

UC Davis

UC Davis Previously Published Works

Title

RNA target highlights in CASP15: Evaluation of predicted models by structure providers

Permalink

<https://escholarship.org/uc/item/4df4b5fs>

Journal

Proteins Structure Function and Bioinformatics, 91(12)

ISSN

0887-3585

Authors

Kretsch, Rachael C

Andersen, Ebbe S

Bujnicki, Janusz M

et al.

Publication Date

2023-12-01

DOI

10.1002/prot.26550

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

RNA target highlights in CASP15: Evaluation of predicted models by structure providers

Rachael C. Kretsch ¹	ORCID: https://orcid.org/0000-0002-6935-518X
Ebbe S. Andersen ²	ORCID: https://orcid.org/0000-0002-6236-8164
Janusz M. Bujnicki ³	ORCID: https://orcid.org/0000-0002-6633-165X
Wah Chiu ^{1,4,5}	ORCID: https://orcid.org/0000-0002-8910-3078
Rhiju Das ^{1,6,7}	ORCID: https://orcid.org/0000-0001-7497-0972
Bingnan Luo ⁸	ORCID: https://orcid.org/0000-0002-9367-4359
Benoît Masquida ⁹	ORCID: https://orcid.org/0000-0002-6565-2099
Ewan K.S. McRae ¹⁰	ORCID: https://orcid.org/0000-0001-9105-2575
Griffin M. Schroeder ^{11,12}	ORCID: https://orcid.org/0000-0001-6354-752X
Zhaoming Su ⁸	ORCID: https://orcid.org/0000-0002-9279-1721
Joseph E. Wedekind ^{11,12}	ORCID: https://orcid.org/0000-0002-4269-4229
Lily Xu ¹³	ORCID: https://orcid.org/0000-0002-1200-8937
Kaiming Zhang ¹⁴	ORCID: https://orcid.org/0000-0003-0414-4776
Ivan N. Zheludev ⁶	ORCID: https://orcid.org/0000-0002-9572-0574
John Moul ^{15,*}	ORCID: https://orcid.org/0000-0002-3012-2282
Andriy Kryshfovych ^{16,*}	ORCID: https://orcid.org/0000-0001-5066-7178

¹ Biophysics Program, Stanford University School of Medicine, Stanford, CA, USA.

² Interdisciplinary Nanoscience Center and Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark.

³ International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland.

⁴ Department of Bioengineering and James H. Clark Center, Stanford University, Stanford, CA, USA.

⁵ Division of CryoEM and Bioimaging, SSRL, SLAC National Accelerator Laboratory, Menlo Park, CA, USA.

⁶ Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA.

⁷ Howard Hughes Medical Institute, Stanford, CA, USA.

⁸ The State Key Laboratory of Biotherapy, Frontiers Medical Center of Tianfu Jincheng Laboratory, Department of Geriatrics and National Clinical Research Center for Geriatrics, West China Hospital, Sichuan University, Chengdu 610044, Sichuan, China.

⁹ UMR 7156, CNRS – Université de Strasbourg, Strasbourg, France.

¹⁰ Center for RNA Therapeutics, Houston Methodist Research Institute, Houston, TX 77030, USA.

¹¹ Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY, 14642, USA.

¹² Center for RNA Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY, 14642, USA.

¹³ Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.

¹⁴ Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230027, China

¹⁵ Department of Cell Biology and Molecular Genetics, Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, Maryland, USA.

¹⁶ Genome Center, University of California, Davis, Davis California, USA.

* Correspondence to jmoult@umd.edu and akryshfovych@ucdavis.edu.

Abstract

The first RNA category of the CASP competition was only made possible because of the scientists who provided experimental structures to challenge the predictors. In this article, these scientists offer a unique and valuable analysis of both the successes and areas for improvement in the predicted models. All ten RNA-only targets yielded predictions topologically similar to experimentally determined structures. For one target, experimentalists were able to phase their X-ray diffraction data by molecular replacement, showing a potential application of structure predictions for RNA structural biologists. Recommended areas for improvement include: enhancing the accuracy in local interaction predictions and increased consideration of the experimental conditions such as multimerization, structure determination method, and time along folding pathways. The prediction of RNA-protein complexes remains the most significant challenge. Finally, given the intrinsic flexibility of many RNAs, we propose the consideration of ensemble models.

