JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Automated RNA Structure Prediction Uncovers a Kink-Turn Linker in Double Glycine Riboswitches

Wipapat Kladwang,[†] Fang-Chieh Chou,[‡] and Rhiju Das*[,†,§]

Departments of [†]Biochemistry, [‡]Chemistry, and [§]Physics, Stanford University, Stanford, California 94305, United States

Supporting Information

ABSTRACT: The tertiary structures of functional RNA molecules remain difficult to decipher. A new generation of automated RNA structure prediction methods may help address these challenges but have not yet been experimentally validated. Here we apply four prediction tools to a class of double glycine riboswitches that can bind two ligands cooperatively. A novel method (BPPalign), RMdetect, JAR3D, and Rosetta 3D modeling give consistent predictions for a new stem P0 and a kink-turn motif. These elements structure the linker between the RNAs' double aptamers. Chemical mapping on the *Fusobacterium nucleatum* riboswitch with *N*-methylisatoic anhydride, dimethyl sulfate and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate probing, mutate-and-map studies, and mutation/rescue experiments all provide strong evidence for the structured linker. Under solution conditions that permit rigorous thermodynamic analysis, disrupting this helix−junction−helix structure gives 120- and 6−30-fold poorer dissociation constants for the RNA's two glycine-binding transitions, corresponding to an overall energetic impact of 4.3 ± 0.5 kcal/mol. Prior biochemical and crystallography studies did not include this critical element due to over-truncation of the RNA. We speculate that several further undiscovered elements are likely to exist in the flanking regions of this and other functional RNAs, and automated prediction tools can play a useful role in their detection and dissection.

Non-coding RNA sequences play critical roles in cellular biochemistry and genetic regulation, and the number known is growing rapidly.[1,2] Many of their behaviors are intimately tied to their 3D conformations, but determining these structures has challenged both experimentalists and modelers, especially with regard to tertiary interactions mediated by non-Watson−Crick base pairs. Recent years have seen the development of automated algorithms for detecting and modeling tertiary structure in new RNAs, especially modular recurrent motifs.[3−6] However, the predictive power of these methods has yet to be demonstrated through rigorous experiments. Here we report the application and chemical validation of RNA structure prediction to discover an important tertiary element in a paradigmatic class of natural RNA riboswitches.

The cooperative binding of small molecules is a fundamental feature of functional biopolymers that has been studied in numerous model systems,[7,8] mostly protein-based until 2004, when Breaker et al. reported ligand-binding cooperativity in an RNA riboswitch that binds two glycine molecules.[9] This discovery has inspired many biophysical studies,[9−14] culminating in the recent publication of crystallographic models of double aptamers,[15,16] but a predictive model for ligand-binding cooperativity has remained elusive. Prior investigations have mostly focused on constructs pared down to minimal sequences required for glycine binding. We therefore sought to explore outer flanking sequences that have so far remained uncharacterized. We started by applying four recently developed RNA structure modeling tools and found remarkably consistent predictions for a new motif.

As input into modeling, we extracted aligned sequences of 360 double glycine riboswitches, collating single-aptamer entries in RFAM,[18] and extending these alignments into 5′ and 3′ flanking regions by 100 nucleotides (nts). Inspired by consensus approaches for protein fold recognition,[19] our new BPPalign tool (Supporting Information (SI)) searches for novel stems by averaging base-pair (bp) probability calculations for Watson−Crick secondary structure across homologues. In addition to new candidate elements in the 3′ region (Figure S1) containing potential "expression platforms",[9,20] we found interactions involving 9 nts preceding the conventional start of the riboswitch. For 160 sequences, including the widely studied riboswitches from *Vibrio cholerae* (*VC*)[9,10,12,13] and *Fusobacterium nucleatum* (*FN*),[11,21] a 3-bp interaction between the riboswitch inter-aptamer linker and nucleotides in the 5′ flanking sequence could occur (Figures S1 and 1). These potential Watson−Crick pairs would involve longer sequence separations than any other helical stem in the riboswitch. We call this putative stem P0. In addition, the nearby P1 stem of the first aptamer could potentially form an extension of three purine/purine pairs (129 sequences; Figure 1) or, in some cases, three Watson−Crick pairs (41 sequences). Finally, while P0 and this extended P1 are contiguous on the 3′ strand, there is an intervening three-residue bulge in the 5′ strand. We noted that the lengths and sequences of these features matched the published consensus for the kink-turn motif.[22] This motif, originally identified through high-resolution crystallography of the ribosome, has been annotated in numerous functional RNAs, including riboswitches for lysine and *S*-adenosylmethionine,[22−24] but never previously reported in glycine riboswitches.

Received: October 6, 2011
Published: December 19, 2011

ACS Publications © 2011 American Chemical Society

dx.doi.org/10.1021/ja2093508 | *J. Am. Chem.Soc.* 2012, 134, 1404−1407

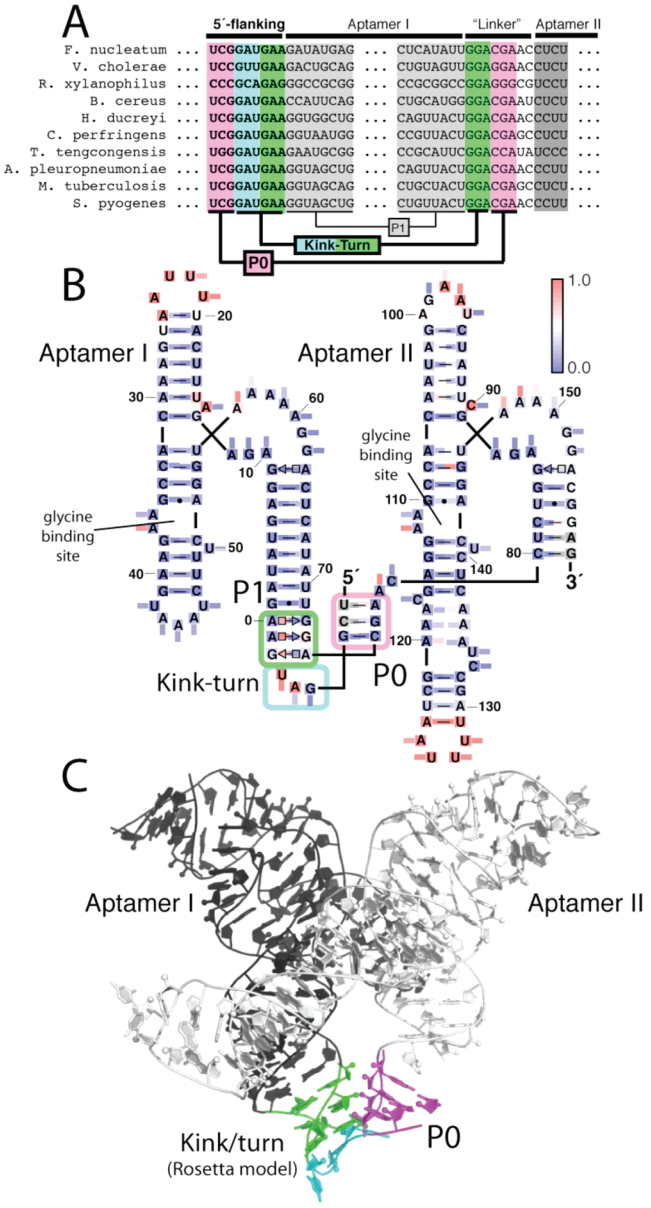


Figure 1. A new stem P0 and kink-turn tertiary element in double glycine riboswitches on a sequence alignment (A), on the secondary structure of the *FN* riboswitch (B), and built into this RNA's 3D crystallographic model (C). Coloring in (B) shows chemical mapping data: SHAPE (on letters), DMS (rectangles at A/C), and CMCT (rectangles at G/U). Rectangles point in to or out of helices for residues that are paired/unpaired in the glycine-bound structure.[16] Note numbering at previously defined start site[11] progresses from −1 to 0 to +1 (rather than from −1 to +1[17]).

Three independent lines of bioinformatic/computational evidence supported the presence of the kink-turn motif. (1) The automated RMdetect software[5] gave a strong-confidence detection of a kink-turn (101 sequences, Figure S2) when given our extended multiple sequence alignment. (2) The JAR3D server[6,25] returned a kink-turn motif when given the putative P0/P1 sequences (Figure S3). (3) We applied Rosetta/FARFAR 3D modeling[4] to replace the linker in the *FN* riboswitch crystallographic model[16] with a kink-turn motif, forming a continuous helical interface with P1. The resulting model (Figure 1C) demonstrated that a kink-turn can bridge the riboswitch aptamers with reasonable geometry and no clashes. We also carried out *de novo* modeling of the entire linker and the added 5′ strand; encouragingly, the lowest-energy models exhibited kink-turn conformations (Figure S4). An additional unexpected concordance was observed: the linker backbone within the modeled kink-turn approximately followed the linker path in the deposited crystallographic electron density and coordinates[16] (4.4 Å C4′ rmsd), despite the absence of pairing partners in the crystallized molecule (Figure S5).

These computational approaches gave consistent predictions but are largely untested, can give false positives (Figures S1 and S2), and do not provide information on the energetic significance of the putative element. Therefore, we carried out experiments to confirm and further characterize the kink-turn motif using mutate-and-map and mutate/rescue trials, with quantitative readouts from nucleotide-resolution chemical mapping. As a model RNA we chose the smallest known glycine riboswitch, the *FN* system, which has been studied by mutation, chemical mapping, and crystallography, albeit in truncated form.[11,14,16,21] Glycine binding by the *FN* riboswitch is less cooperative than the more widely characterized *VC* riboswitch,[9−14] suggesting that the effects of the putative kink/turn on cooperativity might be more sensitively deduced. We focused on an *FN* sequence with a 9-nt natural extension restored to the 5′ end compared to the prior construct (Figure 1B; SI). We called this sequence FN-KTtest. High-throughput measurements of dimethyl sulfate (DMS) and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfonate, (CMCT) modification reported on the chemical accessibilities of Watson−Crick nucleobase edges for A/C and G/U, respectively.[21] Nucleotides outside the linker region served as controls in this analysis; we confirmed that their reactivities in 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, and 10 mM glycine correlated with the burial of bases in the glycine-bound *FN* crystallographic model[15] (Figure 1B). Within the linker region and within the added natural 5′ flanking sequence, the DMS and CMCT reactivities were consistent with the 3D model of the kink-turn which protects the Watson−Crick edges of some bases (nts −4 to −6, and 72 to 77) while exposing others (nts −3 to 0; Figure 1B,C). Further analysis, based on 2′-OH acylation with *N*-methylisatoic anhydride (selective hydroxyl acylation analyzed by primer extension (SHAPE) chemistry[26]), gave protections of linker nts 72−79 (Figure 1B; see also Figure S6), as would be expected for a structured element; these nucleotides gave high SHAPE reactivities in previous studies without the 5′ flanking sequence.[11,21]

