SUPPLEMENTAL TEXT

Summary of sequencing statistics

Over 1000 reads were observed at each concentration tested for 399 of the 522 single point mutants (76%). This is in line with the number of reads that were used in some previous studies that applied a similar sequencing-based approach to examine the activity of ribozyme variants (Kobori et al. 2015; Kobori and Yokobayashi 2016). However, accurate results can be obtained with fewer reads. Assuming a binomial distribution, 1000 reads results in a maximum standard deviation of ~1.6% (when the readthrough efficiency is 50%) for a given ligand concentration. Yet, decreasing this value to 100 reads only increases the theoretical maximum standard deviation to ~5%. Given that each data point is an independent trial and error approximately follows a normal distribution, this is generally sufficient to obtain an accurate curve.

All but one of the single point mutants had over 100 reads at each glycine concentration tested. Even so, all single point mutants fit well to the same modified binding equation (Equation 1, Materials and Methods). U164G was the only construct with fewer than 100 reads at any ligand concentration (6 out of 12; an average of 107.5 reads per glycine concentration). Yet, manual inspection shows that a reasonable curve was generated for this construct despite having the lowest degree of coverage of the point mutants (Supplemental Fig. S6).

In addition to the 522 single mutants, there are 135,459 potential double mutants in our library. As would be expected, these, on average, had substantially fewer reads than the single point mutants. Some of the double mutations were particularly rare due to a bias in mutation rates (Supplemental Table S2), with dual C-to-G/G-to-C mutations being the least common.

Yet, valuable information can be inferred even in cases where the mean number of reads is very low (see main text for an example). Out of the 135,459 potential double mutants in this library, 18,407 (14%) were observed to either have a mean number of reads per glycine concentration of greater than 20 and generated fits with error values for $K_{1/2}$ and amplitude that were smaller than the predicted values or had a mean number of reads per glycine concentration of greater than 50. This level of coverage is sufficient to draw preliminary conclusions about these mutants.

Comparison of data generated by SMARTT and the conventional gel-based assay

Three termination profiles generated by SMARTT were compared to the conventional gel-based assay to assess the agreement of the two approaches (see main text, Fig. 5). WT, C158U, and U88A were selected because they exhibited distinct minimal (Y_{min}), maximal (Y_{max}), and/or half-maximal termination ($K_{1/2}$) values. The fits obtained by SMARTT for WT and C158U exhibited similar, though slightly compressed amplitudes relative to the fits obtained with the gel-based assay. These constructs also displayed similar $K_{1/2}$ and Y_{min} values compared to the gel-based approach (Supplemental Table S3).

U88A did not modulate in response to glycine in the gel-based assay, but displayed a small yet nonzero amplitude by SMARTT (Amp^{SMARTT}: 7 ± 1 %). This mutation was not expected to show a glycine-dependent response in either assay since analogous mutations completely disrupt ligand binding in other constructs (Ruff et al. 2016; Ruff and Strobel 2014). We expect that the observed modulation is due to a low frequency of wild-type sequences that are misclassified as a result of sequencing errors or mutations that arise during the sample

preparation that follows transcription. Consistent with this hypothesis, the $K_{1/2}$ value observed for U88A ($K_{1/2}^{SMARTT}$ of 400 ± 100 μM) was within error of WT ($K_{1/2}^{SMARTT}$: 450 ± 30 μM). This result establishes the lower limit of detection for amplitude values obtained by SMARTT as being ~7%, and indicates that $K_{1/2}$ values obtained for variants with amplitudes near 7% may be influenced by contaminating WT sequences. Therefore, all $K_{1/2}$ values were excluded in our analyses for variants displaying an amplitude under the conservatively chosen threshold of 15%. Altogether, these results demonstrate that data generated by the two assays are highly comparable overall.

