Chapter 3

Design of Ligand-Responsive Small Hairpin RNAs

ABSTRACT

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

INTRODUCTION

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

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

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

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

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

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

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

RESULTS

Design and characterization of a modular shRNA platform

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

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

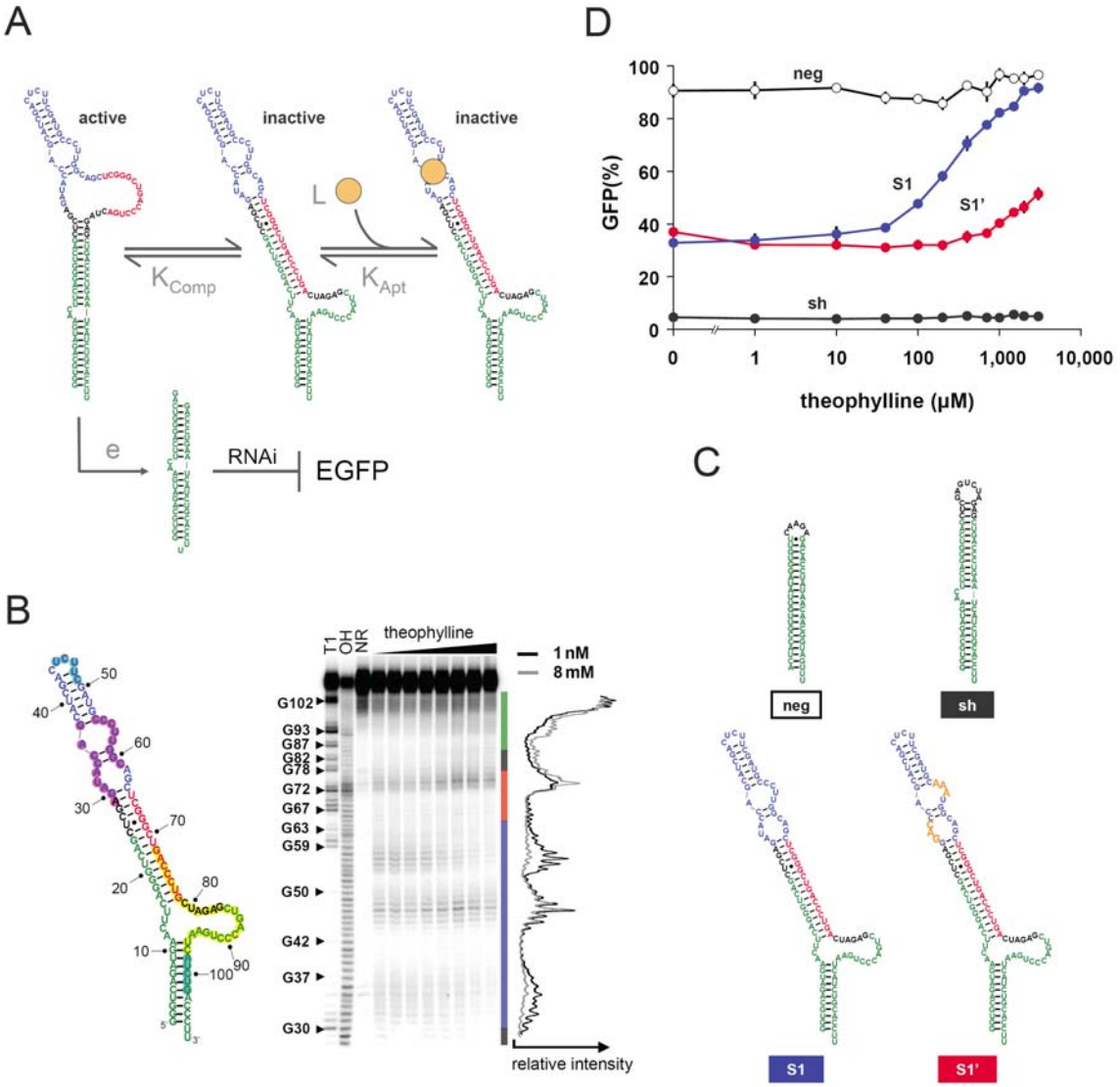


Figure 3.1 Design and characterization of an shRNA switch platform. Color schemes for switches shown in all figures are as follows: shRNA stem, green; aptamer domain, blue; competing strand, red; mutations to aptamer core, orange. (**A**) Sequence and representative structures of shRNA switch S1 and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

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

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

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

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

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

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

Mathematical modeling offers tuning parameters to predictively modulate component transfer functions