1 Introduction

Experimental structural biologists are integral to the success of the Critical Assessment of Techniques for Structure Prediction (CASP) and are increasingly benefitting from the predictive capabilities enabled by experiments like CASP. Complementing the RNA-puzzles efforts for double-blind RNA three-dimensional structure prediction¹⁻⁴, in this first RNA category of CASP (CASP15, 2022), ten RNA and two RNA-protein complexes were suggested as modeling targets by six structure determination groups from four countries. All targets were released for prediction from May to July 2022. Among these, four targets were solved by X-ray crystallography and eight by cryogenic electron microscopy (cryo-EM).

This article follows the tradition of protein CASP target highlight articles⁵⁻¹⁰, where each section provides the accounts of the structure providers and their insights into the accuracy of the best models submitted. All target providers were invited to contribute to this paper, resulting in five sections highlighting nine of the targets (**Table 1**).

The numerical evaluation of CASP15 RNA models is available at the Prediction Center website (https://predictioncenter.org/casp15/results.cgi?tr_type=rna). The detailed evaluation of these predicted models, including direct comparisons and refinement to X-ray and cryo-EM data, are provided elsewhere in this issue^{TBD}.

Table 1: CASP15 RNA targets included into this study.

Target	Name	PDB	Length (nt)	Method	Resolution (Å)	Potential template	Comments	Top 5 score range		
								GDT-TS ¹	IDDT ²	RMSD ³ (Å)
R1107	Human CPEB3 ribozyme	7QR4	69	X-ray	2.83	4PR65	Crystallographic dimer (A2)	63.4543	0.728-0.710	4.52-5.92
R1108	Chimpanzee CPEB3 ribozyme	7QR3	69	X-ray	2.18		Non-crystallographic dimer (A2) A30G mutation from human	64.5-59.8	0.755-0.742	4.49-4.94
R1117	PreQ ₁ riboswitch	8FZA	30	X-ray	2.30	2L1V, 3Q50, 3FU2	Ligand present	86.2-85.3	0.747-0.728	2.01-2.43
R1128	RNA origami	8BTZ	238	cryo-EM	5.39	-	Existence of dynamic range of conformations	51.1-38.3	0.867-0.863	4.33-6.11
R1138		7PTK, 7PTL	720	cryo-EM	5.18, 4.90	KL4: 2D1B	Kinetically trapped "young" state and mature state Existence of dynamic range of conformations	29.9-25.7	0.739-0.729	7.82-10.30
R1149	SARS-CoV-2 5' SL5	-	124	cryo-EM	4.74	-	Multiple models represent experimental uncertainty	43.4-39.3	0.746-0.730	6.88-7.98
R1156	BtCoV-HKU5 5' SL5	-	135	cryo-EM	5.83, 6.59, 7.48, 7.61		Helical bend Multiple models represent experimental uncertainty	51.5-34.6	0.729-0.722	5.37-12.14
R1189	RsmZ-A complex	7YR7	118	cryo-EM	3.80	2MF0	RsmZ-A ₃ (A1B6)	23.1-22.7	0.551-0.549	16.29-16.60
R1190		7YR6		cryo-EM	4.60		RsmZ-A ₂ (A1B4)	26.5-24.2	0.603-0.588	15.96-16.18

¹ Local-Global Alignment¹¹ used to measure Global Distance Test (GDT-TS), based on the average percentage of aligned C4' atoms.

² Local Distance Difference Test (IDDT)¹² calculated over all heavy atoms.

³ Root-Mean Squared Deviation (RMSD) of all heavy-atoms after all heavy-atom superposition, RNA-tools¹³.

⁴ Template for kissing-loop motif only.

5 Additionally, there are many templates for the U1A-protein binding loop.

2 Results

2.1 Human and chimpanzee CPEB3 ribozymes (CASP: R1107 and R1108, PDB: 7QR4 and 7QR3) Provided by Benoît Masquida

Cytoplasmic polyadenylation element binding protein 3 (CPEB3 protein) binds CPEs of mRNAs to regulate poly-A tail extension and translation^{14,15}. It plays a role in memory acquisition and maintenance requiring tight post-transcriptional regulation in mammals. One regulatory mechanism intervenes at the post-translational level and comprises SUMOylation of a lysine residue from an F-actin binding region embedded in the N-terminal prion domain. SUMOylation prevents binding of the protein to actin filaments, and contributes to localization of the protein in P-bodies together with the stalled tissue-dependent mRNA targets. Upon neuronal stimulation, the SUMO tag is removed and the CPEB3 proteins aggregate on F-actin filaments and promote translation of their mRNA targets¹⁶⁻¹⁹.