More stringent tests of the predicted kink-turn structure were achieved with mutate-and-map[14,27,28] experiments (Figure 2). We first disrupted the 5′ end of P0 ($^{5'\text{-UCG-3'}}_{3'\text{-AGC-5'}}$ to $^{5'\text{-AGC-3'}}_{3'\text{-AGC-5'}}$ at nts −8 to −6 and 75 to 77; called MutA), expecting the reactivity of the mutated segment and its base-pairing partner to increase. Indeed, the SHAPE reactivities in both P0 strands and putative P1 purine/purine pairs (nts −2 to 0 and 72−74) increased significantly, in some cases to the levels seen in unpaired loops (compare nts −6, −5, 72−74 with nts 20−25, 96−100) and in the prior construct without the 5′ flanking sequence.[21] SHAPE reactivity in other regions of the glycine riboswitch did not change significantly. We also mutated the 3′ end of P0 ($^{5'\text{-UCG-3'}}_{3'\text{-UCG-5'}}$, called MutB) and observed disruptions in the same regions. Finally, we sought to "rescue" these structural disruptions by implementing both sets of mutations ($^{5'\text{-AGC-3'}}_{3'\text{-UCG-5'}}$, called MutAB). The resulting variant indeed restored SHAPE reactivities to

dx.doi.org/10.1021/ja2093508 | *J. Am. Chem.Soc.* 2012, 134, 1404−1407

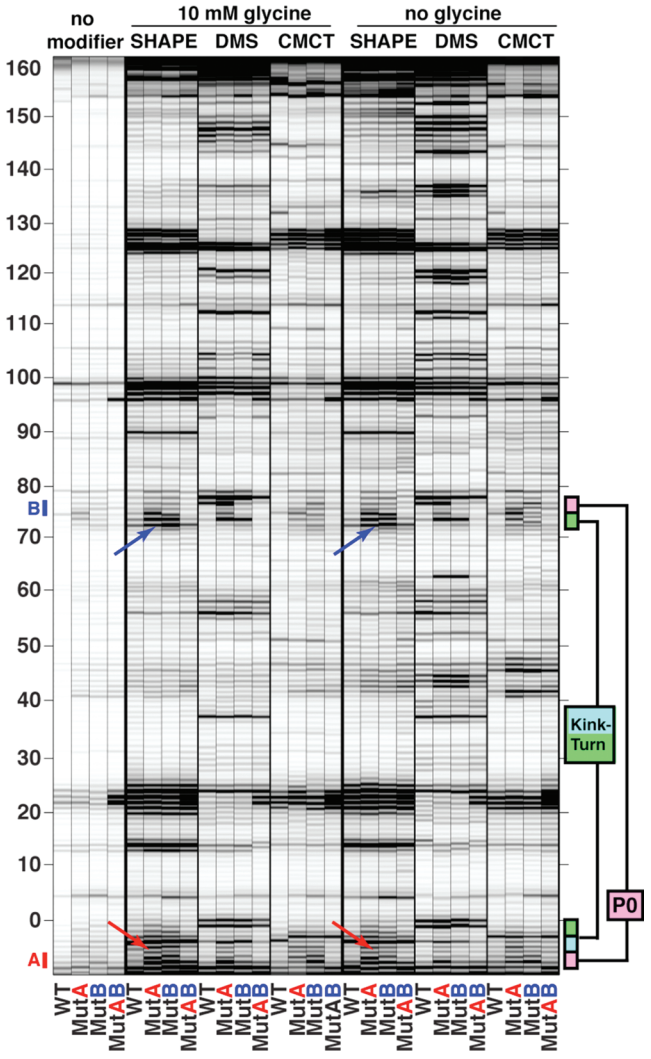


Figure 2. Evidence for the predicted kink-turn tertiary element and P0 stem from disruption of chemical reactivities by mutation (MutA and MutB) and restoration (MutAB) of P0 Watson–Crick pairs. Aligned electrophoretic traces are shown, from chemical mapping at 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 24 °C. Arrows mark SHAPE effects discussed in text.

levels observed for the wild-type FN-KTtest RNA. DMS and CMCT chemical mapping data gave analogous results (Figure 2). We conclude that there is strong biochemical evidence for the P0 stem and P1 purine pairs and, combined with the computational and chemical mapping data above, strong evidence for the kink-turn motif in the *FN* glycine riboswitch. Analogous measurements in the absence of glycine (Figure 2) showed that this motif is also formed in the glycine-free state of the RNA.

We then sought to determine the energetic importance of the kink-turn motif for glycine binding. We monitored glycine-induced conformational changes with DMS chemical mapping (which gave stronger reactivity changes than CMCT or SHAPE). Quantitative thermodynamic comparisons require that ligand-binding events monitored for different molecular variants correspond to transitions between analogous states. In the chemical mapping experiments above, wild-type FN-KTtest and rescued double-mutant MutAB appeared partially folded even in glycine-free conditions, giving DMS protections at nts 43–45, 63, 135–137 compared to single-strand mutants MutA and MutB (Figure S7). We therefore lowered the $MgCl_2$ concentration from 10 to 0.5 mM. Under these conditions, the chemical mapping data for MutA and MutB remained invariant, and the wild-type and MutAB RNAs also gave chemical reactivities identical to those of these mutants (outside the kink-turn linker) in their glycine-free states (Figure S7). For all constructs, two glycine-dependent transitions could be resolved under low $MgCl_2$ conditions (Figure 3), permitting the characterization of the kink-turn's

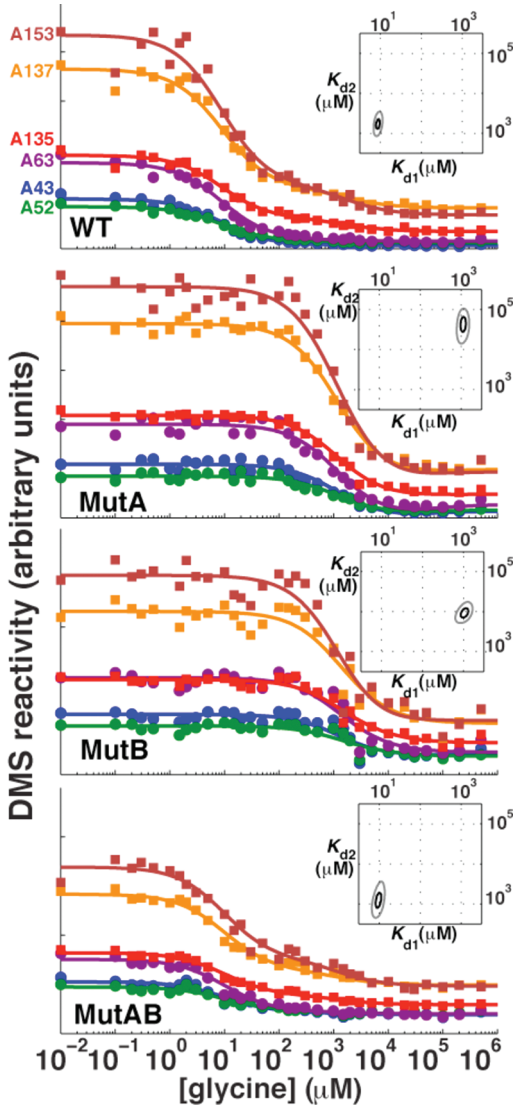


Figure 3. Thermodynamic analysis of glycine binding to riboswitches with (WT, MutAB) and without (MutA, MutB) the kink-turn. Symbols show DMS reactivities for six representative nucleotides (see top panel) at 0.5 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 24 °C; colored lines give calculated reactivities for a model with two glycine-binding events. Insets: Log-likelihood contours for K_{d1} and K_{d2} at 2 (black) and 10 (gray) units below maximum likelihood value.

energetic effect on individual glycine-binding events. The equilibrium is described in terms of glycine-free, single-glycine-bound, and two-glycine-bound states:

$$\text{RNA} + 2\ \text{glycine} \overset{K_{d1}}{\rightleftharpoons} \text{glycine·RNA} + \text{glycine} \overset{K_{d2}}{\rightleftharpoons} (\text{glycine})_2\text{·RNA}$$

We inferred equilibrium constants K_{d1} and K_{d2} and chemical reactivities for each state via a likelihood-based analysis. Further

dx.doi.org/10.1021/ja2093508 | *J. Am. Chem.Soc.* 2012, 134, 1404–1407

discussion of the fitted reactivities, interpretation of K_{d1} and K_{d2}, error estimation from replicates, and comparison to Hill-type fits are given in the SI.

As a baseline, wild-type FN-KTtest gave K_{d1} = 9 ± 0.5 μM and K_{d2} = 1.8 ± 0.4 mM (Figures 3 and S8). MutA and MutB then provided independent tests of the energetic impact of the kink-turn element. They required substantially higher glycine to undergo conformational change, with K_{d1} increased by 120-fold and K_{d2} increased by 6−30-fold (K_{d1} = 1.1 ± 0.2 mM and K_{d2} = 50 ± 20 mM for MutA; K_{d1} = 1.1 ± 0.2 mM and K_{d2} = 10 ± 3 mM for MutB). Disruption of the newly discovered linker thus results in free energy perturbations of 2.8 ± 0.1 kcal/mol and an additional 1.5 ± 0.5 kcal/mol for the two glycine-binding events, respectively. As expected, the double mutant MutAB restored the equilibrium constants to K_{d1} = 10 ± 1 μM and K_{d2} = 1.3 ± 0.5 mM, indistinguishable from the wild-type equilibria within experimental error.

These measurements confirmed unambiguously that this kink-turn motif has a large energetic impact on glycine riboswitch behavior. The behavior of this motif is analogous to the roles of kink-turns in stabilizing tertiary structure and ligand binding within other riboswitches,[24] but is unique here in bridging two homologous aptamers. The inclusion of this motif in future work may lead to more precise chemogenetic data,[11] more easily interpretable X-ray scattering profiles,[10,13] and better-diffracting crystals[15,16] with and without glycine. Interestingly, the kink-turn enhances the affinity but not the thermodynamic cooperativity of glycine binding. In particular, K_{d2}/K_{d1} increased from 10−50 to 130−200 upon kink-turn inclusion, rather than decreasing (strong cooperativity would correspond to $K_{d2}/K_{d1} < 4$; see SI). The glycine riboswitch of *F. nucleatum* may not have evolved to act cooperatively, or it may require different solution conditions, other molecular partners (e.g., kink-turn-binding proteins[29]), or a kinetic mechanism to exhibit ligand-binding cooperativity.[30,31] Given our results herein, we propose another hypothesis: there are secondary and tertiary motifs even farther out in the RNA's flanking regions that modulate the riboswitch's behavior. The interactions may even favor the glycine-free conformation of the RNA, analogous to the tense state in the Monod−Wyman−Changeux theory of allostery.[7,14] Some candidate interactions are listed in SI. There are certainly precedents for flanking sequences playing unexpected roles in functional RNAs (e.g., refs 32,33). We expect automated RNA structure prediction to be a useful tool for uncovering new pieces of the double glycine riboswitch and other functional RNAs that remain mysterious.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Data available at http://rmdb.stanford.edu as entries GLYCFN_KNK_0001, GLYCFN_KNK_0002. Methods, supporting text, and Figures S1−S9 as described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

rhiju@stanford.edu

■ ACKNOWLEDGMENTS

We thank P. Sripakdeevong, C. VanLang, and J. Sales-Lee for discussions; and A. Petrov, C. Zirbel, and N. Leontis for early access to the JAR3D. This work was supported by a Burroughs-Wellcome Career Award at the Scientific Interface (R.D.), Study Abroad Scholarship of Taiwan Government (F.C.), and NSF-CNS-0619516 (BioX^2 cluster).

