Supplementary Information

Automated Physics-based Design of Synthetic Riboswitches from Diverse RNA Aptamers

Amin Espah Borujeni¹, Dennis M. Mishler³†, Jingzhi Wang³, Walker Huso¹, and Howard M. Salis¹,²,*

¹ Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, 16802, USA
² Department of Biological Engineering, The Pennsylvania State University, University Park, PA, 16802, USA
³ Department of Chemistry, Emory University, Atlanta, GA, 30322, USA

† Current address: Department of Molecular Biosciences, Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX, 78712-1639, USA
* The corresponding author may be contacted at salis@psu.edu.

Table of Supplementary Figures

- Supplementary Figure 1: Thermodynamic Equilibrium in a Simplified Intracellular Environment
- Supplementary Figure 2: Schematic of Optimization Algorithm
- Supplementary Figure 3: The Biophysical Model Can Predict the Output of a Combinatorial Library
- Supplementary Figure 4: The Switching Free Energy of 52 Designed Riboswitches
- Supplementary Figure 5: The Biophysical Model Generates Riboswitches with Diverse Sequences
- Supplementary Figure 6: Secondary Structure of a Fluoride Riboswitch in the OFF and ON States
- Supplementary Figure 7: Secondary Structure of a TMR Riboswitch in the OFF and ON States
- Supplementary Figure 8: The Designed Riboswitches Showed Continuous Activation during Growth
- Supplementary Figure 9: Predicted AR_{max} versus Measured AR
- Supplementary Figure 10: Predicted AR_{conc} versus Measured AR for Stable and Metastable Riboswitches
- Supplementary Figure 11: Error Distribution Analysis for Model Predictions
- Supplementary Figure 12: Predicted versus Measured Translation Initiation Rates and Activation Ratios for Individual Aptamers
- Supplementary Figure 13: Correlation Between Errors in Predicted Translation Initiation Rate ON and Translation Initiation Rate OFF
- Supplementary Figure 14: Model Errors versus Individual Gibbs Free Energy Terms
- Supplementary Figure 15: The Measured Activities of All 62 Designed Riboswitches
- Supplementary Figure 16: Secondary Structure of a Dopamine Binding Aptamer
- Supplementary Figure 17: Relative Gene Expression from Different Promoters and Plasmid Copy Numbers
- Supplementary Figure 18: Comparison of Riboswitch Growth Rate and Expression Levels at Different mRNA Levels
- Supplementary Figure 19: A “Perfect” Riboswitch to Test the Limits of Sensing
- Supplementary Figure 20: Accepted and Unaccepted Types of Pseudoknot Base Pairings in the Biophysical Model

Table of Supplementary Notes

- Biophysical Model of Translation Initiation
- Biophysical Model of Riboswitches to Control Translation Initiation
- Thermodynamic Minimization Model to Capture the Effect of Ligand and mRNA Concentration on Riboswitch’s Function
- Riboswitch Calculator: Automated Design of Synthetic Riboswitches
- Incorporation of Pseudoknot Structural Constraint into Riboswitch Calculator
- Combinatorial Library and High Throughput Screening Method
- Volume and Intracellular Concentration of mRNA and Ligands
- The mRNA and Ligand Level Comparison Between cell-free and in vivo Systems for mRNA Dose Response Experiments
- Designing a “Perfect” Riboswitch
- Examples of Data Analysis (DNT, Dopamine, and Thyroxine Riboswitches)
Supplementary Figure 1: Thermodynamic Equilibrium in a Simplified Intracellular Environment. Cartoon shows a simplified cellular environment and the thermodynamic equilibrium between free mRNA (state 1 in Figure 1B), mRNA-ligand complex (state 3 in Figure 1B), and free ligand. Translation is initiated from free mRNA (state 1) and ligand-bound mRNA (state 3) but with different rates. The mRNA at both state 1 and state 3 has similar volume. The Gibbs free energy of each complex is shown. \( \Delta G_{\text{mRNA}} \) and \( \Delta G_{\text{ligand}} \) are calculated from Equation S11 and S12.
Supplementary Figure 2: Schematic of Optimization Algorithm. The inputs to the optimization algorithm are the aptamer’s sequence and structure and the protein coding sequence. Additional constraints such as restriction site sequences and positions can be defined by the user. The optimization starts with the randomized pre- and post-aptamer sequences to generate an initial pool of riboswitches. Next, riboswitches will be selected for random mutations based on predefined probabilities, followed by fitness evaluation using the biophysical model calculations. Only the riboswitches that meet the objective function requirements will be saved for further recombination step. The objective functions can be maximizing $AR_{\text{actual}}$, minimizing $\Delta \Delta G_{\text{mRNA}}$, or reaching targeted translation rates in OFF and ON states. Finally, the selected riboswitches will be sent back to the population for the next round of mutations. At the end of each cycle, a few riboswitches with the highest fitness values are saved as the elite members, and the cycles are repeated until the optimization converges to the optimum elite solutions.
Supplementary Figure 3: The Biophysical Model Can Predict the Output of a Combinatorial Library. (A) Lynch et al. 2007 generated four libraries with different sizes, in which the linker region between the theophylline binding aptamer and the SD region was randomized. The libraries were then screened using high-throughput methods [1] (see supplementary notes for more details). (B) According to our biophysical model calculations, the probability of finding a riboswitch with large predicted AR_max dropped significantly across the entire library variants as the library size increased. (C) The number of experimentally identified riboswitches for each library is compared with the number of potential riboswitches that have large predicted maximum activation ratios (AR_max > 10). (D) The in vivo β-galactosidase levels (Miller units) of 20 identified riboswitches from all four libraries [1] are correlated well with the predicted ΔGtotal at both OFF state (white circles) and ON state (orange circle). Pearson R² = 0.64 and p-value = 7x10⁻¹⁰. (E) The measured activation ratios of most of the 20 identified riboswitches are close to the predicted maximum activation ratio (AR_max). The dashed line is the x=y line.
Supplementary Figure 4: The Switching Free Energy of 52 Designed Riboswitches. The switching free energy ($\Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}$) of 52 designed riboswitches using theophylline, fluoride, and TMR aptamers are shown. Light blue bars correspond to metastable riboswitches and the light green bars are the stable riboswitches. The red asterisks indicate the two outliers in Figures 3CD (Theo-42 and Theo-44). The riboswitches are shown according to their order in the Supplemental Data.
Supplementary Figure 5: The Biophysical Model Generates Riboswitches with Diverse Sequences. (A) The diversity of pre- and post-aptamer regions for 52 designed riboswitches using theophylline, fluoride, and TMR aptamers is shown. The plots are created by the WebLogo software [2]. The last three nucleotide of the post-aptamer region is the ATG start codon. The expected SD region is shown. (B, C, and D) The diversity of pre- and post-aptamer for individual riboswitches with theophylline, TMR, and fluoride aptamers is shown. The first five nucleotides of aptamer are added to the end of pre-aptamer region and the last five nucleotides of aptamer are added to the beginning of post-aptamer. The last three nucleotide of post-aptamer is the ATG start codon. The expected SD region is also depicted. We also calculated the sequence entropy to be 63%, using the pre- and post-aptamer regions of all these 52 designed riboswitches according to formula:

$$\frac{\sum_{i=1}^{N} \left( \sum_{j=A,T,C,G} p_{i,j} \log_2 (p_{i,j}) \right)}{2N},$$

