nature methods

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41592-020-0878-9

In the format provided by the authors and unedited.

Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures

Kalli Kappel[1,8], Kaiming Zhang[2,8], Zhaoming Su[2,6,8], Andrew M. Watkins[3], Wipapat Kladwang[3], Shanshan Li[2], Grigore Pintilie[2], Ved V. Topkar[1], Ramya Rangan[1], Ivan N. Zheludev[3], Joseph D. Yesselman[3,7], Wah Chiu[1,2,4] ✉ and Rhiju Das[1,3,5] ✉

[1]Biophysics Program, Stanford University, Stanford, CA, USA. [2]Department of Bioengineering, James H. Clark Center, Stanford University, Stanford, CA, USA. [3]Department of Biochemistry, Stanford University, Stanford, CA, USA. [4]Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA, USA. [5]Department of Physics, Stanford University, Stanford, CA, USA. [6]Present address: The State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Sichuan, China. [7]Present address: Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE, USA. [8]These authors contributed equally: Kalli Kappel, Kaiming Zhang, Zhaoming Su. ✉e-mail: wahc@stanford.edu; rhiju@stanford.edu

Supplementary information for:

Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures

Kalli Kappel[#], Kaiming Zhang[#], Zhaoming Su[#], Andrew M. Watkins, Wipapat Kladwang, Shanshan Li, Grigore Pintilie, Ved V. Topkar, Ramya Rangan, Ivan N. Zheludev, Joseph D. Yesselman, Wah Chiu*, Rhiju Das*

[#]These authors contributed equally.
*Correspondence to: rhiju@stanford.edu, wahc@stanford.edu

This PDF file includes:

Other supplementary information for this manuscript includes the following:

Supplementary Note 1. RNA preparation.

DNA templates for all molecules except *Tetrahymena* ribozyme were prepared by PCR assembly of DNA oligonucleotides designed with Primerize[1] and purchased from Integrated DNA Technologies (Supplementary Table 5). PCR reactions contained 1× HF buffer (New England Biolabs (NEB)), 0.2 mM dNTPs, 4 μM end primers, 40 nM middle primers, and 40 U/mL Phusion Polymerase. Reactions were first denatured at 98 °C for 30 s. Then for 35 cycles, they were denatured at 98°C for 10 s, annealed at 67 °C (*V. cholerae* glycine riboswitch) or 65°C (all other RNAs), and incubated at 72 °C for polymerase extension for 30 s. Finally, reactions were incubated at 72 °C for 10 min. DNA templates were purified with AMPure XP beads (Beckman Coulter) following the manufacturer's instructions. RNA transcription reactions for all RNAs except *Tetrahymena* ribozyme contained 0.2 μM DNA template, 40 mM Tris·HCl, pH 8.1, 25 mM $MgCl_2$, 3.5 mM spermidine, 0.01% TritonX-100, 40 mM DTT, 4% PEG 8000, 3 mM NTPs, and 5 U/μL T7 RNA polymerase (NEB). Reactions were incubated at 37 °C for 4 h. RNAs were purified with Zymo RNA Clean and Concentrator columns (Zymo Research) following the manufacturer's instructions.

The DNA template for the *Tetrahymena* ribozyme was PCR amplified from the pT7L-21 plasmid[2] using the primers listed in Supplementary Table 5. The reaction contained 1× HF buffer (New England Biolabs (NEB)), 0.2 mM dNTPs, 4 μM primers, 40 pg/μL plasmid, and 40 U/mL Phusion Polymerase. The reaction was performed as described above except with an annealing temperature of 69°C and an extension time of 60 s. The DNA template was purified as described above. The transcription reaction contained 0.2 μM DNA template, 40 mM Tris·HCl, pH 8.1, 25 mM $MgCl_2$, 3.5 mM spermidine, 0.01% TritonX-100, 40 mM DTT, 4% PEG 8000, 3 mM NTPs, and 7.5 U/μLT7 RNA polymerase (NEB). The reaction was incubated at 37 °C for 1 hour. The RNA was then purified by ethanol precipitation as follows. 1/10 × volume of 3 M Na-acetate, pH 5.2, and 1.0 μL 15 mg/mL GlycoBlue (Thermo Fisher Scientific) was added to the transcription reaction and mixed. 3.3 × volume of chilled 100% ethanol was then added, then reactions were placed on dry ice for 20 minutes. The reactions were then spun down in a tabletop centrifuge at maximum speed for 30 minutes, then the supernatant was removed and discarded. The pellet was then washed twice by adding 500 μL 70% chilled ethanol, centrifuging for 10 minutes at maximum speed, and finally removing the supernatant. The pellet was then dried for 5 minutes and resuspended in RNase-free water. Approximately 10 μL of loading buffer containing 10 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue in 95% formamide was then added to the reaction. The *Tetrahymena* ribozyme RNA was then PAGE-purified, as follows. A 8% 29:1 acrylamid:bis, 7 M urea polyacrylamide gel was poured and allowed to set overnight. RNA containing loading buffer described above was loaded onto the gel and run at 25 W for 2 h. The RNA was visualized in the gel by brief exposure to a 254-nm UV lamp, held far from the gel to minimize RNA damage[3]. RNA was eluted from the gel in RNase-free water overnight at 4 °C. RNA was then purified with Zymo RNA Clean and Concentrator columns (Zymo Research).

Supplementary Note 2. Ribozyme activity assays.

<u>*Tetrahymena* ribozyme activity assay</u>

To check that the *Tetrahymena* ribozyme was active, we first combined 150 nM RNA and 50 mM Na-HEPES, pH 8 and heated at 90°C for 3 minutes to denature the RNA. The solution was then cooled at room temperature for 10 minutes. 10 mM $MgCl_2$ was added and the solution was then incubated at 50°C for 30 minutes. The solution was then cooled at room temperature for 10 minutes, then 500 μM GTP was added and the solution was incubated at room temperature

for 5 minutes. 750 nM fluorescently labeled substrate (Cy5-CCCUCUAAAAA, purchased from IDT) was then added and the reaction was incubated at room temperature in the dark for 1 hour. All concentrations listed above are final. 2 μL of the reaction was then added to 10 μL of stop solution containing 20 mM EDTA and 90% formamide, to quench the reaction. The reaction was then heated at 90°C for 7 minutes. Immediately, 2 μL of the reaction was then loaded onto a 20% 29:1 acrylamide:bis, 7 M urea, 1 x TBE polyacrylamide gel, which had been pre-run at 10 W for 1 hour. The gel was run for 35 minutes at 15 W, and then imaged immediately with a Bio-Rad VersaDoc MP 4000 using the Cy5 detection setting.

hc16 activity assay

To check that the hc16 ribozyme was active, we performed ligation reactions as previously described[4], except that we used a fluorescently labeled substrate (Cy5-AUCUUACUU). Briefly, the hc16 RNA was incubated at 85 °C for 2 minutes and then 50 °C for 2 minutes, then diluted in buffer containing $MgCl_2$, KCl, spermidine, Na-HEPES, pH 7.5 and then incubated at 50°C for 10 minutes. The fluorescently labeled substrate was added and incubated in the dark at 50°C for 75 minutes. The final concentrations in the reaction were: 0.1 μM RNA, 0.2 μM substrate, 50 mM $MgCl_2$, 200 mM KCl, 5 mM spermidine, and 50 mM Na-HEPES, pH 7.5. To stop the reaction, 10 μL of 120 mM EDTA, 8 M urea solution was added to 10 μL of the reaction. The stopped reaction was then loaded onto a 20% 29:1 acrylamide:bis, 7 M urea, 1 x TBE polyacrylamide gel, which had been pre-run at 10 W for 1 hour. The gel was then run for 35 minutes at 15 W, then imaged immediately as described above.

24-3 activity assay

The RNA template for the 24-3 primer extension assay was prepared through PCR assembly and *in vitro* transcription as described above. A fluorescently labeled RNA primer was purchased from IDT. Primer and template sequences are listed in Supplementary Table 5. The primer extension assay was performed as described previously[5]. Briefly, reactions were conducted with final concentrations of 0.5 μM fluorescently labeled primer RNA, 0.5 μM template RNA, 0.5 μM 24-3 ribozyme, 0.05% Tween 20, 50 mM Tris pH 8.1, 200 mM $MgCl_2$, and 4 mM UTP, GTP, ATP, and CTP. Briefly, the RNA (in RNase free water) was first heated for two minutes at 80°C, then cooled at 17°C for ten minutes. Primer extension was initiated by adding chilled extension buffer (Tween 20, Tris, $MgCl_2$, NTPs) to the final concentrations listed above. The sample was mixed and allowed to extend at 17°C. The reaction was quenched by adding EDTA to a final concentration of 125 mM. Samples were then ethanol precipitated as described above, and resuspended in 70% formamide, diluted with TBE buffer. The samples were then run on a 20% 29:1 acrylamide:bis, 7 M urea, 1 x TBE polyacrylamide gel, then imaged for FAM fluorescence.

Supplementary Note 3. M2-seq experiments.

DNA templates for RNAs containing 5′ and 3′ buffer sequences were prepared by PCR assembly of DNA oligonucleotides designed with Primerize[1] and purchased from IDT (Supplementary Table 5), except for the *Tetrahymena* ribozyme, which was PCR amplified from the pT7L-21 plasmid using the primers listed in Supplementary Table 5. PCR products were checked by agarose gel electrophoresis. Reactions that produced only the desired product were purified with AMPure XP beads. Other reactions were purified by gel extraction using the Qiagen gel extraction kit following the manufacturer's instructions. Mutations were introduced in the templates through error-prone PCR as described previously[6]. Again, reactions were

purified with AMPure XP beads or by gel extraction. RNA transcription, RNA purification, DMS treatment, ethanol precipitation, reverse transcription, cDNA purification, and library preparation for sequencing were performed as described previously[6]. Primers used for reverse transcription are listed in Supplementary Table 5.