■ REFERENCES

(1) Gesteland, R. F.; Cech, T. R.; Atkins, J. F. *The RNA world: the nature of modern RNA suggests a prebiotic RNA world*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006.
(2) Amaral, P. P.; Dinger, M. E.; Mercer, T. R.; Mattick, J. S. *Science* **2008**, *319*, 1787.
(3) Parisien, M.; Major, F. *Nature* **2008**, *452*, 51.
(4) Das, R.; Karanicolas, J.; Baker, D. *Nat. Methods* **2010**, *7*, 291.
(5) Cruz, J. A.; Westhof, E. *Nat. Methods* **2011**, *8*, 513.
(6) Sarver, M.; Zirbel, C. L.; Stombaugh, J.; Mokdad, A.; Leontis, N. B. *J. Math. Biol.* **2008**, *56*, 215.
(7) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.
(8) Perutz, M. F. *Q. Rev. Biophys.* **1989**, *22*, 139.
(9) Mandal, M.; Lee, M.; Barrick, J. E.; Weinberg, Z.; Emilsson, G. M.; Ruzzo, W. L.; Breaker, R. R. *Science* **2004**, *306*, 275.
(10) Lipfert, J.; Das, R.; Chu, V. B.; Kudaravalli, M.; Boyd, N.; Herschlag, D.; Doniach, S. *J. Mol. Biol.* **2007**, *365*, 1393.
(11) Kwon, M.; Strobel, S. A. *RNA* **2008**, *14*, 25.
(12) Erion, T. V.; Strobel, S. A. *RNA* **2011**, *17*, 74.
(13) Lipfert, J.; Sim, A. Y.; Herschlag, D.; Doniach, S. *RNA* **2010**, *16*, 708.
(14) Kladwang, W.; VanLang, C. C.; Cordero, P.; Das, R. *Nat. Chem.* **2011**, *3*, 954.
(15) Huang, L.; Serganov, A.; Patel, D. J. *Mol. Cell* **2010**, *40*, 774.
(16) Butler, E. B.; Xiong, Y.; Wang, J.; Strobel, S. A. *Chem. Biol.* **2011**, *18*, 293.
(17) Cech, T. R.; Damberger, S. H.; Gutell, R. R. *Nat. Struct. Biol.* **1994**, *1*, 273.
(18) Griffiths-Jones, S.; Moxon, S.; Marshall, M.; Khanna, A.; Eddy, S. R.; Bateman, A. *Nucleic Acids Res.* **2005**, *33*, D121.
(19) Bradley, P.; Malmstrom, L.; Qian, B.; Schonbrun, J.; Chivian, D.; Kim, D. E.; Meiler, J.; Misura, K. M.; Baker, D. *Proteins* **2005**, *61* (Suppl. 7), 128.
(20) Tucker, B. J.; Breaker, R. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 342.
(21) Kladwang, W.; Vanlang, C. C.; Cordero, P.; Das, R. *Biochemistry* **2011**, *50*, 8049.
(22) Goody, T. A.; Melcher, S. E.; Norman, D. G.; Lilley, D. M. *RNA* **2004**, *10*, 254.
(23) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 4214.
(24) Schroeder, K. T.; Daldrop, P.; Lilley, D. M. *Structure* **2011**, *19*, 1233.
(25) Rahrig, R. R.; Leontis, N. B.; Zirbel, C. L. *Bioinformatics* **2010**, *26*, 2689.
(26) Merino, E. J.; Wilkinson, K. A.; Coughlan, J. L.; Weeks, K. M. *J. Am. Chem. Soc.* **2005**, *127*, 4223.
(27) Kladwang, W.; Das, R. *Biochemistry* **2010**, *49*, 7414.
(28) Kladwang, W.; Cordero, P.; Das, R. *RNA* **2011**, *17*, 522.
(29) Turner, B.; Melcher, S. E.; Wilson, T. J.; Norman, D. G.; Lilley, D. M. *RNA* **2005**, *11*, 1192.
(30) Wickiser, J. K.; Cheah, M. T.; Breaker, R. R.; Crothers, D. M. *Biochemistry* **2005**, *44*, 13404.
(31) Lemay, J. F.; Penedo, J. C.; Tremblay, R.; Lilley, D. M.; Lafontaine, D. A. *Chem. Biol.* **2006**, *13*, 857.
(32) Martick, M.; Scott, W. G. *Cell* **2006**, *126*, 309.
(33) Chadalavada, D. M.; Gratton, E. A.; Bevilacqua, P. C. *Biochemistry* **2010**, *49*, 5321.

■ NOTE ADDED IN PROOF

J. Ye and colleagues (RNA, in press) recently report biochemical evidence for leader-linker interactions in *FN*, *VC*, and *Bacillus subtilis* glycine riboswitches.

Supporting Information for **"Automated RNA structure prediction uncovers a missing link in double glycine riboswitches"**

Wipapat Kladwang[1], Fang-Chieh Chou[2], Rhiju Das[1,3,*]

Departments of Biochemistry[1], Chemistry[2], and Physics[3], Stanford University, Stanford, CA 94305, US

This documents contains Supporting Methods, Supporting Text, Supporting References and Supporting Figures S1–S9.

Supporting Methods

Preparation of sequence alignments

The Rfam database (v10.0, January 2010) contained 3368 sequences in the glycine riboswitch entry. Double riboswitches were assembled by searching for sequences drawn from the same genomic source in which the end of one domain was within 9 nucleotides of the start of a second domain. 86% of the Rfam sequences passed this criterion, giving 1453 double riboswitches. We further filtered this alignment to remove marine shotgun metagenome sequences (which appeared highly redundant for this domain), leaving 360 sequences. Flanking sequences (100 nucleotides 5′ and 100 nucleotides 3′ of the Rfam seed boundaries) and any missing inter-domain linker sequences were filled in via *efetch* queries to the Entrez database of the form

```
http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nucleotide
```

All steps were carried out using automated python scripts, available with the *BPPalign* tool, described next

The BPPalign *method*

The *BPPalign* method was inspired by approaches in protein modeling that seek to improve the accuracy of structure predictions by evaluating consensus across several fold recognition algorithms[1] or different homologs[2]. The protocol involved two steps. First, all sequences in the alignment above were folded in the *partition* executable in *RNAstructure*[3] (version 5.3), with the base pairing probability *bpp* matrix saved to disk. Then, the *bpp* data were averaged across all homologs, based on alignment to one reference sequence (here, the *F. nucleatum* riboswitch sequence from NCBI accession number AE009951.2). These aligning and averaging steps were carried out in MATLAB (R2010b), using a script *make_bpp_plot.m*. All sequence handling scripts are available through the SimTK website under the project BPPalign (https://simtk.org/home/bpp_align).

RMdetect

RMdetect[4] (version 0.0.3) was applied to the same multiple sequence alignment as the BPPalign tool above. The command lines used were:

```
rmdetect.py  new_align_no_metagenome.stk > rmdetect_result.txt

rmcluster.py --min-occur=0.05 --min-score=9.0 --min-bpp=0.01
  --min-mi=0.001 --out-dir=./ --fig < rmdetect_result.txt >
  rmdetect_cluster.txt

rmout.py < cluster_004_KT_1.0.res  > cluster_004_KT_1.0_examples.txt
```

We searched for modules CL_1.0 (C-loop), GB_1.0 (G-bulge), KT_1.0 (kink-turn), TGA_1.0 (tandem G/A). In separate runs, we verified the recovery of the newer ARICH modules; these results are not shown since this module's definition was actually based on the glycine riboswitch Rfam alignment. We note that previous work applying RMdetect to the glycine riboswitch did not find the kink-turn reported herein due to the use of truncated single-domain alignments from Rfam.

JAR3D

JAR3D aligns given RNA sequences to hybrid stochastic-context-free-grammar/Markov-random-field models[5-7]. We used a beta version of the JAR3D web server. The server was given the sequence drawn from the *FN* riboswitch, GGAUGAAG*UGGAC, where '*' separates the strands. Nine motifs achieved over 90th percentile score for this sequence, and all were kink-turn motifs or close variants; the top hit (Group_311; percentile 100%) is shown in SI Fig. S3. Similar results were achieved when the entire alignment of sequence homologs in these segments were given to JAR3D.

Rosetta grafting

First, the sequence of the proposed kink-turn in the glycine riboswitch was compared with the ones of known kink turns[8]. We found that the *H. marismortui* 23S ribosomal RNA KT-46 motif has the most similar sequence to the proposed kink-turn sequence in the *FN* glycine riboswitch: they both start with two G-C base pairs, followed by a 3-nucleotide bulge, then 4 non-Watson-Crick (WC) base pairs of the same sequence at the end:

KT in FN riboswitch (3P49)

```
      GAU
5´ CG   GAAG    -7 - 1
3´ GC---AGGU    76 - 71
```

KT-46 from *H. marismortui* 23S rRNA (3CC2)

```
      AUG
5´ GG   GAAG    1311 - 1319
3´ CC---AGGU    1343 - 1338
```

Therefore, we took nucleotides 1310–1323/1334–1344, which included the KT-46 motif, from the PDB coordinates 3CC2.[9] The WC base-pairing region (nts 1320–1323/1334–1337) was aligned to the corresponding region (nts 2–5/67–70) in the glycine riboswitch model 3P49[10]; and we replaced the proposed kink-turn region (nts 0–2/70–77) in the glycine riboswitch with the KT-46 model (nts 1310–1320/1337–1344), with the nucleobase mutated in Rosetta[11,12]. The junction nucleotides (nts 2 and 70) were minimized with variable bond length and bond angle, as in ref[11]. The 2-nucleotide loop connecting the kink turn and the second aptamer (nts. 78–79) was modeled using FARFAR loop modeling[12], generating 100 models each with 5000 fragment insertions. The lowest energy model is shown in Fig. 1C.

