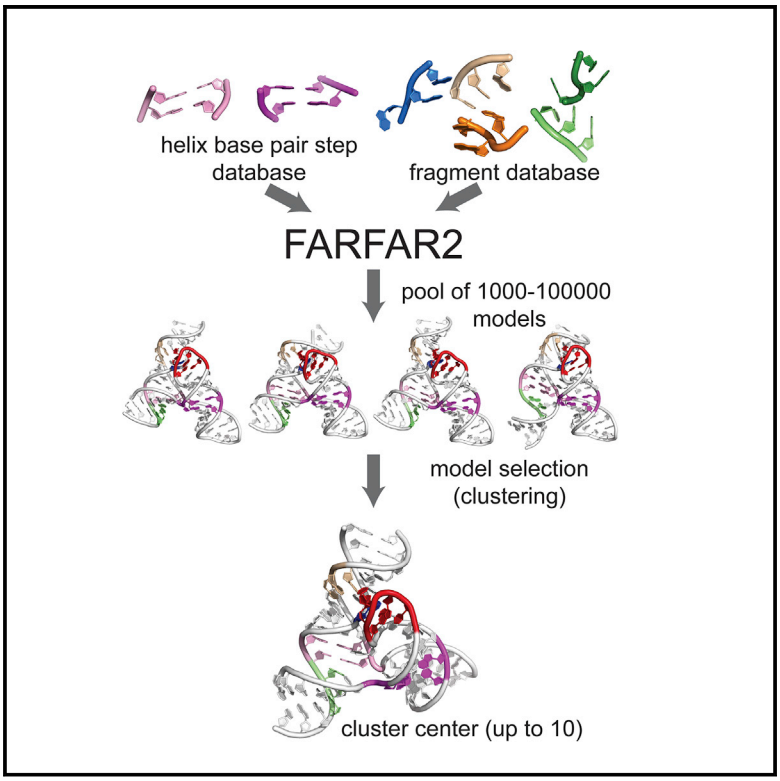
Structure

Resource

FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds

Graphical Abstract



Authors

Andrew Martin Watkins, Ramya Rangan, Rhiju Das

Correspondence

rhiju@stanford.edu

In Brief

Watkins et al. benchmark FARFAR2, an algorithm for RNA structure prediction that consolidates a decade of *ad hoc* strategies. FARFAR2 recovers native-like structures more accurately than original challenge submissions for 16 of 21 RNA-Puzzles and achieves native-like preregistered blind predictions for adenovirus VA-I RNA and five riboswitch complexes.

Highlights

- A new RNA fragment assembly method, FARFAR2, uses RNA-Puzzle-inspired innovations
- FARFAR2 recovers native-like structures on six new blind challenges
- A FARFAR2 webserver is available at https://rosie.rosettacommons.org/farfar2
- A dataset of millions of FARFAR2-predicted models is available for further study

Watkins et al., 2020, Structure *28*, 1–14
August 4, 2020 © 2020 Elsevier Ltd.
https://doi.org/10.1016/j.str.2020.05.011

CellPress

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

Structure

CellPress

Resource

FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds

Andrew Martin Watkins,[1] Ramya Rangan,[2] and Rhiju Das[1,2,3,*]
[1]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, USA
[2]Biophysics Program, Stanford University, Stanford, CA 94305, USA
[3]Lead Contact
*Correspondence: rhiju@stanford.edu
https://doi.org/10.1016/j.str.2020.05.011

SUMMARY

Predicting RNA three-dimensional structures from sequence could accelerate understanding of the growing number of RNA molecules being discovered across biology. Rosetta's Fragment Assembly of RNA with Full-Atom Refinement (FARFAR) has shown promise in community-wide blind RNA-Puzzle trials, but lack of a systematic and automated benchmark has left unclear what limits FARFAR performance. Here, we benchmark FARFAR2, an algorithm integrating RNA-Puzzle-inspired innovations with updated fragment libraries and helix modeling. In 16 of 21 RNA-Puzzles revisited without experimental data or expert intervention, FARFAR2 recovers native-like structures more accurate than models submitted during the RNA-Puzzles trials. Remaining bottlenecks include conformational sampling for >80-nucleotide problems and scoring function limitations more generally. Supporting these conclusions, preregistered blind models for adenovirus VA-I RNA and five riboswitch complexes predicted native-like folds with 3- to 14 Å root-mean-square deviation accuracies. We present a FARFAR2 webserver and three large model archives (FARFAR2-Classics, FARFAR2-Motifs, and FARFAR2-Puzzles) to guide future applications and advances.

INTRODUCTION

Noncoding RNA (ncRNA) molecules are essential to some of biology's most critical and ancient functions, such as translation (the ribosome), splicing (the spliceosome), and control of gene expression levels (riboswitches) (Cech and Steitz, 2014). Many ncRNAs exhibit intricately folded three-dimensional (3D) structures, but orders-of-magnitude more sequences of biologically interesting RNAs have been determined than high-quality RNA structures. For example, there are the thousands of classes of RNA domains that have been curated in the RFAM database that do not have experimentally solved structures (Kalvari et al., 2018). Therefore, computational methods to predict 3D ncRNA structure could be of substantial value. Current computational methods for 3D ncRNA structure prediction include coarse-grained molecular dynamics simulation (Vfold3D [Zhao et al., 2017] and iFoldRNA [Krokhotin et al., 2015]), Monte Carlo (Rosetta FARFAR [Das and Baker, 2007] and SimRNA [Boniecki et al., 2015]), and motif assembly (RNAComposer [Popenda et al., 2012]; 3dRNA [Jian et al., 2017]; MC-Fold [Parisien and Major, 2008]).

Over the last 9 years, these methods have been tested and advanced through the RNA-Puzzles community-wide blind trials (Cruz et al., 2012; Miao et al., 2015, 2017, 2020). RNA-modeling methods using the Rosetta software (Das and Baker, 2007; Das et al., 2010) motivated the launch of these trials (Sripakdeevong et al., 2012) and have achieved the most accurate models for a plurality of cases to date (10 of 21), including ligand-binding riboswitches, ribozymes, and other ncRNAs with complex folds with lengths of 41–188 nucleotides (nt) (Table 1). The primary Rosetta modeling tool underlying these predictions is called FARFAR (Fragment Assembly of RNA with Full-Atom Refinement) and was inspired by Rosetta's protein structure prediction methods. FARFAR first models an RNA structure by stitching together three-residue fragments of previously solved RNA structures whose sequence matches the target sequence. This Monte Carlo process is guided by a low-resolution scoring function that rewards base pairs and base stacks with geometries similar to those seen in previously solved RNA structures (Das and Baker, 2007). Each model is then refined in a high-resolution all-atom scoring function that rewards hydrogen bonds, van der Waals packing of atoms, and other physically important interactions, and the lowest-energy models are clustered to achieve submitted models (Cruz et al., 2012; Das et al., 2010; Miao et al., 2015, 2017).

While consistently successful in RNA-Puzzles, FARFAR modeling has involved problem-specific expert intuition, such as guesses of ligand-binding sites based on inspection of sequence alignments, as well as algorithmic extensions created "on-the-fly" to explore novel ideas inspired by the targets, including different approaches to model A-form helices or tertiary contacts (Cruz et al., 2012; Miao et al., 2015, 2017). As in early days of protein structure prediction (Sali, 1995; Simons et al., 1997), many of these steps have not been well

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

documented, automated, or systematically benchmarked, so the performance and current limitations of FARFAR remain uncertain. Indeed, the original FARFAR study as well as subsequent studies (Das et al., 2010; Ding et al., 2008; Laing and Schlick, 2010; Popenda et al., 2012) suggested that FARFAR would have difficulty sampling native-like folds for RNA molecules larger than 12 nt. In addition to learning whether native-like folds can be discovered through sampling, a systematic FARFAR benchmark is needed to provide modeling pools for groups developing complementary procedures, including recent scoring methods taking advantage of artificial neural networks and evolutionary coupling information (Li et al., 2019; Wang et al., 2019; Weinreb et al., 2016) and alternative high-resolution refinement procedures (Tan et al., 2018; Watkins et al., 2018) that might be pipelined with FARFAR modeling.

This paper seeks to address these gaps in Rosetta RNA modeling. First, we describe how improvement of the general Rosetta codebase as well as consolidation of the FARFAR protocol have enabled development of a streamlined version of the method, named FARFAR2, and an improved webserver on the ROSIE platform (Lyskov et al., 2013; Moretti et al., 2018) at https://rosie.rosettacommons.org/farfar2. We confirm that the method recapitulates and extends prior results on two benchmarks of small RNA folds and RNA submotifs, here revisited as the FARFAR2-Classics and FARFAR2-Motifs model sets. We then present a benchmark of FARFAR2 that revisits every previous RNA-Puzzle for which there exists a deposited experimental structure at the time of this study. This FARFAR2-Puzzles benchmark uses only secondary structure information and template structures employed at the time of the original challenge. Finally, as independent validation, we present blind tests of the method based on three independently solved structures of riboswitch aptamers based on cryoelectron microscopy (cryo-EM) maps (Kappel et al., 2019), an adenoviral ncRNA presented as RNA-Puzzle 24 (Hood et al., 2019), and two T-box riboswitch structures presented as RNA-Puzzles 26 and 27. The results confirm the ability of Rosetta FARFAR2 to sample complex global folds of RNAs while also highlighting current limitations in discriminating the most accurate model among a pool of models, which will likely require complementary approaches to resolve. The input files and output models from this study, including more than 10 million structures available in a single archive, provide a rich resource that we expect to be valuable in developing approaches that extend or go beyond Rosetta FARFAR2.

RESULTS

Consolidated RNA Fragment Assembly Protocol Improves Modeling of Small RNAs

Our core goal in developing the Rosetta FARFAR2 protocol has been to achieve a single application that enables straightforward modeling of complex RNA structures with sizes up to 200 nt, incorporating any available additional knowledge. Our previous attempts to construct a comprehensive modeling pipeline required several manual steps running a series of distinct Rosetta applications, such as pregenerating helix ensembles, set up through a separate Python script (Cheng et al., 2015; Watkins et al., 2019). The FARFAR2 protocol (Figure 1) is designed instead to take input information in as simple a manner as possible into a single Rosetta executable, *rna_denovo*. Analogous to other RNA-modeling packages (Krokhotin et al., 2015; Piatkowski et al., 2016; Popenda et al., 2012), the *rna_denovo* executable now accepts the RNA sequence, the RNA secondary structure in community-standard dot-parentheses notation, and, if available, the names of PDB-formatted files holding template structures of any known submotifs or subdomains.

FARFAR2 also implements four methodological improvements. It uses an updated library of fragments, based on the nonredundant 2018 crystallographic database of 657 RNA structures (Leontis and Zirbel, 2012), which is 15% larger and more diverse than the previous default fragment library from 2009 (Richardson et al., 2008). Score filters during fragment assembly allow recognition of poorly assembled conformations that can be discarded before the computationally expensive all-atom minimization, leading to more efficient use of computational power (Figure 1 and STAR Methods). FARFAR2 also implements a special set of Monte Carlo moves for nucleotides in stacked Watson-Crick pairs (“base-pair steps”) that maintain Watson-Crick geometry of RNA helices while allowing their backbone conformations to be perturbed, drawing on the same crystallographic database (see STAR Methods). Lastly, during the minimization stage, the protocol uses an updated all-atom scoring function developed in a recent study seeking high accuracy on small RNA noncanonical motifs (Watkins et al., 2018).

As initial tests of this FARFAR2 protocol, we measured its performance on two benchmarks involving small RNAs developed in prior work. We revisited the original Rosetta RNA benchmark of 18 small RNA problems (Das and Baker, 2007), generating a set of 25.2M total models (1.4M models per problem) that we term the FARFAR2-Classics dataset. These results confirmed that the updated fragment library, minimization scoring function, and mode of Watson-Crick base-pair modeling give improved results over the original Rosetta fragment assembly method, with the most notable improvements arising from full-atom refinement (Figure S1). Full results are given in Table S2; in particular, the “hires” scoring function, base-pair constraints, and old fragment library taken together represent a reproduction of the performance of the original FARFAR code. We assessed performance using the “native-like” standard applied in the original work (Das and Baker, 2007): achieving folds with root-mean-square deviation (RMSD) to experimental structure of better than 4 Å in the top five cluster centers (5,000 low-energy models clustered with a 3.0 Å radius). By this metric, FARFAR2 succeeded in 15 of 18 cases, better than the original results of 10 of 18. Figures 2A–2E show native-like models achieved by FARFAR2 for the five cases for which the original study did not sample such folds. For 1A4D (Figure 2A), a nuclear magnetic resonance (NMR) structure of the loop D/loop E arm of the *Escherichia coli* 5S rRNA, the FARFAR2 model correctly recovers 11 consecutive base pairs, only one of which is a canonical Watson-Crick base pair. For 1CSL (Figure 2B), the HIV RRE (Rev response element) high-affinity site, FARFAR2 recovers a bent geometry and both “bulged-out” nucleotides. For 1I9X (Figure 2C), the branchpoint duplex from U2 small nuclear RNA, FARFAR2 recovers a nearly atomic-accuracy model (2.5 Å RMSD). For 1KKA (Figure 2D), an NMR structure of the unmodified anticodon stem loop from tRNA-Phe, FARFAR2 obtains a model with a correct geometry for the unusually twisted helix, as well as a

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

CellPress

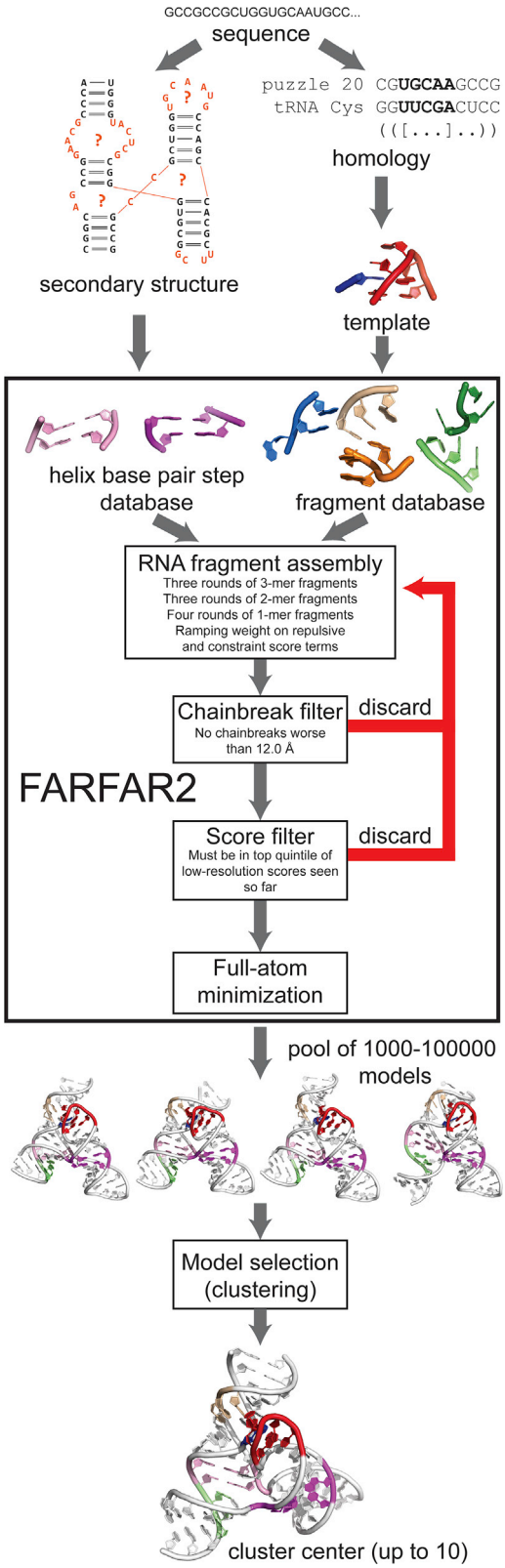


Figure 1. The FARFAR2 Structure Prediction Algorithm

A 3D structure prediction problem is specified by RNA sequence; from that sequence, a consensus secondary structure is obtained from prior literature studies or covariance analysis of sequence alignments (left), and homologies may be identified to previously solved structures (right). The orange areas in the depicted secondary structure diagram represent the regions whose conformations are unknown a priori and whose solution would guide the tertiary structure prediction. Manually identified homologies can also furnish template structures, which are combined by automatic sampling from a base-pair step and fragment database in a low-resolution fragment assembly stage. Subsequent models are filtered to omit trajectories with chain breaks or poor scores, and passing models are subjected to minimization in an all-atom scoring function. Finally, models are chosen from the resulting ensemble through clustering.

geometry for the apical loop that lacks several clashes present in the database-deposited coordinates. Finally, for 2A43 (Figure 2E), a pseudoknot from luteovirus, FARFAR2 recovers the A-minor motif that anchors the pseudoknot fold. We note that for this and the other two benchmarks below, we sought to simulate a realistic prediction scenario by stringently excluding any extended fragments in the fragment library that were structurally similar to the target structure (see Fragment Homology Exclusion in STAR Methods). As a result of this fragment exclusion, FARFAR2 performance in real targets may be underestimated by these benchmarks, and we return to this point below.

As a higher-resolution test, we evaluated FARFAR2 on a benchmark of noncanonical RNA motifs (apical loops, internal loops, junctions, and tertiary contacts) extracted from larger RNA structures (Das et al., 2010). Recently, we reported that a nucleotide-by-nucleotide build-up method called stepwise Monte Carlo (SWM) outperformed FARFAR for intricate noncanonical loops (Watkins et al., 2018). However, for problems with longer loops or that require positioning of distinct helical elements in tertiary contacts, the SWM method transited through physically unreasonable intermediate conformations (Watkins et al., 2018), and FARFAR achieved better RMSD accuracies than SWM in these cases. Here we revisited this comparison with FARFAR2; the resulting 820,000 models (10,000 models per problem) comprise the FARFAR2-Motifs dataset. As observed previously, SWM achieved a 1.5 Å model among the top five cluster centers in more problems than FARFAR2 (42 compared with 37 out of 82; see Figure S2 and methods for clustering details). Nevertheless, cases in which FARFAR2 outperformed SWM supported the continuing use of FARFAR2 for modeling complex RNA folds with long loops or tertiary contacts for which the partners' relative positions are uncertain (Figures 2F–2J). An overtwisted helix P5b from the P4-P6 domain of *Tetrahymena* ribozyme, the loop E motif from *E. coli* 5S rRNA, and the kink-turn motif each involve concomitant modeling of two strands with lengths up to 9 nt, and SWM had difficulty building up complete solutions for these loops (best cluster center RMSDs of 2.7 Å, 1.7 Å, and 2.1 Å, respectively). In contrast, the best of five cluster centers from FARFAR2 did achieve subangstrom recovery of these motifs (0.76 Å, 0.72 Å, and 0.97 Å RMSD, respectively; Figures 2F–2H). For tertiary contacts in which the relative positioning of partners had to be modeled *de novo*, current SWM procedures for docking the partner segments gave poor accuracies, e.g., for an A-minor tertiary contact from the lariat-capping GIR1 ribozyme and from the tetraloop-receptor contact of the P4-P6 RNA (1.8 Å and 3.0 Å, respectively). Fragment-based FARFAR2 recovered these structures with excellent accuracies of 1.2 Å and 0.81 Å RMSD, respectively

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

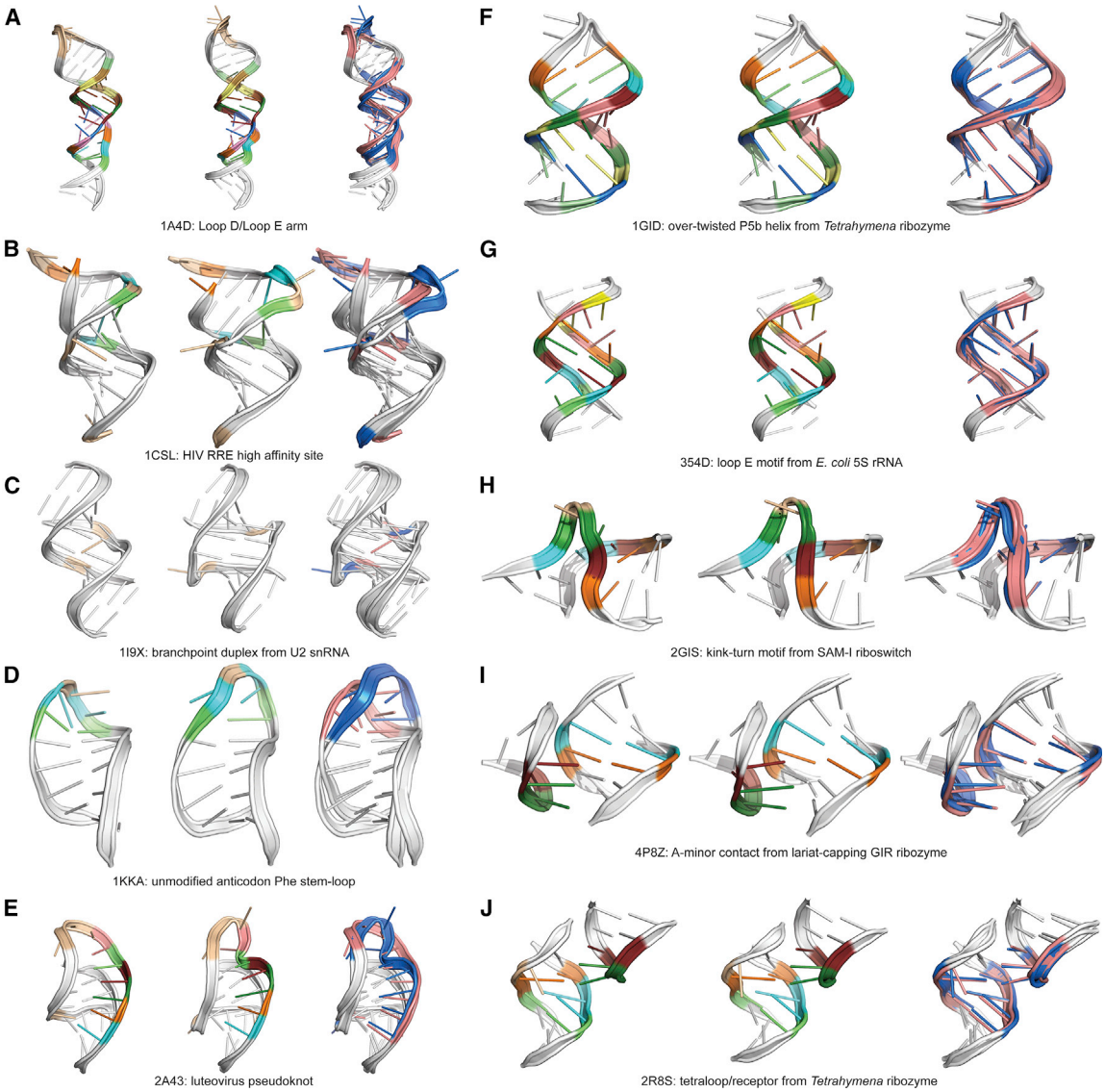


Figure 2. Increased Accuracy of FARFAR2 on Existing Structure Prediction Benchmarks

Cases from the FARFAR2-Classics (A–E) and FARFAR2-Motifs (F–I) benchmarks that saw success from the application of FARFAR2 instead of FARNA and SWM, respectively. In each panel, FARFAR2 model, native structure, and overlay are shown from left to right. In (A) to (E), the FARFAR2 model is the best of five low-energy cluster centers after clustering 5,000 models with a 3.0 Å radius; in (F) to (J), the FARFAR2 model is the best of five low-energy cluster centers after clustering 400 models with a 2.0 Å radius. In each case, model selection conditions reproduce the conditions used in the original publications (Das and Baker, 2007; Watkins et al., 2018). In overlays, the FARFAR2 model is colored in salmon and the experimental structure in blue; in individual structures, recovered noncanonical base pairs are colored in cyan, lime, orange, salmon, and ruby, and recovered bulged residues are colored in wheat. In (A) to (E), residues from prespecified, flexible helices are colored white; in (F) to (J), fixed input residues (mostly helical) are colored white.

