Chemistry & Biology
Article

A High-Throughput Screen for Synthetic Riboswitches Reveals Mechanistic Insights into Their Function

Sean A. Lynch,[1] Shawn K. Desai,[1] Hari Krishna Sajja,[1] and Justin P. Gallivan[1,*]
[1] Department of Chemistry and Center for Fundamental and Applied Molecular Evolution, Emory University, 1515 Dickey Drive, Atlanta, GA 30322, USA
*Correspondence: justin.gallivan@emory.edu
DOI 10.1016/j.chembiol.2006.12.008

SUMMARY

Riboswitches are RNA-based genetic control elements that regulate gene expression in a ligand-dependent fashion without the need for proteins. The ability to create synthetic riboswitches that control gene expression in response to any desired small-molecule ligand will enable the development of sensitive genetic screens that can detect the presence of small molecules, as well as designer genetic control elements to conditionally modulate cellular behavior. Herein, we present an automated high-throughput screening method that identifies synthetic riboswitches that display extremely low background levels of gene expression in the absence of the desired ligand and robust increases in expression in its presence. Mechanistic studies reveal how these riboswitches function and suggest design principles for creating new synthetic riboswitches. We anticipate that the screening method and design principles will be generally useful for creating functional synthetic riboswitches.

INTRODUCTION

Riboswitches are RNA-encoded genetic control elements that regulate gene expression in a ligand-dependent fashion without the need for proteins [1–3]. Riboswitches are comprised of an aptamer domain, which recognizes the ligand, and an expression platform, which couples ligand binding to changes in gene expression [1, 4]. Riboswitches are widespread in prokaryotes [5–8], and more recent studies have revealed that riboswitches also control gene expression in eukaryotes [9, 10]. In addition to natural riboswitches that control gene expression in response to endogenous metabolites, a variety of synthetic riboswitches that respond to nonendogenous small molecules have been developed [11–16]. In principle, synthetic riboswitches can be engineered to respond to any nontoxic, cell-permeable molecule that is capable of interacting with an RNA. As such, synthetic riboswitches represent versatile ligand-dependent gene expression systems with great potential to detect the production of small molecules for applications in directed evolution or metabolic engineering. While methods to select aptamers that bind to small molecule targets are well established [17], general methods of converting these aptamers into riboswitches that function optimally in prokaryotic cells are not. To fully exploit their potential, it is critical to develop new methods to create synthetic riboswitches that function in bacteria.

A variety of methods have been used to create synthetic riboswitches that control eukaryotic transcription or translation. Werstuck and Green created synthetic riboswitches that regulate translation by cloning aptamer sequences into the 5′ untranslated regions (5′UTR) of eukaryotic mRNA sequences. Binding of the ligand increased the strength of the RNA secondary structure in the 5′UTR and reduced the translation of downstream coding regions [16]. This approach was adopted by the groups of Wilson [12], Pelletier [13], and Suess [15] to create synthetic riboswitches that repress eukaryotic protein translation in response to small molecule ligands. Buskirk et al. [18] created an aptamer-based transcriptional control system that promotes ligand-dependent transcription in yeast, and Gaur and coworkers used an aptamer to control ligand-dependent RNA splicing in vitro [19]. In addition to systems where regulation occurs in *cis*, a variety of riboregulators have been used to affect eukaryotic translation in *trans* [20–23].

Given that most natural riboswitches have been identified in prokaryotes, it is perhaps surprising that there are relatively few examples of synthetic riboswitches that function in bacteria. Suess et al. [14] created a riboswitch based on a designed helix-slipping mechanism that activates protein translation in a theophylline-dependent manner in the Gram-positive bacterium *B. subtilis*, and we reported a theophylline-sensitive riboswitch that activates protein translation in the Gram-negative bacterium *E. coli* [11]. While our previously reported synthetic riboswitches were useful in several contexts, three important issues remained unresolved. First, these switches did not completely repress protein translation in the absence of the ligand. Second, the most effective switch showed a signal to background ratio (activation ratio) of approximately eight in the presence of 500 μM theophylline. While this 8-fold increase allowed us to perform both genetic screens

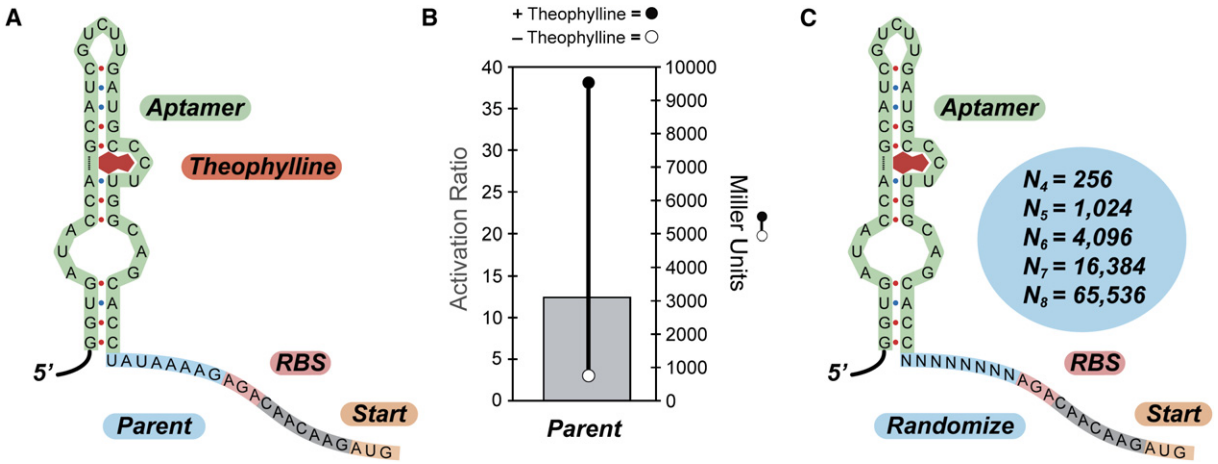

Figure 1. Diagram of the 5′ Region of a Synthetic Riboswitch, the Performance of the Synthetic Riboswitch, and the Randomization Strategy

(A) Sequence of a portion of the 5′ region of the parent synthetic riboswitch with 8 bases separating the aptamer (green) and the ribosome binding site (pink); the AUG start codon (peach) is highlighted. The aptamer is shown in the secondary structure predicted for the theophylline aptamer by *mFold* and confirmed experimentally.

(B) Measures of the activity of the synthetic riboswitch shown in (A) when cloned upstream of the *IS10-lacZ* gene fusion and expressed in *E. coli.* Right axis: β-galactosidase activity in the absence (open circle) or presence of theophylline (1 mM, closed circle). Activities are expressed in Miller units, and the standard errors of the mean are less than the diameters of the circles. Left axis: the activation ratio of the synthetic riboswitch (gray bar), which is determined by taking the ratio of the activities in the presence and absence of theophylline.

(C) Sequence diagram of the randomized synthetic riboswitches. The aptamer is shown in green, the randomized regions in light blue, the ribosome binding site in pink, and the AUG start codon in peach. Theoretical library sizes are shown in the circle.

and selections to detect the presence of theophylline, for more demanding screening applications, an increase in the ratio of signal to background is desirable. Finally, while we determined that these riboswitches operated posttranscriptionally, we were unable to determine their precise mechanisms of action. To address these issues, we developed a high-throughput screen that enables the identification of synthetic riboswitches that display very low background levels of translation in the absence of ligand and dramatically higher signal-to-background ratios. Sequence analysis allowed us to propose and test a model for synthetic riboswitch function and to also explain the mechanism of action of our previously reported synthetic riboswitches. Our results show that starting from a single aptamer, there are many different ways to create functional riboswitches, consistent with the notion that natural riboswitches may be derived from small-molecule-binding RNA aptamers that were subsequently recruited to regulate gene expression. We anticipate that the high-throughput assay described herein will be generally useful for discovering synthetic riboswitches with new ligand specificities and better performance characteristics.