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

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

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

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

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

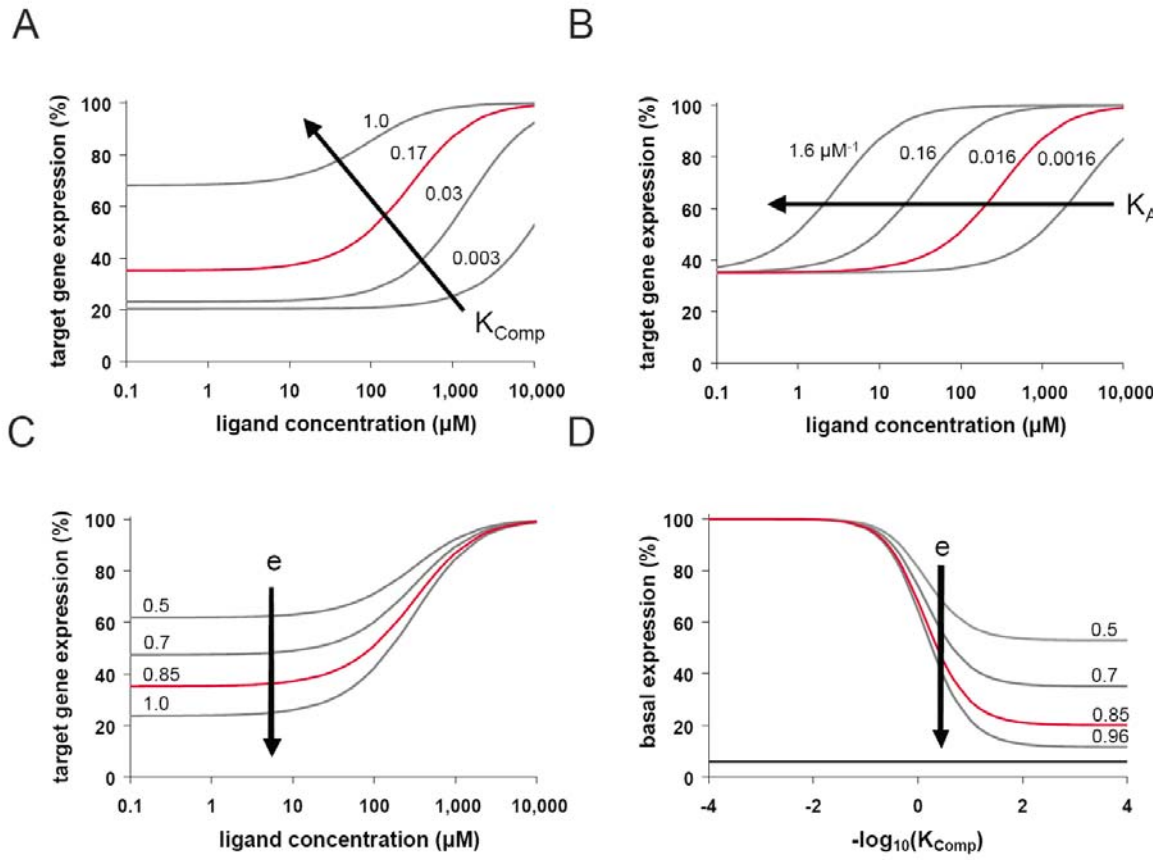


Figure 3.2 Model predicts tuning of the shRNA switch transfer function through variation of identified tuning parameters. Model predictions for the effect on the component transfer function of varying K_{Comp} (**A**), K_{Apt} (**B**), or e (**C**). (**D**) Effect of e on the dependence of basal expression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

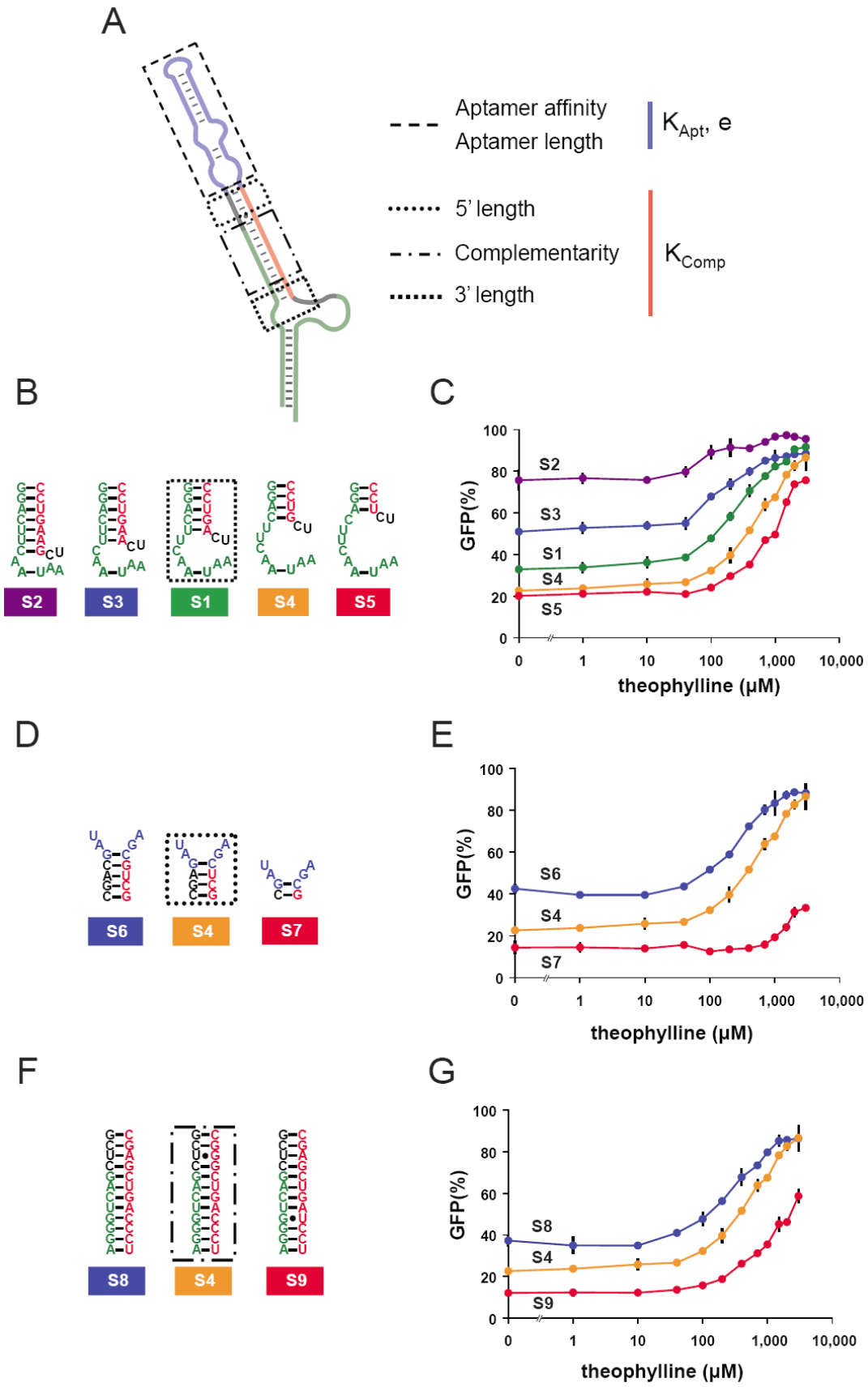

A
Aptamer affinity
Aptamer length
K_{Apt}, e
5' length
Complementarity
3' length
K_{Comp}
B
S2
S3
S1
S4
S5
C
GFP(%)
theophylline (μM)
S2
S3
S1
S4
S5
D
S6
S4
S7
E
GFP(%)
theophylline (μM)
S6
S4
S7
F
S8
S4
S9
G
GFP(%)
theophylline (μM)
S8
S4
S9

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

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

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

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

Aptamer tuning strategies enable predictive alteration of the component transfer function