Determination of the length of the terminator hairpin

The exact length of the *C. tetani* type-1 singlet terminator hairpin is not immediately evident by manual inspection of its RNA sequence. Although base pairing is possible through position U165 (A131-U165), hairpin extension up to position C162 (G134-C162) or U163 (A133-U163) appears to induce termination most frequently. Several observations support this. In the WT construct, termination is most efficient following positions 169-171 (Supplemental Fig. S5A). This would place C162 or U163 as the last base pair to form in the terminator stem prior to termination. This is because termination usually occurs 7-8 nucleotides downstream from the base of this helix (Ray-Soni et al. 2016). Consistent with this, mutations at positions G134, C162, and U163 all caused significant increases in the observed readthrough efficiencies in the absence of glycine (Fig. 5A). Although mutations to A133 (A133-U163) have minimal effects on Y_{min}, this is likely because these mutations instead shift the site of termination upstream ~1 nt (Supplemental Fig. S5B). Mutations to U163 do not shift the site of

termination, but, as mentioned, do affect Y_{min}. Altogether, these observations indicate that the terminator typically extends either to C162 (G134-C162) or U163 (A133-U163) before inducing termination. For simplicity, the terminator hairpin was allowed to extend through U163 when performing free energy calculations of the expression platform.

SUPPLEMENTAL TABLES

TABLE S1. DNA Mutation frequencies within the SMARTT library.

Mutations per molecule	Frequency per ligand concentration[a]
0	26 ± 2%
1	35.3 ± 0.3%
2	23.2 ± 0.7%
3+	15.2 ± 0.8%

[a]Error intervals indicate the standard deviation across ligand concentrations.

TABLE S2. Nucleotide conversion frequencies.

Mutation Type	Normalized frequency[a]
A→G, T→C	20.0%
A→T, T→A	26.9%
A→C, T→G	6.8%
G→A, C→T	23.3%
G→T, C→A	17.5%
G→C, C→G	5.4%

[a]Expected frequency for a sequence with 50% GC content.

TABLE S3. Fit parameters generated with SMARTT and the gel-based assay.

	SMARTT[a]				Gel[a]			
Variant	**$K_{1/2}$ (μM)**	**Y_{min} (%)**	**Y_{max} (%)**	**Amp (%)**	**$K_{1/2}$ (μM)**	**Y_{min} (%)**	**Y_{max} (%)**	**Amp (%)**
WT	450 ± 30	7 ± 1	65 ± 1	58 ± 1	420 ± 20	2 ± 1	88 ± 1	86 ± 1
U88A	400 ± 100	8 ± 1	15 ± 1	7 ± 1	-	5 ± 1	5 ± 1	0 ± 1
C158U	270 ± 20	41 ± 1	79 ± 1	38 ± 1	300 ± 40	46 ± 1	96 ± 1	50 ± 2

[a]Error intervals indicate the standard deviation.

TABLE S4. DNA Template Sequences.

Sample	Sequence[a]
B. subtilis transcription termination template DNA (Gel)	TAATATAAGAAAAATCTGCCCAAATTGCTGTTGCCCAGCATATAGTGA TGATGGTAGGATATGAGTATGTATTTGATGTAAGATATTGCTATAGTA TGTCCATAACAGCATGAAAATATGAGCGAATGACAGCAAGGGGAGA GACCTGACCGAAAACCTCGGGATACAGGCGCCGAAGGAGCAAACT GCGGAGTGAATCTCTCAGGCAAAAGAACTCTTGCTCGACGCAACTC TGGAGAGTGTTTGTGCGGATGCGCAAACCACCAAAGGGGACGTCT TTGCGTATGCAAAGTAAACTTTCAGGTGCCAGGACAGAGAACCTTC ATTTTACATGAGGTGTTTCTCTGTCCTTTTTTGTATGTTTTTTAGCTGC GCCGTTGATAAAGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAG CTTGGCGTAATCATGGTCATAGCTGTTTCC
C. tetani transcription termination template DNA (Gel)	TAATATAAGAAAAATCTGCCCAAATTGCTGTTGCCCAGCATATAGTGA TGATGGTAGGATATGAGTATGTATTTGATGTAAGATATTGCTATAGTA TGTCCATAACAGCATGAAAATATGATCTAGTTGAAGAGTATAAGAGA GATCCTATTTTAAAGGACGCCGAAGGGACAATCTATGTTTATCCCAA TAAAACATAGAGAAATTCTCAGGCAAAAGAATTATACTTTGATAGAC TCTGGAAAGTAAACAGAGAGAGAGCGAACGTGGGGTTTGTTCTCTC TTTATTTTTTTAACAGAGAGGACAAACCTTGGGGTTTGTTCTCTTTTTA TTGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCC
D. hafniense transcription termination template DNA (Gel)	TAATATAAGAAAAATCTGCCCAAATTGCTGTTGCCCAGCATATAGTGA TGATGGTAGGATATGAGTATGTATTTGATGTAAGATATTGCTATAGTA TGTCCATAACAGCATGAAAATATGACTGGATGAGGTTTTCAGGAGAA CAGGGTAAGCTAACCATGATGAACTGAAAACGGACAGAACTCTGG AGAGTTCCGCAAGGACGCCGAAGGGGCAAGACAGCAAAGCTGTTC AATCTCTCAGGCAAAAGGACAGAGCGCATAGTTTGTGAACGATGCT AAGGCAGGCGATAACTGCCAATAAGCGGGTTCGCAAATGCTCGTGTA CCGGAGTCAAACTTAAGATTTGGCTTTTTTTTATTGGTGAAAAGAAAG GAAAAAAACAAAAAAGAAAAGGGGTAAGGCATGCAAGCTTGGCGTAA TCATGGTCATAGCTGTTTCC
C. tetani transcription termination template DNA (Sequencing)	TAATATAAGAAAAATCTGCCCAAATTGCTGTTGCCCAGCATATAGTGA TGATGGTAGGATATGAGTATGTATTTGATGTAAGATATTGCTATAGTA TGTCCATAACAGCATGAAAATATGATCTAGTTGAAGAGTATAAGAGA GATCCTATTTTAAAGGACGCCGAAGGGACAATCTATGTTTATCCCAA TAAAACATAGAGAAATTCTCAGGCAAAAGAATTATACTTTGATAGAC TCTGGAAAGTAAACAGAGAGAGAGCGAACGTGGGGTTTGTTCTCTC TTTATTTTTTTAACAGAGAGGACAAACCTTGGGGTT