Rosetta de novo *structure prediction.*

For fully *de novo* modeling, we removed nucleotides 0, 1 and 71 in the P1 stem, and 72–79 in the linker from the crystallographic model 3P49[10]. We then rebuilt strands –7 to 1 and 71–79 with Rosetta Fragment Assembly of RNA with Full Atom Refinement (FARFAR[11]). 50,000 models were generated with 10,000 fragment insertions each, and the lowest energy 1000 models were clustered with RMSD threshold 2.0 Å. Four of five lowest energy cluster centers recovered the P0 stem, at least 2 of 3 sheared purine pairs, and a sharp bend at the three-nucleotide bulge (SI Fig. S4). One of the cluster centers recovered all three purine pairs and a complete kink-turn conformation; this model gave

an all-atom RMSD of 3.4 Å over the rebuilt nucleotides from the 'graft' model above (SI Fig. S4).

RNA preparation and high-throughput chemical mapping

The full FN-KT RNA sequence included 5′ and 3′ flanking sequences to permit interpretation of chemical mapping data all the way to the beginning and end of the riboswitch domains and to permit facile primer binding; the sequence was:

ggaaauaa**UCG**GAUGAAGAUAUGAGGAGAGAUUUCAUUUUAAUGAAACACCGAAGAAGUAAAUCUUUCAGG
UAAAAAGGACUCAUAUUGGA**CGA**ACCUCUGGAGAGCUUAUCUAAGAGAUAACACCGAAGGAGCAAAGCUAA
UUUUAGCCUAAACUCUCAGGUAAAAGGACGGAGaaaacaaaacaaagaaacaacaacaacaac

The primer binding site is underlined. Sites that were mutated to their complements in MutA, MutB, and MutAB are given in boldface. Flanking sequences are in lowercase. The added flanking sequences were checked in *RNAstructure*[13] and *Viennafold*[14] to give no predicted base pairing with the glycine-binding domain; chemical mapping data for parts of the flanking sequences were observable and gave high reactivity, as expected. Further, agreement of chemical reactivity in the glycine-binding domain (outside the predicted kink-turn) to prior measurements with different flanking sequences confirmed the lack of interaction of flanking sequences with the glycine-binding domain.

Preparation and purification of RNAs, chemical mapping with DMS, CMCT, and NMIA (SHAPE) modification, and the readout of modifications by reverse transcription and capillary electrophoresis were carried out as previously described[11,15]. Briefly, DNA templates for the RNAs were prepared by PCR assembly, with a 20-bp promoter TTCTAATACGACTCACTATA for T7 RNA polymerase, and purified with AMPure magnetic beads (Agencourt, Beckman Coulter). To remove a mispriming product, these DNA samples were further purified by agarose gel electrophoresis, band excision, and purification in Qiaquick microcentrifugation columns following kit instructions. Then, 40 μL *in vitro* RNA transcriptions were purified with MagMax magnetic beads (Ambion). RNAs (at 60 nM concentration) were incubated in the desired solution conditions (e.g., 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, and 10 mM glycine) for 20 minutes before addition of chemical modification reagent (final concentrations: 0.125% DMS, 2.6 mg/mL CMCT, or 3.0 mg/mL N-methyl isatoic acid) and incubation times of 15 minutes (DMS, CMCT) or 30 minutes (NMIA) at 24 °C. We were able to achieve final glycine concentrations of 500 mM before reaching solubility limits; we note here glycine concentrations higher than 10 mM appear to quench SHAPE measurements (NMIA), but DMS and CMCT reactions were robust up to 500 mM glycine (C. VanLang, RD, unpublished data).

Chemical modification reactions were quenched with β-mercaptoethanol (DMS) or lowered pH (Na-MES, pH 6.0); purified by binding to a fluorescent (5′-rhodamine-green-labeled) DNA primer pre-bound to poly-dT beads; reverse transcribed with SuperScript III enzyme; and purified and desalted before co-loading on ABI 3100 or 3730 capillary sequencers with Texas-red-labeled reference ladders. All data were analyzed in the HiTRACE software[16]. Data are deposited in the RNA Mapping Database (http://rmdb.stanford.edu) as entries GLYCFN_KNK_0001 and GLYCFN_KNK_0002. For Fig. 1B (estimates of intrinsic chemical reactivities), data were background-

subtracted, corrected for attenuation of reverse transcription products, and normalized as in ref.[17]. For glycine titration analyses, these extra processing steps were not used in order to avoid introducing additional noise.

Analysis of titrations: thermodynamic formalism and likelihood-based fitting

Our system involves the binding of at most two ligands and can be formally described by the chemical equilibrium

$$\text{RNA+2 glycine} \overset{K_{d1}}{\rightleftharpoons} \text{glycine•RNA+glycine} \overset{K_{d2}}{\rightleftharpoons} (\text{glycine})_2\text{•RNA}$$

or, equivalently, the partition function:

$$Z = 1 + \frac{x}{K_{d1}} + \left(\frac{x}{K_{d1}}\right)\left(\frac{x}{K_{d2}}\right), \qquad (1)$$

where x is the ligand concentration[18]. The terms partition all RNA conformations without ligand bound into one state **0**, all conformations with one ligand bound into state **1**, and all double-ligand-bound conformations into the state **2**. Systems with high binding cooperativity correspond to cases with low detectable population of state **1** at all x, as happens with $K_{d1} > K_{d2}$. A system with no cooperativity (e.g., independent binding to two sites) will give $K_{d2} / K_{d1} = 4$. It is important to note that state **1** may be an admixture of multiple structures (e.g., with different sites bound to the single ligand) or a single structure; a glycine titration on a single RNA cannot distinguish between the two, but comparison of RNA mutants can discriminate models (see caption of SI Fig. S5).

The fraction of conformations within the three states are then:

$$f_{\mathbf{0}} = \frac{1}{Z}$$

$$f_{\mathbf{1}} = \frac{\frac{x}{K_{d1}}}{Z} \qquad (2)$$

$$f_{\mathbf{2}} = \frac{\left(\frac{x}{K_{d1}}\right)\left(\frac{x}{K_{d2}}\right)}{Z}$$

The nucleotide-by-nucleotide chemical mapping data for each RNA variant can be summarized in data matrix D_{ij} where i indexes across the glycine concentrations and j indexes across nucleotides. The predictions for these data, D_{ij}^{pred} are given by the summing the chemical reactivity for each state C_j^a, weighted by the population fractions in each state:

$$D_{ij}^{pred} = \sum_{a=0,1,2} f_a^i C_j^a \quad , \qquad (3)$$

where f_a^i are the population fractions of state *a* at glycine concentration *i.* We carried out fits of D_{ij} to $D_{ij}^{pred}(K_{d1}, K_{d2})$ by grid search over K_{d1} and K_{d2} (each from 10^1 to 10^5 μM, in $10^{0.05}$ increments), obtaining analytical solutions for C_j^a maximum likelihood values. The likelihood function[19] was given by:

$$L \propto \prod_{i,j} \frac{1}{\sigma_j} \exp\left[-\frac{\left(D_{ij} - D_{ij}^{pred}\right)^2}{2\sigma_j^2} \right]$$

Here we have allowed each nucleotide *j* to have a different error. Given K_{d1} and K_{d2} (and thus f_a^i), this likelihood function is optimized by setting its derivatives with respect to each parameter C_j^a to zero, resulting in the following linear equation at each *j*:

$$\sum_{i,a'} f_a^i f_{a'}^i C_j^{a'} = \sum_i f_a^i D_{ij}$$

The errors are then solved as:

$$\sigma_j^2 = \frac{1}{N_i} \sum_i \left(D_{ij} - D_{ij}^{pred}\right)^2$$

The log-likelihood function then simply evaluates to $\log L = -N_i \sum \log \sigma_j + \text{constant}$.

The procedure above gave well-defined solutions for K_{d1} and K_{d2} for the wild type and MutAB glycine titration data, where two distinct transitions could be resolved visually (Fig. 3 and SI Fig. 4). For MutA and MutB data, the transitions were closer together, and the analysis above gave several distinct local likelihood optima, some involving anomalous transitions with K_{d1} or K_{d2} above the largest measured concentrations (0.5 M). To regularize the fits, we assumed that the reactivities for the states **0**, **1**, and **2** were similar to $C_j^a[\text{FIT}]$ measured for the wild type and MutAB constructs (indeed this assumption is required in order to make any thermodynamic comparison). The log-likelihood function was supplemented with a term:

$$L_{regularized} = L \exp\left[-\beta \sum_{a,j} \frac{1}{2\sigma_j^2} \left(C_j^a - C_j^a[\text{FIT}]\right)^2 \right]$$

with β = 0.1; variation of this parameter by 2-fold did not change the results.
To determine final fitted values for K_{d1} and K_{d2} for each construct, $\log L(K_{d1}, K_{d2})$ was summed across replicate measurements taken on different days (two for wild type and MutA; three for MutB and MutAB; each set involved two independent preparations of RNA). See insets to Fig. 3. Values reported in the main text are maximum likelihood values for K_{d1} and K_{d2}. Errors give the bounds of the log-likelihood contour corresponding to a decrement of 2 from the likelihood value; these bounds correspond closely to twice the standard error.[19]

Free energy perturbations from mutations were derived from the Boltzmann relation and the assumed partitition function. If the wild type construct gives best-fit equilibrium constants K_{d1} and K_{d2} and a mutant gives K^*_{d1} and K^*_{d2}, the free energy perturbations for binding the first glycine is:

$$\Delta\Delta G_{\mathbf{0\rightarrow 1}} = k_B T \log(K^*_{d1} / K_{d1})$$

And for binding the next glycine is:

$$\Delta\Delta G_{\mathbf{1\rightarrow 2}} = k_B T \log(K^*_{d2} / K_{d2})$$

The free energy perturbation for binding two glycines to the ligand-free state is the sum of the above two expressions: $\Delta\Delta G_{\mathbf{0\rightarrow 2}} = \Delta\Delta G_{\mathbf{0\rightarrow 1}} + \Delta\Delta G_{\mathbf{1\rightarrow 2}}$.

Supporting Text

Comments on the Hill equation

In some prior work[20-22], thermodynamic measurements were fit to a Hill-like equation:

$$f_{\text{unbound}} = 1/[1+(x/K_d)^n];\ f_{\text{bound}} = (x/K_d)^n/[1+(x/K_d)^n]$$

This fit is useful as a phenomenological characterization in cases of unknown binding stoichiometry. However, the Hill equation does not correspond to a realistic model except in special cases where *n* is an integer; those cases correspond to models which exhibit exactly two states, with 0 or *n* ligands bound, and do not apply here[23]. This issue is well-known in studies of multiple ligand binding by proteins (see, e.g., ref[24] for review of the thermodynamic formalism for hemoglobin oxygen-binding). For the present problem, using the Hill-like equation at different nucleotides or with different probes (e.g., different chemical modifications; gel mobility; x-ray scattering) gives different numerical values for K_d and *n*. In contrast, the thermodynamic model defined above in eqs. (1)–(3) should give the same values (within experimental error) for K_{d1} and K_{d2} with different measurements. We have verified this by, e.g., repeating our analyses for DMS reactivity at just nucleotides 135–150; we arrive at indistinguishable results.

