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.

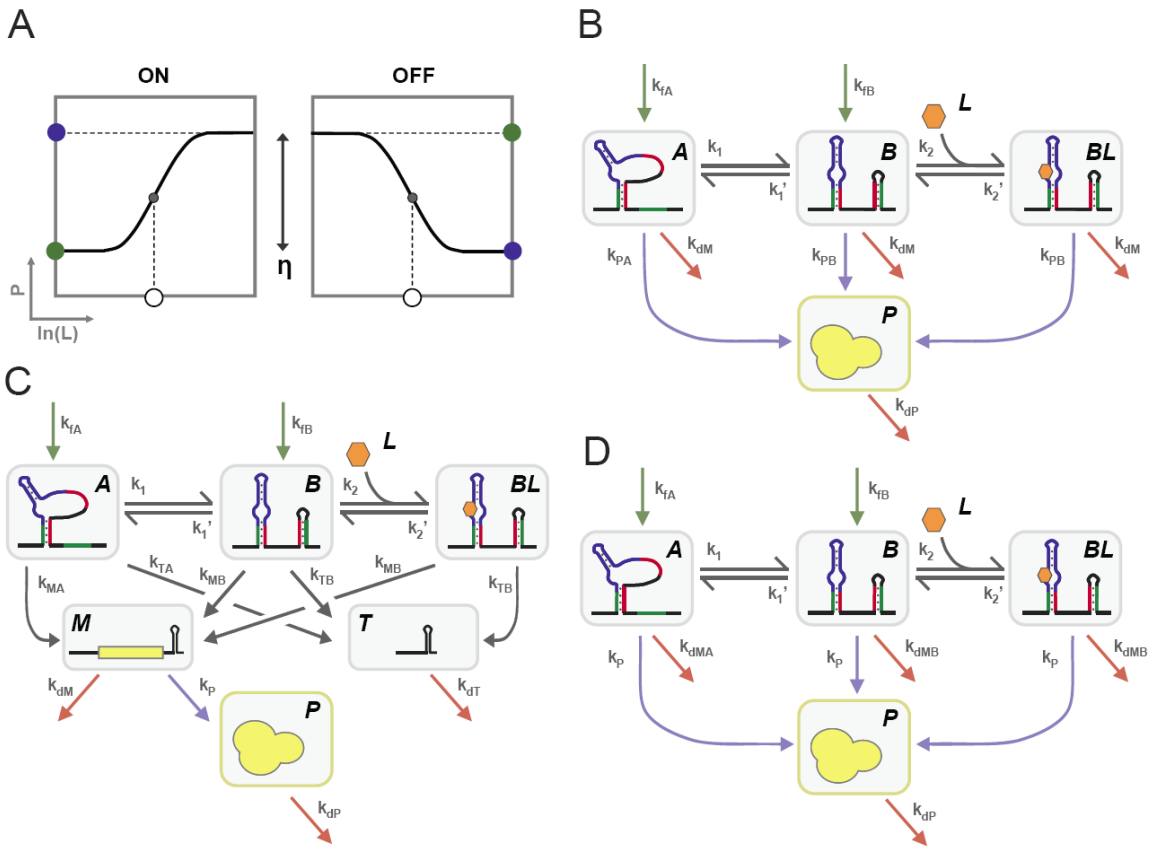

Figure 5.1 Kinetic models for riboswitches functioning through three distinct regulatory mechanisms. (**A**) General response curve relating ligand concentration (L) and regulated protein levels (P) for both ON and OFF behavior. Descriptors important in evaluating riboswitch performance are indicated: 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.