The CPEB3 gene encodes a ribozyme conserved in the mammalian order, embedded in the second intron of the pre-mRNA²⁰. This ribozyme is very similar to the Hepatitis delta virus (HDV) ribozyme, although its characteristic slow cleavage activity both *in vitro* and *in vivo* allows a subtle coupling with splicing. Slowing down catalytic activity using antisense oligonucleotides prevents the formation of a catalytic structure and results in increasing the cellular levels of both CPEB3 mRNA and protein²¹.

RNA constructs of the human and chimpanzee CPEB3 ribozymes modified by insertion of a U1A protein binding motif in place of the wild-type P4 were co-crystallized in the presence of the U1A protein to foster crystal-packing contacts²² (**Table 1**). A30 of the human ribozyme is changed to G30 in the chimpanzee homologue. The crystal structures show an overall organization consistent with that of the HDV ribozyme wherein helix P1 stacks onto helix P4 on one side and helix P2 stacks on helix P3 on the other (**Figure 1A-C**)²³. Ribozyme dimers were obtained in both cases, although the dimer was non-crystallographic for the chimpanzee RNA. The dimerization occurs through the L3 loops of two molecules like a handshake. L3 contains the two residues (U21, C22) involved in formation of the characteristic HDV-ribozyme-like double nested pseudoknot with two residues in J1/4 (G37 and U38). A L3 palindromic sequence stretch 5'-A(23)CGU-3' actually makes dimerization possible and hence prevents formation of a competent catalytic pocket. An additional striking feature of the ribozyme dimers is that the dimerizing L3 loop, which harbors seven nucleotides in the CPEB3 ribozyme, not eight like in the HDV ribozyme, adopts the same conformation as an anticodon loop from tRNAs interacting with the cognate codons during translation²⁴⁻²⁶.

The predicted models for the CPEB3 ribozyme are all monomers. Models with RMSD around 5 Å correctly predict the main secondary structure elements (P1, P2, P3, P4) as well as their relative positions. In the 5 Å RMSD range, the main discrepancies correspond to the residue

conformations belonging to non-helical regions. Although the U1A loop is well predicted, perhaps because known structures could be used as a template, the L3 loop departs significantly from the observed anticodon-like conformation. Instead, a U-turn occurs between U21 and C22 (Compare **Figure 1.D** to **Figure 1E**), which is involved in the P1.1 pseudoknot in the template HDV structures but not in the CPEB3 ribozyme structures. This modeling error is most probably due to the fact that the CPEB3 ribozyme target was treated as a monomer, and accounts the lack of predicted model with RMSD values better than 4.52 Å for the human ribozyme and 5.48 Å for the chimpanzee. Another region of the human ribozyme that was wrongly predicted, independently from the dimerization, is the J1/2 stretch. In the best model R1107TS232_1, generated by Ai chemistry_RNA2 (**Figure 1E**), P1 is closed by a sugar edge-Hoogsteen A8-A30 pair and the nucleotides upstream and downstream from A8 stack on each other, adopting a helical conformation to conduct the strand to the inlet of P2. However, in the crystal structure, the C residue is expelled into solvent and the sugar edge of the contiguous A residue interacts with the Watson-Crick edge of A30 in P1. The situation for the chimpanzee ribozyme is different since a G-C pair is formed at the tip of P1, which was easier to identify.