[a] Promoter is shown in orange, the predicted transcription start site is colored red, the aptamer region is shown in blue, and the predicted U-tract is colored pink.

TABLE S5. Oligos used for SMARTT assays.

DNA Oligo	Pertinent Step	Nextera XT i7 Index Name	Sequence (5' to 3')[a]
3' End DNA Adapter	Adapter Ligation	-	/5rApp/NNNNNCTGTAGGCACCATCAAT /3ddC/
RT Primer	Reverse Transcription	-	**GATTGATGGTGCCTACAG**
Fwd Primer 1	Illumina Adapter PCR – Step 1	-	CTTTCCCTACACGACGCTCTTCCGATC T**GTCCATAACAGCATGAAAATATG**
Rev Primer 1	Illumina Adapter PCR – Step 1	-	GTGACTGGAGTTCAGACGTGTGCTCT TCCGATCT**GATTGATGGTGCCTACAG**
Fwd Primer 2	Illumina Adapter PCR – Step 2	-	AATGATACGGCGACCACCGAGATCTA CACT**CTTTCCCTACACGACGCTC**
Rev Primer 2A (Index 1)	Illumina Adapter PCR – Step 2	N701	CAAGCAGAAGACGGCATACGAGATTC GCCTTA**GTGACTGGAGTTCAGACGTG**
Rev Primer 2B (Index 2)	Illumina Adapter PCR – Step 2	N702	CAAGCAGAAGACGGCATACGAGATCT AGTACG**GTGACTGGAGTTCAGACGTG**
Rev Primer 2C (Index 3)	Illumina Adapter PCR – Step 2	N703	CAAGCAGAAGACGGCATACGAGATTT CTGCCT**GTGACTGGAGTTCAGACGTG**
Rev Primer 2D (Index 4)	Illumina Adapter PCR – Step 2	N704	CAAGCAGAAGACGGCATACGAGATGC TCAGGA**GTGACTGGAGTTCAGACGTG**
Rev Primer 2E (Index 5)	Illumina Adapter PCR – Step 2	N705	CAAGCAGAAGACGGCATACGAGATAG GAGTCC**GTGACTGGAGTTCAGACGTG**
Rev Primer 2F (Index 6)	Illumina Adapter PCR – Step 2	N706	CAAGCAGAAGACGGCATACGAGATCA TGCCTA**GTGACTGGAGTTCAGACGTG**
Rev Primer 2G (Index 7)	Illumina Adapter PCR – Step 2	N707	CAAGCAGAAGACGGCATACGAGATGT AGAGAG**GTGACTGGAGTTCAGACGTG**
Rev Primer 2H (Index 8)	Illumina Adapter PCR – Step 2	N710	CAAGCAGAAGACGGCATACGAGATCA GCCTCG**GTGACTGGAGTTCAGACGTG**
Rev Primer 2I (Index 9)	Illumina Adapter PCR – Step 2	N711	CAAGCAGAAGACGGCATACGAGATTG CCTCTT**GTGACTGGAGTTCAGACGTG**
Rev Primer 2J (Index 10)	Illumina Adapter PCR – Step 2	N712	CAAGCAGAAGACGGCATACGAGATTC CTCTAC**GTGACTGGAGTTCAGACGTG**
Rev Primer 2K (Index 11)	Illumina Adapter PCR – Step 2	N714	CAAGCAGAAGACGGCATACGAGATTC ATGAGC**GTGACTGGAGTTCAGACGTG**
Rev Primer 2L (Index 12)	Illumina Adapter PCR – Step 2	N715	CAAGCAGAAGACGGCATACGAGATCC TGAGAT**GTGACTGGAGTTCAGACGTG**
Rev Primer 2M (Index 13)	Illumina Adapter PCR – Step 2	N718	CAAGCAGAAGACGGCATACGAGATGT AGCTCC**GTGACTGGAGTTCAGACGTG**

