

RNA

De novo design of RNA pseudoknots with deep learning

Jill Townley^{1,2,†}, Wipapat Kladwang^{3,4,†}, David Baker^{5,6}, Hamish M. Blair⁷, Christian A. Choe⁸, Gina El Negr⁹, Andrew Favor⁵, Eli Fisker¹, Daniel B. Haack^{10,11}, Shujun He¹², Jason Hingey¹¹, Po-Ssu Huang^{8,9}, Rui Huang³, Chaitanya K. Joshi^{13,†}, Thomas Karagianes^{1,2,3}, Andrew Kubaney⁵, Pietro Liò¹³, Adamo Mancino¹⁴, Jonathan Romano^{1,2,4}, Boris Rudolfs¹⁰, Nicholas Spellmon¹⁴, Navtej Toor¹⁰, Jigyasa Verma³, Vivian Wu^{3,4}, Zhiheng Yu¹⁴, Eterna Participants^{1,§}, Rhiju Das^{3,4,9,*}

RNA design has been hindered by the limited accuracy of three-dimensional (3D) structure prediction. In this study, we show that intricate RNA structures can be generated with current deep learning tools through accurate de novo design of pseudoknot secondary structures. In an Eterna competition involving 57 pseudoknots, generative artificial intelligence (AI) methods matched experienced human designers in solving most blind challenges, evaluated by single nucleotide-resolution chemical mapping, compensatory mutagenesis, and cryo-electron microscopy. AI-generated molecules with accurate secondary structures formed well-ordered 3D folds stabilized by noncanonical tertiary interactions not modeled during design. Success was guided by an RNet foundation model trained on prior chemical mapping data, suggesting that some difficult RNA design tasks may be tractable without first solving RNA 3D structure prediction.

Complex RNA structures underlie fundamental biological processes, including translation and viral replication as well as emerging biotechnologies and medicines. Building on pioneering efforts in design by human experts (1), recent years have seen steady improvements in the ability to design RNA structures ranging from mRNA vaccines (2) to ribozymes (3, 4) to elaborate RNA origami (5), but these efforts have remained limited to design of simple RNA secondary structures, redesign of previously known RNA structures, or the assembly of a limited repertoire of well-characterized tertiary motifs. Deep learning-based design has been successful in designing RNA molecules with structures related to previously characterized RNA but with highly distinct sequences (6–8). However, de novo design of previously unknown structures, although routine for proteins (9, 10), has not been achieved for RNA. Progress in RNA design has been limited by the poor accuracy of RNA secondary and tertiary structure prediction methods, particularly for synthetic molecules (11–13), for which sequences and structures of evolutionarily related molecules are unavailable.

Pseudoknots are complex secondary structures in which nucleotides in the loops of stems form base pairs with loops outside their stems (14). An increasing number of natural pseudoknots have been shown to function as ribozymes, riboswitches, and ribosomal frameshifting signals through the formation of complex three-dimensional (3D) topologies with well-defined stems (15, 16). However, new pseudoknots have only been automatically designed for a special class of

RNA origami molecules (5, 17) and have not been achieved with general tertiary motif assembly (18). In 2024, human participants of the OpenKnot challenges on Eterna (19) began to achieve accurate design of novel pseudoknots, as assessed and refined through experiments on millions of RNA molecules whose secondary structures were evaluated by their nucleotide-by-nucleotide reactivity to selective 2' hydroxyl acylation analyzed by primer extension (SHAPE) (20–22). At the same time, de novo RNA design algorithms based on deep learning were proposed but remained experimentally untested in their application to previously unseen secondary and tertiary structures (23, 24). These developments motivated the extension of the OpenKnot challenge to explicitly compare the design performance of these new AI algorithms to experienced Eterna participants (tables S1 and S2).

Benchmarking AI design against experienced humans

Round 1 of the extended OpenKnot challenge invited six AI methods to submit designs for SHAPE experiments, in which single-stranded and conformationally flexible nucleotides are preferentially acylated at their 2'-hydroxyl groups and base-paired nucleotides are typically protected and exhibit low reactivity (Fig. 1, A to E). The 17 Round 1 secondary structures were drawn from 11 RNAs whose experimental 3D structures had been deposited in the Protein Databank (PDB) as well as six synthetic pseudoknots proposed by Eterna participants based on earlier partially successful attempts at design (tables S3 to S4 and fig. S1A). For several of the 11 PDB-derived targets, the natural sequences were known to form alternative states. For example, in the Round 1 target W03, a long stem in an RNA sensing cyclic diguanosine monophosphate in *Clostridium acetobutylicum* [c-di-GMP-II riboswitch (25)] appeared mostly unformed in the absence of the small-molecule ligand, as assessed by SHAPE (Fig. 1D). Therefore, even for cases with natural sequences, the challenge was to discover sequences that might stabilize the target pseudoknot for the RNA in solution without stabilizing partners. Up to 20 submissions were accepted from each AI method as well as from each Eterna participant; some methods that required input 3D backbones (table S2) were unable to make designs for the six targets without PDB structures. Designs throughout the OpenKnot challenge were probed in the absence of small-molecule ligands so that pseudoknot formation would reflect the design rather than ligand stabilization; subsequent SHAPE measurements with ligands supported this choice (fig. S2 and supplementary text).

We observed a wide range of performance, with many Eterna designs achieving experimental SHAPE profiles with the patterns of high and low reactivity matching the target profile of unpaired and paired nucleotides (Fig. 1, B and E). The agreement of each design's experimental SHAPE profile with its target secondary structure was quantified using the OpenKnot score, which computes the percentage of nucleotides whose SHAPE reactivity matches expectation (materials and methods); scores range from 0 to 100. A cluster of Eterna designs surpassed a cutoff of success of 90% (Fig. 1, F and G; fig. S3A; and data S1); this cutoff was set on the basis of the inherent experimental uncertainty in the measurements (19) and an error rate of ~10% of SHAPE methods in assigning secondary structure in RNAs with known structures (26, 27). Eterna participants responsible for these high-scoring designs used a combination of in-game tools, intuition, and diverse design strategies (supplementary text). Most AI methods (Rosetta, Rosetta-LoRes, 3DRNA, gRNAde, MPNN-RFdiff; table S2) did not pass the score cutoff and underperformed Eterna participants in Round 1 (Fig. 1, E to H, and supplementary text). One exception was MPNN-fixbb,