RESULTS AND DISCUSSION

We previously observed that changing the length of the sequence separating the theophylline aptamer and the ribosome binding site (RBS) had dramatic effects on both the function and dynamic range of a synthetic riboswitch [11]. However, we could not determine precisely how these changes exerted their effects. We anticipated that by randomizing the RNA sequence in this region and screening the library for function, we could identify improved switches and possibly gain insight into their mechanisms of action. Because the sequence space of RNA is relatively small, most, if not all, potential sequences can be sampled in the context of a simple genetic screening experiment.

Creation of a Library of Randomized Mutants

We previously created synthetic riboswitches by cloning a theophylline-binding aptamer [24–27] at various locations upstream of the ribosome binding site of a β-galactosidase reporter gene (*IS10-lacZ*) that was controlled at the transcriptional level by a weak, constitutively active *IS10* promoter (Figure 1A). To increase the overall signal, we replaced the *IS10* promoter with the stronger P*tac*1 promoter [28]. As expected, the *tac* promoter enhances the level of β-galactosidase expression in the presence of theophylline but also increases the background expression ~80-fold from ~10 Miller units [11] to ~800 Miller units (Figure 1B).

We used cassette-based PCR mutagenesis to create five different libraries in which the distance between the aptamer and the RBS was varied between 4 and 8 bases, and the sequence was randomized fully (Figure 1C) because we previously observed that longer or shorter spacings resulted in poorly functioning switches [11].

High-Throughput Screen for Optimally Functioning Riboswitches

To screen the libraries, *E. coli* were transformed with plasmids harboring randomized sequences of a given length

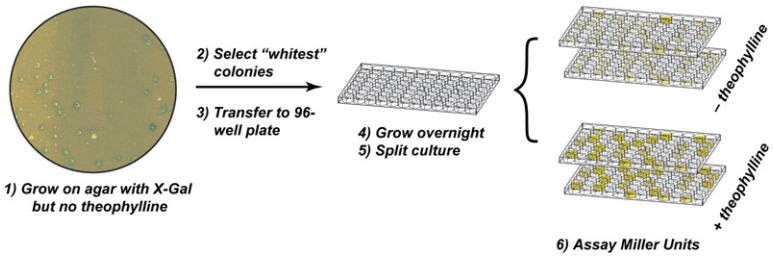

Figure 2. Diagram of the High-Throughput Assay

Candidate riboswitches are identified by plating cells onto selective media containing X-gal but no theophylline. A robotic colony picker identifies the whitest colonies (lowest levels of β-galactosidase activity) and transfers the cells to a 96-well microtiterplate. The culture is grown overnight in selective media, split, and the clones are tested for β-galactosidase activity by using the automated assay described in the text.

and were grown on selective agar plates in the presence of X-gal (the substrate for β-galactosidase), but in the absence of theophylline. Most (~99%) clones appeared blue, indicating that they were expressing β-galactosidase in the absence of theophylline. However, in all libraries, a number of colonies appeared white, suggesting little to no β-galactosidase expression. To identify whether these colonies harbored riboswitches that could activate protein translation in the presence of theophylline, we used a colony-picking robot to isolate the whitest colonies from each plate of approximately 4000 colonies (Figure 2). These clones were inoculated into 96-well microtiterplates containing selective LB media. These cultures were grown overnight and were used to inoculate two new 96-well plates that either contained theophylline (0.5 mM) or did not. These plates were grown at 37°C with shaking for 2–2.5 hr, and β-galactosidase activity was assayed using an adaptation of Miller's method performed either by hand with a multichannel pipettor or by using a robotic liquid-handling system [29–31]. Ratios of the Miller units for cultures grown in the presence of theophylline to those grown without theophylline (the "activation ratio") were utilized to identify functional switches.

To validate the screen, we chose clones that displayed an activation ratio of greater than two and assayed them individually in larger cultures, but we discovered that less than half of these clones functioned as switches. Further analysis indicated that these irregularities were due to a number of factors, including small but significant differences in the growth rates and lysis efficiencies of the cells in the microtiterplates. To improve the accuracy of identifying functional synthetic riboswitches, we performed each assay in duplicate and analyzed the data by an empirically derived procedure in which we retained candidates that: (1) showed an activation ratio of greater than 2.0 in two separate determinations, (2) displayed a minimum level of β-galactosidase activity in the presence of theophylline (an $OD_{420} \geq 0.04$ in the Miller assay, regardless of cell density), (3) grew normally relative to others in the plate (as represented by OD_{600}), and (4) showed consistent results between the two plates. This simple analysis significantly reduced the number of potential candidates, of which greater than 90% were confirmed as functional synthetic riboswitches when assayed individually in larger volumes of culture. For the small number of candidates that were not validated, sequencing often revealed mixed populations of plasmids that were likely introduced during colony picking, and plating cells at lower densities minimized such events.

Figure 3 shows statistics for the libraries and the results from the top validated hits identified in each of these screens. For each sequence length, the absolute levels of β-galactosidase activity in the presence and absence of theophylline (1 mM) are plotted, as well as the ratio of these numbers (activation ratio) for the switches that displayed the highest activation ratios. From these data, it is clear that the high-throughput screening method is capable of identifying riboswitches that display both low background levels of protein expression in the absence of ligand and strong increases in the presence of the ligand. Indeed, the best clone identified (8.1) displays a 36-fold increase in protein expression in the presence of theophylline and a very low level of expression in its absence. To put this in perspective, natural riboswitches that regulate protein translation (though, most of these switches repress protein translation) show repression ratios of about 100 in the presence of the ligand [7]. The discovery of switches that function within a factor of three of natural genetic regulatory elements that have evolved over considerably longer time periods suggests that this screening method is quite effective and that creation of new synthetic riboswitches based on different aptamers may be straightforward.

Sequencing Suggests a Possible Mechanism of Action for Synthetic Riboswitch Function

In addition to identifying synthetic riboswitches that display excellent performance characteristics, these screens provided a wealth of sequence information that allowed us to propose and test models for riboswitch function. Visual sequence analysis revealed several conserved or semi-conserved motifs that are complementary to different regions of the theophylline aptamer. Shown in Figure 4 are the *mFold*-predicted [32, 33] secondary structures of two of the new synthetic riboswitches in the region extending from the 5′ end of the aptamer to the 3′ end of the AUG start codon of the *IS10-lacZ* gene. On the left-hand side are the two minimum energy structures—in both cases, the bases between the aptamer and the RBS are paired with the aptamer sequence, suggesting that translation might be inhibited in the absence of the ligand. On the right-hand side of Figure 4, the experimentally determined secondary structure for the theophylline-binding aptamer [24–27] appears in the calculated

A

	Theoretical Size	Colonies Plated	Clones Assayed	Confirmed Switches
N_4	256	~ 6,000	96	0
N_5	1,024	~ 4,000	192	4
N_6	4,096	~ 19,000	288	16
N_7	16,384	~ 28,000	480	>14
N_8	65,536	~ 59,000	864	>20