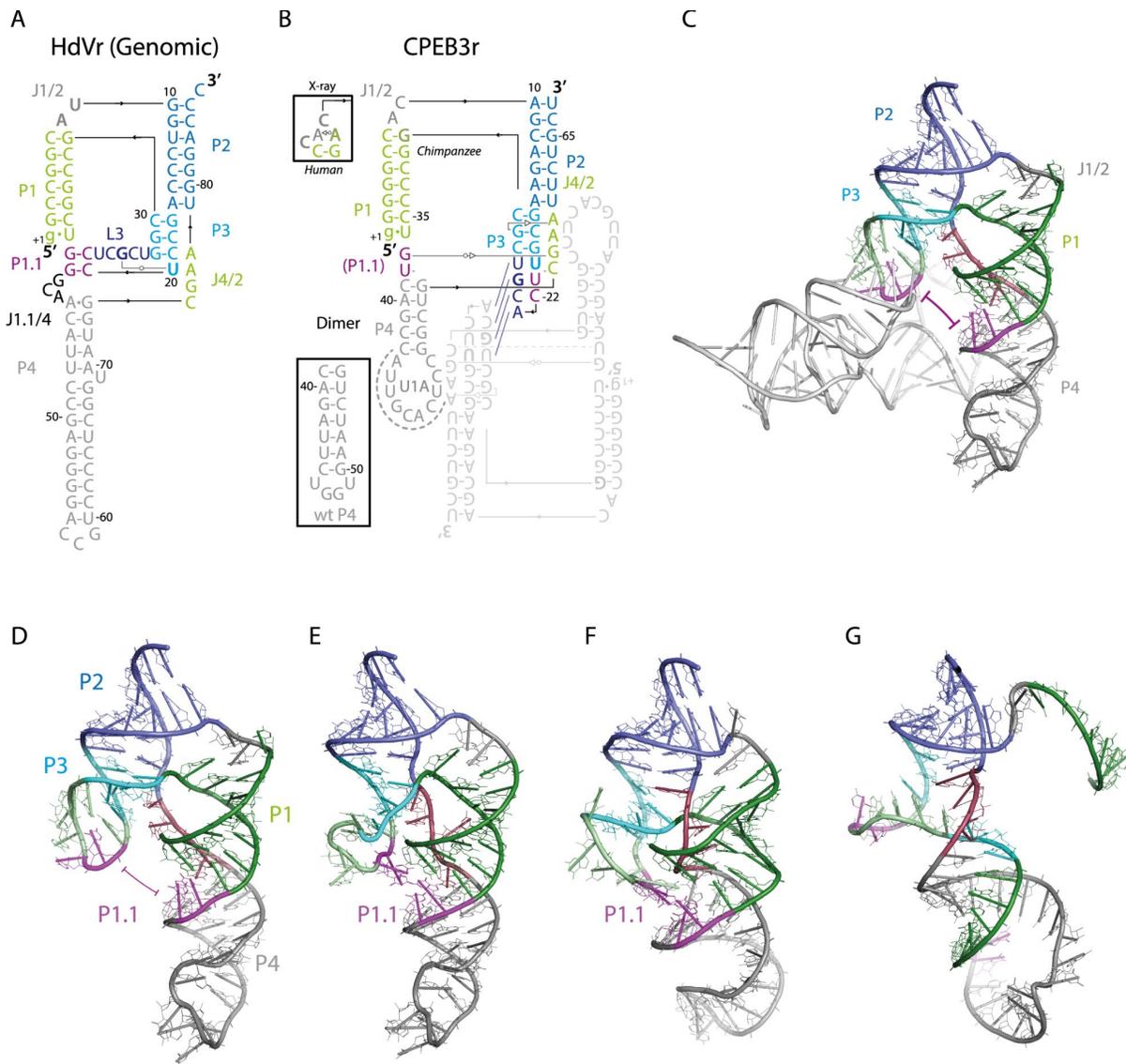


Figure 1: (A) Secondary structure of the HDV ribozyme as deduced from the crystal structure (PDB: 1DRZ)²⁷. Among the characteristic structural elements, P1 forms a nested double pseudoknot together with P1.1 and P2. (B) The secondary structure of the human and chimpanzee CPEB3 ribozymes as deduced from the crystal structures (C) show that the P1.1 element (purple) is not formed and instead the ribozymes form dimers with a neighboring molecule (gray shaded for clarity). The residues are numbered according to the P4 wild type sequence framed. (D-G) Comparison of representative models with the crystal structure of the human CPEB3 ribozyme. (D) In the crystal structure of the human CPEB3 ribozyme (PDB: 7QR4)²³, the distal location of the P1.1 forming elements (purple) is indicated by a symbol (|---|). (E) R1107TS232_1, Ai_chemistry_RNA2, RMSD 4.52 Å; (F) R1107TS054_3, Ultrafold, RMSD 8.13 Å; (G) R1107TS229_1, Yang_server, RMSD 17.92 Å.

