Document S1: Supplemental Information and Modeling Procedures

Pre-optimized transcription-based amplifier design

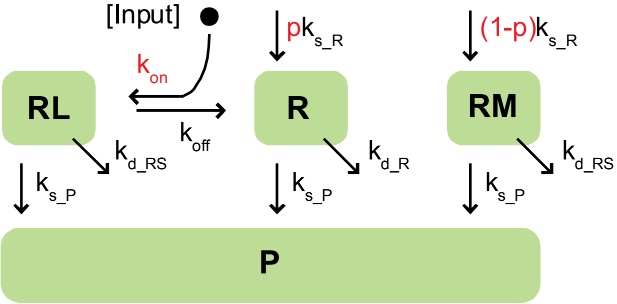
We chose synthetic transcription-based devices as our amplifier for their potential to exhibit a large dynamic range and amplification gain, while remaining orthogonal to our target small-molecule inputs. Transcription-based devices are typically implemented with a *trans* protein factor and *cis* DNA element. The *trans* elements are comprised of a DNA-binding domain (BD) fused to an activation domain (AD) and nuclear localization signal (NLS), whereas the *cis* hybrid promoter comprises operator (OP) sites for BD binding and a minimal promoter (miniP) for transcription initiation. Previous studies have demonstrated the use of various BDs to implement this *trans*-acting transcription initiation mechanism between protein and DNA, including bacterial DNA-binding proteins (Ajo-Franklin et al., 2007), zinc finger motifs (Khalil et al., 2012; Lohmueller et al., 2012), TALEs (Perez-Pinera et al., 2013), and recently CRISPR/dCas9 systems (Farzadfard et al., 2013; Gilbert et al., 2013).

We built a transcription-based amplifier based on this reported architecture using a bacterial DNA-binding protein LexA as the BD (Brent and Ptashne, 1985), a viral transcriptional activator VP16 as the AD (Gossen and Bujard, 1992), and a corresponding hybrid promoter by placing LexA-specific OP sites (LexAO) upstream of a minimal promoter from P_{GAL1} (Brent and Ptashne, 1985). The *trans* element was expressed from a constitutive promoter (P_{GPD}), and a reporter cassette to measure the transcription-based amplifier performance was designed by placing EGFP downstream of the hybrid promoter. Both expression cassettes harboring the *trans* and *cis* elements were placed on the same plasmid (pCS3434) and transformed into the yeast host (CSY3). We characterized the fluorescence expression levels of this pre-optimized transcription-based amplifier and observed no amplification activity (gain ~0.4 protein/protein) and a low-to-medium expression level (Figure S1B). Moreover, the results indicate a bimodal distribution of the EGFP expression among cells; however, both populations of cells are above the autofluorescence level, suggesting that the transcription-based amplifier is functional but exhibits large variability.

Modeling RNA switch-controlled gene regulation

Step 1: Obtaining a general model

To enable a more quantitative understanding of RNA switches, we expanded a reported kinetic model for RNA switches to predict the protein expression level (McKeague et al., 2015). The modified model contains four internal states (**R**, **RL**, **RM** as distinct species of mRNAs; **P** for protein), and one external state ([Input] as a small molecule signal; Supplemental Data Figure 1). **R** represents the state of mRNA with its RNA switch in the cleavable conformation (i.e., no input binding), while **RL** represents the state of mRNA with its RNA switch in the input-bound, non-cleavable conformation. **RM** represents the state of mRNA that has a misfolded RNA switch during transcription, which is not responsive to input induction nor cleavable. **P** represents the level of the protein (encoded by the mRNA), and [Input] represents the small molecule concentration fed into the system. Eight kinetic parameters were used in the model to describe the RNA switch-associated gene expression: the rate constants of synthesis (k_{s_R}, and k_{s_P}), the rate constants of degradation (k_{d_R}, k_{d_RS}, and k_{d_P}), mRNA folding partition coefficient (p), and binding kinetic constants (k_{on} and k_{off}) between the fed input and the sensor of the RNA switch (i.e., aptamer).



Supplemental Data Figure 1. A modified kinetic model of RNA switches with four internal states, one external state, and eight parameters.

Four assumptions were made (without losing the ability to capture the system dynamics) in this kinetic model to simplify the differential equations such that the system becomes computationally solvable. First, we assumed that the binding affinity of the input small molecule to the RNA switch on the mRNA is the same as its binding affinity to the aptamer (K_D) measured at physiologically-relevant Mg^{2+} concentration *in vitro* (McKeague et al., 2015). Second, we assumed that the different species of mRNA (**R**, **RL**, and **RM**) have identical chances of being translated into proteins before their degradation, which is represented in the model with the same protein synthesis rate constant k_{d_P}. Third, the cleaved mRNA degradation rate constant (k_{d_R}) captures two independent processes: mRNA cleavage and decapping-associated 5' → 3' degradation. Since the decapping-associated degradation rate has been reported to be significantly faster than the RNA cleavage rate (Cao and Parker, 2001), these two processes can be simplified into one single rate constant. Last, we assumed that once the mRNA is transcribed into either conformation (**R** or **RM**), conformation interchanges between the two states are unlikely to happen (i.e., the existence of kinetic trap during mRNA transcription and folding) (Russell et al., 2006; Treiber and Williamson, 1999).

The dynamics between these four states can be expressed as a set of differential equations as follows:

$$\frac{d[R]}{dt} = -k_{d\_R}[R] - k_{on}[L][R] + k_{off}[RL] + pk_{s\_R}$$

$$\frac{d[RL]}{dt} = k_{on}[L][R] - k_{d\_RS}[RL] - k_{off}[RL]$$

$$\frac{d[RM]}{dt} = -k_{d\_RS}[RM] + (1-p)k_{s\_R}$$

$$\frac{d[P]}{dt} = k_{s\_P}[R] + k_{s\_P}[RL] + k_{s\_P}[RM] - k_{d\_P}[P]$$

We can re-write this set of differential equations of a linear system as the matrix form:

$$\frac{d}{dt}\begin{bmatrix} R \\ RL \\ RM \\ P \end{bmatrix} = \underbrace{\begin{bmatrix} -k_{d\_R} - k_{on}[Input] & k_{off} & 0 & 0 \\ k_{on}[Input] & -k_{d\_RS} - k_{off} & 0 & 0 \\ 0 & 0 & -k_{d\_RS} & 0 \\ k_{s\_P} & k_{s\_P} & k_{s\_P} & -k_{d\_P} \end{bmatrix}}_{F} \begin{bmatrix} R \\ RL \\ RM \\ P \end{bmatrix} + \underbrace{\begin{bmatrix} pk_{s\_R} \\ 0 \\ (1-p)k_{s\_R} \\ 0 \end{bmatrix}}_{G}$$

If F is invertible, then we can solve the linear system and obtain the time-dependent analytical solution:

$$\begin{bmatrix} R(t) \\ RL(t) \\ RM(t) \\ P(t) \end{bmatrix} = F^{-1}(e^{Ft} - I)G$$

And since the eigenvalues of the matrix F are all negative in the range of our parameter space (see Supplemental Proof 1), we can obtain the steady-state solution by setting the time variable (t) to infinity:

$$\begin{bmatrix} R(\infty) \\ RL(\infty) \\ RM(\infty) \\ P(\infty) \end{bmatrix} = -F^{-1}G$$

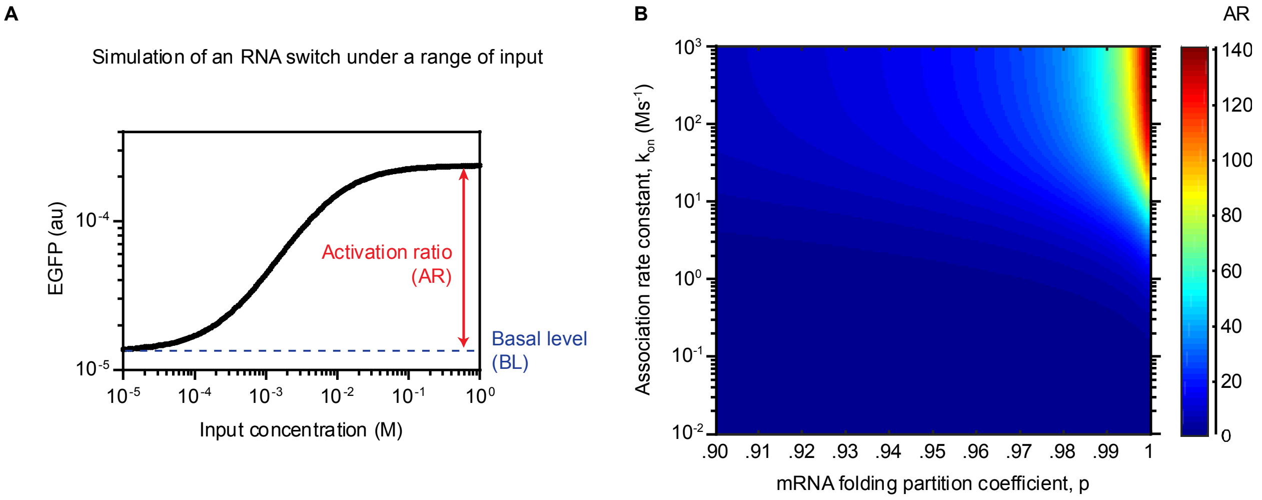
If the encoded gene is a fluorescent protein (e.g., EGFP), we can express the expected fluorescence level by a linear scalar (denoted as r) multiplying the expected protein levels. Previous studies have shown that the cellular fluorescent protein concentration is proportional to the measured fluorescence (Li et al., 2000). The equation then becomes:

$$\begin{bmatrix} R(\infty) \\ RL(\infty) \\ RM(\infty) \\ GFP(\infty)/r \end{bmatrix} = \begin{bmatrix} R(\infty) \\ RL(\infty) \\ RM(\infty) \\ P(\infty) \end{bmatrix} = -F^{-1}G$$