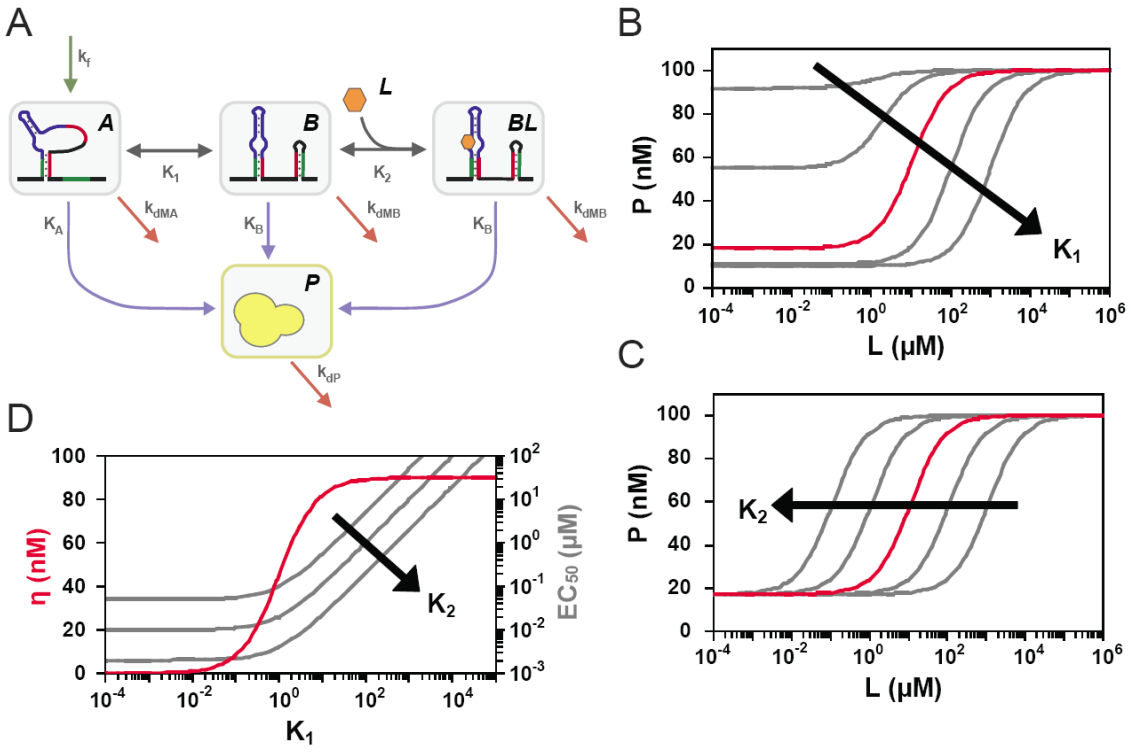

Figure 5.2 Thermodynamically-driven riboswitches display similar tuning properties. (**A**) Simplified molecular description that captures the three riboswitch regulatory mechanisms in the thermodynamically-driven regime. K_1 is the conformational partitioning constant (k_1'/k_1) and K_2 is the aptamer 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).

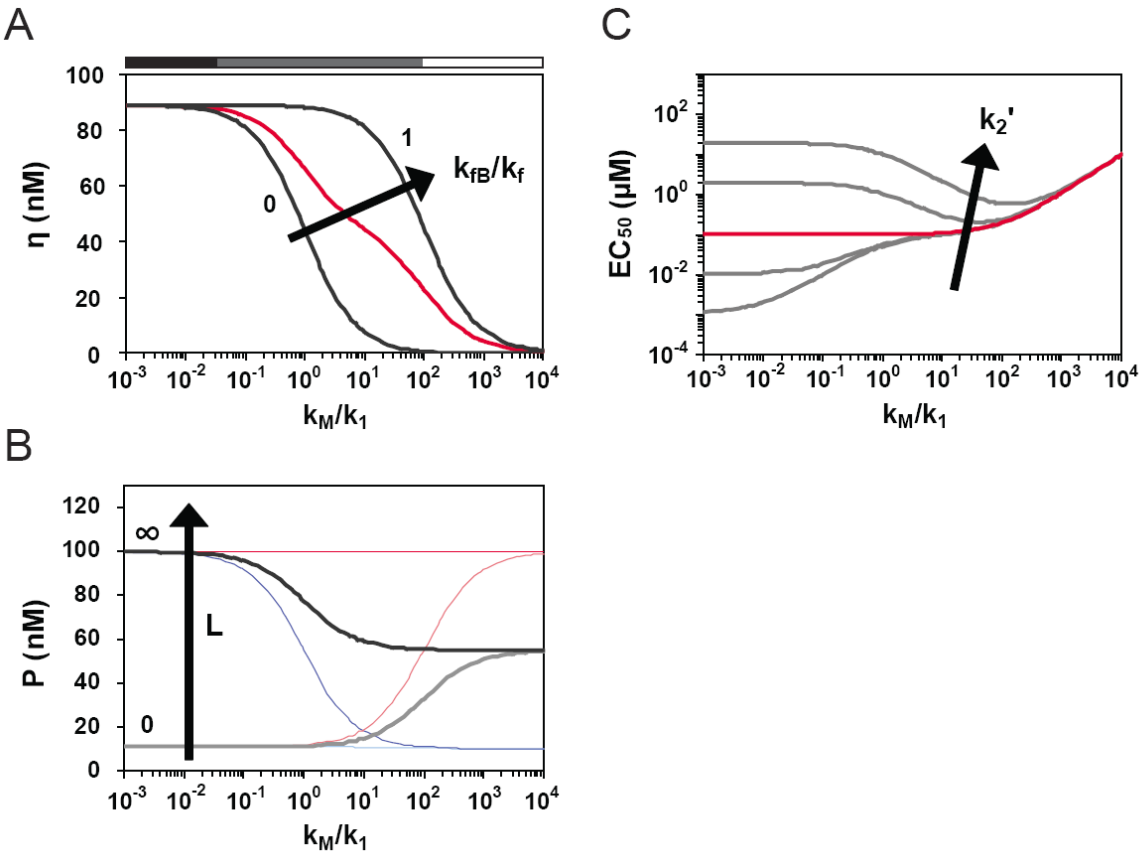

Figure 5.3 Rate competition dictates riboswitch performance. The relative values of the reversible and irreversible rate constants generally establish three operating regimes: thermodynamically-driven (■) when reversible rate constants dominate, kinetically-driven (■) when the rate constants are balanced, 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.