Looking at models with worse accuracy, the increase of the RMSD values up to 10 Å is associated with misfolding of the ribozyme, including strand crossing and also topological differences compared to the crystal structures (**Figure 1F**). In model R1107TS054_3, generated by the Ultrafold server, the connection between P4 and P2 is made on the deep groove side of

P3 instead of on its shallow groove side, which scrambles the catalytic site. This conformation would result from a different folding process since the single strand J4/2 ends up on the other side of P3. For models in this range of RMSD values, accuracy of loop modeling is also worse. The U1A region is for example modeled as a simple loop where a U-turn is mediated adequately to reverse the backbone direction and mediate loop closure without care of individual nucleotide conformations particularly the fact that this loop is bound to the U1A protein.

Beyond RMSD values of 10 Å, aberrant secondary structure elements appear. For example, in model R1107TS229_1 from Yang_server, the two strands forming P3 are split and reorganized around a three-way junction connecting P2, the L3 loop and a P4 element presenting a four base pair extension encompassing the second strand of P3 and the residues from J4/2. This results in a profound reorganization of P4 caused by the interaction with the second strand of P1 (green base-paired region in **Figure 1G**). This h leads to the misfolding of the U1A protein binding site.

To summarize, for models that achieved below 5 Å RMSD, conformational discrepancies are mostly observed for residues belonging to loops (**Figure 1E**). Around 10 Å RMSD values, additional strand crossing events are observed leading to topological differences resulting from different folding pathways (**Figure 1F**). Finally, when close to 20 Å RMSD, shuffling of the strands constitutive of individual helices may generate spurious secondary structure elements resulting in the loss of similarity between models and reference structures (**Figure 1G**).

2.2 Small preQ₁ riboswitch (CASP: R1117, PDB: 8FZA) Provided by Griffin M. Schroeder and Joseph E. Wedekind

Riboswitches are gene-regulatory elements usually located in the 5' untranslated region of bacterial messenger (m)RNA²⁸. Riboswitches regulate downstream genes by use of an aptamer domain that senses a cellular metabolite with high specificity²⁹. Metabolite binding triggers conformational changes in a nearby, gene-regulatory expression platform that induce transcription termination or translation initiation^{30,31}. The cognate ligand is usually a cofactor or metabolite intermediate related to the downstream gene, allowing the riboswitch to maintain bacterial homeostasis through feedback loops²⁹. Importantly, dysregulation of riboswitches has been shown to decrease bacterial fitness, making riboswitches attractive drug targets³².

Of the over 55 classes of validated riboswitches³³, one of the best studied is the prequeosine₁ (preQ₁) sensing family. One³⁴ or two³⁵ preQ₁ metabolites bind per aptamer domain, which adopts a distinct architecture that falls into one of three folding classes. Of these classes, the class I riboswitch is the most widely distributed among bacteria and is the most prevalent preQ₁ riboswitch in the biosphere³⁴. This class can be divided into three subgroups known as types I-III (preQ₁-I-III). Although each subgroup is predicted to fold into an H-type pseudoknot³⁴, we previously demonstrated that types I and II show different aptamer-to-preQ₁ binding stoichiometries despite sharing a common global fold³⁵. At present, little is known about the type III subtype, which is found almost exclusively in proteobacteria³⁴. Accordingly, we determined the

co-crystal structure of a preQ₁-I_{III} (type III class I) riboswitch (PDB: 8FZA, **Table 1**) to ascertain how preQ₁ recognition leads to gene regulation by this clinically relevant riboswitch subclass.

To obtain diffraction-quality crystals, a poorly conserved turn between helix P1 and loop L3 (**Figure 2A**) was modified to yield a small 30-mer construct ideal for structure prediction, which was submitted to CASP15. Consistent with the covariation model³⁴ (**Figure 2A**), the crystal structure revealed a highly compact H-type pseudoknot (**Figure 2B**) featuring two helical regions, P1 and P2, joined by three loop regions, L1-L3. Many groups — most notably the Chen, GeneSilico and Aichemy_RNA2 groups — correctly predicted the global fold. However, we were struck by the fourth model generated by the Chen group (R1117TS287_4) because its all-atom RMSD with our experimental structure was 2.01 Å (**Figure 2B**), the lowest of all CASP15 RNA targets. Major areas of deviation include the sharp L1-P2 bend at Cyt8 located in the ceiling of the binding pocket (RMSD of 6.39 Å at atom O2), Cyt12 in loop L2 (RMSD of 5.67 Å at atom OP1) and the P1-L3 turn (RMSD of 7.68 Å at atom O2' of Ade21), which was modified to promote crystallization. Both Aichemy_RNA2 (R1117TS232_1) and GeneSilico (R1117TS128_1) produced slightly poorer predictions based on global RMSD values of 2.27 Å and 2.43 Å. Like Chen, the latter two models showed difficulties predicting mainchain and base positions at Cyt8, Cyt12 and P1-L3. These pseudoknot loop and turn regions showed substantial conformational differences when comparing co-crystal structures of known type I and II preQ₁ riboswitches^{35–39}. The observation underscores the need for more experimentally-derived templates.