To verify that the model is capable of capturing RNA switching dynamics, we performed a simulation using this eight-parameter *in silico* model over five decades of input concentration (10^{-5} to 10^{0} M in log) with $K_D = 10^{-6}$ M, p = 0.95, and $k_{on} = 10$ Ms^{-1} (see the Supplemental Data Table 1 for all parameters). The simulation results demonstrated a typical RNA switch activity where the gene expression level increases as the input level increases (Supplemental Data Figure 2A). We then proceed to evaluate the effect of p (mRNA partition coefficient) and k_{on} (input-aptamer association rate constant) on the performance of an RNA switch (i.e., activation ratio). We simulated the model across five decades of k_{on} (10^{-2} to 10^{3} Ms^{-1} in log) and a linear range of p (0.9 to 1). The simulation results indicate that the activation ratio can be determined by potentially varying both p and k_{on} (Supplemental Data Figure 2B). Moreover, high activation ratios were observed to be in the top-right corner of the heat map plot, suggesting that a large k_{on} (strong binding between the input signal and aptamer) and a large p (correct folding) are critical for desirable RNA performance.

Supplemental Data Table 1. Summary of parameters used in the RNA switch model.

Name	Description	Value	Reference
k_{s_R}	Synthesis rate of mRNA	10^{-8} M/s	(Beisel et al., 2008)
k_{d_RS}	Degradation rate constant of stable mRNA	$4.2\text{x}10^{-4}$/s	(Cao and Parker, 2001)
k_{d_R}	Degradation rate constant of unstable mRNA	$6.2\text{x}10^{-2}$/s (Determined by sTRSV)	This work
k_{s_P}	Synthesis rate constant of protein	10^{-2}/s	(Beisel et al., 2008)
k_{d_P}	Degradation rate constant of protein	10^{-3}/s	(Beisel et al., 2008)
r	Linear ratio between predicted protein level and measured EGFP fluorescence	$3.2\text{x}10^{-5}$ (Determined by sTRSV inactive)	This work
K_D	Binding affinity between aptamer and target small molecule	theophylline (3.38 μM) tetracycline (30 nM) neomycin (137 nM) (*6R*)-FA (61 nM)	This work and (McKeague et al., 2015)
k_{on}	Association rate constant between small molecule and aptamer	Determined by *in vivo* data	This work
k_{off}	Dissociation rate constant between small molecule and aptamer	Determined by $K_D \times k_{on}$	This work
p	mRNA folding partition coefficient	Determined by *in vivo* data	This work



Supplemental Data Figure 2. (A) Simulation of a typical RNA switch activity. (B) Simulated activation ratio (AR) based on a range of k_{on} and p.

Supplemental Proof 1.

To prove that the steady-state solution exists, we need to show that F has negative eigenvalues. We first write down the derivation formula to derive the eigenvalues (λ's) of the matrix F. The determinant of the characteristic polynomial matrix is set to zero to solve for eigenvalues.

$$F - I\lambda = 0 \Rightarrow \begin{vmatrix} -k_{d\_R} - k_{on}[Input] - \lambda & k_{off} & 0 & 0 \\ k_{on}[Input] & -k_{d\_RS} - k_{off} - \lambda & 0 & 0 \\ 0 & 0 & -k_{d\_RS} - \lambda & 0 \\ k_{s\_P} & k_{s\_P} & k_{s\_P} & -k_{d\_P} - \lambda \end{vmatrix} = 0$$

We can then simplify the characteristic polynomial matrix using matrix decomposition:

$$\Rightarrow (-k_{d\_P} - \lambda)(-k_{d\_RS} - \lambda) \begin{vmatrix} -k_{d\_R} - k_{on}[Input] - \lambda & k_{off} \\ k_{on}[Input] & -k_{d\_RS} - k_{off} - \lambda \end{vmatrix} = 0$$

It can be simply observed that the first two eigenvalues, - k_{d_P} and - k_{d_RS}, are both negative. This leaves us the other two eigenvalues, which can be obtained by solving the determinant of the 2-by-2 matrix.

$$\begin{vmatrix} -k_{d\_R} - k_{on}[Input] - \lambda & k_{off} \\ k_{on}[Input] & -k_{d\_RS} - k_{off} - \lambda \end{vmatrix} = 0$$

$$\Rightarrow (-k_{d\_R} - k_{on}[Input] - \lambda)(-k_{d\_RS} - k_{off} - \lambda) - k_{off}k_{on}[Input] = 0$$

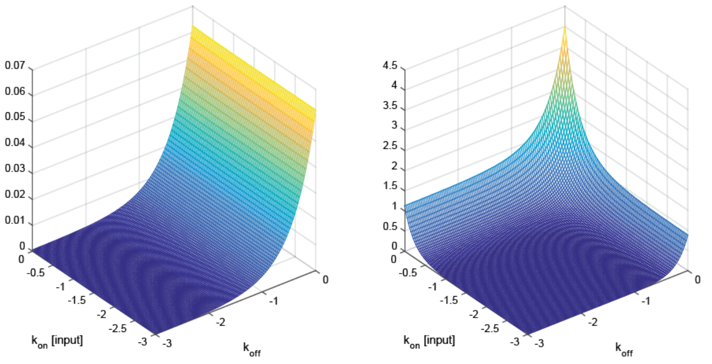
$$\Rightarrow \lambda^2 + \left(k_{d\_R} + k_{on}[Input] + k_{d\_RS} + k_{off}\right)\lambda + k_{d\_R}k_{d\_RS} + k_{off}k_{d\_R} + k_{on}[Input]k_{d\_RS} = 0$$

For this equation to have real negative solutions, the following criteria must hold (as a simple result of quadratic formula):

$$k_{d\_R}k_{d\_RS} + k_{off}k_{d\_R} + k_{on}[Input]k_{d\_RS} > 0 \tag{1}$$

$$\left(k_{on}[Input] + k_{off} + k_{d\_R} + k_{d\_RS}\right)^2 - 4\left(k_{d\_R}k_{d\_RS} + k_{off}k_{d\_R} + k_{on}[Input]k_{d\_RS}\right) > 0 \tag{2}$$

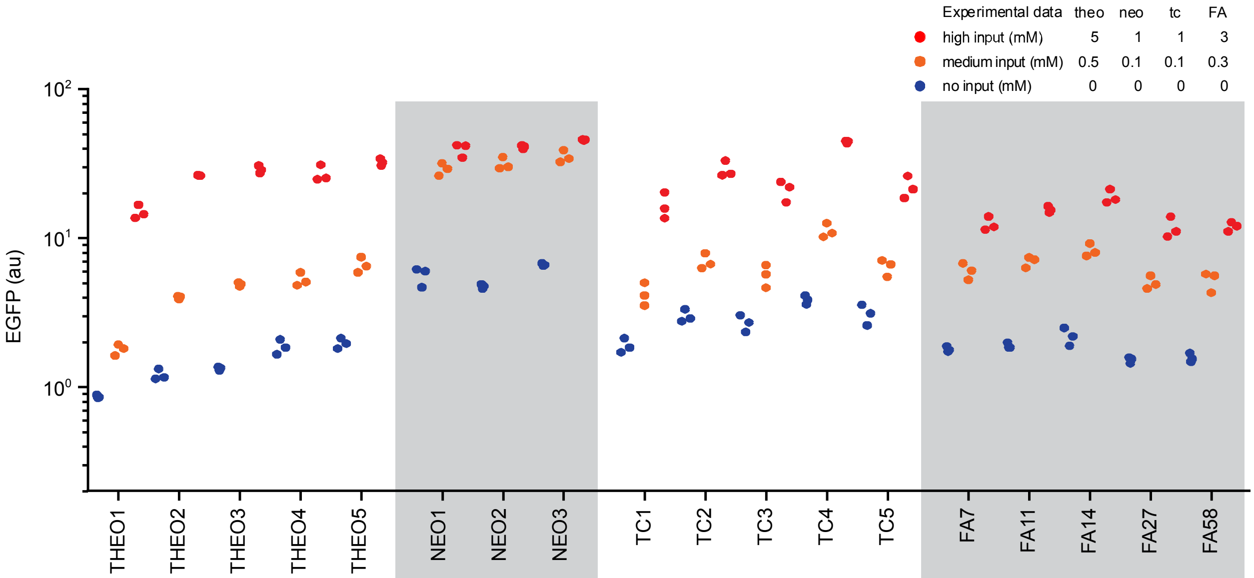
When the two variables, k_{on}[input] and k_{off}, are nonnegative (which is true in our model), the condition (1) and (2) will hold. The two equations are quadratic (or reduced quadratic) with a positive z-intercept (i.e., the function value at (0, 0)) and continuous, positive first-order partial derivatives (i.e., $\partial f/\partial(k_{on}[input]) \geq 0$ and $\partial f/\partial(k_{off}) \geq 0$; f representing either function). This means that for the functions values will continue to increase in both dimensions, with an initial value greater than zero. We plotted the values of both functions against the two variables to illustrate this result (Supplemental Data Figure 3). This shows that the last two eigenvalues are also real and non-positive. Therefore, we can conclude that the steady-state solution exists since F is a matrix with all real negative eigenvalues.