M2-seq data analysis was performed with the pipeline available at https://github.com/ribokit/M2seq. Briefly, multiplexed FASTQ sequence files were demultiplexed into sample-specific files using NovoBarcode (Novocraft Technologies). Demultiplexed, paired-end files were run through the ShapeMapper analysis pipeline[7]. The ShapeMapper-derived mutation strings were converted to simple binary mutation files with muts_to_simple.py. These .simple files were then converted to .rdat 2D mutational profile files with simple_to_rdat.py. RDAT files were then analyzed by scripts in the Biers package (https://ribokit.github.io/Biers/) as follows. Z-scores incorporating DMS treated and untreated RNA for each sample were generated with output_Zscore_from_rdat.m. Secondary structure prediction was performed with ShapeKnots version 6.1 with the M2-seq data using the rna_structure MATLAB wrapper in the Biers software (MATLAB version 2014b was used here). 100 bootstrapping iterations were performed. The secondary structure predicted by ShapeKnots using all of the M2-seq data, but with base pairs with bootstrap probabilities less than 65% removed, was used for the fully automated auto-DRRAFTER modeling[8].

Supplementary Note 4. Complete description of the auto-DRRAFTER pipeline.

The inputs for the auto-DRRAFTER pipeline are an RNA sequence, secondary structure, and cryo-EM map. First, ideal A-form helices are built for all base paired regions of the structure. Helices are then placed in the density map as follows. The density map is first low-pass filtered to 20 Å, to identify "end nodes" or regions of the map that correspond to hairpin loops or helices formed between the very 5' and 3' ends of an RNA (Extended Data Figure 6b). Points are then placed in the density map using e2segment3d in EMAN2 using the command[9]:

```
e2segment3d.py DENS_MAP --pdbout=ouput.pdb --
process=segment.distance:maxsegsep=18:minsegsep=15:thr=MAP_THR
```

where MAP_THR is approximately the highest threshold such that the whole density map remains connected. A graph is then constructed from these points, where each point becomes a node, and nodes are connected by an edge if they are within 20 Å. For points that are initially connected to just one other node, additional edges are added to other points that are within 27 Å and have a "connecting density score" of greater than half of the average connecting density score for sets of points with connecting density scores greater than 0.02. The connecting density score is defined as the average of the density values at five equally spaced points between two nodes. "End nodes" are then defined as nodes that are connected to only one other point or nodes that are connected to multiple points for which the angle formed between a first neighboring point, the node of interest, and a second neighboring point is less than 0.87 radians, for all possible sets of two neighboring points (Extended Data Figure 6b). End nodes are also defined in the secondary structure by converting the secondary structure to a graph, where the helices, loops, and junctions are represented as nodes with edges between elements containing adjacent nucleotides, then identifying nodes that are connected to only one other node (Extended Data Figure 6a).

End nodes in the secondary structure are then mapped to end nodes in the density map (Extended Data Figure 6c). If there are more than two end nodes in the map, or if there are more than 8 consecutive nucleotides that are not base paired in the input secondary structure, then just a single end node in the map is randomly selected for placement of an end node from the secondary structure. If there are exactly two end nodes identified in the map, then all possible mappings between secondary structure end nodes and density map end nodes are considered. For secondary structure end nodes that are hairpins, the 3D structure of the hairpin is modeled onto the adjacent helix through RNA fragment assembly in Rosetta[10] then the helix is fit into the density map at the location of the end node.

Helices are placed into the density map by alignment to a probe helix, which is first fit into the density map as follows. A six base pair probe helix is aligned to an end node in the map and one of its neighbors (referred to as the neighboring map node below). The position of the helix is then optimized by randomly translating and rotating it, keeping the distance between the C1′ atom of the sixth nucleotide of the first strand of the probe helix and the end node less than 15 Å or the starting distance between these points and keeping the distance between the C1′ atom of the sixth nucleotide of the first strand of the probe helix and the neighboring map node greater than or equal to the initial distance between these points minus 8 Å (Extended Data Figure 6b). Each conformation is scored using the Rosetta elec_dens_fast score term[11]. Conformations are accepted if the score of the current conformation is less than the score of the previous conformation. 5000 perturbations are attempted. The entire process is repeated 10 times and the best scoring placement is taken as the final probe helix placement. Helices from the secondary structure are then aligned to the optimized conformation of the probe helix to generate the initial conformations for the auto-DRRAFTER simulations (Extended Data Figure 6c). Alignment is performed using three base pairs of the probe helix, or the number of base pairs in the helix to be fit if it contains fewer than three base pairs.

This process may result in several possible placements of helices within the density map. For each placement, the rest of the RNA is built with RNA fragment assembly in Rosetta, keeping the placed helix fixed throughout the run[10]. The low-resolution and all-atom Rosetta score functions were augmented with the elec_dens_fast score term to monitor agreement with the density map[12].

Modeling is then performed in several rounds (Extended Data Figure 6c-j). Two thousand models are built for each placement in each round. The average pairwise RMSD (convergence) is then computed for the top ten scoring models across all helix placements. If the convergence RMSD is higher than 10 Å, then another round of modeling is performed. To set up the next round of modeling, for each helix placement, the convergence RMSD and average and minimum density scores for each node in the RNA, defined by the RNA secondary structure as described above, are computed over the top ten scoring models. Here, the average density score is defined as the average of the density values at the positions of each of the atoms within an element. The minimum density score is defined as the minimum of the density values at the positions of each of the atoms within an element. For each node, the average and minimum density scores is calculated in each of the top ten scoring models. The highest average density score is then compared to a threshold defined as 40% of the maximum density value in the density map. The minimum density score for the model with the highest average density score is also compared to this threshold. The convergence value is also compared to a threshold defined by the number of nucleotides in the node and whether it is a helix. The threshold was 7.0 Å for helices containing 3 or fewer base pairs, 10.0 Å for 4 base pair helices, 20.0 Å for 5 base pair helices, and 25.0 Å

for helices containing 6 or more base pairs, and 4.0 Å for elements other than helices. If the convergence for the node is below the appropriate threshold and the average and minimum density scores are above the appropriate threshold, then the 3D conformation of this node is extracted from the model in which it had the highest density score and kept fixed for the subsequent rounds. Conformations are not taken from outlier models, defined as models for which $|\bar{C} - C_i| > 3\sigma$, where C_i is defined for a model *i* as the convergence when that model *i* is removed from the top ten scoring models (i.e. convergence computed over the other nine models), $\bar{C}$ is the average of the C_i values for the top ten scoring models and σ is the standard deviation. Nodes that clash with extracted structures for any other node are removed.

When the convergence of the top ten scoring models across all helix placements drops below 10 Å, the process described above is applied to the overall top ten scoring models (rather than the top ten scoring models for each helix placement) to identify regions of the structure that are well-converged and fit well in the density map, with the following differences (Extended Data Figure 6g-h). All conformations are taken from the top scoring model rather than the model with the best average density score. Additionally, the conformations taken from the models from the previous round are used as initial placements for the next round of modeling, but are allowed to move throughout the run. All other parameters remained the same in the subsequent round of modeling.

If the convergence does not drop below 10 Å and the average change in the convergence between the last three consecutive rounds is less than 1.5 Å after seven rounds, then subsequent rounds of modeling were not performed. Typically, this is an indication that one of the initial assumptions of the modeling, such as the RNA secondary structure or initial helix placements, is incorrect and these inputs should be carefully reexamined.

For the final round of modeling, regions of the models that are well-converged and fit well in the density are again identified using the same procedure as for the previous round of modeling (Extended Data Figure 6i-j). These conformations are kept fixed during the fragment assembly stage of the modeling, but allowed to move along with the rest of the structure during the all-atom refinement stage. This last modeling round is performed in parallel in separate half maps, if available.

Supplementary Note 5. HIV-1 DIS blind modeling.

Five sets of HIV-1 DIS models were built as follows. For the first set of models, the center helix was manually fit into the cryo-EM density map (chain A residues 257-262 and chain B residues 257-262), and the rest of the RNA was built through the fragment assembly procedure described here. We then built a second set of models starting from the best scoring model from the first set of models. During this run, all helices were allowed to move independently as rigid bodies and nucleotides in all junctions were allowed to move. During the final refinement stage, all nucleotides were allowed to move. For the third set of models, we started auto-DRRAFTER runs from chain A residues 244-276 and chain B residues 244-276 from the previously solved NMR model of HIV-1 DIS (PDB ID: 2D1A)[13], after first mutating it to the correct sequence. Throughout the run, helices were allowed to move as rigid bodies, though they were not allowed to dock independently (-dock_each_chunk_per_chain false), and nucleotides in junctions were also allowed to move. All nucleotides could move during the final refinement stage. For the fourth set of models, an initial conformation for chain A residues 248-270 and chain B residues 248-270 was taken from the first set of models. The remaining RNA was built with constraints for possible interactions that could occur within the other four RNA junctions. The constraints

were based on configurations observed in initial stepwise Monte Carlo runs of these junctions without the density map[14]. These constraints were also used to build the final set of models. These models were built using stepwise Monte Carlo with the density map.