[a] Annealing region is in bold. Indexes are underlined.

SUPPLEMENTAL FIGURES

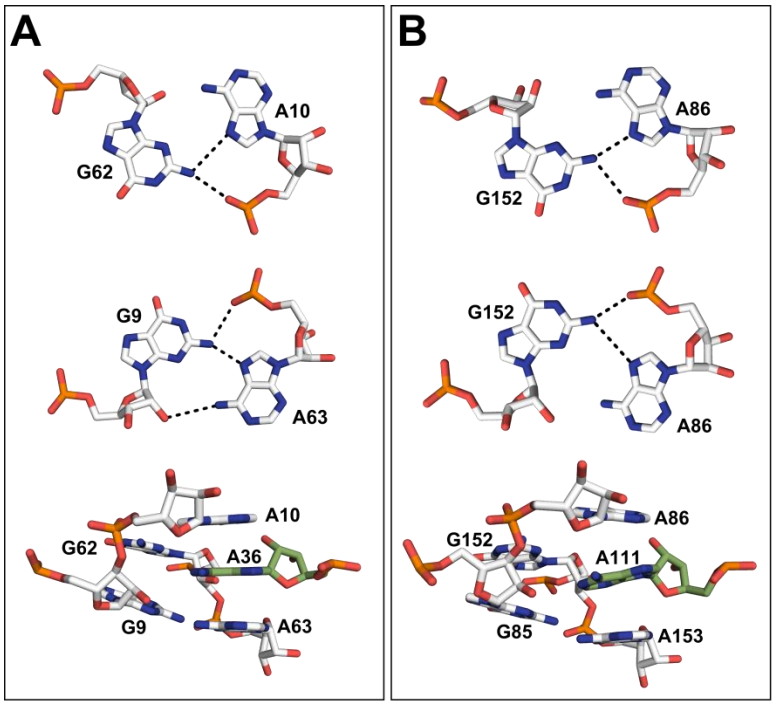

FIGURE S1. Interactions in the purine-rich region of the tandem glycine riboswitch. Noncanonical pairing interactions found at the top of the P1 helix of aptamer 1 (A) and aptamer 2 (B) in the tandem glycine riboswitch crystal structure (PDB ID: 3P49). The intercalating adenine from the P3 helix of each aptamer is shown in green. Numbering is based on the labeling scheme used by Butler et al. (2011).