Supplemental Data Figure 3. Function values of equation (1) and (2) against variables k_{on}[input] and k_{off}.

Step 2: *In vivo* characterization

We proceeded to characterize RNA switches *in vivo*. We identified 18 different RNA switches: five theophylline-responsive (THEO1 to 5), three neomycin-responsive (NEO1 to 3), five tetracycline-responsive (TC1 to 5), five (*6R*)-FA-responsive switches (FA7, 11, 14, 27, 58), and two control ribozymes (sTRSV and sTRSV inactive) (McKeague et al., 2015; Townshend et al., 2015). We inserted each RNA switch into the 3' UTR of the EGFP transcript. Cells harboring each genetic construct were grown in closed-batch culture and back-diluted into three input levels, in three biological replicates. Their fluorescence is measured by flow cytometer after 6 hours of re-growth (See STAR Methods), and the results are shown in Supplemental Data Figure 4.



Statistics of experimental data

	THEO1	THEO2	THEO3	THEO4	THEO5	NEO1	NEO2	NEO3	TC1	TC2	TC3	TC4	TC5	FA7	FA11	FA14	FA27	FA58
AR_{med}	2.1	3.3	3.7	2.8	3.4	5.2	6.6	5.3	2.2	2.3	2.1	2.9	2.1	3.3	3.7	3.8	3.3	3.3
AR_{high}	17.3	21.8	21.7	14.5	16.5	7.0	8.6	6.9	8.7	9.6	7.8	11.5	7.1	6.9	8.2	8.6	7.7	7.6

Supplemental Data Figure 4. *In vivo* characterization of RNA switches. Each dot represents a single experiment replicate. The subscript of AR (i.e., “high” in AR_{high}) indicates the input concentration from which the AR is calculated.

Step 3: Parameter estimation

We used the measured in vivo data to estimate p and k_{on} for each RNA switch. The kinetic parameters in the model can be grouped into switch-specific and non-switch specific parameters. Switch-specific kinetic parameters are ones that contribute to performance of RNA switches when regulating gene expression, such as k_{on}, k_{off}, and p. Most of the non-switch specific parameters have been reported (Supplemental Data Table 1). The two remaining unknown parameters (r and k_{d_R}) can be determined through measuring the fluorescence levels with two non-input responsive controls: sTRSV (wild-type ribozyme) and sTRSV inactive (non-cleaving ribozyme variant). By placing sTRSV inactive in the 3' UTR of the mRNA (instead of RNA switches) to regulate protein expression, all mRNAs become non-cleavable (i.e., **RM** state in the model). The model can thus be reduced into two states (Supplemental Data Figure 5A) to determine the linear scalar between predicted protein level and the measured fluorescence (r).

$$\frac{d}{dt}\begin{bmatrix} R \\ RL \\ RM \\ P \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -k_{d\_RS} & 0 \\ k_{s\_P} & k_{s\_P} & k_{s\_P} & -k_{d\_P} \end{bmatrix} \begin{bmatrix} R \\ RL \\ RM \\ P \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ k_{s\_R} \\ 0 \end{bmatrix}$$

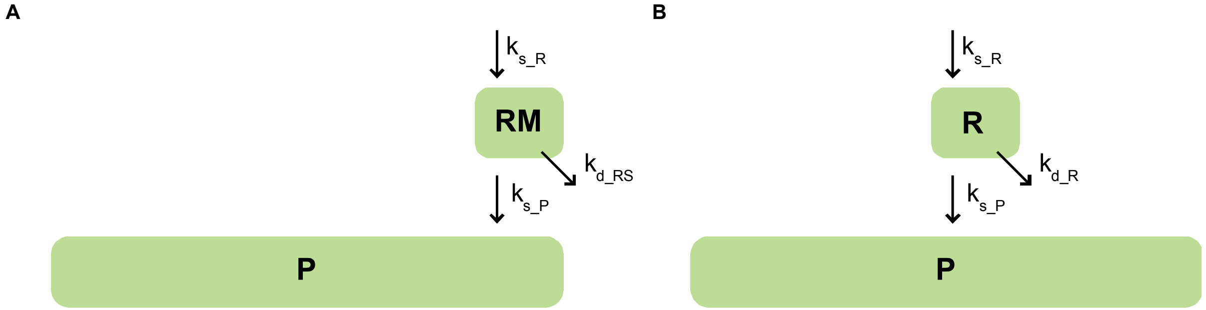
$$\Rightarrow \frac{d}{dt}\begin{bmatrix} RM \\ P \end{bmatrix} = \begin{bmatrix} -k_{d\_RS} & 0 \\ k_{s\_P} & -k_{d\_P} \end{bmatrix} \begin{bmatrix} RM \\ P \end{bmatrix} + \begin{bmatrix} k_{s\_R} \\ 0 \end{bmatrix}$$

The steady-state analytical solution is:

$$\begin{bmatrix} RM(\infty) \\ P(\infty) \end{bmatrix} = -inv\left(\begin{bmatrix} -k_{d\_RS} & 0 \\ k_{s\_P} & -k_{d\_P} \end{bmatrix}\right) \begin{bmatrix} k_{s\_R} \\ 0 \end{bmatrix}$$

$$= -inv\left(\begin{bmatrix} -4.2\times10^{-4} & 0 \\ 1\times10^{-2} & -1\times10^{-3} \end{bmatrix}\right) \begin{bmatrix} 1\times10^{-8} \\ 0 \end{bmatrix} = \begin{bmatrix} 2.38\times10^{-5} \\ 2.38\times10^{-4} \end{bmatrix}$$

The EGFP signal measured under the sTRSV inactive construct has a median value of 76.51 esu; therefore the linear scalar factor (r) between predicted protein concentration and measured fluorescence is 3.2 x 10^5.



Supplemental Data Figure 5. (A) A reduced model for sTRSV control (non-responsive, non-cleavable ribozyme variant). (B) A reduced model for sTRSV (non-responsive, cleavable ribozyme).

To determine k_{d_R}, sTRSV is placed in the 3' UTR of the mRNA, which leads to all mRNAs in **R** state ready for cleavage and degradation (Supplemental Data Figure 5B). We assumed that all sTRSV-regulated mRNAs are being cleaved; i.e., the non-cleaving (mis-folded) portion of the sTRSV-regulated mRNA is negligible, as demonstrated by previous studies (Shepotinovskaya and Uhlenbeck, 2008; Townshend et al., 2015). We can re-write the equations as:

$$\frac{d}{dt}\begin{bmatrix} R \\ P \end{bmatrix} = \begin{bmatrix} -k_{d\_R} & 0 \\ k_{s\_P} & -k_{d\_P} \end{bmatrix} \begin{bmatrix} R \\ P \end{bmatrix} + \begin{bmatrix} k_{s\_R} \\ 0 \end{bmatrix}$$

$$\Rightarrow \frac{d}{dt}\begin{bmatrix} R \\ P \end{bmatrix} = \begin{bmatrix} -k_{d\_R} & 0 \\ k_{s\_P} & -k_{d\_P} \end{bmatrix} \begin{bmatrix} R \\ P \end{bmatrix} + \begin{bmatrix} k_{s\_R} \\ 0 \end{bmatrix}$$

The steady-state analytical solution is:

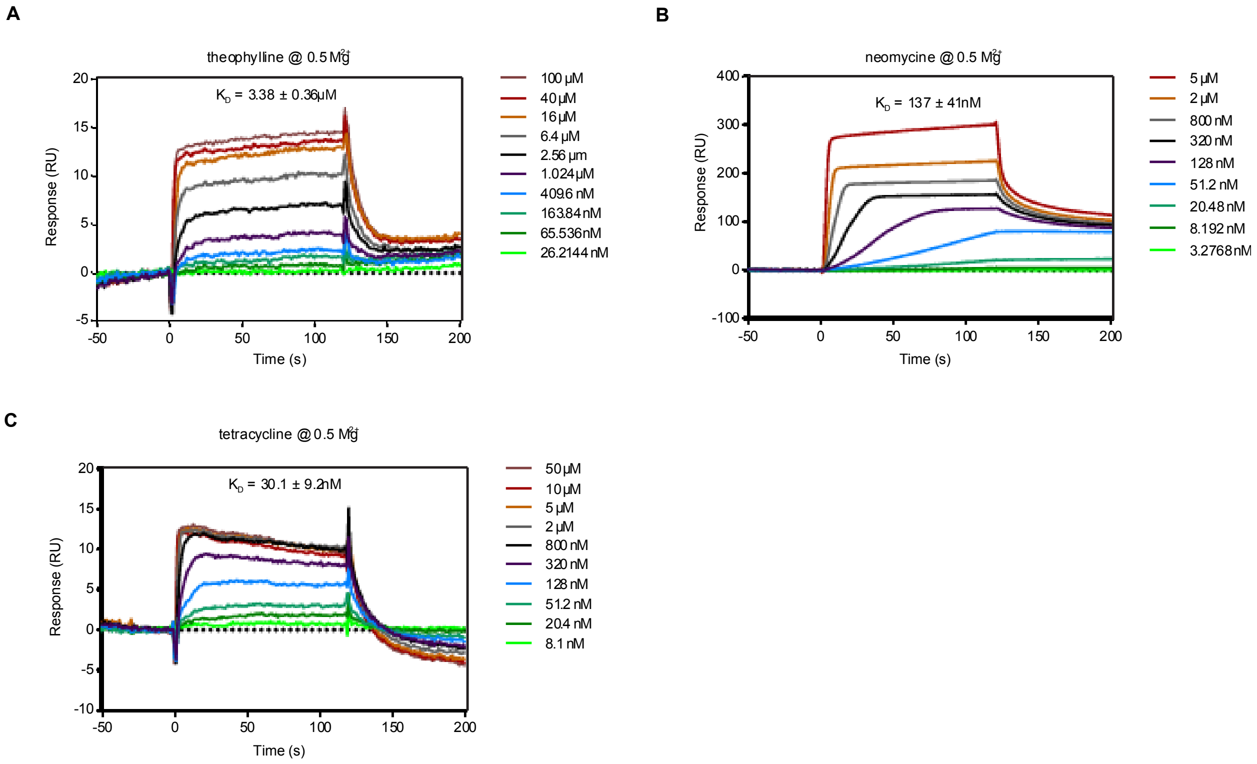
$$\begin{bmatrix} RM(\infty) \\ P(\infty) \end{bmatrix} = -inv\left(\begin{bmatrix} -k_{d\_R} & 0 \\ k_{s\_P} & -k_{d\_P} \end{bmatrix}\right) \begin{bmatrix} k_{s\_R} \\ 0 \end{bmatrix}$$

$$P(\infty) = \left(k_{d\_R}k_{d\_P}\right)^{-1} k_{s\_P}k_{s\_R}$$

$$\Rightarrow k_{d\_R} = k_{s\_P}k_{s\_R}/(k_{d\_P}P(\infty)) = k_{s\_P}k_{s\_R}/(k_{d\_P}EGFP(sTRSV))$$

And the mean value of EGFP measured under the sTRSV construct is 0.52 esu, resulting in the degradation value of $k_{d\_R} = 0.0618\ s^{-1}$.

The switch parameters (k_{on}, k_{off}, and p) in the model are different for each switch, and explain the basic characteristics of RNA switch dynamics. For example, if an RNA switch has a high partition coefficient (p), this suggests that more mRNAs are pre-folded into the non-cleavable **RM** state, resulting in a higher basal level of protein expression. We begin with determining k_{on} and k_{off} of each RNA switch. Following the assumption in the model, we know that the ratio of k_{on} and k_{off} is an aptamer-specific parameter (K_D) that can be explicitly determined through an *in vitro* SPR aptamer-binding assay (Supplemental Data Figure 6; see STAR Methods). Therefore, k_{on} and k_{off} together indeed only offer one degree of freedom in parameter fitting, which leaves the model with two to-be-determined RNA-switch-dependent parameters: p and k_{on}.



Supplemental Data Figure 6. *In vitro* SPR characterization of RNA aptamers at physiologically-relevant Mg^{2+} concentration. SPR sensorgrams are shown for the (A) theophylline aptamer (K_D = 3.38 μM), (B) neomycin aptamer (K_D = 137 nM), and (C) tetracycline aptamer (K_D = 30 nM). *In vitro* binding data between (6R)-folinic acid and its aptamer used to construct RNA switches, and the associated K_D (61 nM) were reported in (McKeague et al., 2015).

Our goal is to determine the two RNA-switch-dependent parameters through measuring the EGFP expression levels at different input concentrations. However, conventional parameter-fitting approaches (such as, linear regression) are either not feasible or precise enough to solve the proposed kinetic model due to its natural non-linear dynamic characteristics. In order to address this challenge without having to further simplify our model, we adopted a convex optimization approach to solve this parameter-fitting question. Convex optimization has been demonstrated as a powerful and systematic approach to solve problems with properly defined system characteristics and obtain statistical inferences based on measured data (Boyd and Vandenberghe, 2004). In the following paragraphs, we will explain how to transform this parameter-fitting question into a standard convex optimization form:

$$\text{Minimize } (Ax - b) \text{ under constraints: } Cx + d = 0$$

We begin with solving the protein level at steady state, which can be expressed as:

$$\frac{d[P]}{dt} = k_{s\_P}[R] + k_{s\_P}[RL] + k_{s\_P}[RM] - k_{d\_P}[P] = 0$$

$$\Rightarrow [P]_L = \frac{k_{s\_P}}{k_{d\_P}}([R]_L + [RL]_L + [RM]_L)$$

$$\Rightarrow [EGFP]_L = r[P]_L = \begin{bmatrix} r\frac{k_{s\_P}}{k_{d\_P}}, & r\frac{k_{s\_P}}{k_{d\_P}}, & r\frac{k_{s\_P}}{k_{d\_P}} \end{bmatrix} \begin{bmatrix} R_L \\ RL_L \\ RM_L \end{bmatrix}$$

This equation establishes the relationship between the measurable EGFP with the three mRNA states in the equation (**R**, **RL**, and **RM**), which can serve as the to-be-minimized objective function for the convex optimization program (as follows). In other words, the optimal values of the three mRNA states that minimize this objective function best describe the kinetic system model.

$$objective\colon min\left(\underbrace{\begin{bmatrix} r\frac{k_{s\_P}}{k_{d\_P}}, & r\frac{k_{s\_P}}{k_{d\_P}}, & r\frac{k_{s\_P}}{k_{d\_P}} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} R_L \\ RL_L \\ RM_L \end{bmatrix}}_{x} - \underbrace{EGFP_L}_{b}\right)$$

We then derive the convex optimization constraints by eliminating the state **P** in the differential equation:

$$\frac{d}{dt}\begin{bmatrix} R_L \\ RL_L \\ RM_L \end{bmatrix} = \begin{bmatrix} -k_{d\_R} - k_{on}[Input] & k_{off} & 0 \\ k_{on}[Input] & -k_{d\_RS} - k_{off} & 0 \\ 0 & 0 & -k_{d\_RS} \end{bmatrix} \begin{bmatrix} R_L \\ RL_L \\ RM_L \end{bmatrix} + \begin{bmatrix} pk_{s\_R} \\ 0 \\ (1-p)k_{s\_R} \end{bmatrix}$$

At steady state, the equation is set to 0 and can be re-written as:

$$\underbrace{\begin{bmatrix} -k_{d\_R} - k_{on}[Input] & k_{off} & 0 \\ k_{on}[Input] & -k_{d\_RS} - k_{off} & 0 \\ 0 & 0 & -k_{d\_RS} \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} R_L \\ RL_L \\ RM_L \end{bmatrix}}_{x} + \underbrace{\begin{bmatrix} pk_{s\_R} \\ 0 \\ (1-p)k_{s\_R} \end{bmatrix}}_{d} = 0$$

This set of equalities should hold for all mRNA states at different input concentrations, which can serve as the constraints for our convex optimization solver. Our goal is to determine the two RNA-switch-dependent parameters (p and k_{on}), under which the constraints are satisfied (i.e., feasible solutions) as well as minimize the objective function. However, there is a quasi-convex term in the constraints (that is, the variable k_{on} is multiplied with mRNA state **R** in the first column of the matrix). Therefore, we implemented an iterative convex feasibility algorithm to solve this quasi-convex problem. This algorithm uses the convex optimization package CVX (Grant and Boyd, 2013) as the core solver. In addition, we chose L1-norm to minimize the data, which has been shown to be robust and less sensitive to outliners or noise (Boyd and Vandenberghe, 2004). For each RNA switch, the pair of p and k_{on} can be found in Supplemental Data Table 2. Since the L1-norm optimization does not generate a "goodness-of-fit" number to show the accuracy of the estimated parameters, we reported the value of our minimization argument (i.e., distance score; with ranges from 300-500 for random pairs of p and k_{on}) as an indicator. Another common approach to validate the estimated parameters is to plug them back into the model and apply Monte Carlo simulation to test if the predicted metrics (i.e., ARs) match the original measurements. The simulation results suggested that the pair of p and k_{on} is capable of capturing the RNA switching dynamics for all identified RNA switches (last two columns in Supplemental Data Table 2).