Candidates for further glycine riboswitch interactions

From the BPPalign output as well as manual inspection of the double glycine riboswitch alignment, we found candidates for additional elements that may be important for riboswitch function. These are under experimental investigation; we list them here to encourage their validation or falsification by other groups.

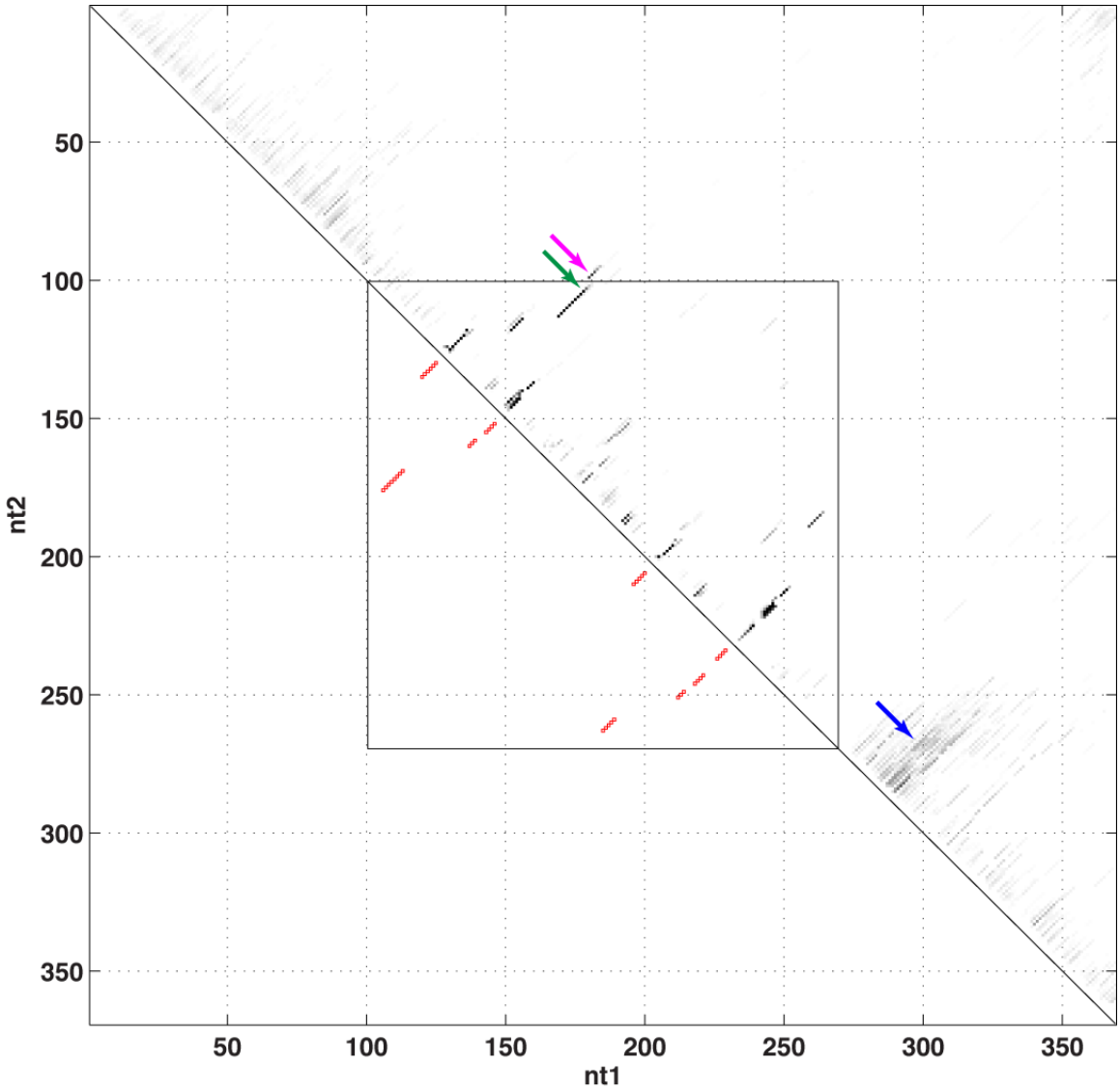
1. Expression platforms. Transcription terminator and translational suppressor stems 3′ of the ligand binding domains are readily identifiable (SI Fig. S1) and in several cases are mutually exclusive with P1 of the second aptamer. In the *FN* riboswich, three potential terminator hairpins with multiple subsequent U-nucleotides can be identified.
2. CUCUC/GAGAG The sequence CUCUC (or similar) appears at the glycine binding site in nearly all riboswitches in both aptamers I and II (e.g., nts 49–53 and nts 138–142 in the *FN* sequence). It is also noteworthy that the P1 stem of aptamer II typically contains sequence CUCU. The sequence GAGAG (or similar) appears in nearly all riboswitches (e.g, nts 9–13 and nts 86–90) in the core 3-way junction; BPPalign (SI. Fig. S1) often pairs this to the CUCU sequences. These interactions are not consistent with the glycine-bound crystallographic structure and therefore may be important in the glycine-free state. However we have not been able to detect covariation across these potentially interacting sequences; they may be conserved for other reasons including stability of the glycine-bound state.
3. Upstream modulators. The BPPalign analysis shows several potential elements 5′ of the domain that may disrupt P1 or P0 through alternative base pairing. It is not clear if these signals are significant, but such interaction may be important in restoring cooperative behavior to the *FN* riboswitch. It is unclear whether these upstream sequence elements are actually present in the riboswitch; there is so far little work in mapping the transcription start sites for natural riboswitches expressed in their native cellular environments.
4. Aptamer 1–3′ interaction (CCCC/GGGG). Nearly all glycine riboswitches have a hairpin extension on the P3 stem of aptamer 1, although this is missing in the most widely studied riboswitches from *F. nucleatum* and *V. cholerae*. Further, nearly all of these P3b stems display a CCCC sequence in their apical loops (with occasional replacement of a C to U). Interestingly, these riboswitches also exhibit the sequence GGGG at their 3′ ends (immediately after P1 for aptamer II). Visual inspection of the available crystallographic models for the *FN* riboswitch and *VCII* dimer suggest that such sequences should be able to interact in the glycine-bound state of these riboswitches. Further evidence for this interaction comes from the observation that sequences without the 3′ GGGG sequence do not have the C-rich P3b loop in aptamer I. This interaction has presumably not been noticed before due its absence in the small glycine riboswitches chosen for study to date.

Supporting References

1. Bradley, P., Malmstrom, L., Qian, B., Schonbrun, J., Chivian, D., Kim, D. E., Meiler, J., Misura, K. M., and Baker, D. (2005) Free modeling with Rosetta in CASP6, *Proteins 61 Suppl 7*, 128-134.

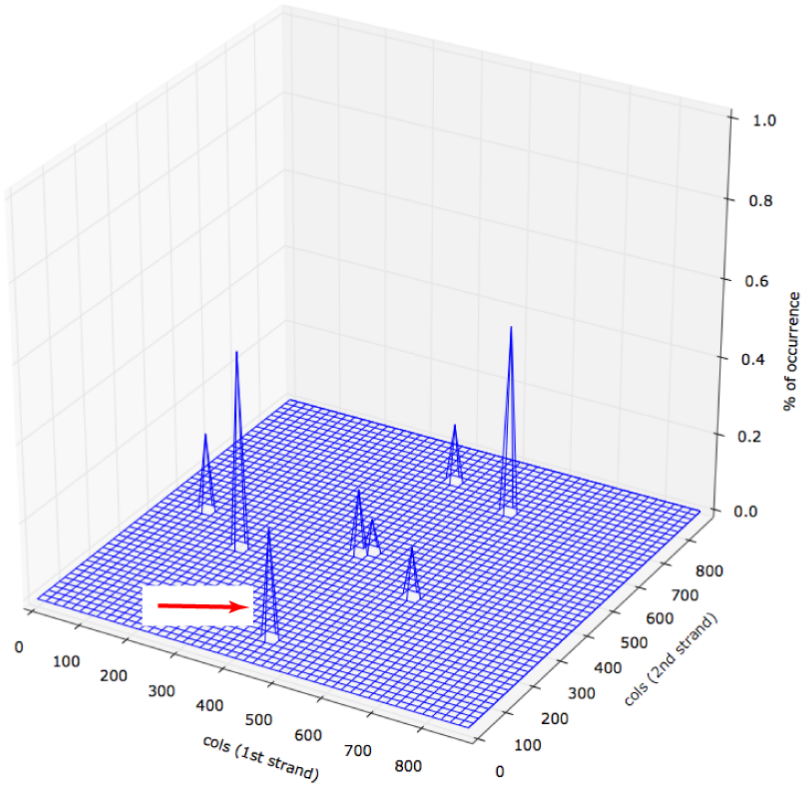
2. Bonneau, R., Strauss, C. E., and Baker, D. (2001) Improving the performance of Rosetta using multiple sequence alignment information and global measures of hydrophobic core formation, *Proteins 43*, 1-11.
3. Reuter, J. S., and Mathews, D. H. (2010) RNAstructure: software for RNA secondary structure prediction and analysis, *BMC Bioinformatics 11*, 129.
4. Cruz, J. A., and Westhof, E. (2011) Sequence-based identification of 3D structural modules in RNA with RMDetect, *Nature Methods 8*, 513-521.
5. Sarver, M., Zirbel, C. L., Stombaugh, J., Mokdad, A., and Leontis, N. B. (2008) FR3D: finding local and composite recurrent structural motifs in RNA 3D structures, *J Math Biol 56*, 215-252.
6. Rahrig, R. R., Leontis, N. B., and Zirbel, C. L. (2010) R3D Align: global pairwise alignment of RNA 3D structures using local superpositions, *Bioinformatics 26*, 2689-2697.
7. Sarver, M. (2006) Structure-Based Multiple RNA Sequence Alignment and Finding RNA Motifs, In *Mathematics/Probability and Statistics*, Bowling Green State University, Bowling Green.
8. Schroeder, K. T., McPhee, S. A., Ouellet, J., and Lilley, D. M. (2010) A structural database for k-turn motifs in RNA, *RNA 16*, 1463-1468.
9. Blaha, G., Gurel, G., Schroeder, S. J., Moore, P. B., and Steitz, T. A. (2008) Mutations outside the anisomycin-binding site can make ribosomes drug-resistant, *Journal of Molecular Biology 379*, 505-519.
10. Butler, E. B., Xiong, Y., Wang, J., and Strobel, S. A. (2011) Structural basis of cooperative ligand binding by the glycine riboswitch, *Chemistry & biology 18*, 293-298.
11. Das, R., Karanicolas, J., and Baker, D. (2010) Atomic accuracy in predicting and designing noncanonical RNA structure, *Nat Methods 7*, 291-294.
12. Sripakdeevong, P., Kladwang, W., and Das, R. (2011) An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
13. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure, *Proc Natl Acad Sci U S A 101*, 7287-7292.
14. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neubock, R., and Hofacker, I. L. (2008) The Vienna RNA websuite, *Nucleic Acids Res 36*, W70-74.
15. Kladwang, W., Cordero, P., and Das, R. (2011) A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA, *RNA 17*, 522-534.
16. Yoon, S., Kim, J., Hum, J., Kim, H., Park, S., Kladwang, W., and Das, R. (2011) HiTRACE: high-throughput robust analysis for capillary electrophoresis, *Bioinformatics 27*, 1798-1805.
17. Kladwang, W., Vanlang, C. C., Cordero, P., and Das, R. (2011) Understanding the Errors of SHAPE-Directed RNA Structure Modeling, *Biochemistry 50*, 8049-8056.
18. Garcia, H. G., Kondev, J., Orme, N., Theriot, J. A., and Phillips, R. (2011) Thermodynamics of biological processes, *Methods in enzymology 492*, 27-59.
19. Edwards, A. W. F. (1972) *Likelihood*, Cambridge University Press, Cambridge.
20. Mandal, M., Lee, M., Barrick, J. E., Weinberg, Z., Emilsson, G. M., Ruzzo, W. L., and Breaker, R. R. (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression, *Science 306*, 275-279.
21. Kwon, M., and Strobel, S. A. (2008) Chemical basis of glycine riboswitch cooperativity, *RNA 14*, 25-34.
22. Erion, T. V., and Strobel, S. A. (2011) Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch, *RNA 17*, 74-84.
23. Weiss, J. N. (1997) The Hill equation revisited: uses and misuses, *Faseb J 11*, 835-841.