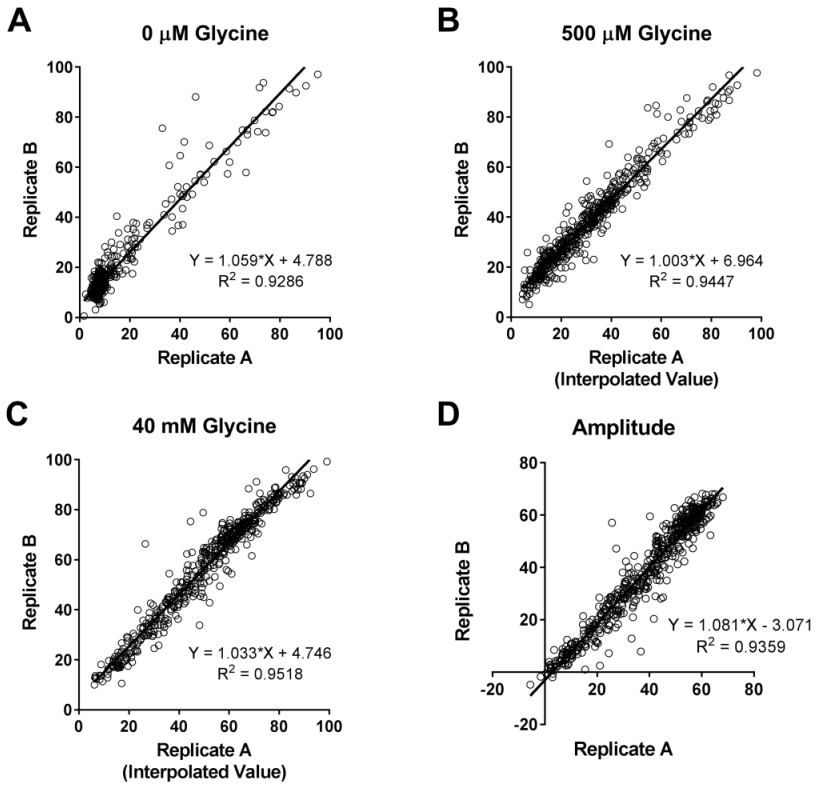

FIGURE S2. Comparison of replicate samples. Prior to performing the full glycine titration with SMARTT (Replicate A), a preliminary evaluation of the assay sampling three glycine concentrations (0 μM, 500 μM, and 40 mM) was performed (Replicate B). Scatter plots are shown comparing the observed readthrough efficiencies for WT and the 522 single point mutants at these concentrations (A-C) as well as their observed amplitudes (D). A direct comparison between the readthrough efficiencies obtained for the two replicates performed at 0 μM glycine is shown in (A). Scatter plots (B) and (C) depict comparisons between the readthrough efficiency observed in Replicate B at 500 μM glycine (B) and 40 mM glycine (C) with the interpolated values obtained in Replicate A at these concentrations (based on the fit parameters generated) as direct comparisons could not be made. The amplitude values used for Replicate B in (D) were estimated by taking the difference between the readthrough efficiency values at 40 mM and 0 μM glycine.