Supplemental Data Table 2. Estimated RNA switch-specific parameters (p and k_{on}) and optimization distance scores. Monte Carlo simulations (N = 5000) were also used to calculate predicted ARs to validate the accuracy of these estimated parameters. The subscript of AR (i.e., "high" in AR_{high}) indicates the input concentration from which the AR is calculated. *Distance score represents the function value of the argument (Ax - b) used in L1-norm minimization, and has a range of 300-500 for random unoptimal pairs of p and k_{on}.

Name	p	k_{on}	Distance score*	Predicted AR_{med}	Predicted AR_{high}
THEO1	1.000	1.64	5.21	2.9	18.2
THEO2	0.996	3.78	2.58	3.7	22.3
THEO3	0.994	4.34	4.94	3.6	20.9
THEO4	0.987	3.46	7.78	2.3	11.8
THEO5	0.985	5.18	5.71	2.8	14.1
NEO1	0.932	195.7	15.2	4.0	2.5
NEO2	0.948	198.9	9.06	4.9	7.2
NEO3	0.923	242.3	7.67	4.0	5.8
TC1	0.987	8.62	9.21	1.7	7.1
TC2	0.973	18.43	9.33	1.8	7.5
TC3	0.975	13.30	9.52	1.6	6.4
TC4	0.960	49.21	5.84	2.5	10.0
TC5	0.967	12.42	11.31	1.5	5.0
FA7	0.988	9.23	5.96	3.3	5.4
FA11	0.987	13.20	8.25	4.0	6.7
FA14	0.986	16.42	12.4	4.5	7.5
FA27	0.992	8.29	8.08	3.6	6.0
FA58	0.991	9.20	7.12	3.7	6.2

Modeling transcription-based amplifiers

Step 1: Obtaining a general model

The dynamics of transcription-based amplifiers have been well studied and the apparent amplification (i.e., the ratio of expression between “with amplifier” and “without amplifier”) can be modeled as a Hill equation (Alon, 2007). We chose to model the amplifier using a standard four-parameter function with: max (the saturation response level, i.e., ON state), min (the leakage response level, i.e., OFF state), n (cooperativity), and EL_{50} (level of that gives half-maximal response, i.e., the inflection point). The equation is shown as follows:

$$y = min + \frac{\max - min}{1 + 10^{n(logEL_{50} - logx)}}$$

in which y represents the expression level with the amplifier, and x represent the expression level without the amplifier.

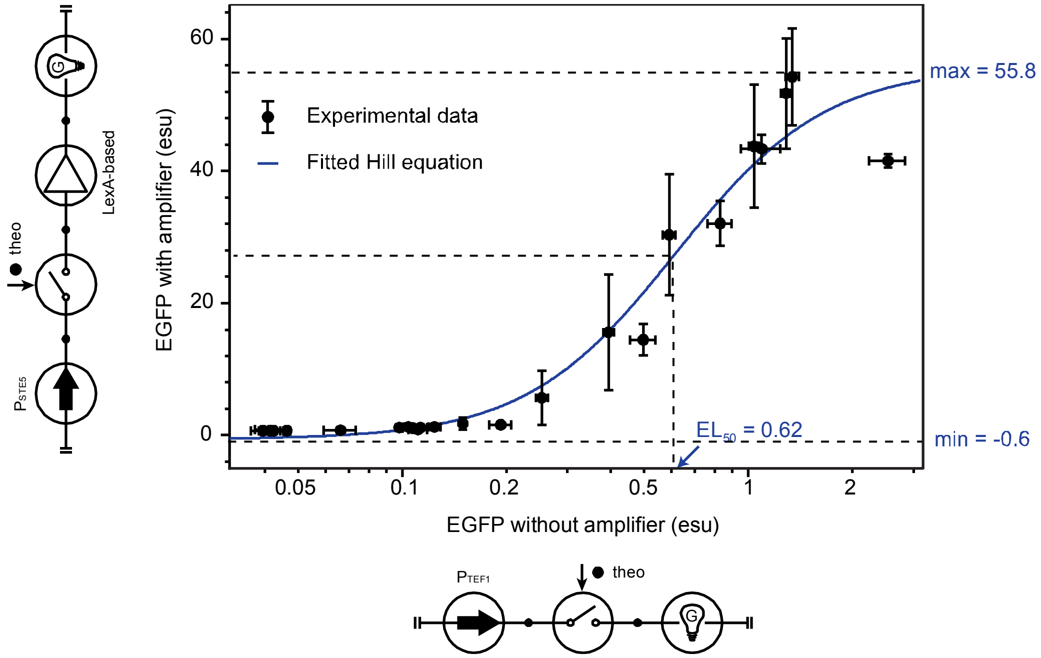
Step 2: *In vivo* characterization

We chose to characterize a LexA-based amplifier (CSY1101 + pCS3442; see the circuit symbol next to the y-axis in Supplemental Data Figure 7), and used two RNA switches (THEO1 and THEO5) in combination with a fine titration of input to generate the EGFP expression at different levels. Ideally the “without amplifier” constructs should have been built correspondingly by using P_{STE5} to regulated EGFP expression with RNA switches; however, the EGFP expression of these constructs was extremely low and noisy, and we were not able to obtain measurement with good quality (data not shown). Instead, we built the “without amplifier” constructs using a strong constitutive promoter (P_{TEF1}) with the same two RNA switches, and then measured the EGFP expression of these constructs before converting into the P_{STE5}-equivalent expression level ($P_{STE5} : P_{TEF1}$ = 1 : 40; see the circuit symbol next to x-axis in Supplemental Data Figure 7).

The experiment was performed in closed-batch culture, by first growing the device-harboring strains overnight, and back-diluting (1:50) into fresh media under different input levels (0, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5, and 10 mM of theophylline). The EGFP expression level was measured after 6 hours of growth using flow cytometry, and the results are presented in Supplemental Data Figure 7.

Step 3: Parameter estimation

We then fit the measured EGFP levels into the model described in Step 1 to estimate the unknown parameters. As shown in Supplemental Data Figure 7, the data exhibit strong fitness to the proposed model (R^2 = 0.93). The associated parameters we obtained for this transcription-based amplifier are: max = 55.8 (esu), min = -0.6 (esu), n = ~2, and EL_{50} = 0.62 (esu). The model was fitted using the default 4-parameter sigmoid fitting process in Prism (GraphPad). With this amplifier model, we can further generate similar models to describe other amplifiers (e.g., amplifiers with different ENs) by simply changing the values of max and min in the equation to the ON and OFF state values of the new transcription-based amplifier (values can be found in Figures 2E and 2F).



Supplemental Data Figure 7. Characterization of the LexA-based amplifier. Each dot represents mean values and error bar represent ± 1 s.d. of three biological replicates. The circuit symbols attached at both axes represent the constructs used to obtain the experimental data.

Modeling His5p-dependent cell growth

Step 1: Obtaining a general model

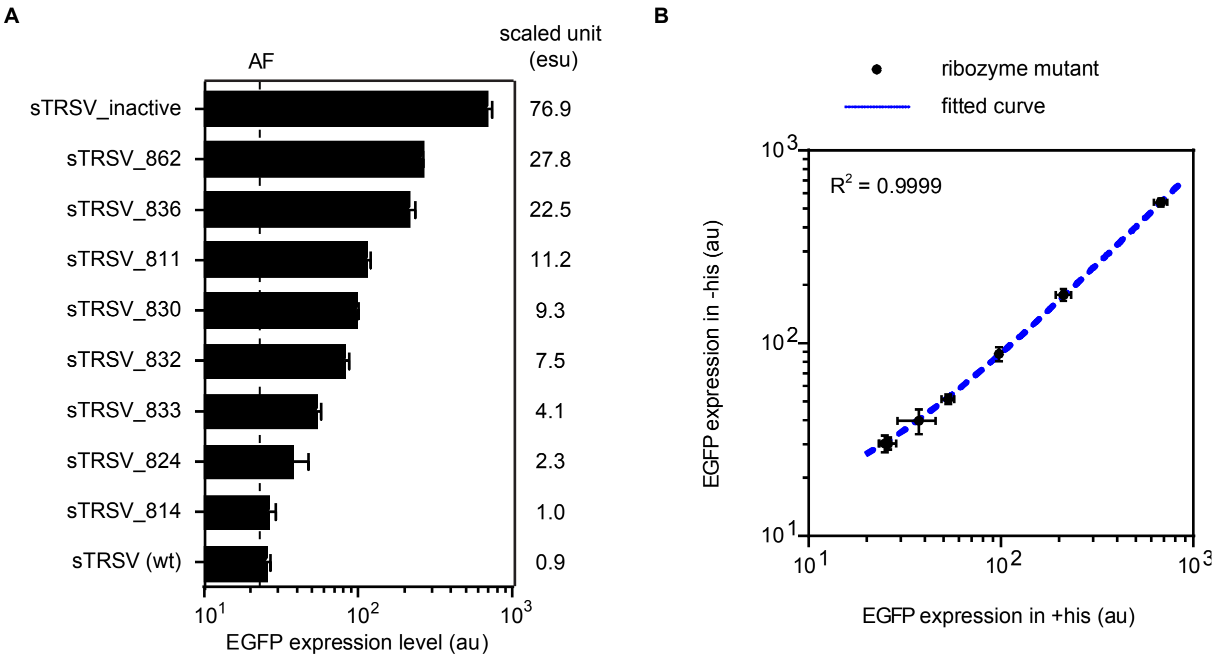
We also chose to use a standard four-parameters function to characterize the His5p-dependent growth (Boer et al., 2010): max (the saturation growth, i.e., wild-type yeast strain growth), min (the basal growth level, i.e., histidine-deficient limited growth), n (cooperativity), and EL_{50} (the His5p level of that gives half-maximal growth). The equation is shown as follows:

$$y = min + \frac{\max - min}{1+10^{n(logEL_{50}-logx)}}$$

in which y represents the with relative growth of yeast (evaluated by normalizing the viable cell counts to yeast with full histidine supply), and x represent the expression level of EGFP-His5p.