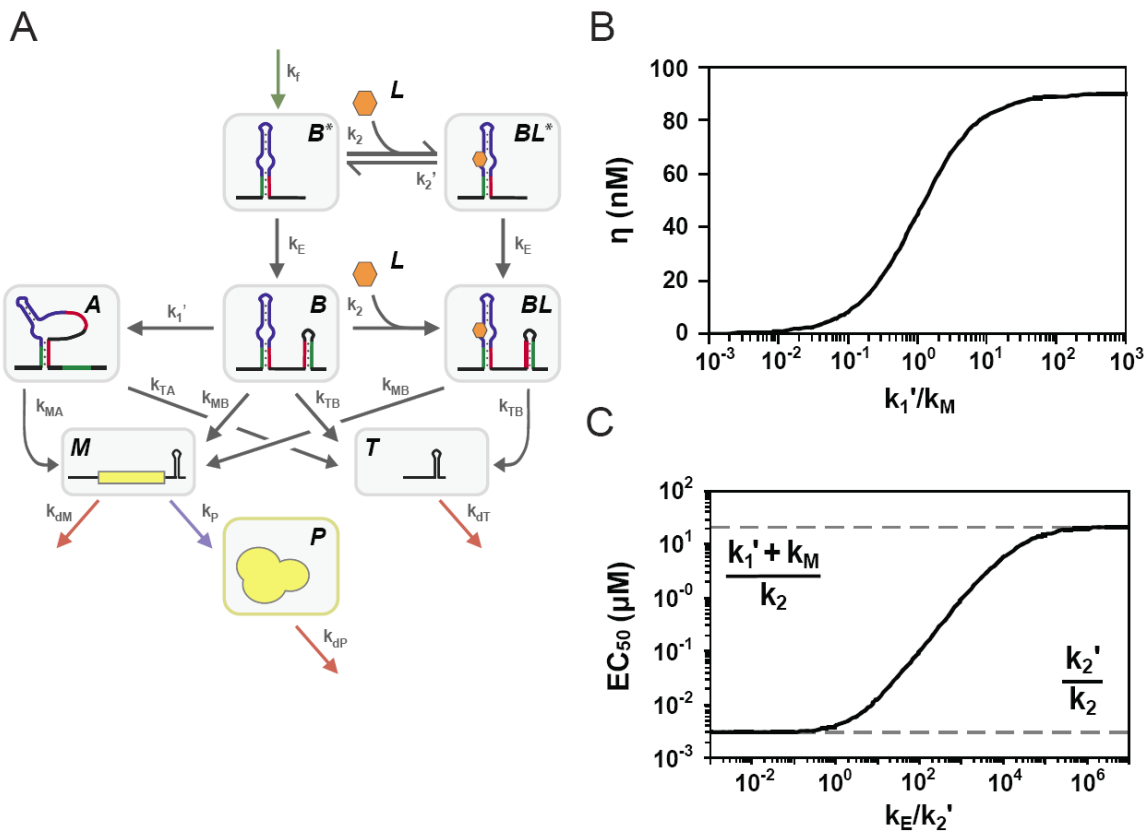

Figure 5.4 Rescuing riboswitch performance in the non-functional regime. (**A**) Molecular description of a non-functional riboswitch functioning through transcriptional termination. The description is identical to that in Figure 5.1C with notable exceptions: the aptamer in conformation B is transcribed 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'.

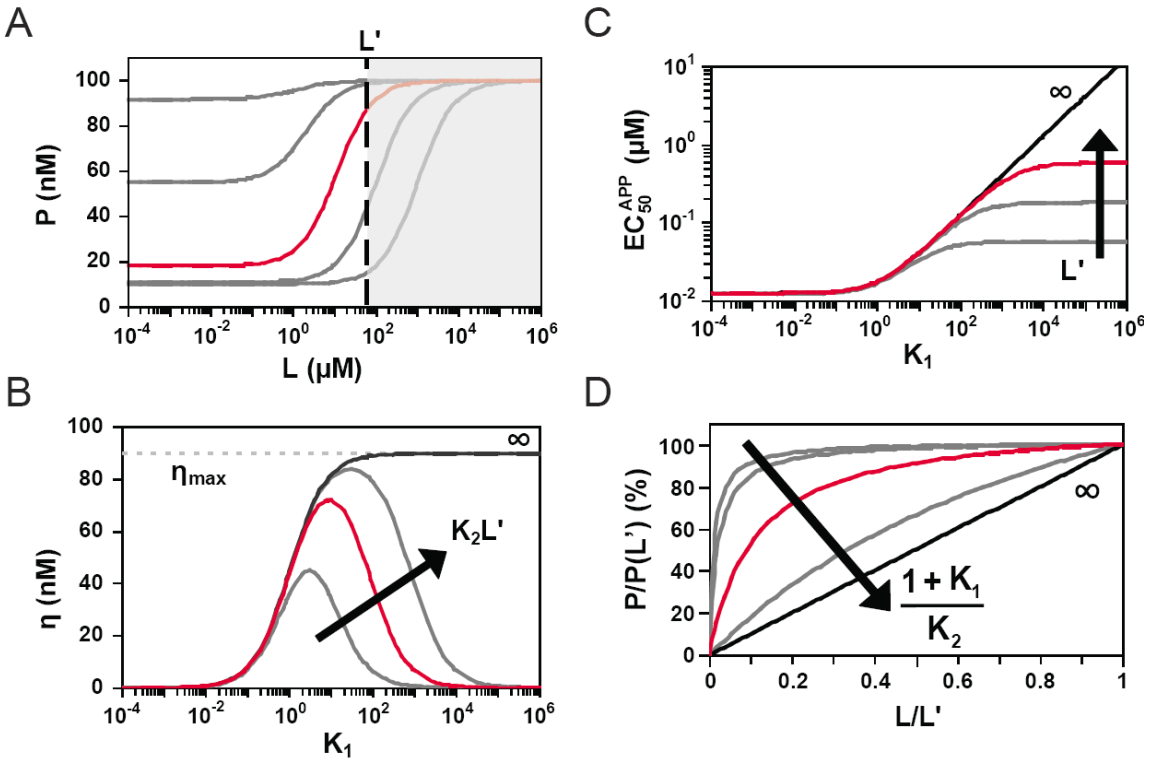

Figure 5.5 Placing an upper limit on the ligand concentration range alters the observed tuning properties. (**A**) Placing an upper limit on the ligand concentration (L') restricts access to the full response curve. This limit affects the dependence of (**B**) the dynamic range (η) and (**C**) the apparent 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.