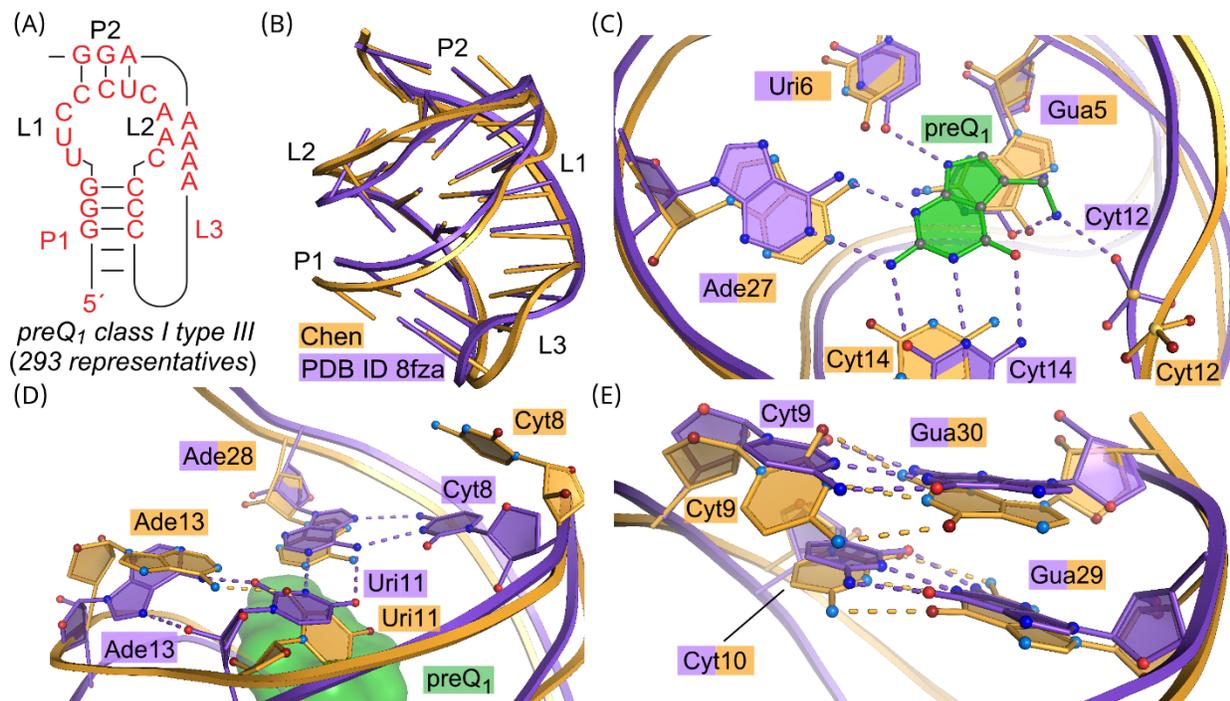


Figure 2 Covariation model and comparison of the preQ₁-I_{III} riboswitch co-crystal structure to the best predicted CASP15 model. (A) Covariation model based on previous data³⁴. (B) Global superposition of the experimental model (PDB:8FZA, purple) with the top prediction model (R1117TS287_4, orange). (C) Close-up view of the preQ₁ binding pocket. The metabolite (green) was derived from the co-crystal structure. (D)

Close-up of the pocket ceiling. (E) The expression platform showing WC pairing of Gua29 and Gua30 of the Shine-Dalgarno sequence.