where p is the number of appearance of nucleotides A, T, C, or G in each position i and N is the total number of nucleotide positions.
Supplementary Figure 6: Secondary Structure of a Fluoride Riboswitch in the OFF and ON States. The predicted secondary structures of a designed Fluoride riboswitch (F-8) are depicted for both OFF and ON (constrained) states. (A) In the absence of ligand (OFF state), the free riboswitch mRNA folds with \( \Delta G_{mRNA,OFF} = -46.1 \), while ribosome binds to the mRNA with following energies: \( \Delta G_{mRNA-rRNA,OFF} = -42.7 \), \( \Delta G_{start,OFF} = -1.2 \), \( \Delta G_{spacing,OFF} = 0.7 \) (spacing = 7 nt), and \( \Delta G_{standby,OFF} = 5.4 \). Overall, the net energy change from the ribosome binding is \( \Delta G_{total,OFF} = 8.3 \). All the Gibbs free energies are in Kcal/mol. (B) In the presence of ligand (ON state), the mRNA folds to a constrained structure, leading to different Gibbs free energies between ligand-bound mRNA and ribosome: \( \Delta G_{mRNA,ON} = -29.9 \), \( \Delta G_{mRNA-rRNA,ON} = -36.6 \), \( \Delta G_{start,ON} = -1.2 \), \( \Delta G_{spacing,ON} = 0.7 \) (spacing = 7 nt), and \( \Delta G_{standby,ON} = 0 \). This increases the net energy release from the ribosome binding to mRNA (\( \Delta G_{total,ON} = -7.24 \)), equivalent of 1090-fold maximum activation in translation initiation rate (\( AR_{max} = \exp(-0.45 \times (-7.24-8.3)) = 1090 \)). However, largely positive \( \Delta \Delta G_{mRNA} = 16.2 \) (\( \Delta \Delta G_{mRNA} + \Delta G_{ligand} = 10.2 \)) lowers the actual activation to 1.5-fold due to metastable binding of ligand to mRNA.
Supplementary Figure 7: Secondary Structure of a TMR Riboswitch in the OFF and ON States. The predicted secondary structures of designed TMR-7 riboswitch are depicted for both OFF and ON (constrained) states. (A) The calculated Gibbs free energies in the OFF state (no ligand) are: $\Delta G_{mRNA,OFF} = -52.5$, $\Delta G_{mRNA-rRNA,OFF} = -53.9$, $\Delta G_{start,OFF} = -1.2$, $\Delta G_{spacing,OFF} = 0$ (spacing = 5 nt), and $\Delta G_{standby,OFF} = 7.8$. Therefore, the total Gibbs free energy change between free mRNA and ribosome-bound mRNA is $\Delta G_{total,OFF} = 5.2$. All the Gibbs free energies are in Kcal/mol. (B) Presence of ligand in the ON state changes the Gibbs free energies between ligand-bound mRNA and ribosome according to: $\Delta G_{mRNA,ON} = -46.3$, $\Delta G_{mRNA-rRNA,ON} = -53.7$, $\Delta G_{start,ON} = -1.2$, $\Delta G_{spacing,ON} = 0$ (spacing = 5 nt), and $\Delta G_{standby,ON} = 0.2$, which leads to much stronger ribosome binding rate ($\Delta G_{total,ON} = -8.4$). Therefore, the maximum activation ratio is $AR_{\text{max}} = \exp(-0.45 \times (-8.4-5.2)) = 456$-fold. Although the binding of ligand to mRNA is stable ($\Delta \Delta G_{mRNA} + \Delta G_{\text{ligand}} = -4.0$), the TMR toxicity limits the amount of ligand availability, and therefore reduces the actual activation ratio to $AR_{\text{actual}} = 6.5$-fold.
Supplementary Figure 8: The Designed Riboswitches Showed Continuous Activation during Growth. (A) The fluorescence per cell (FLPC, Equation 3 in Methods section), measured per time during the serial dilution experiment, is shown for Theo-44 theophylline riboswitch (top) and a no-aptamer control (bottom) upon addition of 2mM Theophylline. (B) The FLPC values for the F-4 fluoride riboswitch (top) and a no-aptamer control (bottom) during the serial dilution experiment and at the presence of 150 mM NaF is shown. The first and second dilutions are marked. The shaded regions are the standard deviation for 2 separate measurements. Averages are the bold lines.
Supplementary Figure 9: Predicted AR$_{\text{max}}$ versus Measured AR. The predicted maximum activation ratios (AR$_{\text{max}}$) are shown against the measured AR for 67 riboswitches using theophylline, fluoride, and TMR binding aptamers. The color representations and ligand concentrations are similar to those in Figures 3CD. Pearson $R^2 = 0.06$ and $p$-value = 0.04, $N = 67$. The dashed line is $x=y$ line.

Supplementary Figure 10: Predicted AR$_{\text{conc}}$ versus Measured AR for Stable and Metastable Riboswitches. The predicted concentration-limited activation ratios (AR$_{\text{conc}}$) are shown against the measured AR for 67 riboswitches using theophylline, fluoride, and TMR binding aptamers. The riboswitches are divided into (A) stable riboswitches (with $\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}} \leq 0$) and (B) metastable riboswitches (with $\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}} > 0$). The color representations and ligand concentrations are similar to those in Figure 3CD. The dashed lines are $x=y$ line. (A) When riboswitches are stable, the model calculated AR$_{\text{conc}}$ is sufficient to predict their function. Spearman $R = 0.72$, $p$-value = $2.5 \times 10^{-5}$, $N = 27$; Pearson $R^2 = 0.62$, $p$-value = $3.5 \times 10^{-6}$, $N = 25$. The two outliers are Theo15 (green circle) and Theo-42 (brown circle). (B) When the riboswitches have metastable ligand binding, their measured AR largely deviates from the predicted AR$_{\text{conc}}$. Spearman $R = 0.27$, $p$-value = 0.09, $N = 40$; Pearson $R^2 = 0.04$, $p$-value = 0.22, $N = 40$. 
Supplementary Figure 11: Error Distribution Analysis for Model Predictions. (top row) The probability plot for normal distribution of errors in predicted translation initiation rates (OFF and ON) and AR_{actual} showed that the errors in AR_{actual} are normally distributed, while the errors in individual translation initiation rates are less normal. Overall, there are 8 outliers out of 67 riboswitches that their error in AR_{actual} is not normal; three of them have over-predicted AR_{actual} and five of them are under-predicted. The red-dashed line joins the first and third quartiles of sample population. (bottom row) The histograms show the distribution of error, the red lines are the corresponding Gaussian distributions. The mean ($\mu_{\log}$) and standard deviation ($\delta_{\log}$) for the log-normal distributions are: 0.14 and 2.07 (for translation initiation rate OFF), 0.09 and 1.74 (for translation initiation rate ON), and -0.05 and 1.27 (for AR_{actual}). Therefore, the mean ($\mu$) and standard deviations ($\delta$) in normal scale are: 1.15 and 7.94 (for translation initiation rate OFF), 1.10 and 5.16 (for translation initiation rate ON), and 0.96 and 2.44 (for AR_{actual}), which are calculated according to formula: $\mu = \exp(\mu_{\log})$ and $\delta = \exp(\mu_{\log}) \times [\exp(\delta_{\log}) - 1]$. 
Supplementary Figure 12: Predicted versus Measured Translation Initiation Rates and Activation Ratios for Individual Aptamers. (top row) The predicted translation initiation rates for individual aptamers (45 theophylline riboswitches, 12 fluoride riboswitches, and 10 TMR riboswitches) are compared to their measured fluorescence or luminescence levels in both OFF and ON states. All the luminescence levels are shown as one tenth of the measured values (arbitrary units). The dashed line is x=y line. Spearman R = 0.65, p-value = 3.2x10^{-12}, N = 90; Pearson R^2 = 0.40, p-value = 3.0x10^{-11}, N = 90 (theophylline riboswitches). Spearman R = 0.21, p-value = 0.32, N = 24; Pearson R^2 = 0.09, p-value = 0.15, N = 24 (fluoride riboswitches). Spearman R = 0.66, p-value = 0.002, N = 20; Pearson R^2 = 0.43, p-value = 0.002, N = 20 (TMR riboswitches). (bottom row) The predicted AR_{actual} is compared to the measured AR for individual aptamers. The black and yellow circles in theophylline plot are the over-predicted outliers (Theo-41, Theo-42, and Theo-44) and under-predicted outliers (Theo-13, Theo-14, Theo-15, Theo-26, and Theo-27), respectively. The dashed line is x=y line. Spearman R = 0.39, p-value = 0.009, N = 45; Pearson R^2 = 0.55, p-value = 1.8x10^{-7}, N = 37 (theophylline riboswitches). Spearman R = -0.18, p-value = 0.57, N = 12; Pearson R^2 = 0.009, p-value = 0.77, N = 12 (fluoride riboswitches). Spearman R = 0.53, p-value = 0.12, N = 10; Pearson R^2 = 0.62, p-value = 0.007, N = 10 (TMR riboswitches).
Supplementary Figure 13: Correlation Between Errors in Predicted Translation Initiation Rate ON and Translation Initiation Rate OFF. Error in predicted translation initiation rate (OFF) is highly correlated with the error in the predicted translation initiation rate (ON) for 88% of the riboswitches. Spearman R = 0.78, p-value < 10^{-30}, N = 67; Pearson R^2 = 0.87, p-value = 3.6x10^{-27}, N = 59 (without 8 outliers). The color code is similar to Figure 3CD. The diagonal dashed line is x=y line. The blue and green circles in the top-left box are the under-predicted theophylline riboswitches with measured AR > 150-fold (Theo-13, Theo-14, Theo-15, Theo-26, and Theo-27). All these five riboswitches have lower OFF expression than their predicted translation initiation rate OFF. The three brown circles in the bottom-right box are theophylline riboswitches that are over-predicted by the model (Theo41, Theo-42, and Theo-44). All these three riboswitches have higher OFF expression than their predicted translation initiation rate OFF.
Supplementary Figure 14: Model Errors versus Individual Gibbs Free Energy Terms. (top row) Errors in predicted activation ratios are compared to the difference in individual Gibbs free energy terms between OFF and ON states as calculated by the biophysical model for all 67 riboswitches: $\Delta \Delta G_X = \Delta G_{X,ON} - \Delta G_{X,OFF}$. These plots show that $\Delta G_{\text{mRNA}}$, $\Delta G_{\text{mRNA-rRNA}}$, and $\Delta G_{\text{standby}}$ are changing greatly between OFF and ON states, although they don’t show any correlation with the error in activation ratio. The three circles shown in the boxes are the over-predicted theophylline riboswitches (Theo-41, Theo-42, and Theo-44). The orange highlighted box shows that these three outliers have large positive $\Delta \Delta G_{\text{mRNA}}$, which could be responsible for their lower-than-predicted ON expression. (three bottom rows) The first row shows the errors in predicted activation ratios against the calculated Gibbs free energy terms in the OFF state (blue circles) and ON state (red circles). The boxes indicate the three over-predicted theophylline riboswitches (Theo-41, Theo-42, and Theo-44). The second row shows the errors in predicted TIR$_{\text{OFF}}$ (OFF state translation initiation rate) with respect to the calculated Gibbs free energy terms in the OFF state. The third row shows the errors in predicted TIR$_{\text{ON}}$ (ON state translation initiation rate) with respect to the Gibbs free energy terms in the ON state. Overall, there is no significant correlation between individual Gibbs free energy terms and error in AR$_{\text{actual}}$, TIR$_{\text{OFF}}$, or TIR$_{\text{ON}}$. However, the highlighted orange box in $\Delta G_{\text{SD-antiSD}}$ plot shows that a largely negative $\Delta G_{\text{SD-antiSD}}$ could be responsible for the higher-than-predicted OFF expression for the three outlier riboswitches.
Supplementary Figure 15: The Measured Activities of All 62 Designed Riboswitches. (A) in vivo expression levels at (white circles) zero and (orange circles) 2 mM theophylline are shown for 24 theophylline riboswitches that control luciferase expression by using either TCT8-4 aptamer (left section) or its mutated version (right section). Activation ratios are shown (gray bars). These riboswitches are transcribed using a tac promoter on a pUC19 vector. Riboswitches are (M) metastable if their switching free energy exceeds 0 kcal/mol. Otherwise, the ligand binding is stable (S). (B) in vivo fluorescence levels at (white circles) zero and (orange circles) 2 mM theophylline are shown for 6 theophylline riboswitches (with mutated aptamer), controlling mRFP1 expression. These riboswitches are transcribed using a J23100 promoter on a pFTV1 vector. A designed control theophylline riboswitch with minimal predicted activation is shown in black bar. The red asterisks indicate outliers in Figures 3CD (Theo-42 and Theo-44). (C) in vivo fluorescence levels at (white circles) zero and (orange circles) 150 mM fluoride and at (white circles) zero and (orange circles) 1 mM DNT are shown for 12 fluoride riboswitches and 3 DNT riboswitches, both controlling mRFP1 expression. The J23100 and AEB-3 promoters drive the transcription of fluoride and DNT riboswitches on a pFTV1 vector, respectively. (D) cell-free transcription-translation expression levels are shown for 5 dopamine...
riboswitches and 2 thyroxine (T4) riboswitches, both controlling luciferase expression. These riboswitches are transcribed using a tac promoter on a pFTV1 vector. The activity of dopamine riboswitches were measured under 5 mM ascorbic acid and with (white circles) zero and (orange circles) 1 mM dopamine. The thyroxine riboswitches were characterized under 100 mM NaCl, 1.5 mM NaOH, and with (white circles) zero and (orange circles) 150 uM thyroxine. (E) Expression levels are shown for 10 TMR riboswitches in \textit{in vivo} and cell-free transcription-translation assays within the (white circles) absence and (orange circles) presence of 20 µM and 30 µM TMR, respectively. These riboswitches control the luciferase expression and are transcribed using a tac promoter on a pUC19 vector. (F) The \textit{in vivo} measured expression levels of TMR riboswitches were compared to their measurements using cell-free transcription-translation assays, showing proportionality. The top two characterized riboswitch variants for each aptamer are shown in \textbf{Figure 4A}. The riboswitch measurements in all the sub-figures are labeled according to their order in the \textbf{Supplemental Data}. 
Supplementary Figure 16: Secondary Structure of a Dopamine Binding Aptamer. Secondary structure of a dopamine binding aptamer (dopa2/c.1) that was used to generate three dopamine riboswitches is shown here. A proposed pseudoknot base pairing is depicted. A shorter RNA aptamer (dopa1.3/C.3) was also used to generate two additional dopamine riboswitches and is shown in Figure 3B in the main text.