¹Eterna Massive Open Laboratory, USA. ²Frameshifter, San Francisco, CA, USA. ³Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. ⁴Howard Hughes Medical Institute, Stanford, CA, USA. ⁵Department of Biochemistry, University of Washington, Seattle, WA, USA. ⁶Howard Hughes Medical Institute, Seattle, WA, USA. ⁷Department of Mathematics, Stanford University, Stanford, CA, USA. ⁸Department of Bioengineering, Stanford University, Stanford, CA, USA. ⁹Biophysics Program, Stanford University, Stanford, CA, USA. ¹⁰Department of Biochemistry and Molecular Biophysics University of California San Diego, La Jolla, CA, USA. ¹¹A-Form Solutions, Inc., San Diego, CA, USA. ¹²Texas A&M University, College Station, TX, USA. ¹³Department of Computer Science and Technology, University of Cambridge, UK. ¹⁴Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. *Corresponding author. Email: rhiju@stanford.edu †These authors contributed equally to this work. ‡Present address: Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA. §Eterna Participants are listed in the supplementary materials.

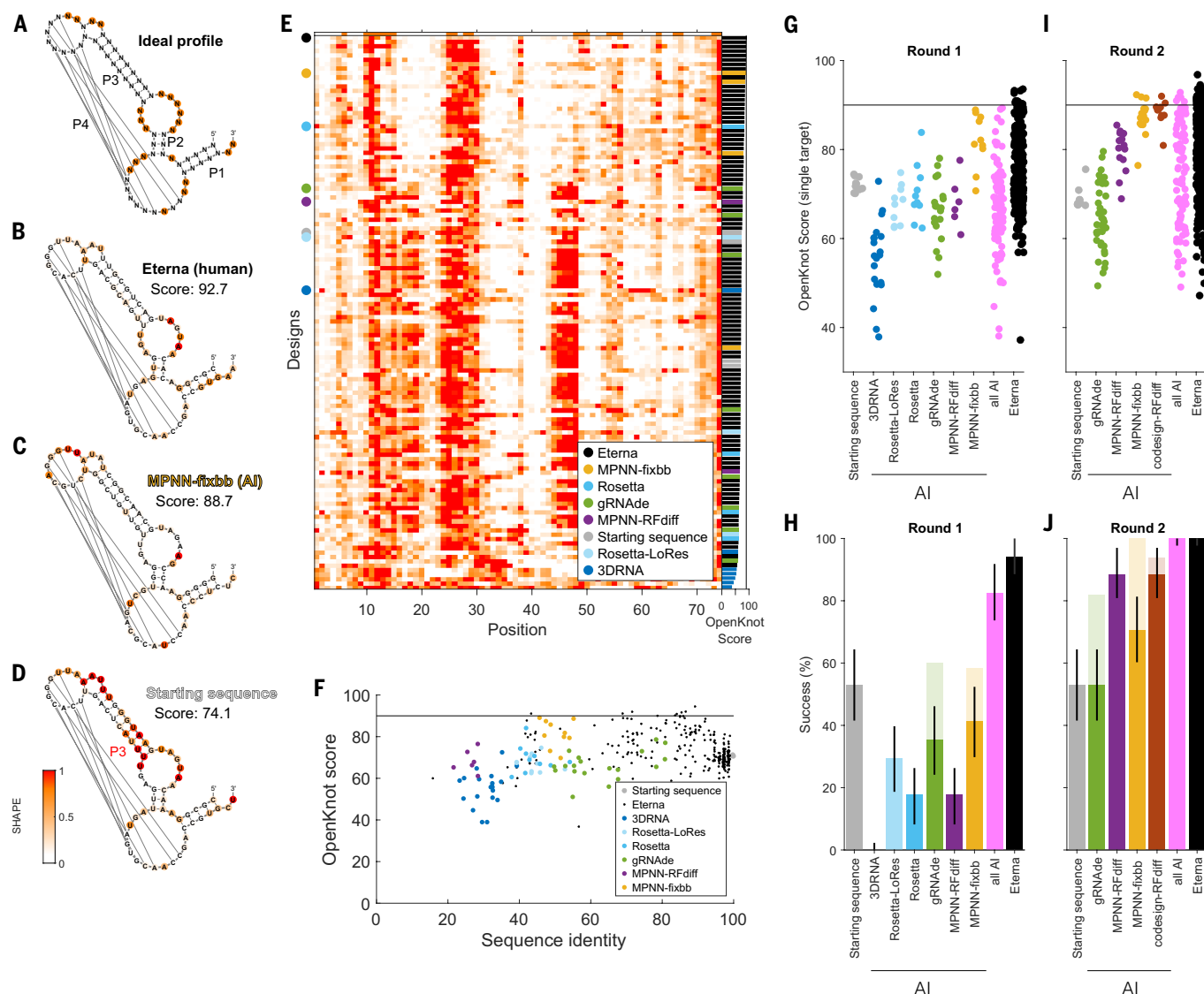


Fig. 1. Performance on 17 starting targets. (A to D) Target secondary structure for W03 (c-di-GMP-II riboswitch), with coloring by ideal SHAPE data (A) and experimental SHAPE data for top designs from (B) Eterna human participants, (C) MPNN-fixbb, and (D) wild-type RNA (natural c-di-GMP-II riboswitch aptamer). The red label in (D) marks the P3 stem where SHAPE reactivity indicates flexibility of the RNA backbone, inconsistent with the target structure. (E) SHAPE reactivity of sequences tested for target W03 in Round 1, demonstrating a wide range of agreement with the ideal target profile (top), as parameterized by the SHAPE-based OpenKnot score (bars, far right). (F) Designs with low (<50%) sequence identity to the starting sequence achieved good agreement with target SHAPE profile (OpenKnot score >90). (G to J) Performance summaries across design methods for [(G and I)] W03 OpenKnot scores and [(H and J)] percentage of targets W01 to W17 for which success (defined here as OpenKnot scores >90) was achieved. In (H) and (J), error bars reflect standard errors, and solid and light-colored bars show success rates over all 17 targets and only on targets on which designs were submitted, respectively. Performance increases were apparent from [(G and H)] Round 1 to [(I and J)] Round 2.

a message-passing neural network that designs sequences for a fixed-input backbone [analogous to ProteinMPNN (28, 29)], which improved on average over the W03 starting sequence (Fig. 1, C and G). Many of the top MPNN-fixbb designs, as well as some top-scoring Eterna designs, were quite different from the starting sequence (sequence identity <50%; see Fig. 1F). In addition, all designs with >90% sequence identity to the starting sequence scored under 80%, suggesting the importance of shifting away from the natural sequence.