Metabolite-binding is an area of functional interest and ligand binding stabilizes the structure by promoting coaxial helical stacking. The co-crystal structure shows high homology to the binding pockets of other preQ₁-I riboswitches^{35–39}. Specificity base⁴⁰ Cyt14 uses *cis* Watson-Crick (WC) pairing to engage preQ₁ (**Figure 2C**). The minor-groove edge of the metabolite is read by Uri6 and Ade27, while the methylamine donates hydrogen bonds to Gua5 and the backbone of Cyt12. The Chen model did not attempt to predict the mode of preQ₁ binding. Rather, their model predicts that metabolite-interacting nucleobases are unpaired, although they are oriented similarly to the bound-state crystal structure (**Figure 2C**). However, Gua5, Uri6, Cyt14 and Ade27 of the Chen model pack more closely in the core while the Cyt12 phosphate bulges outward. Thus, the orientation of these nucleobases in the Chen model prevents hydrogen bond contacts to preQ₁ (**Figure 2C**). Curiously, the Chen apo model does not predict nucleobase incursion into the preQ₁ binding pocket, although this effect was observed previously – along with L2 loop unstacking – in apo-state co-crystal structures of a related *T. tengcongensis* (Tte) preQ₁-III riboswitch^{36,41}. Hence, Chen's apo-state prediction actually resembles a bound state, possibly due to bias from bound-state templates in which the ligand was removed. Nonetheless, gross details of the fold were predicted correctly.

The co-crystal structure of the aforementioned Tte preQ₁-II riboswitch further revealed that the binding pocket ceiling forms a base quartet^{36,37}. We found previously that this quartet plays a key role in the preQ₁-free to bound-state interconversion⁴¹. In our co-crystal structure of the preQ₁-III riboswitch herein, we observe a similar pocket comprising a Cyt8•Ade28-Uri11•Ade13 quartet (**Figure 2D**). Uri11 stacks atop preQ₁, interacts with Ade13 through its sugar edge, and forms a WC pair with Ade28. The Hoogsteen edge of the latter base also interacts with Cyt8 (**Figure 2D**), which is notable because it is part of the Shine-Dalgarno sequence (SDS). Thus, our co-crystal structure provides insight into how preQ₁ recognition leads to gene regulation through sequestration of the SDS in the pocket ceiling.

By contrast, interactions in the pocket ceiling are sparse in the Chen model. Of the six hydrogen bonds observed in our co-crystal structure, only the interaction between N6 of Ade13 and O2 of Uri11 is present (**Figure 2D**). Although the covariation model predicts a WC pair between Uri11 and Ade28³⁴ (**Figure 2A**), the Chen model predicts that Uri11 twisted downward into the preQ₁ pocket and shifted toward Cyt8 (**Figure 2D**) where it cannot hydrogen bond with Ade28. Similarly, Cyt8 adopts a dramatically different orientation that pivots the nucleobase upward and away from the planar rings that compose the pocket ceiling, thereby precluding formation of the Uri11-Ade28-Cyt8 triple (**Figure 2D**). Atop the pocket ceiling, the next two SDS nucleotides, Gua29 and Gua30, form WC interactions in the co-crystal structure consistent with covariation predictions³⁴ (**Figure 2A,E**). This interaction is correct in the Chen model, although Cyt9 shows substantial propeller twist (**Figure 2E**). Overall, the structural basis of gene regulation in the Chen model is largely consistent with predictions from the covariation model³⁴. Albeit, the preQ₁ pocket and ceiling differ in important ways from the experimental coordinates.

The fourth Chen model (R1117TS287_4) is the most accurate RNA prediction in the CASP15 competition according to GDT_TS and RMSD (**Table 1**). This laudable achievement may be due to the similarity of its global fold to known preQ₁-I riboswitch structures^{35–39,41} and the small target size. Although details related to metabolite binding and gene regulation were somewhat obscured (**Figure 2C-E**), the predicted structure (Chen model R1117TS287_4) succeeded as a molecular replacement (MR) search model after minor modifications. Specifically, we removed residue 1 at the 5'-end of the P1 stem and residues 20-23 at the P1-to-L3 turn in the search model (Figure 2b). This search model yielded a translation-function z-score of 8.4 and a log-likelihood gain of 168 in Phenix⁴². These modifications were obvious choices based on their lack of conservation in the covariation model³⁴ and were necessary for crystal packing. Overall, the Chen model and others represent valuable tools to predict the global folds of small RNAs and to facilitate their experimental structure determinations by MR.

