences in this region, resulting in a modular design framework that allows modification of the detected ligand or target gene without compromising regulatory activity. We further engineered circuits comprising clusters of ligand-responsive miRNAs, where the use of cis- or trans-acting miRNAs at variable copy number provided tunable control of gene expression. We also engineered a self-targeting ligand-responsive miRNA circuit based on combined cis and trans regulation that resulted in more stringent regulatory properties, providing a novel system for transgenic control. Our integrated ligand-responsive miRNA framework offers considerable versatility for genetic control and will be broadly useful across applications in biotechnology and medicine.

RESULTS

Extent of structure in the basal segments dictates Drosha processing and gene silencing

miRNAs can be partitioned into four domains (Figure 4.1A) that exhibit unique structural requirements for efficient Drosha processing and RISC activation (Boudreau *et al*, 2008; Han *et al*, 2006; Zeng and Cullen, 2003, 2004, 2005; Zeng *et al*, 2005). For

example, in vitro studies have shown that mutating the sequence of the basal segments to form extensive base pairing interactions abolishes Drosha processing (Han *et al*, 2006; Zeng and Cullen, 2005). These results indicate that efficient Drosha processing requires that the basal segments be single-stranded or unstructured.

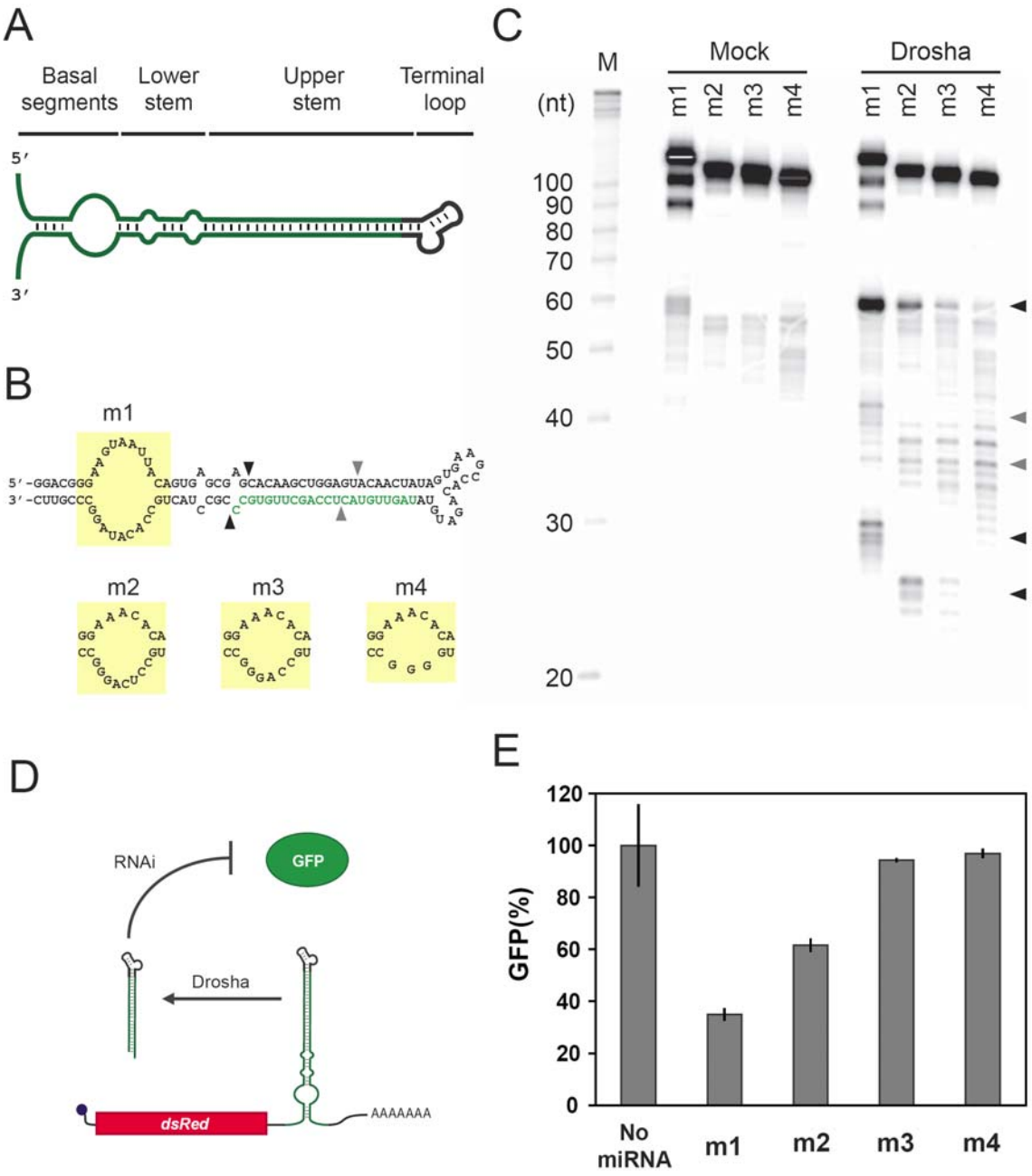


Figure 4.1 Extent of structure in the basal segments dictates miRNA processing and target gene silencing. (**A**) General domains of a pri-miRNA. The upper stem encodes the mature miRNA sequence. The lower stem is ~11 base pairs (bp) in length and designates the Drosha cleavage site. The basal segments comprise the flanking sequences outside of the lower stem. (**B**) Sequence and secondary structures of minimal pri-miRNAs used to examine the effects of bulge size in the basal segments on Drosha processing in vitro. The minimal pri-miRNAs contain a mature miRNA sequence that targets GFP (green text) and differ only in their basal segments. The bulge sequences for each pri-miRNA are indicated in yellow boxes. Black and gray arrows indicate the putative cleavage sites for productive and abortive processing, respectively, based on data from the cleavage assay. (**C**) In vitro Drosha cleavage assay results on the minimal pri-miRNAs with varying bulge sizes. Internally radiolabeled miRNAs were incubated with immunopurified Drosha (Drosha) or mock preparations (Mock), and reaction products were resolved by PAGE. Productive processing is expected to result in an ~61 nt pre-miRNA. Black and gray arrows mark the presumed productive and abortive cleavage products, respectively, from Drosha cleavage as indicated in (B). Gel images are representative of at least two independent experiments. (**D**) Schematic of the in vivo miRNA expression and targeting system. GFP-targeting miRNAs were placed within the 3' UTR of a transcript encoding DsRed-Express that was expressed from the constitutive CMV promoter. (**E**) GFP silencing results from cell culture assays performed with miRNAs exhibiting varying bulge sizes. The population median of GFP fluorescence from transiently transfected HEK 293 cells stably expressing GFP was normalized to that from a construct lacking a miRNA (No miRNA). Error bars represent the standard deviation of two independent transfections.

This requirement suggests a mechanism for regulating Drosha processing, and thus RNAi-mediated gene silencing, by regulating the structure of the basal segments. To utilize this mechanism as a design strategy, we more systematically examined the relationship between bulge size in the miRNA basal segments, in vitro Drosha processing, and in vivo gene silencing.

We performed in vitro Drosha cleavage assays that mimic the first step in miRNA biogenesis to examine the relationship between bulge size and Drosha processing. RNAs with varying bulge sizes in the basal segments (Figure 4.1B) were transcribed in vitro, incubated with immunopurified Drosha, and resolved by PAGE (Figure 4.1C). Decreasing the size of the loop from 18 to 8 nts greatly reduced the appearance of the 61 nt pre-miRNA (Figure 4.1B, black arrows) and increased abortive processing (Figure 4.1B, gray arrows). Abortive processing was previously observed and attributed to DGCR8 recognition of the terminal loop instead of the miRNA basal segments (Han *et al*, 2006).

To examine the relationship between bulge size in the basal segments and gene silencing, we developed a general cell culture assay for miRNA activity (Figure 4.1D). miRNAs designed to target a transcript encoding the green fluorescent protein (GFP) were inserted into the 3' untranslated region (UTR) of a constitutively-expressed transcript encoding the fluorescent protein DsRed-Express. Plasmid DNA encoding the DsRed-Express construct was transfected into HEK 293 cells stably expressing GFP. The level of miRNA-mediated gene silencing was determined by flow cytometry analysis, where DsRed-Express levels were used to distinguish between transfected and untransfected cells. Using the cell culture assay, miRNAs with the same bulges as in the

in vitro experiments were tested for silencing efficiency. Flow cytometry results supported the data from the in vitro Drosha cleavage assays, as smaller bulge sizes showed reduced GFP silencing (Figure 4.1E). Therefore, proper Drosha processing and gene silencing correlate with the size of the bulge in the miRNA basal segments.

Aptamer integration renders miRNA processing sensitive to a small molecule ligand

Aptamers often undergo transitions from unstructured to highly structured conformations upon ligand binding, a phenomenon termed adaptive recognition (Hermann and Patel, 2000). We developed a design strategy based on this phenomenon and the dependence of Drosha processing on the structure of the basal segments to introduce ligand control of miRNA-mediated gene silencing (Figure 4.2A). Through integration of an aptamer into the basal segments of a miRNA, we anticipated that the aptamer-ligand binding interactions would decrease the unstructured nature of that region, thereby inhibiting proper processing and subsequent gene silencing (Figure 4.2B).

We first examined the ability of an aptamer to mediate ligand control of Drosha processing. The theophylline aptamer (Jenison *et al*, 1994; Zimmermann *et al*, 2000) was inserted in the basal segments domain directly adjacent to the miRNA lower stem (Figure 4.2C). The resulting miRNA (th1) was transcribed in vitro and subjected to the Drosha cleavage assay in the presence or absence of theophylline. The primary products of Drosha processing for th1 and a control miRNA were both ~61 nts, the expected size for the pre-miRNA (Han *et al*, 2006; Lee *et al*, 2003) (Figure 4.2D). In addition, the presence of theophylline inhibited proper processing of the aptamer-containing miRNA, resulting in an alternative cleavage pattern similar to that observed from miRNAs with smaller

bulges (Figure 4.1C). The control miRNA exhibited negligible theophylline dependence, suggesting that ligand binding to an aptamer located within the basal segments can control Drosha processing.

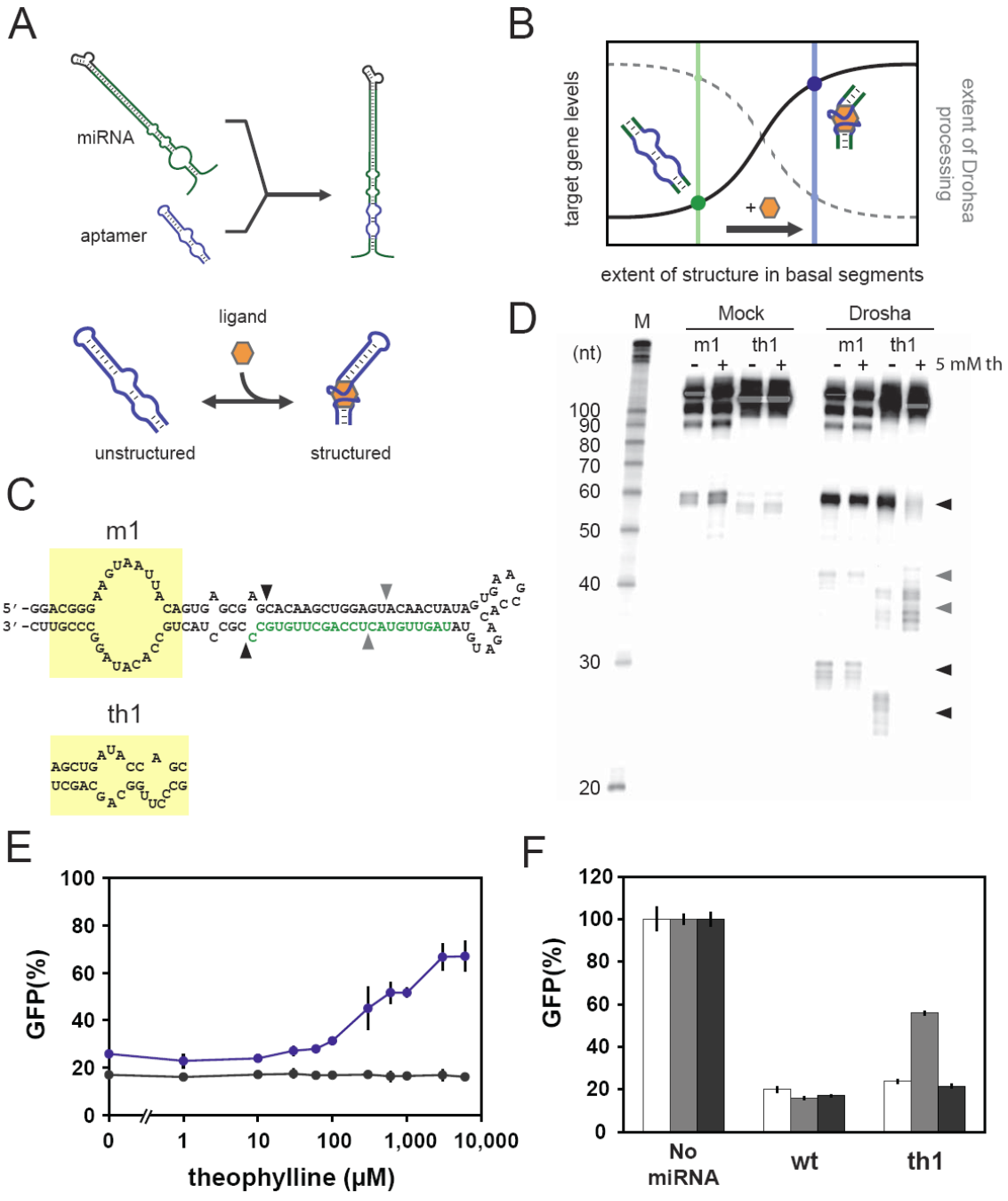


Figure 4.2 Ligand-responsive miRNAs enable ligand-mediated regulation of Drosha processing and gene silencing. (**A**) A design framework for ligand-responsive miRNAs based on integrating the aptamer binding core into the miRNA basal segments directly adjacent to the lower stem. Drosha processing of the miRNA is inhibited through an increase in structure of the basal segments domain resulting from the binding interaction between the aptamer and its cognate ligand. (**B**) Proposed relationship between Drosha processing (dashed gray line), target gene expression levels (black line), miRNA basal segments structure, and ligand concentration for a ligand-responsive miRNA. Unstructured basal segments lead to efficient processing and gene silencing, whereas structured basal segments, resulting from ligand binding and favored as the ligand concentration increases, inhibit Drosha processing. (**C**) Sequence and secondary structures of minimal GFP-targeting pri-miRNAs with a large bulge (m1) or the theophylline aptamer (th1) inserted in the basal segments. Notation follows that indicated in Figure 4.1B. (**D**) In vitro Drosha cleavage assay results for m1 and th1. Internally radiolabeled miRNAs were incubated with immunopurified Drosha (Drosha) or mock preparations (Mock) in the presence or absence of 5 mM theophylline, and reaction products were resolved by PAGE. Black and gray arrows mark the presumed productive and abortive cleavage products, respectively, from Drosha cleavage as indicated in (**c**). Gel images are representative of at least two independent experiments. (**E**) Theophylline response curves for constructs harboring a GFP-targeting miRNA with basal segments containing the theophylline aptamer (th1, blue) or sequence similar to miR-30a (wt, gray) (Figure S4.1). miRNAs were cloned into the plasmid constructs and characterized through the cell culture assays described in Figure 4.1D. The population median of GFP fluorescence from transiently transfected HEK 293 cells stably expressing GFP was normalized to that from a construct lacking a miRNA (No miRNA) at each theophylline concentration. wt was used as a negative control as it results in similar levels of GFP silencing in the absence of theophylline. (**F**) GFP silencing results from cell culture assays performed on th1 and wt transiently transfected in the absence (white) or presence of either 5 mM theophylline (gray) or 1 mM caffeine (black).

GFP levels were normalized to that from a construct lacking a miRNA (No miRNA) transfected under the same conditions. Error bars represent the standard deviation of two independent transfections.

We next examined whether theophylline regulation of Drosha processing resulted in ligand-mediated control of gene silencing. Using the cell culture assay, we tested the silencing efficiency of a GFP-targeting miRNA with a theophylline aptamer in the basal segments (th1). Extensive base pairing below the bulge encoded in the aptamer sequence was incorporated to ensure proper aptamer folding. We also included a GFP-targeting miRNA with basal segments similar to the natural miRNA miR-30a as a control (wt, Figure S4.1). Transient transfections were conducted in the presence of varying concentrations of theophylline. Results show that both miRNA constructs silence GFP with comparable strength in the absence of theophylline (Figure 4.2E). However, the miRNA with the theophylline aptamer mediated a theophylline dose-dependent increase in GFP expression levels, whereas silencing by the miRNA lacking the aptamer was insensitive to theophylline. In addition, silencing by both miRNAs was insensitive to the presence of caffeine (Figure 4.2F), a molecule that differs from theophylline by a single methyl group and binds the theophylline aptamer with a 10,000-fold lower affinity (Jenison *et al*, 1994). The results demonstrate that the observed effect of theophylline on miRNA-mediated gene silencing is specific to the incorporation of the theophylline aptamer in the basal segments of the miRNA.

Framework modularity supports the integration of different aptamer and miRNA targeting sequences

A desirable property of any ligand-responsive regulatory system is modularity. Modular regulatory systems can be readily modified to change the targeted gene or recognized ligand without complete redesign, facilitating the rapid implementation of base designs in diverse applications with varying regulatory needs. While most ligand-responsive RNA regulator designs can be readily modified to target different genes, only a fraction of the developed designs have been shown to support direct insertion of different aptamer sequences (Bayer and Smolke, 2005; Beisel *et al*, 2008; Win and Smolke, 2007).

The targeted gene is specified by the mature miRNA sequence in a miRNA regulatory element. Previous work has shown that modifying the mature miRNA sequence in natural miRNAs is sufficient to target different genes (Zeng *et al*, 2002). Since the aptamer and mature miRNA sequences do not overlap, we expected that each of these elements can be changed independently without affecting the other's function. Using the cell culture assay, we tested two different mature miRNA sequences: one that targets a different location in the GFP encoding transcript (th2) and the other that is partially scrambled (th1'). Flow cytometry results showed GFP silencing and theophylline-dependent up-regulation of gene expression by th2 and negligible silencing by th1' (Figure S4.2). Therefore, the mature miRNA sequence can be modified without compromising the ligand control function encoded within the aptamer sequence.

To examine the modularity of the aptamer sequence, we tested two aptamers that display different lengths and secondary structures: the tetracycline aptamer (Berens *et al*,

2001) and the xanthine aptamer (Kiga *et al*, 1998). The binding core of each aptamer was initially integrated adjacent to the lower stem in place of the bulge (Figure 4.3A), and the resulting miRNAs were tested using the cell culture assay in the presence or absence of the associated ligand. Hypoxanthine was used as a soluble alternative to xanthine that binds the aptamer with comparable affinity (Kiga *et al*, 1998). Control miRNAs that resulted in similar levels of gene silencing as the ligand-responsive miRNAs in the absence of ligand were also tested to determine any non-specific impacts of ligand addition on GFP levels. The miRNA harboring the tetracycline aptamer (tc1) down-regulated GFP and mediated a tetracycline-dependent increase in GFP levels (Figure 4.3B). The control miRNA (m2) delivered a similar extent of silencing with negligible tetracycline dependence, indicating that insertion of the tetracycline aptamer rendered gene silencing sensitive to tetracycline. However, compared to the theophylline aptamer, insertion of the tetracycline aptamer imparted reduced silencing and ligand sensitivity. The altered silencing in the absence of ligand may be attributed to the nature of the unbound aptamer structure, where the tetracycline aptamer folds into a preformed pocket (Muller *et al*, 2006). The altered regulatory response may be attributed to aptamer affinity, the relative membrane permeability of each small molecule, and the extent to which each aptamer adopts a more structured conformation upon ligand binding.

In contrast, insertion of the binding core of the xanthine aptamer (xa1) completely abolished silencing (Figure 4.3C). The size of the bulge in the basal segments of xa1 was similar to m4, the miRNA with the smallest bulge tested (Figure 4.1B), such that the small size of the xanthine binding core may similarly prevent proper Drosha processing. We speculated that an aptamer with a small binding core bulge can be made more

unstructured by including additional bulges, thereby restoring proper processing. Most aptamers selected in vitro contain loops that are separate from the binding core. To decrease the structure of the basal segments without compromising binding activity, we included the loop of the original xanthine aptamer and inverted the aptamer sequence to ensure hypoxanthine binding was proximal to the lower stem (xa2). Results from the cell culture assay showed significant GFP silencing and a hypoxanthine-dependent increase in GFP levels (Figure 4.3C), whereas hypoxanthine had no effect on a control miRNA with a similar silencing strength (m1). To further probe the specificity of ligand dependence, we repeated the cell culture assays using tc1 or xa2 with tetracycline or hypoxanthine. Results showed that GFP levels increased only when each aptamer was paired with its associated ligand (Figure 4.3D).

Engineering ligand-responsive miRNA clusters for tunable genetic control

Multiple miRNAs can be naturally found in clusters within a single transcript, such as miR-17-92 (He *et al*, 2005) and miR-34b-34c (He *et al*, 2007). Each miRNA within a cluster is individually processed such that cells can efficiently regulate multiple miRNAs through a single promoter. This property provides a level of regulatory efficiency that is currently unavailable to shRNAs, which must be paired with individual promoters to increase copy number (Gonzalez *et al*, 2005) or achieve combinatorial control (Beisel *et al*, 2008). By integrating multiple miRNAs into the same transcript, researchers have exploited this architecture to target multiple genes or tune gene silencing (Aagaard *et al*, 2008; Sun *et al*, 2006; Xia *et al*, 2006). Integrating ligand-responsive

miRNAs into clusters would allow tuning of the regulatory response, simultaneous regulation of different targets, and multi-input control.

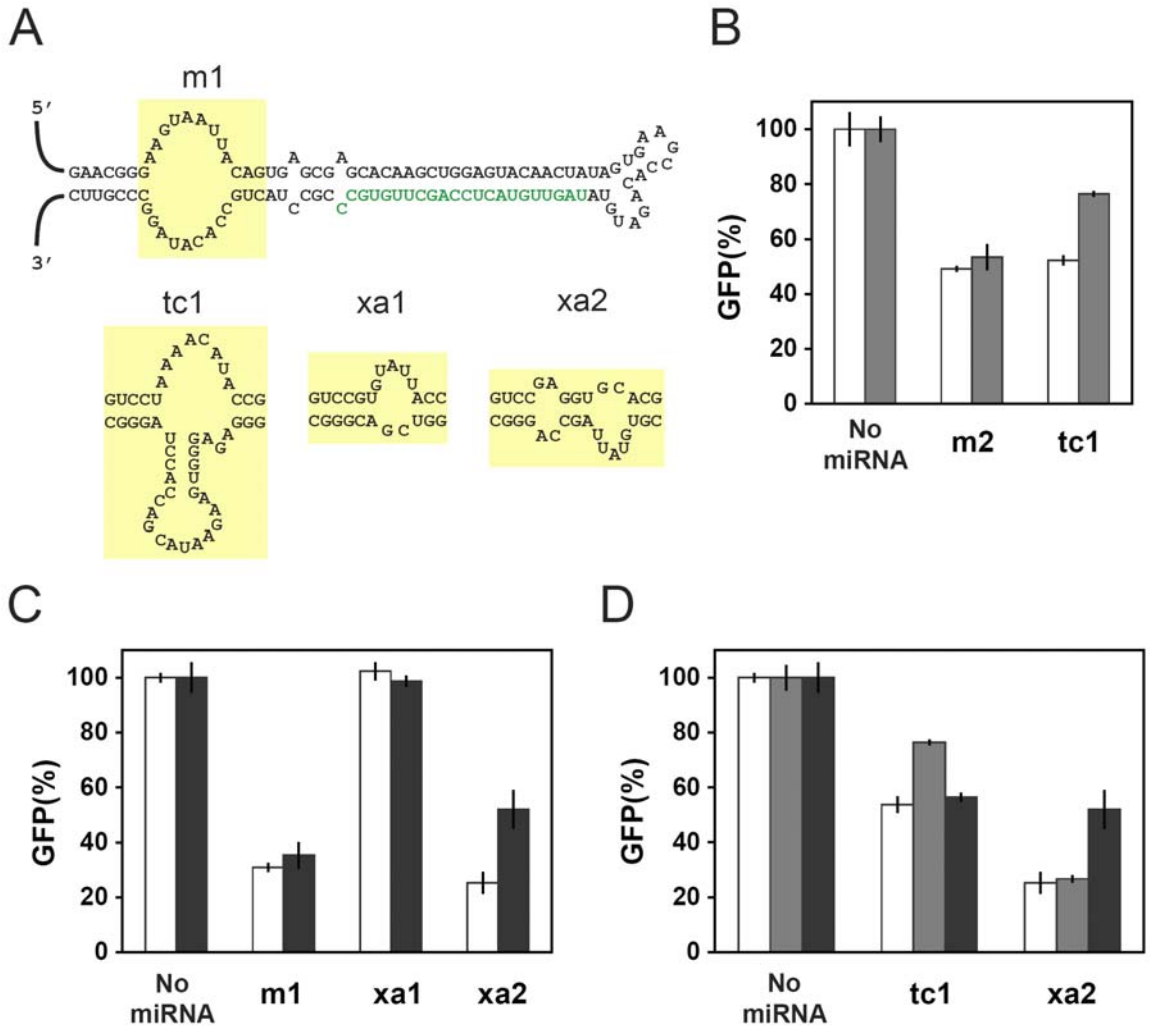


Figure 4.3 Ligand-responsive miRNAs can accommodate different aptamers to tailor the input-responsiveness of the regulatory system. (**A**) Sequence and secondary structures of GFP-targeting pri-miRNAs with the tetracycline (tc1) or xanthine (xa1, xa2) aptamers inserted in the basal segments. The aptamer binding core was inserted for tc1 and xa1, whereas the binding core and loop were inserted for xa2. Notation follows that indicated in Figure 4.1B. (**B-D**) GFP silencing

results for tetracycline- (tc1) and hypoxanthine-responsive (xa1, xa2) miRNA constructs transiently transfected in the absence (white) or presence of either 100 μM tetracycline (gray) or 5 mM hypoxanthine (black). The GFP-targeting miRNAs were cloned into the plasmid constructs and characterized through the cell culture assays described in Figure 4.1D. The population median of GFP fluorescence from transiently transfected HEK 293 cells stably expressing GFP was normalized to that from a construct lacking a miRNA (No miRNA) transfected under the same conditions. m1 and m2 were used as negative controls as they result in similar levels of GFP silencing as xa2 and tc1, respectively, in the absence of ligand. Error bars represent the standard deviation of two independent transfections.

Natural miRNAs in clusters range from being directly adjacent to one another (He *et al*, 2005) to being separated by hundreds of nucleotides (He *et al*, 2007). The variability in spacing may be important in Drosha processing and gene silencing, although the effect of spacing on miRNA activity has not been assessed to date. Therefore, we examined the relationship between the spacer length connecting two ligand-responsive miRNAs and the resulting gene silencing and ligand control. We inserted a second copy of the theophylline-responsive, GFP-targeting miRNA (th1) upstream from the first copy in the 3' UTR of the transcript encoding DsRed-Express with different spacer lengths (Figure 4.4A) and performed the cell culture assay with theophylline. To keep the local sequence around each miRNA consistent, the spacer sequences were identical to the sequence downstream of the first miRNA up to the poly(A) signal. Spacer lengths were measured between the bottoms of each stem below the aptamer and ranged from 2 to 112 nt. Adjacent placement of the miRNAs compromised both silencing and the response to theophylline potentially due to disrupted

miRNA folding or steric hindrance of Drosha processing, whereas increased spacer length restored and even exceeded the silencing activity and theophylline-dependence from a single copy (Figure 4.4B). The results suggest that separating the identical miRNAs a minimal distance improves processing and gene silencing.