Supplementary Figure 17: Relative Gene Expression from Different Promoters and Plasmid Copy Numbers. (A) The expression of luciferase on four different plasmids with increasing copy numbers is shown. The numbers are normalized with respect to pUC19. In all four plasmids, tac promoter drives the transcription process. (B) The relative transcription rates of three promoters are shown using mRFP1 reporter expression level measurements. The transcription rates are normalized with respect to the J23100 promoter.
**Supplementary Figure 18: Comparison of Riboswitch Growth Rate and Expression Levels at Different mRNA Levels.**

(A) The specific growth rates of empty *E. coli* DH10B cell, the cells with plasmids harboring the no-aptamer control mRNA, and the cells with theophylline and fluoride riboswitches controlling mRFP1 production are shown. The black bar is the empty cell, the white bars are mRNAs with weak AEB-3 promoter, and grey bars are mRNAs with strong LmrA promoter. Overall, cells grow slower in 150 mM fluoride than 2 mM theophylline regardless of the riboswitch mRNA. The LmrA promoter decreased the cell’s growth rate by only 13% for Theo-44 theophylline riboswitch, while decreasing it by 20% for no-aptamer control mRNA. Similarly, the cell’s growth rate using LmrA promoter was reduced by 48% and 20% for F-6 and F-11 fluoride riboswitches, while the no-aptamer control mRNA reduced the growth rate by 30%. (B) The expression levels of fluoride and theophylline riboswitches together with the no-aptamer control is shown using AEB-3 (white bars) and LmrA (grey bars) promoters. In all cases, the expression levels of riboswitch mRNAs are much lower than the no-aptamer control mRNAs, providing evidence that cells with riboswitch mRNAs have not reached the maximum translational capacity.
Supplementary Figure 19: A “Perfect” Riboswitch to Test the Limits of Sensing. (A) Optimization was used to design a “perfect” riboswitch using a hypothetical aptamer, which is shown in its ligand-free and ligand-bound states. (B) Model calculations show the effects of the aptamer's binding free energy (affinity) and maximum ligand concentration on the “perfect” riboswitch’s actual activation ratio to illustrate the best possible expression changes under potential sensing scenarios.

Supplementary Figure 20: Accepted and Unaccepted Types of Pseudoknot Base Pairings in the Biophysical Model. (A) A standard pseudoknot base pairing is depicted. The RNA structure will be divided into two canonical base pairings sub-structures (red and blue lines). (B) A non-standard pseudoknot base pairing (green lines) is shown. The RNA structure cannot be divided into separate canonical sub-structures. The energy of this type of pseudoknot base pairing is not calculated by the model.
Supplementary Notes

1- Biophysical Model of Translation Initiation

Statistical thermodynamic modeling relates the mRNA's translation initiation rate, $r$, to the total Gibbs free energy change between mRNA and initiating 30S ribosomal subunit, $\Delta G_{\text{total}}$, according to Equation S1 [4].

$$r \propto \exp(-\beta \Delta G_{\text{total}})$$  \hfill (S1)

$\beta$ is the apparent Boltzmann coefficient and is measured to be 0.45 mol/kcal [4], which has been further verified in other studies [5] and in broadly different bacterial species [6]. The model assumes a two state binding process where the $\Delta G_{\text{total}}$ is the difference between the Gibbs free energy of the final and initial states, according to Equation S2 (Figure 2A).

$$\Delta G_{\text{total}} = \Delta G_{\text{mRNA:rRNA}} + \Delta G_{\text{start}} + \Delta G_{\text{spacing}} + \Delta G_{\text{standby}} - \Delta G_{\text{mRNA}}$$  \hfill (S2)

In the initial state, free mRNA folds into its minimum Gibbs free energy, $\Delta G_{\text{mRNA}}$. In the final state, mRNA undergoes both unfolding and hybridization steps to bind to the 30S ribosomal subunit and form the 30S initiation complex (30SIC). Multiple factors contribute to the Gibbs free energy of the 30SIC in the final state, including the mRNA folding energy, mRNA:16S-rRNA hybridization energy, mRNA folding geometry and site accessibility, mRNA folding location, tRNA-start codon binding energy, and the conformational flexibility of 30S ribosomal subunit. These energetic components are summarized in four Gibbs free energy terms as presented in Equation S2.

