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

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

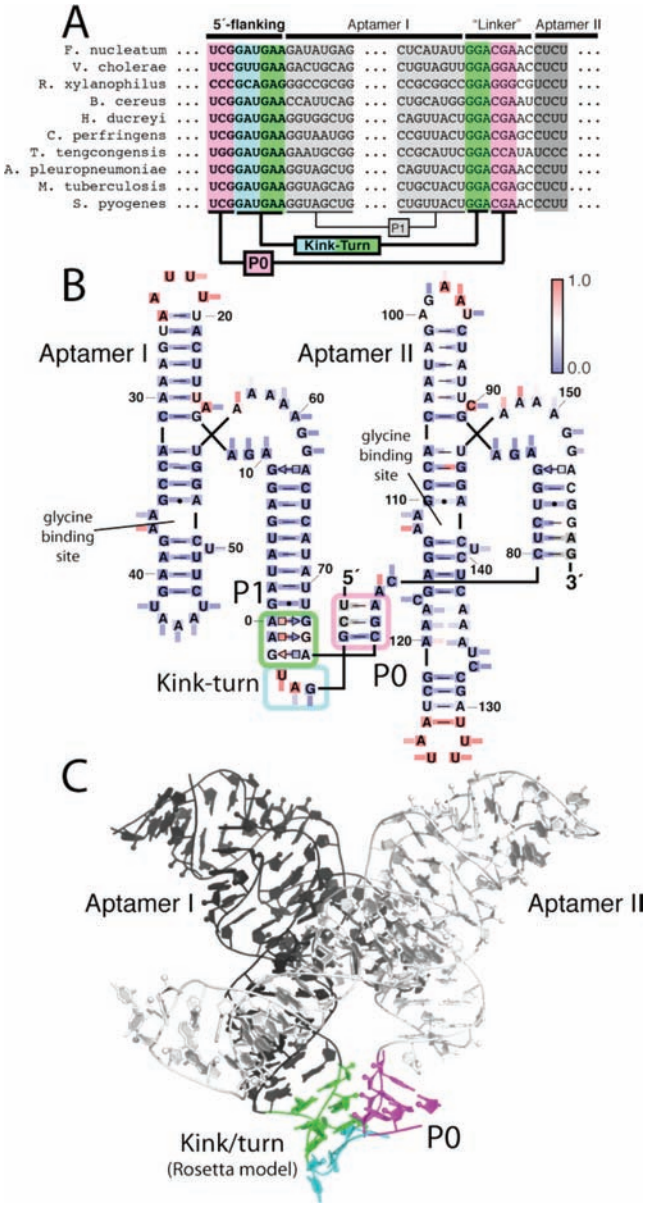

Figure 1. A new stem P0 and kink-turn tertiary element in double glycine riboswitches on a sequence alignment (A), on the secondary structure of the *FN* riboswitch (B), and built into this RNA's 3D crystallographic model (C). Coloring in (B) shows chemical mapping data: SHAPE (on letters), DMS (rectangles 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