Step 2: *In vivo* characterization

We characterized His5p-dependent growth using a fusion EGFP-His5p to enable quantitative measurement. We used a reported mutant library of ribozymes (Townshend et al., 2015) to generate a range of EGFP-His5p protein expression (see Supplemental Data Figure 8 for the expression levels generated by these ribozymes). These ribozymes (see “RNA switches and ribozyme” in Supplemental Data 1) were placed at the 3' UTR of the EGFP-His5p (expressed from the promoter P_{TEF1}) on a plasmid and further transformed into yeast (CSY3). The device-harboring cells were grown overnight to stationary phase with histidine supply, and carefully back-diluted (1:50) into fresh media in the presence or absence of histidine. The cell counts (i.e., growth) and EGFP (i.e., His5p expression) were measured after growing for another 6 hours using flow cytometry. The relative growth (showing on the y-axis in the plot of Figure 4A) was obtained by dividing the viable cell counts of the cells with histidine to that without histidine. The results indicate a clear transition from reduced cell growth (~30%) to resumed growth (~100%) across a two-order of magnitude of His5p expression range (plot of Figure 4A).



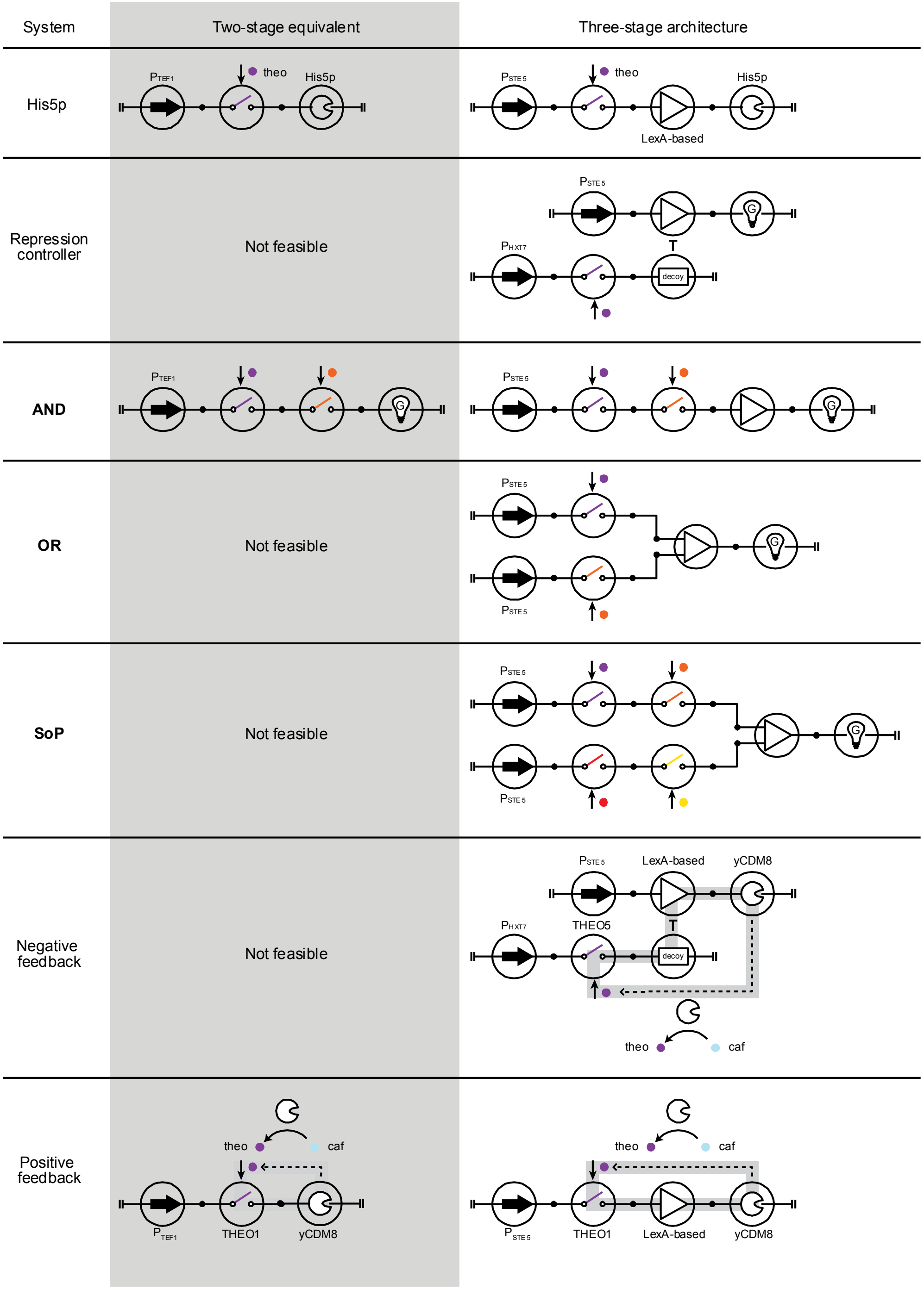
Supplemental Data Figure 8. (A) The EGFP expression level of each mutant ribozyme. Bar represents mean values and error bars represent ± 1 s.d. of three biological replicates. The vertical dash line represents the autofluorescence level ("AF"; measured from CSY3). (B) EGFP expression of each mutant ribozyme in + his (x-axis) and - his (y-axis) media conditions. Each dot represents the mean values and error bars represent ± 1 s.d. of three biological replicates. The equation of the fitted curve is $y = 0.7814x + 11.08$ ($R^2 = 0.9999$; fit with standard linear regression model).

Step 3: Parameter estimation

We then fitted the measured data into the described model. As shown in Figure 4A, the data exhibit strong fitness to the proposed model ($R^2 = 0.97$). The associated parameters were: max = 102% growth, min = 32% growth, $n = 1.8$, and $EL_{50} = 6.1$ (esu). The model was fitted using the default 4-parameter sigmoid fitting process in Prism (GraphPad). The fitted model together with the experimental data are plotted in Figure 4A.

Two-stage device controls

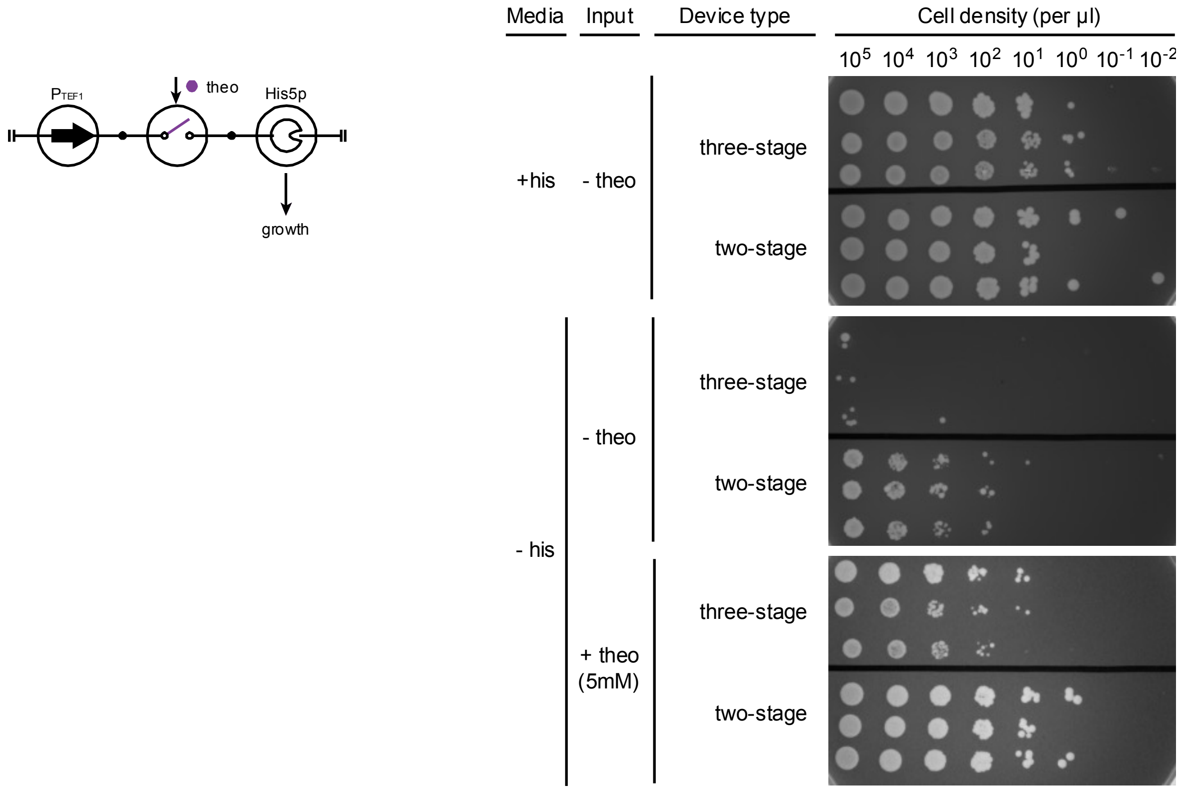
To further support the utility of our design framework, we constructed functionally equivalent devices using a two-stage architecture (i.e., RNA switches directly regulating the actuating stage) as controls. For the various applications demonstrated in this study, only three of seven controllers (His5p, **AND**, and positive feedback) could be implemented with the two-stage architecture (Supplemental Data Figure 9). Controllers with functional mechanisms implemented at the transcriptional level (specifically, repression and **OR**) are currently not feasible in the corresponding two-stage framework.