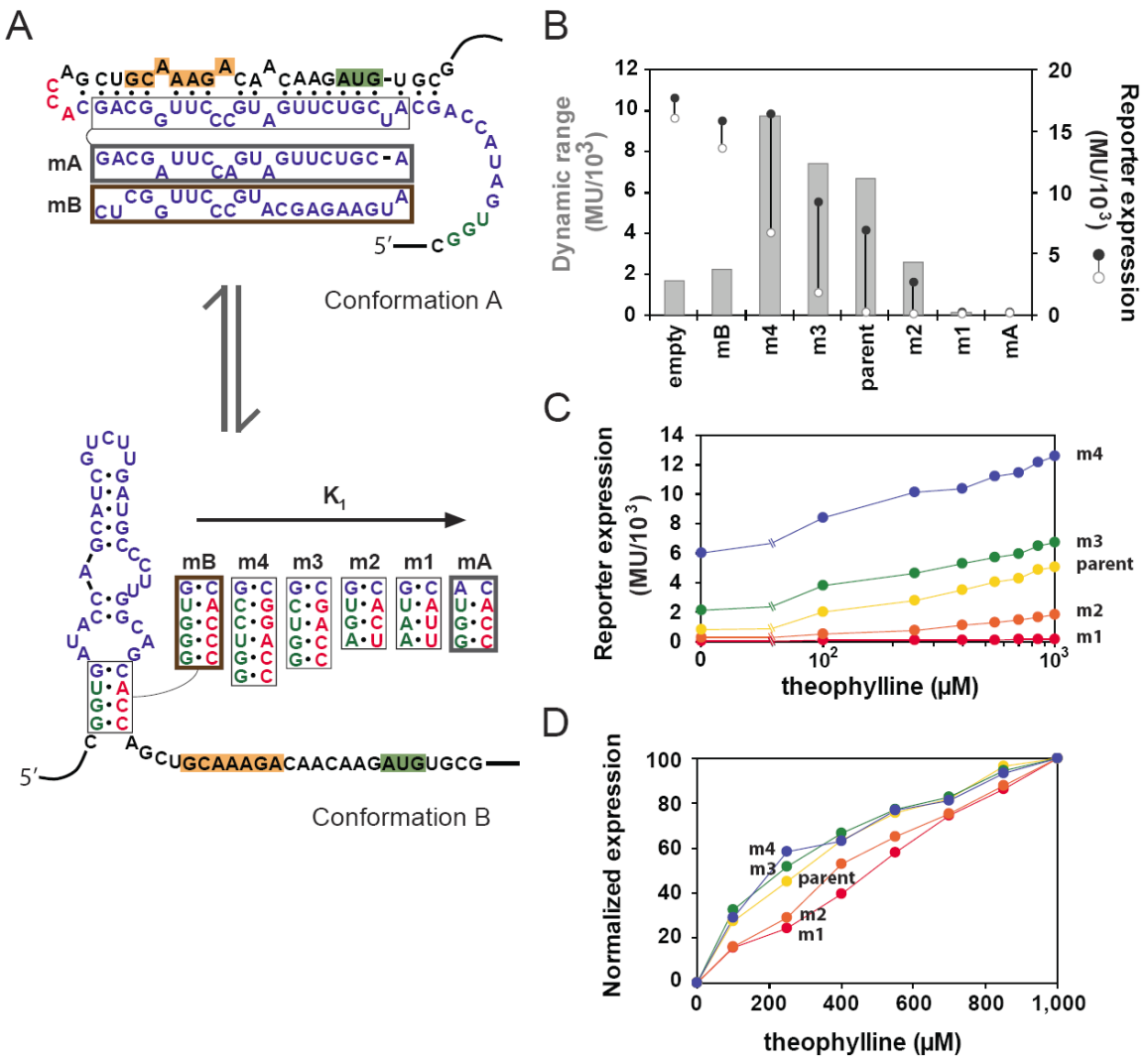

Figure 5.6 Mutational analysis of a synthetic riboswitch supports model predictions. (**A**) Mutations made to the aptamer stem of the parent synthetic riboswitch (m1-4) are anticipated to solely modulate conformational partitioning (K_1). The theophylline-responsive riboswitch controls Tn10-β-Galactosidase 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.