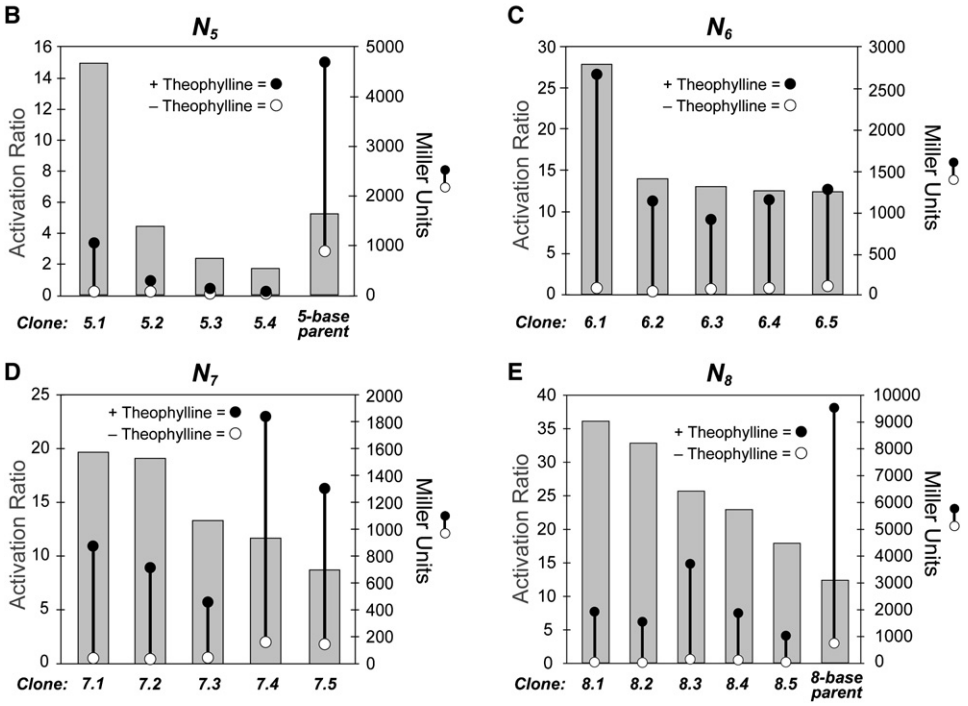

Figure 3. Analysis of Libraries and Results from Screens

(A) Statistics for libraries assayed.

(B–E) Measures of the activities of the synthetic riboswitches identified in the high-throughput screens. Each measurement is done in triplicate. For all panels: right axis, β-galactosidase activity in the absence (open circle) or presence of theophylline (1 mM, closed circle). Activities are expressed in Miller units, and the standard errors of the mean are less than the diameters of the circles. Left axis, the activation ratio of the synthetic riboswitches (gray bar), which are determined by taking the ratio of the activities in the presence and absence of theophylline.

structure and the RBS is not paired—in both cases, the calculated differences in free energy between the two structures are less than the free energy of theophylline binding (−9.2 kcal/mol) [24], suggesting that conversion between the structures is thermodynamically favorable in the presence of theophylline. All of the synthetic riboswitches identified in our screens can be folded into one of the two folding motifs ("blue" or "red") shown in Figure 4 (sequences and their motifs are reported in the Supplemental Data available with this article online).

The folds in Figure 4 suggest a possible switching mechanism in which extensive pairing in the region near the ribosome binding site prevents translation of β-galactosidase in the absence of theophylline. Such behavior is consistent with the studies of de Smit and van Duin who demonstrated that secondary structure near the ribosome binding site dramatically reduces the translation of the mRNA downstream [34–36]. Theophylline binding ($\Delta G_{bind} \sim -9.2$ kcal/mol) could in principle drive the equilibrium toward the structures on the right, in which the known secondary structure of the aptamer is present [25] and the ribosome binding sites are unpaired, which would increase the efficiency of translation.

To test whether base-pairing between the aptamer and a region near the ribosome binding site was important for function, we performed covariance experiments in which we mutated these bases to disrupt the putative base-pairing (Figure 5). Removal of several putative base pairs in clone 8.1 increased β-galactosidase expression 5-fold in the absence of the ligand, while restoring the base-pairing by mutating the aptamer sequence decreased the background level of β-galactosidase expression to its original level, consistent with the pairing hypothesis. Because these mutations also affect theophylline

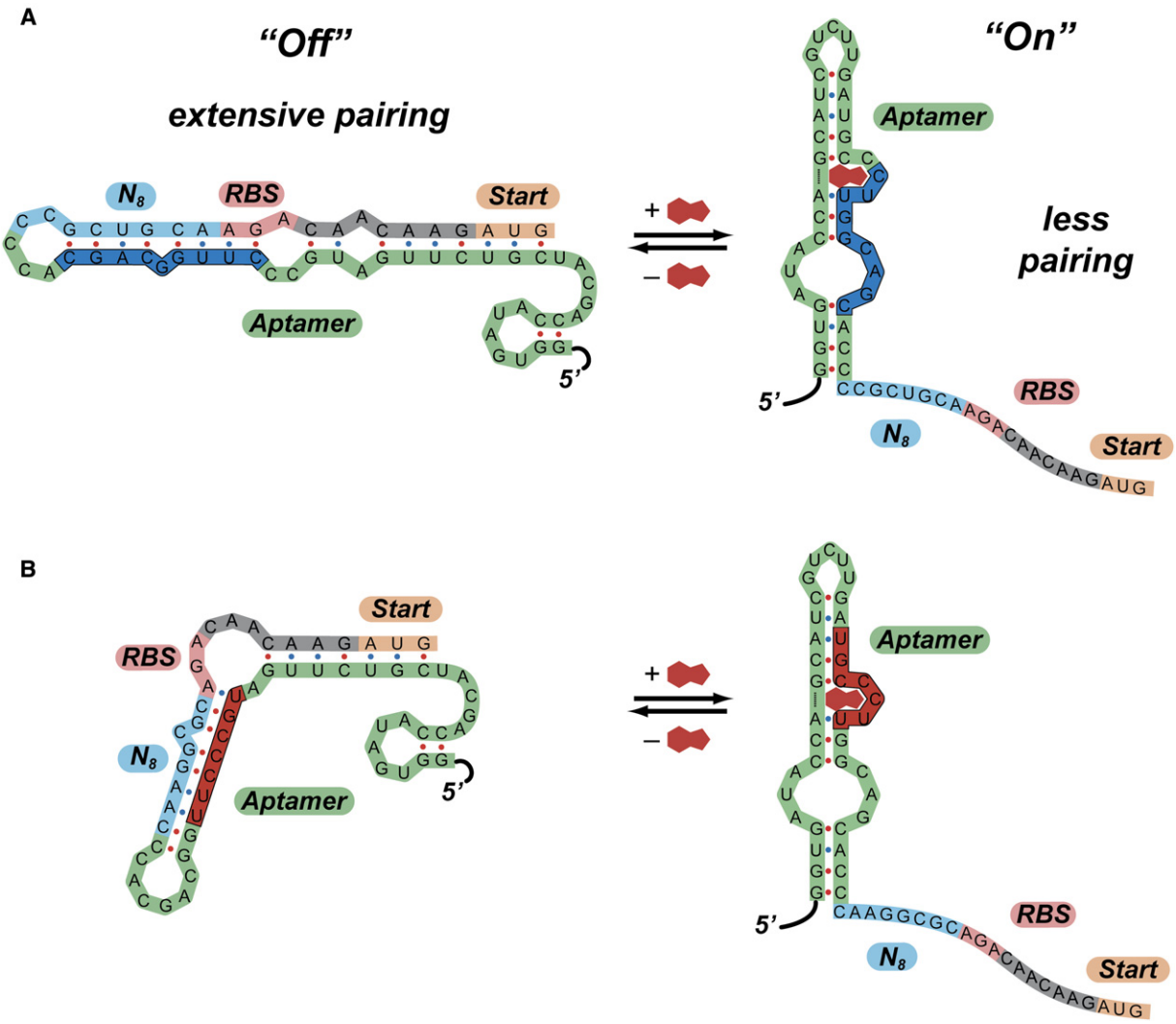

Figure 4. Predicted Mechanisms of Action of Synthetic Riboswitches

(A) Predicted mechanism of action of clone 8.1. In the absence of theophylline (left), the 5′ region can adopt a highly folded structure that extensively pairs the region that includes ribosome binding site to part of the aptamer sequence (shown in dark blue). In the presence of theophylline (right), the secondary structure shifts, such that the ribosome binding site is exposed. The secondary structure shown for the theophylline aptamer predicted by *mFold* and is identical to the structure determined by NMR.