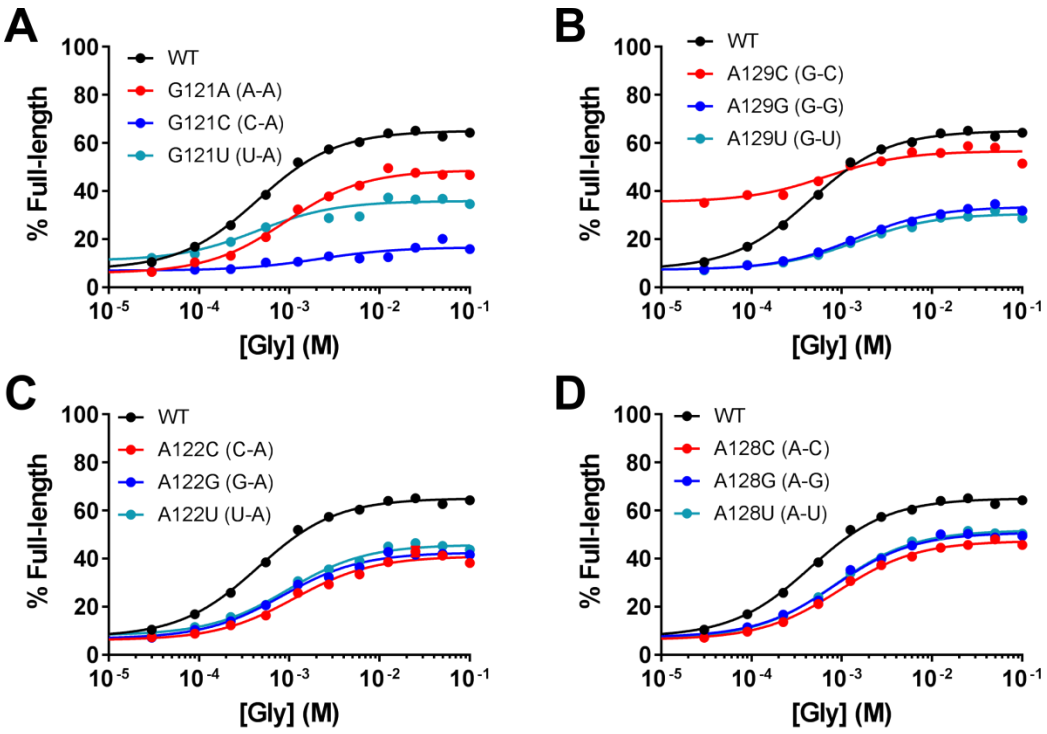

FIGURE S3. Effects of mutating the purine-rich region of the type-1 singlet ghost aptamer. Response profiles are shown for mutants of the putative GA (A&B) and AA (C&D) noncanonical base pairs found in the ghost aptamer of the *C. tetani* type-1 singlet. All mutations, including those that resulted in typical AU, GC, or GU base pairs resulted in defects. Error bars representing the standard deviation do not appear as they would be smaller than the size of each point.

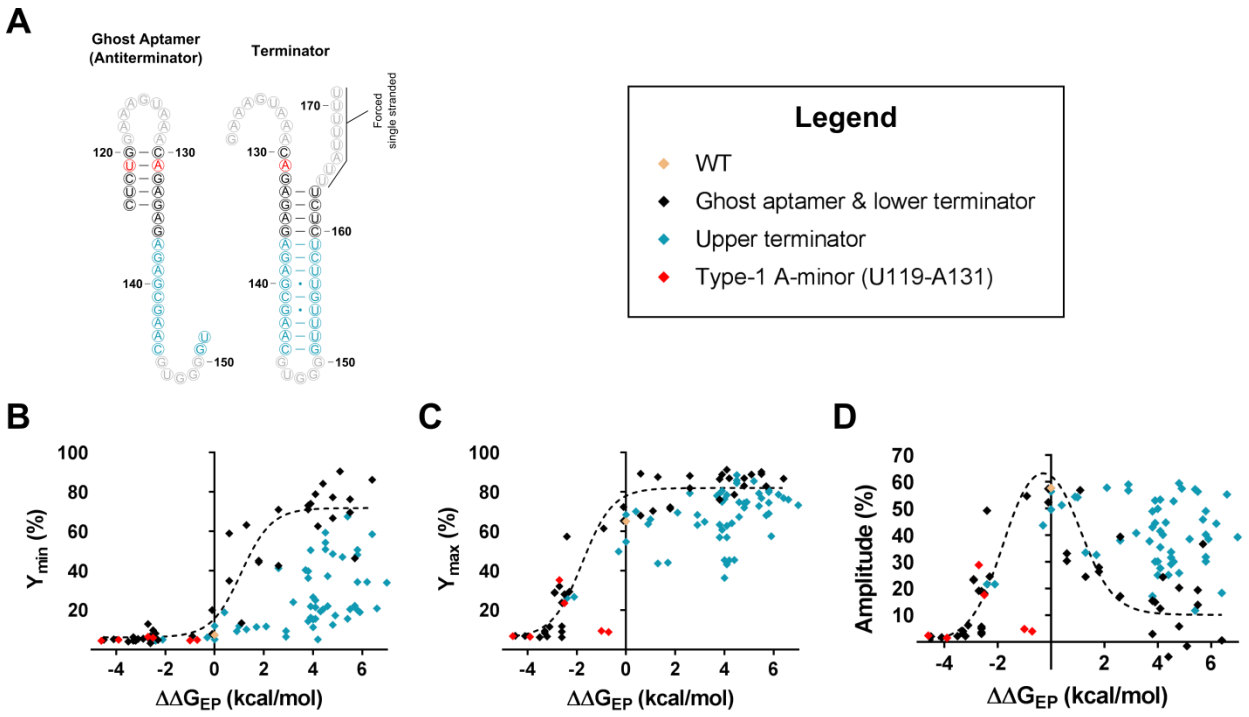

FIGURE S4. Dependence of readthrough efficiencies on $\Delta\Delta G_{EP}$ for the full *C. tetani* expression platform. (A) Predicted secondary structures of the antiterminator and terminator sequences used in the free energy calculations of $\Delta\Delta G_{EP}$. Nucleotides are colored according to plots (B-D). Correlations are shown between $\Delta\Delta G_{EP}$ and Y_{min} (B), Y_{max} (C), and amplitude (D). The dotted lines represent the fits depicted in Fig.7B-D. As described in the text, positions in the upper terminator systematically deviate from these fits.