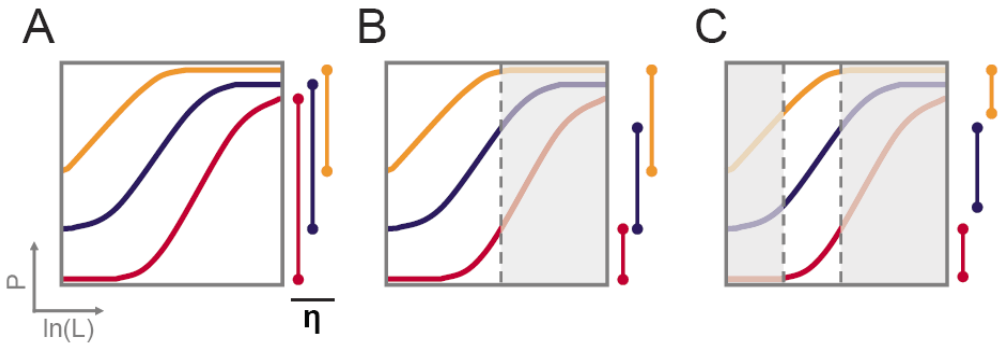

Figure 5.7 The accessibility of the riboswitch response curve depends on application category and associated system restrictions. Categories include an inducible regulatory system with (**A**) no ligand limitations or (**B**) a ligand concentration upper limit, and (**C**) an autonomous regulatory system 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

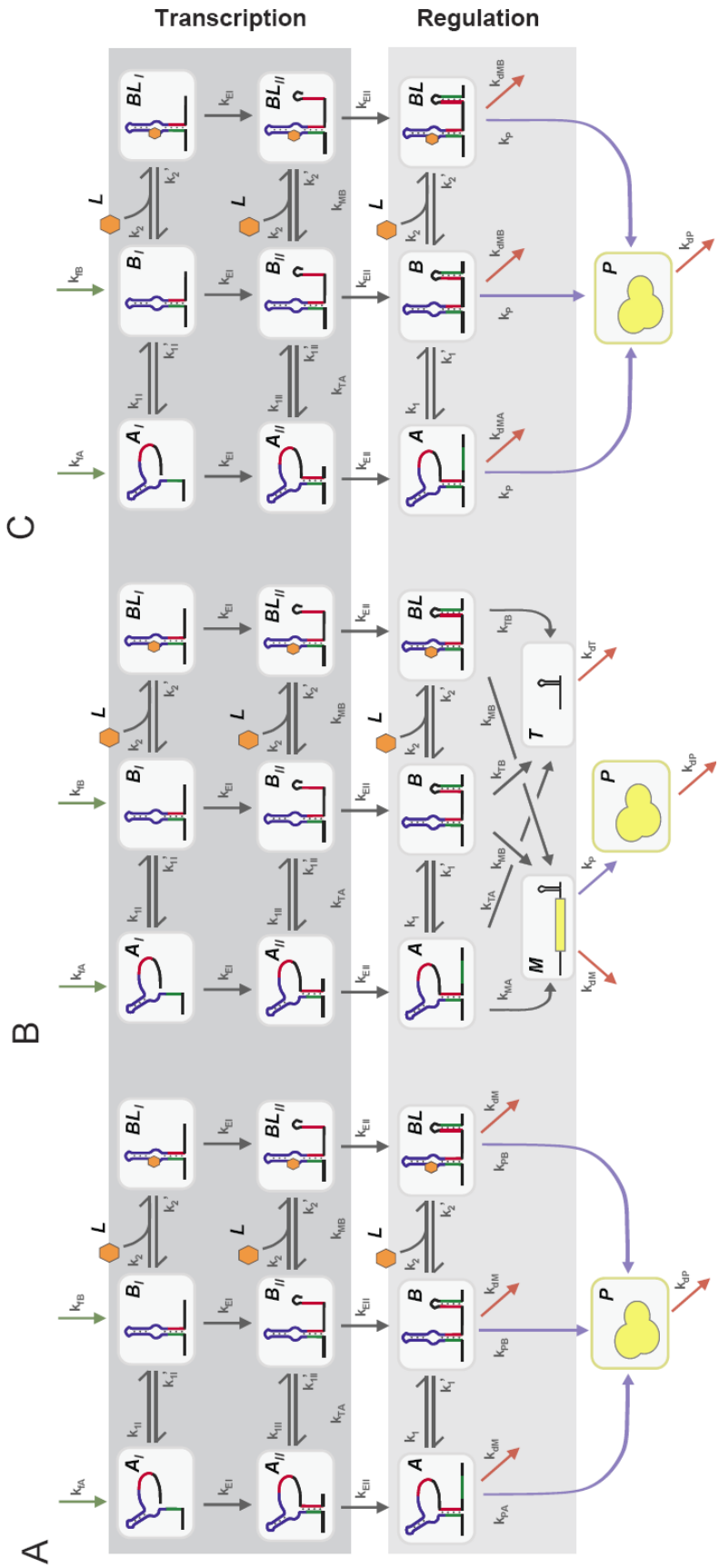
Transcription
Regulation
A
B
C

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.