All five sets of models were refined with a modified ERRASER protocol[15]. Briefly, the entire structure was first minimized with restraints on the phosphate positions using the default ERRASER kinematics and also using the kinematics from the auto-DRRAFTER runs. The full structure was subsequently minimized again without restraints, then individual nucleotides were rebuilt using the standard ERRASER protocol, followed by a final round of minimization over the entire structure.

Supplementary Note 6. Best-case modeling in experimental density maps.

For the ATP stabilizer with and without AMP and the spinach-TTR-3, the P1 helices were initially fit manually in the density map. Additionally, a model of the tetraloop-tetraloop receptor taken from a crystal structure of the P4-P6 domain of the *Tetrahymena* ribozyme (PDB ID: 1GID) was included as a rigid body[16]. The conformation of the tetraloop-tetraloop receptor was refined during the final round of modeling.

For the *F. nucleatum* glycine riboswitch with and without glycine, the initial conformation for residues 1-20, 25-71, and 80-158 was taken from a previously solved crystal structure (PDB ID: 3P49)[17], which was initially fit into the density map. During the runs, helices were allowed to dock independently as rigid bodies, and nucleotides in junctions were allowed to remodel.

For the *Tetrahymena* ribozyme, the initial conformations for nucleotides 98-168, 174-209, 211-234, 240-258, 260-268, 270-276, 305-321, 324-325, 327-331, 405-406 were taken from a previously solved crystal structure (PBD ID: 1X8W)[18], which was initially fit in the density map.

For the SAM-IV riboswitch with and without SAM, P5 (residues 87-91 and 99-102) was manually placed in the density map. Additionally, the structure of the SAM binding pocket (nucleotides 3, 7, 42, 63, 64, and 78) was taken from a previously solved crystal structure of the SAM-I riboswitch (PDB ID: 2GIS)[19]. For models built into the map solved without SAM, these nucleotides were allowed to remodel after the initial round of modeling. For models built into the maps solved with and without SAM, these nucleotides were allowed to move during the final all-atom refinement during the final round of modeling.

For the *V. cholerae* glycine riboswitch with and without glycine, the initial conformation of residues 141-155, 162-187, and 193-224 was taken from a previously solved crystal structure (PDB ID: 3OWI) which was initially fit into the density map[20]. The conformations of these nucleotides were refined during the final round of modeling.

For hc16 and the hc16 product, we built a preliminary model by first fitting P6a, P6b, P5b, the tetraloop-tetraloop receptor (residues 95-100, 167-173, and 192-197), and the part of P1 formed between the substrate and the ribozyme into the density map for the first conformation of the hc16 product. The rest of the RNA was then built with auto-DRRAFTER while keeping these regions fixed. The helices and the tetraloop-tetraloop receptor (residues 95-100, 167-173, and 192-197) from the best scoring preliminary model were used as an initial conformation for auto-DRRAFTER runs for hc16 and the first conformation of the hc16 product. These elements were allowed to move as rigid bodies throughout the runs and were allowed to move during the final round of refinement. For the second conformation of the hc16 product, the complete best scoring preliminary model was used as an initial conformation for auto-DRRAFTER runs. Regions that were not in helices were allowed to remodel. We built models for hc16 and both conformations of the hc16 product with four possible secondary structures in the regions of uncertainty (dashed

lines, Extended Data Figure 9): P10-ext and P11, P10-ext and alt-P11, P12 and P11, P12 and alt-P11. Models with alt-P11 and either P10-ext or P12 were built as described above. Only a final round of modeling was performed for models with P11, starting from the second to final round of models built with alt-P11 and either P10-ext or P12, but with the region around P11/alt-P11 completely rebuilt. The final set of models for these systems contains the top scoring models for each of these possible secondary structures.

For eterna3D-JR_1, we did not have any additional information beyond the M2-seq-based secondary structure and cryo-EM map, so we did not build a best-case set of models.

Ligands were not included in the automated or best-case modeling. The final round of modeling, including refinement into half maps, was performed for all best-case models.

Supplementary Note 7. hc16 mutate-map-rescue experiments

For each base pair tested, we probed two single mutants that would disrupt the putative base pair and a double mutant that should restore base pairing (for a G-C base pair, the three mutations were G→C, C→G, and G→C and C→G; for an A-U base pair, the three mutations were A→U, U→A, and A→U and U→A; for a G-U base pair, the three mutations were G→C, U→G, and G→C and U→G), expect for mutations testing P11 and alt-P11. The specific mutations tested for P11 and alt-P11 were: C9A; G114U; C9A and G114U; G10U; C113A; G10U and C113A; G10U; C115A; G10U and C115A; C11A; G114U; C11A and G114U; C9U and G10U; C9U, G10U, C113A, and G114A; G10U and C11U; G114A and C170A; G10U, C11U, G114A, and C170A. Chemical mapping was performed as described previously[21]. Briefly, RNA was prepared as described above. Primers for PCR assembly of the DNA templates were designed with Primerize and are listed in Supplementary Table 5[22]. Chemical mapping was performed in 50 mM Na-HEPES pH 8.0, 10 mM $MgCl_2$ with 1.06 mg/mL 1M7 in DMSO (all concentrations are final). Reactions were quenched and purified as described previously[21]. cDNA was generated with SuperScript III. Reverse transcription reactions were performed at 48 °C for 40 minutes. cDNA purification was performed as described previously[21]. cDNA was combined with GeneScan 350 ROX dye Size Standard (Life Technologies, 401735) to provide an internal control. Capillary electrophoresis was performed at ELIM Biopharmaceuticals, Inc. with an ABI 3130 or ABI 3730 machine.

Supplementary Results

Screening RNA molecules with native gels

As the first step of the Ribosolve pipeline, RNA molecules were screened on a native gel (Fig. 1a). All of the molecules in our benchmark set ran as sharp bands in our standard buffer conditions (Methods), with the exceptions of the 24-3 ribozyme and downstream peptide riboswitch (Extended Data Figure 1). We envision that this screening step could be used to optimize folding conditions or identify molecules that may be poorly suited to structural characterization with cryo-EM. Our results indicate that it is necessary but not sufficient for RNA molecules to pass this check.

Assessing auto-DRRAFTER accuracy with simulated density maps

To assess the accuracy of auto-DRRAFTER, we benchmarked the method using a set of eight RNA molecules with previously solved crystal structures. We used auto-DRRAFTER to build models into 10 Å resolution simulated density maps. The final models closely resembled the crystal structures, with the best RMSD accuracy of the top ten scoring auto-DRRAFTER

models ranging from 2.3 – 10.0 Å (Supplementary Table 2, Extended Data Figure 10a-h). We also used auto-DRRAFTER to build models for one of these RNA molecules (THF riboswitch) into the 2.9 Å crystallographic density map. The models were slightly more accurate than those built into the simulated density map (RMSD = 3.6 Å for crystallographic density map, RMSD = 4.0 for simulated density map).

Previously, we showed that DRRAFTER model accuracy for RNA coordinates built into cryo-EM maps of large RNPs could be reliably predicted from modeling convergence, which is defined as the average pairwise RMSD over the top ten scoring models[23]. We confirmed that there is also a strong correlation between auto-DRRAFTER modeling convergence and average model RMSD accuracy ($r^2 = 0.95$, two-tailed $p = 9\times10^{-29}$, N=45), suggesting that auto-DRRAFTER model accuracy can also be reliably predicted (Fig. 1c). We also confirmed that convergence correlates with accuracy for individual residues ($r^2 = 0.88$, two-tailed $p < 0.001$, N = 6357), though there are several outliers for residues that have very low convergence, suggesting that per-residue convergence values should be interpreted with caution (Supplementary Figure 2a). Model accuracy is also correlated with real-space correlation coefficients (CC) for the map versus model (full model $r^2 = 0.76$, two-tailed $p = 4\times10^{-15}$, N = 45; per residue $r^2 = 0.48$, two-tailed $p < 0.001$, N = 6357), suggesting that CC can provide an additional check of final auto-DRRAFTER model accuracy (Supplementary Figure 2b, c).

Using convergence and map vs. model CC to assess Ribosolve model accuracy

To assess the accuracy and quality of our models, we looked at the overall and per-residue convergence and CC. Based on the results from our auto-DRRAFTER benchmark on simulated maps, we expect these values to correlate with model accuracy (Fig. 1c, Supplementary Figure 2). We first confirmed that this was the case for the *F. nucleatum* glycine riboswitch, for which a crystal structure of nearly the entire riboswitch has previously been solved. For a set of generally inaccurate models (mean RMSD to the crystal structure = 14.1 Å), the overall convergence was high (20.2 Å), nearly all residues had convergence values above 10 Å, and many residues had CC values below 0.5, indicating that they did not fit well in the density map (Supplementary Figure 6a-c). In contrast, for a set of generally accurate models (mean RMSD to the crystal structure = 4.9 Å), the convergence RMSD over all residues was low (3.2 Å), nearly all individual residues had convergence values below 10 Å, and all residues had CC values above 0.5 (Supplementary Figure 6d-f).

As an initial check of the accuracy of the Ribosolve models, we confirmed that the per-residue convergence values were generally below 10 Å and the per-residue and overall CC values were above 0.5 for all automated and best-case models that converged (overall convergence < 10 Å, Supplementary Figure 7, Supplementary Figure 8). Residues that do not fit these criteria are listed in Supplementary Table 7.