(Figures 2I and 2J). Because such long-looped junctions and tertiary contacts appear frequently in complex RNA folds, these results have motivated us to continue to use FARFAR2 as our default procedure for such modeling cases (Watkins et al., 2018, 2019).

The RNA-Puzzles Benchmark

The aforementioned findings on prediction accuracies for small RNAs and RNA motifs motivated us to test FARFAR2 against larger RNA structures with many motifs and complex folds. Through our participation in the RNA-Puzzles trials (Cruz et al., 2012; Miao et al., 2015, 2017), we have kept records of our strategies for each prediction challenge, including secondary structure predictions, inferences based on homology to prior deposited structures, and functional constraints, such as sites of self-cleavage in ribozyme challenges. We ran FARFAR2 for each of the 21 problems for which an experimental structure is now available (Magnus et al., 2019), again using procedures to ensure that fragments from these structures or their homologs could not be sampled, providing "like-blind" modeling conditions (see Fragment Homology Exclusion in STAR Methods). This set comprised all RNA-Puzzles from RNA-Puzzle 1 through RNA-

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

Puzzle 21 except RNA-Puzzle 16. We treated bound and unbound states of RNA-Puzzle 14 (a riboswitch aptamer for glutamine) as separate problems. The benchmark included both single-stranded and multistranded RNAs, as well as problems for which considerable homology was available and puzzles where we had to start from only our original secondary structure predictions (Table S3). We made no explicit provision for ligand binding, save implicitly when the ligand-binding site was part of a template structure, since augmenting the low-resolution fragment assembly stage of modeling with a concurrent ligand docking protocol would require substantial additions to the FARFAR2 algorithm.

Our primary question was whether FARFAR2 samples native-like global structures for these RNA-Puzzles with complex folds. For each of the RNA-Puzzles problems, we therefore generated 3,000–30,000 FARFAR2 models, involving approximately 6–48 h of computation on 500 CPUs. To evaluate conformational sampling efficiency, we mimicked our typical protocol in RNA-Puzzles trials, in which we inspect approximately 100–200 low-energy models as potential candidates for submission, corresponding to the lowest 1% of models by Rosetta all-atom energy. To assess whether any of the large RNA models were native-like, we translated the 4.0 Å RMSD threshold used in our assessment of the smaller RNAs of the FARFAR2-Classics to these larger challenges, making use of a previous extension to the length-dependent $RMSD_{100}$ metric, which we have found to accord well to visual assessments of native-like folds (Carugo and Pongor, 2008; Kappel and Das, 2019). A 4.0 Å RMSD on FARFAR2-Classics (median length 26 nt, all of which are built *de novo*) is equivalent to a 9.1 Å RMSD on FARFAR2-Puzzles (median length 71 nt that need to be built *de novo*; see STAR Methods), and so we chose RMSD 9.1 Å as our cutoff for most of the RNA-Puzzles. However, the three longest problems were substantially longer (117, 130, and 185 nt must be built for RNA-Puzzles 12, 5, and 7, respectively). Applying the same $RMSD_{100}$ criterion, we used 13.8 Å RMSD to assess whether models of these longest problems were native-like. In addition to setting this absolute RMSD-based cutoff for evaluation, we compared FARFAR2 performance relative to the original performance by all competitors when the blind challenge was originally issued. This latter evaluation enables comparison with expert execution of every leading method for RNA structure prediction by their authors rather than our best-effort reproduction.

Using these evaluation criteria, the FARFAR2 protocol performs well. The protocol is able to sample native-like models within its best 1% by Rosetta all-atom free energy (30–300 models) for 19 of 21 cases. The agreement across all cases is visually apparent, as illustrated in Figure 3. In 16 of 21 RNA-Puzzles, FARFAR2 samples a model within its best 1% by energy closer to native than the best originally submitted model during the actual RNA-Puzzles trial (20–100 models; Table 1 and Figure 4). In two additional cases, FARFAR2 outperforms previous Rosetta-based submissions from our group; detailed depictions of prior models are in Figures S3–S5. These results indicate that FARFAR2 emulates or exceeds prior performance in RNA-Puzzles, and specific cases illustrate the complexity of the folds that are recovered. In the largest and most challenging problems, FARFAR2 models of RNA-Puzzle 7, the Varkud satellite (VS) ribozyme, recapitulated the experimental folds accurately up to one missed interhelical angle (Figure 3). Improvement was also seen in several ribozyme structures of moderate size (Puzzles 15, 17, 19, 20). These molecules—a hammerhead and pistol ribozyme and two bimolecular twister sister constructs—each feature a compact multiway junction and a key tertiary contact. The pistol ribozyme features a pseudoknot, while the other three possess intercalated T loops, and, combined with other interconnections, these features lead to slight under- or overtwisting of helices in experimental structures compared with ideal A-form helices.

These observations suggested that the improvements seen in FARFAR2 relative to the original RNA-Puzzles submissions might be due to improvements in helix modeling through "base-pair step" fragments. A detailed comparison of the effects of base-pair step sampling on Puzzle 21, a guanidinium riboswitch, is shown in Figure S6. A comparison of FARFAR2 performance with base-pair step fragments versus fixed helices across multiple large puzzles (RNA-Puzzles 3, 5, 6, 9, 12, 15, 17, and 18) confirms the efficacy of the base-pair step helix modeling scheme (see Figure S7). To further test the importance of base-pair steps, we repeated modeling of RNA-Puzzle 15, a hammerhead ribozyme, with every helix modeling scheme we have previously employed in Rosetta fragment assembly: none, as in FARNA; guiding formation of helices through harmonic base-pairing constraints (Das et al., 2010); base pairs of fixed geometry and kinematics through the -fixed_stems flag; use of pregenerated ensembles of fixed helices (Cheng et al., 2015), use of single fixed helices of ideal geometry; and base pair step sampling. The results confirmed that base pair step sampling yields the most accurate models (Figure 5). To test the effects of other improvements in Rosetta FARFAR2 compared with the original FARFAR study, we also carried out a test on a sample of the five most recent RNA-Puzzles, emulating the original FARFAR simulation conditions ("hires" scoring function, the original fragment library, and use of harmonic base-pair constraints rather than sampling base-pair steps to model helices). This direct comparison of FARFAR2 with original FARFAR simulation settings indicated strong improvements, with differences in achieved RMSD of up to 7 Å in best 1% by energy RMSD and 8 Å in top-ten cluster RMSD (Table S4). Analysis of subsets of the simulated model ensembles suggests that FARFAR2 simulations demand 10–30 times fewer structures than FARFAR to generate native-like structures (Figure S8).

While we saw improvements by FARFAR2 over RNA-Puzzles efforts in 16 of 21 modeling challenges, in five of these modeling challenges, FARFAR2 did not sample, among its top 1% by energy, a more accurate model than the original RNA-Puzzles submissions. Four of these cases, RNA-Puzzles 12, 13, 14 (bound), and 21, represent structures of riboswitch aptamers with their small-molecule ligands (cyclic diAMP, ZMP, glutamine, and guanidinium, respectively). For these cases, the original modeling involved manual identification of ligand-binding residues from sequence conservation and, in some cases, explicit modeling of the ligand-binding site (Miao et al., 2017, 2020). The fifth case was the Zika xrRNA (exonuclease resistant RNA), and the best submission for RNA-Puzzle 18 was solved with the stepwise Monte Carlo method, which is our preferred method for modeling challenges limited to individual motifs within complex

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

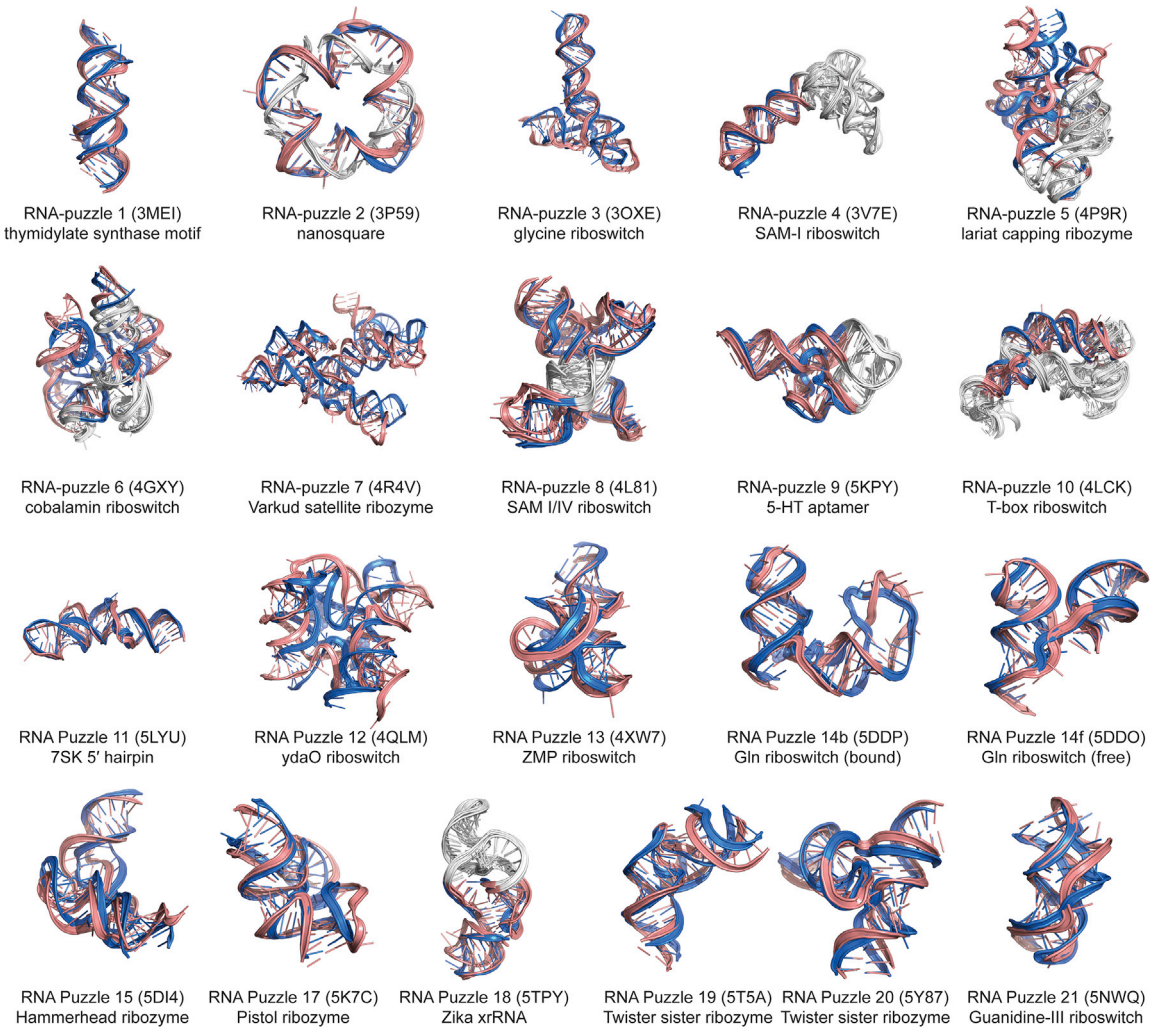


Figure 3. The Best-of-Top-1% RMSD Prediction (Pink) versus Native (Blue) for Each RNA-Puzzle in the FARFAR2-Puzzles Benchmark
White regions are input template structures employed at the original time of modeling and employed in the FARFAR2 simulation.

RNA folds (Watkins et al., 2018). In that case, FARFAR2 still does outperform the best original submission based on the FARFAR implementation available at the time. These results suggest that FARFAR2 successfully automates or improves upon the *ad hoc*, manual steps used with FARFAR for prior RNA-Puzzles challenges but could be improved further if ligand-binding hypotheses and stepwise Monte Carlo could be incorporated into the modeling.

To compare FARFAR2 results with the original RNA-Puzzle submission process more directly, we needed a method of model selection to obtain a final set of ten models from the ensemble of sampled models. Here, we selected a conventional protocol for analyzing problems of this size within Rosetta, clustering the top 400 models with a 5.0 Å cluster radius. (We tested alternative clustering methods and obtained similar or slightly worse results; see STAR Methods.) The best of ten clustered FARFAR2 models outperformed the original 5–10 Das lab submissions in 10 of 21 cases, and, despite no manual intervention in model selection, was less than 1.0 Å worse in RMSD in an additional four. Furthermore, these selected models were native-like (RMSD accuracy better than 9.1 Å for short problems or 13.8 Å for long problems) in 16 of 21 cases. For each of these models, we also compute the fraction of non-Watson-Crick base pairs recovered and the clashscore (Table S6). The best FARFAR2 cluster center outperforms the prior most accurate submission in clashscore in 15 of 21 cases. The best FARFAR2 cluster center has superior non-Watson-Crick base-pair recovery in 9 of 21 cases, equal recovery in one case, and is less than 10% worse in three cases. These results further suggest that FARFAR2 followed by a simple clustering procedure will give similar accuracy to prior ad hoc modeling and model selection procedures used in RNA-Puzzles.

Current Bottlenecks

In addition to testing whether FARFAR2 emulates our RNA-Puzzles modeling protocols for complex RNA folds, the extensive FARFAR2-Puzzles dataset enables analysis of what is currently bottlenecking FARFAR2 accuracy for recovering these folds (Figures S9–S11). Structure prediction algorithms face two potential limitations: inadequate conformational sampling, which

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

Table 1. Results for Each RNA-Puzzle Challenge Revisited in This Work

Puzzle	PDB	Length (Total)[c]	No. of Structures Generated	RNA	FARFAR2 Top 1% Best RMSD (Å)[b]	FARFAR2 Best of 10 Low-E Cluster RMSD (Å)[b]	Best RNA-Puzzle RMSD (All Submissions) (Å)[b]	Best RNA-Puzzle RMSD (Das Submissions) (Å)[b]
1	3MEI	46 (46)	14,052	thymidylate synthase motif	2.03	2.50	3.40	3.40
2	3P59	60 (100)	43,431	nanosquare	2.28	2.71	2.30	2.45
3	3OXE	84 (84)	33,442	glycine riboswitch	7.05	12.41	7.60	12.10
4	3V7E	40 (126)	5,768	SAM-I riboswitch	2.43	2.52	3.40	4.50
5	4P9R	130 (188)	24,903	lariat-capping ribozyme	9.57	13.94	9.58	9.58
6	4GXY	98 (158)	28,859	cobalamin riboswitch	9.98	13.08	11.57	11.57
7	4R4V	185 (185)	7,963	VS ribozyme	15.21	18.52	20.56	20.56
8	4L81	71 (96)	33,086	SAM I/IV	4.65	5.23	4.80	4.80
9	5KPY	40 (71)	18,660	5-HT aptamer	4.54	4.56	6.05	7.70
10	4LCK	83 (177)	5,873	T-box riboswitch	6.31	6.31	6.78	6.78
11	5LYU	57 (57)	41,952	7SK 5′ hairpin	4.43	6.04	4.96	6.83
12	4QLM	117 (117)	35,506	ydaO riboswitch	11.73	13.32	10.15	12.63
13	4XW7	60 (60)	20,297	ZMP riboswitch	5.47	7.13	5.41	5.41
14b	5DDP	61 (61)	24,531	Gln riboswitch (bound)	5.81	6.88	5.79	6.59
14f	5DDO	61 (61)	15,112	Gln riboswitch (free)	3.26	11.85	6.05	6.05
15[a]	5DI4	68 (68)	6,123	hammerhead ribozyme	4.44	5.98	5.30	9.83
17	5K7C	58 (58)	16,529	pistol ribozyme	5.03	6.69	5.15	7.16
18	5TPY	34 (71)	17,091	Zika xrRNA	4.29	5.02	3.15	3.15
19	5T5A	56 (62)	4,499	twister sister ribozyme	4.86	5.16	5.50	8.54
20	5Y87	62 (68)	1,547	twister sister ribozyme	3.03	4.03	4.67	5.94
21	5NWQ	41 (41)	48,146	guanidinium-III riboswitch	4.40	6.04	3.93	3.93

See Table S4 for additional figures of merit.

[a]The native structure in RNA-Puzzle 15 was corrupted by a crystal contact that caused a strand-swapped pairing intractable to prediction. We reconstructed the strand-unswapped monomer and rescored original submissions to present a more realistic modeling challenge.

[b]All-heavy-atom RMSD is calculated over all residues, following superposition over all residues. The 400 models with the lowest (best) energies are clustered.

[c]Length is number of residues modeled *de novo* (not found in input templates); total also counts template residues.

CellPress

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

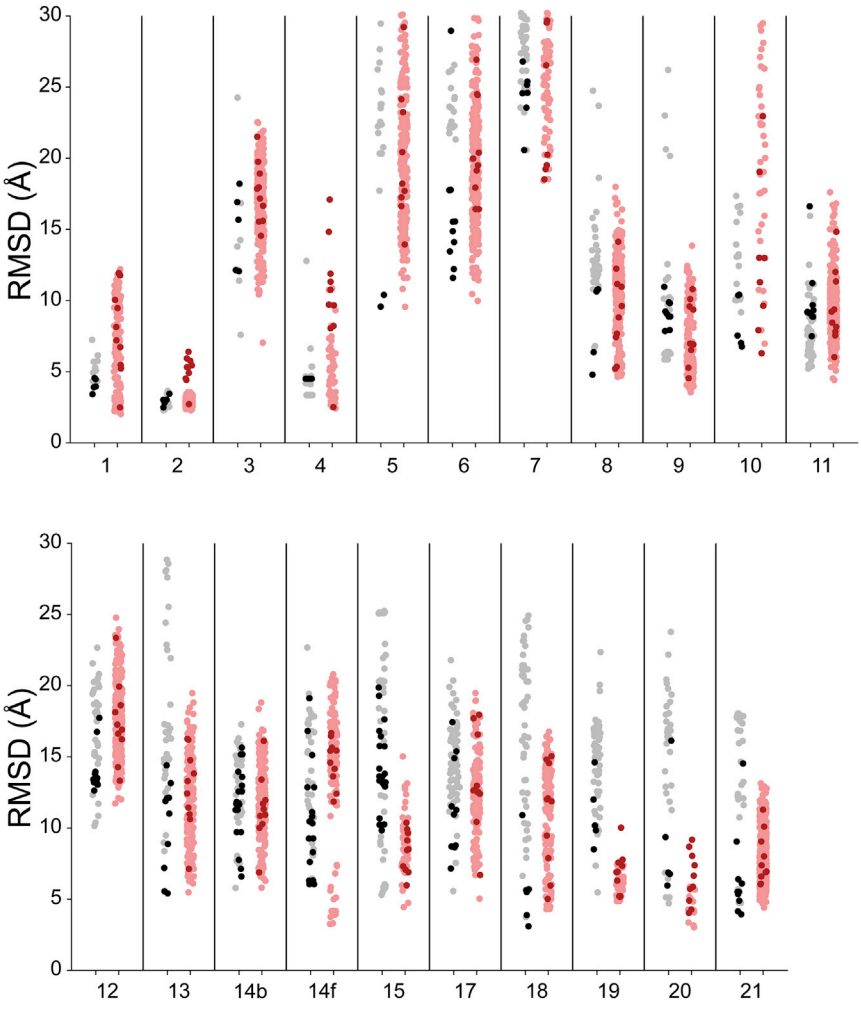


Figure 4. Direct Comparison between Original RNA-Puzzles Submissions from All Groups (Left Points) and FARFAR2 Models (Right Points) for Each Benchmark Case

Among RNA-Puzzles submissions, those from the Das lab (created using manually curated Rosetta models, mostly using FARFAR) are black points; others are gray. Among FARFAR2 models, pink points are the top 1% of models by energy; dark red are cluster centers.

could leave the true global minimum unexplored, and a scoring function that incorrectly favors minima different from the experimental structure. The limitations are not necessarily exclusive, and we evaluated these in turn.

A hallmark of poor conformational sampling in a stochastic modeling procedure is the inability to find similar solutions from independent runs, where similarity here is evaluated as model-to-model RMSD of less than 5 Å. We noted that for modeling problems involving greater than 80 nt built *de novo*, poor conformational sampling was clearly an issue. In each of these six cases (RNA-Puzzles 3, 5, 6, 7, 10, and 12), none of the top ten clusters included more than one model in total (Table S5; Figure S11A). This set of long problems also included models with the worst RMSD accuracy achieved, although in each case the accuracies were still strikingly better than what would be expected from random sampling (Figures 3 and S11B). For nearly all problems smaller than 80 residues, independent FARFAR2 runs produced low-energy models that were clustered together in one of the top ten clusters, and the majority of these cases exhibited better than 5 Å RMSD accuracies and visually correct backbone folds (Figure 3). These results suggest a "rule of thumb": problems involving *de novo* building more than 80 nt will be challenging for FARFAR2 conformational sampling.