$\Delta G_{\text{mRNA:rRNA}}$ accounts for the hybridization energy between Shine-Delgarno like region (SD) of mRNA and the last nine nucleotide of 3'-end of 16S-rRNA (antiSD). It also includes the mRNA refolding energy at regions upstream from SD and downstream from the ribosomal footprint. Based on unpublished data, the ribosomal footprint covers the first 13nt of the coding section. The mRNA region within the ribosomal footprint remains unfolded, as it is loaded into the mRNA channel of 30S subunit. The Gibbs free energy of RNA folding and RNA/RNA hybridization was calculated using the ViennaRNA suite, version 1.8.5 [7]. $\Delta G_{\text{start}}$ is the hybridization energy of tRNA^{fMet} anti-codon loop to the mRNA's start codon at the P-site.
ΔG_{spacing} is the Gibbs free energy penalty for the non-optimal spacing between start codon and the SD-antiSD hybridization region on mRNA (optimal spacing is five nucleotides [8]). The 30S subunit can sustain either a compressed or a stretched state to accommodate different length of spacing region, but with an input of spacing energy penalty that was measured previously [4]. The final energy term, ΔG_{standby}, relates the RNA folding geometry and its site accessibility to the efficiency of mRNA binding to the 30S platform [9].

The 30S platform binds to the mRNA's standby site upstream from the SD prior to hybridization of SD to antiSD. The positively charged surface of 30S platform prefers single stranded regions of mRNA for binding. A standby site at 5' UTR of mRNA can fold into multiple RNA structures, thus its site accessibility for binding to the 30S platform depends solely on the shape and geometry of RNA structures. According to our previous biophysical modeling of standby site [9], the standby site can be divided into multiple standby site modules, where each module contains one hairpin and two adjacent single stranded binding sites. The height of the hairpin and the length of distal and proximal binding sites are denoted as H, D, and P, respectively. We found that the available RNA surface area, As, determined from formula: \( As = 15 + D + P - H \), controls the binding of standby site modules to 30S platform. The 30S platform can sustain a distorted conformational state to accommodate a less accessible standby site module with low As, but with an input of energy penalty ΔG_{distortion}. The ΔG_{distortion} was measured to have an inverse quadratic relationship with As. While 30S platform accommodates different standby site modules, it also minimizes its binding energy penalty to mRNA by partially unfolding RNA structures. This would increase the available RNA surface area and reduces ΔG_{distortion}, but at the expense of inputting unfolding free energy ΔG_{unfolding}. Finally, we found that 30S platform can bind to distant upstream standby site modules and slide on the mRNA surface toward the initiation region. It pushes aside the RNA structures that impede the sliding motion with an input of energy penalty ΔG_{sliding}, which is a function of the size of impeding structures. Overall, the summation of these three Gibbs free energy penalties (Equation S3) represents the ΔG_{standby} for each standby site module. Importantly, we have shown that 30S platform follows the thermodynamic minimization principle and selects the standby site module with the minimum ΔG_{standby} for binding to the mRNA.

\[
\Delta G_{\text{standby}} = \Delta G_{\text{distortion}} + \Delta G_{\text{unfolding}} + \Delta G_{\text{sliding}}
\]  

(S3)
2- Biophysical Model of Riboswitches to Control Translation Initiation

The biophysical model of 30S platform’s binding to the structured 5’ UTR allowed us to predict the effect of any structured 5’ UTR on the translation initiation rate. The core component of this model is the ability to predict the translation initiation rate directly from the RNA folding shape and geometry. Importantly, the ligand-bound RNA aptamers are geometrically similar to the RNA structures that appear in the 5’ UTR and theoretically should control the translation initiation rate in the similar fashion. This special feature of the biophysical model allowed us to develop the automated design of synthetic riboswitches to control the translation initiation rate directly from the RNA aptamers. Notably the automated design does not add any additional parameters to the model. The only input for the model is the structural constraints of RNA aptamers after it is bound to the ligand.

The biophysical model of riboswitches assumes a two state process, an “OFF state” and an “ON state”. In the “OFF state” when the ligand is not present inside the cell, mRNA which carries the RNA aptamer sequence in its 5’ UTR, folds into its minimum free energy $\Delta G_{mRNA, OFF}$. Therefore, the 30S subunit binds to the free mRNA with a total Gibbs free energy change $\Delta G_{total, OFF}$ (Figure 2A) and initiates translation with a rate according to Equation S4.

$$r_{OFF} \propto \exp\left(-\beta \Delta G_{total, OFF}\right)$$  \hspace{1cm} (S4)

In the ON state, however, it is assumed that the excess ligand is present inside the cell and all mRNA molecules are stably bound to ligand. Thus, mRNA folds into its free energy state $\Delta G_{mRNA, ON}$, which is subject to structural constraint of the RNA aptamer. This will pose a constraint on 30S subunit binding to the ligand-bound mRNA, where it should compete with the ligand-bound RNA aptamer for proper assembly on the mRNA. This constrained binding requires new binding free energy $\Delta G_{total, ON}$ and results in new translation initiation rate, according to Equation S5. In the ON state, in addition to $\Delta G_{mRNA, ON}$, the $\Delta G_{mRNA:rRNA, ON}$, $\Delta G_{spacing, ON}$, and $\Delta G_{standby, ON}$ can also be perturbed by the structural constraint of RNA aptamer (Figure 2B).

$$r_{ON} \propto \exp\left(-\beta \Delta G_{total, ON}\right)$$  \hspace{1cm} (S5)

The $r_{ON}$ calculated from Equation S5 is the maximum predicted translation initiation rate
for the ON state, because it assumes that there are excess number of ligand inside the cell and all the mRNA molecules exist in the stable ligand-bound state. The in vivo measured expression levels (at both OFF and ON states) of the previously designed theophylline riboswitches [3] were well-correlated with the predicted Gibbs free energies according to Equations S4 and S5 (Figures 2CDE), demonstrating the model’s accuracy at the stable switching and excess ligand condition.

However, the riboswitch activity decreases as ligand concentration drops. Here, by employing a thermodynamic minimization approach (see below, section 3), we can predict the ON state translation initiation rate, \( r_{\text{ON,conc}} \), as a function of total mRNA and ligand concentrations according to Equation S6. This formulation assumes that mRNA and ligand have reached the equilibrium, and their binding is stable.

\[
r_{\text{ON,conc}} \propto \frac{C_{\text{mRNA,1}}}{C_{\text{mRNA, total}}} \exp(-\beta \Delta G_{\text{total, OFF}}) + \frac{C_{\text{mRNA,3}}}{C_{\text{mRNA, total}}} \exp(-\beta \Delta G_{\text{total, ON}})
\]

(S6)

Since the ligand and mRNA molecules have very different molecular sizes (\(V\)) and occupy different excluded volumes inside the cell, we sought to rewrite \( r_{\text{ON,conc}} \) as a function of volume fraction of mRNA and ligand inside the cell. Therefore, by multiplying both numerators and denominators in Equation S6 by \( V_{\text{mRNA}} / (V_s * C_s) \) we can simplify Equation S6 to Equation S7, in which \( \nu_i \) is the volume fraction of species \( i \) and is calculated from formula: \( \nu_i = (V_i * C_i) / (V_s * C_s) \). “s” stands for solvent, which is assumed to be water.

\[
r_{\text{ON,conc}} \propto \frac{V_{\text{mRNA,1}}}{V_{\text{mRNA, total}}} \exp(-\beta \Delta G_{\text{total, OFF}}) + \frac{V_{\text{mRNA,3}}}{V_{\text{mRNA, total}}} \exp(-\beta \Delta G_{\text{total, ON}})
\]

(S7)

Using the molar conservation of ligand, we can find the following relationship for the volume fraction of mRNA:ligand complex (state 3 in Figure 1B in the main text) as a function of the specific volume and the volume fraction of total and free ligand molecules (Equation S8).

\[
\nu_{\text{mRNA,3}} = \frac{V_{\text{mRNA}}}{V_{\text{ligand}}} \left( \nu_{\text{ligand, total}} - \nu_{\text{ligand, free}} \right)
\]