Combining M2-seq secondary structure determination with auto-DRRAFTER

Auto-DRRAFTER requires RNA secondary structure as an input. For the tests using simulated maps and for most of our Ribosolve benchmark set, the secondary structures were known or hypothesized based on phylogenetic covariance analysis. Secondary structures were not known for the synthetic and *in vitro* evolved RNAs for which only single sequences were available in the literature, precluding covariance analysis. To elucidate the unknown secondary structures and to confirm hypothesized secondary structures, we used a multidimensional chemical mapping technique called mutate-and-map read out by next-generation sequencing

(M2-seq)[6]. M2-seq has previously been shown to be particularly accurate in revealing specific base-pairing interactions, and can now be carried out quickly and in a single pot[6]. The M2-seq-based secondary structures for all RNAs in our benchmark set agreed well with previous models based on sequence covariation, previously solved crystal structures, or designs, where available (Extended Data Figure 7, Extended Data Figure 8, Extended Data Figure 9). These M2-seq-based secondary structures were then used as inputs along with our cryo-EM maps for the auto-DRRAFTER pipeline.

Fully automated and best-case model building for eleven RNAs

For each of the eleven RNAs for which we were able to obtain cryo-EM maps with visible major and/or minor grooves, we built fully automated Ribosolve models without any user intervention as well as "best-case" Ribosolve models using additional information such as placement of sub-structures within the density maps, previously solved crystal structures of parts of the RNA molecules, and modifications to the M2-seq-based secondary structures based on sequence covariation and previously solved crystal structures (see Methods, Extended Data Figure 8). The best-case models for all systems converged well (predicted RMSD accuracy < 6.3 Å, Supplementary Figure 7, Table 1) and the automated models for all systems except the *Tetrahymena* ribozyme, hc16, the hc16 product, and the SAM-IV riboswitch with SAM converged well (predicted RMSD accuracy < 4.8 Å, Supplementary Figure 8, Supplementary Table 6). The M2-seq-based secondary structures used for the auto-DRRAFTER modeling for the SAM-IV riboswitch with SAM and the *Tetrahymena* ribozyme contained base pairs that were incompatible with sequence covariation for the SAM-IV riboswitch and previously solved crystal structures for the *Tetrahymena* ribozyme (Extended Data Figure 8a, j). The majority of these base pairs were automatically flagged as uncertain by the M2-seq analysis (bootstrap probabilities less than 90%). For the best-case modeling, these secondary structures were modified to resolve these inconsistencies, which may explain why the best-case models for these systems converged while the automated models did not (Extended Data Figure 8a, j). We also needed to modify the M2-seq-based secondary structures for hc16 and the hc16 product to be able to build models that converged and fit in the density maps. Again, this required removing base pairs with bootstrap probabilities less than 90%. Further details are described below.

Model building and prospective compensatory mutagenesis for hc16

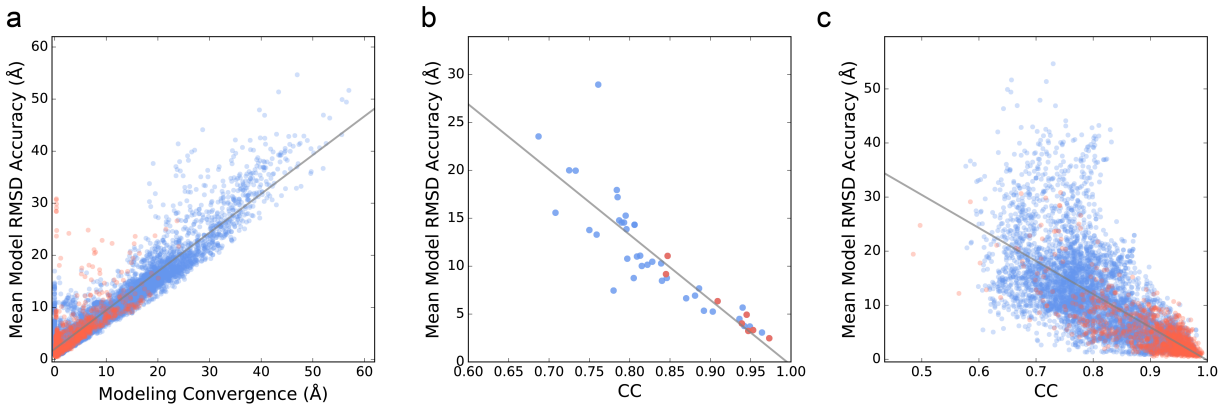
When the automated models for hc16 and the hc16 product initially did not converge, we hypothesized that like for the SAM-IV riboswitch with SAM and the *Tetrahymena* ribozyme, the secondary structure required refinement. Notably, hc16 was evolved *in vitro* from a random library that contained the P4-P6 domain of the *Tetrahymena* ribozyme as a constant scaffold region[4], but the M2-seq-based secondary structure was inconsistent with the presence of this structural element. In particular, the M2-seq data suggested that the P4 stem from the *Tetrahymena* ribozyme was replaced by an alternative helix, which we will refer to as alt-P4 (Extended Data Figure 8l, m). To resolve whether P4 or alt-P4 was formed in hc16 and the hc16 product, we performed mutate-map-rescue experiments[21]. In addition to P4 and alt-P4, we also tested P7 for which we could not see a clear signal in the M2-seq data and a possible alternative, alt-P7. Our results support the formation of the alt-P4 helix observed in the M2-seq data as well as P7, and do not provide evidence for the formation of P4 or alt-P7 (Supplementary Figure 4).

New sets of auto-DRRAFTER models with an updated secondary structure containing alt-P4, P7, and additional previously hypothesized secondary structure elements that were not

present in the M2-seq-based secondary structure (Extended Data Figure 8m) still did not converge and did not fit well in the density maps. Again, we reexamined the secondary structure and, based on the discrepancies between our models and the density maps and regions marked uncertain by M2-seq secondary structure analysis, we hypothesized several modifications to the secondary structure: an extension of the alt-P4 helix that would replace P5, possible additional helices (P10, P12), and possible two base pair pseudoknots (P11/alt-P11) (Extended Data Figure 8n). Additional mutate-map-rescue experiments support the replacement of P5 with alt-P4, and provide some evidence for P10, but do not clearly discriminate between P11 and alt-P11 and P10-ext and P12 (all data deposited in the RMDB). We therefore built auto-DRRAFTER models with each possible secondary structure. The final ensemble of auto-DRRAFTER models, while convergent in global architecture, reflects these residual uncertainties by including ensemble members with different secondary structures.

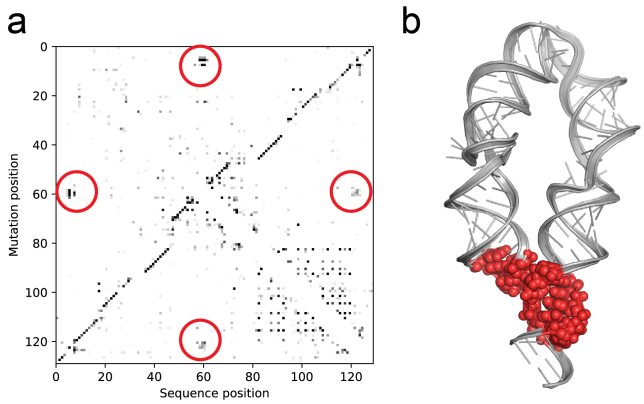
Supplementary Figure 1.

Prototype Eterna3D interface. JavaScript-based web interface used by Eterna participants to rationally design RNA segments that connect the two segments of the tetraloop/tetraloop-receptor contact. Players were able to utilize both two-way junction motifs and helices from a list on the left to build a path from the tetraloop to the tetraloop receptor.



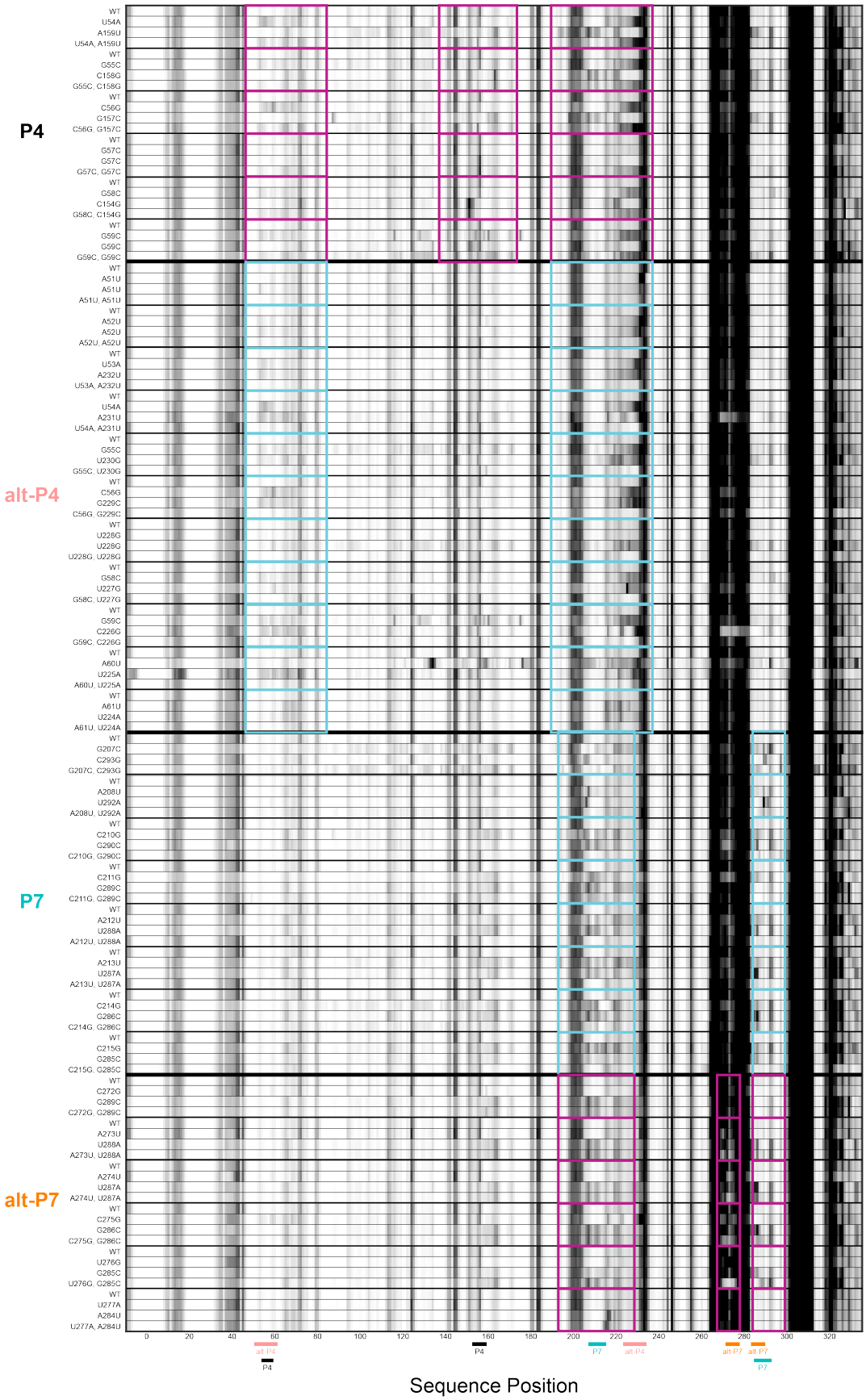
Supplementary Figure 2.