Assessment of scoring function limitations can be carried out by comparing the predicted energy of the experimental structure of a molecule with structurally distinct models. If "decoy" models give better energies than near-native structures, there is a problem with the scoring function. In an effort to exactly match bond geometries and modeling assumptions of our FARFAR2 simulations, we generated near-native models (<1.0 Å RMSD) using additional FARFAR2 simulations that we augmented with restraints to the native structure (see STAR Methods). We found that in slightly more than half of the cases (11 of 21), FARFAR2 decoys far from the experimental structure gave energies better than those achieved in the near-native FARFAR2 simulations that were restrained to produce near-experimental solutions (Figures S9–S11C and Table S5). This result indicates that our scoring function is problematic for modeling complex folds and suggests that more extensive sampling on those cases would not result in more accurate modeling. Surprisingly, we found that even for five of the six larger modeling problems (>80 nt) that currently show poor sampling (noted above), FARFAR2 discovered solutions with lower energies than the near-native structures (Figure S11C). One possible explanation for this result is that our use of FARFAR2 simulations restrained to the experimental structures gives poor energy estimates for "near-native" conformations due to inaccuracies in the experimental structures (many of which were solved before the advent of RNA-specific refinement tools for crystallography) or the use of ideal bond lengths/angles in FARFAR2 simulations. Nevertheless, having an scoring function that is less sensitive to fine conformational details and that more accurately discriminates near-native conformations from non-native conformations of complex RNA folds is an important goal for future work, further discussed below.

Estimating Modeling Error for FARFAR2

The large model datasets collected in the FARFAR2-Puzzles benchmark above provide additional information in predicting model accuracy not just generally but also for specific cases. In actual blind modeling or biological prediction scenarios, it is important to estimate how accurate a predicted FARFAR2 model

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

CellPress

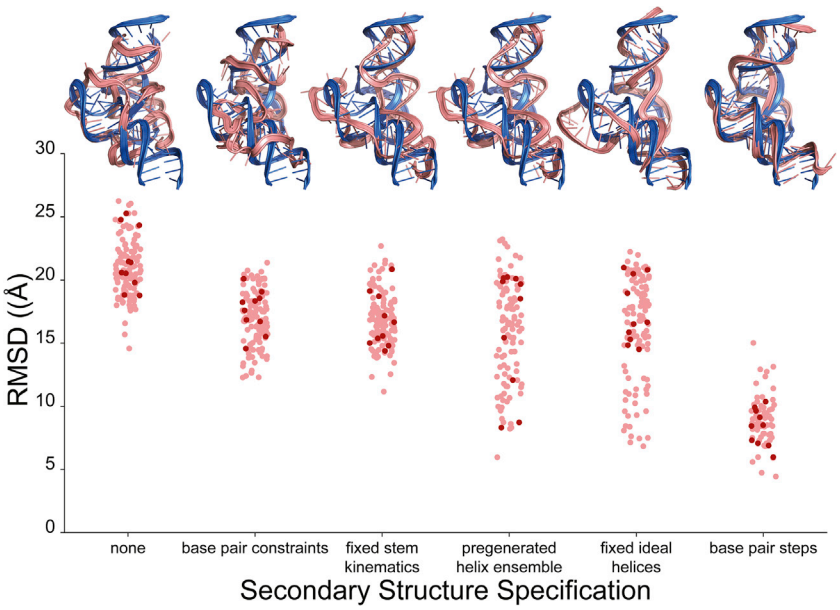


Figure 5. A Detailed Study of Secondary Structure Specification Methods for RNA-Puzzle 15.
The method of secondary structure specification affects the quality of the resulting predicted models, as illustrated on RNA-Puzzle 15, a hammerhead ribozyme. For each method, the native structure is depicted in blue and the best model from ten lowest-energy cluster centers is in pink. In the scatterplots, we show the top 1% by energy of generated models in pink and the ten automatically selected cluster centers in dark red. Only base pair step sampling (used in FARFAR2) can routinely sample models closer to native than 10 Å RMSD. Only the previously developed method that samples from pregenerated ensembles for each helix can approach the quality of base-pair step sampling, with cluster centers found below 10 Å; despite requiring a separate modeling step for each target, even this method is still substantially worse than base-pair step sampling.

is likely to be. Recently, extensions of FARFAR to building coordinates into RNA-protein (Kappel et al., 2018) and RNA-only electron density maps (Kappel et al., 2019) have suggested a promising approach. In those settings, the average pairwise RMSD between the ten lowest-energy models was predictive of the average RMSD-to-native of those same ten models. Stated differently, modeling precision gives not just a lower bound on modeling accuracy but, through an empirical scaling, an estimate of the modeling accuracy. We tested whether a similar relationship would apply in this setting, where models are significantly more diverse due to the absence of electron density.

As shown in Figure 6, the average pairwise RMSD of the top ten FARFAR2 cluster centers does correlate well with the average RMSD to native of those models (R^2 of 0.84). The correlation is weaker than that observed in density guided modeling (gray points, Figure 6, R^2 of 0.94), and the trend is shifted higher so that the same inter-model RMSD corresponds to a worse average RMSD-to-native in the density-free FARFAR2 cases. Nevertheless, the error in the above estimate is itself predictable; the standard deviation of the pairwise RMSDs among top-ten models predicts most of the variance in the RMSD-to-native among those models (with electron density, R^2 = 0.90; without, R^2 = 0.64). These relationships suggested that we would be able to predict ranges of model accuracy in real, blind prediction scenarios, a prospect that we tested in our final study.

Blind Predictions of Six RNA Structures

After testing modeling accuracy of FARFAR2 in the retrospective benchmarks above, we sought to validate that the FARFAR2 method was similarly effective in truly blind modeling challenges. In separate work, we have developed a pipeline for highly accurate, rapid solution of complex RNA folds using a battery of cryo-EM, multidimensional chemical mapping, and automated computational modeling (Kappel et al., 2018, 2019). We saw this study as a valuable opportunity to conduct a series of blind challenges of the FARFAR2 method. We applied FARFAR2 to predict the structures of six natural RNAs – two tandem glycine riboswitches (from *Fusobacterium nucleatum* and *Vibrio cholerae*), two tRNA-sensing T-box riboswitches (for tRNA-Gly from *Geobacillus kaustophilus* and for tRNA-Gly from *Bacillus subtilis*; PDB: 6PMO and 6POM), a metagenomic *Mycobacterium S*-adenosyl methionine-binding SAM-IV riboswitch, and an adenoviral ncRNA virus-associated (VA) RNA I (PDB: 6OL3) (Hood et al., 2019). These modeling challenges were carried out fully blindly of experimental efforts by authors A.M.W. and R.R.; resulting models were preregistered with the Open Science Framework in the case of the two glycine riboswitches; submitted to an "Unknown RFam" RNA-Puzzles challenge in the case of SAM-IV; and submitted as "FARFAR2" predictions for RNA-Puzzles 24, 26, and 27 for VA RNA I and the T-box riboswitches, respectively.

The results of these six blind challenges supported the accuracy of FARFAR2 in a wide range of template-based modeling scenarios and a fully *de novo* modeling scenario (Table S7). The five natural riboswitch aptamers tested use of FARFAR2 in problems where templates were available, but peripheral tertiary domains needed to be built *de novo*. We used a 122-nt template derived from a crystal structure (PDB: 3P49) to build a series of models of the full-length 167-nt *F. nucleatum* glycine riboswitch aptamer (Figure 7A). This template structure previously included a U1a protein-binding loop to facilitate crystallization, which we replaced with the native sequence; it had also omitted a predicted 9-nt kink-turn linker between its two glycine-sensing domains, requiring the deletion and FARFAR2 remodeling of the first 18 nt on the 5′ end, nucleotides 72–91, and the three final 3′ nucleotides, 165–167. We subsequently used part of the lowest-energy *F. nucleatum* solution, threaded with the related sequence from *V. cholerae*, to predict the structure of the 229-nt *V. cholerae* tandem glycine riboswitch, using FARFAR2 to insert a P4 stem into a multiloop of the 5′-aptamer domain and substantially remodel the resulting fold (Figure 7B). Our strategy for the *Mycobacterium* SAM-IV riboswitch used a template 23-nt *S*-adenosylmethionine-binding site from an SAM-I riboswitch

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

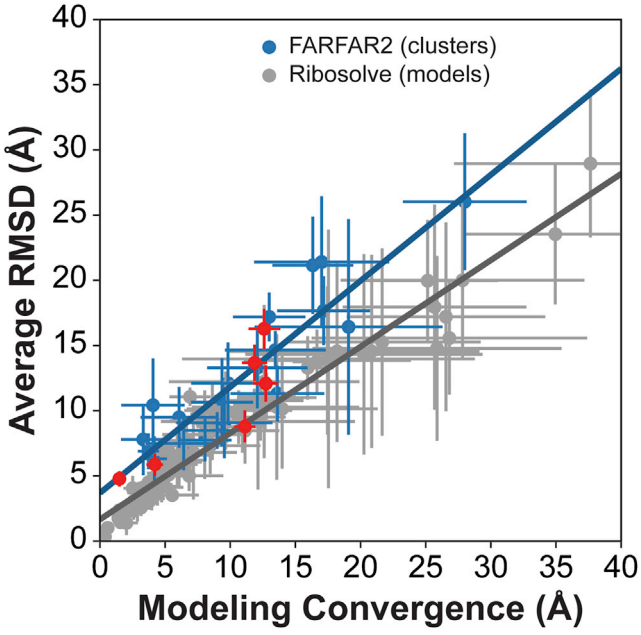


Figure 6. The Degree of Convergence Seen Among Top Models or Clusters Predicts the RMSD Accuracy of Those Models to the Experimental Structure

Convergence (the average pairwise RMSD among the top ten models or clusters) is predictive of modeling accuracy (the average RMSD-to-native of the ten lowest-energy models or clusters) whether with electron density (top ten DRRAFTER and Ribosolve models, gray) or without electron density (top ten FARFAR2 clusters, blue). Lines of best fit: Ribosolve y = 0.66x + 1.65 Å (R^2 = 0.94); FARFAR2 y = 0.81x + 3.69 Å (R^2 = 0.84). Error bars represent standard deviation of pairwise RMSD (x-error) and standard deviation of RMSD to native (y-error), which are themselves related as Ribosolve y-error = 0.76x-error + 0.10 Å (R^2 = 0.90); FARFAR2 y-error = 0.91x-error + 0.09 Å (R^2 = 0.64). Additionally, red points show the results of the six blind challenges attempted using FARFAR2 in this work.

(PDB: 2YGH) and largely followed our previously described procedure for homology modeling of complex RNA folds (Watkins et al., 2019), but used the FARFAR2 protocol instead, i.e., base-pair step sampling for helices, the updated fragment library, and the updated scoring function for full-atom refinement (Figure 7C). For the two T-box riboswitches, we used a template tRNA-Gly (PDB: 4LCK); for the larger *B. subtilis* T-box riboswitch, we could additionally take local template structures from the previously solved T box (PDB: 4LCK) (Figures 7D and 7E). Finally, for the VA RNA I, no template structures were available, so we used only the literature secondary structure (Dzananovic et al., 2017). Using the same model selection procedure as for FARFAR2-Puzzles, we submitted ten models for each problem except the SAM-IV riboswitch, for which we could submit only five models.

Upon unblinding, the FARFAR2 modeling recapitulated the global folds of the RNAs solved independently with experimental crystallography or cryo-EM and mutate-and-map data (Figure 7). The best submitted FARFAR2 models were a 3.0 Å prediction of the *F. nucleatum* glycine riboswitch (2.2 Å over residues modeled directly), a 4.3 Å prediction of the *V. cholerae* glycine riboswitch, a 3.2 Å prediction of the *Mycobacterium* SAM-IV riboswitch, a 14.3 Å prediction of the *G. kaustophilus* T-box riboswitch discriminator region/tRNA-Gly complex, a 10.0 Å prediction of the full-length *B. subtilis* T-box riboswitch/tRNA-Gly complex, and a 7.7 Å prediction of the VA RNA I (Table 2). Interestingly, two of the three cases with models better than 5 Å RMSD involved building RNA segments with total lengths greater than 80 nt (*V. cholerae* glycine and SAM-IV riboswitches), contradicting our 80-nt rule-of-thumb expectation described above. It is possible that stringent exclusion of homologous fragments during FARFAR2 benchmarking (STAR Methods), which we did not need to enforce during blind modeling, may have led to slightly better performance than expected from the benchmarking.

Encouragingly, for each of the specific blind modeling cases, our ability to estimate average modeling accuracy based on convergence observed during modeling was validated. The trend of Figure 6, through which we predict average model accuracy based on the model "convergence," held as well: the models achieved average RMSD accuracies of 4.8, 5.9, 8.8, 16.3, 12.1, and 13.7 Å compared with predicted accuracies of 4.9 ± 0.43, 7.1 ± 1.9, 12.7 ± 6.0, 13.9 ± 6.0, 14.0 ± 3.3, and 13.3 ± 3.4 Å, respectively. These results (plotted in red on Figure 6), validate our empirically derived relationship of convergence observed during modeling with accuracy in actual modeling scenarios. For example, in the two T-box riboswitches there were local inaccuracies with the placement of a peripheral helix (indicated by a red arrow in Figure 7); however, even in these cases, our predicted error estimate based on modeling convergence allowed us to forecast the greater inaccuracies.

DISCUSSION

We have developed and benchmarked a major update to Rosetta fragment assembly of RNA with full-atom refinement (FARFAR2), which automates multiple steps of previous protocols and brings to bear an updated fragment library, flexible treatment of helix base-pair steps, and a refined all-atom scoring function. We find that FARFAR2 achieves native-like structure models on three retrospective benchmarks, including 18 small RNAs, 82 motif-scale challenges, and 21 RNA-Puzzles, and on six prospective blind challenges.

Fragment assembly methods have become a widespread method for structure prediction because of their simple premise: folded biomolecules occupy a special region of conformational space, and new molecules should at least locally resemble previously determined structures. Fragment-based methods have yielded excellent predictions of RNA structure in blind modeling challenges, and FARFAR2's extensions to this system—enabling kinematically realistic helix sampling, expanding the database of crystallographic fragments, and scoring function improvements—have improved its performance, particularly in revisited RNA-Puzzle challenges. Without requiring human intervention following template selection, FARFAR2 samples native-like folds within its 1% lowest-energy structures in most of these challenges. The FARFAR2 structures are consistently more accurate than the best original Das lab submissions and, in the majority of cases, more accurate than the best overall submission. Model selection based on clustering the lowest-energy models is as or more successful in curating accurate structures than previous manual approaches. Nevertheless, our

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

CellPress

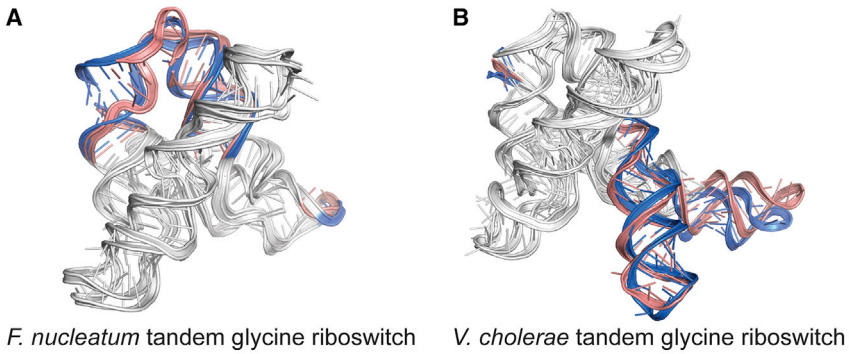


Figure 7. FARFAR2 Results in Six Blind Modeling Challenges

Blind predictions (salmon) of six complex RNA folds (blue) subsequently determined via cryoelectron microscopy or crystallography: the (A) *F. nucleatum* and (B) *V. cholerae* full-length tandem glycine riboswitches, as well as the (C) *Mycobacterium* SAM-IV riboswitch, the (D) *G. kaustophilus* T-box riboswitch/tRNA-Gly, the (E) *B. subtilis* T-box riboswitch/tRNA-Gly, and the (F) adenoviral VA RNA I. Predictions generally achieve nucleotide accuracy. White regions are input template structures. Red arrows in (D, E) mark a peripheral T-box helix with poor local accuracy.

benchmark highlights continuing limitations. Problems involving rebuilding more than 80 nt are not conformationally sampled well by FARFAR2. More generally, across all length ranges, comparison of energies of FARFAR2 models with optimized experimental structures suggests that our scoring function remains limiting. Improved clustering methods, explicit treatment of ligands, final-pass scoring functions, and better refinement procedures may resolve this limitation.

When applied to six blind challenges, FARFAR2 achieved blind predictions of 3 to 14 Å accuracy on newly built residues, showing good agreement with subsequently solved structures from cryo-EM and crystallography and confirming accuracy estimates made during modeling based on convergence of ten lowest-energy models. The convergence-based accuracy estimates introduced in this study should help guide the interpretation of the models and point to cases where more experimental data are required.

The frontiers for accurate RNA 3D structure prediction involve modeling larger problems (more than 80–100 nt of unknown tertiary structure, illustrated here by RNA-Puzzles 5, 12, and 7 [Figure 4]) and problems in which excellent RMSD accuracy, zero clashes, or perfect non-Watson-Crick base-pairing recovery are required (e.g., if the structure is required for subsequent ligand docking, molecular dynamics simulation, or drug design). Fully automated methods that more accurately solve subproblems at the expense of more computation, such as stepwise Monte Carlo (Watkins et al., 2018), may bridge these gaps if pipelined with fragment assembly.

We reiterate a need for better scoring functions to guide discrimination of near-native conformations from non-native conformations, which will be needed to prioritize fragment-assembled models for refinement. Methods that make more active use of residue contact inference from sequence covariation (Weinreb et al., 2016) or artificial neural networks (Kryshtafovych et al., 2019; Li et al., 2019; Wang et al., 2019) show promise for such improved scoring functions.

Beyond the FARFAR2 algorithm itself, this benchmark system and the associated 'decoy' models in the FARFAR2-Classics, FARFAR2-Motifs, and FARFAR2-Puzzles datasets can serve as starting points for the development of the next generation of high-resolution structure prediction algorithms and scoring functions, and we make these resources publicly available through links given in STAR Methods. Finally, to help ensure that the FARFAR2 code can be widely used and benchmarked, we have created a public webserver hosted on the Rosetta Server that Includes Everyone, ROSIE (Moretti et al., 2018). The server provides a simple interface that requires a sequence and secondary structure for inputs, with advanced options to give full access to the diverse configuration settings available for FARFAR2. Models are clustered exactly as in this

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

Table 2. Results for each blind challenge modeled in this work

Challenge	Length (Total)[c]	No. of Structures Generated	RNA	Top 1% Best RMSD (Å)[b]	Best Preregistered Model (Å)[a,b]	Average Model Accuracy (Å)[b]	Predicted Average Model Accuracy (Å)[b]
1	45 (167)	2,681	*F. nucleatum* glycine riboswitch	2.8	3.0	4.8	4.9 ± 0.43
2	145 (229)	3,949	*V. cholerae* glycine riboswitch	4.4	4.3	5.9	7.1 ± 1.9
3	96 (119)	10,833	SAM-IV riboswitch	3.8	3.2	8.8	12.7 ± 6.0
4	59 (130)	52,663	*G. kaustophilus* T-box riboswitch/ tRNA-Gly	9.9	14.3	16.3	13.9 ± 6.0
5	106 (230)	93,800	*B. subtilis* T-box riboswitch/ tRNA-Gly	7.3	10.0	12.1	14.0 ± 3.3
6	112 (112)	87,170	VA RNA I	4.8	7.7	13.7	13.3 ± 3.4

[a]Five models were submitted for the SAM-IV riboswitch blind challenge; ten for the others.
[b]All-heavy-atom RMSD is calculated over all residues, following superposition over all residues.
[c]Length is number of residues modeled *de novo* (not found in input templates); total also counts template residues.

study, and accuracy estimates are presented based on the relations calibrated in this study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- METHOD DETAILS
  - An Automated Fragment Assembly Benchmark
  - Updated Fragment Set
  - Fragment Homology Exclusion
  - Modes of Applying Secondary Structure Information
  - Benchmark Cases
  - FARFAR2 Execution on Benchmark Cases
  - Setting up Helices
  - Restrained Native Simulations
  - Sources of Experimental PDB Structures
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data Analysis
  - Model Selection
  - Measures of Prediction Accuracy
  - Translating RMSD Thresholds

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.str.2020.05.011.

ACKNOWLEDGMENTS

We thank K. Kappel, Z. Su, K. Zhang, and W. Chiu (Stanford) for independent structure determination of three blind challenges. We thank Z. Miao (EBI) and E. Westhof (U. Strasbourg) for organizing the RNA-Puzzles and 'Unknown RFam' challenges. We acknowledge funding from the National Institutes of Health (R21 CA219847 and R35 GM122579), National Science Foundation Graduate Research Fellowship Program under grant no. 1650114 (R.R.), and the Army Research Office W911NF-16-1-0372.

AUTHOR CONTRIBUTIONS

Conceptualization, A.M.W. and R.D.; Methodology, A.M.W. and R.D.; Investigation, A.M.W. and R.R.; Writing – Original Draft, A.M.W.; Writing – Review & Editing, A.M.W. and R.D.; Funding Acquisition, R.D.; Resources, R.D.; Supervision, R.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 10, 2019
Revised: April 27, 2020
Accepted: May 20, 2020
Published: June 9, 2020

REFERENCES

Akiyama, B.M., Laurence, H.M., Massey, A.R., Costantino, D.A., Xie, X., Yang, Y., Shi, P.Y., Nix, J.C., Beckham, J.D., and Kieft, J.S. (2016). Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. Science *354*, 1148–1152.

Amarasinghe, G.K., De Guzman, R.N., Turner, R.B., Chancellor, K.J., Wu, Z.R., and Summers, M.F. (2000). NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the Ψ-RNA packaging signal. Implications for genome recognition. J. Mol. Biol. *301*, 491–511.

Baeyens, K.J., De Bondt, H.L., Pardi, A., and Holbrook, S.R. (1996). A curved RNA helix incorporating an internal loop with G·A and A·A non-Watson-Crick base pairing. Proc. Natl. Acad. Sci. U S A *93*, 12851–12855.