At the completion of Round 1, a neural network RNet, trained on chemical mapping of 1 million previous RNAs (21), was observed to give SHAPE profiles that gave simulated scores largely reproducing experimental scores, especially for poorly performing designs (fig. S4). This observation suggested that RNet modeling might allow for prospective filtering of designs. The 17 Round 1 targets were then posed again to all designers in a Round 2. Eterna designers—who were given

access to RNet in the interactive interface—and AI design methods that were updated to take into account RNet (gRNAde, MPNN-RFdiff, MPNN-fixbb, and the new codesign-RFdiff; materials and methods) did notably better in this round, e.g., on the W03 c-di-GMP-II riboswitch target (Fig. 1I and fig. S3B). Over all 17 targets, all the tested AI methods and Eterna human participants were able to achieve scores above 90 on at least 80% of the targets that they designed (light colored bars, Fig. 1J), and there was no significant difference between methods. For both Rounds 1 and 2, an alternative Z-score-based assessment modeled on procedures used in the Critical Assessment of Structure Prediction (12, 13) gave similar rankings (fig. S5, A and B).

Generalization to previously unseen pseudoknot targets

The results above indicated that AI designers and Eterna participants could learn from explorations in prior designs to improve from

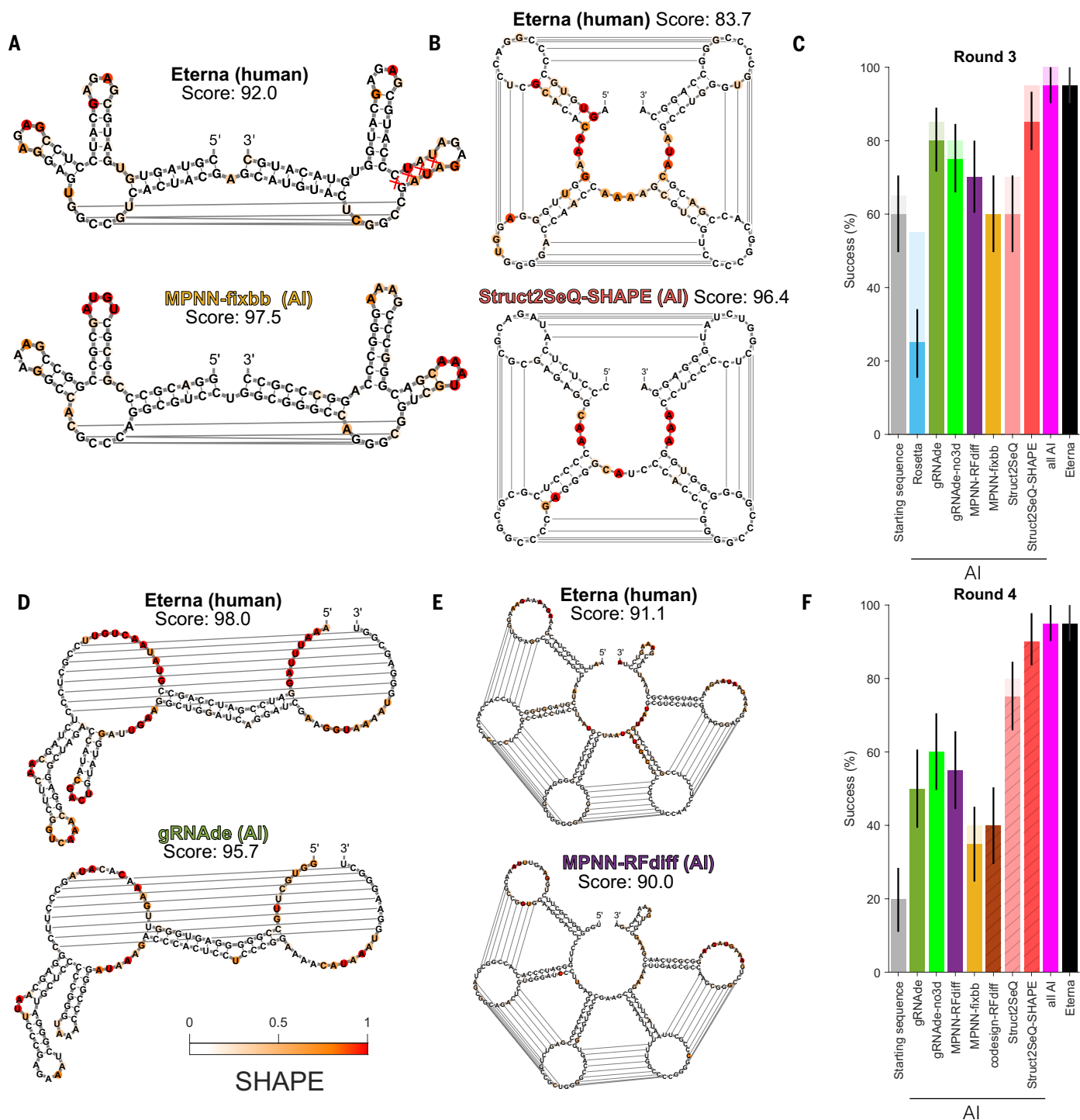


Fig. 2. Performance on 40 new target pseudoknots. (A and B) Designs with highest OpenKnot score for (A) P20 (Kissing Multiloops) and (B) P16 (AK_PK100-3) from Eterna and AI design methods (MPNN-fixbb and Struct2SeQ-SHAPE, respectively). The Struct2SeQ-SHAPE design in (B) is more consistent with an alternative secondary structure (see also figs. S6 and S7). (C) Performance summary across 20 targets of Round 3, with lengths up to 100 nt. (D and E) Designs with highest OpenKnot score for larger Round 4 targets (D) Q08 (Rous Sarcoma Virus) and (E) Q20 (SV_i) from Eterna and AI design methods (gRNAde and MPNN-RFdiff, respectively). (F) Performance summary across 20 targets of Round 4, with lengths between 117 and 240 nt; hatched bars mark methods whose submissions were characterized in an experiment that followed release of initial Round 4 results. In (A), (B), (D), and (E), secondary structures and coloring are for the target designs and experimental SHAPE profiles, respectively. In (C) and (F), solid-colored bars show rates of achieving OpenKnot score >90, after filtering designs for accurate RNet-predicted secondary structures; light-colored bars show rates without the secondary structure filter; and error bars reflect standard errors.