Accuracy estimation from data available at the time of modeling. RMSD accuracy versus (a) convergence or (b and c) real-space CC for the top ten scoring auto-DRRAFTER models built into simulated density maps. (b) RMSDs and CC or convergence values calculated over the full models and (a and c) per residue. Points for models from the final round of auto-DRRAFTER modeling are colored red. Points from all other rounds of modeling are colored blue. Best-fit lines are colored gray and given by (a) y = 0.75x + 2.0, (b) y = -68.1x + 67.7, and (c) y = -60.2x + 61.0.



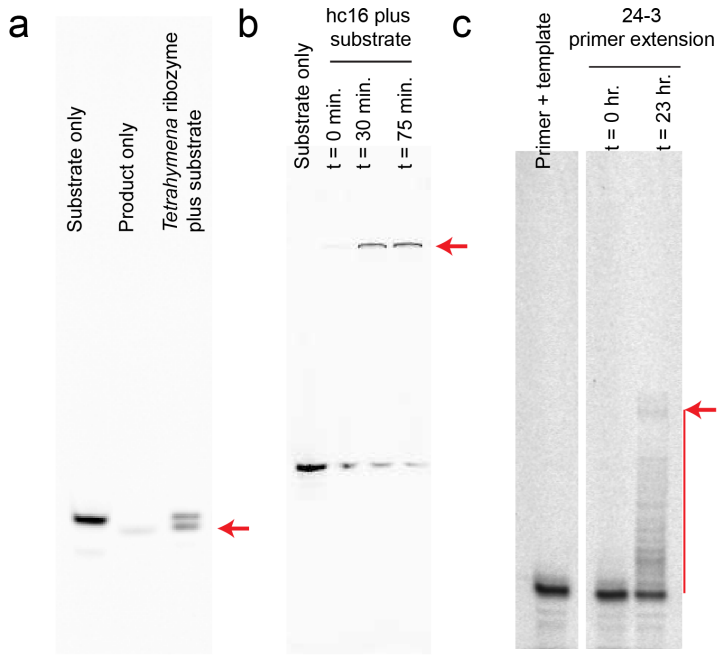
Supplementary Figure 3.

Tertiary contact validates ATP-TTR-3 Ribosolve model. (a) M2-seq Z-score plots with putative tertiary contacts marked with red circles for the ATP-TTR-3 without AMP (n = 712801 sequences). (b) Ribosolve model with nucleotides corresponding to circled regions in (a) shown as red spheres for the ATP-TTR-3 without AMP.


P4
alt-P4
P7
alt-P7
WT
U54A
A159U
U54A, A159U
G55C
C158G
G55C, C158G
C56G
G157C
C56G, G157C
G57C
G57C, G57C
G58C
C154G
G58C, C154G
G59C
G59C, G59C
A51U
A51U, A51U
A52U
A52U, A52U
U53A
A232U
U53A, A232U
A231U
U54A, A231U
U230G
G55C, U230G
C56G
G229C
C56G, G229C
U228G
U228G, U228G
G58C
U227G
G58C, U227G
G59C
C226G
G59C, C226G
A60U
U225A
A60U, U225A
A61U
U224A
A61U, U224A
G207C
C293G
G207C, C293G
A208U
U292A
A208U, U292A
C210G
G290C
C210G, G290C
C211G
G289C
C211G, G289C
A212U
U288A
A212U, U288A
A213U
U287A
A213U, U287A
C214G
G286C
C214G, G286C
C215G
G285C
C215G, G285C
C272G
G289C
C272G, G289C
A273U
U288A
A273U, U288A
A274U
U287A
A274U, U287A
C275G
G286C
C275G, G286C
U276G
G285C
U276G, G285C
U277A
A284U
U277A, A284U
0
20
40
60
80
100
120
140
160
180
200
220
240
260
280
300
320
alt-P4
P4
P4
P7
alt-P4
alt-P7
alt-P7
P7
Sequence Position

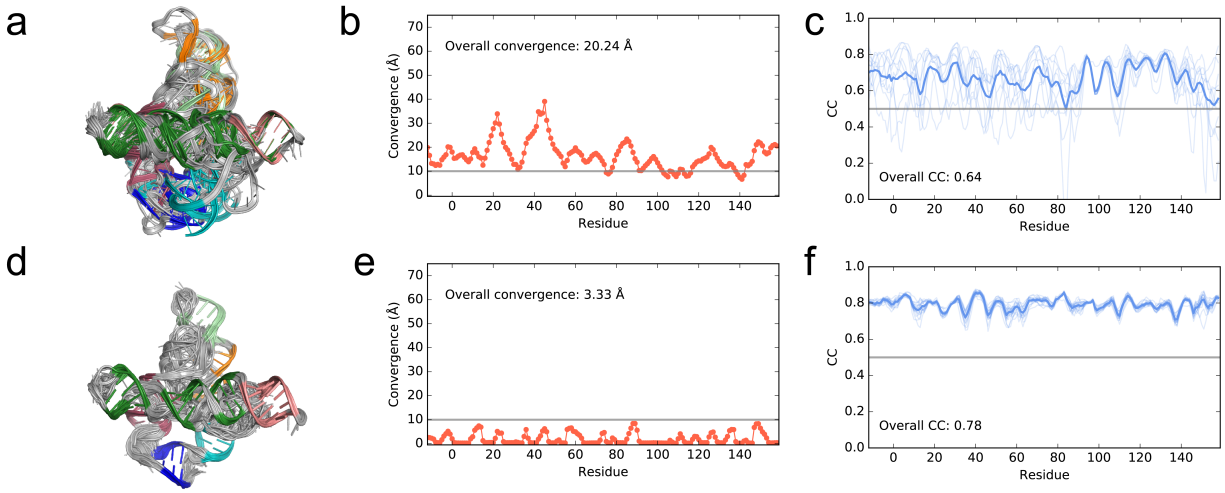
Supplementary Figure 4.

Mutate-map-rescue data testing P4, alt-P4, P7, and alt-P7 in the hc16-product. Boxes denote regions where perturbations are observed. Cyan boxes indicate regions where perturbations are observed in the single mutants and rescue is clearly observed for the double mutants. Magenta boxes indicate regions where clear rescue is not observed.