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

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

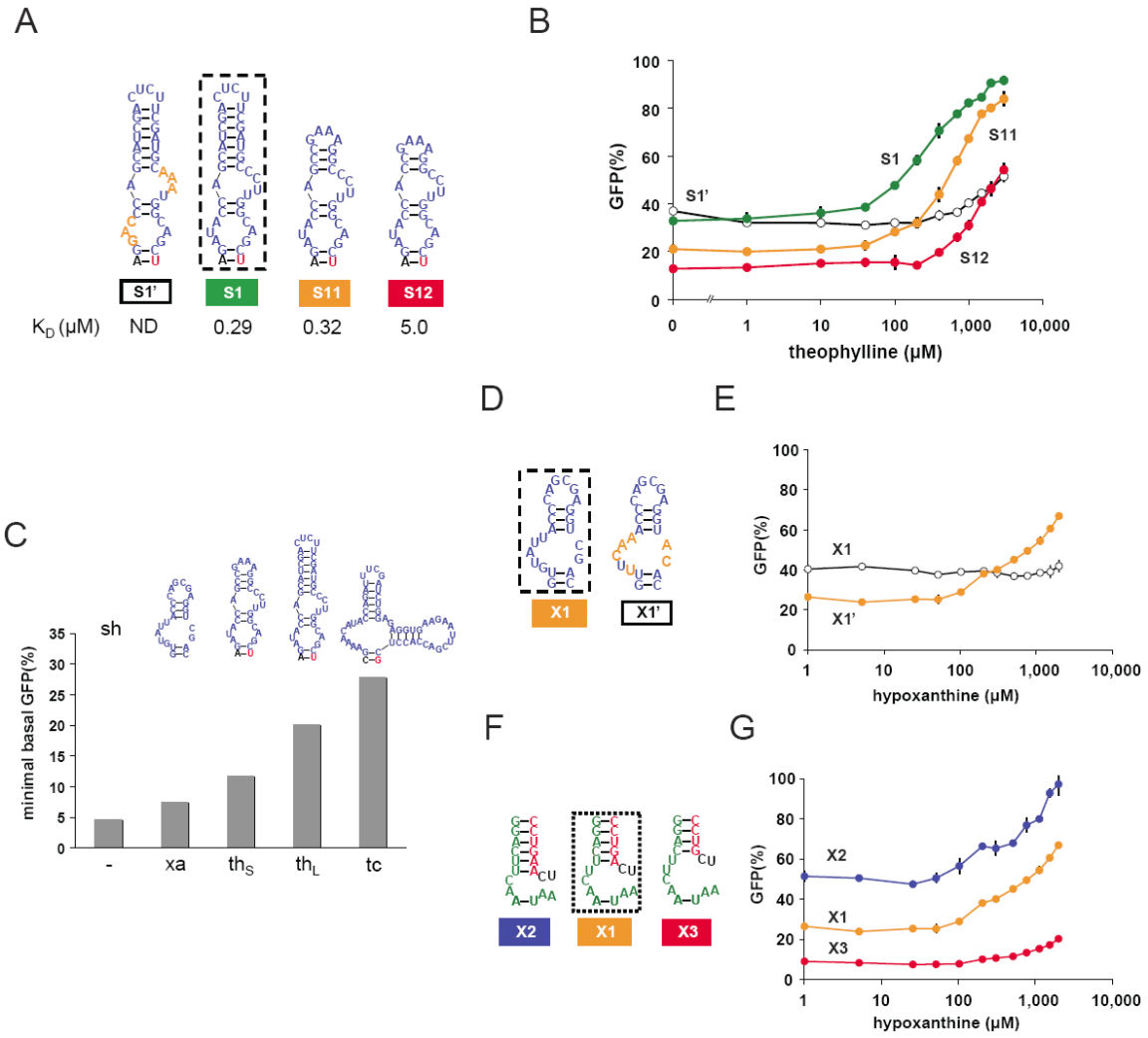


Figure 3.4 Experimental validation of aptamer tuning strategies. (**A**) Theophylline aptamer variants swapped into the equivalent box in Figure 3.3A. Dissociation constants (K_D) as reported previously (Zimmermann et al, 2000) are indicated for each aptamer. (**B**) Tuned theophylline response curves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

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

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

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

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

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

Programming transfer functions by combining competing strand or aptamer tuning strategies

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

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

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

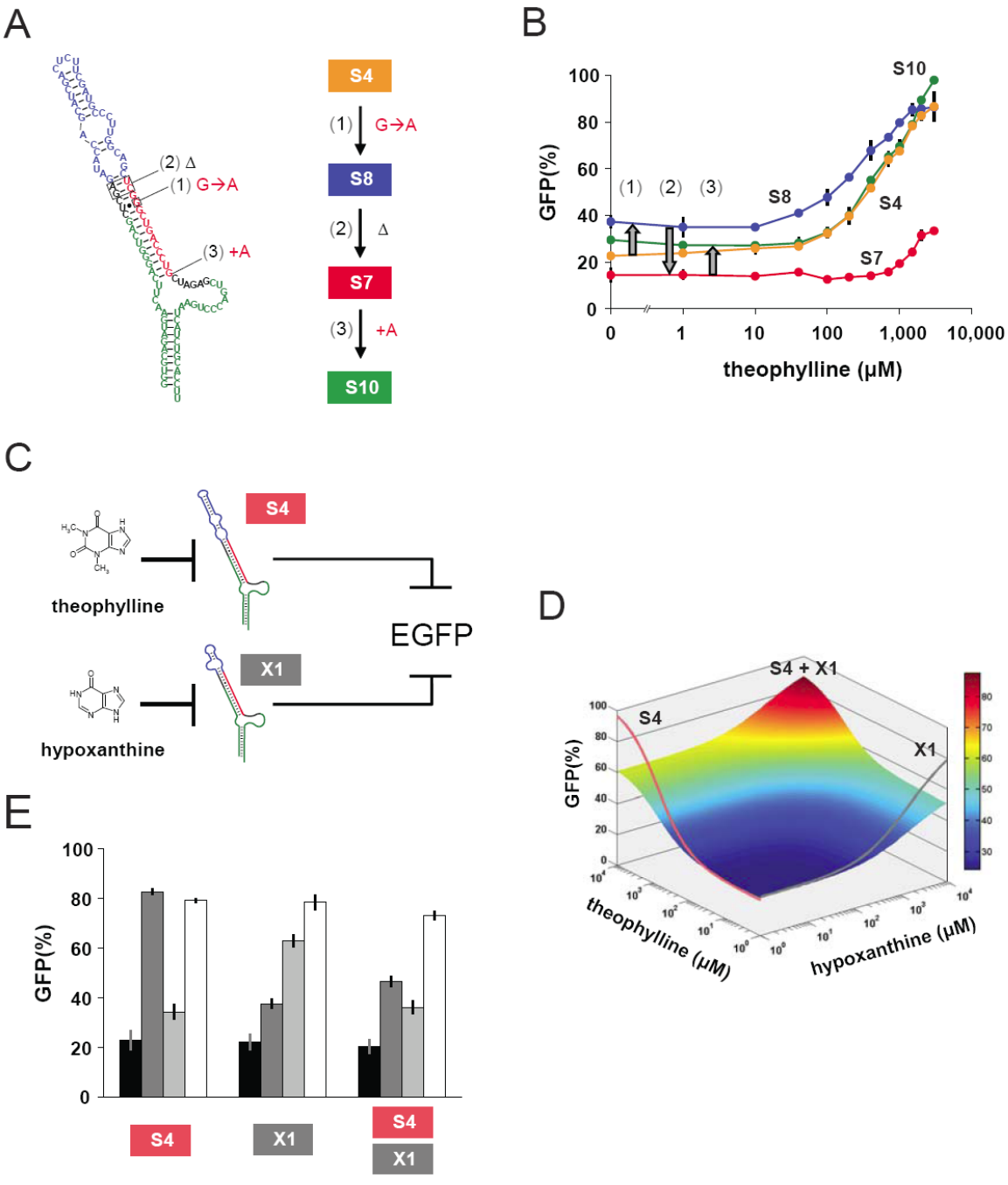


Figure 3.5 Programming transfer functions through combinatorial design strategies. (**A**) Combinatorial tuning strategies enable fine-tuning of the component transfer function. Stepwise nucleotide changes were made to S4, where each change fell under a different competing strand tuning strategy. (**B**)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