(S8)

By substituting Equation S8 into Equation S7, we can simplify the \( r_{\text{ON,conc}} \) calculation to
Equation S9. In this equation, $\Delta G_{\text{total,OFF}}$ and $\Delta G_{\text{total,ON}}$ are calculated using the biophysical model given the sequence of mRNA and structural constraint of aptamer, while all the volume fractions at equilibrium are calculated using the thermodynamic minimization of the box given the specific volume of mRNA and ligand.

$$r_{\text{ON,conc}} \propto \frac{V_{\text{mRNA,1}}}{V_{\text{mRNA, total}}} \exp(-\beta \Delta G_{\text{total,OFF}}) + \frac{V_{\text{mRNA}}}{V_{\text{ligand}}} \left[ \left( \frac{V_{\text{ligand, total}} - V_{\text{ligand, free}}}{V_{\text{mRNA, total}}} \right) \exp(-\beta \Delta G_{\text{total,ON}}) \right]$$  

(S9)

Finally, the concentration-limited riboswitch’s activation ratio $AR_{\text{conc}}$ ($AR_{\text{conc}} = r_{\text{ON,conc}} / r_{\text{OFF}}$) can be calculated at any given mRNA and ligand concentrations, according to Equation S10, in which $AR_{\text{max}}$ is a function of $\Delta \Delta G_{\text{total}}$, which is the difference between $\Delta G_{\text{total,OFF}}$ and $\Delta G_{\text{total,ON}}$ ($AR_{\text{max}} = \exp(-\beta \Delta \Delta G_{\text{total}})$). This equation is also presented as Equation 1 in the main text. The measured activation ratio of TMR riboswitches at limited ligand concentration (20 uM TMR) were well-correlated with the model predictions according to Equations S10 (Supplementary Figures 10A), whereas the model predictions with excess ligand assumption failed to do so (Supplementary Figures 9).

$$AR_{\text{conc}} = \frac{V_{\text{mRNA,1}}}{V_{\text{mRNA, total}}} + \frac{V_{\text{mRNA}}}{V_{\text{ligand}}} \left[ \left( \frac{V_{\text{ligand, total}} - V_{\text{ligand, free}}}{V_{\text{mRNA, total}}} \right) AR_{\text{max}} \right]$$  

(S10)

The biophysical model of riboswitch is a thermodynamic-based model that predicts the interactions between ligand, mRNA, and ribosome at steady-state. Currently, it does not consider the kinetics of mRNA structural switching or dynamics of mRNA-ligand binding. Also other factors including the transport of ligand across the cell membrane, its diffusion inside the cytosolic environment, and the ligand’s toxicity effect on the cell growth have not been considered in this model.

The thermodynamic calculations for $AR_{\text{max}}$ and $AR_{\text{conc}}$ are under the assumption that, at steady-state, the binding of mRNA molecules to the ligands is stable. During the ligand binding event, the mRNA folding conformation does switch from its minimum free energy at the OFF state to a more positive free energy at the ON state, resulting in a positive energy penalty $\Delta \Delta G_{\text{mRNA}}$, according to Equation S11.
\[ \Delta \Delta G_{\text{mRNA}} = \Delta G_{\text{mRNA,ON}} - \Delta G_{\text{mRNA,OFF}} \quad (\text{S11}) \]

However, this energy penalty can be compensated by the free energy input from the binding of ligand to the mRNA, \( \Delta G_{\text{ligand}} \) (negative value). \( \Delta G_{\text{ligand}} \) can be calculated from the dissociation constant of ligand from RNA aptamer according to the Equation S12.

\[ \Delta G_{\text{ligand}} = RT \ln(K_D) \quad (\text{S12}) \]

R is the universal gas constant and T is the temperature in Kelvin. \( K_D \) is the ligand's dissociation constant. A tightly binding ligand to RNA aptamer results in more negative \( \Delta G_{\text{ligand}} \).

When \( \Delta G_{\text{ligand}} \) can fully compensate the free energy penalty \( \Delta \Delta G_{\text{mRNA}} \), the RNA folding switch can occur at thermodynamic equilibrium and the binding would be stable. In contrast, a weak ligand binding with a less negative \( \Delta G_{\text{ligand}} \) cannot compensate a large free energy penalty, and consequently the transition from the OFF state to the ON state would not be stable. We define this condition as a metastable switching condition, where the chance of remaining bound to ligand is a function of the mRNA's switching free energy. We calculate the switching free energy by summing \( \Delta \Delta G_{\text{mRNA}} \) and \( \Delta G_{\text{ligand}} \). When the switching free energy (\( \Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}} \)) is negative, the mRNA-ligand binding is stable, and when the summation is positive the binding becomes metastable.

When a riboswitch is metastable, we assume that the metastable mRNA-ligand complex will partition into two reversible states, an unbound mRNA and a ligand-bound mRNA, according to the change in their Gibbs free energies. Applying a mRNA conservation mole balance, the fraction of unbound and ligand-bound mRNAs can be written as Equation S13 and S14.

\[
\begin{align*}
\text{mRNA}_{\text{unbound}} & \iff \frac{\Delta G}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))} \\
\text{mRNA}_{\text{metastable-complex}} & = \text{mRNA}_{\text{unbound}} + \text{mRNA} : \text{ligand} \\
\text{fraction}_{\text{unbound}} & = \frac{1}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))} \\
\text{fraction}_{\text{mRNA:ligand}} & = \frac{\exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))} \\
\end{align*}
\]

(S13)  

(S14)

Based on the biophysical model, the actual translation initiation rate in the ON state for a
metastable riboswitch can be calculated using three states: free mRNA, unbound mRNA from the metastable complex, and ligand-bound mRNA from the metastable complex, according to Equation S15. Here, the concentration of free mRNA and metastable complex are $C_{\text{mRNA,1}}$ and $C_{\text{mRNA,3}}$, respectively.

$$r_{\text{ON,actual}} \propto \frac{C_{\text{mRNA,1}}}{C_{\text{mRNA, total}}} (r_{\text{OFF}}) + \frac{C_{\text{mRNA,3}}}{C_{\text{mRNA, total}}} \text{fraction}_{\text{unbound}} (r_{\text{OFF}}) + \frac{C_{\text{mRNA,3}}}{C_{\text{mRNA, total}}} \text{fraction}_{\text{ligand}} (r_{\text{ON}})$$  \hspace{1cm} (S15)

With some mathematical rearrangement, Equation S15 will reduce to Equation S16, and then to Equation S17.

$$r_{\text{ON, actual}} \propto \frac{r_{\text{OFF}} + \left[ \frac{C_{\text{mRNA,1}}}{C_{\text{mRNA, total}}} r_{\text{OFF}} + \frac{C_{\text{mRNA,3}}}{C_{\text{mRNA, total}}} r_{\text{ON}} \right] \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}$$  \hspace{1cm} (S16)

$$r_{\text{ON, actual}} = \frac{r_{\text{OFF}} + r_{\text{ON,conc}} \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}$$  \hspace{1cm} (S17)

The $r_{\text{OFF}}$ is calculated from Equation S4 and $r_{\text{ON,conc}}$ is calculated according to Equation S9. Finally, the actual activation ratio “$\text{AR}_{\text{actual}}$” is calculated as the ratio of $r_{\text{ON,actual}}$ and $r_{\text{OFF}}$ (Equation S18).

$$\text{AR}_{\text{actual}} = \frac{1 + \text{AR}_{\text{conc}} \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}$$  \hspace{1cm} (S18)

Again, with some mathematical rearrangements, the Equation S18 will be simplified to Equation S19, which is also presented as Equation 2 in the main text.

$$\text{AR}_{\text{actual}} = 1 + \frac{\exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))}{1 + \exp(-\beta(\Delta \Delta G_{\text{mRNA}} + \Delta G_{\text{ligand}}))} (\text{AR}_{\text{conc}} - 1)$$  \hspace{1cm} (S19)
3- Thermodynamic Minimization Model to Capture the Effect of Ligand and mRNA Concentration on Riboswitch’s Function

The biophysical model of riboswitches assumes excess ligand for binding to mRNA in the ON state. However, it is possible that the concentration of ligands is not in excess and even lower than the mRNA level. This would affect the model predictions for riboswitch’s function. Here we used a thermodynamic minimization approach to predict the translation initiation rate for diverse mRNAs and ligand levels inside the cell.