Round 1 to Round 2, on the same set of 17 targets. As a test of the generalization of the resulting methods, we carried out a Round 3 with 20 new pseudoknots as target secondary structures (Fig. 2, A to C; fig. S1C; and table S5). Compared with Rounds 1 and 2, fewer targets (5 versus 11) were drawn from RNAs with experimental 3D structures in the PDB, to mitigate any effects of training set memorization for the AI methods. An additional five target secondary structures were drawn from Pseudobase, an expert-curated database of natural pseudoknots (30). The remaining 10 were curated from secondary structures proposed by Eterna participants (materials and methods). As an additional test of generality, we posed 20 additional targets with longer lengths, up to 240 nucleotides (nts), drawn from similarly diverse sources, in a parallel Round 4 (Fig. 2, D to F; fig. S1C; and table S5). To enable all methods to make designs for targets that were not associated with PDB structures, modeled 3D structures were provided for targets missing experimental structures (materials and methods and table S5).

SHAPE characterization showed strong support for the performance of AI methods and Eterna participants. Each individual deep learning design method outperformed the starting sequences, whose OpenKnot scores were particularly low for the longer targets in Round 4. When grouped together, both AI methods and Eterna methods achieved OpenKnot scores of 90 in 19/20 targets in Round 3 (Fig. 2C) and in Round 4 (Fig. 2F; see also Fig. 2, C and D). Z-score analysis also supported these conclusions (fig. S5, C and D). Only one target was not designable with OpenKnot score >90 by any method or Eterna participant (P16, AK_PK100-3; Fig. 2B and supplementary text).

Besides the P16 target, successes included a wide variety of elaborate pseudoknots proposed by Eterna participants (e.g., Kissing Multiloops,

Fig. 2A, and SV_i, Fig. 2E) as well as a pseudoknot proposed for a natural RNA with unknown 3D structure (Rous Sarcoma Virus frameshift signal; Fig. 2D). The success of the design methods was particularly notable for the targets with long length in Round 4, for which the starting sequences were successful in only 4 of 20 targets (the starting sequence in Fig. 2F). The leading AI methods included a new variant of graph neural network gRNAde that did not take 3D input, as well as a method new to the challenge called Struct2Seq, a deep Q reinforcement learning method that was trained and guided purely by the RNet model (31) (table S2 and materials and methods). In both Rounds 3 and 4, the differences between the best method from the MPNN, gRNAde, and Struct2Seq were generally not significant ($P > 0.05$; exact paired binomial test on discordant pairs, one-sided). The exception was Struct2Seq-SHAPE, which significantly outperformed the other AI methods in Round 4; this method's designs were submitted after release of initial results, however, allowing additional development time (materials and methods).

Compensatory mutagenesis confirms designed base pairs

The results of Rounds 3 and 4 suggested that both experienced human designers as well as AI methods had improved to the point that they could design natural and non-natural pseudoknots without progressive refinement. However, the evaluation of these designs was based on a predetermined but arbitrary cutoff of 90 for the OpenKnot score and the assumption that SHAPE measurements accurately monitor base pairing status. Caveats for SHAPE interpretation have been described (26, 27, 32, 33), and it is possible that nucleotides whose protection from SHAPE appeared consistent with the target secondary structure were instead forming alternative pairings. To test this possibility, we brought to bear an independent approach based on comparative