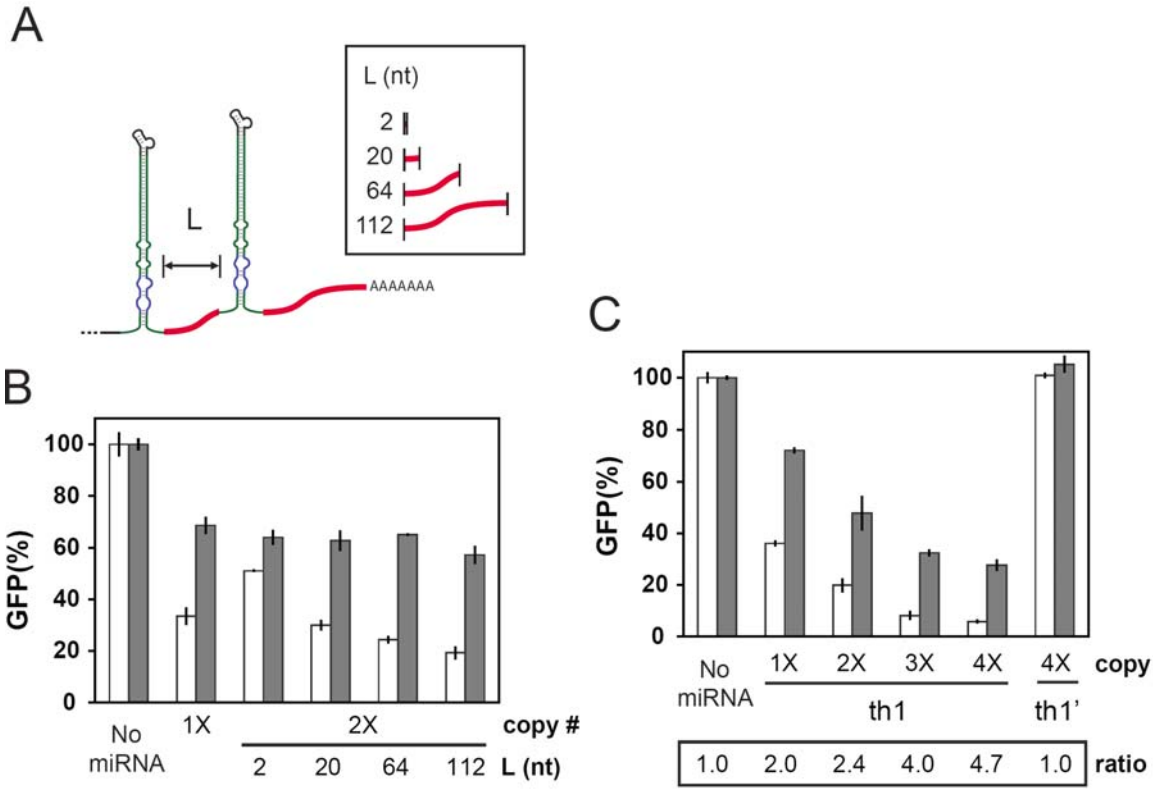


Figure 4.4 Synthetic ligand-responsive miRNA clusters allow tuning of the regulatory response. (**A**) Schematic of a synthetic miRNA cluster in which multiple ligand-responsive miRNAs are placed in the 3' UTR of a transgene encoding transcript. The spacer sequence downstream of each miRNA (indicated in red) was kept consistent, and the spacer length (L) was varied between 2 and 112 nts. Multiple copies (from 1 to 4) of a single miRNA were sequentially inserted. (**B**) The impact of spacer length between two theophylline-responsive GFP-targeting miRNAs (2X,

th1) on gene silencing in the presence (gray) or absence (white) of 5 mM theophylline. The GFP-targeting miRNAs were cloned into the plasmid constructs and characterized through the cell culture assays described in Figure 4.1D. GFP levels are reported as described in Figure 4.3. The GFP silencing from a single-copy theophylline-responsive miRNA construct (1X, th1) is shown for comparison. (**C**) The impact of ligand-responsive miRNA copy number on gene silencing and dynamic range. Multiple copies (#X) of the GFP-targeting (th1) or non-targeting (th1', Figure S4.2) theophylline-responsive miRNAs were cloned into the plasmid constructs described in Figure 4.1D using the largest spacer length tested (112 nt). GFP levels were characterized and reported as described in (B), and the dynamic range is reported as the ratio of GFP levels in the presence and absence of theophylline.

Gene silencing and the dynamic range increased when two copies of a ligand-responsive miRNA were separated by the longest spacer tested (112 nt). As miRNAs in natural clusters are individually processed, we expected that inserting additional copies separated by appropriate spacer lengths would further improve silencing and the theophylline response. Constructs harboring up to four copies of the theophylline-responsive, GFP-targeting miRNA (th1) or four copies of the non-targeting variant (th1') separated by the longest spacer sequence were subjected to the cell culture assay. GFP silencing increased with each additional copy of th1, whereas four copies of th1' had no effect on GFP levels (Figure 4.4C). The addition of each miRNA copy provided more miRNAs for Drosha processing, resulting in greater production of mature miRNAs. The presence of additional miRNA copies also increased the number of opportunities to inhibit Drosha processing, thereby increasing the dynamic range (measured as the ratio between the expression activity in the presence and absence of theophylline).

Therefore, changing the copy number of ligand-responsive miRNAs provides one approach to coordinately tune gene silencing and the dynamic range. The diminished GFP levels in the presence of theophylline may be attributed to the inability to access higher theophylline concentrations due to cytotoxicity (Beisel and Smolke, 2009) and incomplete inhibition of Drosha processing when theophylline is bound to the aptamer.

Ligand-responsive miRNA clusters can regulate endogenous genes

Most of the synthetic ligand-responsive RNA-based regulatory systems are encoded in the target transcript, providing regulation in cis (Desai and Gallivan, 2004; Ogawa and Maeda, 2008; Suess *et al*, 2004; Suess *et al*, 2003; Thompson *et al*, 2002; Win and Smolke, 2007; Yen *et al*, 2004). However, the regulation of endogenous genes through cis regulatory strategies is currently limiting due to the lack of directed recombination technologies. To test whether ligand-responsive miRNAs provide a trans regulatory strategy for the effective regulation of endogenous genes, we cloned a miRNA that targets the endogenous La gene (La1) into the 3' UTR of a transcript encoding DsRed-Express (Figure 4.5A). Under transient transfection conditions, La1 resulted in only 40% knockdown as measured by qRT-PCR (data not shown). To improve silencing, we cloned four copies of La1 or a La-targeting miRNA harboring the theophylline aptamer (La2) separated by the longest spacer sequence into the 3' UTR of the transcript encoding DsRed-Express. We made HEK 293 stable cell lines with single integrands of these constructs in the same genomic locus to reduce variability in the bulk qRT-PCR measurements. The resulting lines were grown in the presence or absence of theophylline for over one week and assayed for relative La transcript levels by qRT-PCR. Four copies

of either La1 or La2 resulted in relatively strong silencing of the La target, and only La2 mediated a theophylline-dependent increase in La transcript levels (Figure 4.5B). The results demonstrate that ligand-responsive miRNAs can control endogenous genetic targets, providing a control strategy that does not physically disrupt the locus of the target gene.

Ligand-responsive miRNAs can control gene expression in cis

Drosha cleavage separates the pre-miRNA from upstream and downstream sequences. When the miRNA is located in the 3' UTR of a transcript, Drosha cleavage is anticipated to separate the coding region from the poly(A) tail, resulting in the prevention of translation and facilitation of mRNA degradation (Dreyfus and Regnier, 2002; Sachs and Varani, 2000). Drosha was recently shown to cleave a naturally-occurring pseudo-miRNA in the transcript of DGCR8 to regulate the activity of the Microprocessor (Han *et al*, 2009). In addition, a recent study showed that introducing a 3' UTR-encoded miRNA down-regulated expression from the transcript harboring the miRNA (Stern *et al*, 2008). As such, transcripts containing 3' UTR-encoded ligand-responsive miRNAs are expected to be down-regulated in a ligand-dependent manner.

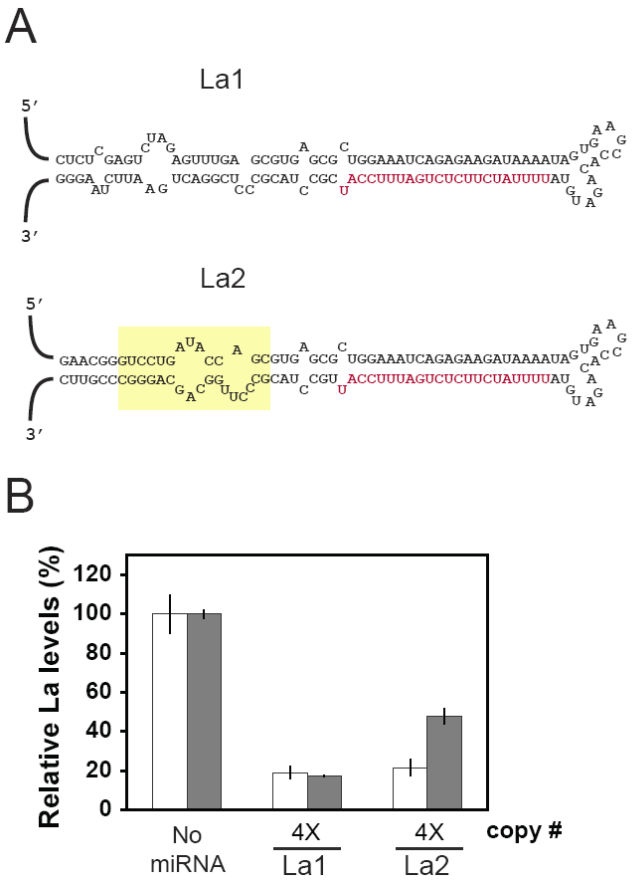


Figure 4.5 Ligand-responsive miRNA clusters can effectively control expression of endogenous gene targets. (**A**) Sequence and secondary structures of miRNAs that target the endogenous La gene. Color schemes are identical to Figure 4.1B, except that the mature miRNA sequence complementary to the La transcript is indicated in red. Sequences similar to miR-30a (La1) or the theophylline aptamer (La2) were inserted into the miRNA basal segments. miRNAs were cloned into the plasmid constructs described in Figure 4.1D at the indicated copy numbers using the largest spacer length tested (112 nt). The resulting constructs were stably transfected into HEK 293-Flp-In cells. (**B**) Relative La transcript levels for stable cell lines expressing the La-targeting miRNA constructs in the presence (gray) or absence (white) of 1.5 mM theophylline. La transcript levels were measured through qRT-PCR and normalized to GAPDH encoding transcript levels as an internal control. Relative levels are normalized to that of cells stably

transfected with the construct lacking a miRNA (No miRNA) grown under the same conditions. Error bars represent the calculated error of quadruplicate qRT-PCR wells of each sample.

To assess the capacity for ligand-responsive miRNAs to regulate gene expression in cis, we measured expression levels for the transcripts harboring the ligand-responsive miRNAs targeting GFP and La. DsRed-Express levels were quantified by flow cytometry under similar conditions as the trans-targeted gene silencing experiments. The effects of increasing copy number of the miRNA on cis regulation were examined for the th1 and th1' series. DsRed silencing increased with increasing copy number of the theophylline-responsive, GFP-targeting miRNA (th1), and two copies were sufficient to introduce ligand regulation (Figure 4.6A). Similar effects were observed with four copies of the non-targeting miRNA (th1'), indicating that transcript silencing and ligand control are independent of downstream processing. Transcript analysis confirmed that the predominant regulatory mechanism of ligand-responsive miRNAs in cis is mRNA destabilization (Figure S4.3). A similar analysis on the stably integrated La-targeting miRNAs indicated that four copies of La1 or La2 resulted in significant DsRed-Express down-regulation (Figure 4.6B). In addition, the miRNA containing the theophylline aptamer (La2) conferred an increase in DsRed levels in the presence of theophylline that was not observed for the construct lacking the theophylline aptamer (La1). These results suggest that ligand-responsive miRNAs can regulate transcripts in cis by modulating Drosha cleavage.

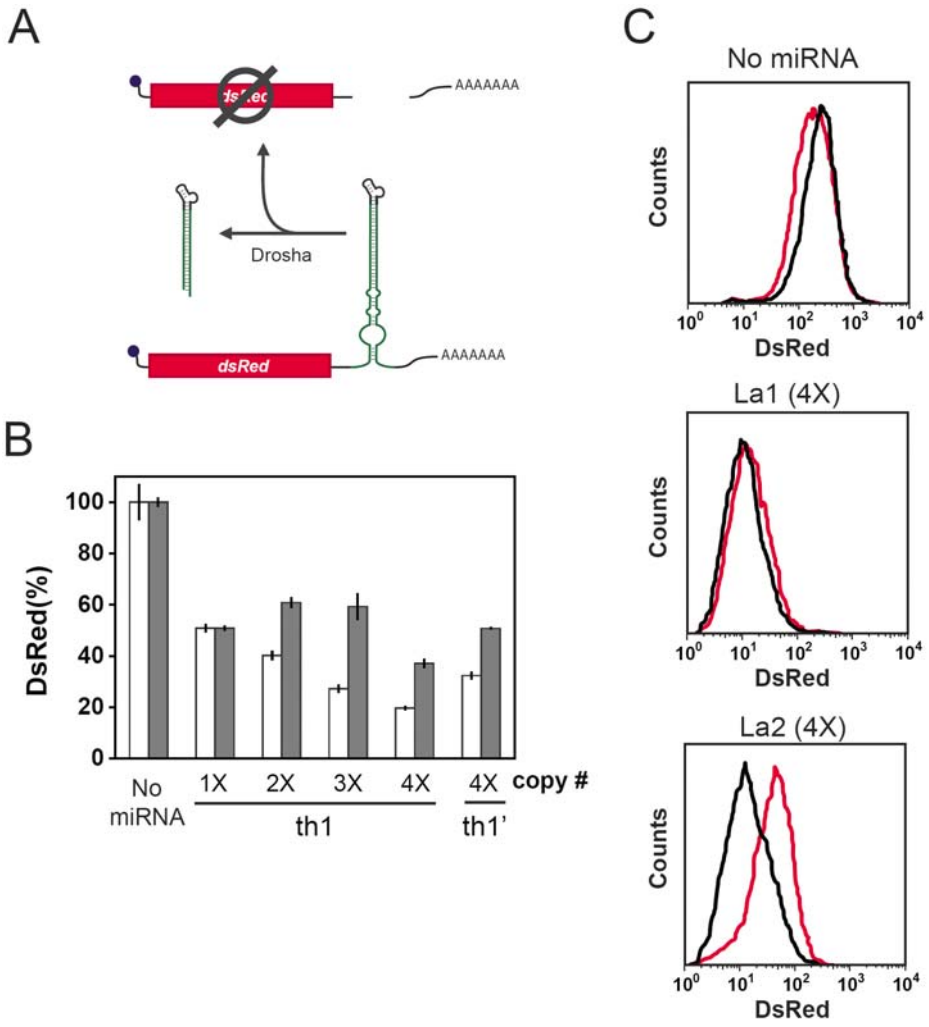


Figure 4.6 Ligand-responsive miRNAs can control transgene expression in cis. (**A**) Schematic of DsRed regulation in cis through miRNA cleavage. Drosha processing of the miRNA located in the 3' UTR separates the coding region from the poly(A) tail, thereby inactivating the transcript. (**B**) The impact of ligand-responsive miRNA copy number on expression of the transgene through regulation in cis in the presence (gray) or absence (white) of 5 mM theophylline. DsRed-Express levels of the constructs tested in Figure 4.4C were characterized through identical cell culture assays. The population mean of DsRed-Express fluorescence from transiently transfected HEK 293 cells stably expressing GFP was normalized to that from a construct lacking any miRNAs (No miRNA) transfected under the same conditions. Error bars represent the standard deviation of

two independent transfections. (**C**) Flow cytometry histograms for DsRed-Express levels from the La-targeting miRNAs. The stable cell lines tested in Figure 4.5B were grown in the presence (red) or absence (black) of 1.5 mM theophylline for over a week prior to flow cytometry analysis. Histograms are representative of two independent experiments.

Self-targeting miRNAs combine trans and cis regulation for a tighter control system

Ligand-responsive miRNAs can regulate genes in trans through RISC targeting and in cis through Drosha cleavage. A control system based on combining both modes of regulation into 'self-targeting miRNAs' may offer tighter regulation while still operating within the 3' UTR of the transgene encoding transcript. We developed a dual-acting miRNA circuit based on the insertion of ligand-responsive miRNAs or miRNA clusters into the 3' UTR of a targeted GFP transcript (Figure 4.7A), where ligand-responsive control of Drosha cleavage is expected to impact direct destabilization of the cleaved transcript and subsequent RISC-mediated inactivation of other target transcripts. The resulting constructs were stably integrated into a single genomic locus in HEK 293 cells to ensure consistent expression for all constructs. Cells were grown in the presence or absence of theophylline for over a week and analyzed by flow cytometry. The self-targeting miRNAs (combined cis and trans mechanism; th1) improved both GFP silencing and the dynamic range as compared to their non-targeting counterparts (cis mechanism; th1') (Figure 4.7B**,** Figure S4.4). Similar effects were observed for the self-targeting miRNA clusters. These results demonstrate the diverse tuning capabilities when encoding ligand-responsive miRNAs in a transcript 3' UTR by implementing self-targeting and non-targeting miRNAs at different copy numbers.

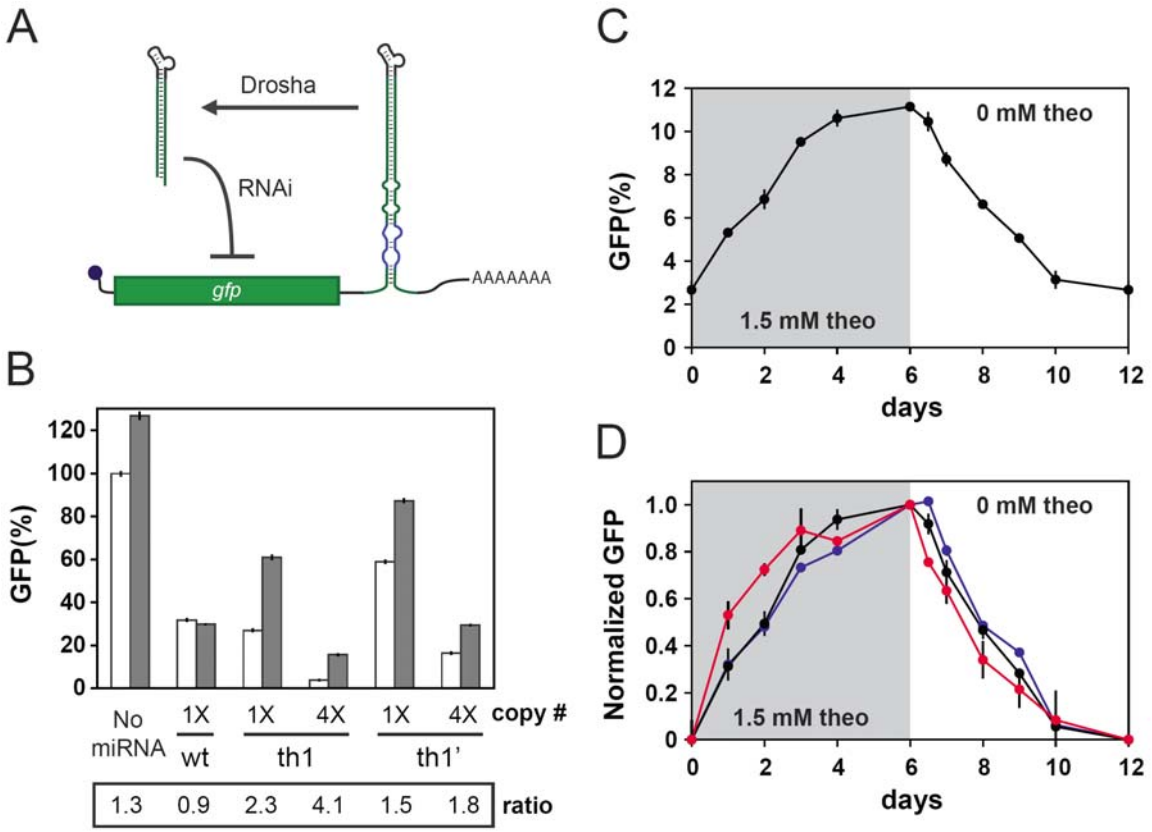


Figure 4.7 Self-targeting miRNAs provide an enhanced regulatory response. (**A**) Schematic of the miRNA regulatory circuit associated with self-targeting miRNAs. Self-targeting miRNAs are located in the 3' UTR of the trans targeted transcript encoding GFP such that Drosha cleavage and RISC targeting down-regulate expression. Both events are inhibited by ligand binding to the aptamer contained in the miRNA basal segments. (**B**) Relative GFP levels for cells stably expressing the self-targeting miRNA constructs grown in the presence (gray) or absence (white) of 1.5 mM theophylline for over a week. Gene silencing from one (1X) or four (4X) copies of a theophylline-responsive self-targeting miRNA (th1) and one or four copies of a theophylline-responsive non-targeting miRNA (th1') were determined, where multiple copies were separated by the largest spacer length (112 nt). One copy of a self-targeting miRNA with basal segments similar to miR-30a (wt) was used as a negative control. The dynamic range is reported as the ratio of GFP levels in the presence and absence of theophylline. (**C**) Temporal response of the relative

GFP levels to a change in ligand concentration for cell lines stably expressing ligand-responsive miRNAs. Representative time course data is shown for cells expressing the miRNA construct containing four copies of th1. Cells were grown in the presence of 1.5 mM theophylline for six days and then transferred to media without theophylline for six days. (**D**) Normalized GFP levels for cell lines stably expressing ligand-responsive miRNAs over time in changing concentrations of ligand. Normalized time course data are shown for cells expressing the miRNA constructs containing four copies of th1' (red), one copy of th1 (blue), or four copies of th1 (black). Time course data were normalized to zero when theophylline was added at the beginning of the time course and one when theophylline was removed in the middle of the time course to compare the dynamics of the approach to steady-state between ligand-responsive miRNAs acting through different regulatory mechanisms. Error bars represent the standard deviation of cells grown in two separate culture wells.

We performed time course studies on the dual-acting miRNA circuits to examine the dynamic properties of these regulatory systems. Cells lines were grown in the presence of theophylline for six days and then grown in media without theophylline for another six days. Cell lines harboring th1 or th1' in single or four copies exhibited increasing GFP levels when grown in the presence of theophylline and reached steady-state levels by day 6 (Figure 4.7C). GFP levels decreased upon removal of theophylline and returned to original levels after 4-6 days of growth in the absence of theophylline, indicating that the genetic control exerted by ligand-responsive miRNAs is reversible.

We further examined the effects of miRNA copy number and regulatory mechanism (cis, dual cis/trans) on the dynamics of the ligand-responsive miRNA regulatory response (Figure 4.7D). Time-course data for th1, th1 (4 copies), and th1' (4

copies) were normalized to lie between zero and one, marking the steady-state levels in the absence or presence of theophylline, respectively. Cells with four copies of the non-targeting miRNA (th1') approached steady-state faster than the self-targeting miRNA (th1), especially upon the addition of theophylline. Cells with either one or four copies of the self-targeting miRNA approached steady state at similar rates. The different rates of approach to steady-state between the non-targeting and self-targeting miRNAs can be explained by the machinery involved in each response. Transcripts harboring the non-targeting miRNAs are regulated directly by Drosha processing, which is rapidly modulated by ligand binding and release. The approach to steady-state is thus set by the rates of transcription and translation when theophylline is added and by GFP turnover and dilution when theophylline is removed. Conversely, transcripts harboring self-targeting miRNAs are regulated by both Drosha processing and subsequent RISC targeting. Therefore, the approach to steady-state levels following theophylline addition is anticipated to be slower since transcript levels will continue to increase after the turnover of activated RISC (Bartlett and Davis, 2006). However, the response to theophylline removal is more similar for non-targeting and self-targeting miRNAs due to the small time lag between Drosha processing and RISC activation in comparison to GFP turnover and dilution. The particular response dynamics associated with the cis and cis/trans regulatory mechanisms offer yet another design parameter when specifying the regulatory performance of ligand-responsive miRNAs.

DISCUSSION

We have developed a novel mode of RNA-based gene regulation in mammalian cells based on synthetic ligand-responsive miRNAs. Ligand-responsive miRNAs function through ligand-mediated regulation of Drosha processing of a pri-miRNA based on modulation of the structured nature of the basal segments region through aptamer-ligand binding interactions. The use of theophylline, tetracycline, and hypoxanthine as effective ligands in this work supports the ability to utilize diverse molecules to control miRNA activity, such as metabolites, metals, and proteins, against which aptamers have been selected (Lee *et al*, 2004). Beyond in vitro selected aptamers, aptamers present in natural riboswitches have also been integrated into RNA-based ligand control systems (Barrick and Breaker, 2007). While most aptamers conform to a standard bulge-loop structure, a significant number adopt alternative structures, such as pseudoknots (Lorsch and Szostak, 1994; Mannironi *et al*, 1997; Tuerk *et al*, 1992; Wilson *et al*, 1998) or dangling ends (Koizumi and Breaker, 2000), that may be incompatible with the ligand-responsive miRNA framework. To address such broader implementation challenges, future work may focus on the development of modified frameworks that accept more diverse aptamer structures, the modification of aptamer selection procedures to preferentially select bulge-loop structures, or the selection of aptamers within RNA-based regulatory systems to identify sequences that function as integrated sensing components (Weigand *et al*, 2008).

We showed that ligand-responsive miRNAs can be readily altered to target different genes by changing the mature miRNA sequence. A future extension of our system will be the incorporation of natural miRNA sequences to exploit natural regulatory networks. Natural miRNAs are key players in diverse cellular processes and

help enact global changes in gene expression by simultaneously targeting hundreds of genes. Therefore, ligand-responsive miRNAs can be directly interfaced with the regulatory architecture controlling complex cellular processes. By implementing ligand-responsive miRNAs that respond to endogenously-expressed molecules, these regulatory molecules will provide a platform for reprogramming cellular state according to the intracellular environment, supporting autonomous approaches to tissue engineering and disease treatment.

We also found that gene silencing and ligand responsiveness were dependent on the spacer length between miRNAs in synthetic miRNA clusters. The variable spacing within natural-occurring miRNA clusters suggests that secondary structure of the intervening sequence may be a key factor in miRNA processability and that spacing may be an evolutionary factor to tune miRNA processing and subsequent gene silencing activity. Future studies that correlate the secondary structures of natural miRNA clusters with the extent of processing of each miRNA will provide further insight into this relationship. From this enhanced understanding, molecular engineers may be able to design synthetic cluster expression platforms that will minimize the cluster length, while maximizing the processing of the encoded miRNAs (Aagaard *et al*, 2008). Scalable synthetic cluster expression platforms will also be important tools in the development of sophisticated gene regulatory circuits based on simultaneous implementation of multiple ligand-responsive miRNAs to achieve finely tuned, combinatorial gene expression control schemes. Such scalable systems will require an understanding of how to engineer molecular platforms that effectively insulate the function of individual miRNA elements within larger clusters and thus an enhanced understanding of RNA structure-function

relationships. However, the practical size of an engineered miRNA cluster may be limited by downstream bottlenecks in the linear pathway of miRNA biogenesis, such as Drosha cleavage, nuclear export, Dicer processing, and RISC loading and targeting.

The linear pathway of miRNA biogenesis offers multiple points at which to engineer ligand-responsive control of RNAi-mediated gene silencing. Besides regulation of Drosha cleavage, RNA-based regulatory frameworks have been reported that regulate nuclear export (Beisel *et al*, 2008) and Dicer processing (An *et al*, 2006; Beisel *et al*, 2008). Since miRNA biogenesis spans the nuclear and the cytoplasmic compartments of the cell, specific regulatory approaches may be more appropriate for a given application depending on the localization properties of the ligand. Furthermore, more sophisticated regulatory strategies can be designed that combine regulatory modes to coordinately modulate multiple points in the processing pathway by integrating aptamers to the same ligand or different ligands.

The synthetic ligand-responsive miRNA systems reported here may foreshadow a prevalent strategy used by natural systems to achieve sophisticated control over miRNA regulatory networks. The successful design of such synthetic systems may suggest relevant architectures to assist in the identification of natural counterparts. Researchers have previously identified mechanisms that modulate miRNA activity based on binding interactions between RNA binding proteins and miRNA loops (Guil and Caceres, 2007) and target sites (Kedde *et al*, 2007), lending further support to the likelihood of an as yet undiscovered natural regulatory strategy based on small molecule control of miRNA activity.