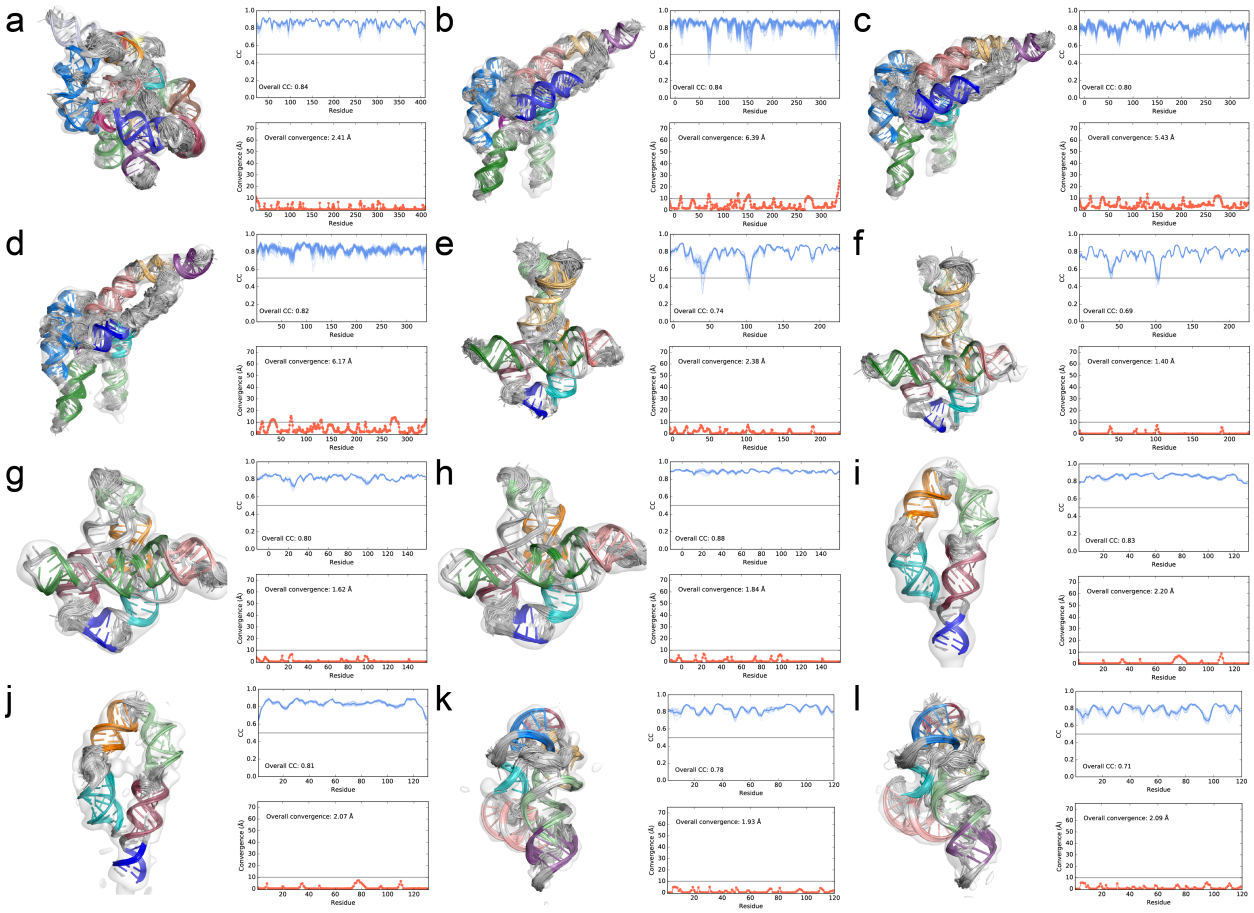
Supplementary Figure 5.

Ribozyme activity assays. (a) *Tetrahymena* ribozyme substrate cleavage (polyacrylamide gel, imaging Cy5, see Methods for details). (b) hc16 ligation reaction (polyacrylamide gel, imaging Cy5, see Methods for details). (c) 24-3 primer extension (polyacrylamide gel, imaging FAM, see Methods for details). Red arrows mark product(s) of ribozyme reactions. Activity assays were performed once per ribozyme.



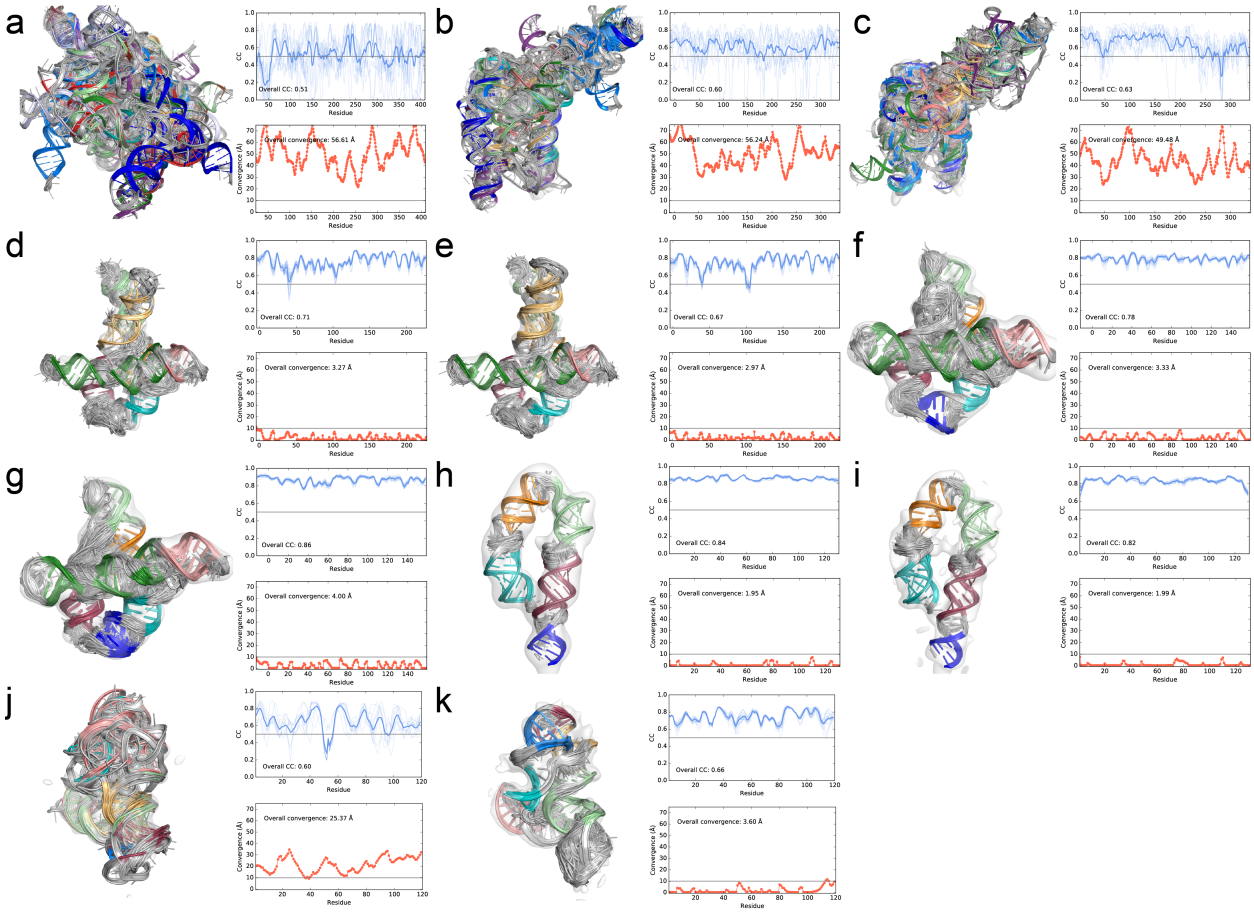
Supplementary Figure 6.

(a) Poorly converged and (d) well-converged auto-DRRAFTER models for the *F. nucleatum* glycine riboswitch with glycine. Helical regions are depicted with bright colors. Non-helical regions are colored black. (b) and (e) Modeling convergence per residue for models in (a) and (d), respectively. (c) and (f) Real-space correlation coefficients (CC) per residue for the top ten scoring auto-DRRAFTER models (thin, light blue lines) for models in (a) and (d), respectively (n = 10 models). Average CC values shown as thick blue lines (n = 10 models).



Supplementary Figure 7.

Best-case auto-DRRAFTER models built into experimental density maps. (left) Top ten scoring automated auto-DRRAFTER models built into each half map shown within the full density map and (top right) the CC values between these models and the opposite half maps (i.e. models built into half map 1 are checked against half map 2; n = 20 models (n = 80 models for b, c, and d)). CC values for each model are shown as light blue lines, the average over all top scoring models is plotted as a thick blue line. The overall CC values reported are averaged over the top ten scoring models for each half map. (Bottom right) auto-DRRAFTER modeling convergence computed between models refined into separate half maps. (a) *Tetrahymena* ribozyme, (b) hc16 product conformation 1, (c) hc16 product conformation 2 (d) hc16, (e) *V. cholerae* glycine riboswitch with glycine, (f) *V. cholerae* glycine riboswitch, (g) *F. nucleatum* glycine riboswitch with glycine, (h) *F. nucleatum* glycine riboswitch, (i) ATP-TTR-3 with AMP, (j) ATP-TTR-3, (k) SAM-IV riboswitch with SAM, and (l) SAM-IV riboswitch. All models are colored as in Fig. 3.



Supplementary Figure 8.

Automated auto-DRRAFTER models built into experimental density maps. (left) Top ten scoring automated auto-DRRAFTER models built into each half map shown within the full density map and (top right) the CC values between these models and the opposite half maps (i.e. models built into half map 1 are checked against half map 2). CC values for each model are shown as light blue lines, the average over all top scoring models is plotted as a thick blue line (n = 20 models). The overall CC values reported are averaged over the top ten scoring models for each half map. (Bottom right) auto-DRRAFTER modeling convergence computed between models refined into separate half maps. (a) *Tetrahymena* ribozyme, (b) hc16 product, (c) hc16, (d) *V. cholerae* glycine riboswitch with glycine, (e) *V. cholerae* glycine riboswitch, (f) *F. nucleatum* glycine riboswitch with glycine, (g) *F. nucleatum* glycine riboswitch, (h) ATP-TTR-3 with AMP, (i) ATP-TTR-3, (j) SAM-IV riboswitch with SAM, and (k) SAM-IV riboswitch. All models are colored as in Fig. 3.

Supplementary Table 1. Cryo-EM data collection and processing statistics.