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

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

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

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

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

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

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

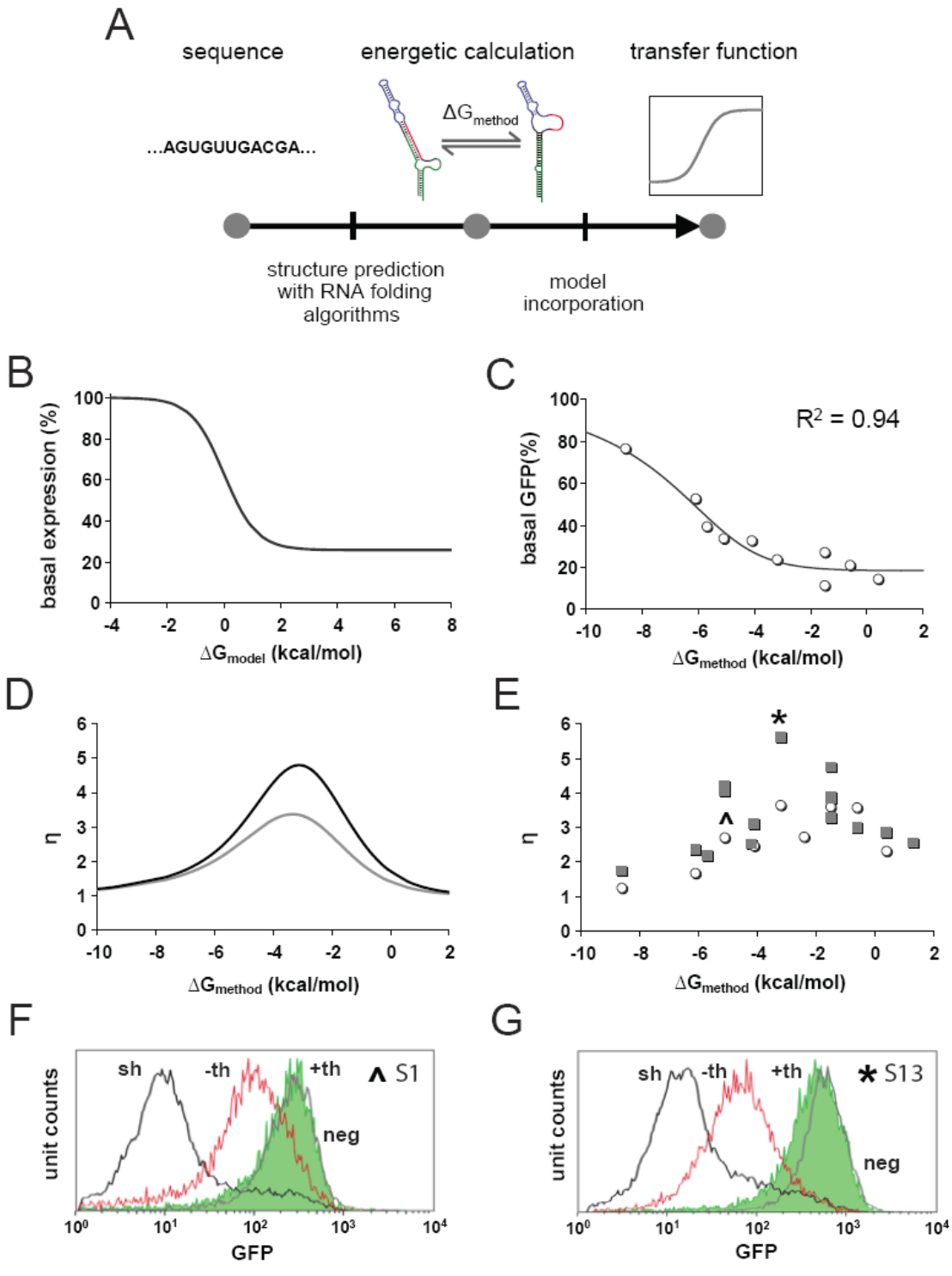


Figure 3.6 Extended model enables sequence-to-transfer function prediction and guides the forward design of optimized shRNA switches. (**A**) General process to convert shRNA switch sequence information into a predicted transfer function. RNA secondary structure algorithms and the method displaying the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

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

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

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

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

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

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

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

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

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

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

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

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

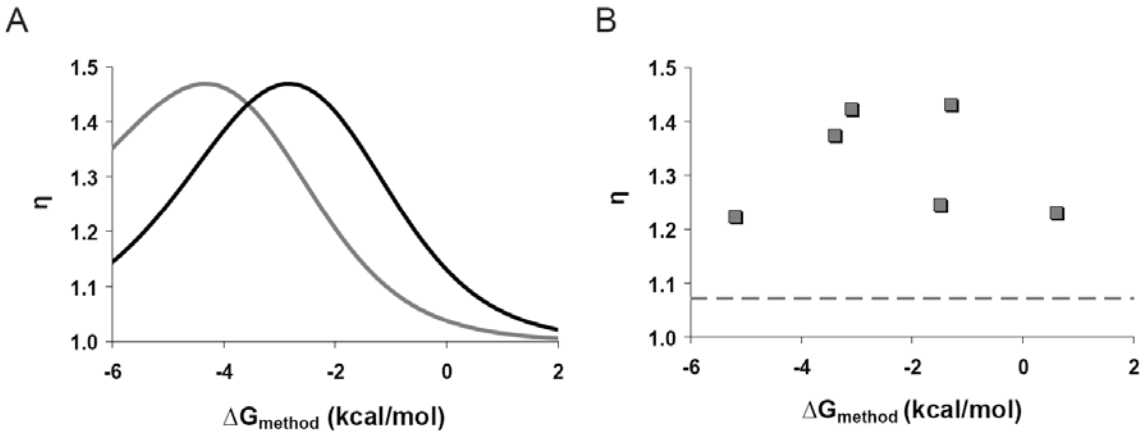


Figure 3.7 Model application guides the forward design of shRNA switches targeting an endogenous gene. (**A**) Extended model predictions for the relationship between ΔG_{method} under the 'Stems' method and dynamic range (η) using empirical parameter values determined from the GFP experiments and experimental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