MATERIALS AND METHODS

Plasmid construction. The coding region of DsRed-Express was initially cloned with the consensus Kozak sequence (CGCCACC) into the NheI/XhoI restriction sites of pcDNA3.1(+). pcDNA3.1(+) contains the constitutive CMV promoter upstream of a multicloning site. Ligand-responsive and control miRNAs reported in Table S4.1 were cloned into XbaI/ApaI downstream of each coding region. To construct synthetic miRNA clusters, additional miRNAs were digested with AvrII/XhoI and iteratively inserted into XhoI/XbaI within the miRNA-containing plasmid. For the 2-nt spacer, the second miRNA (th3) was separately prepared and inserted. For all other spacer lengths, the original miRNA was amplified with a common forward primer and a reverse primer that hybridizes different distances downstream of the miRNA: Sp.fwd, 5'-GTTCCTGTAGACGGCTCTC-3'; Sp1.rev, 5'-AATACCTAGGCTGATCAGCGGGTTT-3'; Sp2.rev, 5'-AA TACCTAGGAGGGGCAAACAACAG-3'; Sp3.rev, 5'-AATA CCTAGGAAAGGACAGTGGGAGTG-3'. To make stable cell lines, the coding region of DsRed-Express was replaced with EGFP and the entire transcript was excised with NheI/NsiI and cloned into the same sites in pcDNA5/FRT (Invitrogen). Prior to insertion, the NsiI site was introduced into this plasmid at position 1524 using site-directed mutagenesis. All restriction enzymes and T4 DNA ligase were purchased from New England Biolabs. All constructs were sequence-verified (Laragen).

RNA preparation. Internally radiolabeled RNAs were transcribed *in vitro* from an annealed template containing the T7 promoter (5'-TTCTAATACGACTCACTATA**G**GG-3', where **G** is the first transcribed nucleotide) using the Ampliscribe T7 transcription kit

(Epicentre) according to the manufacturer's instructions with [α-^{32}P]-GTP. Following transcription and DNase treatment, the transcription product was purified through a NucAway clean-up column (Ambion) according to manufacturer's instructions and gel-purified by PAGE.

Drosha cleavage assays. In vitro assays were conducted as described previously (Lee and Kim, 2007). Briefly, the Drosha complex was immunopurified from 293T cells transiently transfected with pCK-Drosha-FLAG and pCK-DGCR8-FLAG (9:1 mass ratio). Two days post-transfection, cells were lysed using M-PER (Pierce) according to the manufacturer's instructions and the resulting supernatant was incubated with Anti-FLAG M2 affinity beads (Sigma Aldrich) for at least 1 hr at 4°C with rotation. The beads were then washed with the lysis buffer (20 mM Tris-HCl pH 8.0, 100 mM KCl, 0.2 mM EDTA) five times and evenly divided for the in vitro assays (two in vitro reactions from a 10 cm transfection dish). Radiolabeled RNAs (~10^5 cpm, 3 µl) were combined with 0.75 µl RNasin (Promega), 3 µl reaction buffer (64 mM $MgCl_2$), 8.25 µl water, and 15 µl immunopurified Drosha complex. After an incubation of 90 min at 37°C, the reaction was terminated with the addition of 0.5 M sodium acetate and 0.02% sodium dodecyl sulfate (SDS), phenol:chloroform extracted, and ethanol precipitated. Samples were then resuspended in 15 µl RNA loading buffer (95% formamide, 0.02% SDS, 0.025% bromophenol blue, 0.025% xylene cyanol FF) and resolved on a 12.5% denaturing polyacrylamide gel. The RNA decades ladder (Ambion) was used as a size marker.

Cell culture and transfection. 293 and Flp-In-293 cells were maintained in DMEM supplemented with 10% FBS at 37°C in a 5% CO_2–humidified incubator. Transient transfections were conducted with Fugene 6 (Roche) according to the manufacturer's instructions one day after seeding. Immediately prior to transfection, the media was supplemented with the appropriate ligand at the specified concentration. The media was replaced two days post-transfection. Three days post-transfection, the cells were trypsinized and subjected to flow cytometry analysis on a Cell Lab Quanta SC MPL (Beckman Coulter) and the resulting data was analyzed using the Flowjo software (Tree Star). Cells were initially gated for viability by electronic volume and side scatter. GFP and DsRed fluorescence of viable cells were measured through a 525 nm and 760 nm band pass filter, respectively, after excitation with a 488 nm laser. Moderate to high DsRed levels served as a transfection control to gate between transfected and untransfected cells. Transfected cells could be distinguished from untransfected cells even when four miRNAs were present in the 3'UTR of the DsRed-Express encoding transcript (data not shown). GFP levels were calculated as the median fluorescence of the transfected population divided by that of the untransfected population (Beisel *et al*, 2008). All ratios were normalized such that the value for DsRed-Express lacking a miRNA under the same conditions was set to 100%. Reported DsRed measurements are the mean value of the transfected population normalized to the construct lacking a miRNA set to 100%, where the mean value was selected based on the high variability associated with transient plasmid-based expression of fluorescent proteins.

Stable transfection of 293-Flp-In cell lines was performed using the Flp-In recombinase system (Invitrogen) according to the manufacturer's instructions to generate

isogenic stable cell lines. Integrands were selected using 200 μg/ml hygromycin B (Invitrogen), whereas stable cell lines were maintained in 50 μg/ml hygromycin B. The procedure described above for the fluorescence expression analysis of transiently transfected cell populations was used to analyze the stable cell lines with notable exceptions. Normalization of flow cytometry data to untransfected cells was not performed as all cells express the integrated construct, and all data was normalized to cells lacking a miRNA and grown in the absence of theophylline. Normalization to cells grown under the same conditions was not performed since theophylline differentially affected the two negative controls: no miRNA and a self-targeting miRNA lacking the theophylline aptamer. The different effects may be attributed to differences in perturbations induced by theophylline stress on unregulated genes and genes regulated by a self-targeting miRNA as observed for other cellular stressors (Figure S4.5).

qRT-PCR. The following oligos were used for qRT-PCR. La protein (Acc # X13697): La_fwd, 5'-GGTTGAACCGTCTAACAACAG-3'; La_rev, 5'-ATGTCATCAAGAGTTGCATCAG-3'; GAPDH (Acc # NM_002046): GAPDH_fwd, 5'-GAAGGTGAAGGTCGGAGTC-3'; GAPDH_rev, 5'-GAAGATGGTGATGGGATTTC-3'; DsRed-Express: DsRed.fwd, 5'-AAGAAGACTATGGGCTGGGA-3'; DsRed.rev 5'-CGATGGTGTAGTCCTCGTTG-3'; and the Neomycin resistance gene: NeoR.fwd, 5'-ACCTTGCTCCTGCCGAGAAAGTAT-3'; NeoR.rev, 5'-ATGTTTCGCTTGGTGGTCGAATGG-3'. Transcript levels were measured by qRT-PCR under either transient or stable transfection conditions. For transient transfections 293 cells were washed with 1X PBS three days post-transfection and total RNA was extracted using the RNeasy Mini kit (Qiagen)

according to the manufacturer's instructions. For stable transfections, cell lines were grown for over a week in the presence or absence of 1.5 mM theophylline prior to total RNA extraction. Total RNA samples were treated with DNase I at 37°C for 20 minutes and purified using a NucAway column (Ambion). Up to 5 μg of purified RNA was reverse-transcribed using Superscript III reverse transcriptase (Invitrogen) according to the manufacturer's instructions using the reverse primers for each pair of gene target and loading control (DsRed/NeoR, La/GAPDH) followed by the recommended incubation with RNase H. qRT-PCR was conducted with the resulting cDNA on the iCycler iQ system (BioRAD) according to the manufacturer's instructions. Samples were prepared in quadruplicate using the iQ SYBR green supermix and data were analyzed using the iCycler iQ software. The mean of the resulting C_T values for the target gene of each sample were subtracted from the mean C_T value for the control gene. The resulting values were then converted from $\log_2$ to linear scale and normalized to the value for the sample lacking any miRNA transfected with the same concentration of ligand. The reported sample error was calculated using the following expression:

$$\text{Sample Error} = \frac{2^{\text{AVE(Cont)}-\text{AVE(Target)}+\frac{1}{2}[\text{SD(Cont)}+\text{SD(Target)}]} - 2^{\text{AVE(Cont)}-\text{AVE(Target)}}}{\left[2^{\text{AVE(Cont)}-\text{AVE(Target)}}\right]_{\text{Neg}}} \quad (4.1)$$

where AVE and SD are the respective average and standard deviation of each quadruplicate sample, Cont and Target are the loading control and target, respectively, and Neg is the sample lacking a miRNA transfected with the same ligand concentration as the sample in question.

ACKNOWLEDGEMENTS

We thank A. Brown, Y.C. Chen, M. Greenwood-Goodwin, and J. Vowles for constructive comments on the manuscript and V.N. Kim for providing the pCK-Drosha-FLAG and pCK-DGCR8-FLAG plasmids. This work was supported by the Caltech Joseph Jacobs Institute for Molecular Engineering for Medicine (grant to C.D.S.), the Department of Defense (grant to C.D.S.), the National Institutes of Health (fellowship to K.G.H), and the National Science Foundation (fellowship to C.L.B.).

REFERENCES

Aagaard LA, Zhang J, von Eije KJ, Li H, Saetrom P, Amarzguioui M, Rossi JJ (2008) Engineering and optimization of the miR-106b cluster for ectopic expression of multiplexed anti-HIV RNAs. *Gene Ther* **15:** 1536-1549.

An CI, Trinh VB, Yokobayashi Y (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA* **12:** 710-716.

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microRNAs on protein output. *Nature* **455:** 64-71.

Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8:** R239.

Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116:** 281-297.

Bartlett DW, Davis ME (2006) Insights into the kinetics of siRNA-mediated gene silencing from live-cell and live-animal bioluminescent imaging. *Nucleic Acids Res* **34:** 322-333.

Bauer M, Kinkl N, Meixner A, Kremmer E, Riemenschneider M, Forstl H, Gasser T, Ueffing M (2009) Prevention of interferon-stimulated gene expression using microRNA-designed hairpins. *Gene Ther* **16:** 142-147.

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23:** 337-343.

Beisel CL, Bayer TS, Hoff KG, Smolke CD (2008) Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol* **4:** 224.

Beisel CL, Smolke CD (2009) Design principles for riboswitch function. *PLoS Comput Biol* **5:** e1000363.

Berens C, Thain A, Schroeder R (2001) A tetracycline-binding RNA aptamer. *Bioorg Med Chem* **9:** 2549-2556.

Boudreau RL, Martins I, Davidson BL (2009) Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs in vitro and in vivo. *Mol Ther* **17:** 169-175.

Boudreau RL, Monteys AM, Davidson BL (2008) Minimizing variables among hairpin-based RNAi vectors reveals the potency of shRNAs. *RNA* **14:** 1834-1844.

Cai X, Hagedorn CH, Cullen BR (2004) Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10:** 1957-1966.

Desai SK, Gallivan JP (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* **126:** 13247-13254.

Dreyfus M, Regnier P (2002) The poly(A) tail of mRNAs: bodyguard in eukaryotes, scavenger in bacteria. *Cell* **111:** 611-613.

Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* **346:** 818-822.

Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19:** 92-105.

Gonzalez S, Castanotto D, Li H, Olivares S, Jensen MC, Forman SJ, Rossi JJ, Cooper LJ (2005) Amplification of RNAi--targeting HLA mRNAs. *Mol Ther* **11:** 811-818.

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432:** 235-240.

Grimm D, Streetz KL, Jopling CL, Storm TA, Pandey K, Davis CR, Marion P, Salazar F, Kay MA (2006) Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441:** 537-541.

Guil S, Caceres JF (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol* **14:** 591-596.

Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN (2004) The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18:** 3016-3027.

Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125:** 887-901.

Han J, Pedersen JS, Kwon SC, Belair CD, Kim YK, Yeom KH, Yang WY, Haussler D, Blelloch R, Kim VN (2009) Posttranscriptional crossregulation between Drosha and DGCR8. *Cell* **136:** 75-84.

He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, Jackson AL, Linsley PS, Chen C, Lowe SW, Cleary MA, Hannon GJ (2007) A microRNA component of the p53 tumour suppressor network. *Nature* **447:** 1130-1134.

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM (2005) A microRNA polycistron as a potential human oncogene. *Nature* **435:** 828-833.

Hermann T, Patel DJ (2000) Adaptive recognition by nucleic acid aptamers. *Science* **287:** 820-825.

Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-resolution molecular discrimination by RNA. *Science* **263:** 1425-1429.

Kedde M, Strasser MJ, Boldajipour B, Oude Vrielink JA, Slanchev K, le Sage C, Nagel R, Voorhoeve PM, van Duijse J, Orom UA, Lund AH, Perrakis A, Raz E, Agami R (2007) RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* **131:** 1273-1286.

Kiga D, Futamura Y, Sakamoto K, Yokoyama S (1998) An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. *Nucleic Acids Res* **26:** 1755-1760.

Koizumi M, Breaker RR (2000) Molecular recognition of cAMP by an RNA aptamer. *Biochemistry* **39:** 8983-8992.

Lee JF, Hesselberth JR, Meyers LA, Ellington AD (2004) Aptamer database. *Nucleic Acids Res* **32:** D95-100.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425:** 415-419.

Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21:** 4663-4670.

Lee Y, Kim VN (2007) In vitro and in vivo assays for the activity of Drosha complex. *Methods Enzymol* **427:** 89-106.

Lorsch JR, Szostak JW (1994) In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* **33:** 973-982.

Mannironi C, Di Nardo A, Fruscoloni P, Tocchini-Valentini GP (1997) In vitro selection of dopamine RNA ligands. *Biochemistry* **36:** 9726-9734.

McBride JL, Boudreau RL, Harper SQ, Staber PD, Monteys AM, Martins I, Gilmore BL, Burstein H, Peluso RW, Polisky B, Carter BJ, Davidson BL (2008) Artificial miRNAs mitigate shRNA-mediated toxicity in the brain: implications for the therapeutic development of RNAi. *Proc Natl Acad Sci U S A* **105:** 5868-5873.

Muller M, Weigand JE, Weichenrieder O, Suess B (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Res* **34:** 2607-2617.

Ogawa A, Maeda M (2008) An artificial aptazyme-based riboswitch and its cascading system in E. coli. *Chembiochem* **9:** 206-209.

Osborne SE, Ellington AD (1997) Nucleic Acid Selection and the Challenge of Combinatorial Chemistry. *Chem Rev* **97:** 349-370.

Sachs AB, Varani G (2000) Eukaryotic translation initiation: there are (at least) two sides to every story. *Nat Struct Biol* **7:** 356-361.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* **455:** 58-63.

Stern P, Astrof S, Erkeland SJ, Schustak J, Sharp PA, Hynes RO (2008) A system for Cre-regulated RNA interference in vivo. *Proc Natl Acad Sci U S A* **105:** 13895-13900.

Suess B, Fink B, Berens C, Stentz R, Hillen W (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res* **32:** 1610-1614.

Suess B, Hanson S, Berens C, Fink B, Schroeder R, Hillen W (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res* **31:** 1853-1858.

Suess B, Weigand JE (2008) Engineered riboswitches: overview, problems and trends. *RNA Biol* **5:** 24-29.

Sun D, Melegari M, Sridhar S, Rogler CE, Zhu L (2006) Multi-miRNA hairpin method that improves gene knockdown efficiency and provides linked multi-gene knockdown. *Biotechniques* **41:** 59-63.

Thompson KM, Syrett HA, Knudsen SM, Ellington AD (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol* **2:** 21.

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249:** 505-510.

Tuerk C, MacDougal S, Gold L (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A* **89:** 6988-6992.

Weigand JE, Sanchez M, Gunnesch EB, Zeiher S, Schroeder R, Suess B (2008) Screening for engineered neomycin riboswitches that control translation initiation. *RNA* **14:** 89-97.

Wilson C, Nix J, Szostak J (1998) Functional requirements for specific ligand recognition by a biotin-binding RNA pseudoknot. *Biochemistry* **37:** 14410-14419.

Win MN, Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104:** 14283-14288.

Xia XG, Zhou H, Xu Z (2006) Multiple shRNAs expressed by an inducible pol II promoter can knock down the expression of multiple target genes. *Biotechniques* **41:** 64-68.

Yen L, Svendsen J, Lee JS, Gray JT, Magnier M, Baba T, D'Amato RJ, Mulligan RC (2004) Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* **431:** 471-476.

Zeng Y, Cullen BR (2003) Sequence requirements for micro RNA processing and function in human cells. *RNA* **9:** 112-123.

Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* **32:** 4776-4785.

Zeng Y, Cullen BR (2005) Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280:** 27595-27603.

Zeng Y, Wagner EJ, Cullen BR (2002) Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9:** 1327-1333.

Zeng Y, Yi R, Cullen BR (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **24:** 138-148.

Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6:** 659-667.

SUPPLEMENTARY INFORMATION

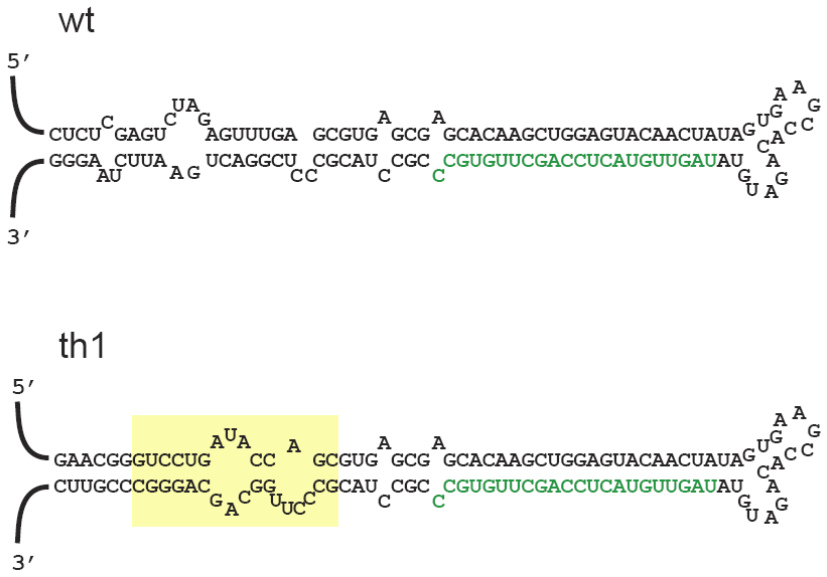


Figure S4.1 Sequence and secondary structures of miRNAs targeting GFP. Basal segments contain sequences that are similar to miR-30a (wt) or the theophylline aptamer (th1). The aptamer insertion site is indicated by the yellow box according to Figure 4.3A and the mature miRNA sequence complementary to the GFP transcript is indicated in green text.

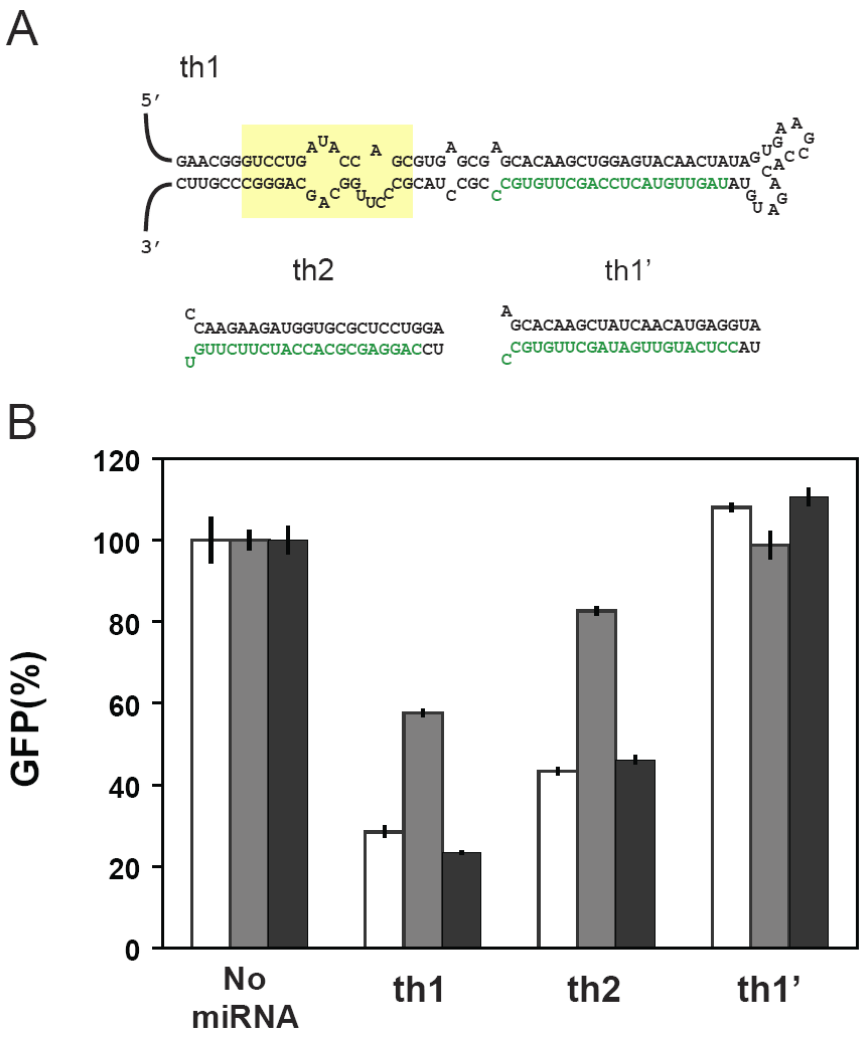


Figure S4.2 Ligand-responsive miRNAs can accommodate different mature miRNA sequences to tailor the gene silencing output of the regulatory system. (**A**) The mature miRNA sequence contained within the upper stem of th1 was modified to target a different sequence within the GFP mRNA (th2) or abolish targeting (th1’). All miRNAs contain the theophylline aptamer in the basal segments. The aptamer insertion site is indicated by the yellow box according to Figure 4.3A, and mature miRNA sequences are indicated in green text. The GFP-targeting miRNAs were cloned into the plasmid constructs and characterized through the cell culture assays described in Figure 4.1D. (**B**) GFP silencing results for theophylline-responsive miRNA constructs (th1, th2, or th1’) transiently transfected in the absence (white) or presence of either 5

mM theophylline (gray) or 1 mM caffeine (black). The population median of GFP fluorescence from transiently transfected HEK 293 cells stably expressing GFP was normalized to that from a construct lacking a miRNA (No miRNA) transfected under the same conditions. Error bars represent the standard deviation of two independent transfections.

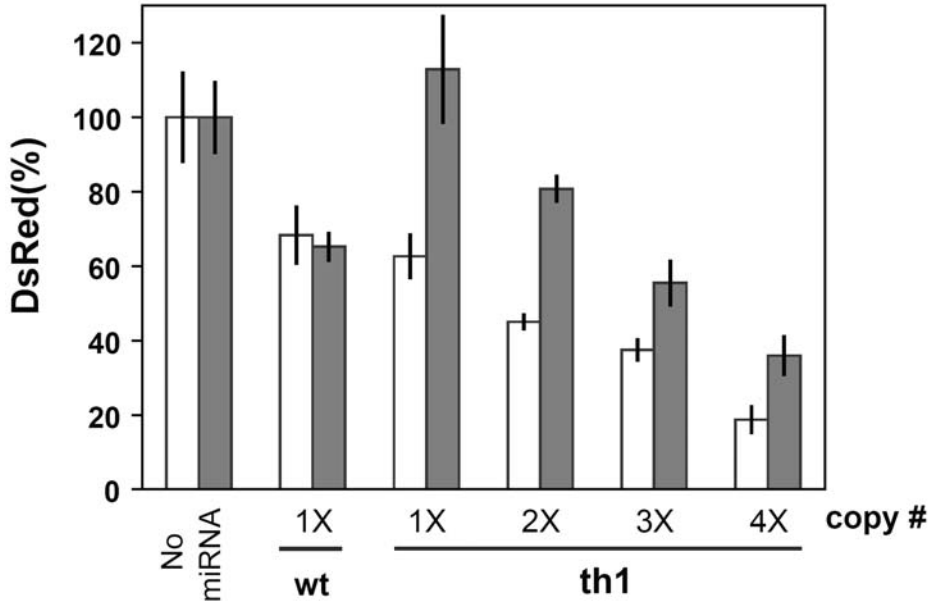


Figure S4.3 Ligand-responsive miRNAs control gene expression in cis through transcript destabilization. Multiple copies (#X) of the GFP-targeting (th1) theophylline-responsive miRNAs were cloned into the plasmid constructs described in Figure 4.1D using the largest spacer length tested (112 nt). wt was used as a negative control to allow direct comparison to Figure 4.4B. HEK 293 cells stably expressing GFP were transiently transfected with these constructs in the presence (gray) or absence (white) of 5 mM theophylline. DsRed transcript levels were measure through qRT-PCR and normalized to transcript levels of the Neomycin resistance gene expressed from the same transfected plasmid. Relative levels were normalized to that of cells transfected with the construct lacking any miRNAs (No miRNA) grown under the same conditions. Error bars represent the calculated error of quadruplicate qRT-PCR wells for each sample.

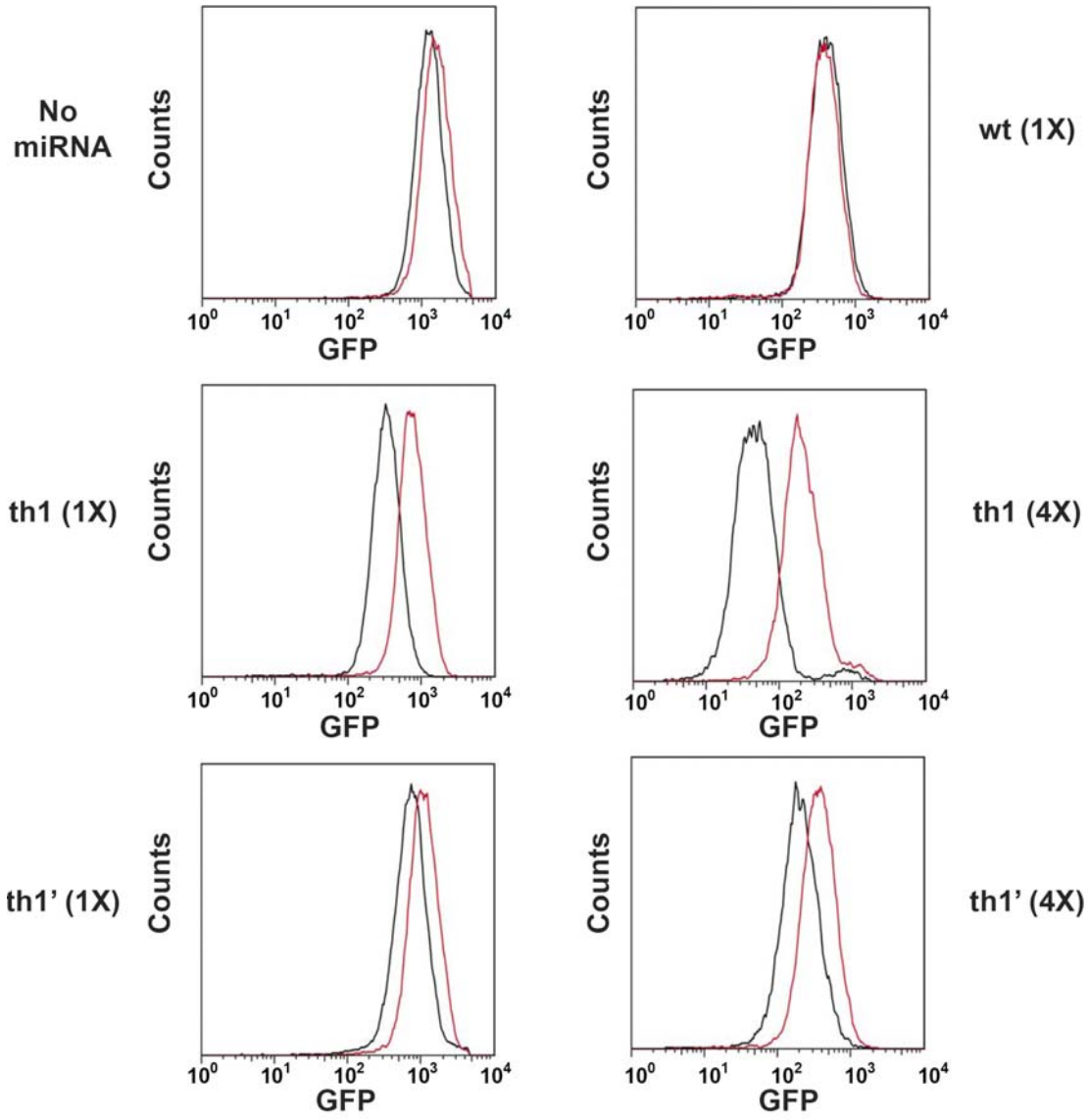


Figure S4.4 Flow cytometry histograms for HEK 293-Flp-In cells stably expressing the self-targeting miRNA constructs. miRNAs were located in the 3' UTR of the trans-targeted transcript encoding GFP. Constructs harboring no miRNAs (No miRNA), one copy of a self-targeting miRNA with basal segments containing sequences similar to miR-30a (wt), one (1X) or four (4X) copies of a theophylline-responsive self-targeting miRNA (th1), and one or four copies of a theophylline-responsive non-targeting miRNA (th1') were characterized, where multiple copies were separated by the largest spacer length tested (112 nt). Stable cell lines were grown for over

one week in the presence (red) or absence (black) of 1.5 mM theophylline prior to flow cytometry analysis.