Baird, N.J., Zhang, J., Hamma, T., and Ferré-D'Amaré, A.R. (2012). YbxF and YlxQ are bacterial homologs of L7Ae and bind K-turns but not K-loops. RNA *18*, 759–770.

Berglund, J.A., Rosbash, M., and Schultz, S.C. (2001). Crystal structure of a model branchpoint-U2 snRNA duplex containing bulged adenosines. RNA *7*, 682–691.

Boniecki, M.J., Lach, G., Dawson, W.K., Tomala, K., Lukasz, P., Soltysinski, T., Rother, K.M., and Bujnicki, J.M. (2015). SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res. *44*, e63.

Butler, E.B., Xiong, Y., Wang, J., and Strobel, S.A. (2011). Structural basis of cooperative ligand binding by the glycine riboswitch. Chem. Biol. *18*, 293–298.

Cabello-Villegas, J., Winkler, M.E., and Nikonowicz, E.P. (2002). Solution conformations of unmodified and A37N6-dimethylallyl modified anticodon stem-loops of Escherichia coli tRNAphe. J. Mol. Biol. *319*, 1015–1034.

Carugo, O., and Pongor, S. (2008). A normalized root-mean-square distance for comparing protein three-dimensional structures. Protein Sci. *10*, 1470–1473.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. (1996). Crystal structure of a group I ribozyme domain: principles of RNA packing. Science *273*, 1678–1685.

Cech, T.R., and Steitz, J.A. (2014). The noncoding RNA revolution - trashing old rules to forge new ones. Cell *157*, 77–94.

Cheng, C.Y., Chou, F.C., Kladwang, W., Tian, S., Cordero, P., and Das, R. (2015). Consistent global structures of complex RNA states through multidimensional chemical mapping. eLife *4*, e07600.

Correll, C.C., Beneken, J., Plantinga, M.J., Lubbers, M., and Chan, Y.L. (2003). The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. Nucleic Acids Res. *31*, 6806–6818.

Cruz, J.A., Blanchet, M.F., Boniecki, M., Bujnicki, J.M., Chen, S.J., Cao, S., Das, R., Ding, F., Dokholyan, N.V., Flores, S.C., et al. (2012). RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. RNA *18*, 610–625.

Dallas, A., and Moore, P.B. (1997). The loop E-loop D region of Escherichia coli 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. Structure *5*, 1639–1653.

Das, R., and Baker, D. (2007). Automated de novo prediction of native-like RNA tertiary structures. Proc. Natl. Acad. Sci. U S A *104*, 14664–14669.

Das, R., Karanicolas, J., and Baker, D. (2010). Atomic accuracy in predicting and designing noncanonical RNA structure. Nat. Methods *7*, 291–294.

Dibrov, S., McLean, J., and Hermann, T. (2011a). Structure of an RNA dimer of a regulatory element from human thymidylate synthase mRNA. Acta Crystallogr. D Biol. Crystallogr. *67*, 97–104.

Dibrov, S.M., McLean, J., Parsons, J., and Hermann, T. (2011b). Self-assembling RNA square. Proc. Natl. Acad. Sci. U S A *108*, 6405–6408.

Ding, F., Sharma, S., Chalasani, P., Demidov, V.V., Broude, N.E., and Dokholyan, N.V. (2008). Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. RNA *14*, 1164–1173.

Dzananovic, E., Astha, Chojnowski, G., Deo, S., Booy, E.P., Padilla-Meier, P., McEleney, K., Bujnicki, J.M., Patel, T.R., and McKenna, S.A. (2017). Impact of the structural integrity of the three-way junction of adenovirus VAI RNA on PKR inhibition. PLoS One *12*, e0186849.

Egli, M., Minasov, G., Su, L., and Rich, A. (2002). Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. Proc. Natl. Acad. Sci. U S A *99*, 4302–4307.

Finger, L.D., Trantirek, L., Johansson, C., and Feigon, J. (2003). Solution structures of stem-loop RNAs that bind to the two N-terminal RNA-binding domains of nucleolin. Nucleic Acids Res. *31*, 6461–6472.

Hofacker, I.L., and Lorenz, R. (2014). Predicting RNA structure: advances and limitations. Methods Mol. Biol. *1086*, 1–19.

Holbrook, S.R., Cheong, C., Tinoco, I., and Kim, S.H. (1991). Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. Nature *353*, 579–581.

Hood, I.V., Gordon, J.M., Bou-Nader, C., Henderson, F.E., Bahmanjah, S., and Zhang, J. (2019). Crystal structure of an adenovirus virus-associated RNA. Nat. Commun. *10*, 2871.

Huang, L., Serganov, A., and Patel, D.J. (2010). Structural insights into ligand recognition by a sensing domain of the cooperative Glycine riboswitch. Mol. Cell *40*, 774–786.

Huang, L., Wang, J., Wilson, T.J., and Lilley, D.M.J. (2017). Structure of the guanidine III riboswitch. Cell Chem. Biol. *24*, 1407–1415.e2.

Ippolito, J.A., and Steitz, T.A. (2000). The structure of the HIV-1 RRE high affinity Rev binding site at 1.6 Å resolution. J. Mol. Biol. *295*, 711–717.

Jian, W., Mao, K., Zhao, Y., Chen, Z., Jianjin, X., Yi, Z., and Xiao, Y. (2017). Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. Nucleic Acids Res. *45*, 6299–6309.

Kacer, V., Scaringe, S.A., Scarsdale, J.N., and Rife, J.P. (2003). Crystal structures of r(GGUCACAGCCC)2. Acta Crystallogr. D Biol. Crystallogr. *59*, 423–432.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. *46*, D335–D342.

Kappel, K., and Das, R. (2019). Sampling native-like structures of RNA-protein complexes through Rosetta folding and docking. Structure *27*, 140–151.

Kappel, K., Liu, S., Larsen, K.P., Skiniotis, G., Puglisi, E.V., Puglisi, J.D., Zhou, Z.H., Zhao, R., and Das, R. (2018). De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. Nat. Methods *15*, 947–954.

Kappel, K., Zhang, K., Su, Z., Kladwang, W., Li, S., Pintilie, G., Topkar, V.V., Rangan, R., Zheludev, I.N., Watkins, A.M., et al. (2019). Ribosolve: rapid determination of three-dimensional RNA-only structures. BioRxiv. https://doi.org/10.1101/717801.

Krokhotin, A., Houlihan, K., and Dokholyan, N.V. (2015). iFoldRNA v2: folding RNA with constraints. Bioinformatics *31*, 2891–2893.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2019). Critical assessment of methods of protein structure prediction (CASP)—round XIII. Proteins *87*, 1011–1020.

Laing, C., and Schlick, T. (2010). Computational approaches to 3D modeling of RNA. J. Phys. Condens. Matter *22*, 283101.

Leonard, G.A., McAuley-Hecht, K.E., Ebel, S., Lough, D.M., Brown, T., and Hunter, W.N. (1994). Crystal and molecular structure of r(CGCGAAUUAGCG): an RNA duplex containing two G(anti)· A(anti) base pairs. Structure *2*, 483–494.

Leontis, N.B., and Zirbel, C.L. (2012). Nonredundant 3D structure datasets for RNA knowledge extraction and benchmarking. In RNA 3D Structure Analysis and Prediction, E. Westhof and N.B. Leontis, eds. (Springer), pp. 281–298.

Li, Y., Hu, J., Zhang, C., Yu, D.-J., and Zhang, Y. (2019). ResPRE: high-accuracy protein contact prediction by coupling precision matrix with deep residual neural networks. Bioinformatics *35*, 4647–4655.

Liu, Y., Wilson, T.J., and Lilley, D.M.J. (2017). The structure of a nucleolytic ribozyme that employs a catalytic metal ion. Nat. Chem. Biol. *13*, 508–513.

Lyskov, S., Chou, F.C., Conchúir, S.Ó., Der, B.S., Drew, K., Kuroda, D., Xu, J., Weitzner, B.D., Renfrew, P.D., Sripakdeevong, P., et al. (2013). Serverification of molecular modeling applications: the Rosetta online server that includes everyone (ROSIE). PLoS One *8*, e63906.

Magnus, M., Antczak, M., Zok, T., Wiedemann, J., Lukasiak, P., Cao, Y., Bujnicki, J.M., Westhof, E., Szachniuk, M., and Miao, Z. (2019). RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. Nucleic Acids Res. *48*, 576–588.

Martinez-Zapien, D., Legrand, P., McEwen, A.G., Proux, F., Cragnolini, T., Pasquali, S., and Dock-Bregeon, A.C. (2017). The crystal structure of the 5'

functional domain of the transcription riboregulator 7SK. Nucleic Acids Res. *45*, 3568–3579.

Mathews, D.H. (2006). RNA secondary structure analysis using RNAstructure. Curr. Protoc. Bioinformatics *13*, 12.6.1–12.6.14.

Meyer, M., Nielsen, H., Oliéric, V., Roblin, P., Johansen, S.D., Westhof, E., and Masquida, B. (2014). Speciation of a group i intron into a lariat capping ribozyme. Proc. Natl. Acad. Sci. U S A *111*, 7659–7664.

Miao, Z., Adamiak, R.W., Blanchet, M.F., Boniecki, M., Bujnicki, J.M., Chen, S.J., Cheng, C., Chojnowski, G., Chou, F.C., Cordero, P., et al. (2015). RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures. RNA *21*, 1066–1084.

Miao, Z., Adamiak, R.W., Antczak, M., Batey, R.T., Becka, A.J., Biesiada, M., Boniecki, M.J., Bujnicki, J.M., Chen, S.J., Cheng, C.Y., et al. (2017). RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. RNA *23*, 655–672.

Miao, Z., Westhof, E., Adamiak, R.W., Antczak, M., Chen, S.J., Cheng, C.Y., Cheng, Y., Chou, F.C., Das, R., Dokholyan, N.V., et al. (2020). RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. RNA. https://doi.org/10.1261/rna.075341.120.

Mir, A., Chen, J., Robinson, K., Lendy, E., Goodman, J., Neau, D., and Golden, B.L. (2015). Two divalent metal ions and conformational changes play roles in the hammerhead ribozyme cleavage reaction. Biochemistry *54*, 6369–6381.

Moretti, R., Lyskov, S., Das, R., Meiler, J., and Gray, J.J. (2018). Web-accessible molecular modeling with Rosetta: the Rosetta online server that includes everyone (ROSIE). Protein Sci. *27*, 259–268.

Pallan, P.S., Marshall, W.S., Harp, J., Jewett, F.C., Wawrzak, Z., Brown, B.A., Rich, A., and Egli, M. (2005). Crystal structure of a luteoviral RNA pseudoknot and model for a minimal ribosomal frameshifting motif. Biochemistry *44*, 11315–11322.

Pan, B., Shi, K., and Sundaralingam, M. (2006). Crystal structure of an RNA quadruplex containing inosine tetrad: implications for the roles of NH2 group in purine tetrads. J. Mol. Biol. *363*, 451–459.

Parisien, M., and Major, F. (2008). The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature *452*, 51–55.

Peselis, A., and Serganov, A. (2012). Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. Nat. Struct. Mol. Biol. *19*, 1182–1184.

Piatkowski, P., Kasprzak, J.M., Kumar, D., Magnus, M., Chojnowski, G., and Bujnicki, J.M. (2016). RNA 3D structure modeling by combination of template-based method ModeRNA, template-free folding with SimRNA, and refinement with QRNAS. Methods Mol. Biol. *1490*, 217–235.

Popenda, M., Szachniuk, M., Antczak, M., Purzycka, K.J., Lukasiak, P., Bartol, N., Blazewicz, J., and Adamiak, R.W. (2012). Automated 3D structure composition for large RNAs. Nucleic Acids Res. *40*, e112.

Porter, E.B., Polaski, J.T., Morck, M.M., and Batey, R.T. (2017). Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. Nat. Chem. Biol. *13*, 295–301.

Ren, A., and Patel, D.J. (2014). C-di-AMP binds the ydaO riboswitch in two pseudo-symmetry-related pockets. Nat. Chem. Biol. *10*, 780–786.

Ren, A., Xue, Y., Peselis, A., Serganov, A., Al-Hashimi, H.M., and Patel, D.J. (2015). Structural and dynamic basis for low-affinity, high-selectivity binding of L-glutamine by the glutamine riboswitch. Cell Rep. *13*, 1800–1813.

Ren, A., Vušurovic, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., Micura, R., and Patel, D.J. (2016). Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. Nat. Chem. Biol. *12*, 702–708.

Richardson, J.S., Schneider, B., Murray, L.W., Kapral, G.J., Immormino, R.M., Headd, J.J., Richardson, D.C., Ham, D., Hershkovits, E., Williams, L.D., et al. (2008). RNA backbone: consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). RNA *14*, 465–481.

Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M., and Scott, W.G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. PLoS Biol. *3*, e5.

Safaee, N., Noronha, A.M., Rodionov, D., Kozlov, G., Wilds, C.J., Sheldrick, G.M., and Gehring, K. (2013). Structure of the parallel duplex of poly(A) RNA: evaluation of a 50 year-old prediction. Angew. Chem. Int. Ed. *52*, 10370–10373.

Sali, A. (1995). Comparative protein modeling by satisfaction of spatial restraints. Mol. Med. Today *1*, 270–277.

Schmitz, U., Behrens, S., Freymann, D.M., Keenan, R.J., Lukavsky, P., Walter, P., and James, T.L. (1999). Structure of the phylogenetically most conserved domain of SRP RNA. RNA *5*, 1419–1429.

Schroeder, K.T., Daldrop, P., and Lilley, D.M.J. (2011). RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. Structure *19*, 1233–1240.

Seetharaman, M., Eldho, N.V., Padgett, R.A., and Dayie, K.T. (2006). Structure of a self-splicing group II intron catalytic effector domain 5: parallels with spliceosomal U6 RNA. RNA *12*, 235–247.

Simons, K.T., Kooperberg, C., Huang, E., and Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. *268*, 209–225.

Sripakdeevong, P., Beauchamp, K., and Das, R. (2012). Why can't we predict RNA structure at atomic resolution? In RNA 3D Structure Analysis and Prediction. Nucleic Acids and Molecular Biology, N.B. Leontis and E. Westhof, eds. (Springer), pp. 43–65.

Suslov, N.B., Dasgupta, S., Huang, H., Fuller, J.R., Lilley, D.M.J., Rice, P.A., and Piccirilli, J.A. (2015). Crystal structure of the Varkud satellite ribozyme. Nat. Chem. Biol. *11*, 840–846.

Szép, S., Wang, J., and Moore, P.B. (2003). The crystal structure of a 26-nucleotide RNA containing a hook-turn. RNA *9*, 44–51.

Tan, D., Piana, S., Dirks, R.M., and Shaw, D.E. (2018). RNA force field with accuracy comparable to state-of-the-art protein force fields. Proc. Natl. Acad. Sci. U S A *115*, E1346–E1355.

Trausch, J.J., Xu, Z., Edwards, A.L., Reyes, F.E., Ross, P.E., Knight, R., and Batey, R.T. (2014). Structural basis for diversity in the SAM clan of riboswitches. Proc. Natl. Acad. Sci. U S A *111*, 6624–6629.

Trausch, J.J., Marcano-Velázquez, J.G., Matyjasik, M.M., and Batey, R.T. (2015). Metal ion-mediated nucleobase recognition by the ZTP riboswitch. Chem. Biol. *22*, 829–837.

Wang, L., Liu, Y., Zhong, X., Liu, H., Lu, C., Li, C., and Zhang, H. (2019). DMFold: a novel method to predict RNA secondary structure with pseudoknots based on deep learning and improved base pair maximization principle. Front. Genet. *10*, 143.

Warner, K.D., Chen, M.C., Song, W., Strack, R.L., Thorn, A., Jaffrey, S.R., and Ferré-D'Amaré, A.R. (2014). Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. Nat. Struct. Mol. Biol. *21*, 658–663.

Watkins, A.M., Geniesse, C., Kladwang, W., Zakrevsky, P., Jaeger, L., and Das, R. (2018). Blind prediction of noncanonical RNA structure at atomic accuracy. Sci. Adv. *4*, eaar5316.

Watkins, A.M., Rangan, R., and Das, R. (2019). Chapter 9. Using Rosetta for RNA homology modeling. In Methods in Enzymology, RNA Recognition, *Vol. 623*, A.E. Hargrove, ed. *(Academic Press), pp. 177–207.*

Weinreb, C., Riesselman, A.J., Ingraham, J.B., Gross, T., Sander, C., and Marks, D.S. (2016). 3D RNA and functional interactions from evolutionary couplings. Cell *165*, 963–975.

Xiong, Y., and Sundaralingam, M. (2000). Two crystal forms of helix II of Xenopus laevis 5S rRNA with a cytosine bulge. RNA *6*, 1316–1324.

Zhang, J., and Ferré-D'Amaré, A.R. (2013). Co-crystal structure of a T-box riboswitch stem i domain in complex with its cognate tRNA. Nature *500*, 363–366.

Zhao, C., Xu, X., and Chen, S.J. (2017). Predicting RNA structure with vfold. Methods Mol. Biol. *1654*, 3–15.

Zheng, L., Mairhofer, E., Teplova, M., Zhang, Y., Ma, J., Patel, D.J., Micura, R., and Ren, A. (2017). Structure-based insights into self-cleavage by a four-way junctional twister-sister ribozyme. Nat. Commun. *8*, 1180.

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Previous RNA-Puzzles Submissions	RNA-Puzzles	http://www.rnapuzzles.org/results/https://github.com/RNA-Puzzles/completed_dataset
Software and Algorithms		
Rosetta	RosettaCommons	https://www.rosettacommons.org/software/academic
PyMOL	Schrödinger	https://www.schrodinger.com/pymol
rna_benchmark package	This paper	https://github.com/DasLab/rna_benchmark

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rhiju Das (rhiju@stanford.edu).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Applications for running FARFAR2 and scripts for working with PDB files for use with this algorithm have been contributed as part of the Rosetta software suite, which is free for academic use (https://rosettacommons.org). FARFAR2 is available as a webserver on ROSIE at https://rosie.rosettacommons.org/farfar2. The scripts used to set up the benchmarks studied in this work are available at https://github.com/DasLab/rna_benchmark. The final raw dataset is deposited with Stanford Library's PURL system at https://purl.stanford.edu/wn364wz7925, and a selection of models – the top 10 clusters for each RNA-Puzzle, along with the native structure and a single restrained native – are available at https://github.com/DasLab/FARFAR2-Puzzles-Clusters.

METHOD DETAILS

An Automated Fragment Assembly Benchmark

Two technical improvements within Rosetta permitted rapid progress on a fully automated fragment assembly protocol. First, FARFAR2 jobs may be fully specified and run using a single command line, rather than requiring pre-configuration using a 'params file' with a complex language (Das and Baker, 2007); example command-lines are given below ('FARFAR2 Execution on Benchmark Cases'). Second, we developed a benchmarking framework (Watkins et al., 2018), available at https://github.com/daslab/rna_benchmark with documentation and instructions.

Updated Fragment Set

We obtained release 3.10 of representative nonredundant 3D structures from BGSU's RNA 3D Hub (Leontis and Zirbel, 2012). We obtained the indicated PDB chains and parsed them into fragment and jump (base-base rigid body transformations for base pairs) database files using the following command lines:

```
rna_database -s *_RNA.pdb -vall_torsions
RICHARDSON_RNA18_2.5_revised.torsions
-cut_at_rna_chainbreak true
-ignore_zero_occupancy false -guarantee_no_DNA true

rna_database -s *_RNA.pdb -jump_database true -cut_at_rna_chainbreak true
-ignore_zero_occupancy false -guarantee_no_DNA true
-out:file:o RICHARDSON_RNA18_2.5_revised.jump
```

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

Fragment Homology Exclusion

Benchmarks of fragment assembly approaches can give misleading overestimates of accuracy if fragments of the experimental ('native') structure are included during modeling. In order to better mimic blind prediction scenarios in which the experimental structure is not available at the time of modeling, we ensured that our fragment libraries were free of contamination from the native structures employed in the benchmarks through a "fragment homology exclusion" option implemented in FARFAR and controlled by command-line options; see also (Watkins et al., 2018). In this mode, all six-nucleotide contiguous stretches of RNA to be built are extracted from the experimental structure, and any fragments in the library that are deemed too similar to the experimental structure are excluded as possible contamination from that experimental structure or a close homolog. More specifically, these ranges of structure are compared by heavy-atom RMSD to every fragment in the fragment library with matching purine/pyrimidine content, and fragments with RMSD less than 1.2 Å from the experimental structure are eliminated from consideration.

Modes of Applying Secondary Structure Information

In most modeling cases we have encountered, models of secondary structure were previously available based on expert analysis, thermodynamic modeling packages (Hofacker and Lorenz, 2014; Mathews, 2006), and/or sequence alignment information. Such known secondary structure (see Table S4) can now be specified either via command-line or in an input text file. The FARFAR2 protocol can apply this secondary structure information in six ways. First, to replicate the original FARNA style of simulation, one may specify no secondary structure at all. Second, a specified secondary structure can provide energetic "base pair constraints" (see FARFAR2 Execution on Benchmark Cases below) that tend to draw paired residues together, introduced in (Das et al., 2010). Base pair constraints consist of harmonic restraints placed on the distance between corresponding Watson-Crick edge atoms (1.9 Å with a standard deviation of 0.1 Å) as well as a harmonic restraint placed on the distance between sugar C1 atoms (10.5 Å with a standard deviation of 1.0 Å) to favor paired, rather than stacked-and-tilted, conformations. The influence of these constraints are ramped up over each round of low-resolution modeling, then applied at full strength during minimization. Third, provision of the -fixed_stems flag makes base pairing geometry rigid in Rosetta's kinematic representation. Fourth and fifth, helical stems may be generated and then provided as fixed inputs to the simulation, either as individual ideal helices or as ensembles mimicking thermal fluctuations, respectively, as described in (Cheng et al., 2015; Miao et al., 2015). This procedure is automated by the `rna_benchmark` system but may be carried out manually using scripts from the Rosetta tools directory (see Setting up Helices below). Sixth, helix flexibility may be simulated directly in a kinematically realistic way by sampling from a library of "base pair steps". Base pair step moves keep the base paired secondary structure fixed while sampling orientation changes between consecutive base pairs seen in the crystallographic database; Rosetta's implementation is generalized to permit realistic sampling of stems with interrupting nucleotide bulges on one side, as well. This mode had not been previously described or tested but is now the default mode for Rosetta's `rna_denovo` application, due to its superior performance in the benchmarking described in the main text and Supplemental Information.