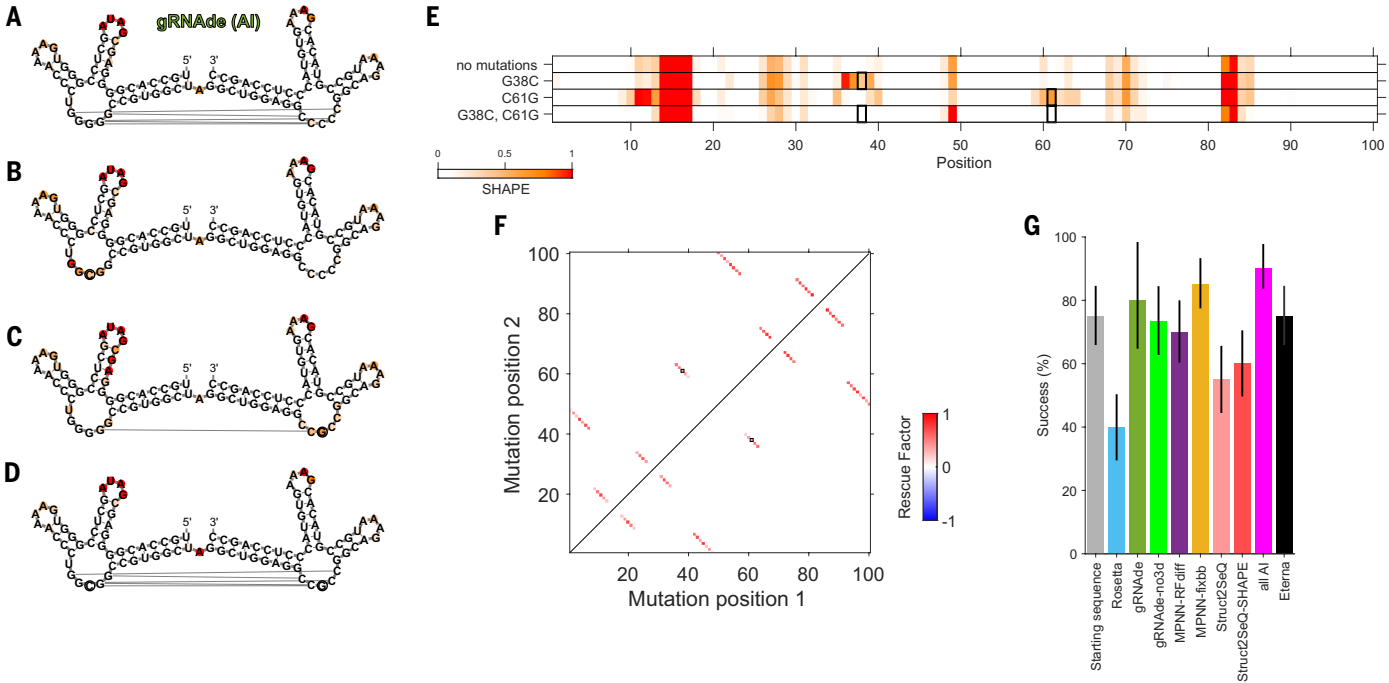


Fig. 3. Compensatory mutagenesis (mutate-map-rescue, M2R-seq) tests accuracy of individual base pairs. (A) SHAPE data for gRNAde design with highest OpenKnot Score for Round 3 target P20 (Kissing Multiloops), colored onto target pseudoknot secondary structure. (B to D) SHAPE profiles for single mutants (B) G38C and (C) C61G showed disruptions near sites of mutations (outlined circles) that were rescued in (D) the compensatory double mutant, which restored a G38-C61 base pair as a C38-G61 base pair. In (A) to (D), to aid visualization, depicted secondary structures are those modeled for the sequences by RNet-SS. (E) Same SHAPE profiles as in (A) to (D) stacked on each other to aid visualization of profile restoration upon compensatory mutagenesis. (F) Experimental rescue factor values for all target base pairs for the design in (A). Black rectangles in (E) and (F) outline bases 38 and 61 and 38-61 base pair mutated in (B) to (D). (G) Performance summaries across all design methods, based on the fraction of the 20 Round 3 targets in which M2R-seq verified the formation of at least 80% of target stems.

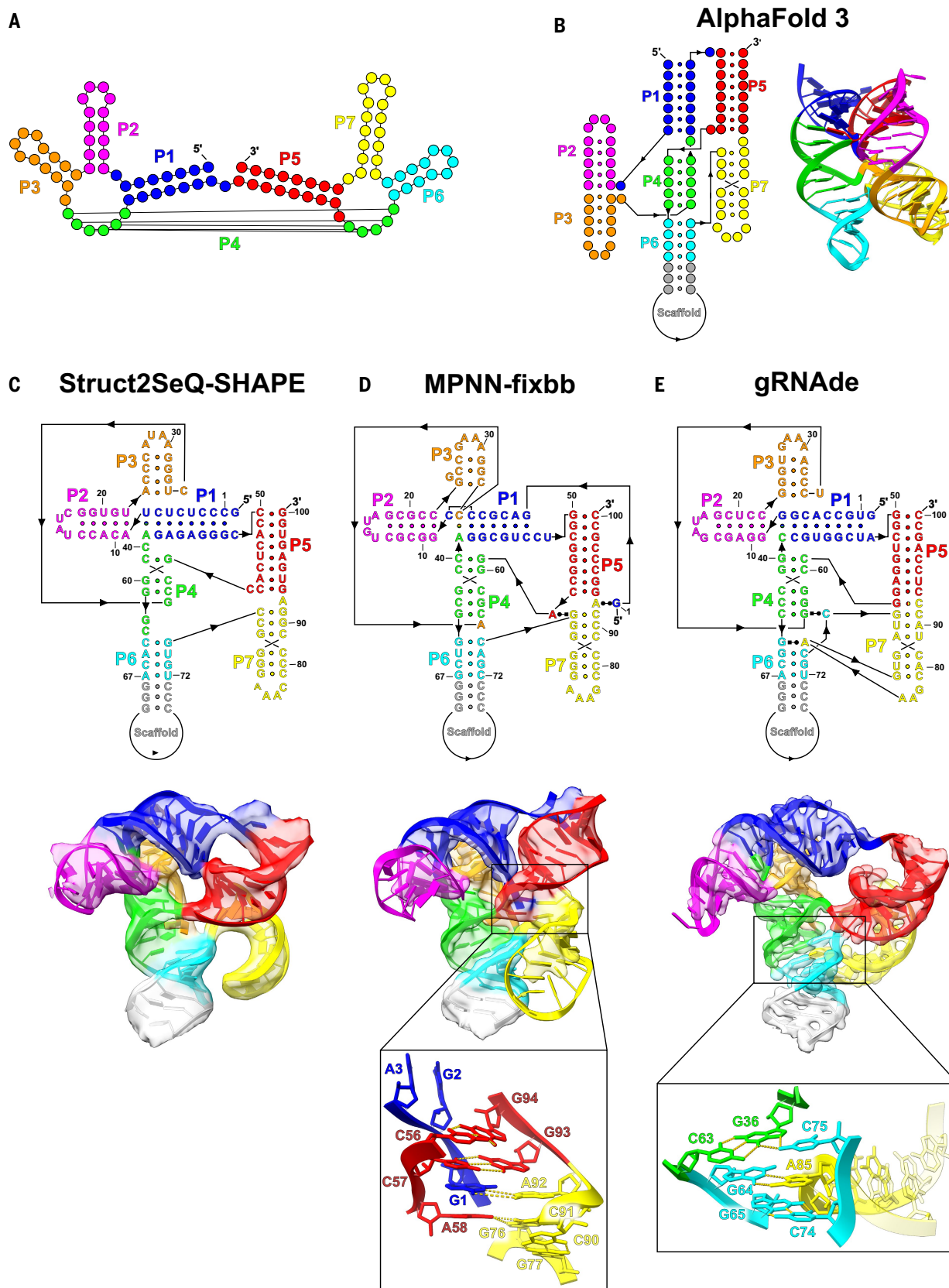


Fig. 4. Cryo-EM of AI-designed pseudoknotted RNA. (A) Secondary structure of Kissing Multiloops (target P20 in Round 3), colored by stem. (B) AlphaFold 3 3D-informed secondary structure and predicted model, colored by stem. (C to E) For the tested designs from (C) Struct2SeQ-SHAPE, (D) MPNN-fixbb, and (E) gRNAde, cryo-EM-derived secondary structures (top) show high accuracy in recovering the target pseudoknot secondary structure (A), while cryo-EM maps (unsharpened) and fitted coordinates, colored by stem, (bottom) highlight distinct 3D topologies from AlphaFold 3 prediction and noncanonical interactions [insets under (D) and (E)].

mutagenesis, called mutate-map-rescue readout by high throughput sequencing (M2R-seq) (33, 34), illustrated for a gRNade design of P20 (Kissing Multiloops) from Round 3 in Fig. 3, A to F. If two nucleotides form a base pair in a putative RNA structure, mutation of either nucleotide should disrupt the pair and produce SHAPE perturbations near the site of mutation (Fig. 3, A to C; fig. S6 shows detailed data for representative designs); reactivity changes at the unmutated partner can also occur but are not always evident because the unmutated nucleotide may remain stacked or form alternative interactions (33, 35, 36). However, if the two nucleotides are mutated at the same time so as to flip the original base pair (e.g., C-G to G-C), the RNA's structure should be “rescued” and the SHAPE profile should return to that of the unmutated RNA (compare Fig. 3D to Fig. 3A, and bottom and top rows of Fig. 3E).