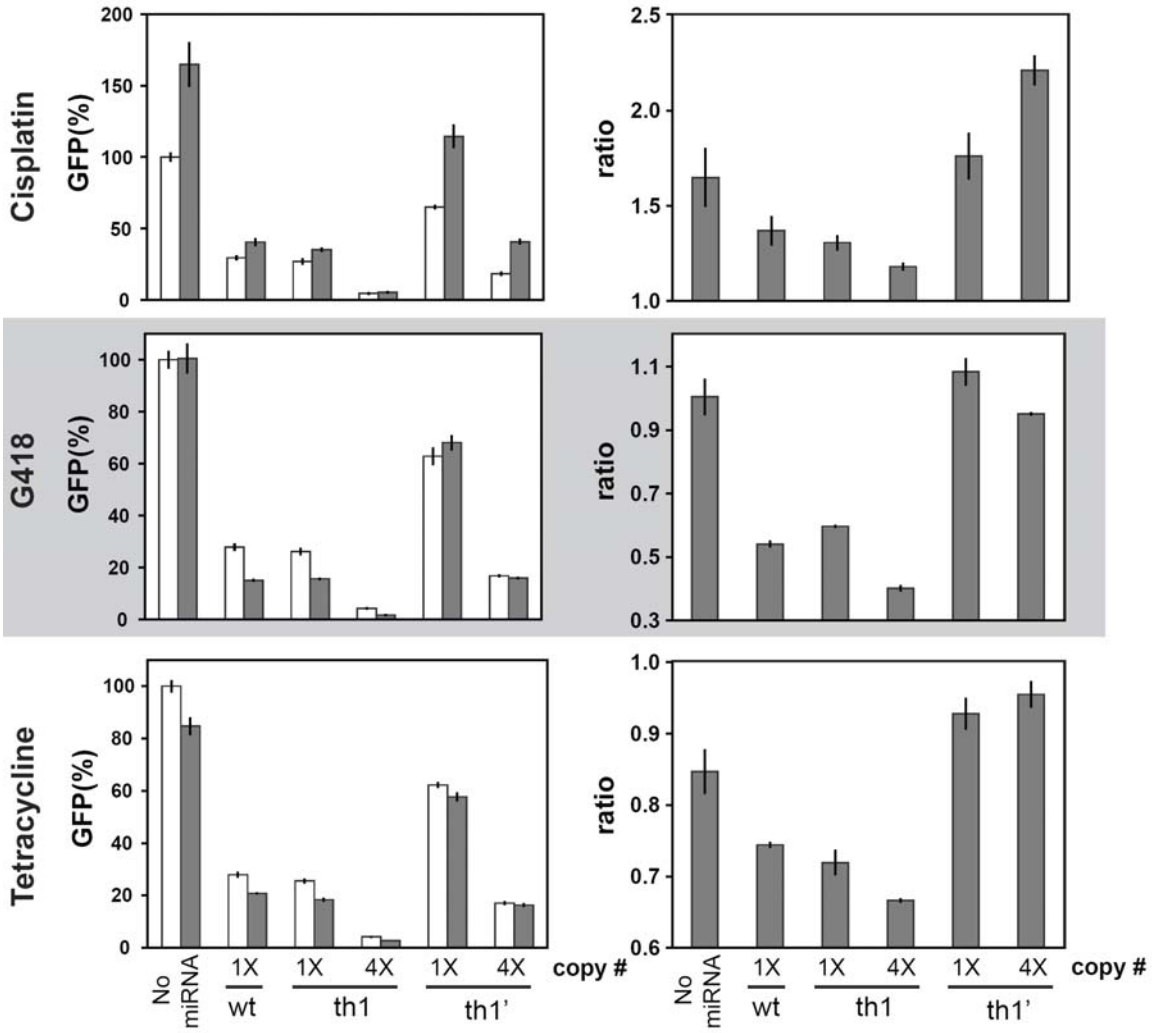


Figure S4.5 Self-targeting miRNAs confer differential response to cellular stress. miRNAs were located in the 3' UTR of the trans-targeted transcript encoding GFP. Constructs harboring no miRNAs (No miRNA), one copy of a self-targeting miRNA with basal segments containing sequences similar to miR-30a (wt), one (1X) or four (4X) copies of a theophylline-responsive self-targeting miRNA (th1), and one or four copies of a theophylline-responsive non-targeting miRNA (th1') were characterized, where multiple copies were separated by the largest spacer

length tested (112 nt). Stable cell lines were grown for over one week in the presence (gray) or absence (black) of 1 μg/ml cisplatin, 900 μg/ml G418, or 500 μM tetracycline prior to flow cytometry analysis. The right column reports the ratio of GFP levels in the presence and absence of theophylline. Error bars represent the standard deviation of cells grown in two separate culture wells.

Table S4.1 Sequences for ligand-responsive and control miRNAs. Each sequence is written 5' to 3' and represents the final construct cloned into XbaI and ApaI within pcDNA3.1(+). th3 was cloned into XhoI and XbaI in pcDNA3.1(+) already containing th1 to test the efficacy of two miRNAs separated by a minimal spacer. Color codes: gray, restriction sites; blue, aptamer; green, designed guide strand sequence. The database # is included for plasmid requests.

Name	Sequence	Aptamer	Database #
wt	TCTAGAGTTTGACAGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCTGCCTACTGCCTCGGACTGAATTCATAGGGCCC	None	pCS351
m1	TCTAGAACGGGAAGTAATTACAGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCTGCCTACTGCCACATAGGGCCC	None	pCS1246
m2	TCTAGAACGGGAAACACAGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCTGCCTACTGCCTCGGGCCC	None	pCS1215
m3	TCTAGAACGGGAAACACAGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCTGCCTACTGCCGGGCCC	None	pCS1241
m4	TCTAGAACGGGAAACACAGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCTGCCTACTGGGGCCC	None	pCS1242

Table S4.1 cont'd

Name	Sequence	Aptamer	Database #
th1	TCTAGAACGGGTCCTGATACCAGCGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCCGCCTACGCCCTTGGCAGCAGGGCCC	Theophylline	pCS1183
th1'	TCTAGAACGGGTCCTGATACCAGCGTGAGCGAGCACAAGCTATCAACATGAGGTAGTGAAGCCACAGATGTACCTCATGTTGATAGCTTGTGCCCGCCTACGCCCTTGGCAGCAGGGCCC		pCS1258
th2	TCTAGAACGGGTCCTGATACCAGCGTGAGCGCCAAGAAGATGGTGCGCTCCTGGAGTGAAGCCACAGATGTCCAGGAGCGCACCATCTTCTTGTCGCCTACGCCCTTGGCAGCAGGGCCC		pCS1664
th3	TCTAGACGCCAGAATGATACCAGCGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCCGCCTACGCCCTTGGCAGCATTCTGGCGCCTAGG		pCS1229
tc1	TCTAGAACGGGTCCTAAAACATACCGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCTGCCTACGGAGAGGTGAAGAATACGACCACCTAGGGCCC	Tetracycline	pCS1217
xa1	TCTAGAACGGGTCCGTGTATTACCTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCCGCCTAGGTCGACGGGCCC	Xanthine	pCS1218
xa2	TCTAGAACGGGTCCGAGGTCGACGTGAGCGAGCACAAGCTGGAGTACAACTATAGTGAAGCCACAGATGTATAGTTGTACTCCAGCTTGTGCCCGCCTACGTGTATTACCCAGGGCCC		pCS1244
La1	TCTAGAGTTTGACAGTGAGCGCTGGAAATCAGTGAAGATAAAATAGTGAAGCCACAGATGTATTTTATCTTCACTGATTTCCATTGCCTACTGCCTCGGACTGAATTCATAGGGCCC	None	pCS1676
La2	TCTAGAACGGGTCCTGATACCAGCGTGAGCGCTGGAAATCAGTGAAGATAAAATAGTGAAGCCACAGATGTATTTTATCTTCACTGATTTCCATCGCCTACGCCCTTGGCAGCAGGGCCC	Theophylline	pCS1677

Chapter 5

Design Principles for Riboswitch Function

ABSTRACT

Scientific and technological advances that enable the tuning of integrated regulatory components to match network and system requirements are critical to reliably control the function of biological systems. RNA provides a promising building block for the construction of tunable regulatory components based on its rich regulatory capacity and our current understanding of the sequence-function relationship. One prominent example of RNA-based regulatory components is riboswitches, genetic elements that mediate ligand control of gene expression through diverse regulatory mechanisms. While characterization of natural and synthetic riboswitches has revealed that riboswitch function can be modulated through sequence alteration, no quantitative frameworks exist to investigate or guide riboswitch tuning. Here, we combined mathematical modeling and experimental approaches to investigate the relationship between riboswitch function and performance. Model results demonstrated that the competition between reversible and irreversible rate constants dictates performance for different regulatory mechanisms. We also found that practical system restrictions, such as an upper limit on ligand concentration, can significantly alter the requirements for riboswitch performance, necessitating alternative tuning strategies. Previous experimental data for natural and synthetic riboswitches as well as experiments conducted in this work support model predictions. From our results, we developed a set of general design principles for synthetic riboswitches. Our results also provide a foundation from which to investigate how natural riboswitches are tuned to meet systems-level regulatory demands.

INTRODUCTION

The breadth of function exhibited by biological systems provides a foundation from which to develop solutions to global challenges including sustainability, renewable energy production, material synthesis, and medical advancement. Underlying these systems-level functions are regulatory components that evaluate molecular information in the extracellular and intracellular environments and ultimately translate that information into phenotypic responses over varying time scales. The properties of individual regulatory components and genetic networks composed of these components are tuned to control critical functions, including survival in fluctuating environments (Acar *et al*, 2008; Bennett *et al*, 2008), minimization of energy expenditure in metabolism (Dekel and Alon, 2005; Zaslaver *et al*, 2004), developmental fate assignment (Suel *et al*, 2007), and proper information transmission through regulatory cascades (Hao *et al*, 2008; Levchenko *et al*, 2000; Yokobayashi *et al*, 2002). To more effectively approach the reliable construction of synthetic biological systems, it is critical to advance our understanding of the degree to which individual component properties are tuned in natural systems, the underlying mechanisms that support tuning of biological components, and the effect of tuned components on resulting systems-level functions.

Riboswitches are RNA-based regulatory components that mediate ligand control of gene expression. Natural riboswitches have been identified in all three kingdoms of life (Barrick and Breaker, 2007) and primarily function by sensing a variety of essential cofactors, amino acids, and nucleotides and regulating the expression levels of proteins in associated metabolic pathways (Winkler, 2005). Riboswitches typically exploit three properties of RNA to translate changes in ligand concentration to changes in the

expression of a target protein: specific and high affinity ligand binding by aptamer sequences, formation of distinct functional conformations primarily dictated by base-pairing interactions, and diverse gene expression regulatory mechanisms based on the central location of mRNA in the process of gene expression. With the exception of the *glmS* ribozyme (Klein and Ferre-D'Amare, 2006; Winkler *et al*, 2004), natural riboswitches function through a general mechanism in which the RNA molecule can primarily adopt two conformations and ligand binding to the formed aptamer in one conformation biases partitioning toward the ligand-bound conformation. Each conformation is associated with differential regulatory activities such that increasing ligand concentrations either increase (ON behavior) or decrease (OFF behavior) gene expression depending on which conformation contains the formed aptamer. Synthetic riboswitches have been constructed based on this functional mechanism to expand on the regulatory potential exhibited by natural riboswitches (Isaacs *et al*, 2006; Suess and Weigand, 2008). There has been significant interest in engineering riboswitches as tailored ligand-responsive genetic control elements by integrating aptamers selected *in vitro* (Osborne and Ellington, 1997) against diverse molecular ligands appropriate for different applications.

Natural and synthetic riboswitches have been demonstrated to be highly tunable regulatory components. Targeted nucleotide changes in synthetic riboswitches can shift the response curve (Bayer and Smolke, 2005; Beisel *et al*, 2008; Isaacs *et al*, 2004; Lynch *et al*, 2007; Win and Smolke, 2007). Studies of natural riboswitches functioning through transcriptional termination found that the time lag between transcription of the aptamer and the terminator stem can tune the effective ligand concentration at which a half-

maximal response is achieved (EC_{50}) (Wickiser *et al*, 2005a; Wickiser *et al*, 2005b). These previous studies explored tuning in limited contexts by focusing on one aspect of riboswitch function (EC_{50}) for one type of regulatory mechanism. However, advancing the characterization or design of new riboswitches requires a general quantitative framework that applies broadly to different regulatory mechanisms. Due to the link between RNA secondary structure and function and the relative ease with which RNA molecules can be modeled, riboswitches present an interesting class of regulatory components through which researchers can examine links between physical composition, tuned component response properties, and resulting systems-level behavior.

In this study, we employed mathematical modeling to explore how the dynamics of riboswitch function dictate its performance, where performance is evaluated based on the response curve quantitatively linking ligand concentration and protein levels. To draw general conclusions regarding riboswitch performance, we considered three representative regulatory mechanisms: transcriptional termination, translational repression, and mRNA destabilization. Parameter space for all three mechanisms was surveyed in order to understand the relationship between model parameters and riboswitch performance. Our results show that the competition between reversible and mechanism-specific irreversible rate constants primarily dictates riboswitch performance and response curve tuning properties. Complete dominance of irreversible rate constants renders a riboswitch non-functional, although ligand binding during transcription can rescue riboswitch performance. Our results also demonstrate that placing an upper limit on the ligand concentration alters the observed tuning properties such that a maximum dynamic range exists for intermediate conformational partitioning. Model predictions are

supported by published experimental data and new data obtained through the modification of a synthetic riboswitch. We provide a set of design principles for the construction of synthetic riboswitches based on our modeling results. In addition, our results lend insights into the inherent flexibility and potential biological relevance of tuning of natural riboswitches.

RESULTS

Kinetic modeling of riboswitch function

We started with a detailed molecular description of riboswitch function (Figure S5.1) that accounts for folding and ligand binding during discrete steps of transcription. Three regulatory mechanisms were considered: translational repression, transcriptional termination, and mRNA destabilization. Translational repression occurs through ribosome binding site (RBS) sequestration in a double-stranded secondary structure that prevents ribosome recruitment. Transcriptional termination occurs through a rho-independent mechanism such that hairpin formation directly upstream of a polyuridine stretch induces dissociation of the transcript from the template and the polymerase. We also considered the regulatory mechanism of a recently-described synthetic riboswitch that undergoes ribozyme self-cleavage (Win and Smolke, 2007), thereby initiating mRNA destabilization (Collins *et al*, 2007). In these examples, two inter-converting conformations (A/B) are associated with differential protein levels subject to the specified regulatory mechanism. Ligand binding to the formed aptamer harbored in B promotes conformational stabilization, thereby increasing the prevalence of B.

We assigned a rate constant to each mechanistic step in the models to yield a detailed relationship between ligand concentration (L) and protein levels (P). In all models, transcriptional initiation produces a partial-length riboswitch in either conformation A (k_{fA}) or B (k_{fB}) to reflect transcriptional folding. Transcription is broken into discrete steps that represent different sequence lengths. Each step determines the extent of conformational switching (k_1, k_1'), the ability to bind and release ligand (k_2, k_2'), and the rate of progression to the next step (k_E). For transcriptional termination, riboswitches effectively choose between termination (k_{TA}, k_{TB}) and extension (k_{MA}, k_{MB}) after transcription of the terminator stem. To ensure that both conformations make the decision with the same frequency, we set the sum of termination and extension rate constants for each conformation equal to a single parameter k_M:

$$k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB} \tag{6.1}$$

Following transcription of the full-length riboswitch for translational repression and mRNA destabilization or extension through the terminator stem for transcriptional termination, the transcript can be translated into protein (k_{PA}, k_{PB}) or undergo degradation (k_{dMA}, k_{dMB}). A single constant is assigned when the rate constants are equal between conformations (k_P, k_{dM}). Values for the rate constants can vary widely depending on the organism and regulatory mechanism (Table 5.1). Therefore, we explored how each rate constant contributes to riboswitch performance.

Table 5.1 Estimated ranges for parameter values based on previous experimental and computational studies. Rates for endogenous mRNA degradation (norm) and ribozyme cleavage (rib) are separately described. [α]Reflects the time to reach the termination stem following transcription initiation and is dependent on the rate of polymerase extension, pausing, the length of the transcribed sequence, and nucleotide concentration. [β]Includes mRNA or protein degradation and dilution due to cell division. [γ]Observed upper limit.

Parameter	Units	Value Range	References
k_f	M/s	$10^{-13} – 10^{-8}$	(Voigt *et al*, 2005)
k_P	1/s	$10^{-4} – 10^{1}$	(Voigt *et al*, 2005)
k_E^{α}	1/s	$10^{-2} – 10^{2}$	(Pan and Sosnick, 2006; Wickiser *et al*, 2005b)
k_1, k_1'	1/s	$10^{-3} – 10^{3}$	(Lee *et al*, 2007; Su *et al*, 2005; Zarrinkar *et al*, 1996; Zhuang *et al*, 2000)
k_2	1/M·s	$10^{3} – 10^{8}$	(Greenleaf *et al*, 2008; Kensch *et al*, 2000; Lang *et al*, 2007; Rieder *et al*, 2007; Wickiser *et al*, 2005a; Wickiser *et al*, 2005b; Win *et al*, 2006)
k_2'	1/s	$10^{-3} – 10^{1}$	(Greenleaf *et al*, 2008; Kensch *et al*, 2000; Wickiser *et al*, 2005a; Wickiser *et al*, 2005b; Win *et al*, 2006)
k_M	1/s	$10^{-2} – 10^{-1}$	(Crothers *et al*, 1974; Wickiser *et al*, 2005b)
k_{dM} (norm)$^{\beta}$	1/s	$10^{-5} – 10^{-2}$	(Bernstein *et al*, 2002; Leclerc *et al*, 2002; Narsai *et al*, 2007; Selinger *et al*, 2003)
k_{dM} (rib)$^{\gamma}$	1/s	10^{-1}	(Emilsson *et al*, 2003)
k_{dP}^{β}	1/s	$10^{-5} – 10^{-2}$	(Belle *et al*, 2006; Corish and Tyler-Smith, 1999)

Riboswitch performance was evaluated under steady-state conditions for both ON and OFF behaviors by calculating a collection of performance descriptors that define the response curve (Figure 5.1A): EC_{50}, dynamic range (η) defined as the difference between

high and low protein levels, basal protein levels (P(L = 0)), and ligand-saturating protein levels (P(L → ∞)). While the dynamic range can be alternatively defined as the ratio of high and low protein levels, we selected the difference definition based on the mathematical symmetry between the equations representing ON and OFF behaviors (Text S1).

Transcription can be considered as a discrete multistep process (Figure S5.1). The conformations that can form at each step depend on the ordering of elements along the riboswitch sequence, such as the relative location of the aptamer or gene regulatory elements. Matching the number of steps and parameter values to particular sequence configurations becomes burdensome and restricts the elucidation of general principles. Therefore, we simplified the transcription process by assuming that synthesized transcripts appear in either conformation A or B and are immediately subject to conformational partitioning, ligand binding, and the regulatory mechanism (Figure 5.1B–D). As a result, the outcome of the transcription process is reflected by biased folding into either conformation A (k_{fA}) or conformation B (k_{fB}). This simplification excludes ligand binding during transcription, which has been demonstrated for natural riboswitches functioning through transcriptional termination (Wickiser *et al*, 2005a; Wickiser *et al*, 2005b). Therefore, we separately accounted for ligand binding during transcription in our analyses.

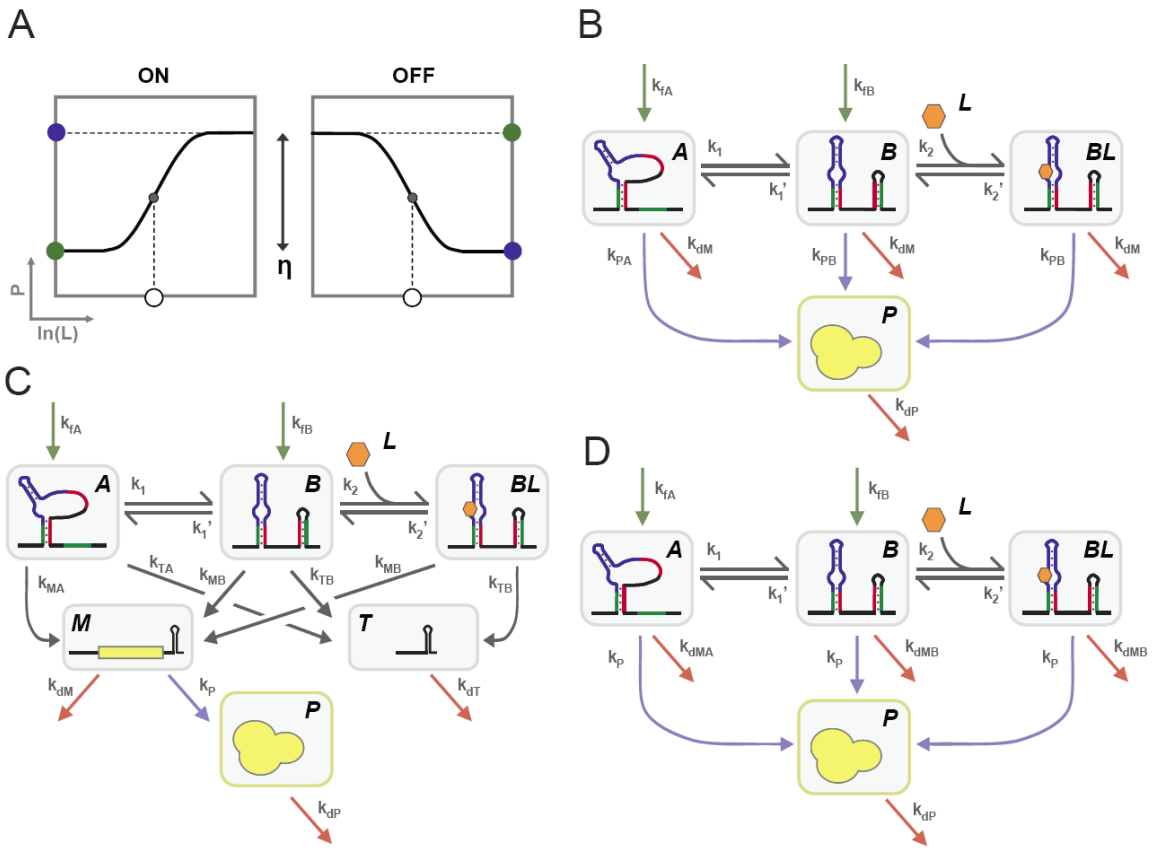


Figure 5.1 Kinetic models for riboswitches functioning through three distinct regulatory mechanisms. (**A**) General response curve relating ligand concentration (L) and regulated protein levels (P) for both ON and OFF behavior. Descriptors important in evaluating riboswitch performance are indicated: dynamic range (η), EC_{50} (○), basal protein levels (●), and ligand-saturating protein levels (●). Gene regulatory mechanisms include (**B**) translational repression, (**C**) transcriptional termination, and (**D**) mRNA destabilization. All riboswitches can reversibly switch between conformations A and B that display different regulatory activities or different rates of degradation. Conformation B contains a formed aptamer that can reversibly bind ligand. Models assume negligible ligand binding during transcription. Green arrows designate mRNA synthesis with biased transcriptional folding, red arrows designate species degradation, and blue arrows designate translation that is proportional to mRNA levels. Under transcriptional termination (**C**), riboswitches effectively choose between termination to form a truncated product (T) and extension to form the full-length mRNA (M). To ensure both conformations make the

decision with the same frequency, we designated a rate constant k_M equal to the sum of the rate constants for extension and termination for either conformation ($k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB}$).

Competition between reversible and irreversible rate constants suggests three operating regimes

We first derived expressions for the performance descriptors – dynamic range, EC_{50}, and basal and ligand-saturating levels – for riboswitches functioning through each regulatory mechanism (Text S5.2). From these derivations two common parameters emerged: the partitioning constant ($K_1 = k_1'/k_1$) and the ligand association constant ($K_2 = k_2/k_2'$). K_1 reflects the relative stability of conformation A and is present in all performance descriptor expressions. K_2 reflects the affinity between the aptamer and its cognate ligand and is only present in the expression for EC_{50}.

For all regulatory mechanisms, K_1 and K_2 reflect reversible conformational switching and ligand association, the core processes of riboswitch function. These processes are opposed by irreversible events that deplete the abundance of both conformations: mRNA degradation for translational repression and mRNA destabilization, and the riboswitch's decision to terminate or extend for transcriptional termination. The ratio between the rate constants for irreversible and reversible events is prevalent in all expressions for the performance descriptors (Text S5.2). This ratio is encapsulated in two reduced parameters γ_1 and γ_2:

$$\gamma_1 = \left(1 + \frac{k_{iA}}{k_1}\right)^{-1}, \tag{6.2}$$

$$\gamma_2 = \left(1 + \frac{k_{iB}}{k_2'}\right)^{-1}. \tag{6.3}$$

Here, k_{iA} and k_{iB} represent the irreversible rate constants for conformation A or B, respectively, for translational repression (k_{dM}), transcriptional termination (k_M), and mRNA degradation (k_{dMA} and k_{dMB}).

Three operating regimes can be defined based on the ratio of reversible and irreversible rate constants within γ_1 and γ_2. The first regime occurs when both reversible rate constants dominate (γ_1, γ_2 converge to one), the second begins when either of the reversible rate constants is balanced with the associated irreversible rate constant (either γ_1 or γ_2 is less than one), and the third begins when the irreversible rate constant k_{iA} dominates over k_1 (γ_1 converge to zero). Each regime is generally determined by the competition between reversible and irreversible rate constants. We next evaluated the tuning properties of each regime for all regulatory mechanisms.

Riboswitches display similar tuning trends in the thermodynamically-driven regime

For dominating reversible rate constants ($\gamma_1 = \gamma_2 = 1$), a riboswitch molecule can sample both conformations and bind and unbind ligand many times before the irreversible event occurs. We define this regime as 'thermodynamically-driven' in accord with previous uses of this term in the study of natural riboswitches (Breaker, 2008; Rieder *et al*, 2007; Wickiser *et al*, 2005a), since energetics dictate the prevalence of each conformation.

In the thermodynamically-driven regime, riboswitch function is captured for the three regulatory mechanisms by a general molecular description (Figure 5.2A). The associated response curve is captured by a single equation that includes the partitioning constant (K_1), the aptamer association constant (K_2), mRNA degradation rate constants

(k_{dMA}, k_{dMB}), and representative regulatory activities of conformations A (K_A) and B (K_B):

$$P = \frac{k_f}{k_{dP}} \cdot \frac{K_A K_1 + K_B(1 + K_2 L)}{k_{dMA} K_1 + k_{dMB}(1 + K_2 L)} \tag{6.4}$$

The values of K_A and K_B depend on the selected regulatory mechanism and are provided in Supplementary Text S5.2.

Parameter variation has a unique effect on the response curve for both ON (Figure 2B,C) and OFF behaviors (Figure S5.2). Increasing K_1 stabilizes conformation A, resulting in more riboswitch molecules adopting this conformation. Since conformation A has lower regulatory activity for ON behavior ($K_A < K_B$), basal levels decrease. Concomitantly, EC_{50} increases as higher ligand concentrations are required to offset the decreased abundance of conformation B. Increasing K_2 reduces EC_{50} as expected when aptamer affinity is modulated. However, K_2 has no effect on dynamic range and ligand-saturating levels since we assumed sufficient ligand can be added to saturate the response curve. Previous mutational studies of two synthetic riboswitches (Bayer and Smolke, 2005; Beisel *et al*, 2008) support these model predictions. However, these studies examined trans-acting mechanisms, calling into question whether model insights apply to cis-acting mechanisms. Finally, rate constants distinct from the core processes of riboswitch function such as transcription initiation (k_f) and protein decay (k_{dP}) affect both basal levels and dynamic range by modulating the steady-state mRNA and protein levels.