A pool of total mRNA transcripts and total ligands that co-exist inside the cell can partition into three different configurations: free mRNA (state 1 in Figure 1B), free ligand, and ligand-bound mRNA complex (state 3 in Figure 1B) (Supplementary Figure 1). Each of these three configurations has specific concentration C, volume V, and Gibbs free energy $\Delta G$. We simplify the intracellular environment into a water-filled, constant-volume box containing the three above species. For simplicity we also assume that the free ribosomes inside the cell are in excess and their number remain constant when binding to the riboswitches. Therefore, the concentration of each species can be estimated using a thermodynamic minimization approach, in which the total Gibbs free energy of the box, $\Delta G_{\text{box}}$ (Equation S20), is minimized subject to total mRNA and ligand molar conservation [10]. We should mention that, when Dirks et al 2007 [10] calculated the mole fraction of each species inside the box, they assumed that the species in the box have similar sizes. That assumption is correct when dealing with multiple strands of RNA where all have roughly similar sizes. However, in the case of ligand and mRNA binding, the size of ligand can be 1000-fold smaller than the mRNA. Therefore, the similar size assumption would not be correct. Therefore, we modified the original derivation in order to incorporate the volume of species in the $\Delta G_{\text{box}}$ formulation.

$$\Delta G_{\text{box}} = \Delta G_{\text{ref}} + \frac{1}{\beta} \sum_{i=1}^{N} v_i \left( \log v_i - \log Q_{v,i} - 1 \right)$$ (S20)

$\Delta G_{\text{ref}}$ is the Gibbs free energy of reference and $\beta$ is the apparent Boltzmann coefficient from [4]. $v_i$ is the volume fraction of each species $i$ and is related to the molar fraction of specie $i$, $x_i$, according to Equation S21.
\[ v_i = \frac{V_i}{V_s} x_i \quad (S21) \]

\( v_i \) is the volume of specie \( i \) and \( V_s \) is the volume of solvent. Here we assumed that the solvent is water and as the reference we set the \( V_s \) to be 1. Thus, by dividing \( V_i \) to \( V_s \), we normalize the volume of species to the volume of water molecules. The molar fractions \( x_i \) is written as Equation S22, where \( C_s \) is the molar concentration of solvent (water molecules), which is 55.14 mol/lit at 37°C.

\[ x_i = \frac{C_i}{C_s} \quad (S22) \]

In Equation S20, \( Q_{v,i} \) is the adjusted partition function of specie \( i \) for its volume, and is calculated according to Equation S23. \( Q_i \) is the partition function of specie \( i \), which is calculated according to Equation S24, where \( \Delta G_i \) is the Gibbs free energy for specie \( i \) (Supplementary Figure 1).

\[ Q_{v,i} = Q_i \left( \frac{1}{v_i} \right) \quad (S23) \]

\[ Q_i = \exp(-\beta \Delta G_i) \quad (S24) \]

The total mRNA and ligand molar conservation can be written as:

\[ x_{\text{mRNA,total}} = x_{\text{mRNA,1}} + x_{\text{mRNA,3}} \quad (S25) \]

\[ x_{\text{ligand,total}} = x_{\text{ligand,free}} + x_{\text{mRNA,3}} \quad (S26) \]

Finally, the 30S ribosomal subunit can bind to the mRNA and initiate translation from two separate configurations; the free mRNA (state 1) using a binding Gibbs free energy \( \Delta G_{\text{total,OFF}} \), and ligand-bound mRNA (state 3) using a binding Gibbs free energy \( \Delta G_{\text{total,ON}} \). Therefore, the translation initiation rate per total mRNA level in the presence of ligand can be calculated according to Equation S9.
4- Riboswitch Calculator: Automated Design of Synthetic Riboswitches

The Riboswitch Calculator offers two design modes for the automated design of riboswitches to control translation initiation: a reverse engineering mode and a forward engineering mode. For both two modes, four inputs are necessary for the calculations: the RNA aptamer’s sequence, the RNA aptamer’s structural constraint upon binding to the ligand (in the “dot-bracket” format), the binding free energy $\Delta G_{\text{ligand}}$ or its equivalent dissociation constant $K_D$ of the ligand to RNA aptamer, and the protein coding section for the gene of interest. Here, we describe these two modes in detail.

Reverse engineering mode: this mode of the Riboswitch Calculator evaluates the performance of a given riboswitch sequence by calculating the following energetic terms: $\Delta G_{\text{total,OFF}}$ for the OFF state, $\Delta G_{\text{total,ON}}$ of the ON state by incorporating the structural constraint of the RNA aptamer, the $\Delta G_{\text{ligand}}$ calculated from the $K_D$ of ligand to RNA aptamer (Equation S12), and $\Delta \Delta G_{\text{mRNA}}$ according to Equation S11. Finally the translation initiation rate at OFF and ON states (Equation S4 and S5) and as a result the riboswitch’s maximum activation ratio ($AR_{\text{max}}$) will be calculated. It also predicts the concentration-limited activation ratio ($AR_{\text{conc}}$) and actual activation ratios ($AR_{\text{actual}}$) using the given mRNA and ligand concentration and the calculated $\Delta \Delta G_{\text{mRNA}}$ (Equations S10 and S19). This mode of Riboswitch Calculator has been validated using 15 synthetic theophylline riboswitches that controlled the translation of luciferase gene [3] (Figure 2CDE).

Forward engineering mode: this mode of the Riboswitch Calculator is able to design riboswitch sequences from any given RNA aptamer sequence in order to control translation initiation of a gene of interest with large dynamic range. To design the riboswitch, the input RNA aptamer sequence will be located within the 5’ UTR of the gene of interest and the sequences upstream (pre-aptamer) and downstream (post-aptamer) of the RNA aptamer in the 5’ UTR will be carefully optimized to achieve the design objectives. For cloning purposes, proper restriction sites can be inserted in the designed sequences. Importantly, since the DNA sequence space to design the pre- and post-aptamer regions is substantially large, we use an optimization algorithm to search for the functional riboswitches (Figure 3A, Supplementary Figure 2). The optimization algorithm starts with an initial sequence for pre-aptamer and post-aptamer regions and continues with rounds of mutations (substitution, insertion, and deletion, each with a pre-defined chance) until reaching the design objectives. At each mutation, the core biophysical model of the Riboswitch Calculator evaluates the $\Delta G_{\text{total,OFF}}$, $\Delta G_{\text{total,ON}}$, $\Delta G_{\text{ligand}}$, and $\Delta \Delta G_{\text{mRNA}}$ by applying the structural constraint of the input RNA aptamer. Mutations that meet the specified
objective function requirements such as minimizing the $\Delta \Delta G_{\text{mRNA}}$, maximizing the dynamic range of riboswitch’s function ($AR_{\text{max}}$), or reaching a user-defined OFF or ON expression targets, will be selected for next rounds. This mode of Riboswitch Calculator has been validated using 62 designed riboswitches that bind to theophylline, TMR, fluoride, DNT, dopamine, and thyroxine, and control the translation of luciferase and mRFP1 reporter proteins (Figures 2, 3 and 4).

5- Incorporation of Pseudoknot Structural Constraint into Riboswitch Calculator

The biophysical model of structured 5' UTR was parameterized using canonical secondary structures. Although many SELEX based RNA aptamers have canonical secondary structures, most of natural RNA aptamers contain pseudoknot base pairing [11], such as add-A and xpt-G aptamer [12], met-A aptamer [13], met-H aptamer [14], and c-di-GMP binding 84-Cd aptamer [15]. In order to generalize the Riboswitch Calculator to all types of RNA aptamers, we developed an automated method to calculate the folding energy of the pseudoknot structures based on their given structural constraint. The algorithm takes the pseudoknot structural constraint in the form of “pseudoknots dot and bracket”. It calculates the pseudoknot structure folding energy according to Equation S27.

$$
\Delta G_{PK} = \Delta G_{\text{assembly}} + \Delta G_{\text{coaxial}} + \Delta G_{\text{hyb}} - T \Delta S_{PK}
$$ (S27)

$\Delta G_{\text{assembly}}$ is the free energy released from initiating the pseudoknot base pairing between two hairpins and is set to 1.3 kcal/mol, according to Cao and Chen [16-17]. $\Delta G_{\text{coaxial}}$ is coaxial stacking energy between two stems that have 0 or 1 nt distance. The coaxial stacking energy is a function of closing base pairs and their neighboring base pairs. The related energies for coaxial stacking energy calculations, according to Cao and Chen [16], were taken from the DotKnots algorithm [18-19]. $\Delta G_{\text{hyb}}$ is the free energy contribution of all the base pairs in the pseudoknot structures. To calculate this base pairing energy, the pseudoknot structure will be automatically divided into two canonical structures (Supplementary Figure 20A), where their folding energies are calculated, according to Equation S28, using Vienna RNA suite, version 1.8.5. It is important to note that our model only calculates the folding energy of the standard pseudoknot base pairings, as the energy parameters for the non-standard pseudoknot base pairings are not available yet (Supplementary Figure 20B).
\[ \Delta G_{\text{hyb}} = \Delta G_{\text{stem}} - \Delta G_{\text{loop}} \]  