Benchmark Cases

The three benchmarks evaluated were intended to evaluate the performance of the FARFAR2 'best practices' on challenges of qualitatively different scales. The original FARNA benchmark and the stepwise Monte Carlo benchmark examined relatively small structures, though many representative examples in the latter case were fairly complex. In contrast, the RNA-Puzzles benchmark, like the six blind challenges undertaken to test this method, examined entire folded RNAs, typically with tertiary contacts between elements distal in secondary and tertiary structure.

The FARFAR2-Classics benchmark comprised 18 challenges, each of which was either a single-stranded stem-loop or a duplex. Two structures that overlapped exactly with the FARFAR2-Motifs benchmark (see below) were omitted from the set of 20 structures (1J6S, a G-quadruplex; 1ZIH, a GCAA tetraloop) used to benchmark the original FARNA algorithm (Das and Baker, 2007). In each case, the challenge structure was the entire RNA solved experimentally. These cases were approached in four distinct modes of secondary structure specification: none, "base pair constraint" generation, fixed helical "chunks," and "base pair step" sampling. Either the prior default fragment library or one generated for this work were used to sample loop nucleotides and, if applicable, base pair steps. Finally, results for each challenge were either left un-minimized or optimized in one of two scoring functions: the original FARFAR refinement scoring function (hereafter "hires") or the modern RNA scoring function developed for stepwise Monte Carlo (originally termed *rna_res_level_energy4.wts*, hereafter called "res4") (Watkins et al., 2018). (The same low-resolution structures were minimized to produce the final data set.) As in the original FARNA paper, 50,000 structures were generated for each simulation. Additionally, as a control, we replicated the exact parameters of the original FARNA work, including an eleven-year-old fragment library based only on the *H. marismortui* 23S ribosome (PDB: 1JJ2). This replication performed better on the majority of cases, suggesting some level of Rosetta simulation improvement and bug fixes since 2007 (see Figure S1, Table S1). For this comparison, we had to reduce the fragment homology exclusion radius from 1.2 Å to 0.5 Å or else we were unable to discover fragments for cases 1CSL, 1DQF, 1ESY, 1I9X, 1KD5, 1Q9A, 1QWA, 1XJR, 2A43, or 2F88. Informed by a comparison of these simulation conditions, we decided to take the SWM "res4" scoring function and the updated fragment libraries as the FARFAR2 standard. We also chose to specify secondary structure through base pair steps. Though the advantage over fixed helical stems was not obvious across the FARNA benchmark, we anticipated that more complex tertiary environments, such as those found in typical whole structure prediction cases, would benefit further from some form of helix flexibility, and this was later confirmed (see main text on FARFAR2-Puzzles; and Figures S6 and S7). These optimal FARFAR2 parameters also proved robust to variations in other simulation settings. We repeated the whole benchmark set with the final FARFAR2 settings but disabled all low-resolution filters (which restart simulations that have generated

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

structures with chainbreaks, missing base pairs, or bad scores). We also tested reducing the ten-round fragment assembly schedule to only one or two rounds, and we obtained similar results (Figure S1D). None of these configuration variants ought to be taken as best practices on their own, however, particularly for larger RNA folds. Setting up only one simulation round, for example, increases the computational expense more than fourfold on average, and while score, base pair, and chainbreak filters may not be essential for simple folds that are quick to energy-minimize, we reasoned that they would prevent needless computation on large RNAs.

The FARFAR2-Motifs benchmark set was chosen because it was a direct expansion of the earlier FARFAR benchmark (Das et al., 2010), augmented with challenges that demonstrated specific features of FARFAR2 (such as 'aligned' cases, which permit expert specification of the relative orientation of a subset of the native structure; and support for chemically modified nucleotides). This expansion originated when developing the stepwise Monte Carlo (SWM) algorithm and included single-stranded loops, two-way and multi-way junctions, tertiary contacts, and motifs that exist outside of a Watson-Crick context (such as quadruplexes (Pan et al., 2006) and parallel strands (Safaee et al., 2013)). We found that SWM frequently produced excellent native RMSDs. We generated 10,000 structures for each benchmark case: fewer than for FARFAR2-Classics because most of the above structures were smaller and structurally simpler, and because there are more than four times as many benchmark cases. Though SWM delivers performance unattainable by fragment methods, strong performance on this benchmark set would suggest the ability of fragment assembly methods to provide good starting points for higher-resolution simulation at a fraction of the computational cost required.

The FARFAR2-Puzzles benchmark set includes all RNA-Puzzles for which solutions and submissions were available at the time of this study: that is, each of the first 21 except for RNA-Puzzle 16, and including both bound and unbound states of RNA-Puzzle 14. In some cases, a large proportion of the problem was known approximately by homology, while in others only the helices were available as starting information. Each problem was run to generate several thousand structures, aiming for substantial sampling comparable to efforts deployed for actual RNA-Puzzles challenges.

FARFAR2 Execution on Benchmark Cases

We tested multiple combinations of settings for the FARFAR2 protocol. We present here the relevant command lines for each possible configuration, as operated on the 157D challenge (an RNA duplex containing two G-A base pairs) from FARFAR2-Classics). (The command line argument to run with a particular set of conditions, such as secondary structure specification vs. number of rounds, are mutually independent of each other and can be recombined.)

Updated Recommended Default Settings

Before presenting details of benchmark conditions, here is an example of how one might specify a modeling problem using the FARFAR2 defaults. Suppose one is concerned with the aforementioned RNA duplex. First, prepare a FASTA-formatted file, `157d_orig.fasta`:

```
>157D_orig A:1-12 B:13-24
cgcgaauuagcgcgcgaauuagcg
```

Then, prepare a file containing the dot -bracket notation secondary structure, `157d.secstruct`. Since only the first line of the secondary structure file is read, one may include the sequence as the second line as a useful point of reference.

```
(((.(((((.(((,))).))))).)))
cgcgaauuagcg,cgcgaauuagcg
```

Finally, execute the following command:

```
rna_denovo -fasta 157d_orig.fasta -secstruct_file 157d.secstruct-minimize_rna true
```

One may provide a parameter `-nstruct` to control how many structures are generated by each execution of the above command; the parameter `-out:file:silent` names the output file where results are accumulated.

FARNA Replication

An exact replication of the original FARNA benchmark required a special flag to use the original fragment library containing only a single ribosome crystal structure (PDB 1JJ2). The specification of a single base pair (as found in `157d_orig_START1_157d.pdb`) is the minimum necessary information to seed simulation of a bimolecular RNA.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-s 157d_orig_START1_157d.pdb -minimize_rna false -cycles 20000 -nstruct 20
-use_1jj2_torsions true
```

No secondary structure specification, *no minimization*, *old fragments*

This run replicates the FARNA benchmark conditions but uses the fragment set that was in regular use by 2010.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-s 157d_orig_START1_157d.pdb -minimize_rna false -cycles 20000 -nstruct 20
-fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR
-jump_library_file 1jj2_RNA_jump_library.dat
-vall_torsions RICHARDSON_RNA09.torsions -bps_moves false
```

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

No secondary structure specification, *no minimization*, *updated fragments*
This run uses an updated set of fragments generated for this work.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-s 157d_orig_START1_157d.pdb -minimize_rna false -cycles 20000 -nstruct 20
-fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR -bps_moves false
```

No secondary structure specification, *minimization under 'rna_hires.wts'*, *updated fragments*
This run uses full-atom minimization in the 'hires' scoring function.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-s 157d_orig_START1_157d.pdb -minimize_rna true -cycles 20000 -nstruct 20
-fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR
-score:weights rna/denovo/rna_hires.wts -bps_moves false
```

No secondary structure specification, *minimization under 'rna_res_level_energy4.wts'*, *updated fragments*
This run uses full-atom minimization in the preferred 'res4' scoring function.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-s 157d_orig_START1_157d.pdb -minimize_rna true -cycles 20000 -nstruct 20
-fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR -bps_moves false
```

Base pair constraints, *minimization under 'rna_res_level_energy4.wts'*, *updated fragments*
This run specifies a secondary structure to generate base pair constraints; it does not require a starting PDB.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-secstruct " (((.(((((.(((,))).))))).)))" -minimize_rna true -cycles 20000
-nstruct 20 -fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR -bps_moves false
```

Fixed stem kinematics, minimization under 'rna_res_level_energy4.wts', updated fragments

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-secstruct "(((.(((((.(((,))).))))).)))" -minimize_rna true -cycles 20000
-nstruct 20 -fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR -bps_moves false -fixed_stems
```

Helix ensembles, minimization under 'rna_res_level_energy4.wts', updated fragments

```
rna_denovo -in:file:silent helix1.out helix2.out helix3.out -fasta 157d_orig.fasta -native
157d_orig_NATIVE_157d.pdb
-secstruct "(((.(((((.(((,))).))))).)))" -minimize_rna true -cycles 20000
-nstruct 20 -fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR -bps_moves false
```

Base pair steps, *minimization under 'rna_res_level_energy4.wts'*, *updated fragments*
This run specifies a secondary structure to sample base pair steps; it does not require a starting PDB.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-secstruct " (((.(((((.(((,))).))))).)))" -minimize_rna true -cycles 20000
-nstruct 20 -fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR
```

Fixed helical inputs, *minimization under 'rna_res_level_energy4.wts'*, *updated fragments*
This run uses multiple fixed input stems.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-s 157d_chunks_HELIX1_157d.pdb 157d_chunks_HELIX2_157d.pdb 157d_chunks_HELIX3_157d.pdb
-minimize_rna true -cycles 20000 -nstruct 20
-fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR -bps_moves false
```

Base pair steps, *minimization under 'rna_res_level_energy4.wts'*, *updated fragments*, *no filters*
This run turns off filters (for score and chain closure) used during the low-resolution phase to improve the quality of output structures at the cost of additional run time.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-secstruct " (((.(((((.(((,))).))))).)))" -minimize_rna true -cycles 20000
-nstruct 20 -fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR
-no_filters
```

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

Base pair steps, *minimization under 'rna_res_level_energy4.wts'*, *updated fragments*, *one round of fragment assembly*

Typical runs of fragment assembly go through ten rounds: three rounds of 3-mer fragment insertion, three of 2-mers, and four of 1-mers, while ramping the weights of particular score terms. Setting the number of total rounds to one or two makes the protocol much more abrupt in its transitions but does not affect other simulation settings.

```
rna_denovo -fasta 157d_orig.fasta -native 157d_orig_NATIVE_157d.pdb
-secstruct " (((.(((((.(((,))).))))).)))" -minimize_rna true -cycles 20000
-nstruct 20 -fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR
-rounds 1
```

Setting up Helices

To use fixed helical inputs for a FARFAR structure prediction challenge without using the rna_benchmark setup scripts, a Python script is available within Rosetta tools:

```
rna_helix.py -seq ggg ccc
```

will create a three base pair helix of sequence GGGCCC, numbered 1-6 with no chain letter. To impose some reasonable numbering:

```
renumber_pdb_in_place.py helix.pdb A:1-3 B:1-3.
```

Restrained Native Simulations

To generate the near-native ensembles of models that were analyzed to determine energy gaps for FARFAR2-Puzzles, we wished to replicate the exact simulation conditions from the benchmark. In particular, it was important to employ the same input template structures, because the accuracy of those templates provides a lower limit to the accuracy of the simulation as a whole. To this end, we made two modifications to the benchmark runs. First, we added harmonic restraints between every heavy-atom in the model and its corresponding position in the native crystal structure, with a standard deviation of 1.0 Å and no penalty for deviations under 4.0 Å. This was achieved via adding to the command line:

```
-align_pdb native.pdb -rmsd_screen 4.0
```

Second, we removed the flags that were specifically removing all fragments homologous to the native structure from the database to ensure discovery of conformations near the experimental ('native') structure, i.e., the removal of the flags `-fragment_homology_rmsd 1.2 -exclusion_match_type MATCH_YR`.

Finally, due to the extensive coordinate restraints, we discovered that Rosetta's L-BFGS minimizer would occasionally halt due to mismatched numeric/analytic derivatives (common when derivatives are very high due to limited floating point precision). To prevent this issue from resulting in anomalously high energies, we minimized the resulting structures once more using the `rna_minimize` application and the same flags.

Sources of Experimental PDB Structures

The structure sources for the motif benchmark have been described previously (Watkins et al., 2018). The structures for the FARNA benchmark are as follows: PDB codes 157D (Leonard et al., 1994); 1A4D (Dallas and Moore, 1997); 1CSL (Ippolito and Steitz, 2000); 1DQF (Xiong and Sundaralingam, 2000); 1ESY (Amarasinghe et al., 2000); 1I9X (Berglund et al., 2001); 1KD5 (Kacer et al., 2003); 1KKA (Cabello-Villegas et al., 2002); 1L2X (Egli et al., 2002); 1MHK (Szép et al., 2003); 1Q9A (Correll et al., 2003); 1QWA (Finger et al., 2003), 1XJR (Robertson et al., 2005); 255D (Holbrook et al., 1991); 283D (Baeyens et al., 1996); 28SP (Schmitz et al., 1999); 2A43 (Pallan et al., 2005), 2F88 (Seetharaman et al., 2006).

The structures for the RNA-Puzzles benchmark are:

PDB codes 3MEI; 3MEI (Dibrov et al., 2011a); 3P59 (Dibrov et al., 2011b); 3OXE (Huang et al., 2010); 3V7E (Baird et al., 2012); 4P9R (Meyer et al., 2014); 4GXY (Peselis and Serganov, 2012); 4R4V (Suslov et al., 2015); 4L81 (Trausch et al., 2014); 5KPY (Porter et al., 2017); 4LCK (Zhang and Ferré-D'Amaré, 2013); 5LYU (Martinez-Zapien et al., 2017); 4QLM (Ren and Patel, 2014); 4XW7 (Trausch et al., 2015); 5DDP (Ren et al., 2015); 5DDO (Ren et al., 2015); 5DI4 (Mir et al., 2015); 5K7C (Ren et al., 2016); 5TPY (Akiyama et al., 2016); 5T5A (Liu et al., 2017); 5Y87 (Zheng et al., 2017), 5NWQ (Huang et al., 2017).

Structures used as templates in the blind predictions: 3P49 (Butler et al., 2011), 2YGH (Schroeder et al., 2011), 4TS2 (Warner et al., 2014), 1GID (Cate et al., 1996), 4LCK (Zhang and Ferré-D'Amaré, 2013).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data Analysis

FARFAR2 simulations produce a compressed Rosetta-format file called a 'silent file' representing each trajectory endpoint. These files may be turned into PDB-format coordinate files using a Rosetta executable, `extract_pdbs`. They also hold scoring information; each line beginning with `SCORE:` either describes what scoring terms were used in generating the silent file or the values for that particular structure. Using these 'score-lines', the programs `grep` and `awk`, and the GNU coreutils `sort` and `wc`, we sorted

Please cite this article in press as: Watkins et al., FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds, Structure (2020), https://doi.org/10.1016/j.str.2020.05.011

the silent file by total score and by RMSD, and thereby selected the best-RMSD structure, the best-score structure, and the best-RMSD structure from among the top 1% by RMSD. As an example, the following command sorts the 'score-lines' by total energy, takes the top 500 lines, re-sorts those lines by RMSD, and prints out the 'tag' of the best-RMSD structure from those top 500 by score.

```
grep "^SCORE:" farna_rebuild.out | grep -v description | sort -nk2 | head -n 500 | sort -nk24 | head -n 1 | awk '{print $NF}'
```

The following Rosetta command will extract that specified model `$TAG` as a PDB file called `$TAG.pdb`.

```
extract_pdbs -in:file:silent farna_rebuild.out -tags $TAG
```

Model Selection

In order to produce a selection of models for comparison to RNA-Puzzles submissions, we attempt a variety of clustering strategies. The following command line will cluster the lowest energy $NN models with a $RR Å radius:

```
rna_cluster -in:file:silent farna_rebuild.out -out:file:silent farna_rebuild.clustered.out
-native farna_rebuild.clustered.out -native 157d_orig_NATIVE_157d.pdb -nstruct $NN-cluster_radius $RR
```

We chose three values for `$NN`: 400 (the program default), 1% of the decoys, or 2% of the decoys. We also surveyed possible values for `$RR`: for constant values, we chose 2.0, 3.0, 5.0, and 7.0; we also employed an 'adaptive' clustering radius that responds to how tightly converged the top models are for a trajectory – for this radius, we chose 0.5, 1.0, or 1.5 times the average pairwise RMSD among the top ten models for the run in question. A 5.0 Å radius and 400 models were the original choice and also ended up being the best-performing choice by a slight margin.

For comparisons with the original FARNA algorithm analysis of the FARFAR2-Classics benchmark, the above clustering command was used with `$NN` = 5000 and `$RR` = 3.0 (Das and Baker, 2007). For comparisons with the original SWM execution of the FARFAR2-Motifs benchmark, the above clustering command was used with `$NN` = 400 and `$RR` = 2.0 (Watkins et al., 2018).

Since biologically relevant RNA molecules may adopt multiple relevant conformations, FARFAR2 users specifically interested in producing a conformational ensemble for further analysis may wish to use a cluster radius small enough that expected suboptimal conformational variations would not be "hidden" by lower energy structures falling within the clustering radius.

Measures of Prediction Accuracy

We measure prediction accuracy in the benchmarks using heavy-atom root-mean-squared deviation (RMSD). RMSD is obtained by aligning the model structure to the native structure and taking the average distance between all heavy-atoms (i.e., non-hydrogens) in the structure.

Translating RMSD Thresholds

The $RMSD_{100}$ measure allows the comparison of RMSDs on different structure sizes by scaling to 100 residues (Carugo and Pongor, 2008):

$$RMSD_{100}(n) = \frac{RMSD_n}{-1.3 + 0.5 \ln n}$$

An extension of this measure (Kappel and Das, 2019) permits scaling between any numbers of residues:

$$RMSD_m(n) = RMSD_n \frac{-1.3 + 0.5 \ln m}{-1.3 + 0.5 \ln n}$$

We applied this to convert RMSD 4.0 Å (the threshold treated as 'native-like' in the original FARNA publication and used for the *FARFAR2-Classics* benchmark in this work) to a threshold for *FARFAR2-Puzzles*. The *FARFAR2-Classics* benchmark had a median length of 26 residues modeled *de novo*; the median *FARFAR2-Puzzles* case featured 71. Because of the presence of a few long RNA-Puzzle problems, we also separately considered 18 "short problems" with median 68 residues modeled *de novo*, and 3 "long problems" with median 130. Applying the above formula to convert (4.0 Å, 26) yields (9.2 Å, 71) or else (9.1 Å, 68); (13.8 Å, 130). We use the tighter 9.1 Å threshold for either "overall" or short problems alone; and 13.8 Å for the largest three problems. These thresholds also corresponded well to visual assessment of whether models were 'native-like' for each of the size ranges (see, e.g., Figures 3, 5, and 7).

Structure, Volume *28*

Supplemental Information

FARFAR2: Improved *De Novo* Rosetta Prediction of Complex Global RNA Folds

Andrew Martin Watkins, Ramya Rangan, and Rhiju Das

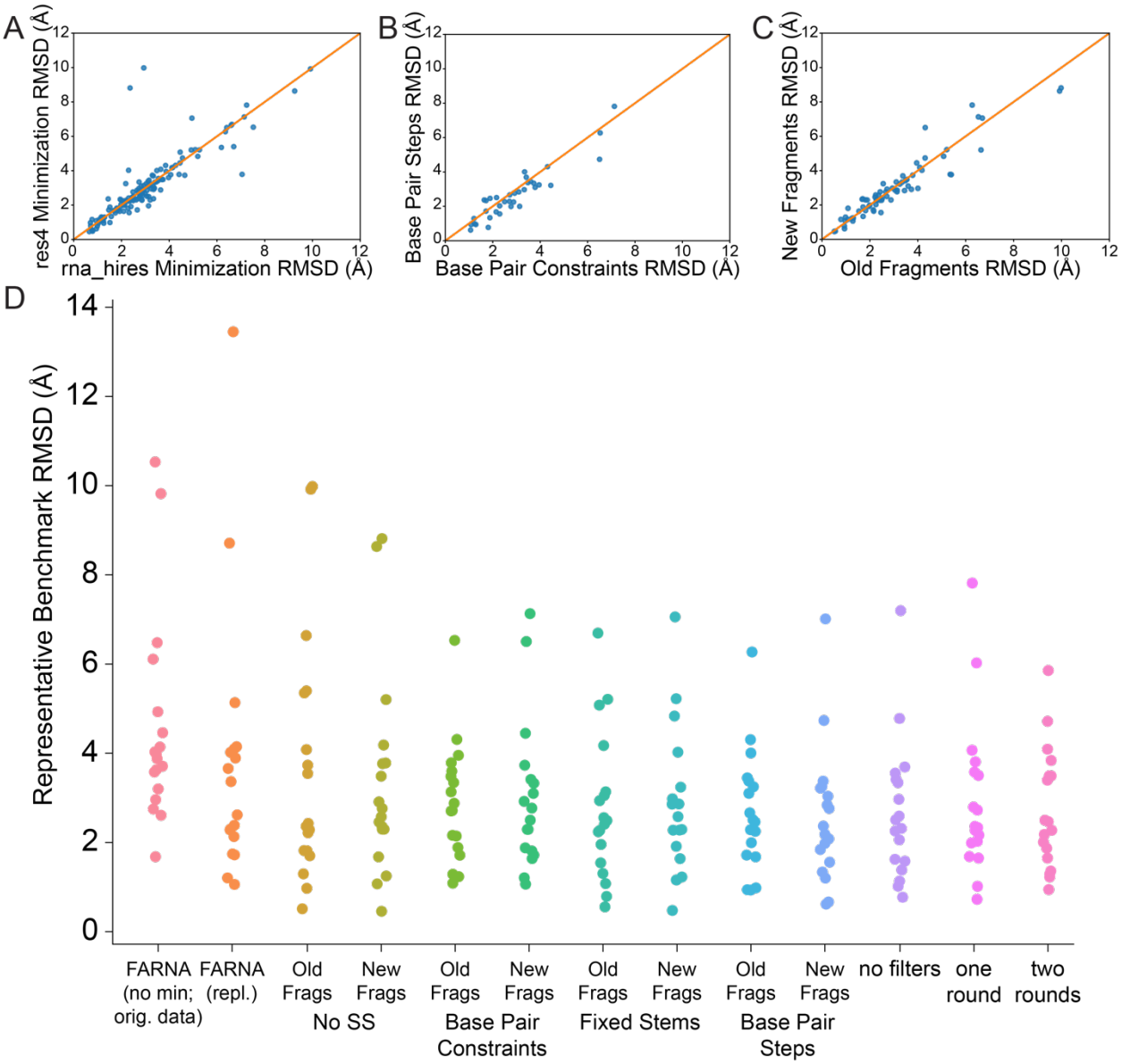


Figure S1. Related to Figure 1. FARFAR2 performance on the Classics benchmark. We conducted many pairwise comparisons on the top-1%-RMSD values of the Classics benchmark, where we altered only one variable between the setting chosen to define FARFAR2 and the complementary FARFAR/FARNA condition. (A) Optimization in the energy function originally developed for SWM (y-axis) was more effective than the original FARFAR energy function (x-axis), giving better RMSD in 91 of 144 possible comparisons. It was also considerably more effective than no minimization at all (not shown), giving better RMSD in 102 cases, often by considerable margins. (B) Providing secondary structure information, particularly as fixed helices or base pair steps (y-axis), offered substantial advantages over energetic

restraints (x-axis) or no secondary structure information. Fixed stems produced a superior RMSD in 68 of 108 possible such comparisons against other secondary structure specification methods, while base pair steps produced a superior RMSD in 76 of 108 comparisons. (C) Some benefit was also seen by using a newly obtained fragment library (y-axis) over the original FARFAR fragment library (x-axis), giving better results in 43 of 72 comparisons and providing the greatest advantage in the hardest problems. (D) A full comparison of simulation results using "res4" minimization, versus FARNA simulation and controls using the FARFAR2 simulation parameters but no filters or only one or two rounds of fragment assembly.