DISCUSSION

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

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

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

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

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

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

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

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

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

MATERIALS AND METHODS

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

ACKNOWLEDGEMENTS

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

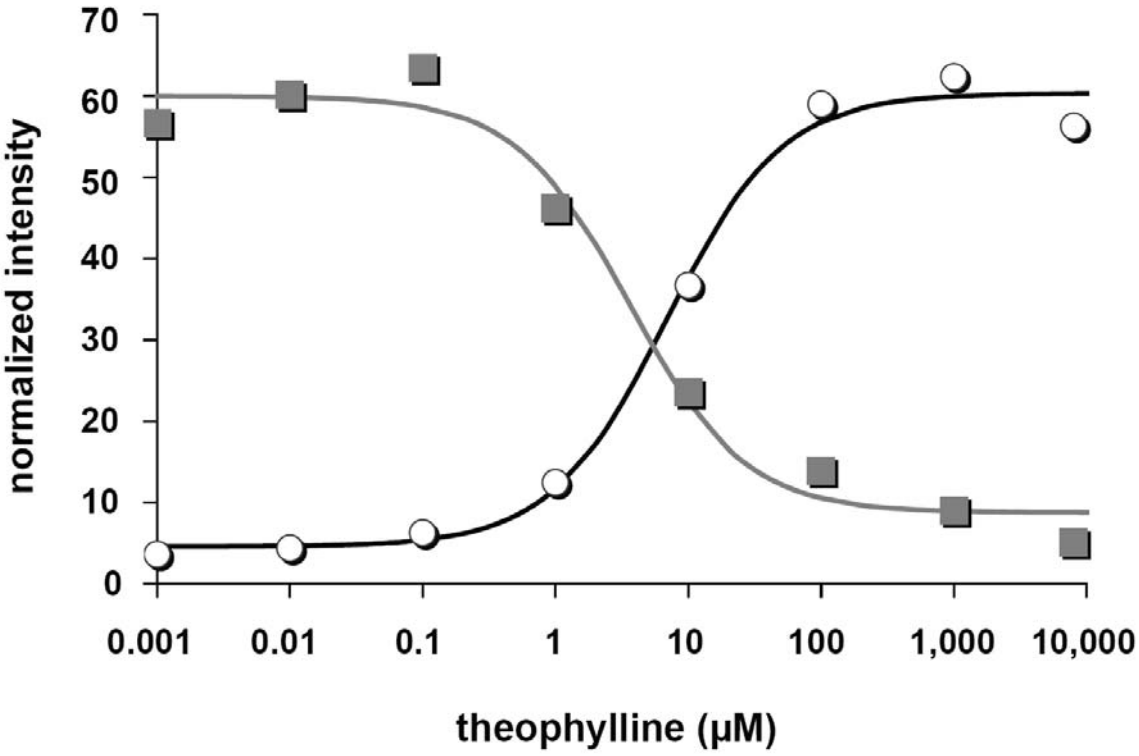
Win MN, Smolke CD (2007) From the cover: a modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci USA* **104**: 14283–14288.

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

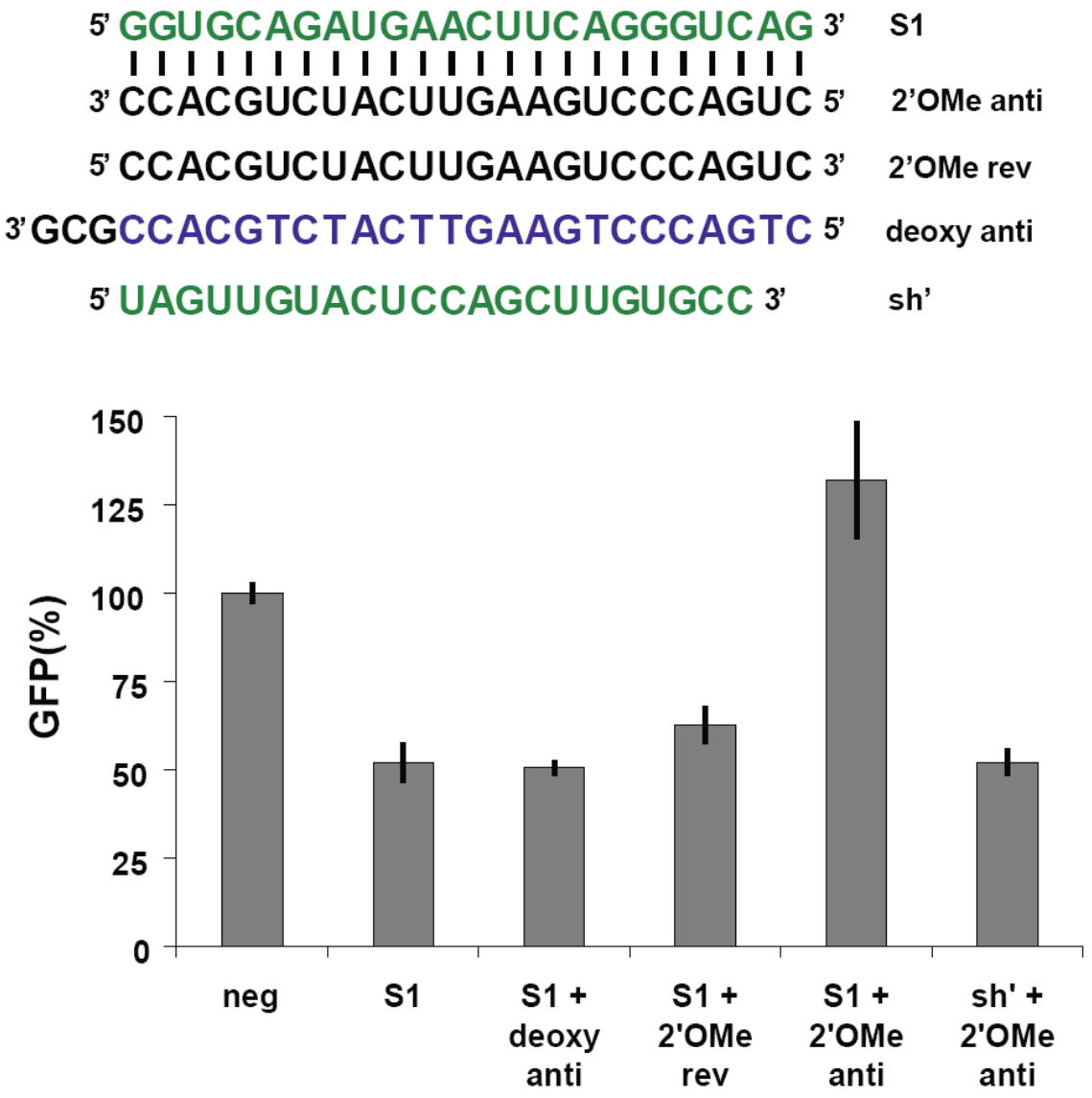
Zeng Y, Cullen BR (2004) Structural requirements for pre-microRNA binding and nuclear export by Exportin 5. *Nucleic Acids Res* **32**: 4776–4785.