(S28)

In this equation, \( \Delta G_{\text{stem}} \) is the folding energy of the RNA hairpin and \( \Delta G_{\text{loop}} \) is the energy penalty for closing the loop (as a result of entropy loss), which is equivalent of the free energy of hairpin loop’s closing base pair. On the other hand, the effect of entropy change upon pseudoknot formation is included within \( \Delta S_{\text{PK}} \) (Equation 27). A virtual bond model has been proposed by Cao and Chen [16-17, 20], in which the entropy of pseudoknot formation is related to the length of hairpin stem and hairpin loops within the pseudoknot structure. Therefore, our algorithm first determines the effective length of hairpin stem and hairpin loops. Next, \( \Delta S_{\text{PK}} \) is calculated using the entropy functions provided by Cao and Chen for different types of pseudoknots, such as linear-loop [16-17] and loop-loop [20], and also by using the fitted entropy functions available in DotKnots algorithm [18-19]. \( T \) in Equation S27 is temperature in kelvin unit.

Overall, our automatic method is capable of calculating the folding energy of standard pseudoknot structures just by using their structural constraints. These pseudoknots can contain bulges or mismatches in their hairpin stems, or can have extra hairpin stems on their loop or on their hairpin stems.

6- Combinatorial Library and High Throughput Screening Method

Lynch et al. 2007 [1] generated four different combinatorial libraries (\( N_5 \), \( N_6 \), \( N_7 \), and \( N_8 \)), where they randomized the linker region between theophylline binding aptamer and the SD region of IS10-LacZ fusion protein. Using high-throughput screening, they identified a handful of riboswitches with wide range of activation ratios (Supplementary Figure 3A). Using our developed biophysical model, we sought to predict the function of all the individual members of these libraries and identify which riboswitch is silent and which one is functional. Therefore, we performed our calculations for all the riboswitch variants in the \( N_5 \), \( N_6 \), \( N_7 \), and \( N_8 \) libraries. Interestingly, we found that the chance of finding a functional riboswitch from these libraries dropped dramatically as the library size increased (Supplementary Figure 3B). Moreover, our predictions show that the number of predicted functional riboswitches largely exceeded the number of experimentally identified ones as the library size increased, demonstrating a technological challenges for the high-throughput screening methods (Supplementary Figure 3C). Finally, the in vivo measured \( \beta \)-galactosidase levels (Miller units) at both OFF (zero
Theophylline) and ON (2mM Theophylline) states, for 20 identified and reported riboswitches in these libraries, were correlated with the model predicted $\Delta G_{\text{total}}$ values (Supplementary Figure 3D). The biophysical model was able to accurately predict the maximum activation ratio $\text{AR}_{\text{max}}$ for most of these riboswitches (Supplementary Figure 3E).

7- Volume and Intracellular Concentration of mRNA and Ligands

The average volume of *E. coli* cells is estimated to be $10^{-15}$ liters. As the water molecules occupy 75% of the cellular volume [21], the free volume inside the cell is approximately $10^{-16}$ liters. The volume of one molecule of water is $3 \times 10^{-26}$ liters. Therefore, there are roughly $10^9$ water molecules (solvents) inside the cell, which is much greater than the number of mRNA or ligand inside the cell ($C_s >> C_i$). The size of the luciferase mRNA (around 1800 nucleotides) is almost the same size as the 16S ribosomal RNA. To estimate this mRNA’s molar volume, we use as a reference the 30S ribosome, which has a volume of about $3 \times 10^{-21}$ liters. We then estimate that luciferase mRNA is as large as $10^{-21}$ liters, or 10,000 fold larger than water molecules ($V_{\text{luciferase}}/V_{\text{water}} = 10000$). Similarly, since the mRFP1 mRNA is almost half size of the luciferase gene (900 nucleotides), we estimate that the molar volume of mRFP1 mRNA is around $5 \times 10^{-22}$ liters ($V_{\text{mRFP1}}/V_{\text{water}} = 5000$).

Theophylline has density and molecular weight of 0.63 g/ml and 180.2 g/mol, respectively. The volume of one molecule of theophylline is around $5 \times 10^{-25}$ liters, which is around 10-fold larger than water molecules ($V_{\text{Theophylline}}/V_{\text{water}} = 10$). Similarly, as TMR has two additional aromatic rings compared to theophylline, we assume that its molar volume is twice as big as theophylline and 20-fold larger than water molecules ($V_{\text{TMR}}/V_{\text{water}} = 20$). Finally, we assume that the elemental fluoride ion has the same molar volume as water molecule ($V_{\text{Fluoride}}/V_{\text{water}} = 1$). These estimates can be found in Supplementary Data.

The luciferase gene was transcribed from pUC19 vector and mRFP1 from pFTV1 vectors. The pUC19 vector has around 75 copies at 37°C [22] and pFTV1 (ColE1 based) has between 40-60 copies inside *E. coli* Top10 or DH10B [23]. The tac promoter drives the production of luciferase gene on pUC19, while J23100 promoter, which is slightly weaker than tac promoter, drives the transcription of mRFP1 gene on pFTV1. However, the longer length of luciferase gene compared to mRFP1 may lead to a lower transcription efficiency for luciferase gene due to higher chance of RNA polymerase dissociation at longer distances. Therefore, we estimate that there are around 5,000 and 10,000 mRNA molecules inside the cell carrying the luciferase and mRFP1 genes, respectively. These estimates are equivalent of ~70 and ~170
copies of mRNA per pUC19 and pFTV1 plasmid DNAs, respectively, which fall in the range of mRNA copies generated from other natural promoters across genome [24-25]. Transferring the luciferase expression cassette (which includes the tac promoter, 5' UTR, and luciferase gene) to other plasmid vectors with different copy numbers (pBAC, p15A, and pFTV1) results in different range of mRNA levels. Based on our measurements, we estimated that mRNA molecules for luciferase gene are present in around 170, 3730, and 3130 copies under pBAC, p15A, and pFTV1 plasmids, respectively (Supplementary Figure 17A). In addition, the mRNA levels for mRFP1 gene are estimated to be 1,000 and 50,000 molecules when using either the AEB-3 or LmrA promoters on pFTV1 plasmid, respectively (Supplementary Figure 17B). Finally, we assume that the intracellular ligand concentration is almost 100-fold less than the extracellular ligand concentrations. For the case of theophylline, this is slightly higher than its reported intracellular concentration [26]. However, for different ligands, this estimated constant ratio can be different, depending on their partition coefficient value across the cell membrane.

8- The mRNA and Ligand Level Comparison Between cell-free and in vivo Systems for mRNA Dose Response Experiments

For the mRNA dose response assays (Figure 5B), 30 uM TMR in a 10 ul in vitro S30 extract reaction is equivalent of around 2x10^{14} TMR molecules. Addition of a given concentration of DNA (A ng/ul) to the test tube results in 2xAx10^{8} DNA molecules inside the reaction (the DNA plasmid has around 4000 base pairs and the molecular weight of one DNA base pair is 660 g/mol). If we assume that the transcription process within the E. coli S30 extract produces on average 10 mRNA transcripts per DNA template, we will have 2xAx10^{10} mRNA molecules inside the reaction mixture. In our cell-free transcription-translation assay, A varied from 0.1 to 100. Therefore, the ratio of total ligand (2x10^{14}) to total mRNA (2xAx10^{10}) varies from 10^{5} to 10^{2}. Therefore, we can approximate that in all cases the ligand is present in excess, and the number of free ligand is almost equal to the total ligand level (C_{ligand,total} \approx C_{ligand,free}). Using the mass action kinetic formulation and the dissociation constant for TMR aptamer, we obtain the ratio of free mRNA to the ligand bound mRNA, according to Equation S29.

\[
K_D = \frac{C_{\text{ligand, free}}}{C_{\text{mRNA, free}}} \approx \frac{C_{\text{ligand, total}}}{C_{\text{mRNA, ligand}}} \tag{S29}
\]

With 30uM total TMR and K_{D} of 60 nM, the concentration of mRNA-ligand complex is 500-fold
larger than free mRNA, hence $C_{mRNA, total} \sim C_{mRNA-ligand}$. This explains the increase in activation ratio as the mRNA level increases (Figure 5A, dilute volume). It is noteworthy that addition of mRNA molecules does not reduce the free volume in the test tube (Figure 5A, dilute volume).