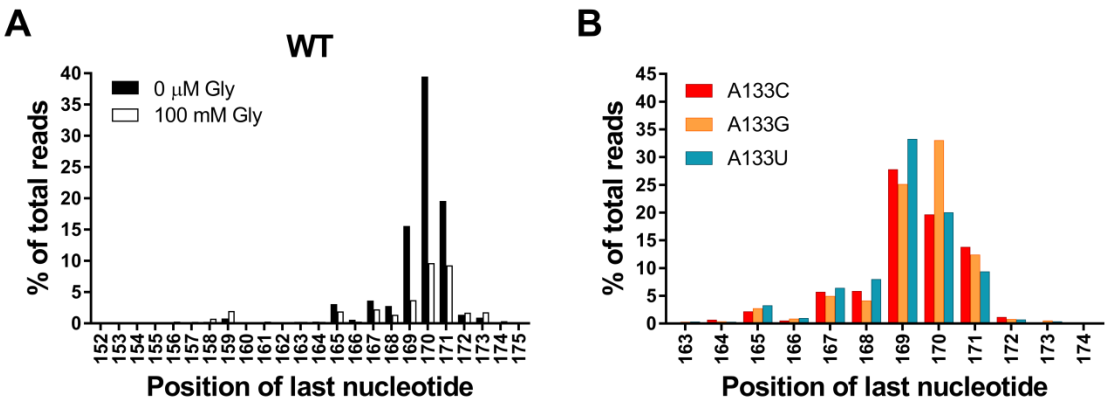

FIGURE S5. Termination efficiencies by position. (A) Termination efficiencies following each nucleotide position between positions 152-175 of the WT construct obtained in the absence of glycine and at 100 mM glycine. Termination occurs most readily following position 170 in the absence of glycine. (B) Termination efficiencies following each nucleotide between positions 163-174 for the three variants of A133 obtained in the absence of glycine. Mutations to A133 result in increased rates of termination following position 169.

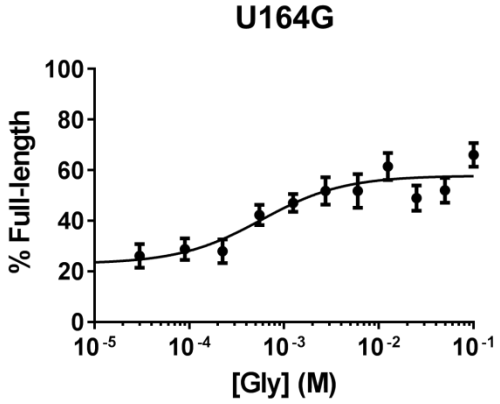

FIGURE S6. Response profile of the single point mutant with the lowest coverage. U164G had the lowest coverage of all single mutants and was the only point mutant for which less than 100 reads was observed at any concentration (6 out of 12; an average of 107.5 reads per glycine concentration). Manual inspection demonstrates that a reasonable fit is produced with this level of coverage. Error bars represent the standard deviation at each point.

SUPPLEMENTAL REFERENCES

Butler EB, Xiong Y, Wang J, Strobel SA. 2011. Structural Basis of Cooperative Ligand Binding by the Glycine Riboswitch. *Chem Biol* **18**: 293–298.

Kobori S, Nomura Y, Miu A, Yokobayashi Y. 2015. High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res* **43**: e85–e85.

Kobori S, Yokobayashi Y. 2016. High-Throughput Mutational Analysis of a Twister Ribozyme. *Angew Chem Int Ed* n/a-n/a.

Ray-Soni A, Bellecourt MJ, Landick R. 2016. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu Rev Biochem* **85**: 319–347.

Ruff KM, Muhammad A, McCown PJ, Breaker RR, Strobel SA. 2016. Singlet glycine riboswitches bind ligand as well as tandem riboswitches. *RNA* **22**: 1728–1738.

Ruff KM, Strobel SA. 2014. Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy. *RNA* **20**: 1775–1788.