Supplemental Data Figure 9. Circuit symbols of functionally equivalent devices between two-stage and three-stage architecture. Three-stage controllers with functional mechanisms implemented at the transcriptional level (i.e., repression and **OR**) are not feasible in the corresponding two-stage framework.

1. Two-stage His5p controller

We began by demonstrating how devices with the two-stage architecture regulate the live-death phenotype. We constructed a two-stage version of the THEO1-based His5p-activator (circuit symbol in Supplemental Data Figure 10) by placing the THEO1 switch in the 3' UTR of the same phenotypic actuator His5p. Instead of using the original very weak controller-driving promoter (P_{STE5}), we chose a strong promoter (P_{TEF1}), which has ~40 times stronger expression level. We believe that if we implement the two-stage controller using P_{STE5}, the experimental result (i.e., difference in the absence/presence of the input) will be weak or barely detectable, leading to an unfair comparison between the two-stage and three-stage architectures.



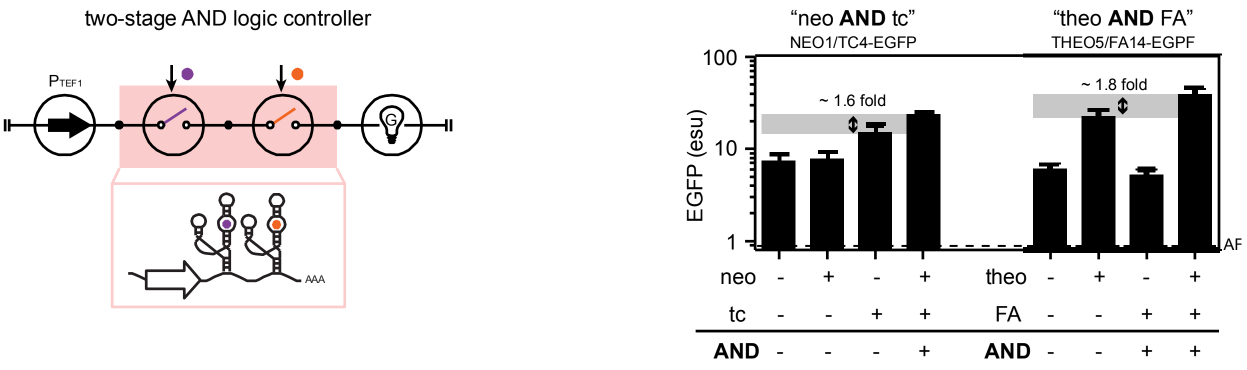
Supplemental Data Figure 10. Circuit symbol of the two-stage phenotypic controller using His5p (left). A serial dilution plating experiment was conducted to assess input-controlled cellular growth of both a two-stage and three-stage THEO1-based His5p-activator (right). The assay was conducted as described in Figure 4B in three biological triplicates.

The two-stage device expression cassette was placed in a low copy plasmid, and transformed into our yeast host. Cell harboring the two-stage (CSY3 + pCS3473) and three-stage His5p-activator (CSY1115 + pCS3442-THEO1) were characterized side-by-side using the serial dilution plating assay as described in Figure 4B. The same three media conditions were used for plating: +his/-theo, -his/-theo and -his/+theo, with the same cell density range (from 10^5 to 10^{-2} viable cells per μL) in three biological replicates. The results indicate that the three-stage His5p-activator exhibits a similar 4 orders of resumed growth range with no leakage growth as observed in Figures 4B and S3B (Supplemental Data Figure 10). In contrast, the two-stage His5p-activator was observed to exhibit a reduced input-inducible range (~1-2 orders of magnitude in cell density) and significantly higher basal growth in the absence of the input. Improvements to increase the performance of this two-stage device (e.g., decrease the basal leakage growth) can only be achieved by changing either the RNA switch or the controller-driving promoter. However, the RNA switch (THEO1) used in this controller is already one of the lowest basal level RNA switches that we have designed (with a basal expression level close to the wide type sTRSV) (Townshend et al., 2015). Moreover, changing the controller-driving promoter will result in a reduced ON-state (i.e., the resumed growth level), and potentially decrease the input-inducible range. These observed challenges associated with designing a two-stage RNA switch-based device support the limited dynamic ranges and lack of design freedom with this architecture.

2. Two-stage AND controller

We constructed two-stage **AND** controllers by moving the AND1 (NEO1/TC4) and AND2 (THEO5/FA14) switches to the 3' UTR of the EGFP output (circuit symbol in Supplemental Data Figure 11). We used P_{TEF1} as the controller-driving promoter for better visualization, and each two-stage **AND** cassette was placed on a low copy plasmid (pCS3474 and pCS3475, respectively). The experiment was conducted as described in Figure 5B in three biological triplicates.

We observed a ~4-fold reduction in Boolean ARs for both two-stage **AND** controllers (1.6 and 1.8; see Supplemental Data Figure 11) compared with the corresponding three-stage **AND**s (7.6 and 8.5; see Figure 5B). Also the two-stage **AND**s exhibit increased basal leakage expression (significantly higher than autofluorescence level), which is consistent with the experimental results observed with the two-stage His5p-activator. Despite the general reduction in device performance, the response of the two-stage **AND**s (specifically, the AND2 construct) under input combination became less ideal as **AND** logic. This observed input disparity can be a small molecule-dependent effect or the switching activity affected by the order of the RNA switches when placed in tandem. We believe the use of the transcription-based amplifier (i.e., three-stage architecture) may have helped shape this response disparity among inputs (i.e., a non-linear Hill response curve that can make any expression below the threshold lower), which improved the performance of the **AND** controller.



Supplemental Data Figure 11. Circuit symbol of the two-stage **AND** controller (left). Two two-stage **AND** controllers were characterized (right). Bars represent the mean values and error bars represent ± 1 s.d. of three biological replicates. The Boolean AR is indicated.

3. Two-stage positive feedback system

We designed a two-stage positive feedback system using the THEO1 switch to directly regulate yCDM8 (Figure S6C). We chose P_{TEF1} as the controller-driving promoter, and placed the two-stage PFB cassette on a low copy plasmid (pCS3476) before transforming the construct into our yeast host (CSY3). We characterized this two-stage positive feedback system in a chemostat for 400 hours (~ 2.4 weeks), side-by-side with its corresponding three-stage design (CSY1118 + pCS3442-THEO1; Figure 6C). The designed two-stage PFB system was observed to be non-functional; specifically, the ON state (~ 100 μM) failed to be maintained after the second induction phase (i.e., no observed bistability). In contrast, the three-stage PFB system exhibited bistability and was able to hold the induced ON state during the observed hold phase (~ 72 hours).

Glossary of Key Terms

Activation ratio (AR): the ratio of gene expression levels in the presence and absence of input signals, only defined for input-controlled devices or systems. The subscript of AR (e.g., “high” of AR_{high}) specifies the concentration of the input signal.

Actuating stage: a stage in a multi-stage genetic system that produces actuators to execute desired functions (e.g., cellular growth, production of a secondary metabolites). It is typically used as the last stage in a genetic control system (Wang et al., 2013).

Amplification gain (or “Gain”): the ratio of gene expression levels in the presence and absence of the amplifier (Figure S1A) and is a unitless measure.

Basal level (BL): the gene expression level of an input-controlled device in the absence of the input.

Boolean activation ratio (Boolean AR): the ratio of gene expression levels between the lowest level of the ON states and the highest level of the OFF states, which represents the worst-case AR of a Boolean controller. The ON and OFF states are defined according to the function of the Boolean controller.

Dynamic range (DR): the ratio of gene expression levels between ON (with *trans* element) and OFF (without *trans* elements) states of a transcription-based device (Figure S1A).

Fall time: the time of a system to decrease expression level from its ON state to mid-point level (i.e., half-DR level) upon input removal. (Note: the term is defined differently in this study than that in an electronic system).

Half dynamic range level (“half-DR level”): the midpoint (i.e., averaged) value between the ON and OFF states. The concept is similar to the idea of half-maximum level described in previous work (Rosenfeld et al., 2002) to evaluate system response times, but takes into account the expression level due basal leakage in a genetic system.