24. Eaton, W. A., Henry, E. R., Hofrichter, J., and Mozzarelli, A. (1999) Is cooperative oxygen binding by hemoglobin really understood?, *Nature structural biology 6*, 351-358.
25. Kladwang, W., VanLang, C. C., Cordero, P., and Das, R. (2011) A two-dimensional mutate-and-map strategy for non-coding RNA structure, *Nat Chem 3*, 954-962.

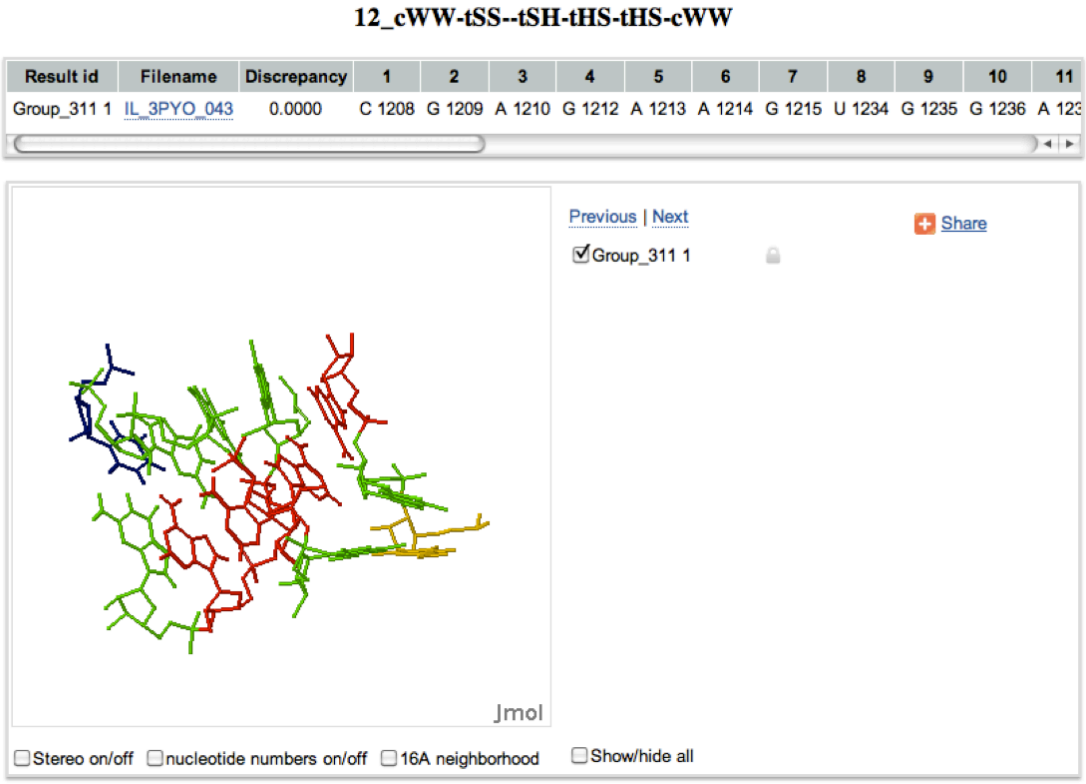


Supporting Figure S1. Results of the BPPalign method on the glycine riboswitch sequences. Grayscale gives average predicted Watson-Crick base pairing probability *bpp* across 360 glycine riboswitches, at positions that were aligned to the *FN* riboswitch sequence in the Rfam alignment. White represents pairs with average *bpp* <0.005; black represents pairs with average *bpp* ≥ 0.1. The large box gives the conventional bounds of the glycine-binding domains, e.g., as defined in Rfam or in previous studies[21,25]. Red squares give base pairs used to seed the Rfam alignment. Magenta arrow points to new predicted stem ("P0"); green arrow points to extension of P1. The 3-nt vertical spacing between these features represents a bulge of conserved length. Other features in the plot correspond to the assumed seed stems (c.f. red squares below diagonal and dark stripes across diagonal); 3′ stems involved in 'expression platforms' (translation repressor and transcription terminator stems; blue arrow); and likely false positives involving sequences conserved for other reasons (e.g., glycine binding) that happen to be complementary. The latter features are under further investigation (see Supporting Text, "Candidates for further glycine riboswitch interactions").

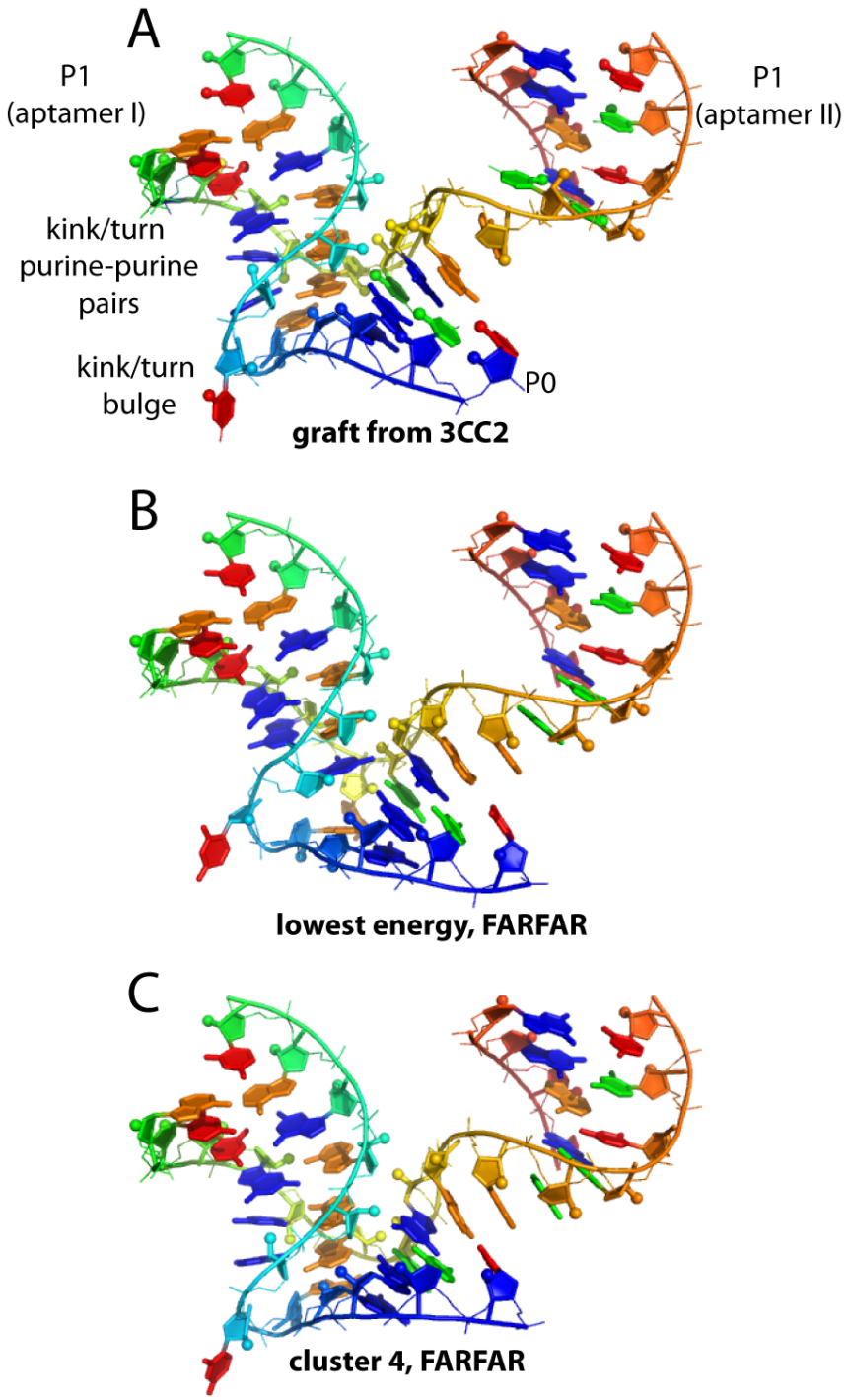
```
Cluster:     1 -> model: CL_1.0     count:    183 occur_(%):  50.70 score:  10.142 bpp: 0.018 MI:   0.207 H:   0.208 cols:  227  357
Cluster:     2 -> model: CL_1.0     count:     52 occur_(%):  12.81 score:   9.232 bpp: 0.014 MI:   0.178 H:   0.207 cols:  578  356
Cluster:     3 -> model: CL_1.0     count:    192 occur_(%):  48.47 score:  10.062 bpp: 0.151 MI:   0.227 H:   0.241 cols:  583  703
Cluster:     4 -> model: TGA_1.0     count:     75 occur_(%):  20.06 score:  10.550 bpp: 0.294 MI:   0.279 H:   0.305 cols:   98  442
Cluster:     5 -> model: KT_1.0     count:    101 occur_(%):  28.13 score:  16.033 bpp: 0.183 MI:   0.201 H:   0.185 cols:  438   96
Cluster:     6 -> model: TGA_1.0     count:     62 occur_(%):  16.71 score:  11.553 bpp: 0.093 MI:   0.103 H:   0.086 cols:  422  443
Cluster:     7 -> model: TGA_1.0     count:     54 occur_(%):  15.04 score:  11.767 bpp: 0.086 MI:   0.033 H:   0.051 cols:  436  773
Cluster:     8 -> model: TGA_1.0     count:     32 occur_(%):   8.36 score:  11.028 bpp: 0.052 MI:   0.271 H:   0.464 cols:  442  466
```