Figure 3F summarizes “rescue factors” (33) for every target base pair in this P20 design. The data showed strong evidence for each stem being formed (conversely, fig. S7 shows a negative example, target P16). M2R-seq data were acquired across all 20 Round 3 targets for the top design from Eterna and each AI method (more than 10,000 sequences, single and double mutants); and we evaluated stem-wise recovery [materials and methods and (32, 33)]. Across 17 of 20 targets, at least one design achieved 100% recovery of all target stems by our M2R-seq rescue-factor criterion. For 19 of 20 targets, there was at least one AI or Eterna design with at least 80% M2R-seq stem-wise recovery (data S1). With this criterion, AI methods were again competitive with Eterna (Fig. 3G; 90 and 75% designs successful, respectively; difference not significant). The differences between AI methods were not significant, except that each of the deep learning methods significantly outperformed Rosetta ($P < 0.05$; exact paired binomial test on discordant pairs, one-sided). Overall, the results from compensatory mutagenesis analysis supported the picture from the OpenKnot score analysis based on SHAPE data: Like experienced human RNA designers, AI methods were able to design complex pseudoknot secondary structures, including novel targets, with consistently high accuracy.

Cryo-EM reveals new 3D folds and noncanonical interactions

To learn whether these previously unseen pseudoknots might correspond to new 3D structures, we subjected several designs to cryo-electron microscopy (cryo-EM). Out of the non-natural secondary structures in Round 3, we prioritized designs for the Eterna-proposed target P20 (Kissing Multiloops; Fig. 2A; Fig. 3, A to F; and Fig. 4A) owing to the short lengths of single-stranded linkers, which we reasoned would restrict the RNA's 3D conformations sufficiently to enable their structural characterization through cryo-EM. Modeling in AlphaFold 3, trRosettaRNA, and other RNA 3D structure prediction algorithms also suggested that the P20 designs would form well-defined folds, although 3D modeling confidence was poor (Fig. 4B). To increase the visibility of the designs in cryo-EM micrographs and to break any pseudosymmetry that might preclude high-resolution refinement, we embedded the top P20 designs inside a recently developed circularly permuted group II intron scaffold (37).

Four designs (the starting sequence, Rosetta, Eterna, and Struct2Seq) appeared aggregated or unfolded by size-exclusion chromatography during purification and were therefore not subjected to microscopy. These were also the designs that exhibited M2R-seq stem-wise recoveries <80%, suggesting that base-pairing stability was important for retaining folded samples. The remaining four were imaged successfully (Fig. 4, C to E; figs. S8 to S11; table S6; and movies S1 and S2). The MPNN-RFdiff design, which gave an M2R-seq stem-wise recovery of 80%, formed a dimer under conditions used for cryo-EM (materials and methods); all seven target stems were present but P4 base pairs were formed between molecules rather than within each molecule (figs. S8 to S11, movie S2, and supplementary text).

The other three designs, from Struct2Seq-SHAPE, MPNN-fixbb, and gRNade, had M2R-seq stem-wise recoveries of 100%. All three resolved

as monomers in cryo-EM, with map resolutions of 5.2, 4.8, and 3.6 Å, respectively, which improved after map masking and sharpening to 4.8, 4.0, and 2.9 Å, respectively. In all three maps, stems P1 to P7 and interconnecting linkers were clearly visible before sharpening (Fig. 4, C to E). These maps enabled the modeling of all coordinates, initially by unbiased manual tracing that was blind to the molecular sequence. The coordinate building confirmed that all designs formed all seven stems of the target secondary structure with high base-pair level accuracy ($F_1 = 0.89, 0.95$, and 0.95 , respectively) and that the three folds were not homologous to prior RNA structures and did not match the AlphaFold 3 models. Detailed inspection of the two highest-resolution designs revealed additional noncanonical pairs, base triplets, and tertiary interactions (Fig. 4, D and E, insets), which did not appear to be explicitly designed by the AI methods (supplementary text).

RNA design without 3D structure prediction

Reliable de novo design of complex RNA pseudoknots is now achievable through deep learning. In the OpenKnot AI challenge, automatic AI methods became competitive with experienced human designers from Eterna in less than 1 year, with both kinds of approaches solving over 95% of 57 pseudoknot design targets, as tested by SHAPE mapping and compensatory rescue experiments involving ~50,000 sequences. Progress in the analogous problem of de novo protein design has been driven by the use of accurate computational methods for 3D protein structure prediction (9, 10, 38), but such tools remain unavailable for RNA (11–13). Instead of relying on 3D structure prediction tools, the MPNN, gRNade, and Struct2Seq frameworks here leveraged a model RNet trained on prior Eterna chemical mapping data that were mainly sensitive to secondary structure. Aside from this shared use of RNet, these three AI frameworks used different variants of deep learning (table S2), and our results were unable to discriminate with statistical confidence which of these AI methods was better than others. Instead, these three frameworks may be best deployed in combination because sequence-space analysis shows that AI methods sample tight, mutually disjoint regions, providing complementary coverage of design space (fig. S14). Further experimental tests of these and other emerging methods (7, 39–41) will be needed to clarify which frameworks will be most appropriate for these and future RNA design tasks.

The AI-generated molecules probed here by cryo-EM displayed intricate noncanonical tertiary interactions, which are critical for sophisticated functions in natural RNA molecules (42–44) but were not predicted a priori. Predictive modeling and design of such high-resolution details will be important for designing RNA catalysts and aptamers and therefore remains an important challenge. Nevertheless, the appearance of noncanonical interactions without explicit design suggests that other problems, such as redesign of machines like the ribosome and RNA polymerase ribozymes (45, 46) or discovery of nonredundant folds to augment RNA structure databases (47), might be immediately accelerated through computational design guided by accurate secondary structure prediction, without also requiring high accuracy in prospectively modeling 3D interactions in atomic detail.