Linearity (in an amplifier): the ability of an amplifier to produce output signal that accurately reflects (i.e., proportional) its input. The linearity/non-linearity we showed in this study is an indirect measurement. Instead of directly displaying the input/output relationship of a genetic amplifier, we plot the relationship between “with” and “without” amplifiers. The linearity/non-linearity we show in this study is an indirect measurement, as a direct measurement of linearity would require determining intracellular input concentrations.

Percentage increase (or “incr”): the percentage increase of gene expression levels observed by inserting an epigenetic enhancer.

Processing stage: a stage in a multi-stage genetic system that transmits genetic signals between stages. The process can involve signal amplification (“amplifiers”), signal inversion (“inverters”), or signal buffering (“buffers”). It is typically used as the middle stage in a genetic control system (Wang et al., 2013).

Reduction ratio (RR): the ratio of gene expression levels in the absence and presence of input signals, which is essentially the multiplicative inverse of AR.

Rise time: the time of a system to increase expression level from its OFF state to mid-point level (i.e., half-DR level) upon input induction. (Note: the term is defined differently in this study than that in an electronic system).

Sensing stage: a stage in a multi-stage genetic system that is responsible for detecting the level of an input signal. It can also transduce the information encoded in the input signal to genetic signals (e.g., small molecule concentrations to mRNA levels). It is typically used as the first stage in a genetic controller (Wang et al., 2013).

Stage: a conceptual functional component in a genetic system, and is a result of decomposing the design goal into a series of genetic information operations (i.e., stages) to enable bottom-up assembly. A stage in a genetic system is commonly actualized with genetic device(s).

Stage level matching (or “level matching”): the process of matching the input and output levels between stages in a system (e.g., in terms of gene expression levels or enzymatic activities), such that the output level from the upstream stage is sufficient, but not excessive, to be used as an input to its downstream stage (modified from (Wang et al., 2013)). Using the three-stage EGFP-activator as an example, it is divided into three stages with respective input/output as: small molecule/translation activity level (i.e., mRNA half-life) in the first stage, translation activity level/transcription activity level (i.e., PoPs) in the second stage, and transcription activity level/protein actuator in the third stage. The stage level-matching process involves tuning the RNA-associated rates to obtain an optimal translation activity level (between the first and second stage), at which we can observe a controllable transcription activity level (third stage). Since the translation and transcription activity levels are not readily directly quantifiable with current experimental techniques, they are represented by the corresponding *cis*-regulated coding genes (i.e., the *trans* element of the transcription-based amplifier and EGFP, respectively). Therefore, this stage level-matching process can also be viewed as tuning the level of the *trans* element of the amplifier (through changing the synthesis, degradation, and ligand binding rates) to achieve an optimally controlled range of EGFP expression.

Supplemental Information References

Ajo-Franklin, C.M., Drubin, D.A., Eskin, J.A., Gee, E.P., Landgraf, D., Phillips, I., and Silver, P.A. (2007). Rational design of memory in eukaryotic cells. Genes & development *21*, 2271-2276.

Alon, U. (2007). An introduction to systems biology : design principles of biological circuits (Boca Raton, FL: Chapman & Hall/CRC).

Beisel, C., Bayer, T., Hoff, K., and Smolke, C. (2008). Model-guided design of ligand-regulated RNAi for programmable control of gene expression. Mol Syst Biol *4*, 224.

Boer, V.M., Crutchfield, C.A., Bradley, P.H., Botstein, D., and Rabinowitz, J.D. (2010). Growth-limiting intracellular metabolites in yeast growing under diverse nutrient limitations. Mol Biol Cell *21*, 198-211.

Boyd, S.P., and Vandenberghe, L. (2004). Convex optimization (Cambridge, UK ; New York: Cambridge University Press).

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729-736.

Cao, D., and Parker, R. (2001). Computational modeling of eukaryotic mRNA turnover. Rna *7*, 1192-1212.

Chang, A.L., McKeague, M., Liang, J.C., and Smolke, C.D. (2014a). Kinetic and equilibrium binding characterization of aptamers to small molecules using a label-free, sensitive, and scalable platform. Analytical chemistry *86*, 3273-3278.

Chang, A.L., McKeague, M., and Smolke, C.D. (2014b). Facile characterization of aptamer kinetic and equilibrium binding properties using surface plasmon resonance. Methods in enzymology *549*, 451-466.

Farzadfard, F., Perli, S.D., and Lu, T.K. (2013). Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas. Acs Synth Biol *2*, 604-613.

Feilotter, H.E., Hannon, G.J., Ruddell, C.J., and Beach, D. (1994). Construction of an improved host strain for two hybrid screening. Nucleic Acids Res *22*, 1502-1503.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods *6*, 343-345.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc *2*, 31-34.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A *89*, 5547-5551.

Grant, M., and Boyd, S. (2013). CVX: Matlab software for disciplined convex programming.

Gueldener, U., Heinisch, J., Koehler, G.J., Voss, D., and Hegemann, J.H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. Nucleic Acids Res *30*, e23.

Hecht, B., Muller, G., and Hillen, W. (1993). Noninducible Tet repressor mutations map from the operator binding motif to the C terminus. Journal of bacteriology *175*, 1206-1210.

Khalil, A.S., Lu, T.K., Bashor, C.J., Ramirez, C.L., Pyenson, N.C., Joung, J.K., and Collins, J.J. (2012). A synthetic biology framework for programming eukaryotic transcription functions. Cell *150*, 647-658.

Kisker, C., Hinrichs, W., Tovar, K., Hillen, W., and Saenger, W. (1995). The complex formed between Tet repressor and tetracycline-Mg2+ reveals mechanism of antibiotic resistance. Journal of molecular biology *247*, 260-280.

Li, J., Wang, S., VanDusen, W.J., Schultz, L.D., George, H.A., Herber, W.K., Chae, H.J., Bentley, W.E., and Rao, G. (2000). Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter. Biotechnol Bioeng *70*, 187-196.

Lohmueller, J.J., Armel, T.Z., and Silver, P.A. (2012). A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. Nucleic Acids Res *40*, 5180-5187.

McKeague, M., Wang, Y.H., and Smolke, C.D. (2015). In vitro screening and in silico modeling of RNA-based gene expression control. Acs Chem Biol *10*, 2463-2467.

Miller, A.W., Befort, C., Kerr, E.O., and Dunham, M.J. (2013). Design and use of multiplexed chemostat arrays. Journal of visualized experiments : JoVE, e50262.

Perez-Pinera, P., Ousterout, D.G., Brunger, J.M., Farin, A.M., Glass, K.A., Guilak, F., Crawford, G.E., Hartemink, A.J., and Gersbach, C.A. (2013). Synergistic and tunable human gene activation by combinations of synthetic transcription factors. Nat Methods *10*, 239-242.

Ramos, J.L., Martinez-Bueno, M., Molina-Henares, A.J., Teran, W., Watanabe, K., Zhang, X., Gallegos, M.T., Brennan, R., and Tobes, R. (2005). The TetR family of transcriptional repressors. Microbiology and molecular biology reviews : MMBR *69*, 326-356.

Reijmers, L.G., Perkins, B.L., Matsuo, N., and Mayford, M. (2007). Localization of a stable neural correlate of associative memory. Science *317*, 1230-1233.

Rosenfeld, N., Elowitz, M.B., and Alon, U. (2002). Negative autoregulation speeds the response times of transcription networks. Journal of molecular biology *323*, 785-793.

Russell, R., Das, R., Suh, H., Travers, K.J., Laederach, A., Engelhardt, M.A., and Herschlag, D. (2006). The paradoxical behavior of a highly structured misfolded intermediate in RNA folding. Journal of molecular biology *363*, 531-544.

Shepotinovskaya, I.V., and Uhlenbeck, O.C. (2008). Catalytic diversity of extended hammerhead ribozymes. Biochemistry *47*, 7034-7042.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Thodey, K., Galanie, S., and Smolke, C.D. (2014). A microbial biomanufacturing platform for natural and semisynthetic opioids. Nat Chem Biol *10*, 837-844.

Townshend, B., Kennedy, A.B., Xiang, J.S., and Smolke, C.D. (2015). High-throughput cellular RNA device engineering. Nat Methods.

Treiber, D.K., and Williamson, J.R. (1999). Exposing the kinetic traps in RNA folding. Current opinion in structural biology *9*, 339-345.

Trenchard, I.J., and Smolke, C.D. (2015). Engineering strategies for the fermentative production of plant alkaloids in yeast. Metabolic engineering *30*, 96-104.

Wang, Y.H., Wei, K.Y., and Smolke, C.D. (2013). Synthetic biology: advancing the design of diverse genetic systems. Annu Rev Chem Biomol *4*, 69-102.

Zhao, H., and Arnold, F.H. (1997). Optimization of DNA shuffling for high fidelity recombination. Nucleic Acids Res *25*, 1307-1308.

Zhou, X., Symons, J., Hoppes, R., Krueger, C., Berens, C., Hillen, W., Berkhout, B., and Das, A.T. (2007). Improved single-chain transactivators of the Tet-On gene expression system. BMC Biotechnol *7*, 6.