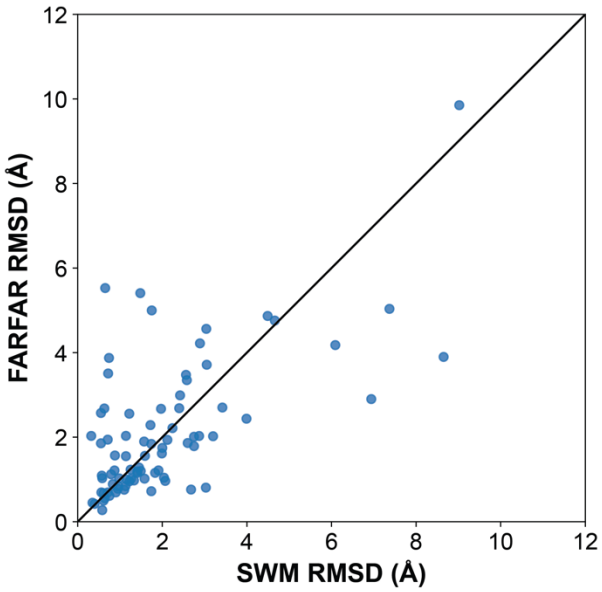


Figure S2. Related to Figure 1. FARFAR2 yields a superior RMSD to SWM among its five low-energy cluster centers in 44 of 82 benchmark cases. FARFAR2's particular advantage lies in cases where *both* SWM and FARFAR2 fail to obtain 1.5 Å RMSD accuracy: of the 6 cases where SWM provides worse than 5.0 Å RMSD, FARFAR2 provides a superior RMSD in 5. In contrast, among the 42 cases where SWM achieves better than a 1.5 Å RMSD, FARFAR2 obtains superior RMSD in only 19.

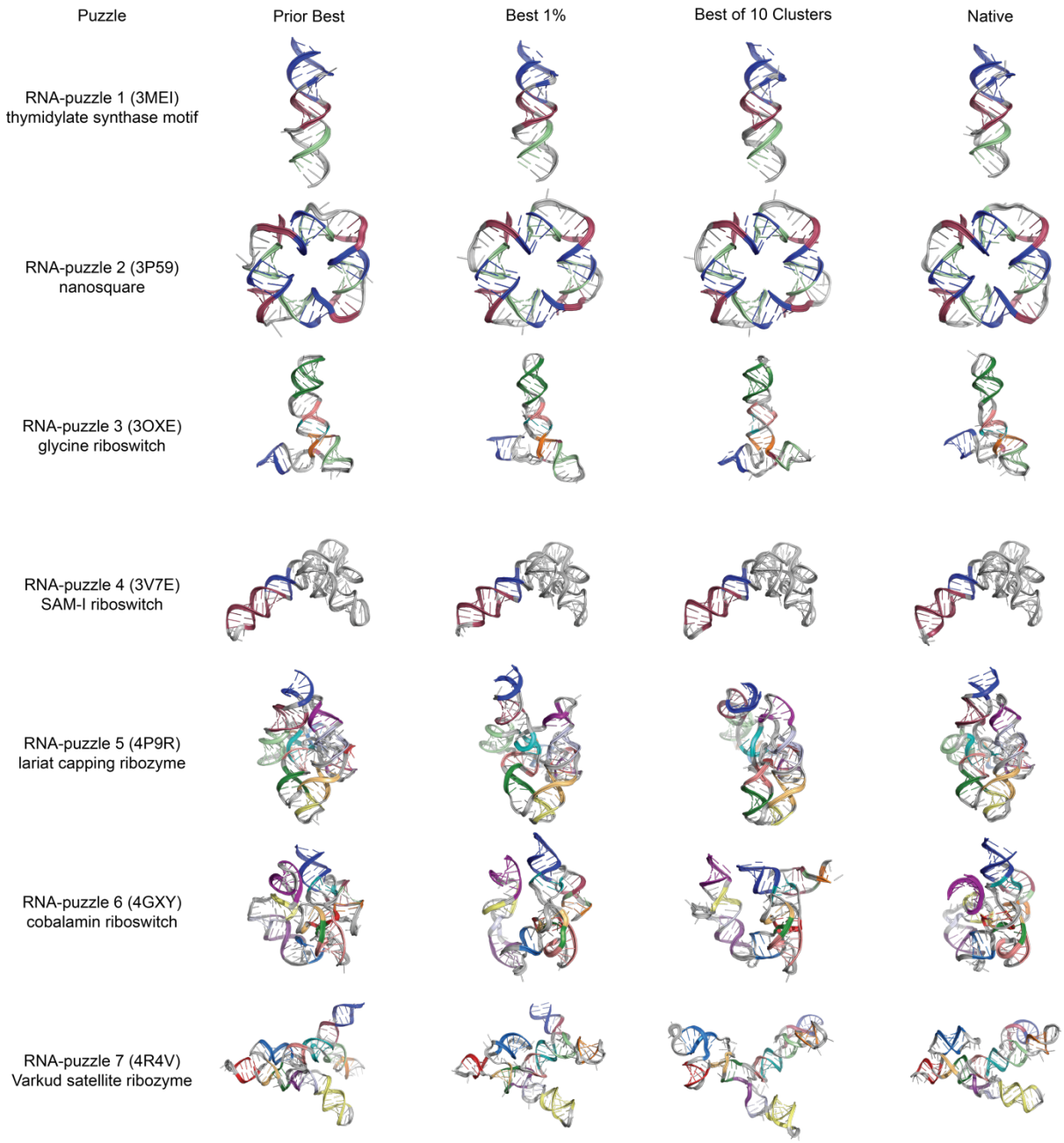


Figure S3. Related to Figure 3. Detailed depictions of FARFAR2-Puzzles benchmark cases 1-7, including the best originally submitted model, the FARFAR2 model with lowest RMSD in the top 1% of models overall, the lowest RMSD cluster center among the top 10 by energy, and the native structure. Models are colored to highlight distinct secondary structure elements.

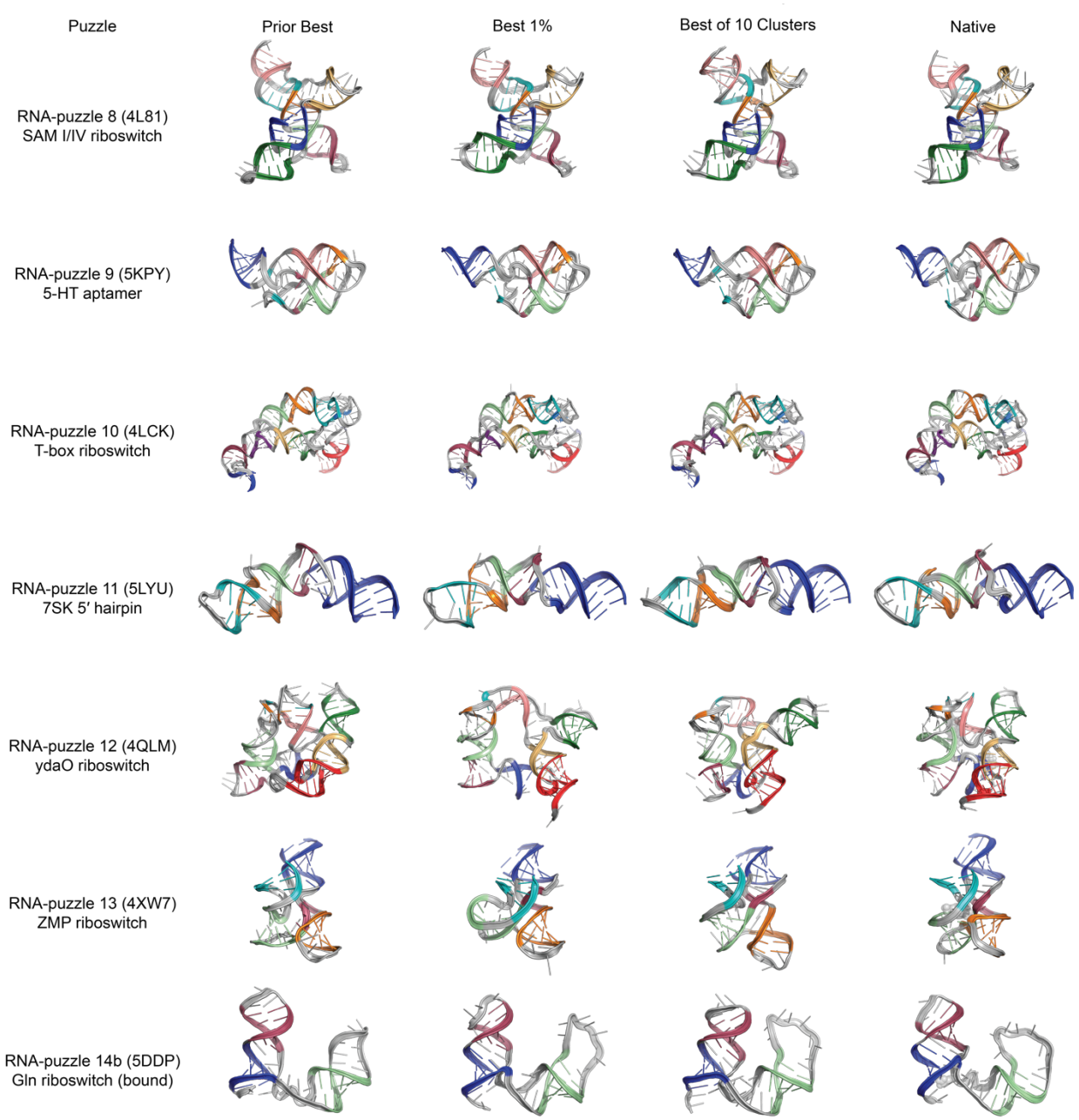


Figure S4. Related to Figure 3. Detailed depictions of FARFAR2-Puzzles benchmark cases 8-14b, including the best originally submitted model, the FARFAR2 model with lowest RMSD in the top 1% of models overall, the lowest RMSD cluster center among the top 10 by energy, and the native structure. Models are colored to highlight distinct secondary structure elements.

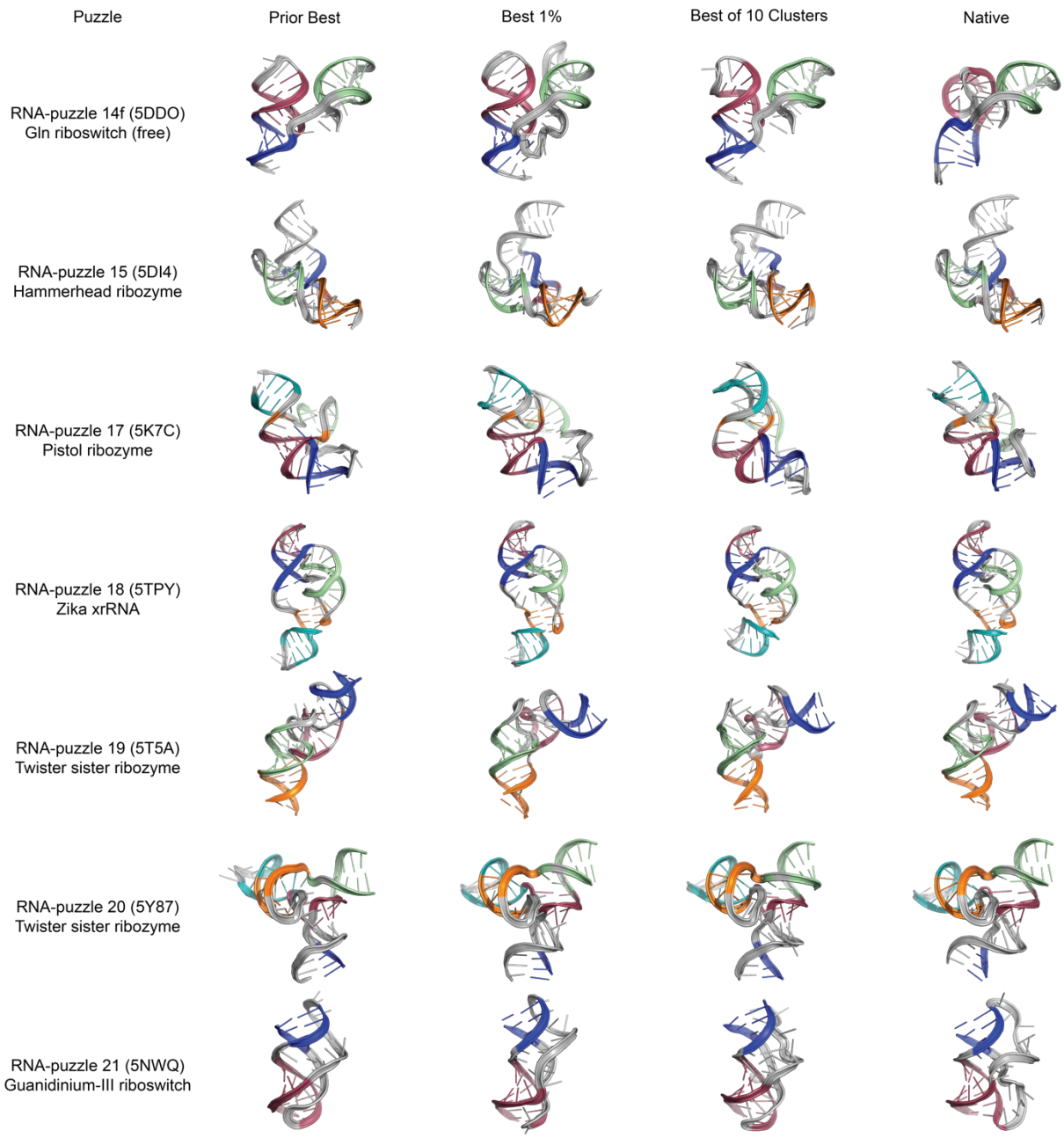


Figure S5. Related to Figure 3. Detailed depictions of FARFAR2-Puzzles benchmark cases 14f-21, including the best originally submitted model, the FARFAR2 model with lowest RMSD in the top 1% of models overall, the lowest RMSD cluster center among the top 10 by energy, and the native structure. Models are colored to highlight distinct secondary structure elements.

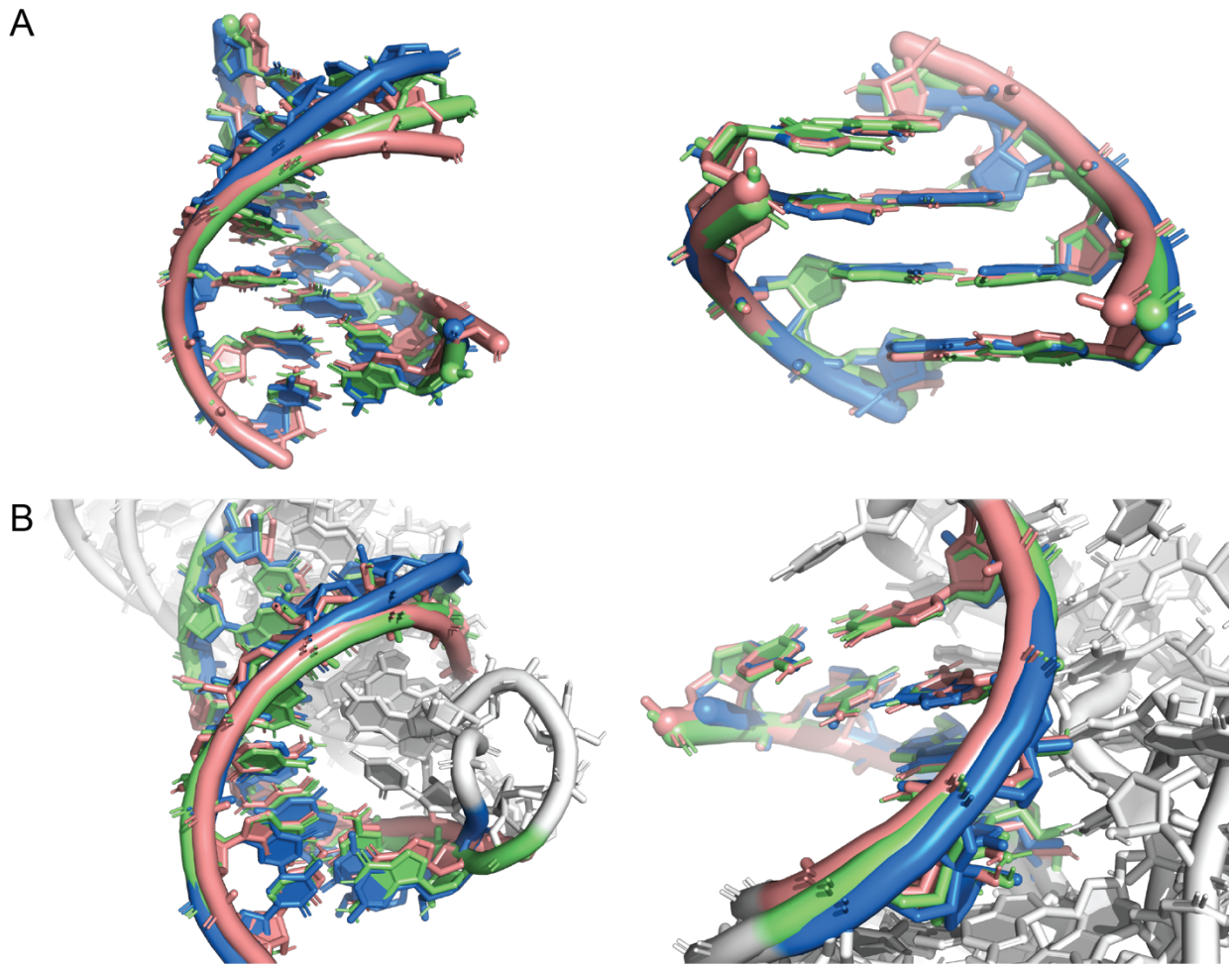


Figure S6. Related to Figure 5. Base pair steps provide the benefits of pre-generated helical ensembles. We conducted a detailed comparison of the performance of fixed helices versus base pair step sampling on both helices of RNA-Puzzle 21 (left and right). (A) We directly compared the two approaches for RNA-Puzzle 21 (native structures in blue). For both helices, fixed stems (pink) yielded a worse RMSD to the native helix conformation than base pair step sampling (green) (1.6 Å > 1.3 Å; 0.8 Å > 0.5 Å). (B) These improvements represented geometrically important flexibility, when viewed in the context of the full structure: the tightly pseudoknotted structure of Puzzle 21 features no external stacking on either helix, and deviations from ideality are essential within each helix in order to relieve strain and achieve conformations where the helix termini have no residues stacked upon them.

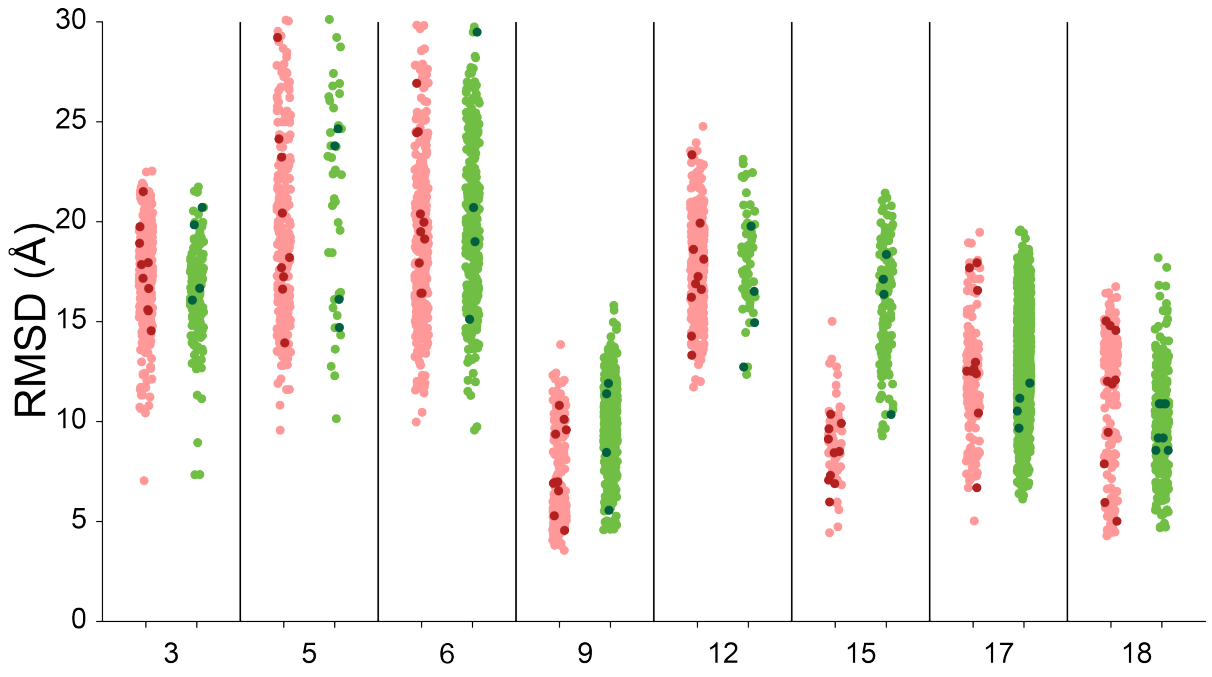


Figure S7. Related to Figure 5. A direct comparison of eight puzzles run with base pair step sampling (i.e., standard FARFAR2; pink points with red cluster centers) with an identical protocol using fixed helical stems (preserving the scoring function and fragment library; light green points with dark green cluster centers) confirms the utility of explicit sampling of helical flexibility.