REFERENCES AND NOTES

1. L. Jaeger, A. Chworos, *Curr. Opin. Struct. Biol.* **16**, 531–543 (2006).
2. H. Zhang et al., *Nature* **621**, 396–403 (2023).
3. R. Yamagami, M. Kayedkhordeh, D. H. Mathews, P. C. Bevilacqua, *Nucleic Acids Res.* **47**, 29–42 (2019).
4. D. Szokoli, N. E. Nwosu, L. M. Glatt, H. Mutschler, *ChemBiochem* **26**, e202500356 (2025).
5. C. Geary, G. Grossi, E. K. S. McRae, P. W. K. Rothmund, E. S. Andersen, *Nat. Chem.* **13**, 549–558 (2021).
6. S. Sumi, M. Hamada, H. Saito, *Nat. Methods* **21**, 435–443 (2024).
7. F. Wong et al., *Nat. Comput. Sci.* **4**, 829–839 (2024).
8. A. T. Merchant, S. H. King, E. Nguyen, B. L. Hie, *Nature* **649**, 1–10 (2025).

9. D. Listov, C. A. Goverde, B. E. Correia, S. J. Fleishman, *Nat. Rev. Mol. Cell Biol.* **25**, 639–653 (2024).
10. C. Armer *et al.*, *Proteins* **93**, 2005–2014 (2025).
11. W. Wang *et al.*, *Nat. Commun.* **14**, 7266 (2023).
12. R. Das *et al.*, *Proteins* **91**, 1747–1770 (2023).
13. R. C. Kretsch *et al.*, *Proteins* **94**, 192–217 (2026).
14. E. Westhof, L. Jaeger, *Curr. Opin. Struct. Biol.* **2**, 327–333 (1992).
15. J. Aruda, S. L. Grote, S. Rouskin, *Curr. Opin. Struct. Biol.* **88**, 102912 (2024).
16. Y. Chen *et al.*, *Nat. Chem. Biol.* **17**, 601–607 (2021).
17. E. K. S. McRae *et al.*, *Nat. Nanotechnol.* **18**, 808–817 (2023).
18. J. D. Yesselman *et al.*, *Nat. Nanotechnol.* **14**, 866–873 (2019).
19. J. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2122–2127 (2014).
20. T. Marinus, A. B. Fessler, C. A. Ogle, D. Incarnato, *Nucleic Acids Res.* **49**, e34 (2021).
21. S. He *et al.*, Ribonanza: deep learning of RNA structure through dual crowdsourcing, *bioRxiv* 2024.02.24.581671 [Preprint] (2024). <https://doi.org/10.1101/2024.02.24.581671>.
22. K. E. Deigan, T. W. Li, D. H. Mathews, K. M. Weeks, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 97–102 (2009).
23. C. K. Joshi *et al.*, gRNAde: Geometric Deep Learning for 3D RNA inverse design (2024); <https://openreview.net/forum?id=lvw3UgeVxS>.
24. A. FAVOR *et al.*, De novo design of RNA and nucleoprotein complexes, *bioRxiv* 2025.10.01.679929 [Preprint] (2025). <https://doi.org/10.1101/2025.10.01.679929>.
25. K. D. Smith, C. A. Shanahan, E. L. Moore, A. C. Simon, S. A. Strobel, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 7757–7762 (2011).
26. C. W. Leonard *et al.*, *Biochemistry* **52**, 588–595 (2013).
27. W. Kladwang, C. C. VanLang, P. Cordero, R. Das, *Biochemistry* **50**, 8049–8056 (2011).
28. A. Kubaney *et al.*, RNA sequence design and protein-DNA specificity prediction with NA-MPNN, *bioRxiv* 2025.10.03.679414 [Preprint] (2025). <https://doi.org/10.1101/2025.10.03.679414>.
29. J. Dauparas *et al.*, *Science* **378**, 49–56 (2022).
30. F. H. van Batenburg, A. P. Gulyaev, C. W. Pleij, J. Ng, J. Oliehoek, *Nucleic Acids Res.* **28**, 201–204 (2000).
31. S. He, Q. Sun, Struct2Seq: RNA inverse folding with Deep Q-Learning, *bioRxiv* 2026.01.16.700031 [Preprint]. (2026). <https://doi.org/10.64898/2026.01.16.700031>.
32. E. B. Arnold *et al.*, *Nucleic Acids Res.* **53**, gkaf290 (2025).
33. S. Tian, P. Cordero, W. Kladwang, R. Das, *RNA* **20**, 1815–1826 (2014).
34. S. Tian, W. Kladwang, R. Das, *eLife* **7**, e29602 (2018).
35. W. Kladwang, C. C. VanLang, P. Cordero, R. Das, *Nat. Chem.* **3**, 954–962 (2011).
36. W. Kladwang, P. Cordero, R. Das, *RNA* **17**, 522–534 (2011).
37. D. B. Haack *et al.*, *Nat. Commun.* **16**, 880 (2025).
38. B. Kuhlman *et al.*, *Science* **302**, 1364–1368 (2003).
39. R. Anand *et al.*, *arXiv:2406.13839 [q-bio.BM]* (2025).
40. N. S. C. Merleau, M. Smerlak, *BMC Bioinformatics* **23**, 335 (2022).
41. S. Najeh, K. Zandi, N. Kharma, J. Perreault, *RNA* **29**, 764–776 (2023).
42. Q. Vicens, J. S. Kieft, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2112677119 (2022).
43. R. C. Kretsch *et al.*, *Proteins* **94**, 51–78 (2026).
44. J. H. Cate *et al.*, *Science* **273**, 1678–1685 (1996).
45. Y. Shulgina *et al.*, *Nat. Commun.* **15**, 10627 (2024).
46. C. K. Joshi *et al.*, Generative inverse design of RNA structure and function with gRNAde, *bioRxiv* 2025.11.29.691298 [Preprint] (2025). <https://doi.org/10.1101/2025.11.29.691298>.
47. W. Wang, B. Su, Z. Peng, J. Yang, *Nat. Biotechnol.* **44**, 205–214 (2026).
48. R. Das, eternagame/OpenKnotAIDesignData: v4.5.2 — M2 Single-Mutant SHAPE Data, Zenodo, (2026); <https://doi.org/10.5281/zenodo.20101887>.
49. A. FAVOR, OpenKnot RNA Pseudoknots – RFDpoly 3D Structures, version v1, Zenodo (2026); <https://doi.org/10.5281/zenodo.18666308>.
50. J. Townley *et al.*, OpenKnotAI: Code for RNA library design, SHAPE data processing, OpenKnot scoring, and RNA design methods, version v1, Zenodo (2026); <https://doi.org/10.5281/zenodo.20649966>.