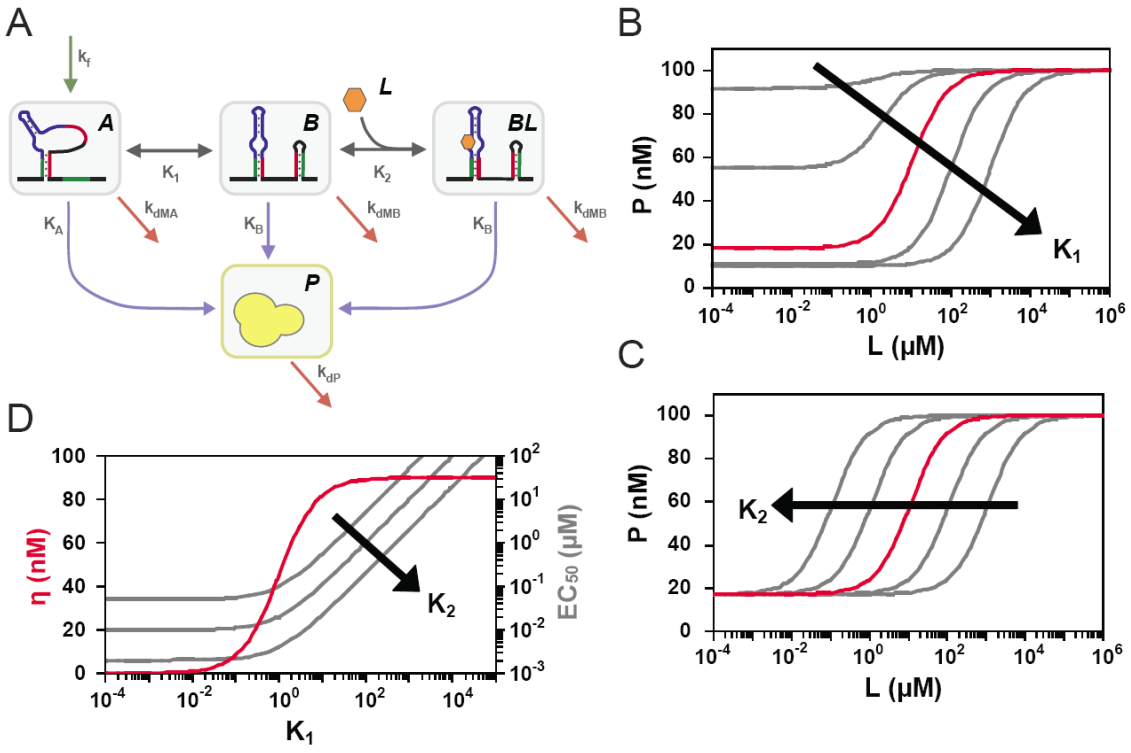


Figure 5.2 Thermodynamically-driven riboswitches display similar tuning properties. (**A**) Simplified molecular description that captures the three riboswitch regulatory mechanisms in the thermodynamically-driven regime. K_1 is the conformational partitioning constant (k_1'/k_1) and K_2 is the aptamer association constant (k_2/k_2'). Coloring is the same as in Figure 1B-D. K_A and K_B reflect the regulatory activity of conformations A and B, respectively, and are specific to each regulatory mechanism: translational repression (k_{PA}, k_{PB}), transcriptional termination ($k_P \cdot k_{MA}/k_M$, $k_P \cdot k_{MB}/k_M$), and mRNA destabilization (k_P, k_P). (**B**) K_1 affects both basal levels and EC_{50}. (**C**) K_2 only affects EC_{50}. Parameter values for red response curves: $K_1 = 10$; $K_2 = 1/\mu M$; $K_A = 10^{-3}/s$; $K_B = 10^{-2}/s$; $k_f = 10^{-11}$ M/s; $k_{dP} = 10^{-3}/s$; $k_{dMA} = k_{dMB} = 10^{-3}/s$. (**D**) Biased conformational partitioning toward conformation B maximizes the dynamic range (η) at the cost of increased EC_{50}.

Stabilizing conformation A (increasing K_1) improves the dynamic range to an upper limit set by the regulatory activities (K_A, K_B) and mRNA degradation rate constants (k_{dMA}, k_{dMB}) associated with each conformation (Figure 5.2D). While all four parameters affect the dynamic range, k_{dMA} and k_{dMB} also impact the dependence of the dynamic range on conformational partitioning (Figure S5.3). This latter effect results from the dominant influence of the larger mRNA degradation rate on steady-state mRNA levels, which can be countered by biasing partitioning toward the more stable conformation. Therefore, when conformation A degrades faster (higher k_{dMA}, ON behavior), less partitioning toward conformation A (lower K_1) is required to separate basal and ligand-saturating levels, whereas more partitioning toward conformation A (higher K_1) is required when conformation B degrades faster (higher k_{dMB}, OFF behavior). As a result, thermodynamically-driven riboswitches functioning through mRNA destabilization require more (OFF behavior) or less (ON behavior) partitioning toward conformation A to achieve a larger dynamic range. In contrast, riboswitches functioning through translational repression and transcriptional termination display similar trends in dynamic range as a function of conformational partitioning for ON and OFF behaviors as the degradation rate constants are the same for each conformation.

EC_{50} is also dependent on the value of K_1 according to the following relationship:

$$EC_{50} = \frac{1 + K_1 \frac{k_{dMA}}{k_{dMB}}}{K_2} \quad . \tag{6.5}$$

EC_{50} approaches the lower limit set by the aptamer dissociation constant when riboswitches principally adopt conformation B (low K_1). Therefore, although stabilizing conformation A (increasing K_1) can improve the dynamic range, excessive stabilization

can be detrimental due to the increase in EC_{50}. As a result, tuning strategies based on increasing K_1 require higher ligand concentrations to access the improved dynamic range. The ratio of the mRNA degradation rate constants in the expression for EC_{50} offsets the modified dependence of the dynamic range on K_1 for riboswitches functioning through mRNA destabilization. Therefore, riboswitches functioning through any of the regulatory mechanisms exhibit the same trade-off between EC_{50} and dynamic range.

Riboswitches display expanded tunability with reduced performance in the kinetically-driven regime

The second regime begins when either of the irreversible rate constants balances the associated reversible rate constant (either γ_1 or γ_2 is between zero and one). We call this regime the 'kinetically-driven' regime in accord with uses of this term in the study of natural riboswitches (Wickiser *et al*, 2005a; Wickiser *et al*, 2005b), where performance is driven by kinetics over energetics. In this regime, riboswitch molecules have fewer opportunities to sample both conformations and bind and release ligand before the irreversible event occurs, where the number of opportunities is governed by the competition between reversible and irreversible rate constants. Since γ_1 is coupled to K_1 and k_{fA} while γ_2 is coupled to K_2, both γ_1 and γ_2 are anticipated to have a significant impact on the response curve and impart several tuning properties distinct to this regime. We initially use riboswitches functioning through transcriptional termination to highlight two of these tuning properties (Figure 5.3A–C).

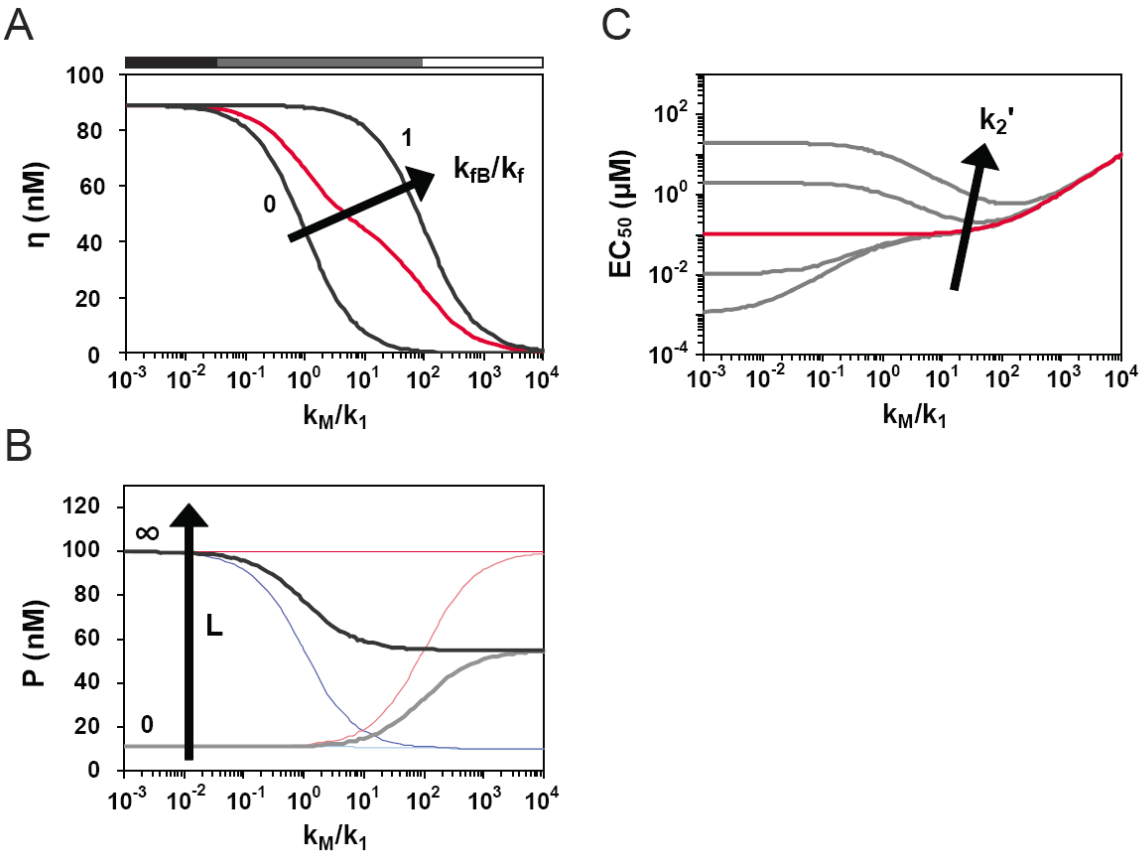


Figure 5.3 Rate competition dictates riboswitch performance. The relative values of the reversible and irreversible rate constants generally establish three operating regimes: thermodynamically-driven (■) when reversible rate constants dominate, kinetically-driven (■) when the rate constants are balanced, and non-functional (□) when irreversible rate constants dominate. Regimes are qualitatively marked for dynamic range and basal and ligand-saturating levels according the ratio of the rate constants for terminator stem formation (k_M) and the progression from conformation A to conformation B (k_1). Effect of varying k_M on (**A**) dynamic range, (**B**) basal protein levels and ligand-saturating protein levels, and (**C**) EC_{50} for riboswitches functioning through transcriptional termination. In (B), colored pairs show basal (light) and ligand-saturating (dark) protein levels for complete (red line), balanced (black line), and negligible (blue line) transcriptional folding into conformation B. Parameter values for all curves

in (A) and (B) and red curve in (C): $k_1 = 10^{-1}$/s;k_1' = 10/s; $k_2 = 10^6$/M·s; k_2' = 10^{-1}/s; $K_A = k_P \cdot k_{MA}/k_M = 10^{-3}$/s; $K_B = k_P \cdot k_{MB}/k_M = 10^{-2}$/s; $k_f = 10^{-11}$ M/s; $k_{dP} = 10^{-3}$/s.

First, irreversible rate constants modulate all performance descriptors, often at a cost to riboswitch performance. As the rate constant for terminator stem formation (k_M) increases, riboswitch molecules become trapped in a given conformation after transcriptional folding or conformational switching as reflected in γ_1. This effect reduces the dynamic range (Figure 5.3A) and shifts basal and ligand-saturating levels according to the extent of transcriptional folding (Figure 5.3B). The reduction in dynamic range can be offset by increasing the overall mRNA and protein abundance through modulation of the rates of transcription (k_f), translation (k_P), and mRNA (k_{dM}) and protein (k_{dP}) degradation. However, such changes also increase basal levels.

γ_1 and γ_2 both impact EC_{50} according to the following relationship:

$$EC_{50} = \frac{1 + K_1 \gamma_1 \frac{k_{dMA}}{k_{dMB}}}{K_2 \gamma_2} . \tag{6.6}$$

Since γ_1 and γ_2 reflect the ratios of k_M/k_1 and k_M/k_2', respectively, the relationship between EC_{50} and k_M depends on both conformational switching (k_1) and ligand release (k_2') (Figure 5.3C). Increasing k_M, which reduces both γ_1 and γ_2, can increase or decrease EC_{50} based on the opposing contributions of γ_1 and γ_2. Reducing γ_1 decreases EC_{50} by restricting the time available to switch between conformations, while reducing γ_2 increases EC_{50} by decreasing the half-life of the ligand-aptamer complex (BL). Therefore, the relative values of k_1 and k_2' must be known to predict the effect of modulating the irreversible rate constant on EC_{50}.

Second, biased transcriptional folding can modulate the relationship between irreversible rate constants and the dynamic range (Figure 5.3A). When transcriptional folding is biased toward conformation A ($k_{fB}/k_f = 0$), riboswitch molecules must have sufficient time to switch between conformations to maintain activity. Therefore, the dynamic range declines as k_M approaches and surpasses k_1. In contrast, when transcriptional folding is biased toward conformation B ($k_{fB}/k_f = 1$), riboswitch molecules must switch to conformation A before the irreversible event occurs. In this case, k_M must exceed the sum $2k_1 + k_1$' to reduce the dynamic range. As a result, biasing transcriptional folding toward conformation B in the kinetically-driven regime increases the dynamic range.

A third tuning property is associated with riboswitches functioning through mRNA degradation. The rate constants for mRNA degradation (k_{dMA}, k_{dMB}) are responsible for both the irreversible event and the steady-state basal and ligand-saturating levels, resulting in complex tuning properties (Figure S5.4). Increasing either k_{dMA} (ON behavior) or k_{dMB} (OFF behavior) initially improves the dynamic range by separating the steady-state basal and ligand-saturating levels. However, the impact of larger values of k_{dMA} and k_{dMB} depends on riboswitch behavior. For ON behavior, if a riboswitch predominantly folds into conformation A during transcription ($k_{fB}/k_f = 0$), then values of k_{dMA} in excess of k_1 diminish the dynamic range as conformation A is degraded before it can switch conformations. However, if a riboswitch predominantly folds into conformation B during transcription ($k_{fB}/k_f = 1$), then the dynamic range plateaus as each molecule either binds ligand or irreversibly switches to conformation A. In contrast, transcriptional folding has a negligible impact on the relationship between the dynamic

range and k_{dMB} for OFF behavior, since molecules that adopt conformation A will switch to conformation B before undergoing degradation. Furthermore, as observed for the thermodynamically-driven regime, more partitioning toward conformation A (higher K_1) is required to counteract the influence of k_{dMB} on mRNA steady-state levels. Increasing k_{dMB} eventually dominates basal levels when partitioning is maintained, leading to a loss in the dynamic range (Figure S5.3).

As such, a tailored design approach is required to account for the difference between ON and OFF behavior for kinetically-driven riboswitches functioning through mRNA destabilization. Transcriptional folding is a key tuning parameter for ON behavior and should be the predominant focus before tuning the degradation rate of conformation A. In contrast, transcriptional folding can be largely ignored for OFF behavior, and the degradation rate of conformation B must be properly tuned to optimize the dynamic range.

Rescuing riboswitch performance in the non-functional regime

Higher irreversible rate constants require increased ligand concentrations to achieve a diminishing change in protein levels. We define the ‘non-functional’ regime as one in which riboswitches are effectively trapped in the conformation formed during transcriptional folding ($\gamma_1 = 0$). In this regime, ligand has a negligible effect on performance. The fast time scales of terminator stem formation and mRNA cleavage may drive riboswitches functioning through these regulatory mechanisms into this regime.

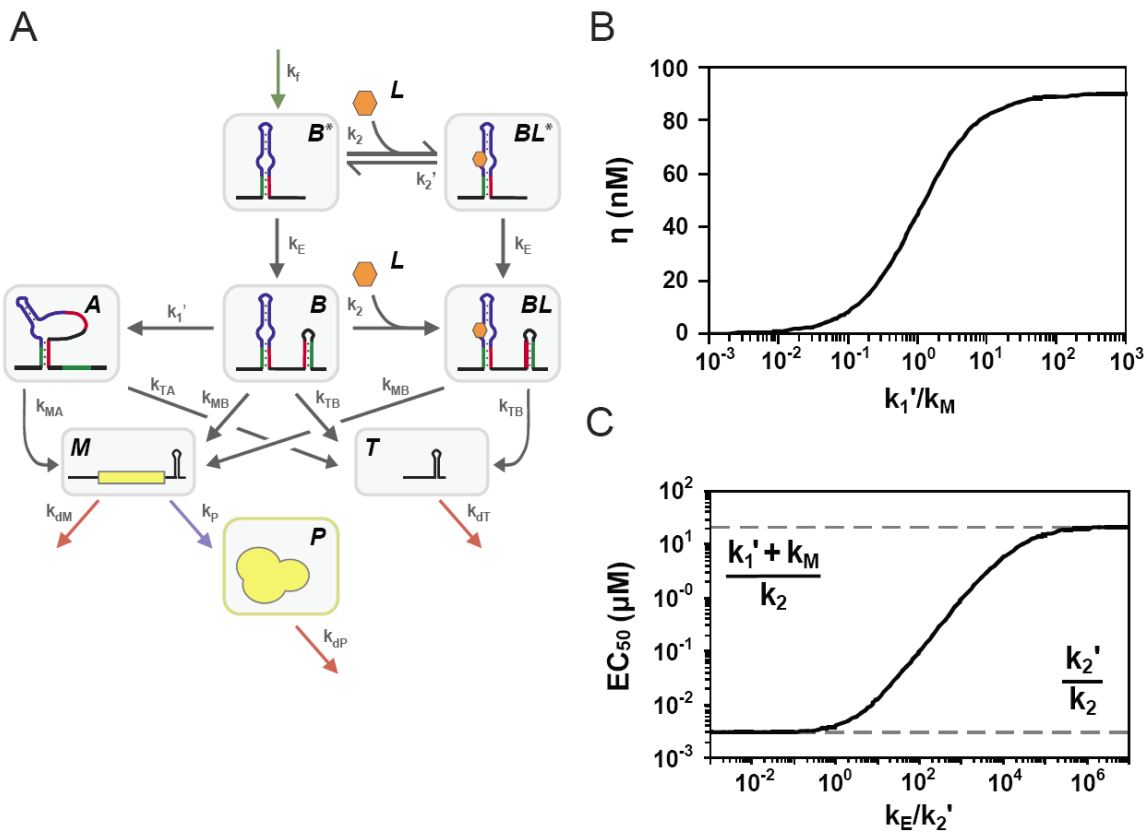


Figure 5.4 Rescuing riboswitch performance in the non-functional regime. (**A**) Molecular description of a non-functional riboswitch functioning through transcriptional termination. The description is identical to that in Figure 5.1C with notable exceptions: the aptamer in conformation B is transcribed first (B*) and can reversibly bind and release ligand before the terminator stem is transcribed (k_E), and terminator stem formation ($k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB}$) occurs much faster than ligand release (k_2') or the progression from conformation A to B (k_1). (**B**) The competition between terminator stem formation (k_M) and the progression from conformation B to A (k_1') determines the dynamic range. (**C**) EC_{50} can be tuned independent from the dynamic range. The accessible range of EC_{50} values is bounded by the aptamer association constant ($K_2 = k_2/k_2$'), the rate constant for the progression from conformation B to A (k_1'), and the rate constant for terminator stem formation (k_M). EC_{50} is tuned over this range by the rate constant representing the delay between aptamer formation and transcription of the terminator stem (k_E). Parameter

values: $k_2 = 10^6/M{\cdot}s$; $k_2' = 3{\cdot}10^{-3}/s$; $K_A = k_P{\cdot}k_{MA}/k_M = 10^{-3}/s$; $K_B = k_P{\cdot}k_{MB}/k_M = 10^{-2}/s$; $k_f = 10^{-11}$ M/s; $k_{dP} = 10^{-3}/s$; $k_{dM} = 10^{-3}/s$; $k_1' + k_M = 20/s$.

Our analysis of the kinetically-driven regime revealed that performance can be preserved by biasing transcriptional folding toward conformation B and ensuring that k_1' exceeds the irreversible rate constant k_{iA}. However, these approaches do not alleviate the increased EC_{50} caused by the reduced half-life of the ligand-aptamer complex (BL) when γ_2 approaches 0. As a potential solution, studies of natural riboswitches have suggested that ligand binding during transcription can preserve EC_{50} (Wickiser *et al*, 2005a; Wickiser *et al*, 2005b). Therefore, we examined the effect of ligand binding during transcription under the assumption that conformation B is solely available ($k_{fB}/k_f = 1$) prior to polymerase extension (k_E) through the gene regulatory element responsible for the irreversible event.

We examined ligand binding during transcription for riboswitches functioning through transcriptional termination (Figure 5.4A). We assumed that terminator stem formation (k_M) occurs much faster than ligand release (k_2') and the progression from conformation A to B (k_1) to limit consideration to non-functional riboswitches. Under these assumptions, the dynamic range is dependent on the ratio of read-through efficiencies for conformations A (k_{MA}/k_M) and B (k_{MB}/k_M), the progression from conformation B to A (k_1'), and the rate of terminator stem formation (k_M). The dynamic range is maximized when conformational progression occurs much faster than terminator stem formation (Figure 5.4B) as predicted from our analysis of the kinetically-driven regime (Figure 5.3A). An *in vitro* study of the *ribD* FMN riboswitch operating through transcriptional termination yielded a reduced dynamic range when removing the

polymerase pause site in the terminator sequence, increasing the nucleotide concentration, and withholding the NusA protein responsible for increasing polymerase residence time at pause sites (Wickiser *et al*, 2005b). These manipulations are expected to reflect an increase in k_M and thus support our model predictions. If increasing k_1' above k_M maximizes the dynamic range, riboswitches operating in this regime are expected to display strong stabilization of conformation A reflecting rapid progression from conformation B. In support of this claim, full-length riboswitches operating under transcriptional termination strongly prefer the aptamer-disrupted conformation and exhibit negligible ligand binding affinity (Lemay *et al*, 2006; Rieder *et al*, 2007; Wickiser *et al*, 2005b).

EC_{50} tuning properties are strikingly different for riboswitches in which ligand binding during transcription allows for improved performance than those for thermodynamically-driven riboswitches. EC_{50} depends on model parameters in Figure 5.4A according to the following relationship:

$$EC_{50} = \frac{1}{2k_2}\left[\sqrt{4(k_1' + k_M)(k_2' + k_E) + (k_1' + k_M + k_2' + k_E)^2} - (k_1' + k_M + k_2' + k_E)\right]. \quad (6.7)$$

Both ligand release (k_2') and the time necessary to transcribe the sequences required for the formation of conformation A (k_E) have a significant impact on the value of EC_{50} (Figure 5.4C). Interestingly, tuning of k_E decouples EC_{50} and basal levels such that EC_{50} can equal the aptamer dissociation constant (k_2/k_2') without impacting the dynamic range. In contrast, the EC_{50} of a thermodynamically-driven riboswitch approaches the aptamer dissociation constant as conformation B is stabilized, resulting in a concomitant decrease in the dynamic range (Figure 5.2D). A previous theoretical study of the *pbuE* adenine riboswitch using experimentally measured kinetic rates also concluded that modulating

polymerase extension time can tune EC_{50} when the extension time is not significantly slower than ligand release (Wickiser *et al*, 2005a).

Restricting the ligand concentration upper limit alters observed tuning properties

In our analyses thus far, we assumed that the maximum ligand concentration always saturates the response curve. However, studies of synthetic riboswitches have demonstrated that the response curve may not be saturated by the accessible upper limit in ligand concentration (Figure 5.5A) due to various system properties including aptamer affinity, ligand solubility, permeability of the ligand across the cell membrane, and cytotoxicity of the ligand (An *et al*, 2006; Bayer and Smolke, 2005; Beisel *et al*, 2008; Desai and Gallivan, 2004; Suess *et al*, 2003; Win and Smolke, 2007). Furthermore, natural riboswitches may regularly function in response to physiologically-relevant changes in metabolite concentrations that are much smaller than the ~1000-fold range necessary to access the full riboswitch response curve. To assess the effect of establishing an upper limit to the ligand concentration, we evaluated the response curve descriptors for a maximum ligand concentration of L'. An apparent EC_{50} (EC_{50}^{APP}) was calculated according to protein levels at L = 0 and L'.

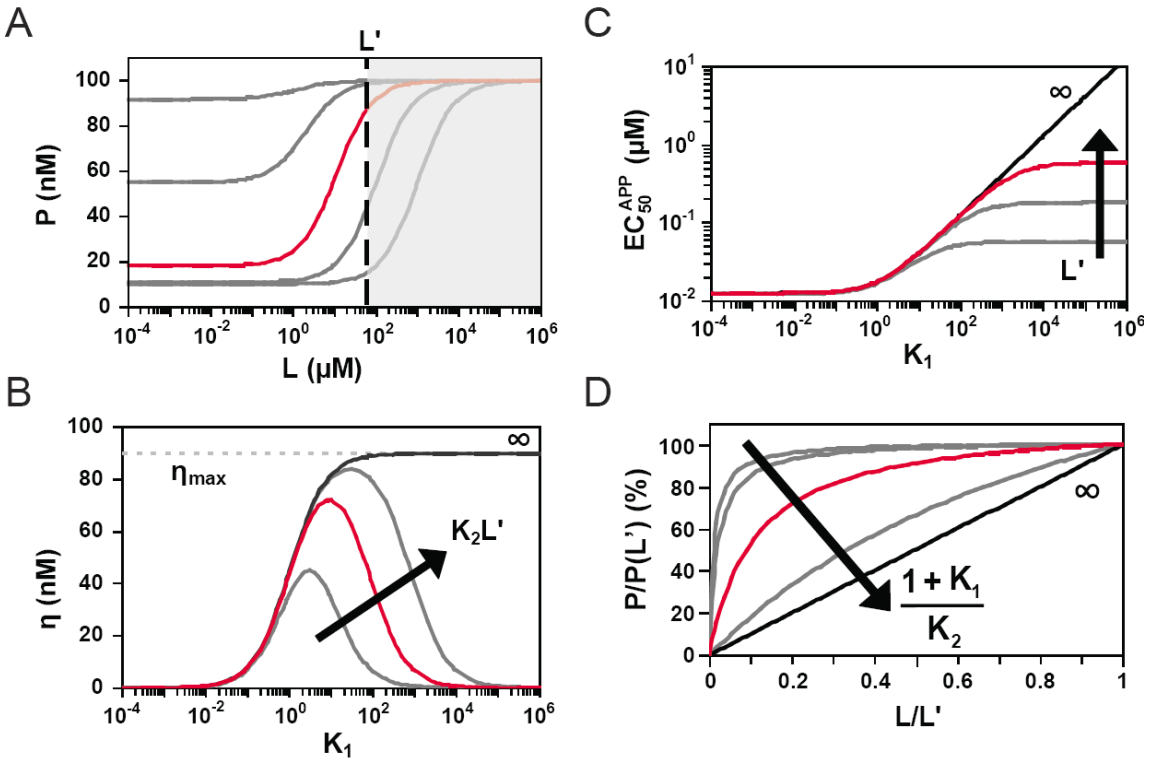


Figure 5.5 Placing an upper limit on the ligand concentration range alters the observed tuning properties. (**A**) Placing an upper limit on the ligand concentration (L') restricts access to the full response curve. This limit affects the dependence of (**B**) the dynamic range (η) and (**C**) the apparent EC_{50} (EC_{50}^{APP}) on the conformational partitioning constant (K_1) and the aptamer association constant (K_2). The maximum dynamic range (η_{max}) is proportional to the difference between regulatory activities for conformations A (K_A) and B (K_B) normalized to the respective degradation rate constants k_{dMA} and k_{dMB}. (**D**) Normalized response curves for fixed L' and increasing values of $(1 + K_1)/K_2$, which equals EC_{50} under ligand-saturating conditions. Parameter values are identical to those reported in Figure 5.2 with L' = 60 μM.

Restricting L' alters the dependence of the dynamic range (Figure 5.5B) and the apparent EC_{50} (Figure 5.5C) on model parameters as illustrated for riboswitches operating in the thermodynamically-driven regime. L' acts as a system restriction that

prevents access to the full response curve such that increasing K_1 shifts the actual EC_{50} beyond L', thereby reducing the maximum dynamic range that can be achieved. This behavior was recently observed for a trans-acting synthetic riboswitch operating under a limited ligand concentration range (Acar *et al*), supporting model predictions. Reflecting this behavior, the apparent EC_{50} has the following dependence:

$$EC_{50}^{APP} = \frac{1 + K_1 \frac{k_{dMA}}{k_{dMB}}}{K_2} \left(2 \frac{1 + K_1 \frac{k_{dMA}}{k_{dMB}}}{K_2 L'} + 1 \right)^{-1}, \qquad (6.8)$$

where the apparent EC_{50} converges to L'/2 as expected for a linear response when L' is below the EC_{50} for an unbounded ligand concentration range (Figure 5.5D). Our modeling results demonstrate that restricting the ligand concentration upper limit reduces riboswitch performance and establishes a unique relationship between dynamic range and conformational partitioning. In addition to serving as a design constraint for synthetic riboswitches, natural riboswitches may inherently operate under defined limits in ligand concentration. Future experiments may focus on measuring the physiologically-relevant metabolite concentration range experienced by natural riboswitches to examine what section of the response curve is utilized.