In contrast to the cell-free transcription-translation reaction, the mRNA and ligand molecules inside the cell are present in very limited amounts. As explained in previous section, the number of transcribed mRNA molecules from pUC19 plasmid is roughly 5,000 molecules. In addition, assuming that the intracellular concentration of TMR is around 0.2 uM, there are roughly 120 molecule of TMR present inside the cell, which is in the same order of magnitude as the mRNA molecules. Therefore, to determine the fraction of ligand-bound mRNAs we need to use thermodynamic minimization approach shown in Equation S20. Interestingly, since the total volume of cell is limited, increasing the number of mRNA molecules reduces the free volume, which further impacts the mRNA-ligand binding fractions (Figure 5A, crowded volume).

Overall, our thermodynamic minimization calculation is able to predict the fraction of free-mRNA, free-ligand, and mRNA-ligand complex as the total mRNA level increases. More importantly, this model can correctly predict the decrease in activation ratio as the mRNA level increases (Figure 5C).

9- Designing a “Perfect” Riboswitch

By combining the biophysical model of riboswitch function together with the optimization algorithm, we investigated whether we could design a so-called “perfect” riboswitch that used a best-possible, hypothetical aptamer to bind a ligand and maximally regulate gene expression. Using this design, we investigated how the maximum available ligand concentration and the aptamer's binding affinity determined the best-possible sensing scenarios. We chose a short 22nt hypothetical aptamer sequence (Figure 6, Supplementary Figure 19A) and highly optimized its pre-aptamer, post-aptamer, and coding section to generate a “perfect” riboswitch that exhibited an extremely large activation ratio ($AR_{max} = 16000$), but with a large switching free energy, $\Delta \Delta G_{mRNA} = 15.2$ kcal/mol. Using Equation 2, we predicted the in vivo actual activation ratio $AR_{actual}$ for this riboswitch at a wide range of $\Delta G_{ligand}$ (from pico-molar to milli-molar ligand binding affinities) and maximum ligand concentrations, and by assuming $V_{mRNA}/V_{water} = 10000$, $V_{ligand}/V_{water} = 20$, and total mRNA concentration = 0.16 $\mu$M (equivalent of 50 mRNA molecules inside the cell). We found that riboswitch activation was extremely high when using a high-affinity aptamer and when the ligand is plentiful ($AR_{actual} > 1000$-fold). When picomolar ligand levels must be sensed, a “perfect” riboswitch will activate expression by 20-fold using a high-
affinity aptamer, and 3-fold using a very low-affinity aptamer. These calculations illustrate the outer bounds of riboswitch sensing regardless of the aptamer sequence or structure (Supplementary Figure 19B).

10- Examples of Data Analysis (DNT, Dopamine, and Thyroxine Riboswitches)

Riboswitch-specific activation ratios were calculated identically for all riboswitches. Here, we provide specific example calculations for the DNT, dopamine, and thyroxine riboswitches to illustrate the procedure.

DNT riboswitches and no-aptamer controls were characterized in bacterial assays where the media was supplemented with 1% (v/v) DMSO and either 0 or 1 mM DNT. The addition of DNT non-specifically increased reporter expression. The no-aptamer control's mRFP1 fluorescence was 1765 au without DNT and 8510 au with 1 mM DNT, yielding an activation ratio of 4.82-fold. Under the same buffer and ligand conditions, the DNT-3 riboswitch's mRFP1 fluorescence was 22 au without DNT and 1176 au with 1 mM DNT, yielding an activation ratio of 53.31. In our data analysis, we eliminate the effects of non-specific expression changes by dividing the ON/OFF activation ratio of the riboswitch by the ON/OFF activation ratio of the no-aptamer control. The riboswitch-specific activation ratio for DNT-3 is therefore 53.31 / 4.82 = 11.10.

The dopamine riboswitches and no-aptamer control were characterized in cell-free transcription-translation assays supplemented with 5 mM ascorbic acid. The no-aptamer control's luciferase expression level (luminescence) was 124680 au without dopamine and 76340 au with 1 mM dopamine, resulting in an activation ratio of 0.61. The Dopa-5 riboswitch's luciferase expression level was 10794 without dopamine and 12898 au with 1 mM dopamine, yielding an activation ratio of 1.19. The riboswitch-specific activation ratio for Dopa-5 was therefore 1.19 / 0.61 = 1.95.

The thyroxine riboswitches and no-aptamer control were characterized in cell-free transcription-translation assays supplemented with 1.5 mM NaOH and 100 mM NaCl. The no-aptamer control's luciferase expression level was 24659 au without thyroxine and 9104 au with 150 uM thyroxine, yielding a 0.37-fold activation ratio. The T4-2 thyroxine riboswitch's luciferase expression level was 1773 au without thyroxine and 1543 au with 150 uM thyroxine, yielding an activation ratio of 0.87. The riboswitch-specific activation ratio for T4-2 was therefore 0.87 / 0.37 = 2.35.
References

1. Lynch, S.a., et al., A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. Chemistry & biology, 2007. 14: p. 173-84.

2. Crooks, G.E., et al., WebLogo: A Sequence Logo Generator. 2004: p. 1188-1190.

3. Mishler, D.M. and J.P. Gallivan, A family of synthetic riboswitches adopts a kinetic trapping mechanism. Nucleic acids research, 2014. 42: p. 6753-61.

4. Salis, H.M., E.a. Mirsky, and C.a. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. Nature biotechnology, 2009. 27: p. 946-50.

5. Hao, Y., et al., Quantifying the sequence-function relation in gene silencing by bacterial small RNAs. Proceedings of the National Academy of Sciences of the United States of America, 2011. 108: p. 12473-8.

6. Farasat, I., et al., Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. Molecular Systems Biology, 2014. 10:731, p. 1-18.

7. Gruber, A.R., et al., The Vienna RNA web suite. Nucleic acids research, 2008. 36: p. W70-4.

8. Chen, H., et al., Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. Nucleic acids research, 1994. 22: p. 4953-7.

9. Esphah Borujeni, A., A.S. Channarasappa, and H.M. Salis, Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic acids research, 2014. 42: p. 2646-59.

10. Dirks, R.M., et al., Thermodynamic Analysis of Interacting Nucleic Acid Strands. SIAM Review, 2007. 49: p. 65-88.

11. Weinberg, Z., et al., Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome biology, 2010. 11: p. R31.

12. Serganov, A. and D.J. Patel, Amino acid recognition and gene regulation by riboswitches. Biochimica et biophysica acta, 2009. 1789: p. 592-611.

13. Corbino, K.a., et al., Evidence for a second class of S-adenosylmethionine riboswitches and other regulatory RNA motifs in alpha-proteobacteria. Genome biology, 2005. 6: p. R70.

14. Wang, J.X., et al., Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. Molecular cell, 2008. 29: p. 691-702.

15. Lee, E.R., et al., An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science (New York, N.Y.), 2010. 329: p. 845-8.

16. Cao, S. and S.-J. Chen, Predicting RNA pseudoknot folding thermodynamics. Nucleic acids research, 2006. 34: p. 2634-52.

17. Cao, S. and S.-J. Chen, Predicting structures and stabilities for H-type pseudoknots with interhelix loops. RNA (New York, N.Y.), 2009. 15: p. 696-706.

18. Sperschneider, J. and A. Datta, DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. Nucleic acids research, 2010. 38: p. e103.

19. Sperschneider, J., A. Datta, and M.J. Wise, Heuristic RNA pseudoknot prediction including intramolecular kissing hairpins. Rna-a Publication of the Rna Society, 2011. 17(1): p. 27-38.

20. Cao, S. and S.-J. Chen, Structure and stability of RNA/RNA kissing complex: with application to HIV dimerization initiation signal. RNA (New York, N.Y.), 2011. 17: p. 2130-43.

21. Kamihira, M. and M. Taniguchi, Sterilization of Microorganisms Supercritical Carbon Dioxide. 1987.

22. Lin-Chao, S., W.T. Chen, and T.T. Wong, High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. Molecular Microbiology, 1992. 6: p. 3385-93.

23. Lutz, R. and H. Bujard, Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic acids research, 1997. 25: p. 1203-10.

24. Li, G.-W., et al., Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell, 2014. 157: p. 624-35.

25. So, L.-H., et al., General properties of transcriptional time series in Escherichia coli. Nature genetics, 2011. 43: p. 554-60.

26. Desai, S.K. and J.P. Gallivan, Genetic screens and selections for small molecules based on a
synthetic riboswitch that activates protein translation. Journal of the American Chemical Society, 2004. **126**: p. 13247-54.