(B) Predicted mechanism of action of clone 8.2. In this case, the region containing the ribosome binding site pairs to a different region of the aptamer in the "off" state (shown in red). This demonstrates that several motifs may function as synthetic riboswitches.

binding, these mutants were not active in the presence of theophylline [26].

In light of these results that suggest that pairing in the region between the aptamer and the ribosome binding site is important for riboswitch function, we revisited the activation data from our previously reported synthetic riboswitches [11]. In creating those switches, we did not explicitly engineer or screen for any particular sequence in this region. Reexamination of our previously reported riboswitch (Figure 1A) revealed that a 7 base sequence located 27 bases after the start codon of the *IS10-lacZ* reporter gene (UUUCUCU) is the precise reverse complement of the bases between the aptamer and the ribosome binding site. To test whether these regions pair to suppress gene expression in the absence of theophylline, we deleted the N-terminal IS10 fusion (pSALWT-ΔIS10). If bases within the IS10 sequence pair to suppress gene expression in the absence of theophylline, deleting this region should increase gene background levels of expression. Consistent with this model, deletion of the IS10 sequence increased the β-galactosidase activity in the absence of theophylline approximately 15-fold (from ~800 to over 12,000 Miller units), confirming that the previously reported synthetic riboswitches depended on the presence of the IS10 sequence to function.

The models proposed in Figure 4 predict that all of the necessary elements to control gene expression are located in the 5′UTR, suggesting that the sequence of

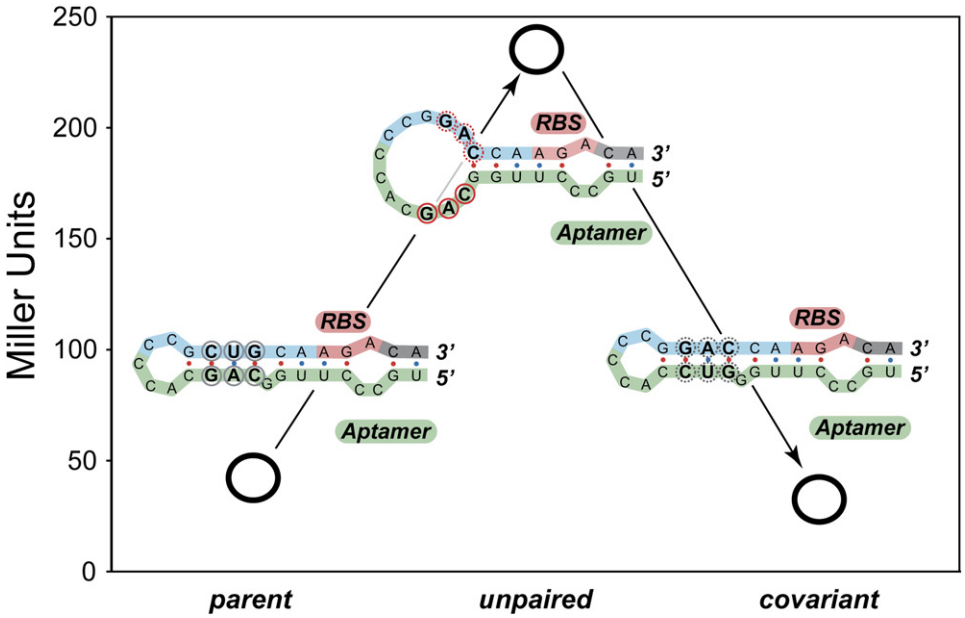

Figure 5. Results from Covariance Experiments

β-galactosidase activities in the absence of theophylline (open circles) for the parent synthetic riboswitch (clone 8.1) (left), a triple mutant that unpairs the region near the ribosome binding site (center), and a mutant that restores the pairing (right). Activities are expressed in Miller units (shown on the left axis); the standard errors of the mean are less than the diameters of the circles. For each construct, the predicted secondary structure is shown in the region near the ribosome binding site. The parent and covariant maintain the pairing and show low background levels of β-galactosidase activity. The mutant unpairs this region and leads to leaky expression in the absence of theophylline.

the downstream gene should not impact the function of the switch. To test this, we replaced the entire 5′UTR of the "leaky" pSALWT-ΔIS10 construct with the 5′UTR from clone 8.1. Gratifyingly, this construct (pSKD8.1-ΔIS10) behaved nearly identically to clone 8.1, with low background levels of β-galactosidase expression in the absence of theophylline and a strong increase in β-galactosidase expression in the presence of theophylline (data not shown). Furthermore, we cloned the entire 5′UTR from clone 8.1 upstream of several reporter genes that lack the N-terminal IS10 fusion, including *gfp*, *dsRED*, and *cat*. These switches all function well and do not depend on the sequence of the gene downstream (S.A.L., H.K.S., and J.P.G., unpublished data). Taken together, these experiments and the covariance experiments above strongly implicate a functional role for the region between the aptamer and the RBS, and the results are consistent with the RNA folds shown in Figure 4.

A Model for Synthetic Riboswitch Function

Our data suggest that these synthetic riboswitches display low background levels of translation in the absence of ligand and robust increases in the presence of the ligand for two reasons. The first is that in the absence of the ligand, the ribosome binding site is paired in such a way that translation is minimized. The second is that ligand binding drives the RNA to a conformation in which the RBS is unpaired. As described by de Smit and van Duin [34–36], sequences with pairing near the ribosome binding site typically display reduced translation rates because the 30S subunit of the ribosome binds most efficiently to single-stranded regions of RNA. Since translation is the slow step in protein production, the position of a pre-equilibrium between RNA structures that have the ribosome binding site paired or unpaired will factor directly into the rate expression for protein synthesis. Sequences that strongly favor pairing of the ribosome binding site are likely to show minimal levels of protein translation in the absence of ligand, since translation is very inefficient when the RBS is paired. Even if such sequences equilibrate with higher-energy structures where the RBS is unpaired (and translation is efficient), the relative population of these high-energy states, and thus the overall translation rate, will be low. The relative stabilities of these structures will dictate the background levels of translation in the absence of the ligand. However, if addition of the ligand shifts the population of mRNAs to a structure(s) where many of the ribosome binding sites are unpaired, this model predicts robust increases in protein translation from these efficiently translated mRNAs.

Figure 6A shows the predicted secondary structures and free energies of the region of the mRNA of clone 8.1 extending from the 5′ end of the aptamer to the 3′ end of the start codon. In the minimum-energy structure ("Off," left), the RBS is paired. In the middle is the lowest energy structure (ΔΔG = +5.5 kcal/mol; see the Supplemental Data for the folding protocol) in which the ribosome binding site is predicted to be unpaired ("On"); this is also the lowest energy structure in which the secondary structure of the aptamer is present. On the right is the same

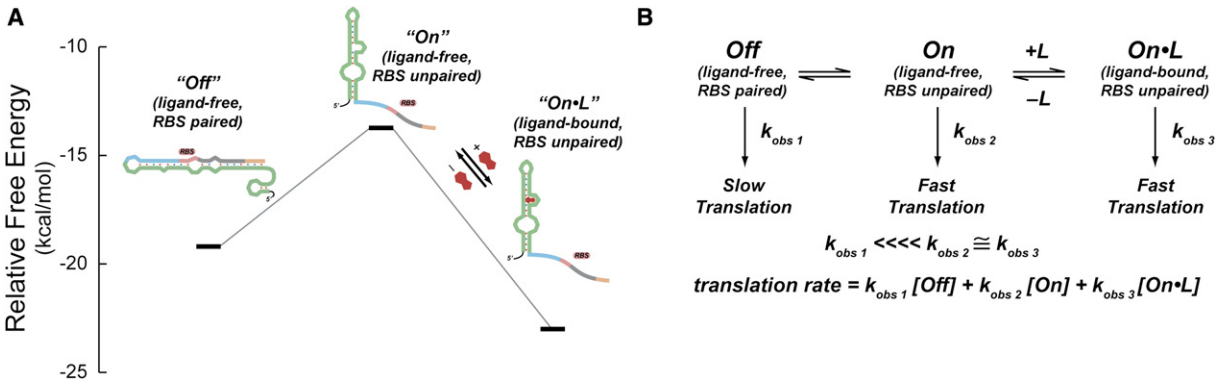

Figure 6. Model for Synthetic Riboswitch Function

(A) Predicted structures and free energies of a portion of the 5′ region of clone 8.1 determined by *mFold*. To represent the free energies of the bound structures, we added the experimentally determined free energy of theophylline binding (−9.2 kcal/mol) to the free energy of the unliganded structure determined by *mFold* since the secondary structure of the mTCT-8-4 aptamer does not change upon ligand binding.