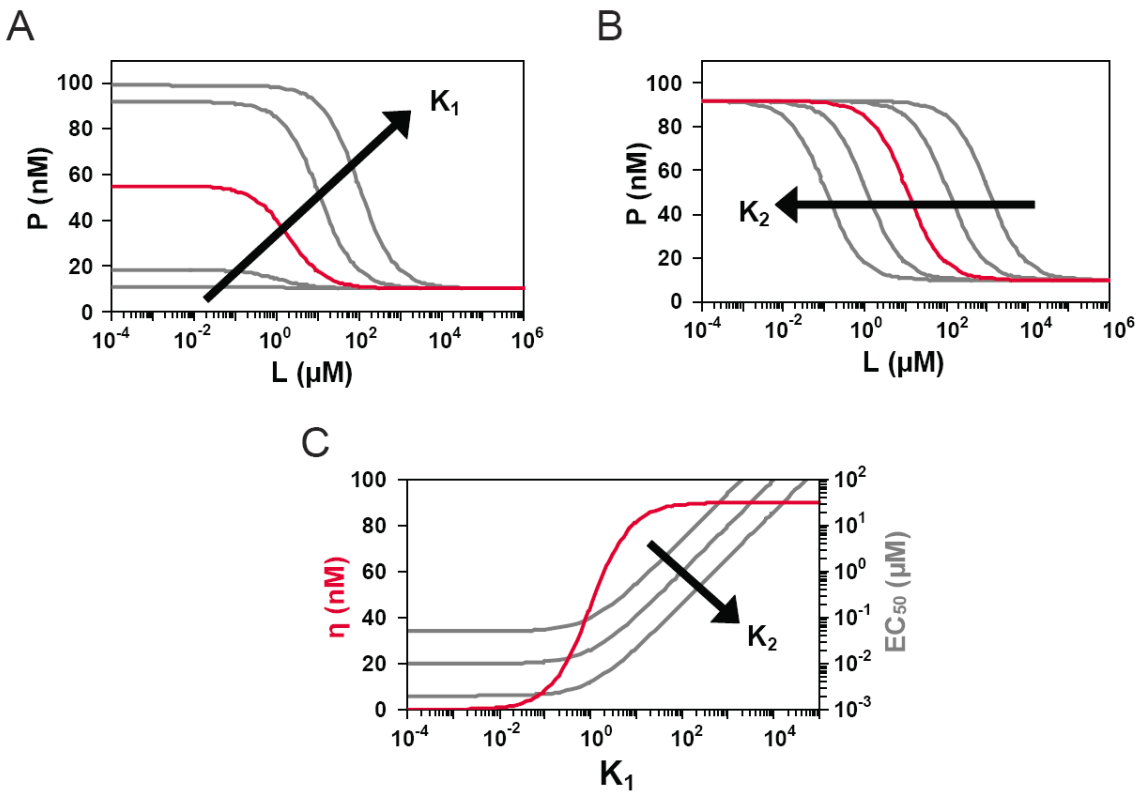

Figure S5.2 Thermodynamically-driven riboswitches exhibiting OFF behavior display similar tuning properties to riboswitches exhibiting ON behavior. K_1 is the conformational partitioning constant (k_1'/k_1) and K_2 is the aptamer association constant (k_2/k_2'). (**A**) K_1 affects both basal levels 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.

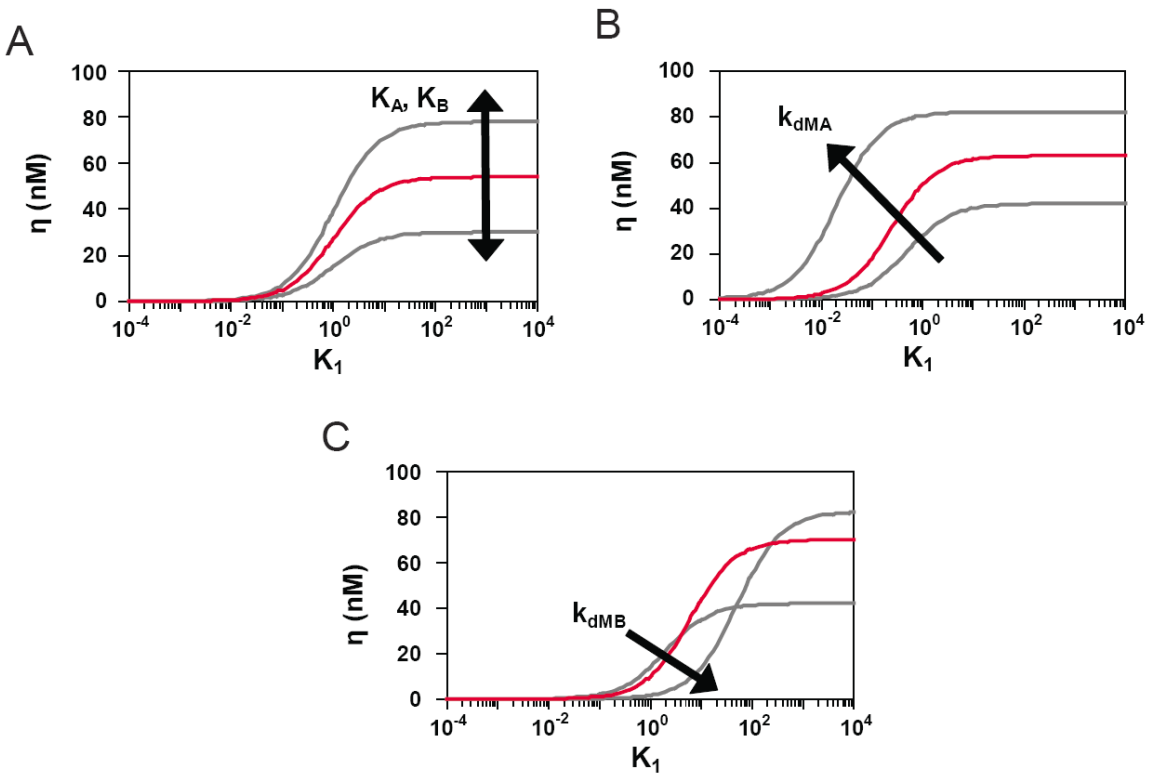

Figure S5.3 The mechanism-specific regulatory activities dictate differential tuning properties for thermodynamically-driven riboswitches. (**A**) Modulation of K_A and K_B affects the maximum dynamic range (η) for ON and OFF behaviors. K_A and K_B are modulated for riboswitches functioning through translational 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.