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

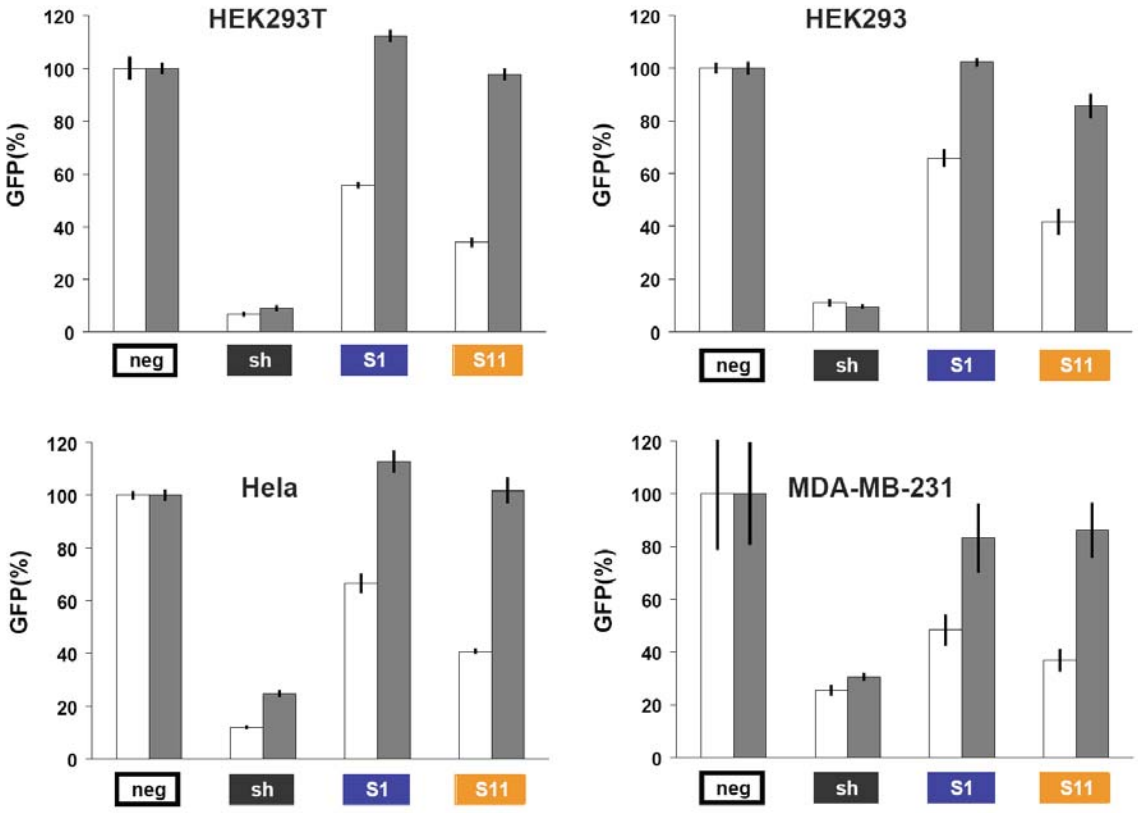
SUPPLEMENTARY INFORMATION



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

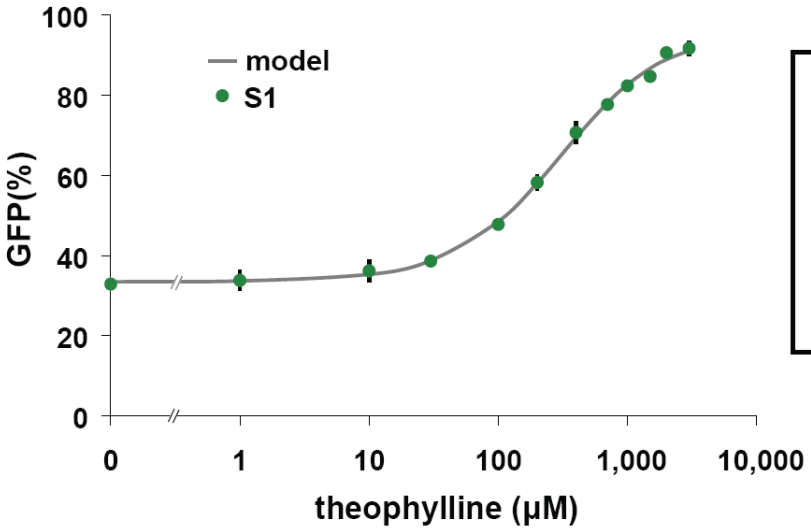


Supplementary Figure S3.2 Antisense inhibition of guide strand activity represses RNAi-mediated silencing of GFP. Cells stably expressing EGFP were cotransfected with the designated oligo and a plasmid that expresses a scrambled shRNA (neg), shRNA switch S1, or an shRNA targeting a different region of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



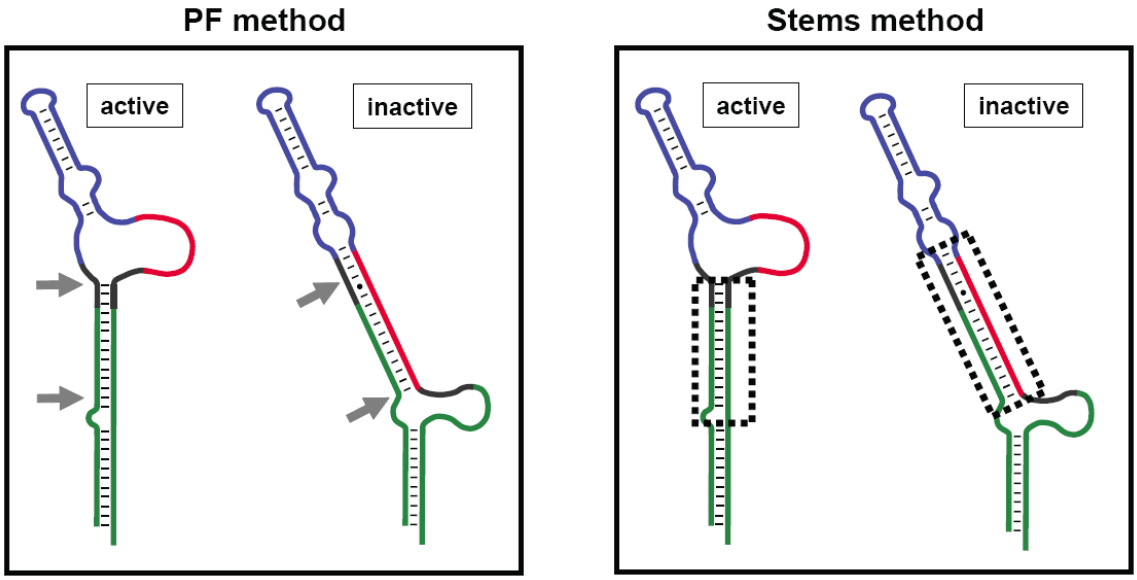
Supplementary Figure S3.3 Functionality of shRNA switches in different cell lines. Cells lacking endogenous EGFP expression were cotransfected with the shRNA construct and an EGFP expression plasmid in the presence (■) or absence (□) of 3 mM theophylline. Mean fluorescence values were normalized to red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