(B) Kinetic model for synthetic riboswitch function. In the "Off" state, the ribosome binding site is paired and translation is slow (represented by the pseudofirst-order rate constant $k_{1\ obs}$, which accounts for ribosome binding and translation). The "Off" state is proposed to be in equilibrium with the ligand-free "On" state, in which the ribosome binding site is unpaired and translation is relatively fast (represented by the pseudofirst-order rate constant $k_{2\ obs}$, which is much greater than $k_{1\ obs}$). Since this state is not highly populated, background translation is minimal. Addition of ligand (L) shifts the equilibrium to the ligand-bound "On•L" state, in which the ribosome binding site is unpaired and translation is relatively fast (represented by the pseudofirst-order rate constant $k_{3\ obs}$, which is also much greater than $k_{1\ obs}$).

calculated fold, but the free energy has been lowered by 9.2 kcal/mol, the experimentally determined free energy of theophylline binding ("On•L").

An equilibrium model predicts that in the absence of the ligand, the majority of the RNA population adopts the "Off" structure in which the RBS is paired because the "On" structure is significantly uphill in free energy. Addition of theophylline provides the thermodynamic driving force to shift the equilibrium toward the "On•L" structure. These observations can be represented in the model shown in Figure 6B, where the pseudofirst-order rate constants ($k_{obs\ 1}$, $k_{obs\ 2}$, and $k_{obs\ 3}$) include the rates of ribosome binding and the initiation of translation and the concentration of the 30S ribosomal subunit. The studies of de Smit and van Duin [34–36] predict that the translation efficiency from the "Off" structure, in which the ribosome binding site is paired, would be significantly lower than the efficiency from the structures in which the RBS is unpaired ("On" and "On•L"), thus $k_{obs\ 1} <<< k_{obs\ 2} \approx k_{obs\ 3}$. Thus, in the absence of ligand, the overall translation rate would be low, and the rate would increase in the presence of the ligand.

While this model qualitatively fits our experimental observations, determining accurate rate constants and mRNA concentrations in vivo is very challenging. Nevertheless, we can test elements of the model, such as the ability of RNA structures to equilibrate in the presence of the ligand. Demonstrating that an mRNA transcript can undergo a ligand-inducible shift from an untranslated state into a translated state in vivo would provide evidence for the equilibrium model proposed above and would also argue against a kinetic model, in which the committed step for translation occurs during transcription (i.e., a transcript is translated only if the appropriate ligand is present at the time of transcription) (Figure 7A) [37, 38].

To test these models, *E. coli* that were constitutively transcribing a riboswitch-controlled *lacZ* gene were grown in the absence of theophylline. Theophylline was added and the β-galactosidase expression at various time points was assayed using Miller's method (Figure 7B). β-galactosidase activity (as measured by A_{420}) first appears approximately 200 s after the addition of theophylline and continues to increase at a steady rate. During the 200 s before β-galactosidase activity appears, several events must occur: (1) theophylline must enter the cell, (2) it must bind to the mRNA, (3) it must induce switching, (4) the "switched" mRNA must be translated, and (5) β-galactosidase must fold and become active. Liang et al. have shown β-galactosidase activity in *E. coli* appears within 60 s upon using IPTG to initiate transcription of *lacZ* [39]. Their results show that transcription rate of *lacZ* must be at least 57 nt/s (3420 nt/60 s) and that the subsequent acts of translation, folding, and appearance of β-galactosidase activity must collectively take less than 60 s [39]. After accounting for the 60 s required for translation, folding, and the appearance of enzymatic activity, the fact that we do not observe β-galactosidase activity until 200 s after theophylline addition suggests that collectively theophylline entry, binding, and switching may take up to 140 s. Although this delay could be consistent with a kinetic barrier for interconversion between a nontranslatable mRNA and a translatable mRNA, we cannot yet determine whether such conformational changes are responsible for the switching behavior.

An alternative to a theophylline-driven equilibrium model is a "kinetic" mechanism whereby the fate of an mRNA is determined at the time of transcription. Much of the data thus far can also be explained in terms of a cotranscriptional kinetic model in which the majority of the mRNA pool folds into a non- or weakly translatable