	Tetra-hymena ribozyme EMD-21840 PDB-6WLS	hc16 product EMD-21835 PDB-6WLN	hc16 EMD-21836 PDB-6WLO	ScaRNA6	*V. cholerae* glycine riboswitch with glycine EMD-21842 PDB-6WLU	*V. cholerae* glycine riboswitch (apo) EMD-21841 PDB-6WLT	RB1 5' UTR	24-3 ribozyme	*F. nucleatum* glycine riboswitch with glycine EMD-21834 PDB-6WLM	*F. nucleatum* glycine riboswitch (apo) EMD-21833 PDB-6WLL	U1 snRNA	Eterna3D-JR_1	Spinach-TTR-3	ATP-TTR-3 with AMP EMD-21831 PDB-6WLJ	ATP-TTR-3 (apo) EMD-21832 PDB-6WLK	SAM IV riboswitch (with SAM) EMD-21839 PDB-6WLR	SAM IV riboswitch (apo) EMD-21838 PDB-6WLQ	Down-stream peptide ribo-switch with gluta-mine
Magni-fication	130k	130k	130k	130k	215k	130k	130k	130k	130k	130k	130k	130k	130k	130k	130k	130k	130k	130k
Voltage (kV)	200	200	200	300	300	300	200	300	200	200	200	200	200	200	200	300	300	200
Electron exposure (e–/Å^2)	5	5.7	5.7	9	7.5	10.6	9	7.6	5	5	9	9.2	9.2	5	5	7.6	7.6	5
Defocus range (μm)	-0.4 – -1.2	-0.4 – -1.2	-0.4 – -1.2	-1.5 – -3.5	-1.2 – -2.8	-1.2 – -2.8	-1.5 – -3.5	-1.5 – -3	-0.4 – -1.2	-0.4 – -1.2	-1.5 – -3.5	-1.5 – -3.5	-1.5 - - 3.5	-0.4 – -1.2	-0.4 – -1.2	-1.5 – -3.5	-1.5 – -3.5	-0.4 – -1.2
Pixel size (Å)	1.07	1.07	1.07	1.06	0.65	1.06	1.07	1.06	1.07	1.07	1.07	1.07	1.07	1.07	1.07	1.06	1.06	1.07
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1	C1
Initial particle images (no.)	233,365	287,851	228,610	36,308	509,061	792,328	51,618	353,447	677,706	380,061	46,983	77,647	169,120	183,317	445,539	849,704	854,888	514,337
Final particle images (no.)	74,621	29,191	21,263	19,363	193,317	230,891	27,143	123,217	35,578	20,269	18,277	51,690	67,474	39,136	71,045	225,303	260,244	32,797
Map resolution (Å)	6.8	10.0	11.0	N/A	5.7	4.8	N/A	N/A	7.4	10.0	N/A	N/A	N/A	9.6	10.0	4.8	4.7	N/A
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	5-11	8-14	8-14	N/A	4-6	4-6	N/A	N/A	7-11	7-11	N/A	N/A	N/A	10-18	10-18	4-6	4-6	N/A
B-factor (Å^2)	553	1055	1002	N/A	393	317	N/A	N/A	415	840	N/A	N/A	N/A	1398	1501	303	238	N/A
RNA length (nt)	388	349	338	265	231	231	189	180	171	171	168	135	130	130	130	119	119	65
Microscope	Talos Arctica	Talos Arctica	Talos Arctica	Titan Krios	Titan Krios	Titan Krios	Talos Arctica	Titan Krios	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica	Talos Arctica	Titan Krios	Titan Krios	Talos Arctica
Phase plate	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	Yes
No. micrographs	752	1414	1012	860	4700	6600	1050	2260	1311	959	880	600	1100	934	977	3200	2860	983

Supplementary Table 2.
Accuracy of auto-DRRAFTER models built into simulated density maps.

System	No. nucleotides	Convergence (Å)	Mean RMSD (Å)	Min RMSD (Å)	CC	No. modeling rounds	Final round modeling performed?
THF riboswitch (3SUX)	101	3.03	4.02	3.81	0.94	3	Yes
Bacterial SRP Alu domain (4WFL)	105	2.10	2.48	2.25	0.97	4	Yes
FMN riboswitch (3F2Q)	112	12.12	9.17	5.92	0.85	9	Yes
SAM-I riboswitch (4KQY)	119	2.97	3.26	2.98	0.95	4	Yes
c-di-AMP riboswitch (4QK8)	122	2.93	3.33	3.00	0.95	4	Yes
Tetrahymena ribozyme P4-P6 domain (1GID)	158	6.36	6.34	5.24	0.91	7	Yes
Lysine riboswitch (3DIL)	174	4.21	4.93	4.51	0.95	5	Yes
Lariat capping ribozyme (4P8Z)	188	6.92	11.07	9.99	0.85	8	Yes

Supplementary Table 3.
Convergence and CC for best-case models.

System	No. nts	Map resolution (Å)	Convergence (Å)	Convergence half 1	Convergence half 2	CC full map	CC half map 1	CC half map 2	No. modeling rounds
Tetrahymena ribozyme	388	6.8	2.4	2.4	2.4	0.80	0.83	0.83	3
hc16 product	349	10.0	6.4	6.3	6.6	0.79	0.84	0.84	3
hc16 product conformation 2	349	10.0	5.4	5.4	5.5	0.79	0.80	0.80	3
hc16	338	11.0	6.2	6.2	6.1	0.81	0.82	0.82	3
V. cholerae glycine riboswitch with glycine	231	5.7	2.4	2.4	2.4	0.73	0.74	0.74	5
V. cholerae glycine riboswitch apo	231	4.8	1.4	1.4	1.3	0.67	0.69	0.69	4
F. nucleatum glycine riboswitch with glycine	171	7.4	1.6	1.5	1.6	0.80	0.80	0.80	3
F. nucleatum glycine riboswitch apo	171	10.0	1.8	1.9	1.7	0.89	0.88	0.88	3
ATP-TTR-3 with AMP	130	9.6	2.2	1.9	2.4	0.79	0.83	0.84	3
ATP-TTR-3 apo	130	10.0	2.1	2.2	1.7	0.62	0.81	0.81	3
SAM-IV riboswitch with SAM	119	4.8	1.9	1.9	1.9	0.76	0.77	0.78	3
SAM-IV riboswitch apo	119	4.7	2.1	2.1	2.0	0.71	0.71	0.70	3

Supplementary Table 4.
Accuracy of Ribosolve models built into experimental density maps.

System	How models were built	Residues	Convergence (Å)	Predicted RMSD Accuracy (Å)	RMSD Accuracy (Å)
F. nucleatum glycine riboswitch with glycine aptamer 1	Fully *de novo*	1-20, 25-71	3.2	4.3	3.8
F. nucleatum glycine riboswitch with glycine aptamer 2	Fully *de novo*	80-158	3.4	4.5	5.1
F. nucleatum glycine riboswitch with glycine other regions (not previously crystallized)	Fully *de novo*	-12-0, 21-24, 72-79	3.3	4.4	N/A
V. cholerae glycine riboswitch with glycine aptamer 2	Fully *de novo*	141-155, 162-187, 193-224	2.3	3.8	3.3
V. cholerae glycine riboswitch with glycine, regions not previously crystallized	Fully *de novo*	-4-140, 156-161, 188-192, 225-226	3.6	4.6	N/A
V. cholerae glycine riboswitch without glycine aptamer 2	Fully *de novo*	141-155, 162-187, 193-224	2.4	3.9	3.6
V. cholerae glycine riboswitch with glycine, regions not previously crystallized	Fully *de novo*	-4-140, 156-161, 188-192, 225-226	3.2	4.3	N/A
SAM-IV riboswitch without SAM	Fully *de novo*	All	3.6	4.6	4.9
HIV-1 DIS	Fully *de novo*	All	4.3	5.0	4.0
THF riboswitch[1]	Fully *de novo*	1-7, 10-101	2.5	3.9	3.6
Bacillus subtilis T-box–tRNA complex	Fully *de novo*	T-box: 13-42, 44-86, 88-114, 118-133, 140-167, 172-182 tRNA: 1-75	4.4	5.1	5.3
Bacillus subtilis T-box–tRNA complex	Fully *de novo*	T-box: 43, 87, 115-116, 134-139, 168-171	5.1	5.5	N/A
Tetrahymena ribozyme core	Piece built from previous crystal structure	98-168, 174-209, 211-234, 240-258, 260-268, 270-276,	1.9	3.6	3.5

		305-321, 324-325, 327-331, 405-406			
Tetrahymena ribozyme peripheral elements	Piece built *de novo*	22-97, 277-304, 332-404, 407-409	2.7	4.1	N/A
SAM-IV riboswitch without SAM	Piece built from previous crystal structure	All	2.1	3.6	3.8
SAM-IV riboswitch with SAM, regions outside the binding pocket	Piece built *de novo*	1-2, 4-6, 8-41, 43-62, 65-77, 79-119	1.9	3.6	2.4
SAM-IV riboswitch with SAM, binding pocket	Piece built from previous crystal structure	3, 7, 42, 63-64, 78	2.3	3.8	2.1
F. nucleatum glycine riboswitch with glycine aptamer 1	Piece built from previous crystal structure	1-20, 25-71	0.8	2.9	1.8
F. nucleatum glycine riboswitch with glycine aptamer 2	Piece built from previous crystal structure	80-158	1.3	3.2	2.7
F. nucleatum glycine riboswitch with glycine, regions not previously crystallized	Piece built *de novo*	-12-0, 21-24, 72-79	3.2	4.4	N/A
V. cholerae glycine riboswitch without glycine aptamer 2	Piece built from previous crystal structure	141-155, 162-187, 193-224	0.1	2.5	0.2
V. cholerae glycine riboswitch without glycine, regions not previously crystallized	Piece built *de novo*	-4-140, 156-161, 188-192, 225-226	2.9	4.2	N/A
V. cholerae glycine riboswitch without glycine aptamer 2	Piece built from previous crystal structure	141-155, 162-187, 193-224	0.1	2.4	0.4
V. cholerae glycine riboswitch without glycine, regions not previously crystallized	Piece built *de novo*	-4-140, 156-161, 188-192, 225-226	1.7	3.4	N/A

[1]Models built into the 2.9 Å crystallographic density.