ACKNOWLEDGMENTS

We thank C. Geary (Heidelberg) for advice on design targets, J. Nicol (Eterna) for advice on padding designs for chemical mapping, R. C. Kretsch (Stanford) for advice on target selection for cryo-EM, A. Espeleta for assisting development of 3DRNA, J. Shendure (U. Washington) and lab for sharing primer sequences for library preparation, G. Meissner (Janelia) for project coordination, the AI@HHMI initiative for support, and Nvidia DGX Cloud and NSF NAIIR Pilot (allocation NAIIR240281) for engineering support for Struct2Seq in Round 4. Claude was used during preparation of the revised manuscript for editorial assistance with text and figure captions and with computational analyses supporting figs. S2, S6, and S14; all AI-assisted output was reviewed and verified by the authors. **Funding:** This work was supported by NIH grant R35GM122579 (R.D.), NIH grant R35GM141706 (N.T.), NIH grant R01AI165433 (S.H.), NIH grant R01GM147893 (P.-S.H.), NIH grant U19AI181881 (A.K.), the Howard Hughes Medical Institute (R.D., D.B.), NSF grant 2330652 (R.D.), NSF grant 255097 (N.T.), the W. M. Keck Foundation (N.T.), the A*STAR Singapore National Science Scholarship (C.K.J.), the Qualcomm Innovation Fellowship (C.K.J.), a University of Cambridge Dawn HPC Pioneer Project grant (C.K.J.), a Texas A&M X grant (S.H.), and the Merck Research Laboratories (MRL) Scientific Engagement and Emerging Discovery Science (SEEDS) Program (P.-S.H.). **Author contributions:** J.T., E.F., and R.D. designed Eterna OpenKnot targets, and J.T., J.R., and T.K. coordinated target deployment, design collection, and scoring on Eterna. C.K.J., P.L., A.F., A.K., G.E.N., S.H., D.B., P.-S.H., C.A.C., and Eterna Participants submitted sequences. W.K., H.M.B., V.W., R.H., J.V., and R.D. designed, collected, and analyzed SHAPE experiments. D.B.H., J.H., B.R., N.T., N.S., A.M., and Z.Y. designed and carried out cryo-EM measurements and modeling. R.D. drafted the manuscript with input from all authors. All authors other than J.T., W.K., Eterna Participants, and R.D. are listed in the byline alphabetically. **Competing interests:** The authors declare that they have no competing interests. **Data, code, and materials availability:** SHAPE profiles and modeled 3D structures for Rounds 1–4, M2, and M2R experiments are publicly available (48, 49) and at the RNA Mapping Database (<https://rmdb.stanford.edu>) under the following accession IDs: OK45LIB_2A3_0000 (Round 1), OK6LIB_2A3_0000 (Round 2), OK7ALIB_2A3_0000 (Round 3), and OK7BLIB_2A3_0000 and OK7BLIB_2A3_0001 (Round 4). Cryo-EM structures for designs of P20 (Kissing Multiloops) are available in the PDB under accession IDs 10ZT (gRNAde design, Mol9), 10ZU (MPNN-fixbb design, Mol14), 11EH (Struct2Seq-SHAPE, Mol13), and 11AG (MPNN-RFdiff dimer, Mol11). Cryo-EM maps are available at EMDB under accession IDs EMD-75574 (gRNAde design, Mol9), EMD-75575 (MPNN-fixbb design, Mol14), EMD-75648 (Struct2Seq-SHAPE, Mol13), and EMD-75584 (MPNN-RFdiff dimer, Mol11). Cryo-EM images are available at EMPIAR under accession IDs EMPIAR-13759 (gRNAde design, Mol9), EMPIAR-13762 (MPNN-fixbb design, Mol14), EMPIAR-13765 (Struct2Seq-SHAPE, Mol13), and EMPIAR-13766 (MPNN-RFdiff dimer, Mol11). Codes are available at (50). Materials are available upon request. **License information:** Copyright © 2026 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. <https://www.science.org/about/science-licenses-journal-article-reuse>. This article is subject to HHMI's Open Access to Publications policy. HHMI lab heads have previously granted a nonexclusive CC BY 4.0 license to the public and a sublicensable license to HHMI in their research articles. Pursuant to those licenses, the Author Accepted Manuscript (AAM) of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

SUPPLEMENTARY MATERIALS

[science.org/doi/10.1126/science.aeg6829](https://doi.org/10.1126/science.aeg6829)
Materials and Methods; Supplementary Text; Figs. S1 to S14; Tables S1 to S8; References (51–58); Movies S1 and S2; Data S1; MDAR Reproducibility Checklist

Submitted 23 February 2026; accepted 30 June 2026

10.1126/science.aeg6829



De novo design of RNA pseudoknots with deep learning

Jill Townley, Wipapat Kladwang, David Baker, Hamish M. Blair, Christian A. Choe, Gina El Nesr, Andrew Favor, Eli Fisker, Daniel B. Haack, Shujun He, Jason Hingey, Po-Ssu Huang, Rui Huang, Chaitanya K. Joshi, Thomas Karagianes, Andrew Kubaney, Pietro Liò, Adamo Mancino, Jonathan Romano, Boris Rudolfs, Nicholas Spellmon, Navtej Toor, Jigyasa Verma, Vivian Wu, Zhiheng Yu, Eterna Participants, and Rhiju Das

Science **393** (6814), . DOI: 10.1126/science.aeg6829

Editor's summary

RNA molecules can fold into intricate three-dimensional shapes that drive much of their biology, but designing new structured RNAs from scratch has remained out of reach. Townley *et al.* show that complex RNA structures called pseudoknots can now be designed reliably using artificial intelligence (AI). In a year-long competition on the citizen science platform Eterna, AI methods solved more than 95% of 57 design challenges, matching the performance of expert human players (see the Perspective by Mustoe and Guo). Cryo-electron microscopy revealed that the molecules folded into entirely new three-dimensional architectures, sometimes featuring intricate interactions that the AI had not been instructed to build. RNA design has thus entered the deep-learning era. —Di Jiang

View the article online

<https://www.science.org/doi/10.1126/science.aeg6829>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2026 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works