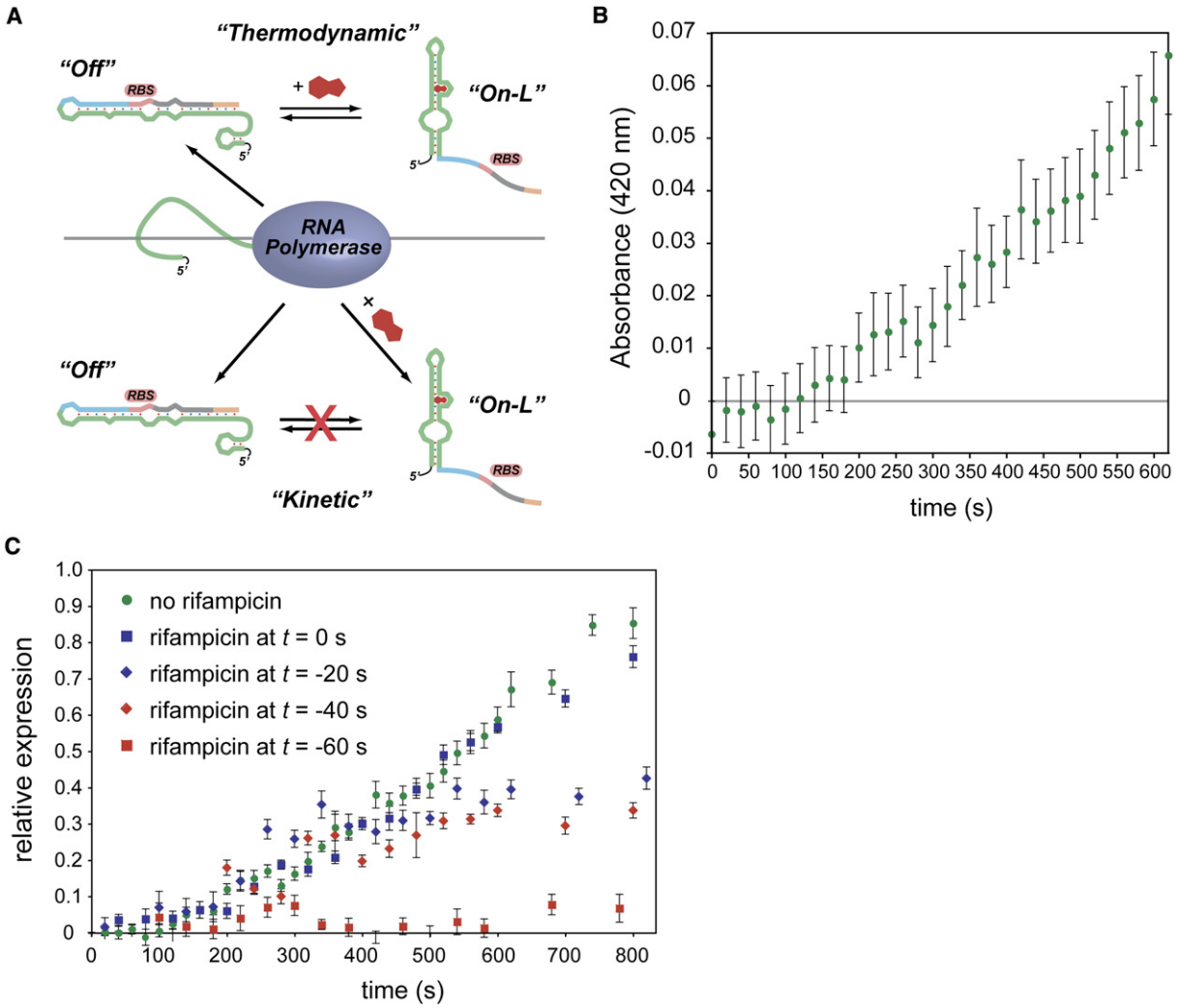

Figure 7. Experiments to Determine Switching Mechanism

(A) Top, a thermodynamic model for riboswitch function. Newly synthesized RNA adopts a folded conformation that prevents translation ("Off"). Addition of theophylline can drive the conformation to the "On-L" conformation. Bottom, a kinetic model in which a newly synthesized RNA adopts either an "Off" or an "On-L" conformation depending on whether ligand is present during transcription; however, the "Off" and "On-L" conformations do not equilibrate.

(B) Time course for appearance of β-galactosidase activity (measured by ONPG hydrolysis) for clone 8.1. Theophylline was added at t = 0 to a final concentration of 3 mM; significant activity appears at 200 s.

(C) Time course for appearance of β-galactosidase activity (measured by ONPG hydrolysis) for clone 8.1. Theophylline was added at t = 0 to a final concentration of 3 mM; rifampicin was added at the times shown in the legend. Rifampicin does not eliminate β-galactosidase activity when added up to 40 s prior to the addition of theophylline, supporting the thermodynamic model shown in (A) in which theophylline can induce expression even in the absence of active transcription. Error bars are ± SD.

conformation when theophylline is not present (e.g., the "Off" conformation shown in Figure 7A). If theophylline can induce some fraction of the mRNA to fold into the highly translatable "On-L" conformation, theophylline-dependent activation of translation could occur regardless of whether the translatable and nontranslatable mRNA conformations can interconvert. Given that (1) *lacZ* transcription is fast (~60 nt/s) [39], (2) both mRNA folding and translation occur cotranscriptionally in *E. coli*, (3) the entire "functional unit" of the riboswitch is located within the first 100 transcribed bases, and (4) the proposed "On" and "Off" conformations are both small hairpins, it is extremely likely that the fate (translatability) of an mRNA is determined within seconds of the initiation of transcription. As such, theophylline would have to be present during transcription for β-galactosidase activity to be observed.

To test whether β-galactosidase expression is dependent on the presence of theophylline while transcription is active, we used the antibiotic rifampicin to halt transcription and asked whether subsequent addition of theophylline could induce β-galactosidase expression. Rifampicin prevents RNA polymerase from entering the

elongation phase following the initiation of transcription, but it has no effect on RNA polymerase once it has reached the elongation phase [39]. The quick action of rifampicin (~5 s), coupled with the fast elongation rate of *E. coli* RNA polymerase (~60 nt/sec) [39], suggests that within 10 s of rifampicin addition, no new transcripts are being produced, and any elongating transcripts are ~100 nt beyond the 5′UTR.

Figure 7C shows the theophylline-induced expression of β-galactosidase from cultures harboring switch 8.1 in which rifampicin was added 0, 20, 40, or 60 s prior to the addition of theophylline. While adding rifampicin 20 or 40 s prior to adding theophylline reduces the overall level of β-galactosidase activity, as would be expected from the inhibition of mRNA synthesis, the onset of β-galactosidase activity occurs at the same time (~200 s) as it does in the absence of rifampicin. Assuming that rifampicin acts within 5 s and that *E. coli* RNA polymerase transcribes mRNA at 60 nt/s [39], by adding rifampicin 40 s prior to theophylline, no new transcripts should be synthesized, and any elongating transcripts should be >1800 nt long, far beyond the 100 nt 5′UTR containing the riboswitch, at the time that theophylline is added. Though we cannot easily determine how quickly theophylline enters the cell and reaches an effective concentration, we can consider two limiting scenarios: theophylline enters essentially immediately or theophylline entry is delayed by up to 200 s. If theophylline enters the cell immediately, the cotranscriptional model predicts that β-galactosidase activity should begin to appear within 60 s, which is the time required for translation and folding of β-galactosidase. However, β-galactosidase activity does not appear until nearly 200 s after the addition of theophylline, which argues against the cotranscriptional model if theophylline entry is fast. If theophylline entry is slow (up to 140 s), the cotranscriptional model becomes even less tenable because after 140 s, all transcripts should be completed, regardless of the time of addition of rifampicin. The fact that β-galactosidase activity continues to increase steadily long after the addition of theophylline, and >300 s after the addition of rifampicin (Figure 7C), argues against a cotranscriptional model for the activation of expression and is more consistent with a model in which theophylline induces a conformational change in an existing mRNA.

Since the data support an equilibrium model in which theophylline induces conformational changes, one might expect that ligand-dependent changes could be observed through RNA footprinting studies. To test this possibility, we performed in vivo footprinting experiments by using the SHAPE method developed by Weeks and coworkers [40]. In the SHAPE method, *N*-methylisatoic anhydride (NMIA) modifies the 2′-OH of RNA depending on the local nucleotide flexibility [40]. These modifications of an RNA template cause the enzyme reverse transcriptase to pause and dissociate, leading to truncations at the site immediately 3′ to the modified ribose. By comparing the products of a reverse-transcription reaction against a known sequence standard by gel electrophoresis, one can map the location of modifications, as well as the extent of modification.

We added NMIA to cells harboring a synthetic riboswitch that were grown in the presence or absence of 1 mM theophylline. The modified total RNA was extracted and used as a template for reverse transcription. The transcription reactions were separated by gel electrophoresis and the pausing due to NMIA modification in the presence and absence of ligand was analyzed. Though we observed NMIA modification of RNA that was not present in an untreated control, we were not able to visualize substantial differences in NMIA modification between cells grown in the presence or absence of theophylline.

SHAPE experiments integrate the total RNA population, not just the fraction of mRNAs that are potentially translatable, and it may be difficult to detect a small increase in translatable mRNAs against a large background of nontranslatable mRNAs. To assess the fraction of mRNAs that are translating in vivo, we compared the levels of β-galactosidase expression between a riboswitch-containing construct (clone 8.1) and a construct that shared the same ribosome binding site from 8.1 but lacked the aptamer. The construct lacking the aptamer expressed approximately six times more β-galactosidase than the riboswitch-containing construct grown in the presence of 1 mM theophylline. This is consistent with the idea that theophylline induces translation of ~16% of the transcripts, but that ~84% may remain inactive. Since SHAPE (and other methods that integrate results over entire populations) cannot distinguish between translating and nontranslating mRNAs, the majority of labeled RNA would be expected to be nontranslating, and a small increase in translating RNAs will likely not be visible. While other in vitro experiments such as nuclease digestion under thermodynamically controlled conditions (renaturing, long time scales) might glean further structural information, it is not immediately clear that this information would be directly relevant to the actual behavior inside the cell. Finally, we note that for an "on" switch, there is no a priori reason to assume that full activation of translation is necessary or even desirable for function in either synthetic or natural riboswitches. Indeed, we have observed riboswitches that activate more fully (e.g., the parent switch in Figure 3E), but such switches may face a trade off between leaky expression and high-absolute levels of activation.