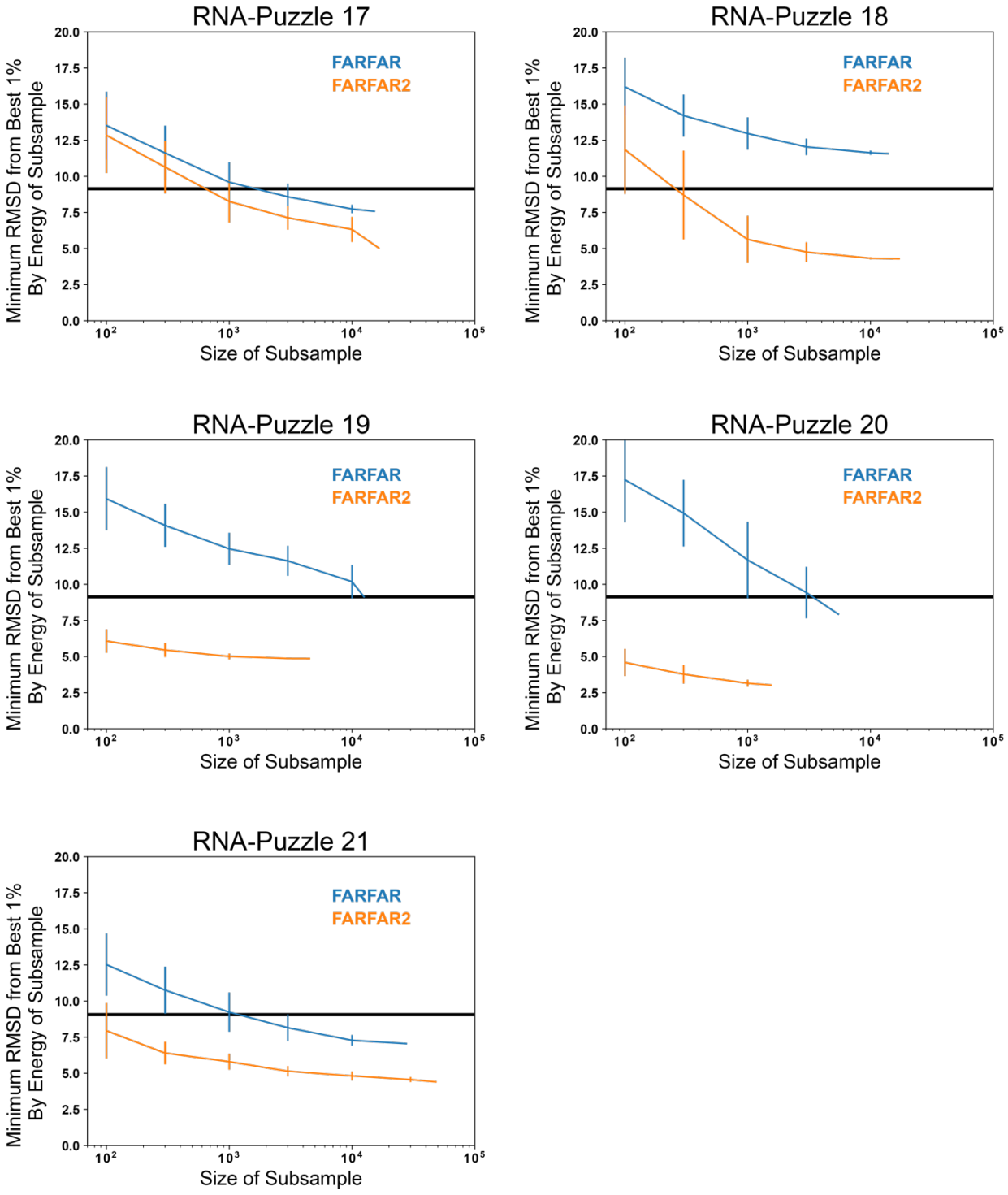


Figure S8. Related to Figure 3. Drawing subsamples from the respective model sets suggests that FARFAR2 (orange) begins to sample native-like (< 9.1 Å) models within its lowest energy percentile more rapidly than a reproduction of original FARFAR with only base pair constraint inputs. The results are especially stark on structures with substantial structure from input

templates (RNA-Puzzles 19, 20, 21), but FARFAR requires 10x or 100x as many structures to obtain comparable results in the remaining cases as well.

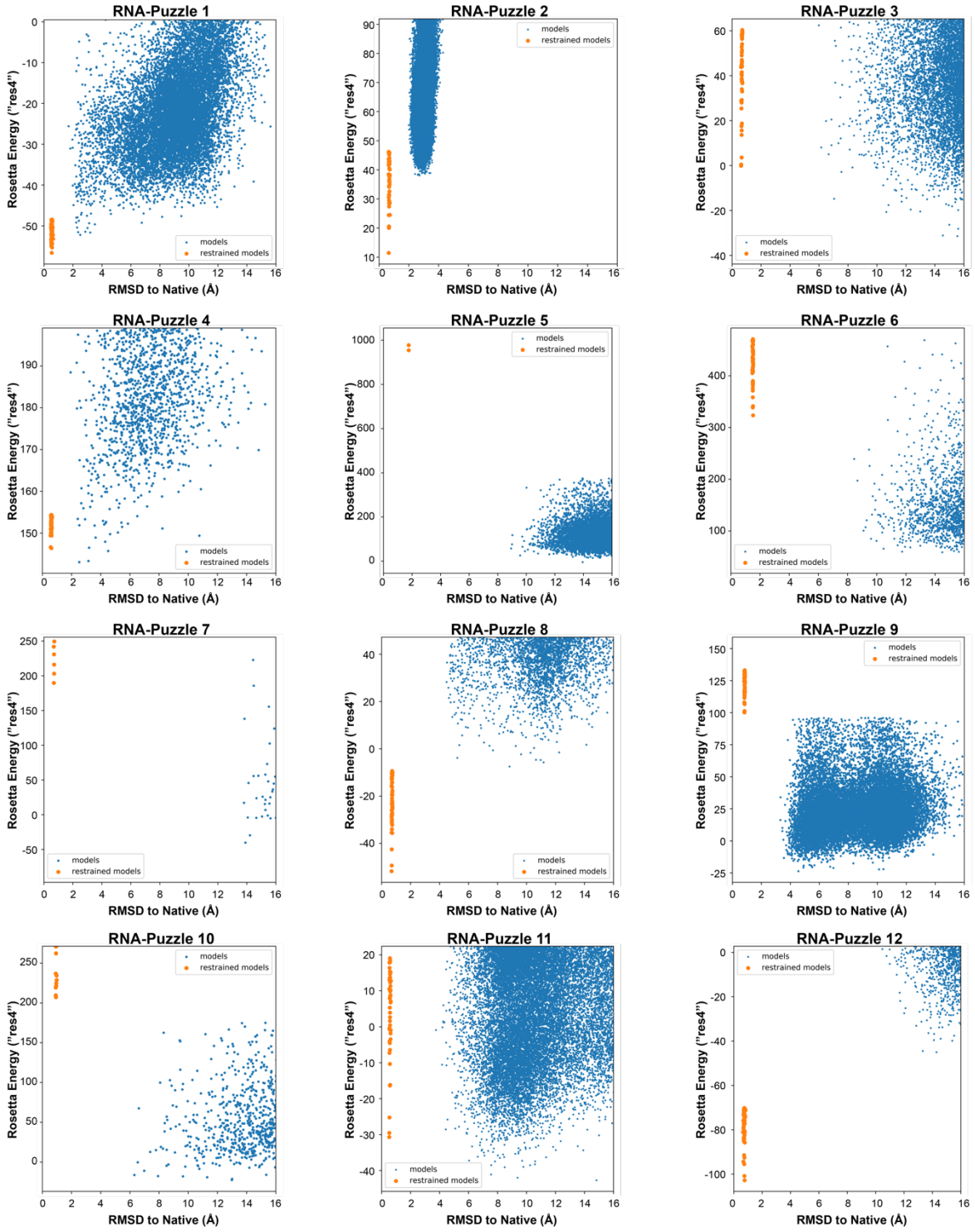


Figure S9. Related to Figure 3. For RNA-Puzzles 1-12, heavily restrained simulations (orange) forced to resemble the native reliably achieve sub-Ångstrom RMSDs, but often exhibit higher energies than the FARFAR2 model ensembles (blue).

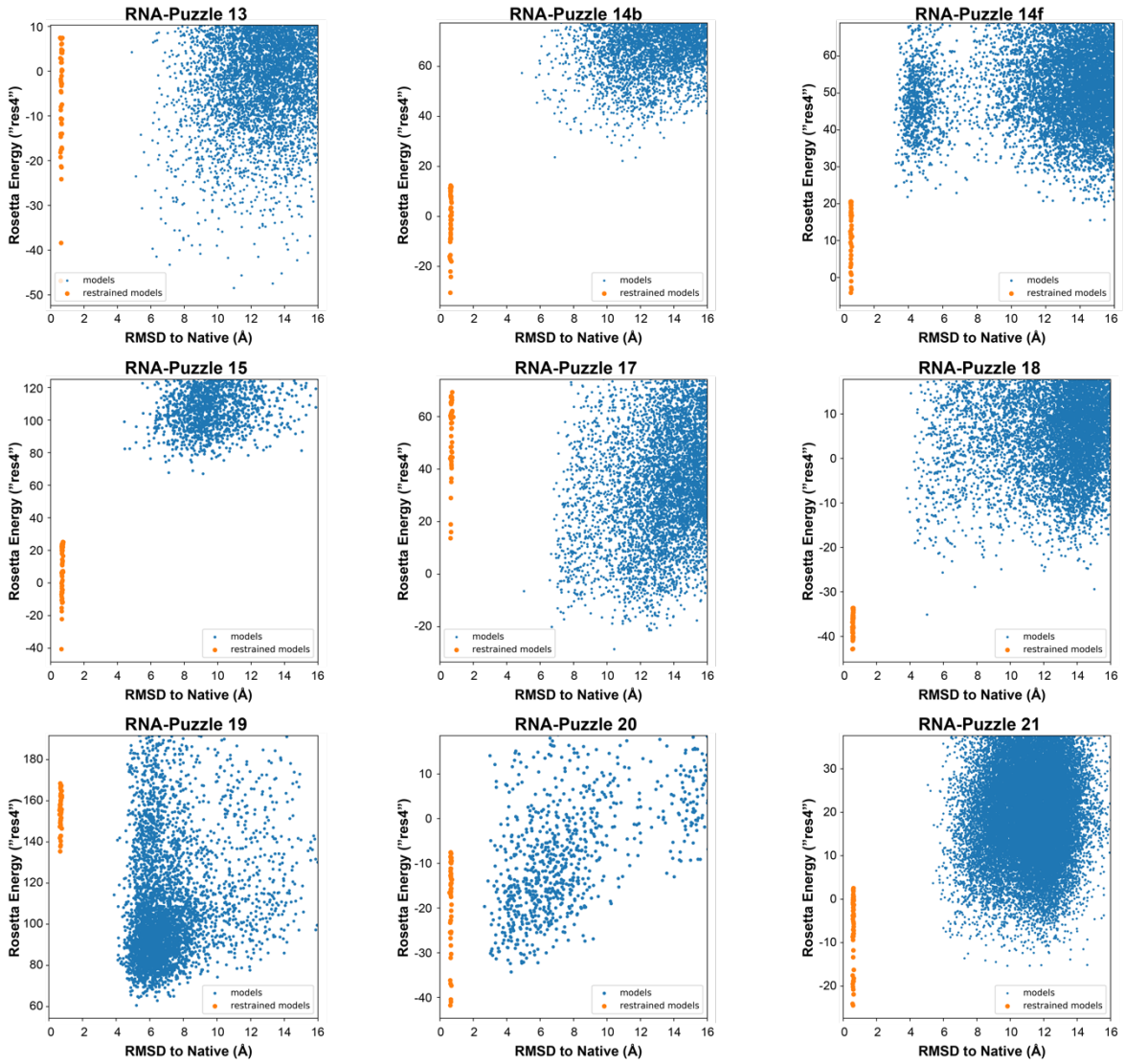


Figure S10. Related to Figure 3. For RNA-Puzzles 13-21, heavily restrained simulations (orange) forced to resemble the native reliably achieve sub-Ångstrom RMSDs, but often exhibit higher energies than the FARFAR2 model ensembles (blue).

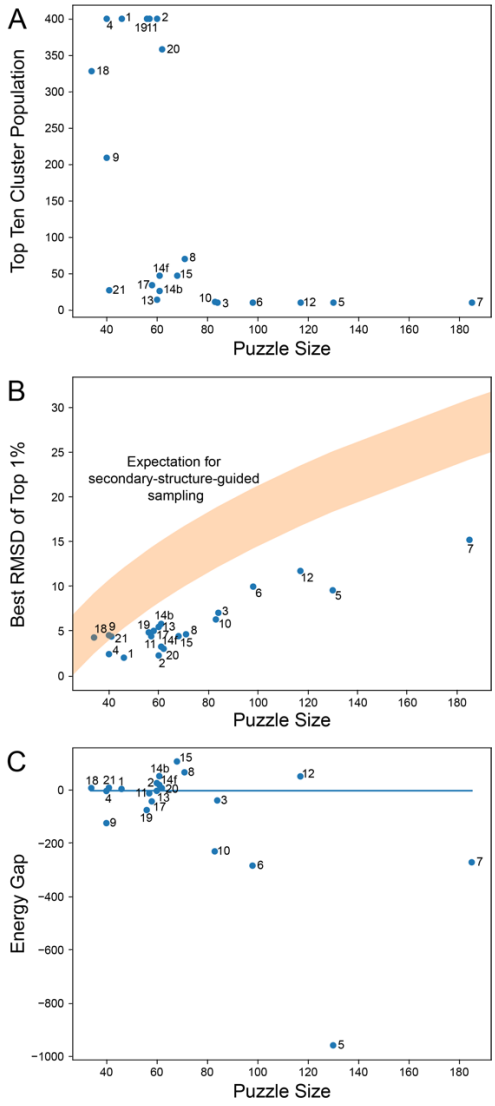


Figure S11. Related to Figure 3. (A) Larger puzzles tended to feature worse sampling, as judged by the total population of their top 10 clusters, though some smaller puzzles remain undersampled. (B) Nonetheless, across all puzzle sizes, the top 1% RMSDs from FARFAR2 simulations are superior to the threshold for a significant RNA structure prediction with specified secondary structure, which is given by the Weeks-Dokholyan power law relationship $5.1\ N^{0.41}$ - 15.8, where the shaded orange band represents a 95% CI (Hajdin, Ding, Dokholyan, & Weeks, 2010). (C) Larger puzzles also more often feature significantly positive or negative energy gaps.

Table S1. Related to Figure 1. Fragment assembly performance depends on multiple parameters optimized in this study. The performance of FARFAR2 (gauged by the lowest RMSD obtained from the 1% lowest energy models) was measured on the original FARNA benchmark to compare scoring functions, secondary structure input, fragment sets, and simulation guidance parameters. This study determined the ideal simulation parameters for use in the remaining larger-scale benchmarks.

																									no filters	-rounds 1	-rounds 2
Fragment library	Old												New												New	New	New
Mode of SS input	None			Fixed helix input			Secondary structure constraints			Base pair steps			None			Fixed helix input			Secondary structure constraints			Base pair steps			Base pair steps	Base pair steps	Base pair steps
Minimization scoring function	n/a	hires	rna4	n/a	hires	rna4	n/a	hires	rna4	n/a	hires	rna4	n/a	hires	rna4	n/a	hires	rna4	n/a	hires	rna4	n/a	hires	rna4	rna4	rna4	rna4
157d	1.385	0.797	0.513	1.363	0.763	0.556	1.288	1.288	1.288	2.003	1.258	0.941	0.839	0.653	0.457	1.470	0.786	0.476	1.065	1.065	1.065	1.644	0.947	0.618	1.018	1.651	0.944
1a4d	5.266	6.185	5.352	3.370	2.884	2.556	5.581	7.052	3.786	3.381	2.823	3.101	4.785	4.412	3.782	3.863	3.035	3.240	6.818	4.651	3.732	3.751	2.737	3.036	3.341	3.503	3.495
1csl	2.472	2.275	1.814	2.303	3.133	1.956	2.160	2.160	2.160	3.004	3.066	2.508	2.021	1.971	2.297	2.323	2.399	1.914	1.639	2.094	1.639	2.592	3.341	2.368	2.592	2.164	2.273
1dqf	0.968	0.699	0.972	0.704	0.708	0.791	1.064	0.938	1.087	1.028	1.029	0.933	1.621	1.242	1.072	0.699	0.768	1.159	1.645	1.842	1.810	0.929	0.989	0.666	0.772	0.726	1.226
1esy	3.141	2.334	2.217	1.866	1.743	1.543	2.788	3.170	3.134	2.269	2.087	1.995	2.459	2.459	2.459	1.965	1.925	1.637	2.770	2.834	2.770	2.048	2.130	1.842	2.059	1.989	1.871
1i9x	1.866	1.801	1.819	2.504	3.264	2.937	1.887	1.947	1.887	2.425	2.753	2.462	1.675	1.715	1.677	2.643	3.382	2.978	2.853	2.857	2.293	2.285	2.320	1.558	2.315	2.361	3.396
1kd5	3.000	2.686	2.362	3.128	2.240	2.240	2.112	2.143	2.143	2.826	1.676	1.677	3.402	2.785	2.576	3.215	2.794	2.579	2.128	2.660	1.718	2.636	1.947	1.979	1.383	2.334	1.651
1kka	3.785	3.546	3.546	5.233	5.263	5.210	3.603	3.688	3.482	3.262	3.111	3.377	3.888	3.739	3.486	5.155	5.090	5.223	3.616	3.531	3.413	3.215	3.331	3.252	3.693	3.579	3.493
1l2x	2.415	2.935	9.981	2.255	2.302	2.410	2.715	2.885	2.715	2.270	2.996	2.664	2.261	2.360	8.813	2.293	2.196	2.276	2.449	2.922	2.922	2.434	2.483	2.763	3.414	2.727	2.499
1mhk	7.539	6.591	6.639	5.770	4.458	5.079	6.451	4.264	4.310	3.668	3.848	4.305	5.114	4.921	5.204	5.473	5.195	4.835	6.988	6.409	6.505	5.809	4.556	4.735	4.780	6.025	4.716
1q9a	4.156	3.645	4.082	4.013	4.111	4.173	3.954	3.869	3.954	4.085	2.178	3.248	4.184	4.144	4.184	3.989	2.298	4.024	4.295	4.448	4.448	4.008	3.234	3.217	1.581	4.067	4.091
1qwa	3.378	3.293	3.733	2.747	2.928	3.136	3.291	3.456	3.344	2.869	3.480	4.005	3.613	3.345	2.911	2.928	2.652	2.863	3.328	3.397	3.328	2.358	2.729	3.375	2.967	2.793	3.840
1xjr	12.32	9.913	9.919	7.659	6.623	6.694	8.115	7.520	6.531	7.127	6.352	6.270	11.00	9.254	8.636	5.848	4.950	7.057	8.375	7.141	7.130	7.178	7.241	7.012	7.197	7.815	5.855
255d	1.354	1.282	1.296	1.556	1.146	1.078	1.221	1.234	1.234	1.488	1.517	0.982	1.339	1.173	1.249	1.237	1.070	1.160	1.208	1.208	1.208	1.162	1.155	1.201	1.135	1.016	1.290
283d	1.651	1.780	1.702	4.086	1.163	1.305	1.566	1.469	1.714	1.746	1.865	1.717	1.723	1.443	2.357	4.993	1.167	1.226	1.619	1.516	1.879	2.548	1.572	1.337	1.623	1.684	1.361
28sp	2.554	2.428	2.428	3.008	2.491	2.269	2.705	2.705	2.705	2.622	2.485	2.285	2.819	2.731	2.762	2.775	2.468	2.291	2.501	2.501	2.501	2.550	2.484	2.081	2.508	2.271	2.181
2a43	6.679	6.701	5.399	3.326	3.341	3.055	3.973	3.973	3.598	3.250	3.155	3.448	3.919	4.063	3.767	3.176	3.069	2.859	3.100	3.100	3.100	2.986	2.722	2.839	3.556	3.807	2.468
2f88	2.822	2.856	2.285	2.156	3.143	2.491	3.976	3.639	2.879	1.896	2.053	2.250	2.876	2.608	2.298	2.333	2.388	2.278	4.334	3.201	2.293	2.294	1.977	2.183	2.256	2.026	2.008

Table S2. Related to Figure 1. Comparison of fragment assembly performance using the modern scoring function on the 'motif-scale' benchmark set versus the performance of SWM on the same benchmark. More models were generated for the FARFAR version of the benchmark, but comparable computational time was required in each case. Unlike in the original SWM work, the relevant metric compared is the best RMSD sampled from among the 1% of models with the best energy.