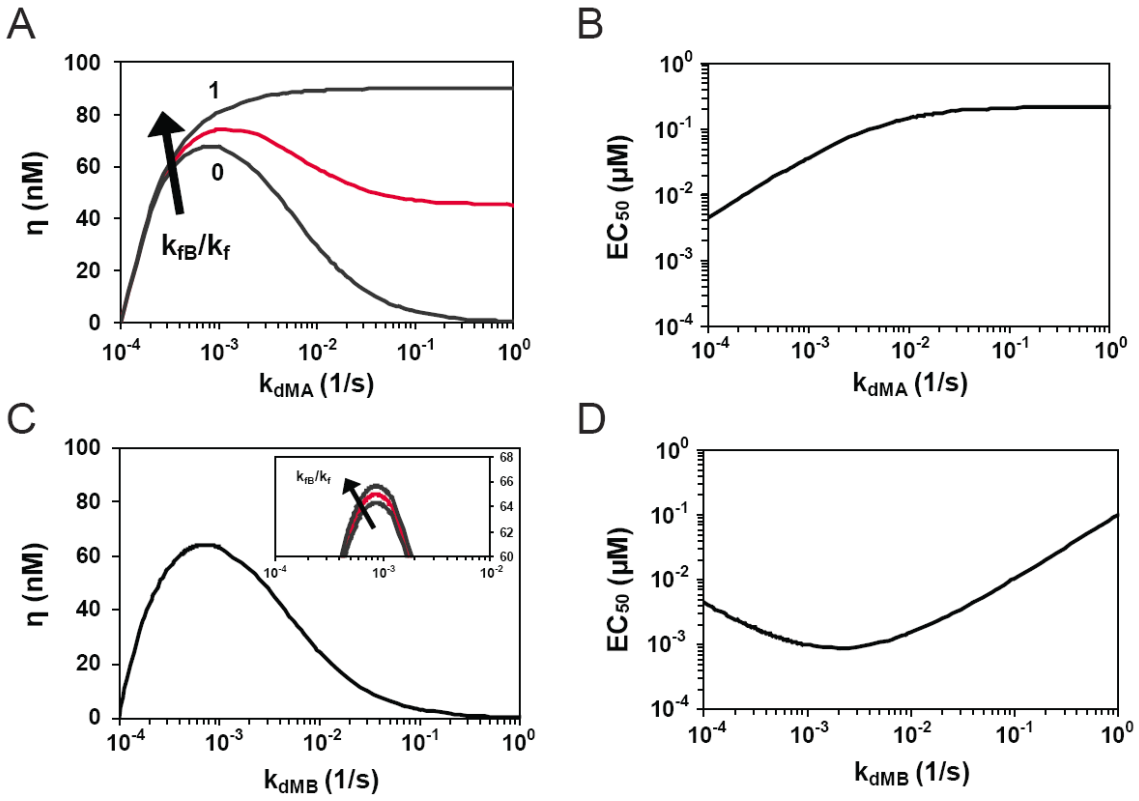

Figure S5.4 Distinction between tuning properties for ON and OFF behaviors for riboswitches functioning through mRNA destabilization. Dynamic range (η; **A**,**C**) and EC_{50} (**B**,**D**) display different dependencies on the dominant mRNA degradation rate constant for ON (k_{dMA}; **A**,**B**) and 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.