Application of tuning strategies to a synthetic riboswitch supports model predictions

To begin evaluating how the predicted tuning trends apply to both natural and synthetic riboswitches, we physically manipulated a recently-described synthetic riboswitch functioning through translational repression that up-regulates gene expression (ON behavior) in the presence of theophylline (Lynch *et al*, 2007) (Figure 5.6A). Under

the naming convention from Figure 1A, conformation A comprises a base-paired structure between the aptamer and RBS, while conformation B includes a formed aptamer and a single-stranded RBS. This riboswitch was selected because it closely resembles natural riboswitches functioning through translational repression, experimental data suggest that this riboswitch operates in the thermodynamically-driven regime (Lynch *et al*, 2007), the ligand concentration upper limit does not saturate the response curve (Desai and Gallivan, 2004), and the demonstration that different sequences yield different response curves suggests riboswitch tuning (Lynch *et al*, 2007). A theophylline concentration of 1 mM was used as an upper limit, as exceeding this concentration inhibits cell growth. In studies performed by Lynch and coworkers, sequences associated with desirable response curves were identified by randomization of the RBS and screening for variants with low basal activity and a large activity increase in the presence of theophylline. Since the randomized sequence was located in a region responsible for conformational partitioning and translation, mutations most likely reflect simultaneous modulation of K_A, K_B, and K_1. We therefore sought to introduce directed mutations to solely modulate individual model parameters and test model predictions for a thermodynamically-driven riboswitch with a ligand concentration upper limit that prevents response curve saturation.

We examined two model predictions that could not be supported with available experimental data for cis-acting riboswitches: (1) solely modulating conformational partitioning (K_1) affects both EC_{50} and basal levels (Figure 5.2B), and (2) the dynamic range can be optimized by modulating K_1 when the ligand concentration upper limit does not saturate the response curve (Figure 5.5B). We modulated K_1 by introducing

systematic mutations into the aptamer stem while preserving the RBS sequence (m1-4; Figure 5.6A). Mutant sequences were ordered with increasing K_1 based on the energetic difference between conformations predicted by the RNA folding algorithm mfold (Beisel *et al*, 2008). The mutations were not anticipated to significantly affect aptamer affinity (K_2) (Jenison *et al*, 1994; Zimmermann *et al*, 2000) or translational efficiency for either conformation (K_A, K_B). Additional mutants were examined that are predicted to entirely favor either conformation A (mA) or conformation B (mB) to establish the regulatory activity of either conformation. Riboswitch performance was evaluated by measuring β-Galactosidase levels over a range of theophylline concentrations.

The introduced mutations altered the response curve in agreement with model predictions (Figure 5.6B–D). Protein levels in the presence and absence of theophylline correlated with the relative stability of conformation A. Furthermore, complete stabilization of conformation A (mA) and conformation B (mB) established respective lower and upper limits for the observed expression levels. As predicted for a non-saturating value of L', an intermediate conformational partitioning value optimized the dynamic range to a value that was below the maximum dynamic range (η_{max} = 15,600 MU) (Figure 5.6B), and EC_{50} approached 0.5 mM (L'/2) for increased stabilization of conformation A (Figure 5.6C,D). Dynamic range optimization is clearly observed when evaluating the ratio of high and low protein levels, which is predicted to display the same qualitative tuning behavior (Figure S5.5). The data agree with our model predictions for K_1 modulation in the thermodynamically-driven regime under conditions where the ligand concentration upper limit does not saturate the response curve, although we cannot rule out the possibility that stabilization of conformation A inadvertently drove the

riboswitch into the kinetically-driven regime. The introduction of the aptamer sequence to the regulatory element decreased the regulatory activity of conformation B as observed when comparing protein levels for mB and a construct harboring only the RBS and aptamer basal stem (empty; Figure 5.6B). Our previous construction and characterization of a trans-acting synthetic riboswitch functioning through RNA interference (Beisel *et al*, 2008) also showed sub-maximum dynamic range optimization when the ligand concentration was limiting and compromised activity of the regulatory element due to introduction of the aptamer element of the riboswitch. Thus, the results support the extension of our model predictions to synthetic riboswitches. In addition, our modeling results may have direct implications for the performance and tuning of natural riboswitches based on the similarity between the synthetic riboswitch examined here and natural riboswitches operating under translational repression.

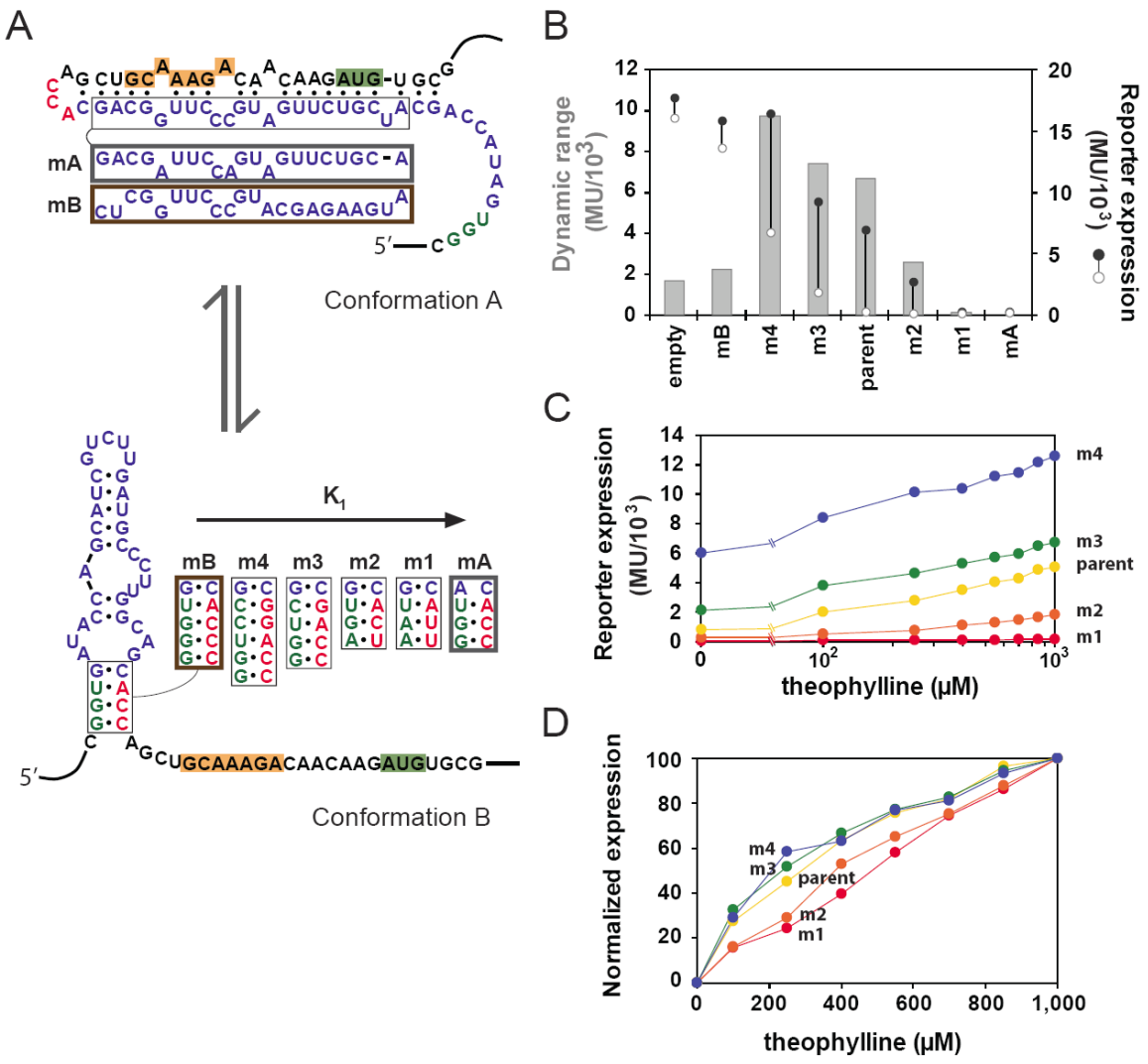


Figure 5.6 Mutational analysis of a synthetic riboswitch supports model predictions. (**A**) Mutations made to the aptamer stem of the parent synthetic riboswitch (m1-4) are anticipated to solely modulate conformational partitioning (K_1). The theophylline-responsive riboswitch controls Tn10-β-Galactosidase levels through RBS sequestration, thereby repressing translation. Mutations were also introduced to lock the riboswitch in either conformation A (mA, gray box) or conformation B (mB, brown box). The RBS and start codon are highlighted in orange and green, respectively. (**B**) β-Galactosidase assay results are reported in Miller Units (MU) for each riboswitch variant in the presence (○) or absence (●) of 1 mM theophylline. Dynamic range (η) is

calculated as the difference between high and low expression levels, where all values were below the theoretical maximum of 15,600 MU as determined by the difference between mB with theophylline and mA without theophylline. The positive control construct (empty) harbors only the RBS and aptamer basal stem. A slight increase in β-Galactosidase activity was observed in the presence of theophylline for the control construct. (**C,D**) Theophylline response curves for riboswitch variants: parent (yellow), m1 (red), m2 (orange), m3 (green), and m4 (blue). (C) Raw data and (D) normalized data illustrate the predicted shift in both basal levels and EC_{50} for increasing stabilization of conformation B. Data represent independent measurement of triplicate samples, where the standard error was below 5% of each mean value.

DISCUSSION

The competition between reversible and irreversible rate constants establishes three operating regimes with distinct tuning properties. Therefore, measuring the reversible and irreversible rate constants is critical when predicting the impact of parameter modulation on the response curve. While well-established methods allow measurement of the rates of mRNA degradation and ligand binding and release, measuring the rates of RNA folding and conformational inter-conversion is currently an active area of research. New technologies are emerging that allow the measurement of kinetic folding rates: site-specific incorporation of aminopurines (Lang *et al*, 2007; Rieder *et al*, 2007), single-molecule force experiments (Greenleaf *et al*, 2008; Li *et al*, 2008; Woodside *et al*, 2006), and single-molecule fluorescence resonance energy transfer (Lee *et al*, 2007). Studies of natural and synthetic riboswitches that apply these approaches may yield a comprehensive understanding of the relationship between riboswitch function and performance (Al-Hashimi and Walter, 2008).

An alternative approach to measuring conformational switching relies on parameter predictions with RNA folding algorithms. Most algorithms calculate the free energy of individual conformations and can be used to estimate the value of K_1 for a riboswitch sequence (Mathews *et al*, 2004; Parisien and Major, 2008). Algorithms have also been developed that provide estimates of the rate constants for conformational switching (k_1, k_1') (Danilova *et al*, 2006). By employing these algorithms, sequences can be rapidly screened *in silico* to identify riboswitches with tuned conformational partitioning according to model predictions. Mutations that impact other parameters, such as mutations to the RBS sequence that affect regulatory activity, can also be screened *in silico* to evaluate the impact on secondary structure and conformational partitioning. However, these algorithms are often inaccurate when predicting RNA folding *in vivo*, requiring modified approaches (Beisel *et al*, 2008) or the development of more advanced algorithms (Parisien and Major, 2008).

Design principles for synthetic riboswitches

Synthetic riboswitches can be divided into two categories based on the intended application: inducible regulators and autonomous regulators. The applicable category depends on the identity and source of the detected ligand and requires distinct approaches to riboswitch design. We provide the following design principles assembled from our modeling results to guide the design of synthetic riboswitches as inducible or autonomous regulatory systems.

The desired properties of inducible regulatory systems include large dynamic ranges, low basal expression levels, and titratable control over expression levels.

Selecting an effective regulatory mechanism is critical since numerous factors reduce the dynamic range, such as conformational partitioning, dominating irreversible rates, upper limits to ligand concentration, and reduced gene regulatory efficiencies from the incorporation of other riboswitch elements (Beisel *et al*, 2008; Win and Smolke, 2007). A design that is biased toward forming the disrupted-aptamer conformation (high K_1) will generally increase the dynamic range, although such strategies require higher ligand concentrations to modulate protein levels. The rates of events separate from core riboswitch processes, such as transcription, translation, and protein decay, can be modulated to increase the dynamic range difference at the expense of increased basal levels.

The selected regulatory mechanism will likely dictate the values of the irreversible rate constants and thus the operating regime. In support of this, studies on natural riboswitches have suggested a consistent pairing between translational repression and the thermodynamically-driven regime (Rieder *et al*, 2007) and transcriptional termination and the non-functional regime with ligand binding during transcription (Lang *et al*, 2007; Lemay *et al*, 2006; Rieder *et al*, 2007; Wickiser *et al*, 2005a; Wickiser *et al*, 2005b). Therefore, the design of inducible regulatory systems may rely on the tuning properties associated with each regime. While thermodynamically-driven riboswitches generally provide for the largest dynamic range, kinetically-driven and non-functional riboswitches can be designed to perform similarly using insights from our modeling efforts. In general, placing the aptamer toward the 5' end of the riboswitch sequence will preserve the dynamic range by biasing transcriptional folding toward conformation B. The exception is OFF-behaving riboswitches acting through mRNA destabilization,

which are insensitive to biased transcriptional folding (Figure S5.4). In addition, introducing pause sites and ensuring rapid conformational switching from the aptamer-formed conformation (k_1') will allow kinetically-driven and non-functional riboswitches to exploit ligand binding during transcription, thereby decreasing the amount of ligand required to induce gene expression.

In many practical applications, system restrictions will limit the accessible range of the response curve (Figure 5.7A,B). Such limitations need to be addressed through parameter tuning in order to access the appropriate section of the response curve. For most biological systems, a predominant restriction is a limit to the maximum ligand concentration. In situations where the maximum ligand concentration does not saturate the response curve, designs for thermodynamically-driven riboswitches should be based on intermediate conformational partitioning values (K_1) that achieve a suboptimal maximum dynamic range. An alternative strategy is the design of non-functional riboswitches that bind ligand during transcription, which can respond to ligand at lower concentrations without sacrificing the dynamic range.

Genes often exist in regulatory networks that dictate cellular phenotype such that complex relationships exist between the expression levels of individual genes and systems-level functions. To effectively regulate these genes with synthetic riboswitches, a variety of tuning strategies must be employed to tune the response curve to operate within system restrictions. The properties of the regulated gene, its integration into a biological network, and the ultimate systems-level functions must be considered. One can envision an application-specific regulatory niche that defines the acceptable ranges of basal and maximum-ligand expression levels for proper system performance (Figure 7B). For

example, the properties of a given system may require that the riboswitch be tuned to minimize basal expression over maximizing dynamic range, such as when the regulated enzyme exhibits high activity or cytotoxicity.

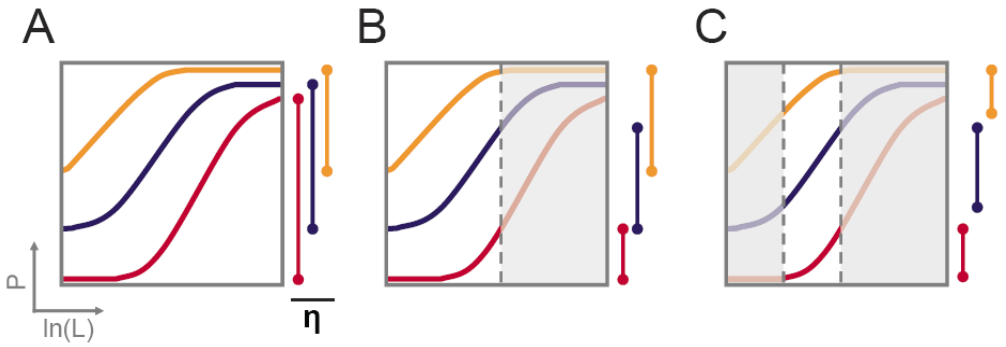


Figure 5.7 The accessibility of the riboswitch response curve depends on application category and associated system restrictions. Categories include an inducible regulatory system with (**A**) no ligand limitations or (**B**) a ligand concentration upper limit, and (**C**) an autonomous regulatory system with defined lower and upper limits for the ligand concentration. The accessible dynamic range (η) for each response curve depends on the system restrictions. The properties of other components in the network will dictate which riboswitch design best meets performance requirements. For example, under the autonomous regulatory system (C) the red curve may be more appropriate for the regulation of cytotoxic genes, the orange curve may be more appropriate for the regulation of enzymes with low catalytic activity, and the blue curve may be more appropriate for regulatory networks that require a large change in protein levels.

The engineering of synthetic riboswitches that act as autonomous regulatory systems presents an even greater design challenge. Here, the upper and lower ligand concentrations that the system fluctuates between establish the accessible section of the

response curve such that the regulatory niche is further restricted (Figure 5.7C). For example, riboswitches responsive to an endogenous central metabolite will likely be operating under a defined concentration range characteristic of the organism and the environment. In this case, the response curve must be tuned to place the desired expression levels at the limits of this defined concentration range by modulating the appropriate performance descriptors. Depending on system restrictions, proper tuning of riboswitches acting as autonomous control systems may require minimization of basal levels, operation across higher expression levels, or maximization of the change in expression levels.

Many parameters can potentially be modulated to tune the response curve. However, current practical considerations favor the modulation of a subset of these parameters in the laboratory. As one example, a given riboswitch may require a higher EC_{50} value to meet the performance requirements. Aptamer affinity (K_2), conformational partitioning (K_1), and the irreversible rates associated with the gene regulatory mechanism can be modulated to increase EC_{50}. However, rational modulation of aptamer affinity is restrictive since most mutations effectively abolish ligand binding, while the method and ease of modulating irreversible rates depend on the regulatory mechanism. Modulating conformational partitioning is an attractive approach since simple base-pairing interactions principally establish each conformation. However, conformational partitioning also impacts basal levels and the dynamic range, such that other parameters may need to be modulated to compensate for any undesired changes. Thus, the effective design of synthetic riboswitches requires knowledge of the relationship between

riboswitch sequence and model parameters and may require the coordinated modulation of multiple parameters to meet application-specific performance requirements.

The relationship between riboswitch sequence and model parameters depends in part on the composition framework used in the riboswitch design. A synthetic riboswitch can be designed such that parameters map to individual domains (Bayer and Smolke, 2005; Beisel *et al*, 2008; Win and Smolke, 2007) or multiple domains (Lynch *et al*, 2007; Topp and Gallivan, 2007, 2008). Each design strategy offers distinct advantages depending on whether rational design or random screening is used to select riboswitch sequences. Individual domain mapping strategies allow for insulated control over each parameter and domain swapping without requiring redesign of the riboswitch, thereby presenting significant advantages for rational design approaches. Multiple domain mapping strategies may be more desirable for random screening approaches, where assigning multiple parameters to a single sequence domain can reduce the number of randomized nucleotides required to sufficiently sample parameter space.

Evolutionary implications for tuning in natural riboswitches

Natural riboswitches primarily serve as key autonomous regulators of diverse metabolic processes (Winkler, 2005). Recent characterization of eleven known S-adenosylmethionine riboswitches in *Bacillus subtilis* demonstrated that these riboswitches exhibit a diverse range of values for basal expression levels, EC_{50}, and dynamic range (Tomsic *et al*, 2008), suggesting that natural riboswitches are finely tuned to match their occupied regulatory niche. However, this study is the only one to date to characterize the response curves of multiple natural riboswitches responsive to the same

ligand. Two questions emerge from these observations and our modeling results that underlie the biological utilization of natural riboswitches as dynamic regulators of metabolism: (1) how finely tuned are natural riboswitches to their regulatory niche, and (2) what sequence modifications are associated with response curve tuning?

Understanding the extent to which natural riboswitches are tuned to their regulatory niches will provide insights into riboswitch utilization and the underlying principles of genetic regulatory control. Similar to the tuning of synthetic riboswitches to match their intended regulatory niche, investigating the extent and biological relevance of natural riboswitch tuning requires knowledge of the functional properties of the regulated genes and their contribution to cellular fitness. Furthermore, the typical ligand concentration range encountered in the intracellular environment designates the operational section of the response curve, such that determining this range is critical to advancing our understanding of natural riboswitch tuning within regulatory niches.

The composition of a natural riboswitch dictates the relationship between its sequence and model parameters. One way to gain insights into this relationship is investigating sequence deviations between natural riboswitches in the same organism or different organisms that recognize the same ligand and employ the same regulatory mechanism. Using the response curve as a basis of comparison, these mutations may be neutral or shift the response curve in line with modulation of single or multiple parameters. Identifying which parameters are modulated will provide insights into how accessible each parameter is to random point mutations and how evolution effectively tunes the response curve through parameter modulation. Advances in our understanding of the biological utilization of natural riboswitches will enable researchers to better define

regulatory niches in a biological system and more effectively design synthetic riboswitches to match these niches. Beyond riboswitch design and implementation, insights into the fine-tuning of natural regulatory components and networks will enable the construction of biological networks that reliably control systems-level functions.

MATERIALS AND METHODS

Mathematical modeling. All modeling assumptions and methods are fully described in Text S5.2. Briefly, time-dependent differential equations were generated using mass action kinetics to describe each mechanistic step in the simplified molecular description of riboswitch function for translational repression, transcriptional termination, and mRNA degradation. The resulting equations were simplified by assuming steady-state conditions. Relevant tuning properties were identified based on the impact of model parameters on the response curve descriptors, including dynamic range (η) defined as the difference between high and low protein levels, ligand concentration to induce a half-maximal response (EC_{50}), basal protein levels ($P(L=0)$), and maximum-ligand protein levels ($P(L \rightarrow L'$ or $\infty)$).

Plasmid construction. pSAL8.3 served as the base plasmid for all experimental studies (Lynch *et al*, 2007). A theophylline-dependent synthetic riboswitch functioning through translational repression resides between the upstream P_{tac1} promoter and the downstream Tn10-β-Galactosidase fusion gene. Mutant sequences were cloned into the unique KpnI and HindIII restriction sites located directly upstream of the riboswitch and approximately 200 nucleotides into the fusion gene coding region. Primers harboring

mutant sequences (Table S1) and a 5' KpnI site were used to PCR amplify the 5' untranslated region extending through the HindIII restriction site. The resulting PCR product was digested with KpnI/HindIII, ligated into pSAL8.3 digested with the same restriction enzymes, and transformed into *Escherichia coli* strain DH10B. Assembled plasmid constructs were verified by sequencing (Laragen, Inc.). All molecular biology reagents and enzymes were obtained from New England Biolabs.

β-Galactosidase activity assay. β-Galactosidase assays were conducted using *E. coli* DH10B cells harboring the pSAL8.3 plasmid mutants based on modifications to previously described protocols (Lynch *et al*, 2007; Zhang and Bremer, 1995). Cells harboring each construct were grown overnight in Luria-Bertani (LB) broth supplemented with 50 μg/ml ampicillin. Overnight cultures were back-diluted into three separate wells containing 500 μl LB broth with 50 μg/ml ampicillin and the appropriate concentration of theophylline and grown at 37°C for 3 hrs with shaking at 210 RPM. Approximately 3 μl of the overnight culture was added to each well. Following the 3-hr incubation with shaking, optical density was recorded by transferring 175 μl into a 96-well microplate with a μClear bottom (Greiner) and measuring on a Safire fluorescence plate reader (Tecan). Cells were lysed by mixing 20 μl of culture with 80 μl permeabilization solution (100 mM Na_2HPO_4, 20 mM KCl, 2 mM $MgSO_4$, 0.6 mg/ml CTAB, 0.4 mg/ml sodium deoxycholate, and 5.4 μl/ml β-mercaptoethanol) and mixed at room temperature for approximately 10 min. In a fresh 96-well microplate, 25 μl of the lysed culture was mixed with 150 μl substrate solution (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 1 mg/ml ONPG, and 5.4 μl/ml β-mercaptoethanol). ONPG hydrolysis was

stopped with the addition of 75 µl 1 M Na_2CO_3 when a faint yellow color was observed. Absorbance at 420 nm was then measured on the fluorescence plate reader and protein levels were calculated in Miller Units (MU):

$$\mathrm{MU} = 1000 \cdot \frac{\mathrm{ABS}_{420}}{(0.025\,\mathrm{ml}) \cdot \mathrm{t} \cdot \mathrm{ABS}_{600}}\ , \tag{6.9}$$

where t is in minutes and absorbance values reflect the difference between each sample and blank media. The MU value of cells carrying a blank plasmid was also subtracted from each sample measurement.

ACKNOWLEDGEMENTS

We thank T. S. Bayer for critical reading of the manuscript, J.P. Gallivan for providing plasmid pSAL8.3.

REFERENCES

Acar M, Mettetal JT, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40:** 471-475.

Al-Hashimi HM, Walter NG (2008) RNA dynamics: it is about time. *Curr Opin Struct Biol* **18:** 321-329.

An CI, Trinh VB, Yokobayashi Y (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA* **12:** 710-716.

Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8:** R239.

Bayer TS, Smolke CD (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nat Biotechnol* **23:** 337-343.

Beisel CL, Bayer TS, Hoff KG, Smolke CD (2008) Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol* **4:** 224.

Belle A, Tanay A, Bitincka L, Shamir R, O'Shea EK (2006) Quantification of protein half-lives in the budding yeast proteome. *Proc Natl Acad Sci U S A* **103:** 13004-13009.

Bennett MR, Pang WL, Ostroff NA, Baumgartner BL, Nayak S, Tsimring LS, Hasty J (2008) Metabolic gene regulation in a dynamically changing environment. *Nature*.

Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99:** 9697-9702.

Breaker RR (2008) Complex riboswitches. *Science* **319:** 1795-1797.

Cheah MT, Wachter A, Sudarsan N, Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447:** 497-500.

Collins JA, Irnov I, Baker S, Winkler WC (2007) Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev* **21:** 3356-3368.

Corish P, Tyler-Smith C (1999) Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Eng* **12:** 1035-1040.

Crothers DM, Cole PE, Hilbers CW, Shulman RG (1974) The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA. *J Mol Biol* **87:** 63-88.

Danilova LV, Pervouchine DD, Favorov AV, Mironov AA (2006) RNAKinetics: a web server that models secondary structure kinetics of an elongating RNA. *J Bioinform Comput Biol* **4:** 589-596.

Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436:** 588-592.

Desai SK, Gallivan JP (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc* **126:** 13247-13254.

Emilsson GM, Nakamura S, Roth A, Breaker RR (2003) Ribozyme speed limits. *RNA* **9:** 907-918.

Greenleaf WJ, Frieda KL, Foster DA, Woodside MT, Block SM (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science* **319:** 630-633.

Hao N, Nayak S, Behar M, Shanks RH, Nagiec MJ, Errede B, Hasty J, Elston TC, Dohlman HG (2008) Regulation of cell signaling dynamics by the protein kinase-scaffold Ste5. *Mol Cell* **30:** 649-656.

Isaacs FJ, Dwyer DJ, Collins JJ (2006) RNA synthetic biology. *Nat Biotechnol* **24:** 545-554.

Isaacs FJ, Dwyer DJ, Ding C, Pervouchine DD, Cantor CR, Collins JJ (2004) Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22:** 841-847.

Jenison RD, Gill SC, Pardi A, Polisky B (1994) High-resolution molecular discrimination by RNA. *Science* **263:** 1425-1429.

Kensch O, Connolly BA, Steinhoff HJ, McGregor A, Goody RS, Restle T (2000) HIV-1 reverse transcriptase-pseudoknot RNA aptamer interaction has a binding affinity in the low picomolar range coupled with high specificity. *J Biol Chem* **275:** 18271-18278.

Klein DJ, Ferre-D'Amare AR (2006) Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313:** 1752-1756.

Lang K, Rieder R, Micura R (2007) Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res* **35:** 5370-5378.

Leclerc GJ, Leclerc GM, Barredo JC (2002) Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines. *Cancer Cell Int* **2:** 1.

Lee TH, Lapidus LJ, Zhao W, Travers KJ, Herschlag D, Chu S (2007) Measuring the folding transition time of single RNA molecules. *Biophys J* **92:** 3275-3283.

Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA (2006) Folding of the adenine riboswitch. *Chem Biol* **13:** 857-868.

Levchenko A, Bruck J, Sternberg PW (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc Natl Acad Sci U S A* **97:** 5818-5823.

Li PT, Vieregg J, Tinoco I, Jr. (2008) How RNA unfolds and refolds. *Annu Rev Biochem* **77:** 77-100.

Lynch SA, Desai SK, Sajja HK, Gallivan JP (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem Biol* **14:** 173-184.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101:** 7287-7292.