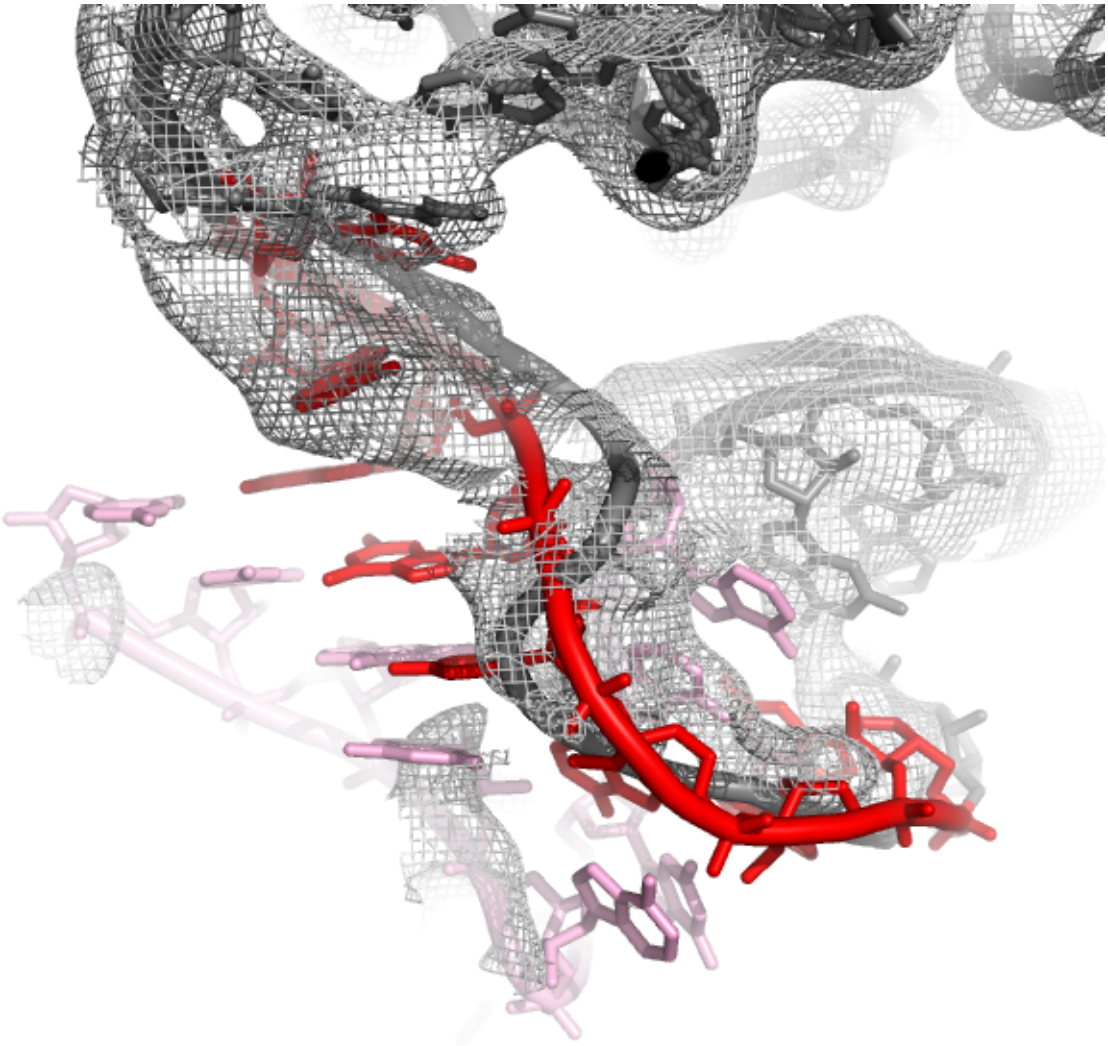
Supporting Figure S2. Results of the *RMdetect* algorithm on sequence alignment of 360 glycine riboswitches. Arrows point to kink-turn prediction. Additional tandem G/A ('TGA') predictions are partially correct or part of kink-turn purine-pairs motif. The C-loop motif ('CL') predictions appear to be false positives.



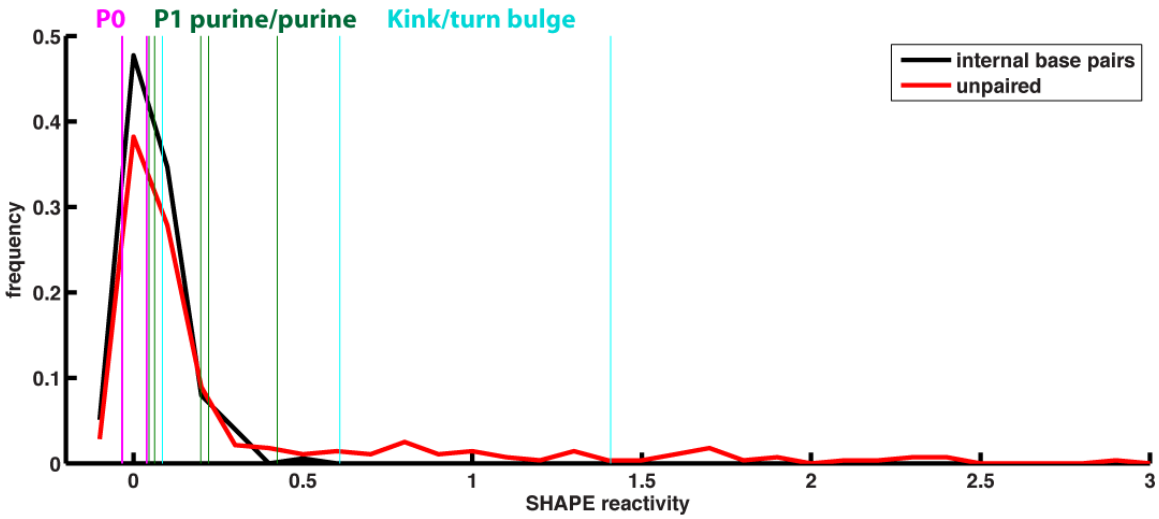
Supporting Figure S3. Results of the *JAR3D* algorithm on putative kink-turn element from the *FN* double glycine riboswitch. A screen-shot of the top hit is shown.



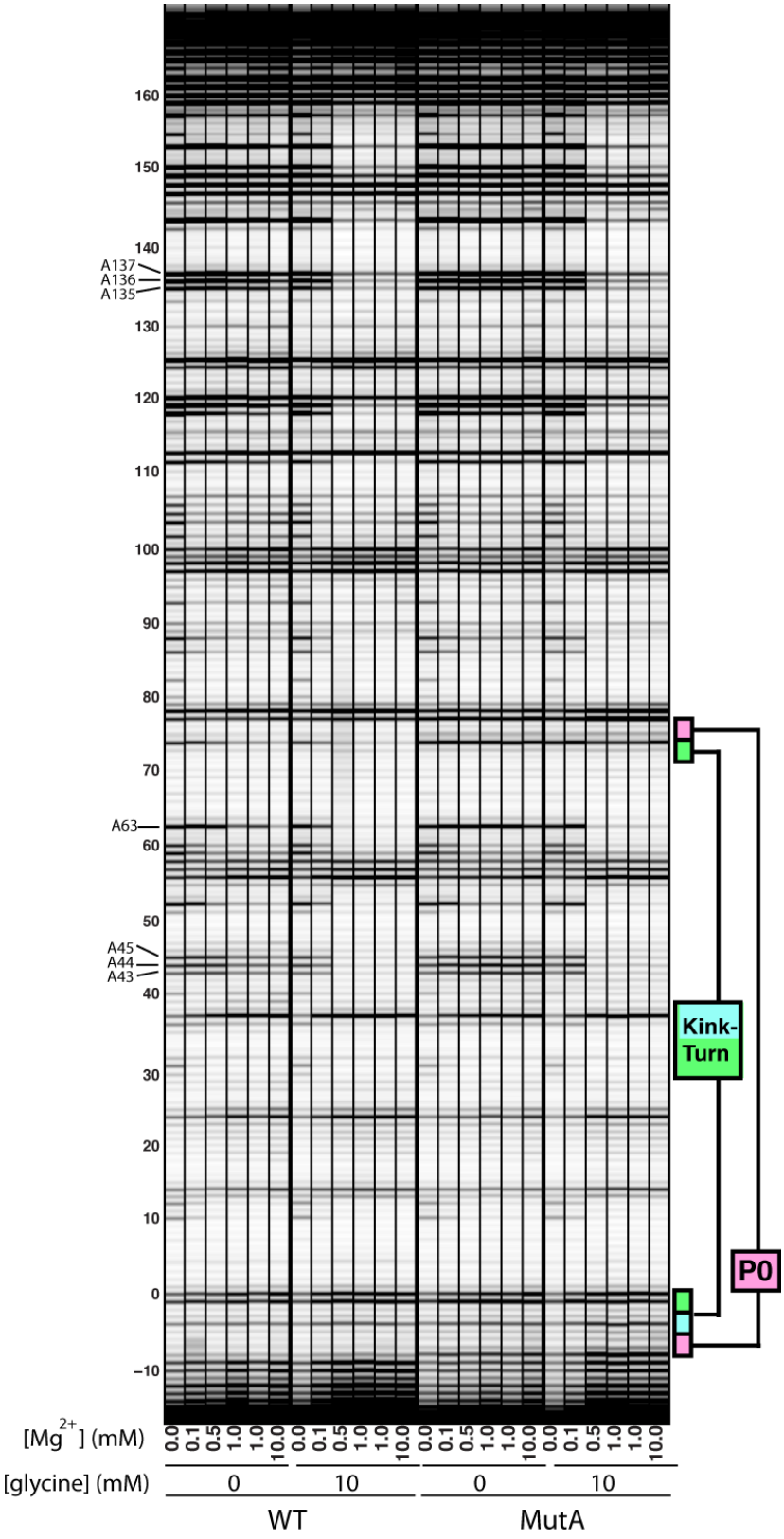
Supporting Figure S4. Rosetta models for the new elements of the glycine riboswitch. (A) Graft of the kink-turn from *H. marismortui* large ribosomal subunit crystallographic model (KT-46) into *FN* riboswitch model (PDB ID: 3P49), optimized with Rosetta. (B) Lowest energy model from *de novo* FARFAR modeling for the entire 8-nt linker strand and 9 added nucleotides; model recovers two purine-purine pairs in P1 and the novel three-Watson-Crick-base-pair stem P0. (C) Fourth lowest energy cluster from *de novo* FARFAR modeling fully recovers the complete kink-turn motif and P0. Bases are colored according to identity (A, orange; C, green; G, blue; U, red), and backbones are colored according to position in the motif. Riboswitch coordinates beyond the P1 stems of the double aptamers are not shown.