Possible Design Implications for Synthetic Riboswitches

While the generality of this equilibrium model will be investigated in future studies, in its current form, the model may provide guidelines for the design of synthetic riboswitches that activate protein translation. In particular, the model suggests that it will be critical to balance the relative free energies of "off state," the ligand-free "on state," and the ligand-bound "on state." Since the energy differences of the ligand-free "on state" and the ligand-bound "on state" are dictated by the free energy of ligand binding to the aptamer, this puts thermodynamic constraints on

potential structures for the "off state." If the free energy of the "off state" is close to that of the ligand-free "on state" (and there is interconversion), background translation in the absence of the ligand will be high. However, if the "off state" is very stable, binding of ligand may not result in a sufficient population of the ligand-bound "on state" to allow translation, even if this state is highly translatable. We anticipate that this model may help guide the design of future synthetic riboswitches.

Possible Evolutionary Implications for Natural Riboswitches

Our results show that beginning with an aptamer sequence that recognizes a small molecule, it is relatively straightforward to create synthetic riboswitches. Indeed, we have discovered a variety of high-performing synthetic riboswitches where both the length and composition of the sequence between the aptamer and the ribosome binding site vary considerably. Furthermore, we have identified two functional motifs where the bases between the aptamer and the ribosome binding site pair to distinct regions of the aptamer, and retrospective analysis of a previously reported synthetic riboswitch reveals a third motif where these bases pair to part of the coding sequence to repress translation in the absence of the ligand. Taken together, our results show that there are many straightforward ways to generate synthetic riboswitches by inserting an aptamer sequence into the 5′UTR of a gene. Via relatively few base changes, the dynamic ranges of these switches can be improved dramatically to give sensitive ligand-dependent genetic control elements. It has been suggested that natural riboswitches may be molecular fossils from an RNA world [7], where RNA sequences that once served to bind to ligands (perhaps as cofactors for RNA catalysis), have since been co-opted for the purpose of gene regulation. Our results show that it is relatively straightforward on the laboratory timescale to convert an aptamer into a synthetic riboswitch. We suggest that by using similar mechanisms, nature may have evolved riboswitches from pre-existing aptamers.

SIGNIFICANCE

Riboswitches are RNA-based genetic control elements that activate or repress gene expression in a small-molecule-dependent fashion without the need for proteins. The ability to create synthetic riboswitches that control gene expression in response to any desired small molecule could enable the development of sensitive genetic screens and selections to detect the presence of small molecules, as well as the creation of designer genetic control elements for studying cell function or conditional gene therapy. Riboswitches require an aptamer to recognize the desired molecule and a mechanism for converting this sensing event into a change in gene expression. While methods to discover aptamers that selectively bind small molecules are well established, methods of converting these aptamers to synthetic riboswitches that function optimally in cells are not. We have presented an automated high-throughput screening method that identifies synthetic riboswitches that show extremely low background levels of gene expression in the absence of ligand and robust increases in expression in the presence of the desired ligand. These synthetic riboswitches not only show superior performance relative to previously reported switches but also compare favorably to natural riboswitches. Sequence information allowed us to propose and test a model for riboswitch function that explains the mechanism of action for these and previously described synthetic riboswitches. Furthermore, we have proposed a simple model for synthetic riboswitches that activate protein translation that may guide future experiments. Our results show that beginning with a known aptamer, it is straightforward to create distinct families of synthetic riboswitches, which is consistent with the idea that natural riboswitches may have evolved from small-molecule-binding aptamers that were retained from an RNA world and subsequently recruited to become genetic regulation elements. We anticipate that this high throughput method will be generally useful for converting in vitro-selected aptamers into robust performing synthetic riboswitches.

EXPERIMENTAL PROCEDURES

General Considerations

All plasmid manipulations utilized standard cloning techniques [41]. All constructs have been verified by DNA sequencing at the NSF-supported Center for Fundamental and Applied Molecular Evolution at Emory University. Purifications of plasmid DNA, PCR products, and enzyme digestions were performed with kits from QIAGEN. Theophylline, *o*-nitrophenyl-β-D-galactopyranoside (ONPG), ampicillin, and chloramphenicol were purchased from Sigma. X-gal was purchased from US Biological. Synthetic oligonucleotides were purchased from IDT. All experiments were performed in *E. coli* TOP10 F′ cells (Invitrogen) cultured in media obtained from EMD Bioscience.

Construction of Randomized Libraries

Libraries were constructed by oligonucleotide-based cassette mutagenesis. Mutagenic primers with degenerate regions were designed to create cassettes with randomized sequences of appropriate lengths between the mTCT8-4 theophylline aptamer [24] and the RBS of the *IS10-lacZ* reporter gene. Full descriptions of the mutagenesis strategies, primer sequences, and plasmid features are available in the Supplemental Data.

Library Screens

Library transformations were plated on large (241 mm × 241 mm) bioassay trays from Nalgene containing LB/agar (300 ml) supplemented with ampicillin (50 μg/ml) and X-Gal (25 mg dissolved in 4.0 ml dimethyl formamide, final concentration 0.008%). Cells were plated to achieve a final density of ~4,000 colonies/plate. Cells were grown for 14 hr at 37°C, followed by incubation at 4°C until blue color was readily visible.

The whitest colonies from each plate were picked by using a Genetix QPix2 colony picking robot and were inoculated in a 96-well microtiter plate (Costar), which contained LB media (200 μl/well) supplemented with ampicillin (50 μg/ml). The plate was incubated overnight at 37°C with shaking (180 rpm). The following day, four 96-well plates (two sets of two) were inoculated with 2 μl of the overnight culture. The first set of plates contained 200 μl LB supplemented with ampicillin (50 μg/ml). The second set of plates contained 200 μl LB

supplemented with both ampicillin (50 μg/ml) and theophylline (0.5 mM). Plates were incubated for approximately 2.5 hr at 37°C with shaking (210 rpm) to an OD_{600} of 0.085–0.14 as determined by a Biotek microplate reader. These values correspond to an OD_{600} of 0.3–0.5 with a 1 cm path length cuvette.

A high-throughput microtiterplate assay for β-galactosidase activity was adapted from previously described methods [29–31]. Cultures were lysed by adding Pop Culture solution (Novagen, 21 μl, 10:1, Pop Culture: lysozyme [4 U/ml]), mixed by pipetting up and down, and allowed to stand at room temperature for 5 min. In a fresh plate, 15 μl of lysed culture was combined with Z buffer (132.25 μl, 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol [pH 7.0]). This was followed by addition of ONPG (29 μl, 4 mg/ml in 100 mM NaH_2PO_4). ONPG was allowed to hydrolyze for approximately 20 min or until faint yellow color was observed. The reaction was quenched by the addition of Na_2CO_3 (75 μl of a 1 M solution). The length of time between substrate addition and quenching was recorded and the OD_{420} for each well was determined. The Miller units were calculated by using the following formula:

$$\text{Miller units} = OD_{420}/(OD_{600} \times \text{hydrolysis time} \times [\text{volume of cell lysate}/\text{total volume}])$$

Ratios of the Miller units for cultures grown in the presence or absence of theophylline represent an "activation ratio." The initial pool of candidate switches comprised clones that showed an activation ratio of greater than 2.0 in two separate determinations. Candidates that did not display a minimum activity in the presence of theophylline (an $OD_{420} \geq 0.04$) in either determination were eliminated from consideration. As a final check, we visually inspected the data for aberrations, such as cultures that grew especially slowly or quickly (as represented by OD_{600}), or for cultures with dramatically different results between the two plates. Clones that were identified as potential switches were subcultured and assayed as previously described.