Supplementary Table 5. (separate Excel-formatted file)
DNA and RNA sequences used in this study.

Supplementary Table 6.

Assessing automated models: convergence, CC, and RMSD to best-case models.

System	No. nts	Map resolution (Å)	Convergence	Convergence half-1 (Å)	Convergence half-2 (Å)	Mean RMSD (Å)	Min. RMSD (Å)	CC full	CC half-1	CC half-2	No. modeling rounds	Final round modeling performed?
Tetrahymena ribozyme	388	6.8	56.6	N/A	N/A	38.6	38.5	0.51	-	-	7	No
hc16 product	349	10.	56.2	N/A	N/A	43.8	43.5	0.60	-	-	7	No
hc16	338	11.	49.5	N/A	N/A	48.0	47.6	0.63	-	-	7	No
V. cholerae glycine riboswitch with glycine	231	5.7	3.3	3.3	3.2	4.4	4.3	0.69	0.71	0.71	12	Yes
V. cholerae glycine riboswitch	231	4.8	3.0	3.1	2.8	4.0	3.9	0.66	0.67	0.67	5	Yes
F. nucleatum glycine riboswitch with glycine	171	7.4	3.3	3.2	3.4	4.3	4.2	0.76	0.77	0.78	4	Yes
F. nucleatum glycine riboswitch apo	171	10.	4.0	4.3	3.6	8.3	8.2	0.85	0.86	0.86	4	Yes
ATP-TTR-3 with AMP	130	9.6	2.0	1.9	2.0	3.8	3.5	0.8	0.84	0.85	3	Yes
ATP-TTR-3 apo	130	10.	2.0	1.9	1.9	4.2	4.0	0.63	0.82	0.81	4	Yes
SAM-IV riboswitch with SAM	119	4.8	25.4	N/A	N/A	21.5	21.4	0.6	-	-	7	No
SAM-IV riboswitch apo	119	4.7	3.6	3.5	3.5	5.6	5.6	0.66	0.67	0.66	5	Yes

Supplementary Table 7.
Nucleotides with convergence above 10 Å or CC below 0.5.

System	Nucleotides with convergence > 10 Å	Nucleotides with CC < 0.5
Tetrahymena ribozyme best-case	22	none
Tetrahymena ribozyme automated	22-409	22, 23, 24, 25, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 102, 103, 104, 105, 110, 111, 112, 113, 114, 115, 116, 149, 150, 151, 152, 153, 154, 169, 170, 171, 172, 173, 175, 176, 177, 178, 184, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 288, 289, 290, 305, 306, 307, 308, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 349, 350, 351, 360, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 381, 382, 383, 384, 385, 386, 387, 388, 391, 392, 393, 394, 395, 396, 397
hc16 product best-case	12, 13, 14, 69, 70, 71, 72, 129, 130, 131, 147, 148, 149, 150, 151, 152, 153, 154, 203, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 332, 333, 334, 335, 336, 337, 338	none
hc16 product automated	-10-338	170, 171, 181, 182, 183, 184, 271, 272, 274
hc16 product conformation 2 best-case	12, 13, 36, 37, 38, 39, 40, 70, 71, 116, 129, 130, 131, 203, 204, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280	none
hc16 best-case	12, 13, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 68, 69, 70, 71, 72, 128, 129, 130, 217, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 336, 337, 338	none
hc16 automated	1-338	46, 47, 241, 242, 243, 244, 245, 246, 247, 248, 264, 265, 266, 271, 278, 279, 280, 281, 282, 283, 284, 285, 286, 338
V. cholerae glycine riboswitch with glycine best-case	none	104
V. cholerae glycine riboswitch with glycine automated	-4	none
V. cholerae glycine riboswitch apo best-case	none	103, 104
V. cholerae glycine riboswitch apo automated	none	101, 102, 103, 104, 105
F. nucleatum glycine riboswitch with glycine best-case	none	none

F. nucleatum glycine riboswitch with glycine automated	none	none
F. nucleatum glycine riboswitch apo best-case	none	none
F. nucleatum glycine riboswitch apo automated	none	none
ATP-TTR-3 with AMP best-case	none	none
ATP-TTR-3 with AMP automated	none	none
ATP-TTR-3 apo best-case	none	none
ATP-TTR-3 apo automated	none	none
SAM-IV riboswitch with SAM best-case	none	none
SAM-IV riboswitch with SAM automated	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119	50, 51, 52, 53, 54, 55, 56, 96, 97
SAM-IV riboswitch apo best-case	none	none
SAM-IV riboswitch apo automated	113, 114, 115	none

Supplementary Video 1. (separate file)
RNA structures determined by the Ribosolve pipeline.

References

1. Tian, S., Yesselman, J. D., Cordero, P. & Das, R. Primerize: automated primer assembly for transcribing non-coding RNA domains. *Nucleic Acids Research* **43**, W522-W526, doi:10.1093/nar/gkv538 (2015).
2. Zaug, A. J., Grosshans, C. A. & Cech, T. R. Sequence-Specific Endoribonuclease Activity of the Tetrahymena Ribozyme - Enhanced Cleavage of Certain Oligonucleotide Substrates That Form Mismatched Ribozyme Substrate Complexes. *Biochemistry* **27**, 8924-8931, doi:10.1021/bi00425a008 (1988).
3. Kladwang, W., Hum, J. & Das, R. Ultraviolet Shadowing of RNA Can Cause Significant Chemical Damage in Seconds. *Sci Rep-Uk* **2**, doi:10.1038/srep00517 (2012).
4. Jaeger, L., Wright, M. C. & Joyce, G. F. A complex ligase ribozyme evolved in vitro from a group I ribozyme domain. *Proceedings of the National Academy of Sciences* **96**, 14712-14717 (1999).
5. Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proceedings of the National Academy of Sciences* **113**, 9786-9791 (2016).
6. Cheng, C. Y., Kladwang, W., Yesselman, J. D. & Das, R. RNA structure inference through chemical mapping after accidental or intentional mutations. *P Natl Acad Sci USA* **114**, 9876-9881, doi:10.1073/pnas.1619897114 (2017).
7. Smola, M. J., Rice, G. M., Busan, S., Siegfried, N. A. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* **10**, 1643-1669, doi:10.1038/nprot.2015.103 (2015).
8. Hajdin, C. E. *et al.* Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *P Natl Acad Sci USA* **110**, 5498-5503, doi:10.1073/pnas.1219988110 (2013).
9. Tang, G. *et al.* EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).
10. Das, R., Karanicolas, J. & Baker, D. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* **7**, 291-294, doi:10.1038/nmeth.1433 (2010).
11. DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W. & Baker, D. Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. *J Mol Biol* **392**, 181-190, doi:10.1016/j.jmb.2009.07.008 (2009).
12. DiMaio, F. & Chiu, W. Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps. *Resolution Revolution: Recent Advances in Cryoem* **579**, 255-276, doi:10.1016/bs.mie.2016.06.003 (2016).
13. Baba, S. *et al.* Solution RNA structures of the HIV-1 dimerization initiation site in the kissing-loop and extended-duplex dimers. *J Biochem* **138**, 583-592, doi:10.1093/jb/mvi158 (2005).
14. Watkins, A. M. *et al.* Blind prediction of noncanonical RNA structure at atomic accuracy. *Sci Adv* **4**, doi:10.1126/sciadv.aar5316 (2018).
15. Chou, F. C., Sripakdeevong, P., Dibrov, S. M., Hermann, T. & Das, R. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nature Methods* **10**, 74-U105, doi:10.1038/Nmeth.2262 (2013).
16. Cate, J. H. *et al.* Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273**, 1678-1685, doi:10.1126/science.273.5282.1678 (1996).

17. Butler, E. B., Xiong, Y., Wang, J. & Strobel, S. A. Structural basis of cooperative ligand binding by the glycine riboswitch. *Chem Biol* **18**, 293-298, doi:10.1016/j.chembiol.2011.01.013 (2011).
18. Guo, F., Gooding, A. R. & Cech, T. R. Structure of the Tetrahymena ribozyme: Base triple sandwich and metal ion at the active site. *Molecular Cell* **16**, 351-362, doi:10.1016/S1097-2765(04)00592-1 (2004).
19. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**, 1172-1175, doi:10.1038/nature04819 (2006).
20. Huang, L. L., Serganov, A. & Patel, D. J. Structural Insights into Ligand Recognition by a Sensing Domain of the Cooperative Glycine Riboswitch. *Molecular Cell* **40**, 774-786, doi:10.1016/j.molcel.2010.11.026 (2010).
21. Tian, S. Q., Cordero, P., Kladwang, W. & Das, R. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *Rna* **20**, 1815-1826, doi:10.1261/rna.044321.114 (2014).
22. Tian, S. & Das, R. Primerize-2D: automated primer design for RNA multidimensional chemical mapping. *Bioinformatics* **33**, 1405-1406, doi:10.1093/bioinformatics/btw814 (2017).
23. Kappel, K. *et al.* De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nature methods* **15**, 947 (2018).