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

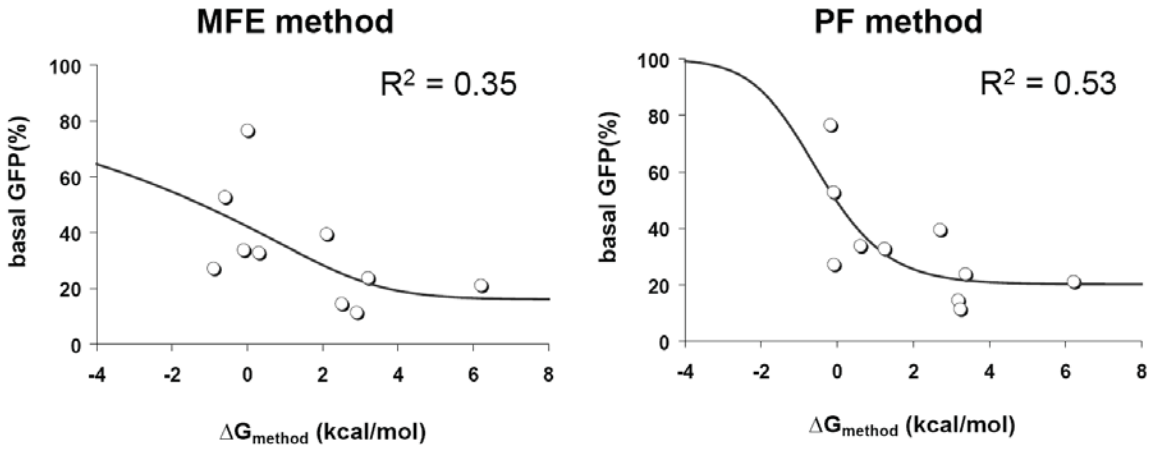


K_{Comp}	0.17
K_{Apt} (μM^{-1})	0.016
e	0.85
f_{shRNA}	0.94
h	1.33

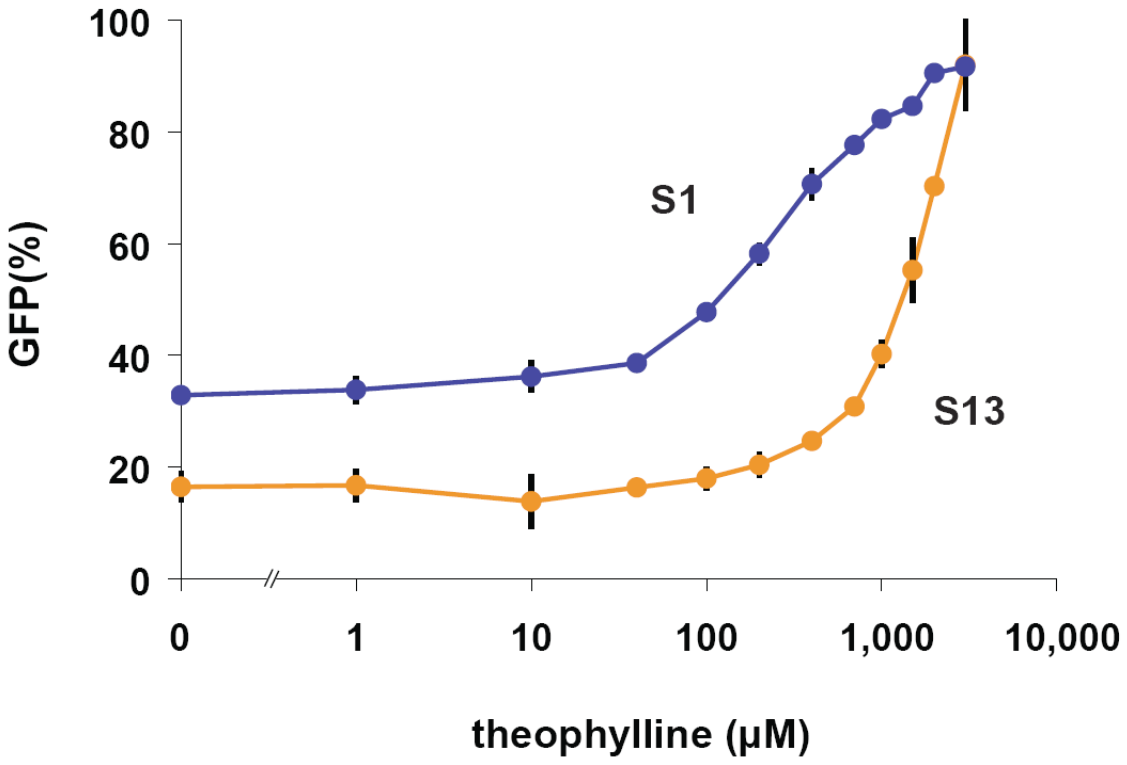
Supplementary Figure S3.4 Derived model equation and model fit of theophylline response curve for S1 data from Figure 3.1D. See Supplementary Text S3.1 for model derivation. The value of f_{shRNA} was set by a separate transfection experiment with the original shRNA targeting EGFP (sh) under the same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