1

Benchmark Case	FARFAR 1% RMSD (A)	SWM 1% RMSD (A)	FARFAR Best of 5 Cluster RMSD	SWM Best of 5 Cluster RMSD
Trans-Helix Loops				
5P_j12_leadzyme	2.56	0.82	2.56	1.22
5P_p1_m_box_riboswitch	2.64	2.13	2.68	0.63
3P_j55a_group_I_intron	0.27	1.72	0.42	0.40
5P_j55a_group_I_intron	3.28	3.25	3.87	0.74
hepatitis_C_virus_ires_IIa	3.23	3.11	3.47	2.56
j24_tpp_riboswitch	1.38	1.22	3.51	0.72
j23_group_II_intron	1.77	1.11	2.57	0.55
j31_glycine_riboswitch	5.53	0.54	5.53	0.65
l1_sam_II_riboswitch	0.67	3.14	0.89	0.83
l2_viral_rna_pseudoknot	1.76	4.28	1.94	0.71
23s_rrna_44_49	1.20	0.71	1.23	1.25
23s_rrna_531_536	4.61	1.49	5.41	1.48
23s_rrna_2534_2540	1.97	7.28	2.90	6.94
23s_rrna_1976_1985	8.19	12.18	9.82	15.27
23s_rrna_2003_2012	7.54	9.85	9.85	9.02
Total cases: 15	*6*	*9*	*3*	*12*
Apical Loops				
gcaa_tetraloop	1.36	1.18	1.55	1.14
uucg_tetraloop	2.57	2.29	2.03	1.14
gagua_pentaloop	0.75	2.94	0.76	1.10
anticodon_phe	1.30	2.82	2.21	2.24
Total cases: 4	*2*	*2*	*2*	*2*
Two-Way Junctions, Fixed				
puzzle1_alt_fixed	1.98	0.409	2.29	1.72
srp_domainIV_fixed	0.50	1.034	0.69	0.90
srl_fixed	0.54	0.707	0.69	0.56
kink_turn_fixed	0.80	2.02	1.28	1.45
j55a_P4P6_fixed	1.78	3.20	1.86	0.55
P5b_connect	0.76	3.63	0.76	2.68

gg_mismatch_fixed	1.87	0.32	2.03	0.32
tandem_ga_imino_fixed	1.07	0.86	1.21	0.87
tandem_ga_sheared_fixed	0.37	0.61	0.50	0.61
hiv_rre_fixed	0.36	0.87	0.45	0.35
j44a_p4p6_fixed	0.75	0.56	1.03	0.58
just_tr_P4P6_fixed	0.66	0.63	0.68	0.61
r2_4x4_fixed	0.72	1.44	1.20	1.49
loopE_fixed	0.51	1.74	0.72	1.74
Total cases: 14	*9*	*5*	*6*	*8*
Three-Way Junctions, Fixed				
hammerhead_3WJ_cat_fixed	3.25	1.62	3.71	3.05
hammerhead_3WJ_precat_fixed	1.16	1.60	1.90	1.57
VS_rbzm_P2P3P6_fixed	0.76	0.74	1.09	0.57
VS_rbzm_P3P4P5_fixed	1.21	1.91	1.21	1.91
hammerhead_3WJ_cat_OMC_fixed	2.85	2.16	4.56	3.04
Total cases: 5	*2*	*3*	*1*	*4*
Tertiary Contacts, Fixed				
tl_tr_P4P6	0.55	1.07	0.55	0.64
hammerhead_tert_fixed	0.96	2.45	1.00	1.16
kiss_add_fixed	2.28	3.24	2.69	2.40
kiss_add_L2_fixed	0.52	1.50	0.69	0.71
kiss_add_L3_fixed	1.42	2.11	1.57	0.88
puzzle18_zika_PK	2.67	2.01	2.67	1.97
gir1_p2.1p5_kiss_fixed	1.19	1.94	1.62	1.99
gir1_p2p9_gaaa_minor_fixed	0.83	2.02	1.02	1.58
t_loop_fixed	0.78	2.27	0.78	0.95
t_loop_modified_fixed	2.37	1.24	0.98	1.33
Total cases: 10	*8*	*2*	*7*	*3*
Two-Way Junctions, Aligned				
gg_mismatch	1.12	2.65	1.12	0.79
tandem_ga_imino	0.89	1.32	1.03	0.98
tandem_ga_sheared	0.49	1.06	0.61	0.75

hiv_rre	1.78	3.59	1.94	2.12
j44a_p4p6	1.01	4.75	1.56	1.59
just_tr_P4P6	0.74	2.47	0.96	1.22
cg_helix	0.23	0.63	0.28	0.58
puzzle1	0.91	3.26	0.84	0.96
srp_domainIV	0.94	2.88	1.01	1.26
r2_4x4	1.84	3.34	1.84	1.74
gagu_forcesyn_blockstackU	4.65	5.64	4.87	4.49
srl_free_bulgedG	4.59	6.38	4.76	4.66
j55a_P4P6_align	0.85	2.81	1.04	2.04
kink_turn_align	0.79	3.01	0.97	2.07
loopE	0.91	5.28	1.75	2.00
Total cases: 15	*15*	*0*	*10*	*5*
Three-Way Junctions, Aligned				
hammerhead_3WJ_precat	2.83	10.74	4.18	6.09
VS_rbzm_P2P3P6_align	0.73	1.29	0.84	1.13
VS_rbzm_P3P4P5_align	1.04	2.40	1.86	2.60
hammerhead_3WJ_cat_OMC_align	2.56	2.89	4.22	2.89
puzzle18_zika_3WJ_extraminres	2.52	4.36	2.99	2.42
Total cases: 5	*5*	*0*	*3*	*2*
Tertiary Contacts				
gaaa_minor_dock	1.02	2.26	1.19	1.41
gir1_p2.1p5_kiss	1.48	3.25	2.01	2.75
gir1_p2p9_gaaa_minor	1.17	2.60	1.16	1.83
tl_tr_P4P6_dock	0.81	6.83	0.81	3.03
kiss_add_PK_dock	2.07	3.46	3.35	2.58
t_loop_align	2.02	4.16	2.02	3.20
hammerhead_tert_align	3.08	7.87	3.90	8.65
t_loop_modified_align	1.93	3.66	2.44	3.99
Total cases: 8	*8*	*0*	*7*	*1*
Non-Helix Embedded				
cg_helix_Zform	5.00	10.76	5.00	1.75

g_quadruplex_fixed	1.58	3.57	1.79	2.75
g_quadruplex_inosine_fixed	1.84	2.43	2.03	2.87
bru_gag_tetraplex	3.30	2.78	2.70	3.42
parallel_AA	0.97	1.22	1.16	1.41
bulged_tetraplex	4.95	7.67	5.04	7.37
Total cases: 6	*5*	*1*	*5*	*1*
Overall: 82	*60*	*22*	*44*	*38*

1

Table S3. Related to Table 1. Exact inputs (sequence and RNA templates) used in modeling for each RNA-puzzle problem.

Puzzle	Secondary Structure	Inputs
1	ccgccgcgccaugccuguggcgg,ccgccgcgccaugccuguggcgg (((((((..((((...(((((((,)))))))..))))...)))))))	Just secondary structure
2	ccggaggaacuacugccggcagccuccggaggaacuacugccggcagccuccggaggaac uacugccggcagccuccggaggaacuacugccggcagccu {{{{{(((.....(((aaaa)))))){{{<<<(((.....(((}}}}))))))[[[[(((... ..(((>>>>)))))aaaa(((.....(((]]]]))))))	Four of eight chains were provided with puzzle specification. Three 'corners' modeled as HCV IRES loops.
3	cucuggagagaaccguuuaaucggucgccgaaggagcaagcucugcgcauaugcagagug aaacucucaggcaaaaggacagag (((((......(((.....)))))))(((...((((...(((((((((....))))))))) ...)))...)))......))))))	Just secondary structure
4	ggcuuaucaagagagguggagggacuggcccgaugaaacccggcaaccacuagucuagcg ucagcuucggcugacgcuaggcuaguggugccaauuccugcagcggaaacguugaaagau gagcca .....(((((.(((((((((((....)))))))))))).))))).....	3IQP, the template at the time, provides residues A:1-47 A:88-126
5	gguuggguugggaaguaucauggcuaaucaccaugaugcaaucggguugaacacuuaauu ggguuaaaacgguggggggacgaucccguaacauccguccuaacggcgacagacugcacgg cccugccucuuagguguguccaaugaacagucguuccgaaaggaagcauccgguauccca agacaauc (((((..(((((..((((((((((...[[[.)))))))))..((((.(((.((((..... (.........((..(((...]]]))).....))..... ..).....))))).)))..........((((....)))).....))))...))))))..))))	Template structure 3BO3 (group I intron) used in original Das lab modeling for A:62-70 A:96-122 A:148-153; A:55-60 A:124-129 A:132-135; they are thus omitted from the secondary structure.
6	cggcaggugcucccgac,gucgggaguuaaaagggaagccggugcaaguccggcacgguc ccgccacugugacggggagucgccccucgggaugugccacuggcc,ggccgggaaggcgg aggggcggcgaggauccggagucaggaaaccugccugccg (((((((.(((.(((.,.)).)))).....((....(((((.....)))))[[((.. ...)).(((...(((...(((((((.((.....(((.((((((,))))))...))).)).)))))).....)))..]]))).....))))))))))	Template structures 2YIE (FMN aptamer) and 2GIS (SAM riboswitch) used in original Das lab modeling provided A:5-8 A:32-35 A:37-39 A:158-164; A:81-83 A:138-143; A:41-61 A:67-75 A:147-149.
7	gcgcugugucgcaaucugcgaagggcgucgucggcccaagcgguaguaagcagggaacuc accuccaaugaaacacauugucguagcaguugacuacuguuaugugauuggguagaggcua agugacgguauuggcguaagccaauaccgcagcacagcacaagcccgcuugcgagauuac agcgc ((((((((((.....((((...((((.....))))..)))).((((((.(((..((((((..(((((.....))))).((((((.....)))))).....))).))).(((. .(((.((((((((((....))))))))))...)))))).....))))))))..))).))))))))	Just secondary structure.
8	ggaucacgaggggagaccccggcaaccugggacggacacccaaggugcucacaccggag acgguggauccggcccgagagggcaacgaaguccgu (((((....((((....))))(((..((((((.[[[[..))).))).))))..(((((... .))))))))))).((((....))))......]]]].	Template structures from original Das lab modeling for SAM binding site using 2YGH, as described previously.

9	ggacacugaugaucgcguggauauggcacgcauugaauuguuggacaccguaaauguccuaacacgugucc ((((((.....(.(((((.....[[)))))).........(((((]].....)))))).....)))))	Template structures from original Das lab modeling: kissing loop from 2XNW; T-loop from the tRNA Phe structure 3L0U.
10	ugcgaugagaagaagaguauuaaggauuuacuaugauuagcgacucuaggauagugaaagcuagaggauaguaaccuuaagaaggcacuucgagca,aguaguucagugguagaacaccaccuugccaaggugggggucgcggguucgaaucccgucu (((.....((((.......((((((..(((((((.........((((((..........))))))).)))))))))))...[[[.))))..))), ((..((((....{..)))).((((((..]]]..)))))).....(((((..}....))))))))	Template structures from original Das lab modeling: a kink-turn from 2GIS provided A:1-10, A:90-96; a loopE motif from 5S rRNA 354D furnished A:16-20, A:80-83; a double T-loop motif from 4JRC supplied A:37-43, A:53-60; and part of a tRNA from 2K4C delivered B:6-18, B:20-3,0 B:38-66.
11	gggaucugucaccccauugaucgccuucgggcugaucuggcuggcuaggcgggucccc ((((((((((.(((((..(((((((....))).)))))))..)).))))))))))	Just secondary structure
12	gaucgcugaacccga,aggggcgggggacccag,ggggcgaaucucuuccgaaaggaagaguaggguuacuccuucgacccgagcccgucagcuaaccucgcaagcguccgaaggagaa ...((((...(((..,.))))))((((((....(..,((((((....(((((((....)))))))))))..(((((.[[[[[[[[)))))..))))..)..).....))))))..))))...]]]]]]]]..	Just secondary structure
13	gggucgugacuggcgaacaggugggaaaccaccggggagcgaccc,gccgcccgccugggc (((((((..(([[[[....(((((....)))))).))..)))))))),(((...]]]]..))))	Just secondary structure
14b	cguugacccaggaaacugggcggaaguaaggcccauugcacuccgggccugaagcaacgcg (((((.(((((....))))).........((((((..........))))))....))))). .	Just secondary structure (original puzzle omits U1A loop in RMSD calculations)
14f	cguuggcccaggaaacugggu,aguaaggcccauugcacuccgggccugaagcaacgcu ((((((((((....))))))),....((((((..........))))))....)))))..	Just secondary structure (original puzzle omits U1A loop in RMSD calculations)
15	ggguacuuaagcccacugaugagucgcugggaugcgacgaaacgccca,gggcgucugggcaguaccca ..........((((((....).(((((......)))))...((((((.,))))))..))))))......... .	Just secondary structure
17	cgugguuagggccacguuaaauaguugcuuaagcccuaagcguugau,aucaggugcaa ((((((((((])))))).........((((.....]]]]]]....(((.,.))))...))))	Just secondary structure
18	gggucaggccggcgaaagucgccacaguuuggggaaagcugugcagccuguaacccccccacgaaaguggg ....(((((((((....)))))).((((((..[[[[...))))))..))))))...]]]](((((....)))))	Template structures from original Das lab modeling from 4PQV.
19	gcagggcaaggcccaguccgugcaagccgggaccgcccc,ggggcgcggcgcucauuccugc (((((...(((((..((((((.......))))))((((((,))))))..)))..))....))))))	Template structures from original Das lab modeling: threaded T-loop from 1B23, with intercalating A:8 modeled in. (Other suspected intercalators like A:9 and B:15 score strictly worse.)
20	acccgcaaggccgacggc,gccgccgcuggugcaaguccagccacgcuucggcgugggcgcucaugggu	Template structures from original Das lab modeling: threaded T-loop from puzzle 19 (5T5A), renumbered.

	((((...((((((.((((,)))).)(((((.......)))))(((((....)))))))).))...))))	
21	ccggacgaggugcgccguacccggucaggacaagacggcgc [[[[.......(((((((..]]]].........)))))))	Just secondary structure, no guanidinium.

1

2

Table S4. Related to Table 1. Comparison of performance between a reproduction of FARFAR original modeling conditions (where secondary structure was modeled using base pair constraints) versus FARFAR2.

Puzzle	FARFAR2 Top 1% Best RMSD	FARFAR2 Best of 10 Cluster RMSD	FARFAR Top 1% Best RMSD	FARFAR Best of 10 Cluster RMSD
17	5.03	6.69	7.58	11.43
18	4.29	5.02	11.57	11.78
19	4.86	5.16	9.14	13.44
20	3.03	4.03	7.94	7.94
21	4.40	6.04	7.05	10.26

Table S5. Related to Table 1. Energy gaps between the minimum energies sampled in artificially restrained near-native simulations and FARFAR2 models for each FARFAR2-Puzzles benchmark case. A positive energy gap indicates that the best energy achieved in FARFAR2 *de novo* modeling is higher (worse) than the energy observed for near-native conformations; a negative energy gap indicates that the best energy from the FARFAR2-Puzzles benchmark is also the best energy overall. The total population in the top 10 clusters (maximum: 400) also indicates whether the benchmark case was thoroughly sampled.

Puzzle	RNA	Energy Gap	Total Top 10 Cluster Population
1	thymidylate synthase motif	4.4	400
2	nanosquare	26.6	400
3	glycine riboswitch	-38.7	10
4	SAM-I riboswitch	-3.3	400
5	lariat capping ribozyme	-958.5	10
6	cobalamin riboswitch	-283.6	10
7	VS ribozyme	-270.9	10
8	SAM I/IV	67.0	70
9	5-HT aptamer	-123.8	209
10	T-box riboswitch	-229.9	11
11	7SK 5′ hairpin	-12.0	400
12	ydaO riboswitch	51.7	10
13	ZMP riboswitch	-2.7	14
14b	Gln riboswitch (bound)	52.7	26
14f	Gln riboswitch (free)	19.6	47
15	hammerhead ribozyme	107.7	47
17	pistol ribozyme	-42.2	34
18	Zika xrRNA	7.8	328
19	twister sister ribozyme	-74.8	400
20	twister sister ribozyme	7.5	358
21	guanidinium-III riboswitch	9.1	27

Table S6. Related to Table 1. Detailed results for each RNA-Puzzle challenge revisited in this work.

Puzzle	FARFAR2 Top 1% Best by RMSD†			FARFAR2 Best of 10 Cluster by RMSD†			Best RNA-puzzle RMSD (All Submissions)†		
	RMSD	F_{NWC}	Clashscore*	RMSD	F_{NWC}	Clashscore	RMSD	F_{NWC}	Clashscore
1	2.03	0.90	2.72	2.50	0.86	4.07	3.40	0.90	0.00
2	2.28	0.00	11.31	2.71	0.00	10.99	2.30	0.50	14.23
3	7.05	0.11	4.79	12.41	0.00	2.58	7.60	0.22	0.00
4	2.43	1.00	22.16	2.52	1.00	19.69	3.40	0.88	1.97
5	9.57	0.25	3.64	13.94	0.26	2.32	9.58	0.20	9.93
6	9.98	0.20	7.82	13.08	0.20	10.35	12.28	0.27	27.88
7	15.21	0.00	5.72	18.52	0.14	2.69	20.72	0.50	11.25
8	4.65	0.50	10.27	5.23	0.75	5.45	4.80	0.75	13.77
9	4.54	0.64	2.64	4.56	0.50	1.76	5.86	0.29	18.53
10	6.31	0.70	5.54	6.31	0.65	5.54	6.78	0.70	13.82
11	4.43	0.00	0.55	6.04	0.14	1.10	5.22	0.00	0.55
12	11.73	0.15	3.71	13.32	0.23	4.24	10.15	0.00	12.61
13	5.47	0.00	7.73	7.13	0.67	6.19	5.41	0.33	10.85
14b	5.81	0.00	6.61	6.88	0.00	4.06	5.79	0.50	11.65
14f	3.26	0.67	7.00	11.85	0.00	3.23	6.05	0.83	16.24
15*	4.44	0.75	8.24	5.98	0.75	6.40	5.30	0.50	5.91
17	5.03	0.00	3.24	6.69	0.00	2.16	7.13	0.11	6.53
18	4.29	0.33	3.92	5.02	0.50	4.79	3.15	1.00	0.43
19	4.86	0.67	15.07	5.16	0.67	11.52	5.50	0.33	18.97
20	3.03	0.67	3.21	4.03	0.67	2.75	6.80	0.33	37.37
21	4.40	0.00	4.52	6.04	0.00	0.75	3.93	0.11	13.53

*Clashes may come from input template structures that were not employed in the previously submitted modeling.

†Heavyatom RMSD is calculated over all residues, following superposition over all residues.

Table S7. Related to Table 2. Exact inputs (sequence and RNA templates) used in modeling for each blind challenge modeled in this work.

Blind challenge	Secondary Structure	Inputs
F. nucleatum glycine riboswitch	ucggaugaagauaugaggagagauuucauuuuaaugaaacaccgaagaag uaaaucuuucagguaaaaaggacucauauuggacgaaccucuggagagcu uaucuaagagauaacaccgaaggagcaaagcuaauuuuagccuaaacucu cagguaaaaggacggag (((….…((((((((.....((((((....))))))).(((...((((.. ...))))..)))........))))))))...)))..(((((.....((((((...)))))).(((....(((....((((....))))).....))). ..))).......)))))	3P49 furnished residues A:19-29 A:34-71 A:92-164.
V. cholerae glycine riboswitch	uccguugaagacugcaggagagugguuguuaaccagauuuuaacaucuga gccaaauaacccgccgaagaaguaaaucuuucaggugcauuauucuuagc cauauauuggcaacgaauaagcgaggacuguaguuggaggaaccucugga gagaaccguuuaaucggucgccgaaggagcaagcucugcgcauaugcaga gugaaacucucaggcaaaaggacagagga [[[......((((((((.....((((((...(((((.....)))))). ....)))))))..(((....(((.....)))...)))((.((((((...(((.....)))).)))))))).....)))))))...]]]..(((((.. ..(.((((.....)))))(((...((((...(((((((.....)))))))....))).).).)))......)))))..	Threaded template from lowest energy model of *F. nucleatum* glycine riboswitch modeling above: residues A:1-29 A:34-64 A:68-91 become A:1-29 A:56-86 A:125-148.
Mycobacterium SAM-IV riboswitch	ggucaugagugccagcgucaagccccggcuugcuggccggcaacccucca accgcgguggggugcccgggugaugaccagguugaguagccgugacggc uacgcggcaagcgcggguc ((((....(.((((((....((.[[[[[]).)))))))(((..(((.((. .{{{.))..)))..)))]]]]]....))))..((((.((((((.....))))))).))))).....}}}...	SAM binding site from 2YGH provided A:1-9 A:38-43 A:63-66 A:76-79, with SAM as B:120
G. kaustophilus T-box riboswitch/tRNA-Gly	gcggaaguaguucagugguagaacaccaccuugccaaggugggggucgcg gguucgaaucccgucuuccgcucca, gaaagugggugcgcguuuggcgcaucaacucgggguggaaccgcgggagcu acgcucucgucccgag (((((..... .....)))))[[[[, ...(((((((.....)))))))....((((]]]]...((((((.... .....)))))))))..	Threaded tRNA template from 4LCK: residues B:6-66
B. subtilis T-box riboswitch/tRNA-Gly	gcggaaguaguucagugguagaacaccaccuugccaaggugggggucgcg gguucgaaucccgucuuccgcucca, guugcagugagagaaagaaguacuugcguuuaccucaugaaagcgaccuu agggcgguguaagcuaaggaugagcacgcaacgaaaggcauucuugagca auuuuaaaaaagaggcugggauuuuguucucagcaacuaggguggaaccg cgggagaacucucgucccua (((((..... .....)))))[[[[, .((((.....((((.....(((((((....(((((.....(((((((.....)))))))))))).)))))))........)))))...)))	Threaded tRNA template from 4LCK: residues B:6-66; threaded T-box template from 4LCK: residues A:2-12 A:19-29 A:34-74 A:76-84 A:87-89 A:94-101

	).....((((((.....)))))).....(((((]]]]...(((((....))))))))))	
VA RNA I	ggaccucgcaaggguaucauggcggacgaccgggguucgaaccccggaucccggccguccgccgugauccaugcgguuaccgcccgcgugucgaacccaggugugcgaggucc ((((((((((.((((.(((((((((((.(((((.((........)).)))))).)))))))))))..(((((((..[[[..)))))))....)))))..]]].))))))))))	Just secondary structure