Narsai R, Howell KA, Millar AH, O'Toole N, Small I, Whelan J (2007) Genome-wide analysis of mRNA decay rates and their determinants in Arabidopsis thaliana. *Plant Cell* **19:** 3418-3436.

Osborne SE, Ellington AD (1997) Nucleic Acid Selection and the Challenge of Combinatorial Chemistry. *Chem Rev* **97:** 349-370.

Pan T, Sosnick T (2006) RNA folding during transcription. *Annu Rev Biophys Biomol Struct* **35:** 161-175.

Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452:** 51-55.

Rieder R, Lang K, Graber D, Micura R (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. *Chembiochem* **8:** 896-902.

Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C (2003) Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res* **13:** 216-223.

Su LJ, Waldsich C, Pyle AM (2005) An obligate intermediate along the slow folding pathway of a group II intron ribozyme. *Nucleic Acids Res* **33:** 6674-6687.

Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* **315:** 1716-1719.

Suess B, Hanson S, Berens C, Fink B, Schroeder R, Hillen W (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res* **31:** 1853-1858.

Suess B, Weigand JE (2008) Engineered riboswitches - Overview, Problems and Trends. *RNA Biol* **5**.

Tomsic J, McDaniel BA, Grundy FJ, Henkin TM (2008) Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. *J Bacteriol* **190:** 823-833.

Topp S, Gallivan JP (2007) Guiding bacteria with small molecules and RNA. *J Am Chem Soc* **129:** 6807-6811.

Topp S, Gallivan JP (2008) Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem* **9:** 210-213.

Voigt CA, Wolf DM, Arkin AP (2005) The Bacillus subtilis sin operon: an evolvable network motif. *Genetics* **169:** 1187-1202.

Wickiser JK, Cheah MT, Breaker RR, Crothers DM (2005a) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44:** 13404-13414.

Wickiser JK, Winkler WC, Breaker RR, Crothers DM (2005b) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* **18:** 49-60.

Win MN, Klein JS, Smolke CD (2006) Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res* **34:** 5670-5682.

Win MN, Smolke CD (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A* **104:** 14283-14288.

Winkler WC (2005) Riboswitches and the role of noncoding RNAs in bacterial metabolic control. *Curr Opin Chem Biol* **9:** 594-602.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428:** 281-286.

Woodside MT, Anthony PC, Behnke-Parks WM, Larizadeh K, Herschlag D, Block SM (2006) Direct measurement of the full, sequence-dependent folding landscape of a nucleic acid. *Science* **314:** 1001-1004.

Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A* **99:** 16587-16591.

Zarrinkar PP, Wang J, Williamson JR (1996) Slow folding kinetics of RNase P RNA. *RNA* **2:** 564-573.

Zaslaver A, Mayo AE, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette MG, Alon U (2004) Just-in-time transcription program in metabolic pathways. *Nat Genet* **36:** 486-491.

Zhang X, Bremer H (1995) Control of the Escherichia coli rrnB P1 promoter strength by ppGpp. *J Biol Chem* **270:** 11181-11189.

Zhuang X, Bartley LE, Babcock HP, Russell R, Ha T, Herschlag D, Chu S (2000) A single-molecule study of RNA catalysis and folding. *Science* **288:** 2048-2051.

Zimmermann GR, Wick CL, Shields TP, Jenison RD, Pardi A (2000) Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. *RNA* **6:** 659-667.

SUPPLEMENTARY INFORMATION

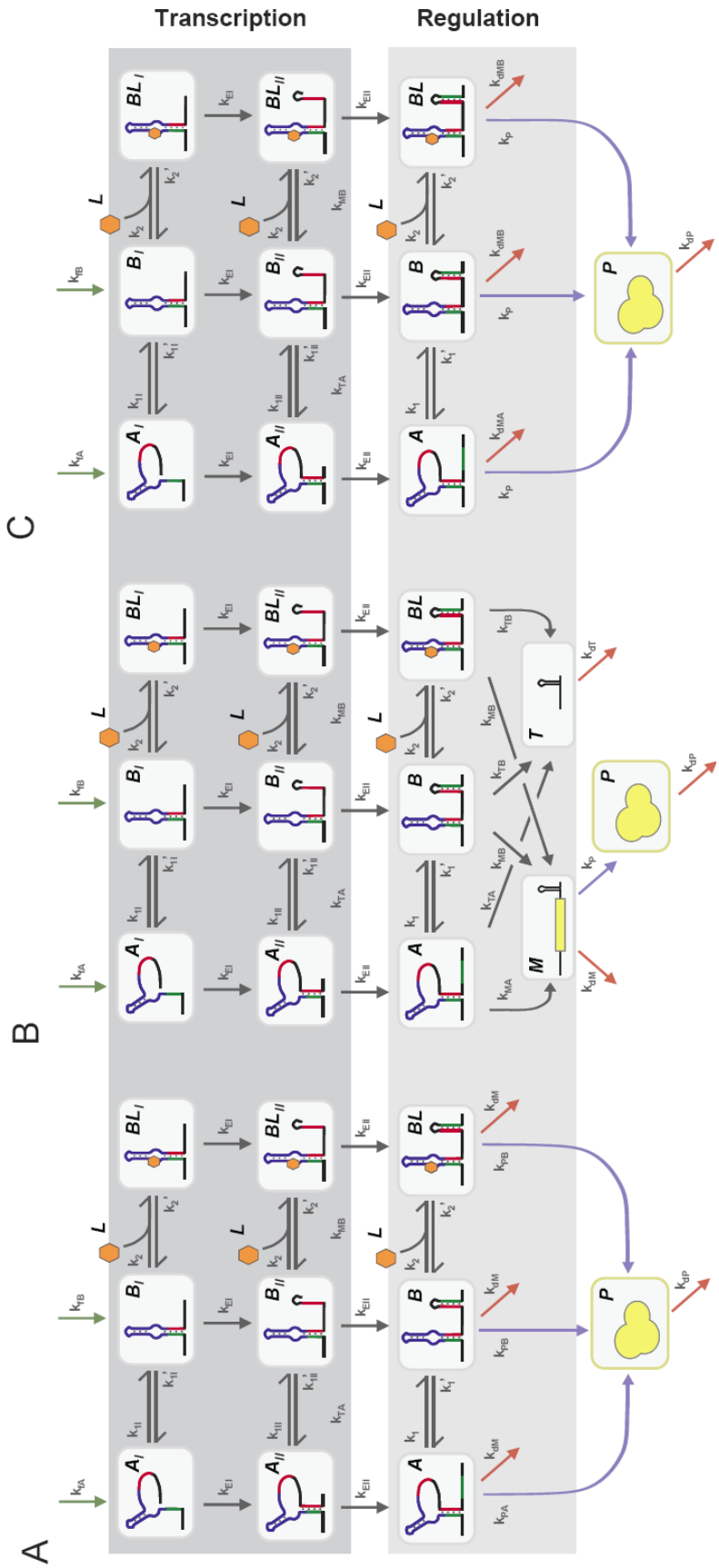

Transcription
Regulation
A
B
C

Figure S5.1 Detailed molecular description of riboswitch function. Molecular descriptions shown for riboswitches functioning through (**A**) translational repression, (**B**) transcriptional termination, and (**C**) mRNA destabilization. All riboswitches can reversibly switch between conformations A and B that display different regulatory activities or different rates of degradation. Conformation B contains a formed aptamer that can reversibly bind ligand (L). Transcription is represented as two discrete steps designated by the subscripts I and II, although the model can be extended to include more or less steps. Riboswitches at each step may switch between conformations or reversibly bind ligand depending on the transcribed sequence. The lag between steps is captured by the rate constant k_E. Once the full riboswitch sequence is transcribed, the riboswitch is susceptible to the regulatory mechanism that controls protein (P) production. Green arrows designate mRNA synthesis with biased transcriptional folding, red arrows designate species degradation, and blue arrows designate translation that is proportional to mRNA levels. Under transcriptional termination (**B**), riboswitches effectively choose between termination to form a truncated product (T) and extension to form the full-length mRNA (M). To ensure both conformations make the decision with the same frequency, we designated a rate constant k_M equal to the sum of the rate constants for extension and termination for either conformation ($k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB}$). We assume that the riboswitch sequence immediately prior to termination or extension cannot undergo degradation.

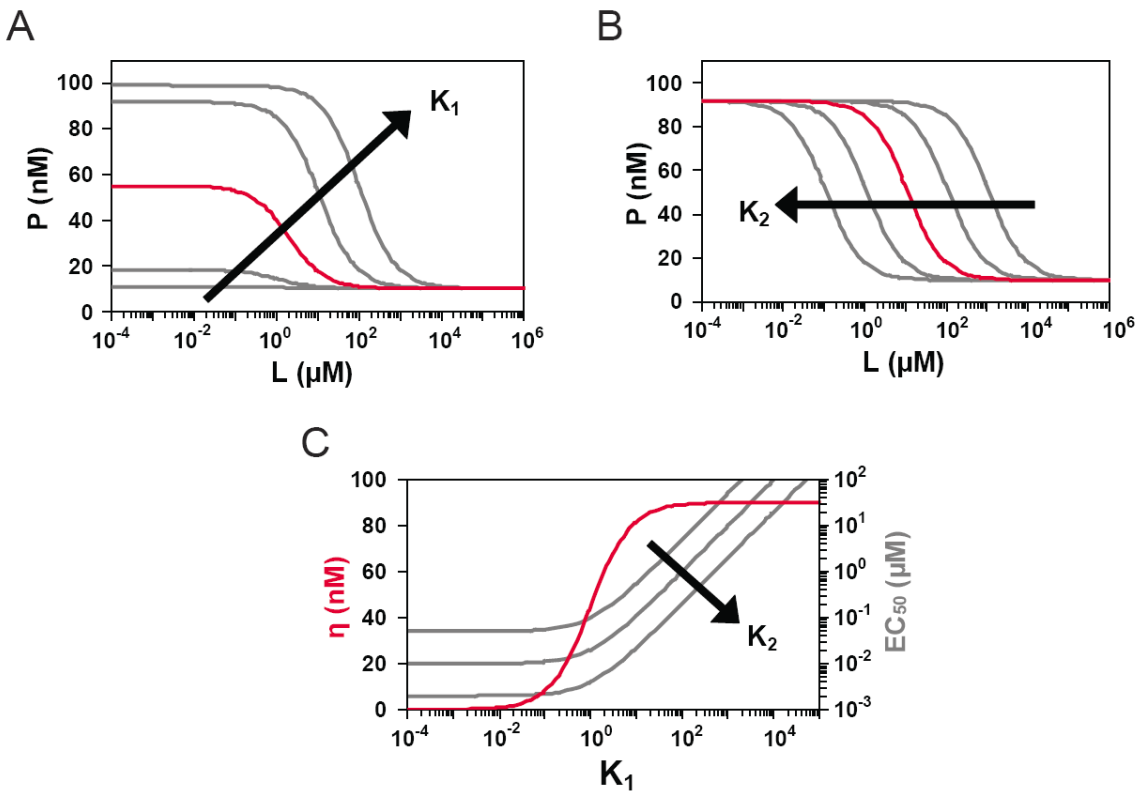


Figure S5.2 Thermodynamically-driven riboswitches exhibiting OFF behavior display similar tuning properties to riboswitches exhibiting ON behavior. K_1 is the conformational partitioning constant (k_1'/k_1) and K_2 is the aptamer association constant (k_2/k_2'). (**A**) K_1 affects both basal levels and EC_{50}. (**B**) K_2 only affects EC_{50}. (**C**) Biased conformational partitioning toward B maximizes the dynamic range at the cost of an increased EC_{50}. Parameter values are identical to those reported in Figure 5.2, except $K_A = 10^{-2}$/s; $K_B = 10^{-3}$/s.

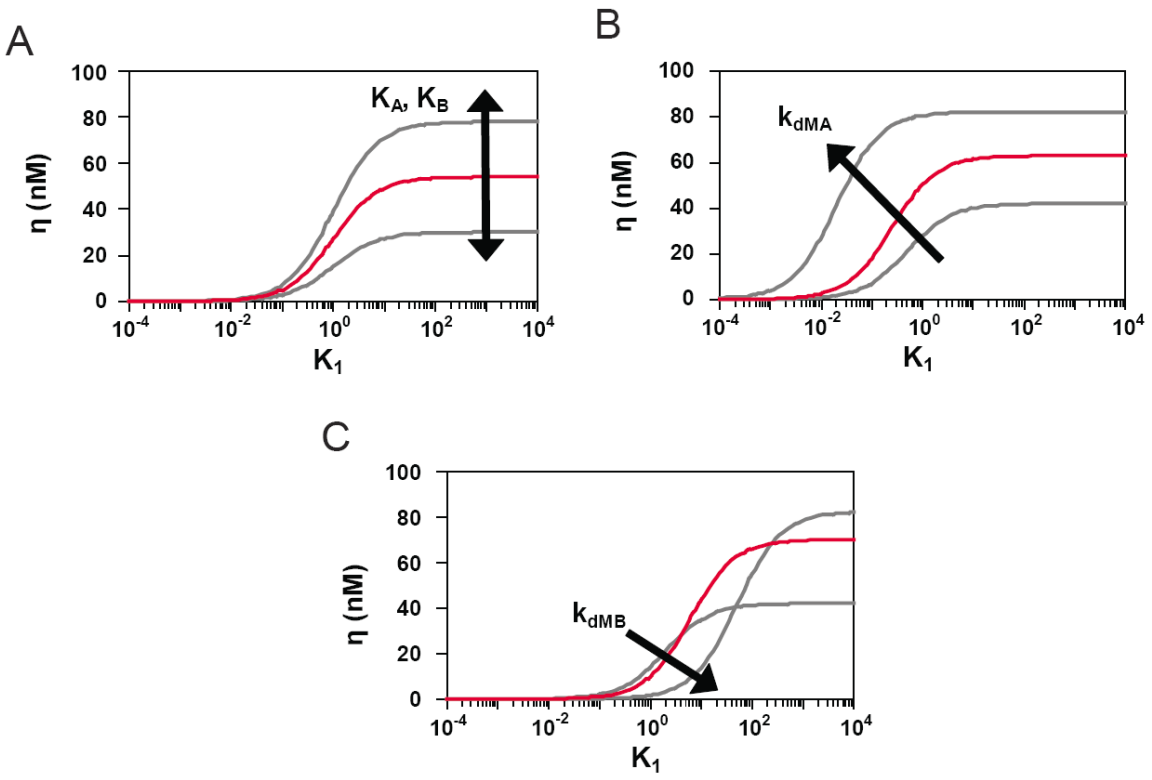


Figure S5.3 The mechanism-specific regulatory activities dictate differential tuning properties for thermodynamically-driven riboswitches. (**A**) Modulation of K_A and K_B affects the maximum dynamic range (η) for ON and OFF behaviors. K_A and K_B are modulated for riboswitches functioning through translational repression and transcriptional termination. Parameter values for red curve in (**A**): $K_A = 10^{-3}$/s; $K_B = 10^{-2}$/s; $k_f = 6 \cdot 10^{-12}$ M/s; $k_{dP} = 10^{-3}$/s; $k_{dMA} = k_{dMB} = 10^{-3}$/s. Both dynamic range and its dependence on K_1 change when irreversible rates are modulated, showing different trends for (**B**) ON and (**C**) OFF behaviors. The degradation rate constants k_{dMA} and k_{dMB} impact steady-state mRNA levels, thereby influencing the dynamic range. Parameter values for red curves in (**B**) and (**C**): $K_A = K_B = 1.4 \cdot 10^{-2}$/s; $k_f = 6 \cdot 10^{-12}$ M/s; $k_{dP} = 10^{-3}$/s; $k_{dMA} = 6 \cdot 10^{-3}$/s and $k_{dMB} = 10^{-3}$/s for ON behavior; $k_{dMA} = 10^{-3}$/s and $k_{dMB} = 6 \cdot 10^{-3}$/s for OFF behavior.

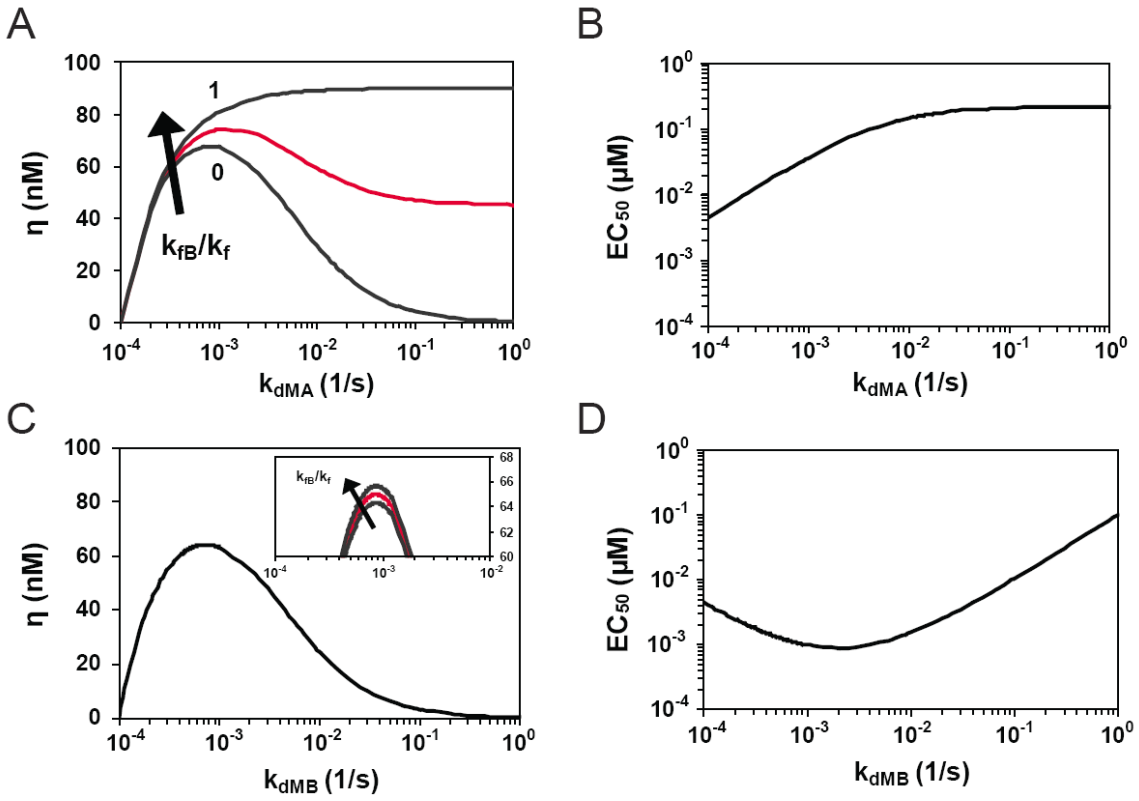


Figure S5.4 Distinction between tuning properties for ON and OFF behaviors for riboswitches functioning through mRNA destabilization. Dynamic range (η; **A**,**C**) and EC_{50} (**B**,**D**) display different dependencies on the dominant mRNA degradation rate constant for ON (k_{dMA}; **A**,**B**) and OFF (k_{dMB}; **C**,**D**) behaviors. Biased transcriptional folding significantly affects riboswitches displaying ON behavior. Riboswitches displaying OFF behavior show a negligible dependence on transcriptional folding (**C**, inset) for the selected parameter values. Parameter values: $k_1 = 5 \cdot 10^{-3}$; $k_1' = 2 \cdot 10^{-1}$; $k_2 = 10^6$/M·s; $k_2' = 10^{-3}$/s; $k_P = 10^{-3}$/s; $k_f = 10^{-11}$M/s; $k_{dP} = 10^{-3}$/s; $k_{dMA} = 10^{-4}$/s for OFF behavior; $k_{dMB} = 10^{-4}$/s for ON behavior.

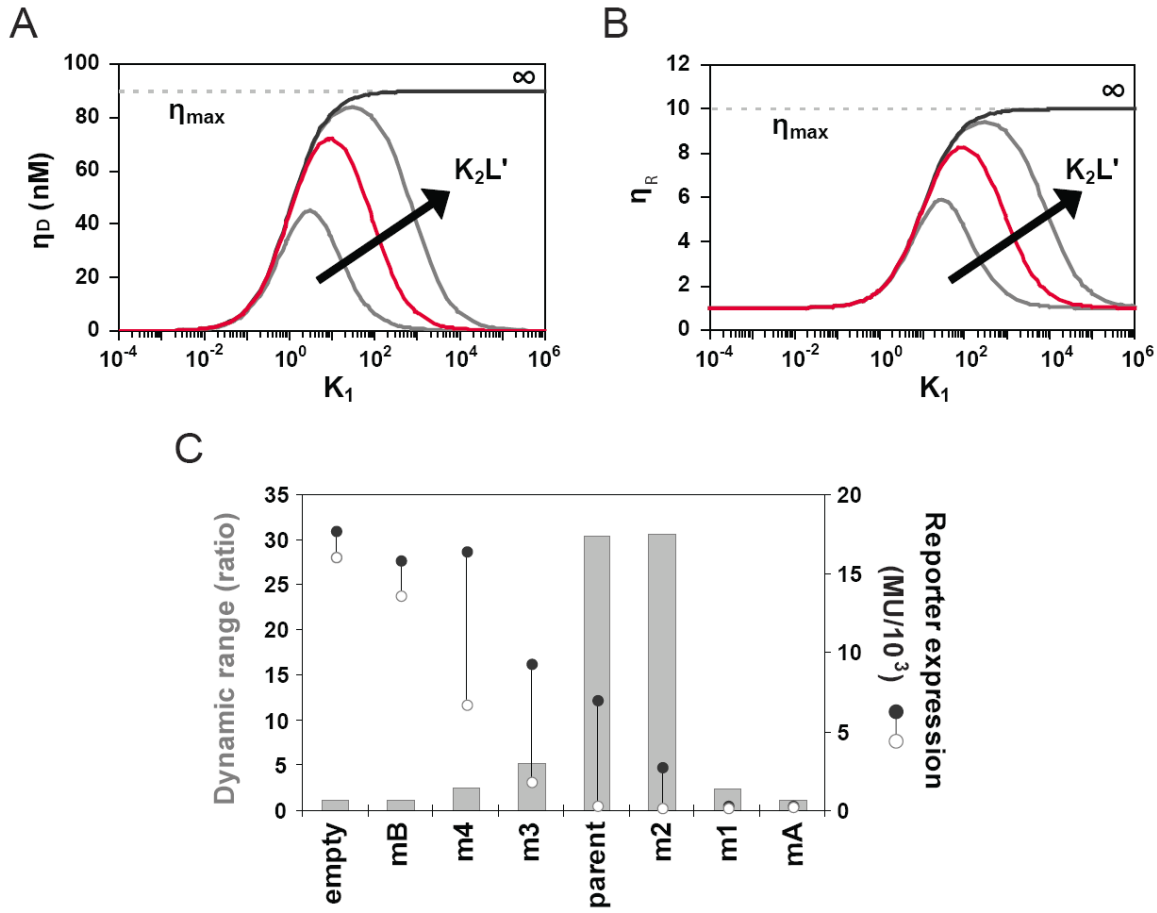


Figure S5.5 The dynamic range difference and ratio exhibit qualitatively similar tuning properties. Model predictions for the dynamic range difference (η_D, **A**) and ratio (η_R, **B**) when subjected to a ligand concentration upper limit (L'). In the absence of a ligand concentration upper limit, the dynamic range converges on a maximum (η_{max}). The optimum value of the conformational partitioning constant (K_1) is higher for the dynamic range ratio as the ratio favors lower basal levels. Increasing the aptamer association constant (K_2) or L' improve the suboptimal dynamic range maximum. Parameter values are identical to those reported in Figure 5.5, and notation is identical to that used in Figure 5.5B. (**C**) β-Galactosidase assay results from Figure 5B, where the dynamic range is calculated as the ratio of β-Galactosidase levels in the presence (○) and absence (●) of 1 mM theophylline. The positive control construct (empty) harbors only the RBS and aptamer basal stem. A slight increase in β-Galactosidase activity was observed in

the presence of theophylline for the control construct. The experimental data follow the general trends predicted from the model, including the higher optimum K_1 value for the dynamic range ratio as compared to the dynamic range difference. β-Galactosidase levels are reported in Miller Units (MU). Data represent independent measurement of triplicate samples, where the standard error was below 5% of each mean value.

Table S5.1 Sequence variants of pSAL8.3 and associated β-Galactosidase levels reported in Miller Units (MU). Database # is included for plasmid requests. The start codon is shown in green and point mutations are shown in blue. To generate each variant, the 5' end of the β-Galactosidase coding region was amplified and cloned into KpnI/HindIII of pSAL8.3 using primers 5'-AATAGGTACC-[Seq]-TGCGAACTC-3' and 5'-CGACGGG ATCGATCCCCCC-3', where [Seq] is the designated sequence in the table.

Name	Sequence	LacZ levels (MU)		Database #
		0 mM	1 mM	
parent	-GGT--GATACCAGCATCGTCTTGATGCCCTTGGCAGC--ACC-AGCTGCAAAGACAACAAG**ATG**	227	6876	pCS1301
m1	-**AA**T--GATACCAGCATCGTCTTGATGCCCTTGGCAGC--A**TT**-AGCTGCAAAGACAACAAG**ATG**	98	234	pCS1315
m2	-**A**GT--GATACCAGCATCGTCTTGATGCCCTTGGCAGC--AC**T**-AGCTGCAAAGACAACAAG**ATG**	87	2653	pCS1314
m3	-GGTC-GATACCAGCATCGTCTTGATGCCCTTGGCAGC-GACC-AGCTGCAAAGACAACAAG**ATG**	1778	9199	pCS1326
m4	-GGTCCGATACCAGCATCGTCTTGATGCCCTTGGCAGCGGACC-AGCTGCAAAGACAACAAG**ATG**	6667	16380	pCS1327
mA	-GGT--**A**ATACCAGCA-CGTCTTGATG**A**CCTT**A**GCAGC--ACC-AGCTGCAAAGACAACAAG**ATG**	170	188	pCS1326
mB	GGGT--GATACCAGCAT**GAAGAGC**ATGCCCTTGGC**TC**C--ACCCAGCTGCAAAGACAACAAG**ATG**	13557	15796	pCS1325
empty	-GGT--GATA----------AGC--ACC-AGCTGCAAAGACAACAAG**ATG**	15980	17661	pCS1324

Supplementary Text S5.1

ALTERNATIVE DEFINITIONS FOR DYNAMIC RANGE

The performance of an inducible regulatory system, such as a riboswitch, can be fully defined by a small collection of response curve descriptors: the dynamic range, effective inducer concentration to achieve a half-maximal response (EC_{50}), basal or ligand-saturating protein levels, and hill coefficient. Of these descriptors, the dynamic range is the most popular single measure of performance when comparing systems. The dynamic range can be reported as either the ratio of high to low protein levels (η_R) or the difference between these levels (η_D). While equally acceptable, one calculation method may be more appropriate than the other depending on the character of the response curve and the performance requirements of the system in which the riboswitch will be integrated. Measuring the performance of a riboswitch by its dynamic range without regard to other descriptors can inappropriately bias the selection of a suitable regulatory system for a given application. For example, a dynamic range ratio of 5 has a very different functional meaning for a system with a basal protein level of 1 molecule per cell than 50 molecules per cell. Dynamic range ratio values are biased to favor minimized basal protein levels, whereas dynamic range difference values are biased to favor larger absolute changes in protein levels.

Both measures of dynamic range are used to address performance requirements for the intended application. In most applications, the inducible regulatory system mediates switching between two phenotypic states determined by expression levels of the

regulated genes. The transition between these two states is application-dependent with regards to the regulated gene expression threshold to switch phenotypes and the sensitivity around this threshold. Therefore, the selected inducible regulatory system must allow gene induction or repression across this threshold, requiring basal and ligand-saturating levels outside the range sensitive to transition. Additional restrictions on upper and lower basal and ligand-saturating levels may exist due to the impact of excessively high or low protein levels on cellular fitness and function. Since these factors will be application-specific, the performance properties of the regulatory system will most likely need to be tuned. As discussed in the main text, the kinetics of riboswitch function can be modulated to tune the riboswitch response curve. The relationships between parameters that can be effectively modulated, dynamic range, and other response curve descriptors can be used to tune a riboswitch to meet application-specific performance requirements.

As shown in Table S5.2 for a thermodynamically-driven riboswitch, η_R and η_D display differential dependence on the regulatory activities of conformations A (K_A) and B (K_B). Calculation of η_R is the same for all riboswitch mechanisms and maintains K_A/K_B, such that this value is dimensionless and insensitive to parameter modulations that equally affect K_A and K_B. The drawback to using η_R in computational analyses is that the equations for ON and OFF behaviors are not equivalent, such that the predicted tuning properties for η_R require designation of either ON or OFF behavior. Specifically, η_R is linearly dependent on K_A/K_B for OFF behavior and is maximized for an intermediate value of K_A/K_B for ON behavior.