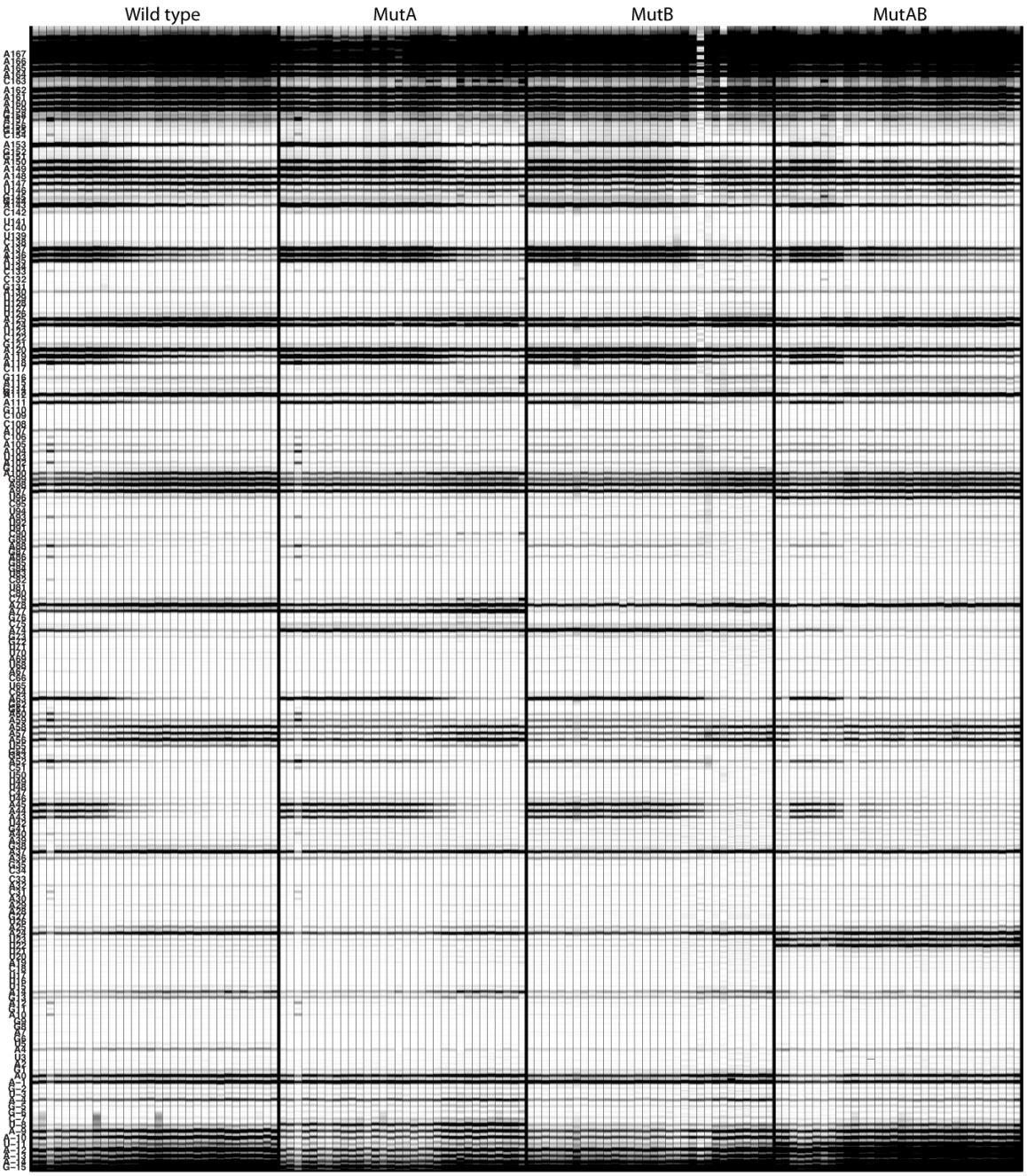
Supporting Figure S5. Unexpected concordance of remodeled linker structured by the kink-turn motif and crystallographic coordinates of unstructured linker. The Rosetta 3D model of the newly predicted inter-aptamer kink-turn (red and pink) reorganizes the inter-aptamer linker (red) through pairing with nine nucleotides (pink) 5′ of the conventional (minimal) glycine riboswitch boundary. The rebuilt linker path still approximately traces the path of the linker in the original glycine riboswitch coordinates (black) as well as the observed electron density (gray mesh).



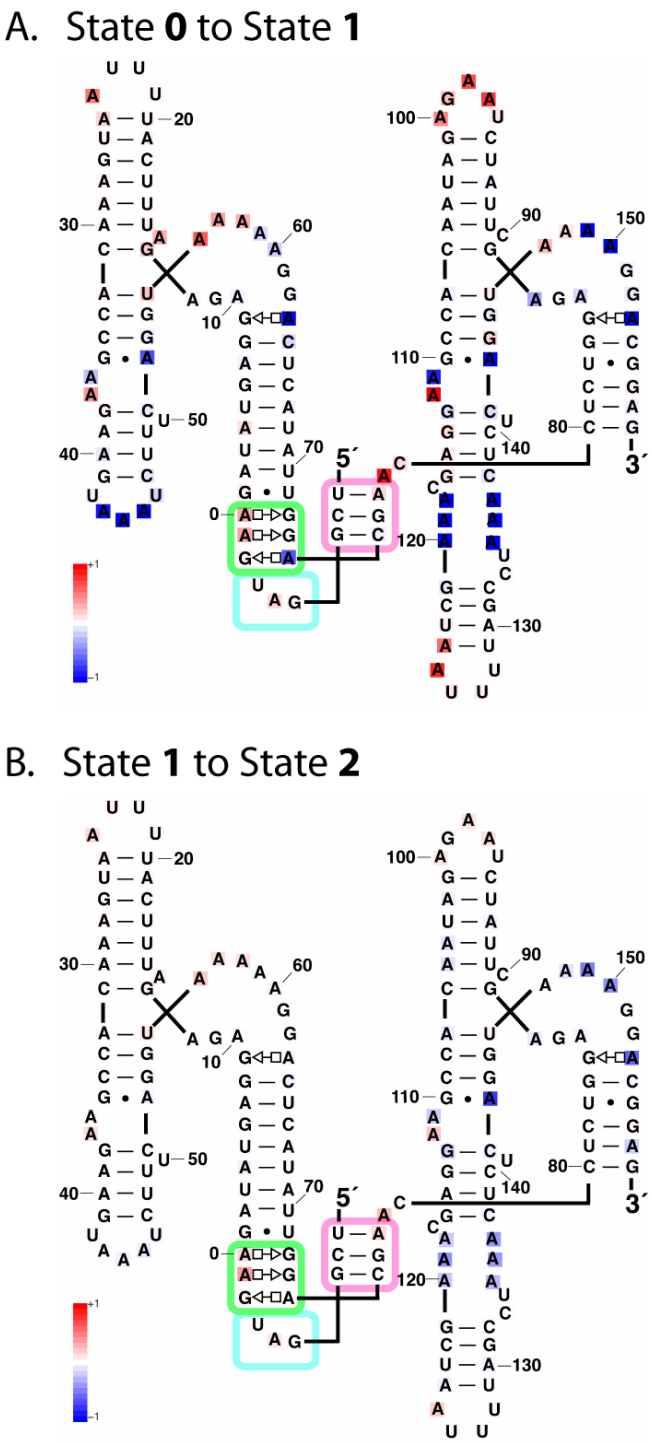
Supporting Figure S6. SHAPE reactivities of newly discovered linker element in the *FN* glycine riboswitch. Reactivity data for the FN-KT construct at the putative P0 stem (magenta, nts –6 and 75-77) fall within the distribution of reactivities seen for nucleotides that form Watson-Crick base pairs in the available crystal structure 3PN4 (black curve), while reactivities at the kink/turn bulge (cyan; nts –5 to –3) are higher, consistent with nucleotides unpaired in the available crystal structure (red curve). SHAPE data at the proposed purine/purine paired extension of aptamer P1 (green; nts –2 to 0 and 72–74) are intermediate in value. Data were taken at 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, and 10 mM glycine at 24 °C.



Supporting Figure S7. Effects of Mg^{2+} on *FN* riboswitch folding. Electropherograms give DMS reactivity data for FN-KT construct harboring the kink-turn linker (WT) and with mutations disrupting the linker (MutA). In the presence of glycine (10 mM) the WT and MutA RNAs have indistinguishable reactivities at all [$MgCl_2$] at regions outside the mutated linker. Note that as [Mg^{2+}] increases past 0.1 mM, both constructs show substantial protections in DMS reactivity, consistent with global secondary structure formation. In the absence of glycine, the same 0.1 mM Mg^{2+} transition occurs. At higher Mg^{2+} (1-10 mM), the WT RNA is seen to undergo further transitions (see marked A residues at 43–45, 63, and 135–137) that are not seen for the MutA RNA at [Mg^{2+}] $\leq$ 10 mM $MgCl_2$. Choosing a background of 0.5 mM Mg^{2+} for glycine titration permits the comparison of WT and MutA thermodynamics without the complication of this additional Mg^{2+}-induced, kink-turn-favored state. All measurements were in the presence of 50 mM Na-HEPES, pH 8.0 at 24 °C. Data at 1.0 mM Mg^{2+} were repeated on different days to demonstrate reproducibility.



Supporting Figure S8. DMS mapping data for glycine titrations of riboswitches and mutated variants. Capillary electropherograms, aligned by HiTRACE, are shown for the wild type, MutA, MutB, and MutAB variants of the FN-KT RNA, probed at 24 °C in 0.5 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0. Within each set, glycine concentrations were (from left to right in μM): 0, 0.1, 0.2, 0.3, 0.5, 1, 1.5, 2, 3, 5, 10, 15, 20, 30, 50, 100, 150, 200, 300, 500, 1000, 1500, 2000, 3000, 5000, 10,000, 20,000, 30,000, 50,000, 100,000, 200,000, and 500,000. Minor bands from nuclease contaminants are visible in wild type and MutA (lane 3). One replicate titration for each mutant is shown, and correspond to the data given in main text Fig. 3. Additional replicates (total of two for wild type and MutA; total of three for MutB and MutAB) were carried out for each RNA variant and used to determine fitted equilibrium constants.



Supporting Figure S9. DMS reactivity changes during glycine binding. Coloring gives change in reactivity between glycine-free state and single-glycine-bound state (A) and between single-glycine-bound state and double-glycine-bound state (B). Blue and red represent protections and increasing reactivity, respectively. The second transition involves conformational transitions mainly in aptamer II. The simplest model then is that the first transition involves glycine binding to the aptamer I site and concomitant formation of inter-aptamer contacts; and the second transition involves glycine binding to the aptamer II site. Supporting this model, disruption of the kink-turn and P0 in the inter-aptamer linker gives a 100-fold shift in the glycine requirement for the concerted inter-domain change.