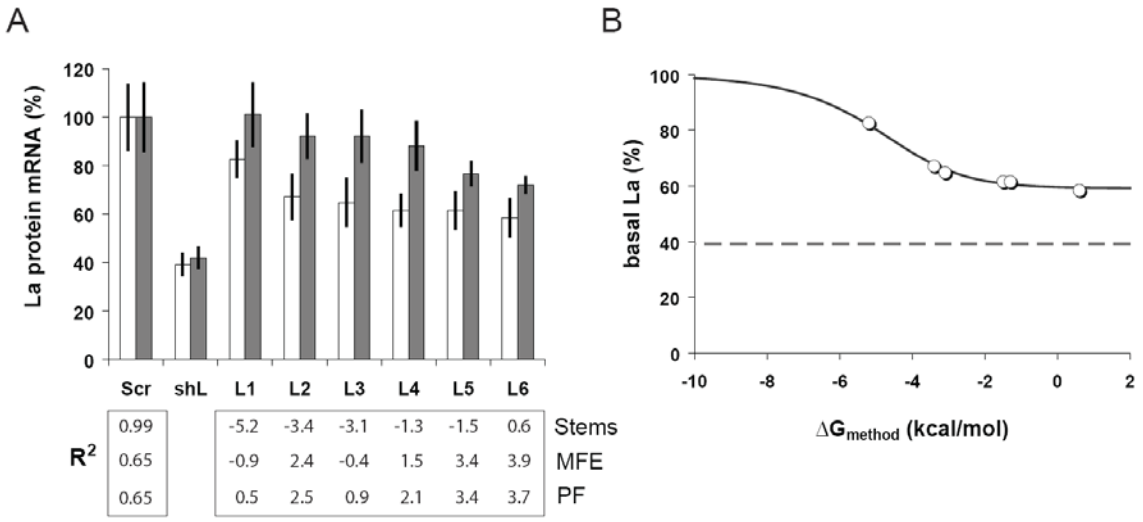
Supplementary Figure S3.5 Depiction of methods to calculate ΔG from shRNA switch sequence information. Base-pairing probabilities of base-pairs designated by arrows were used for the PF method. The boxed sections designate the major stem in the active and inactive conformations used under the Stems method.



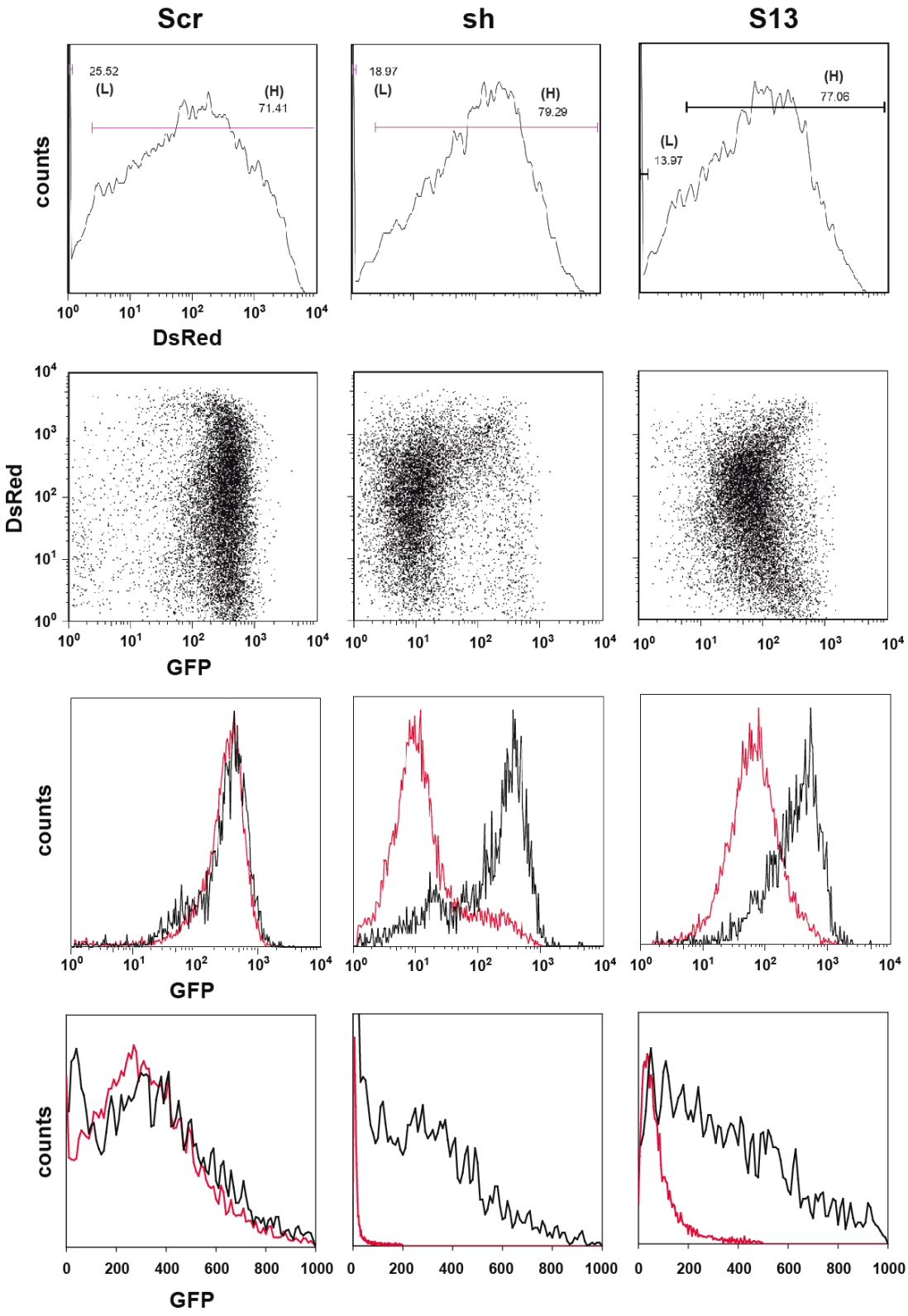
Supplementary Figure S3.6 Alternative methods to relate shRNA switch sequence and in vivo basal expression levels. ΔG was calculated (ΔG_{method}) for shRNA switches S1-10 using RNA secondary structure prediction algorithms. Plots relating ΔG_{method} and measured basal expression level for shRNA switches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



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



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



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

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

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

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

Supplementary Table S3.1 cont'd.

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

Supplementary Table S3.1 cont'd.

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

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

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

Supplementary Table S3.2 cont'd.

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

Supplementary Text S3.1

DERIVATION OF MATHEMATICAL MODEL

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Supplementary Table S3.3 Description of model parameters.

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

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

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

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

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

REFERENCES

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Supplementary Text S3.2

FREE ENERGY CALCULATIONS AND MODEL EXTENSION

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

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

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

Free energy calculation

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

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

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

MFE method

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

PF method

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

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

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

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

Stems method

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

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

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

Model extension

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

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

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

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

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

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

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

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

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

REFERENCES

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

Supplementary Methods

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

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

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