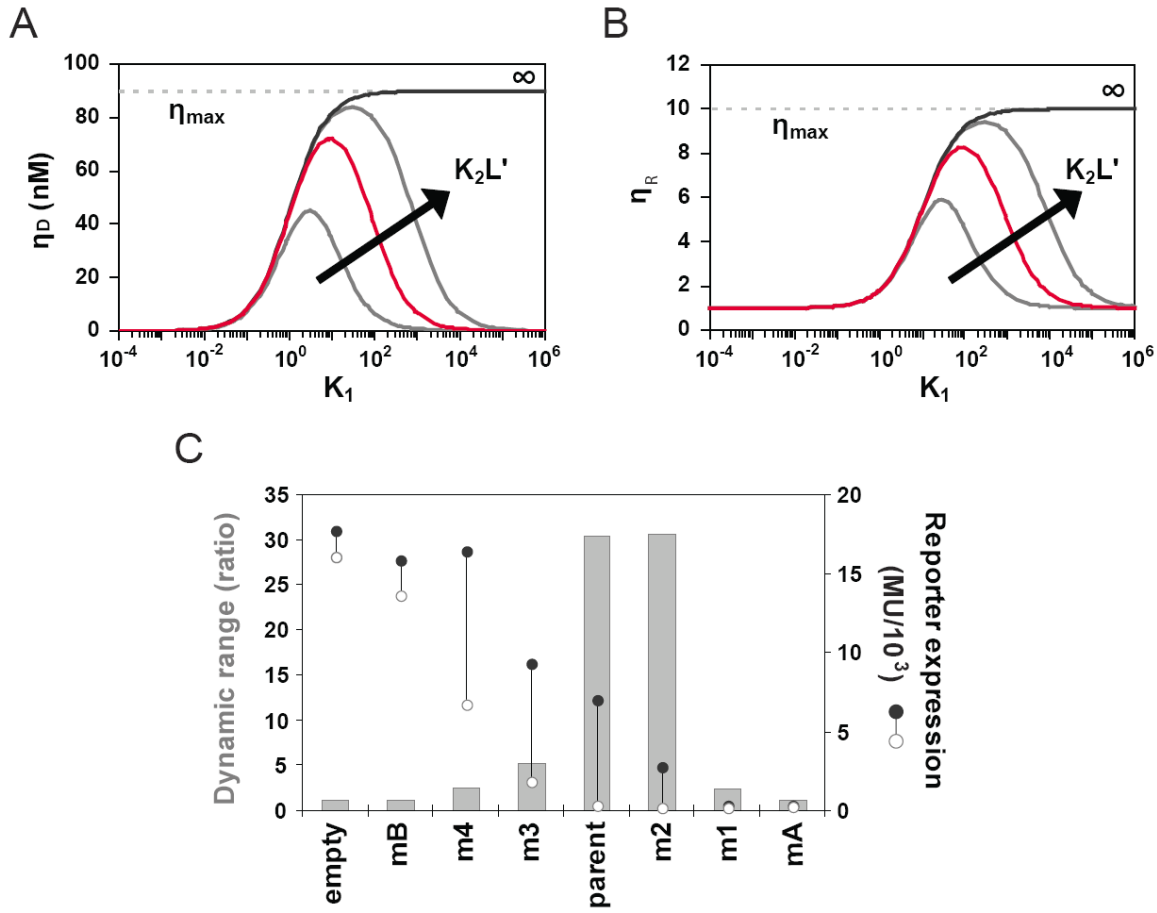

Figure S5.5 The dynamic range difference and ratio exhibit qualitatively similar tuning properties. Model predictions for the dynamic range difference (η_D, **A**) and ratio (η_R, **B**) when subjected to a ligand concentration upper limit (L'). In the absence of a ligand concentration upper limit, the 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.

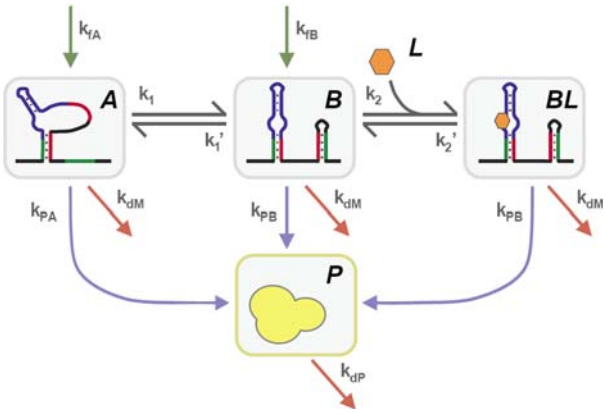

Figure 6.1B Molecular description for a riboswitch operating under translational repression. The riboswitch appears in either conformation A (k_{fA}) or B (k_{fB}) and reversibly switches between conformations (k_1, k_1') and undergoes irreversible degradation (k_{dM}) independent of conformation. Conformation 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}).

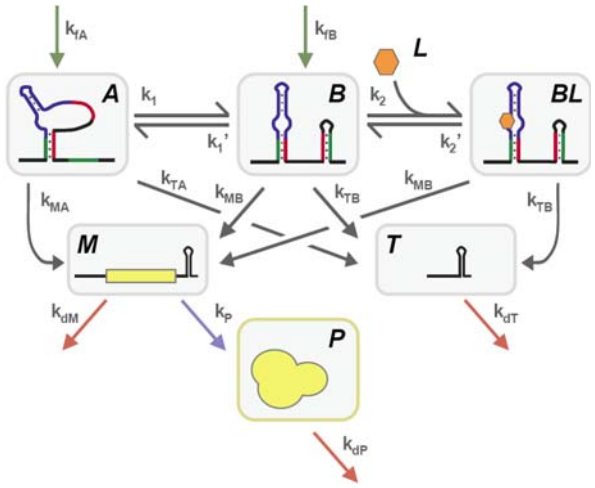

Figure 6.1C Molecular description for a riboswitch operating under transcriptional termination. The riboswitch appears in either conformation A (k_{fA}) or B (k_{fB}) as an intermediate in the transcriptional process and reversibly switches between conformations (k_1, k_1'). Conformation B reversibly 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}).

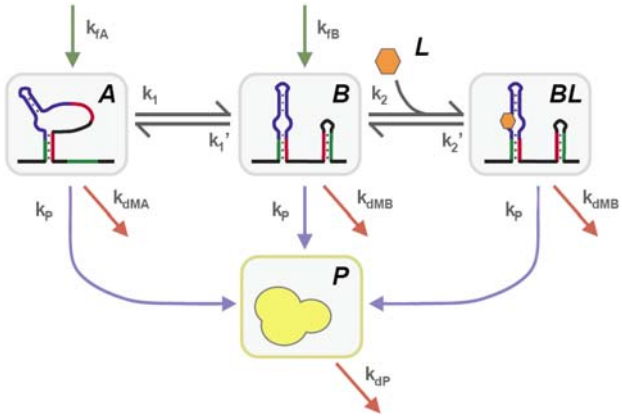

Figure 6.1D Molecular description for a riboswitch operating under mRNA destabilization. The riboswitch appears in either conformation A (k_{fA}) or B (k_{fB}) and reversibly switches between conformations (k_1, k_1'). Conformation B reversibly binds (k_2) and releases (k_2') the cognate ligand (L). 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.