In contrast, the calculation of η_D requires units and the ratio of the irreversible rates between conformations B and A. However, the equations for ON and OFF

behaviors are equivalent, such that any elucidated tuning properties are applicable to both behaviors. This property simplifies the computational analyses and facilitates the elucidation of general design principles. We therefore reported dynamic range as the difference between high and low protein levels in the main text. However, we also calculated the tuning properties based on the dynamic range ratio for riboswitches operating in the thermodynamically-driven, kinetically-driven, and non-functional regime. In all cases the qualitative tuning properties were similar for η_D and η_R (data not shown).

Table S5.2 Comparison of dynamic range calculations for a thermodynamically-driven riboswitch. Dynamic range is calculated as either the difference between (η_D) or the ratio of (η_R) high and low protein levels. Definitions of the model parameters are provided in Figure 5.1A-C and Text S2.

Riboswitch behavior	**η_D**	**η_R**
ON ($K_B > K_A$)	$\frac{k_f}{k_{dP}} \frac{K_1}{K_1 + \frac{k_{dMB}}{k_{dMA}}} \left(\frac{K_B}{k_{dMB}} - \frac{K_A}{k_{dMA}} \right)$	$1 + \left(\frac{K_B}{K_A} \frac{k_{dMA}}{k_{dMB}} - 1 \right) \frac{K_1}{\frac{K_B}{K_A} \frac{k_{dMA}}{k_{dMB}} + K_1}$
OFF ($K_A > K_B$)	$\frac{k_f}{k_{dP}} \frac{K_1}{K_1 + \frac{k_{dMB}}{k_{dMA}}} \left(\frac{K_A}{k_{dMA}} - \frac{K_B}{k_{dMB}} \right)$	$1 + \left(\frac{K_A}{K_B} \frac{k_{dMB}}{k_{dMA}} - 1 \right) \frac{K_1}{1 + K_1}$

Supplementary Text S5.2

DERIVATION OF MATHEMATICAL MODELS

The mathematical models used to investigate riboswitch performance were derived from molecular descriptions reflecting translational repression, transcriptional termination, and mRNA destabilization (Figure 5.1B-D). Each description tracks an mRNA encoding the riboswitch and regulated gene(s) from birth to death using kinetic rates to separate molecular species. We assumed that after transcription each riboswitch can reversibly fold into two distinct conformations designated as A and B, which neglects complex folding paths, kinetic traps, and misfolding in order to avoid a cumbersome model. Conformation B includes a formed aptamer such that ligand reversibly binds this conformation to generate a ligand-riboswitch complex (BL). While ligand binding during transcription may contribute to riboswitch performance (Wickiser *et al*, 2005a; Wickiser *et al*, 2005b), we initially neglected this phenomenon and designated independent transcription rate constants for conformations A (k_{fA}) and B (k_{fB}) to reflect biased transcriptional folding.

For ease of analysis, we include each molecular description from the main text along with an explanation of the associated assumptions. Table S5.3 contains definitions and base values of all rate constants included in the models.

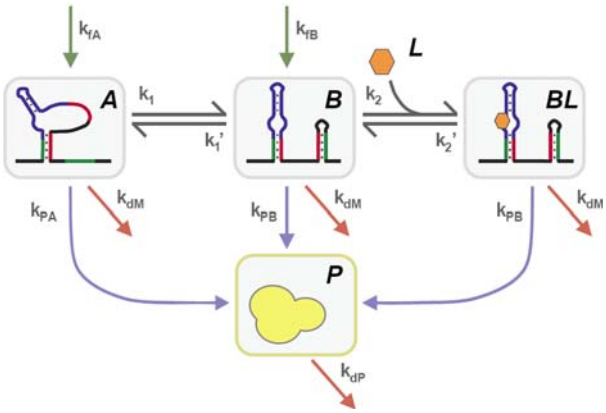


Figure 6.1B Molecular description for a riboswitch operating under translational repression. The riboswitch appears in either conformation A (k_{fA}) or B (k_{fB}) and reversibly switches between conformations (k_1, k_1') and undergoes irreversible degradation (k_{dM}) independent of conformation. Conformation B reversibly binds (k_2) and releases (k_2') the cognate ligand (L). The two conformations direct translation dependent on the strength and accessibility of the encoded ribosome binding site (k_{PA}, k_{PB}). We assumed translation does not affect conformational partitioning or mRNA degradation. Once translated, the protein (P) undergoes degradation (k_{dP}).

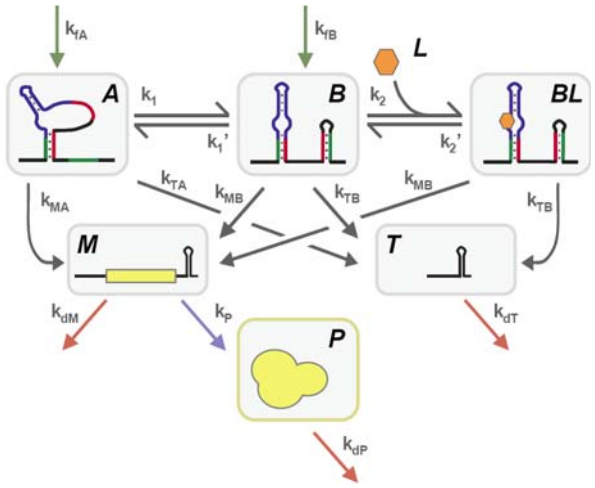


Figure 6.1C Molecular description for a riboswitch operating under transcriptional termination. The riboswitch appears in either conformation A (k_{fA}) or B (k_{fB}) as an intermediate in the transcriptional process and reversibly switches between conformations (k_1, k_1'). Conformation B reversibly binds (k_2) and releases (k_2') the cognate ligand (L). Both conformations choose between termination (k_{TA}, k_{TB}) and polymerase extension (k_{MA}, k_{MB}) at the same rate ($k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB}$). The full transcript (M) is translated into protein (k_P) and undergoes degradation (k_{dM}). Once translated, the protein (P) undergoes degradation (k_{dP}).

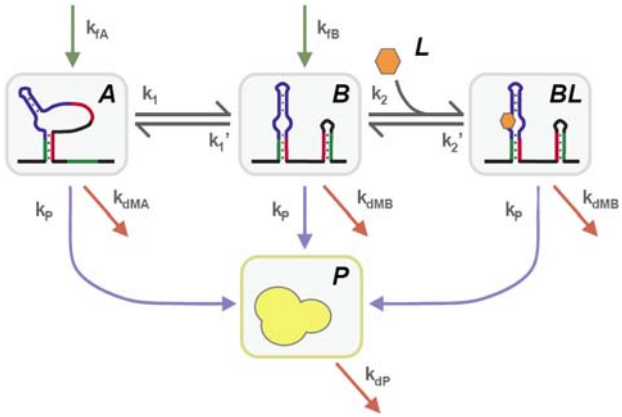


Figure 6.1D Molecular description for a riboswitch operating under mRNA destabilization. The riboswitch appears in either conformation A (k_{fA}) or B (k_{fB}) and reversibly switches between conformations (k_1, k_1'). Conformation B reversibly binds (k_2) and releases (k_2') the cognate ligand (L). The two conformations direct translation at the same rate (k_P) and undergo degradation at different rates (k_{dMA}, k_{dMB}) based on the regulatory mechanism. We assumed translation does not affect conformational partitioning or mRNA degradation. Once translated, the protein (P) undergoes degradation (k_{dP}).

Table S5.3 Rate constants used in all models.

Rate constant	Description	Base value
k_f	Sum of k_{fA} and k_{fB}. Represents total rate of transcription initiation	10^{-11}/s
k_{fA}	Rate of appearance of conformation A	$5 \cdot 10^{-12}$/s
k_{fB}	Rate of appearance of conformation B	$5 \cdot 10^{-12}$/s
k_E	Represents time transcribe full-length riboswitch	10^{-2}/s
k_1	Conformational switching from A to B	10^{-1}/s
k_1'	Conformational switching from B to A	10^{1}/s
k_2	Ligand binding rate	10^{6}/M·s
k_2'	Ligand release rate	10^{-1}/s
k_M	Represents time to decide whether to terminate transcription or continue elongation after full transcription of the full-length riboswitch	10^{-1}/s
k_{MA}	Rate of polymerase extension for conformation A	$9.1 \cdot 10^{-3}$/s
k_{MB}	Rate of polymerase extension for conformation B	$9.1 \cdot 10^{-2}$/s
k_{TA}	Rate of transcriptional termination for conformation A	$9.1 \cdot 10^{-2}$/s
k_{TB}	Rate of transcriptional termination for conformation B	$9.1 \cdot 10^{-3}$/s
k_{dM}	Transcript degradation rate	10^{-3}/s
k_{dMA}	Transcript degradation rate associated with conformation A	10^{-3}/s
k_{dMB}	Transcript degradation rate associated with conformation B	10^{-3}/s
k_{dT}	Truncated transcript degradation rate	10^{-3}/s
k_P	Translation rate	10^{-2}/s
k_{PA}	Translation rate for conformation A	10^{-3}/s
k_{PB}	Translation rate for conformation B	10^{-2}/s
k_{dP}	Protein degradation rate	10^{-3}/s

We generated an expression relating ligand concentration (L) and protein levels (P) for each regulatory mechanism and used these expressions to evaluate riboswitch performance. Ordinary differential equations were generated for each molecular species from the associated molecular description assuming mass action kinetics. Each equation was set equal to zero to evaluate performance under steady-state conditions. A set of performance descriptors (dynamic range; EC_{50}; basal levels; and ligand-saturating levels) was then calculated to explore the relationship between model parameters and riboswitch performance. Dynamic range was calculated as the difference between high and low expression levels, although calculation of the dynamic range as a ratio of these two

values is equally valid (Text S5.1). Table S5.4 contains general equations for the response curve descriptors for each regulatory mechanism and Table S5.5 contains mechanism-specific parameters.

Table S5.4 General equations for performance descriptors. Performance descriptors include dynamic range difference (η), EC_{50}, basal levels (P(L=0)), and ligand-saturating levels (P(L→∞)). Parameters relate to molecular descriptions in Figure 5.1B-D. K_1 is the conformational partitioning constant ($K_1 = k_1'/k_1$) and K_2 is the aptamer association constant ($K_2 = k_2/k_2'$). Mechanism-specific irreversible rates k_{IA} and k_{IB}, competition ratios γ_1 and γ_2, and conformational activities K_A and K_B are described in Table S5.5. The absolute value sign reflects the sign change between ON ($K_B > K_A$) and OFF ($K_A > K_B$) behaviors.

Performance descriptor	General equation
η	$\frac{k_f}{k_{dP}} \cdot \frac{K_1\gamma_1}{K_1\gamma_1 + \frac{k_{dMB}}{k_{dMA}}} \left(\frac{k_{fA}\gamma_1 + k_{fB}}{k_f} \right) \left\| \frac{K_A}{k_{dMA}} - \frac{K_B}{k_{dMB}} \right\|$
EC_{50}	$\frac{1 + K_1\gamma_1 \frac{k_{dMA}}{k_{dMB}}}{K_2\gamma_2}$
P(L=0)	$\frac{k_f}{k_{dP}} \cdot \left[\frac{k_{fA}}{k_f} \frac{K_A}{k_1} \gamma_1 + \frac{K_A K_1 \gamma_1 + K_B}{k_{dMA} K_1 \gamma_1 + k_{dMB}} \left(\frac{k_{fA}\gamma_1 + k_{fB}}{k_f} \right) \right]$
P(L→∞)	$\frac{k_f}{k_{dP}} \cdot \left[\frac{k_{fA}}{k_f} \frac{K_A}{k_1} \gamma_1 + \frac{K_B}{k_{dMB}} \left(\frac{k_{fA}\gamma_1 + k_{fB}}{k_f} \right) \right]$

Table S5.5 Mechanism-specific parameters for translational repression (TR), transcriptional termination (TT), and mRNA destabilization (MD). k_{iA} and k_{iB} represent the irreversible rates, K_A and K_B represent the regulatory activity of conformations, and γ_1 and γ_2 represent the competition between reversible and irreversible rates. The letters A and B in each suffix reflect the associated conformation. Parameters relate to molecular descriptions in Figure 5.1B-D. The irreversible rate for transcriptional termination (k_M) is the same for conformations A and B and is equal to the sum of the rates associated with termination (k_{TA}, k_{TB}) and read-through (k_{MA}, k_{MB}) ($k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB}$).

Model parameter	Regulatory Mechanism		
	TR	**TT**	**MD**
$\mathbf{k_{iA}}$	k_{dM}	k_M	k_{dMA}
$\mathbf{k_{iB}}$			k_{dMB}
$\mathbf{K_A}$	k_{PA}	$k_P \frac{k_{MA}}{k_M}$	k_P
$\mathbf{K_B}$	k_{PB}	$k_P \frac{k_{MB}}{k_M}$	
$\boldsymbol{\gamma_1}$	$\frac{k_1}{k_1 + k_{dM}}$	$\frac{k_1}{k_1 + k_M}$	$\frac{k_1}{k_1 + k_{dMA}}$
$\boldsymbol{\gamma_2}$	$\frac{k'_2}{k'_2 + k_{dM}}$	$\frac{k'_2}{k'_2 + k_M}$	$\frac{k'_2}{k'_2 + k_{dMB}}$

The majority of our modeling efforts assumed that ligand binding during transcription was not a contributing factor. For thermodynamically-driven riboswitches, this is a valid assumption since the conformational equilibrium attained in the presence of ligand is not affected by extra time or opportunities for ligand binding. However, as

discussed in the main text, restoring function to a non-functional riboswitch requires the contribution of ligand binding during transcription. We focused our investigation of this phenomenon on transcriptional termination, since experimental data have shown that certain natural riboswitches functioning through this mechanism are non-functional based on our definition. The derived response curve descriptors for the molecular description of a non-functional riboswitch functioning through transcriptional termination (Figure 5.4A) are presented in Table S5.6.

Table S5.6 Performance descriptors for a non-functional riboswitch functioning through transcriptional termination. Performance descriptors include dynamic range difference (η), EC_{50}, basal levels (P(L=0)), and ligand-saturating levels (P(L→∞)). Parameters relate to the molecular description in Figure 4A. The absolute value sign reflects the sign change between ON ($k_{MB} > k_{MA}$) and OFF ($k_{MA} > k_{MB}$) behaviors. The rate constant for terminator stem formation (k_M) is the same for conformations A and B and is equal to the sum of the rates associated with termination (k_{TA}, k_{TB}) and read-through (k_{MA}, k_{MB}) ($k_M = k_{MA} + k_{TA} = k_{MB} + k_{TB}$).

Performance descriptor	General equation
η	$\frac{k_f k_P}{k_{dP} k_{dM}} \cdot \frac{k_1'}{k_1' + k_M} \left\lvert \frac{k_{MA}}{k_M} - \frac{k_{MB}}{k_M} \right\rvert$
EC_{50}	$\frac{1}{2k_2}\left[\sqrt{4(k_1' + k_M)(k_2' + k_E) + (k_1' + k_M + k_2' + k_E)^2} - (k_1' + k_M + k_2' + k_E)\right]$
P(L=0)	$\frac{k_f k_P}{k_{dP} k_{dM}} \left(\frac{k_1'}{k_1' + k_M} \frac{k_{MA}}{k_M} + \frac{k_M}{k_1' + k_M} \frac{k_{MB}}{k_M} \right)$
P(L→∞)	$\frac{k_f k_P}{k_{dP} k_{dM}} \cdot \frac{k_{MB}}{k_M}$

The above analyses assumed that the ligand concentration can saturate the response curve. However, a practical upper limit to the ligand concentration (L') exists for any regulatory system. In the main text we discussed the implications of setting a

ligand concentration upper limit for thermodynamically-driven riboswitches on the observed tuning properties. All response curve descriptors were evaluated under conditions in which the upper limit to the ligand concentration does not saturate the response curve. The associated equations are presented in Table S7.

Table S5.7 Performance descriptors for a thermodynamically-driven riboswitch subjected to a ligand concentration upper limit (L'). The general thermodynamically-driven riboswitch was used as the basis for these calculations. Parameters relate to the molecular description in Figure 5.2A. Dynamic range (η) and the apparent EC_{50} (EC_{50}^{APP}) reflect a restricted ligand concentration range between 0 and L'.

Performance descriptor	General equation
η	$\frac{k_f}{k_{dP}} \cdot \frac{K_1}{K_1 + \frac{k_{dMB}}{k_{dMA}}} \left(\frac{K_2 L'}{1 + \frac{k_{dMA}}{k_{dMB}} K_1 + K_2 L'} \right) \left\| \frac{K_A}{k_{dMA}} - \frac{K_B}{k_{dMB}} \right\|$
EC_{50}^{APP}	$\frac{1 + K_1 \frac{k_{dMA}}{k_{dMB}}}{K_2} \left(2 \frac{1 + K_1 \frac{k_{dMA}}{k_{dMB}}}{K_2 L'} + 1 \right)^{-1}$

REFERENCES

Wickiser JK, Cheah MT, Breaker RR, Crothers DM (2005a) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44:** 13404-13414.

Wickiser JK, Winkler WC, Breaker RR, Crothers DM (2005b) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* **18:** 49-60.

Chapter 6

Conclusions and Future Prospects

Design of ligand-responsive RNAi effectors

To extend riboswitch design to include RNAi, we rendered three RNAi effectors in mammalian cells – siRNAs, shRNAs, and miRNAs – ligand-responsive through the rational integration of an aptamer domain. Ligand binding induced a local or global conformational change, thereby modulating the extent of processing by the RNAi machinery and subsequent gene silencing. The resulting riboswitches can be broadly applied by identifying an appropriate guide strand sequence, selecting an aptamer against the molecule of choice that will control silencing, and following the prescribed design approach detailed for siRNAs (Chapter 2), shRNAs (Chapter 3), or miRNAs (Chapter 4). The benefit of each design is the relative ease in selecting new gene targets and recognized molecules based on the demonstrated modularity of the guide strand sequence and aptamer domain and the ability to tune riboswitch performance through strategies specific to each design. Additional benefits include the propensity for forward design using available computational tools and the ability to program multiple input-output functions through the rational integration of riboswitches that recognize different biomolecules and target different genes.

The experimentally validated designs all displayed ON behavior, where target gene expression levels increased in the presence of the recognized ligand. Of equal value would be the design of ligand-responsive RNAi effectors displaying OFF behavior. The

focus on ON behaving riboswitches in this work reflects the relative ease in physically disrupting rather than improving biological activity – in this case processing by the RNAi machinery. With the exception of the miRNA designs, the principles underlying ligand control of RNAi effectors could be readily extended to generate OFF-acting designs. The major caveat is the alternative behavior associated with a regulatory mechanism that relies on catalytic cleavage, such as irreversible processing by the RNAi machinery. As discussed in Chapter 5, the speed of the irreversible regulatory event can reduce riboswitch performance if it dominates over the rate of reversible events such as conformational switching or ligand binding and release. Designing riboswitches with biased conformational folding during transcription can relieve the reduced performance, although OFF-acting designs are insensitive to this approach. Therefore, the relative speed of the reversible events and processing by the RNAi machinery may restrict the accessible performance for OFF-acting designs.

At first glance, all three designs for ligand-responsive RNAi effectors may be interchangeably used. However, the redundancy is short-lived when considering the location within the RNAi pathway, the elucidated modes of tuning, and potential toxicity issues. Since the RNAi pathway spans the nucleus and the cytoplasm, localization of the recognized ligand would dictate the selected design. For instance, the miRNA design is better suited to recognize nuclear proteins while the siRNA design is better suited to recognize cytoplasmic proteins. In addition, the tuning strategies varied between designs, where the predominant strategy in the siRNA and shRNA designs was modifying the competing strand complementarity while the miRNA design relied on introducing different miRNA copies. Finally, recent insights into the cytotoxicity induced by RNAi

effectors suggests that the shRNA design may introduce bottlenecks in the RNAi pathway not associated with siRNAs or miRNAs, such that the shRNA designs may be less appropriate for therapeutic applications. These critical factors must be weighed when selecting the most appropriate design for ligand control of gene silencing.

Broad implementation of ligand-responsive RNAi effectors

The lack of superficial restrictions on the recognized ligand and targeted gene opens the described designs to diverse applications. For instance, riboswitches could be readily designed that respond to small molecules with ideal pharmacokinetic properties such as high solubility, tissue permeability, and serum stability, and low immunogenicity and cytotoxicity. RNAi-mediated targeting of any selected gene could then be controlled by adding or removing this small molecule either in cell culture or organisms with negligible side effects. Another application is the development of smart therapeutics – molecules that can decipher between diseased and healthy cells and only enact the therapeutic response in the diseased cells. Aptamers could be selected against known disease biomarkers, such as oncogenic, viral, or mutant proteins or perturbed metabolite levels. Selected aptamers could then be integrated into any of the ligand-regulated RNAi effectors that target essential genes or other therapeutically-relevant targets such that cell death will be induced only in the presence or absence of the biomarker. When the diseased state is typified by a quantitative increase or decrease in the marker, performance tuning will be essential to ensure that the typical perturbation in the disease marker induces a robust change in the therapeutic effect from modulated gene silencing.

Current challenges in riboswitch design

Before realizing the full implementation of ligand-responsive RNAi effectors and other riboswitches, a few major challenges must be overcome in the design process that should be the immediate focus of future research. First, aptamer selection must be advanced to yield aptamers with in vivo binding capacity that function in the context of riboswitches. Most available aptamers were selected under buffer conditions that do not resemble the intracellular environment, where the binding capacity of these aptamers is severely compromised under intracellular conditions (Lorsch and Szostak, 1994). Furthermore, a significant fraction of aptamers do not fit the standard stem-bulge-loop structure and instead form pseudoknots and dangling ends (Kim and Jeong, 2003; Koizumi and Breaker, 2000; Tuerk *et al*, 1992). These nonstandard aptamer conformations are incompatible with current riboswitch designs and may require alternative design schemes to accommodate these structures.

Besides conducting aptamer selections under conditions that mimic the intracellular environment, aptamers could be selected in the context of riboswitches. The precedence was established in the context of allosteric ribozymes with the in vitro selection of aptamers against nucleoside 3',5'-cyclic monophosphate compounds (Koizumi *et al*, 1999). While transitioning to an in vivo selection scheme is confounded by the reduced screening capacity in cells, recent work has suggested an intermediate approach that combines in vitro and in vivo aptamer selections (Weigand *et al*, 2008). Continued efforts may yield standard procedures for the selection of aptamers that function in the context of riboswitches in vivo, thereby facilitating the process of design and implementation.

A second challenge is predicting how a particular riboswitch sequence will behave when expressed in cells. Understanding the associated sequence-function relationship would greatly improve efficacy and reduce the time investment in riboswitch design. One major contributor to sequence-function relationships is RNA folding, where recent computational advances have offered algorithms for the prediction of RNA secondary and tertiary structure with improving accuracy and resolution (Mathews *et al*, 2004; Parisien and Major, 2008; Sharma *et al*, 2008). As these algorithms progress toward atomistic models of RNA folding under equilibrium conditions, the next major area of development is investigating the dynamics of folding and activity. RNA is a naturally dynamic molecule that can transition between many different conformations rather than adopting a single energetically favorable structure. Furthermore, RNAs transcribed by cells are unlikely to reach thermodynamic equilibrium because of RNA processing and degradation. Chapter 5 provides a coarse investigation of the contribution of dynamics to riboswitch activity, which may serve as a foundation in which to further explore the in vivo events that occur in the lifetime of an RNA molecule. By accurately accounting for the energetics of RNA folding and cellular interactions, we can move toward the development of in silico frameworks that reliably predict the behavior of an RNA sequence in cells. Such frameworks would be integral to the design process in the construction of riboswitches and other noncoding RNAs.

As more riboswitch designs become available, a third research challenge will be implementing riboswitches toward different applications. While riboswitches can seemingly render any gene sensitive to any intracellular molecule, care must be taken when identifying and pursuing applications suitable for riboswitch design. Paramount

considerations include the intracellular concentration range of the recognized molecule, the swing in gene expression required to robustly induce the prescribed phenotype, and the performance limitations of the riboswitch. These quantitative parameters must be specified as design goals or restrictions, marking a progression beyond the simple designation of ON or OFF behavior associated with riboswitches.

The ensuing challenge is identifying riboswitch sequences that meet the identified performance needs of the selected application. Simple screening and selection strategies have been developed that could be applied to identify riboswitch sequences with appropriate activity (Fowler *et al*, 2008; Lynch and Gallivan, 2009; Muranaka *et al*, 2009; Nomura and Yokobayashi, 2007; Topp and Gallivan, 2008; Weigand *et al*, 2008; Wieland and Hartig, 2008). However, the progression to larger systems in higher organisms with more complex phenotypes excludes simple screening strategies. In contrast, elucidated sequence-function relationships that link riboswitch performance and systems-level behavior would provide valuable tools to screen riboswitch sequences in silico.

The future of RNA design

The recent construction of synthetic RNAs and the characterization of natural noncoding RNAs continue to reveal the rich and versatile nature of RNA. The myriad of abilities afforded to this ubiquitous biomolecule are ripe for further engineering efforts, raising the question where RNA-based components should be utilized in the design of more sophisticated biological systems. An excellent case study is the natural world. In examples ranging from bacterial small RNAs to miRNAs, RNA is predominantly utilized

in combination with protein regulators to orchestrate cellular processes (Shalgi *et al*, 2007; Shimoni *et al*, 2007). These observations argue against the construction of systems entirely composed of RNA or protein regulators. Toward the construction of hybrid systems, recent engineering efforts have combined protein and RNAi-based regulation with an improved dynamic range and logic capabilities (Deans *et al*, 2007; Greber *et al*, 2008; Rinaudo *et al*, 2007). More studies in the natural utilization of RNA are required to fully comprehend how and why nature utilizes this biomolecule. Insights can be immediately translated into design strategies as we continue engineering biological systems.

REFERENCES

Deans TL, Cantor CR, Collins JJ (2007) A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130:** 363-372.

Fowler CC, Brown ED, Li Y (2008) A FACS-based approach to engineering artificial riboswitches. *Chembiochem* **9:** 1906-1911.

Greber D, El-Baba MD, Fussenegger M (2008) Intronically encoded siRNAs improve dynamic range of mammalian gene regulation systems and toggle switch. *Nucleic Acids Res* **36:** e101.

Kim MY, Jeong S (2003) RNA aptamers that bind the nucleocapsid protein contain pseudoknots. *Mol Cells* **16:** 413-417.

Koizumi M, Breaker RR (2000) Molecular recognition of cAMP by an RNA aptamer. *Biochemistry* **39:** 8983-8992.

Koizumi M, Soukup GA, Kerr JN, Breaker RR (1999) Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol* **6:** 1062-1071.

Lorsch JR, Szostak JW (1994) In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* **33:** 973-982.

Lynch SA, Gallivan JP (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res* **37:** 184-192.

Mathews DH, Disney MD, Childs JL, Schroeder SJ, Zuker M, Turner DH (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **101:** 7287-7292.

Muranaka N, Sharma V, Nomura Y, Yokobayashi Y (2009) An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Res* **37:** e39.

Nomura Y, Yokobayashi Y (2007) Reengineering a natural riboswitch by dual genetic selection. *J Am Chem Soc* **129:** 13814-13815.

Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452:** 51-55.

Rinaudo K, Bleris L, Maddamsetti R, Subramanian S, Weiss R, Benenson Y (2007) A universal RNAi-based logic evaluator that operates in mammalian cells. *Nat Biotechnol* **25:** 795-801.

Shalgi R, Lieber D, Oren M, Pilpel Y (2007) Global and Local Architecture of the Mammalian microRNA-Transcription Factor Regulatory Network. *PLoS Comput Biol* **3:** e131.

Sharma S, Ding F, Dokholyan NV (2008) iFoldRNA: three-dimensional RNA structure prediction and folding. *Bioinformatics* **24:** 1951-1952.

Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, Biham O, Margalit H (2007) Regulation of gene expression by small non-coding RNAs: a quantitative view. *Mol Syst Biol* **3:** 138.

Topp S, Gallivan JP (2008) Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem* **9:** 210-213.

Tuerk C, MacDougal S, Gold L (1992) RNA pseudoknots that inhibit human immunodeficiency virus type 1 reverse transcriptase. *Proc Natl Acad Sci U S A* **89:** 6988-6992.

Weigand JE, Sanchez M, Gunnesch EB, Zeiher S, Schroeder R, Suess B (2008) Screening for engineered neomycin riboswitches that control translation initiation. *RNA* **14:** 89-97.

Wieland M, Hartig JS (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl* **47:** 2604-2607.