Rifampicin Assay

An aliquot of a saturated overnight culture (500 μl) was used to inoculate 50 ml of LB supplemented with ampicillin (50 μg/ml). The fresh culture was grown with shaking (300 rpm, 37°C) to an OD_{600} of 0.4–0.5 (approximately 2.75 hr). The culture was split into two 16.6 ml fractions. Theophylline (50 mg/ml in DMSO, warmed to 37°C) and rifampicin (50 mg/ml in DMSO) were added as appropriate to one of the cultures such that the final concentrations were 3.0 M theophylline, 250 μg/ml rifampicin; the other culture was used as a control. Aliquots of the cultures (4 × 100 μl each) were collected at 20, 40, or 100 s intervals by using a Freedom-EVO liquid-handling system (TECAN) and were transferred to individual wells of a 96-well microtiter plate that were prefilled with Pop Culture (10 μl). A 30 μl aliquot from each well of lysed cells was added to 119 μl of Z buffer in a microtiter plate (see above). ONPG (29 μl, 4 mg/ml in 100 mM NaH_2PO_4) was added and was allowed to hydrolyze for 1 hr at 30°C, followed by quenching with Na_2CO_3 (75 μl of a 1 M solution). The OD_{420} for each well was determined.

Supplemental Data

Supplemental Data include the sequences of all synthetic riboswitches, experimental details of the NMIA labeling protocol, and details of the computational RNA-folding procedure and are available at http://www.chembiol.com/cgi/content/full/14/2/173/DC1/.

ACKNOWLEDGMENTS

We are grateful to the Research Corporation, the Seaver Institute, the Arnold and Mabel Beckman Foundation, the National Institutes of Health (R01-GM074070), and Emory University for financial support. S.A.L. is a National Science Foundation Problems and Research to Integrate Science and Mathematics program Fellow. S.K.D. is grateful to Emory University for providing a George W. Woodruff Fellowship. J.P.G. is a Beckman Young Investigator. We thank Pavan Bang for technical assistance, Professor Dale Edmondson for helpful discussions, and Drs. Doug Grubb and Wayne Patrick for critical readings of the manuscript. DNA sequencing was performed at the National Science Foundation-supported Center for Fundamental and Applied Molecular Evolution (NSF-DBI-0320786).

Received: May 3, 2006
Revised: November 10, 2006
Accepted: December 13, 2006
Published: February 23, 2007

REFERENCES

1. Winkler, W.C., and Breaker, R.R. (2005). Regulation of bacterial gene expression by riboswitches. Annu. Rev. Microbiol. *59*, 487–517.
2. Nudler, E., and Mironov, A.S. (2004). The riboswitch control of bacterial metabolism. Trends Biochem. Sci. *29*, 11–17.
3. Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet. *20*, 44–50.
4. Tucker, B.J., and Breaker, R.R. (2005). Riboswitches as versatile gene control elements. Curr. Opin. Struct. Biol. *15*, 342–348.
5. Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002). Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell *111*, 747–756.
6. Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. (2002). Genetic control by a metabolite binding mRNA. Chem. Biol. *9*, 1043–1049.
7. Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature *419*, 952–956.
8. Winkler, W.C., Cohen-Chalamish, S., and Breaker, R.R. (2002). An mRNA structure that controls gene expression by binding FMN. Proc. Natl. Acad. Sci. USA *99*, 15908–15913.
9. Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., and Breaker, R.R. (2003). An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev. *17*, 2688–2697.
10. Thore, S., Leibundgut, M., and Ban, N. (2006). Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. Science *312*, 1208–1211.
11. Desai, S.K., and Gallivan, J.P. (2004). Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J. Am. Chem. Soc. *126*, 13247–13254.
12. Grate, D., and Wilson, C. (2001). Inducible regulation of the *S. cerevisiae* cell cycle mediated by an RNA aptamer-ligand complex. Bioorg. Med. Chem. *9*, 2565–2570.
13. Harvey, I., Garneau, P., and Pelletier, J. (2002). Inhibition of translation by RNA-small molecule interactions. RNA *8*, 452–463.
14. Suess, B., Fink, B., Berens, C., Stentz, R., and Hillen, W. (2004). A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res. *32*, 1610–1614.
15. Suess, B., Hanson, S., Berens, C., Fink, B., Schroeder, R., and Hillen, W. (2003). Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res. *31*, 1853–1858.
16. Werstuck, G., and Green, M.R. (1998). Controlling gene expression in living cells through small molecule-RNA interactions. Science *282*, 296–298.
17. Roth, A., and Breaker, R.R. (2004). Selection in vitro of allosteric ribozymes. Methods Mol. Biol. *252*, 145–164.

18. Buskirk, A.R., Landrigan, A., and Liu, D.R. (2004). Engineering a ligand-dependent RNA transcriptional activator. Chem. Biol. *11*, 1157–1163.

19. Kim, D.S., Gusti, V., Pillai, S.G., and Gaur, R.K. (2005). An artificial riboswitch for controlling pre-mRNA splicing. RNA *11*, 1667–1677.

20. Bayer, T.S., and Smolke, C.D. (2005). Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat. Biotechnol. *23*, 337–343.

21. Isaacs, F.J., Dwyer, D.J., Ding, C., Pervouchine, D.D., Cantor, C.R., and Collins, J.J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. Nat. Biotechnol. *22*, 841–847.

22. Isaacs, F.J., Dwyer, D.J., and Collins, J.J. (2006). RNA synthetic biology. Nat. Biotechnol. *24*, 545–554.

23. Isaacs, F.J., and Collins, J.J. (2005). Plug-and-play with RNA. Nat. Biotechnol. *23*, 306–307.

24. Jenison, R.D., Gill, S.C., Pardi, A., and Polisky, B. (1994). High-resolution molecular discrimination by RNA. Science *263*, 1425–1429.

25. Zimmermann, G.R., Jenison, R.D., Wick, C.L., Simorre, J.P., and Pardi, A. (1997). Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. Nat. Struct. Biol. *4*, 644–649.

26. Zimmermann, G.R., Shields, T.P., Jenison, R.D., Wick, C.L., and Pardi, A. (1998). A semiconserved residue inhibits complex formation by stabilizing interactions in the free state of a theophylline-binding RNA. Biochemistry *37*, 9186–9192.

27. Zimmermann, G.R., Wick, C.L., Shields, T.P., Jenison, R.D., and Pardi, A. (2000). Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. RNA *6*, 659–667.

28. de Boer, H.A., Comstock, L.J., and Vasser, M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. Proc. Natl. Acad. Sci. USA *80*, 21–25.

29. Jain, C., and Belasco, J.C. (2000). Rapid genetic analysis of RNA-protein interactions by translational repression in *Escherichia coli*. Methods Enzymol. *318*, 309–332.

30. Bignon, C., Daniel, N., and Djiane, J. (1993). Beta-galactosidase and chloramphenicol acetyl transferase assays in 96-well plates. Biotechniques *15*, 243–244.

31. Griffith, K.L., and Wolf, J.R.E. (2002). Measuring [beta]-galactosidase activity in bacteria: cell growth, permeabilization, and enzyme assays in 96-well arrays. Biochem. Biophys. Res. Commun. *290*, 397–402.

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

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

34. de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc. Natl. Acad. Sci. USA *87*, 7668–7672.

35. de Smit, M.H., and van Duin, J. (1994). Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. J. Mol. Biol. *244*, 144–150.

36. de Smit, M.H., and van Duin, J. (1994). Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. J. Mol. Biol. *235*, 173–184.

37. Wickiser, J.K., Cheah, M.T., Breaker, R.R., and Crothers, D.M. (2005). The kinetics of ligand binding by an adenine-sensing riboswitch. Biochemistry *44*, 13404–13414.

38. Wickiser, J.K., Winkler, W.C., Breaker, R.R., and Crothers, D.M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol. Cell *18*, 49–60.

39. Liang, S.T., Ehrenberg, M., Dennis, P., and Bremer, H. (1999). Decay of rplN and lacZ mRNA in *Escherichia coli*. J. Mol. Biol. *288*, 521–538.

40. Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005). RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). J. Am. Chem. Soc. *127*, 4223–4231.

41. Sambrook, J., and Russell, D.W. (2001). Molecular cloning: a laboratory manual, Third Edition, *Volume 1* (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).