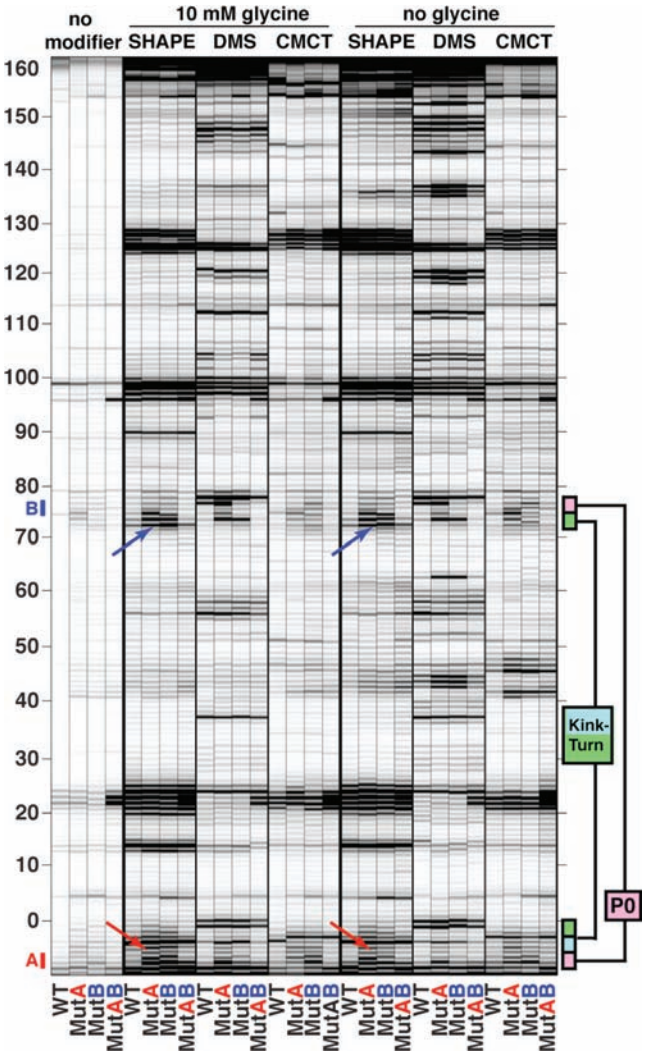

Figure 2. Evidence for the predicted kink-turn tertiary element and P0 stem from disruption of chemical reactivities by mutation (MutA and MutB) and restoration (MutAB) of P0 Watson–Crick pairs. Aligned electrophoretic traces are shown, from chemical mapping at 10 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0,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

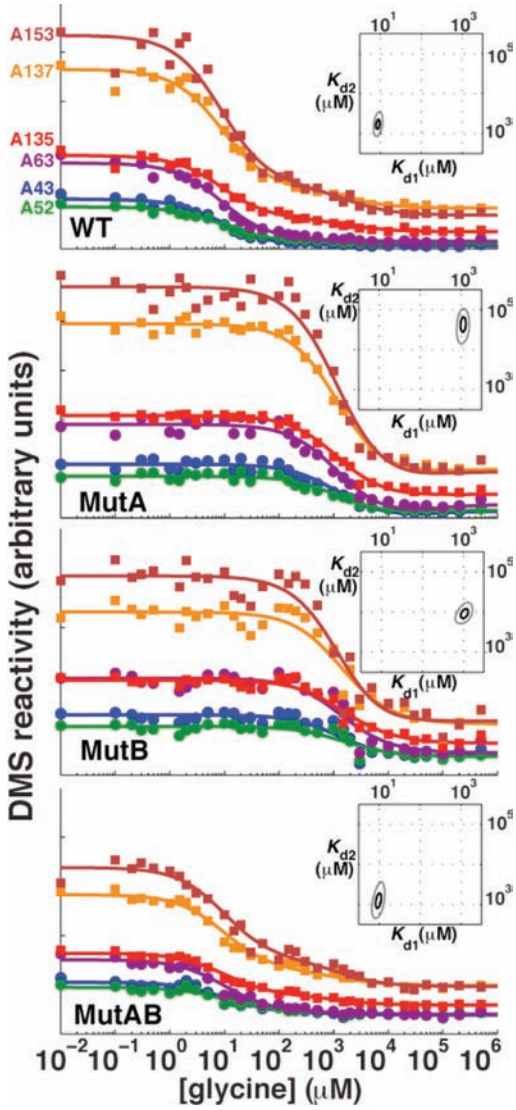

Figure 3. Thermodynamic analysis of glycine binding to riboswitches with (WT, MutAB) and without (MutA, MutB) the kink-turn. Symbols show DMS reactivities for six representative nucleotides (see top panel) at 0.5 mM $MgCl_2$, 50 mM Na-HEPES, pH 8.0, 24 °C; colored lines give calculated reactivities for 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.