2.3 RNA origami (CASP: R1128 and R1138, PDB: 8BTZ, 7PTK, and 7PTL) Provided by Ewan K.S. McRae and Ebbe S. Andersen

RNA origami are tertiary structures that are designed to fold during transcription. The RNA origami architecture utilizes coaxially stacked 4-way junctions and internal pseudoknots (Kissing loops (KLs)) to create a network of helical components from a single strand of RNA⁴³. Recent improvements to the automated design software for RNA origami (ROAD) have allowed us to rapidly generate many unique new design patterns and easily incorporate RNA aptamers into the designs⁴⁴. Keen to validate the fidelity of our designer RNA from *in silico* to *in vitro*, we pursued structural determination of our co-transcriptionally folded and natively purified RNA using cryogenic electron microscopy⁴⁵. During this process we encountered numerous deviations between our designed structures and our experimentally determined structures, notably in the twist, bend, and topological arrangement of helices.

In our opinion, synthetic nucleotide sequences represent a particularly interesting class of targets for structural prediction contests because they have very little sequence homology to known RNA structures, forcing predictions to rely first on the principles of RNA folding rather than homology modeling. Furthermore, the motifs that we do borrow from known structures (i.e. KLs and aptamers) are often different in the context of a larger RNA and in solution than they are in isolation in a crystal structure⁴⁶.

Our design process includes validation of our sequences by comparing the predicted secondary structures from Vienna RNA⁴⁷ and NUPACK^{48,49} to our designs. This, coupled with the almost entirely base-paired nature of our designs, means that prediction of the correct base pairing arrangement (i.e. secondary structure) should be trivial and the real challenge lies in correctly predicting the topological arrangement of helical elements and subtle deviations from ideal A-form helix.

Here we compare the best model from the top 5 groups (lowest global RMSD to our model) to two of our target submissions. Not surprisingly, the base pairing was almost always

correctly predicted. The most frequent deviations from our model were typically the result of missing pseudoknot interactions (i.e., long range tertiary interactions) or incorrect modeling of the 4-way junctions. However, we were thrilled to see that at least two groups consistently modeled our targets *in silico* with excellent agreement to our experimental models. In some cases, their predictions were closer to the empirical structure than our initial design.

Our simplest target was a 238 nucleotide RNA (**Figure 3A**) comprising three helical domains connected by two four-way junctions and a paranemic crossover (PX) (CASP:R1128 PDB: 8BTZ)⁵⁰. Out of the best models from the top 5 prediction groups, only two accurately modeled the topology (**Figure 3B,C**), two failed to find the PX (**Figure 3D,E**) and one modeled the PX but failed to coaxially stack the 5' and 3' end helices into a single continuous helix (**Figure 3F**). To accommodate this tight packing of three crossovers, it appears that at least one of the helices must adopt a slight bend. In our cryo-EM reconstruction this is helix 2. The two best CASP predictions show a more noticeable bending of helices 1 & 3 (**Figure 3C**) or just helix 3 (**Figure 3B**). It is our opinion that these alternate bends are likely sampled as part of the dynamic range of conformations adopted by the RNA in solution.

Our largest target was a 720 nucleotide sequence (CASP:R1138 PDB:7PTK,7PTL), that was designed to form a hexagonal arrangement of 6 parallel helices, connected by 10 four-way junctions and 5 internal KLs. The RNA structure was found to have an unusually stable folding intermediate (PDB:7PTK & **Figure 3G**) that persists in solution for several hours after transcription. This early conformation has the final “latching” helix laying across the other helices; we followed the transition from this early state to a matured state using small-angle X-ray scattering (SAXS) and determine the half-life of the early structure to be ~10 hours, after which it rearranges into a more compact structure with the latch helix more parallel to the rest of the bundle, but still not completely parallel as we designed it (PDB: 7PTL & **Figure 3H**)⁴⁵.

Among the top 5 predictions for the bundle, one group did not model any of the KLs (**Figure 3M**) and another group found 4 out of 5 KLs, seemingly missing the 5th KL by not accounting for the curvature induced by the crossover seams that allows the latch helix to make the final KL and by having incorrect helical stacking across the four-way junctions in the 5' half of the bundle (**Figure 3L**). Two groups modeled all KLs and predicted the curvature from the crossover seams almost exactly as we designed *in silico* (**Figure 3J,K**). Most excitingly for us, the A_ichemy_RNA2 group produced a model that more closely matches the empirical structure of the mature conformation than our initial design did (**Figure 3I**)! None of the groups predicted the early conformation.

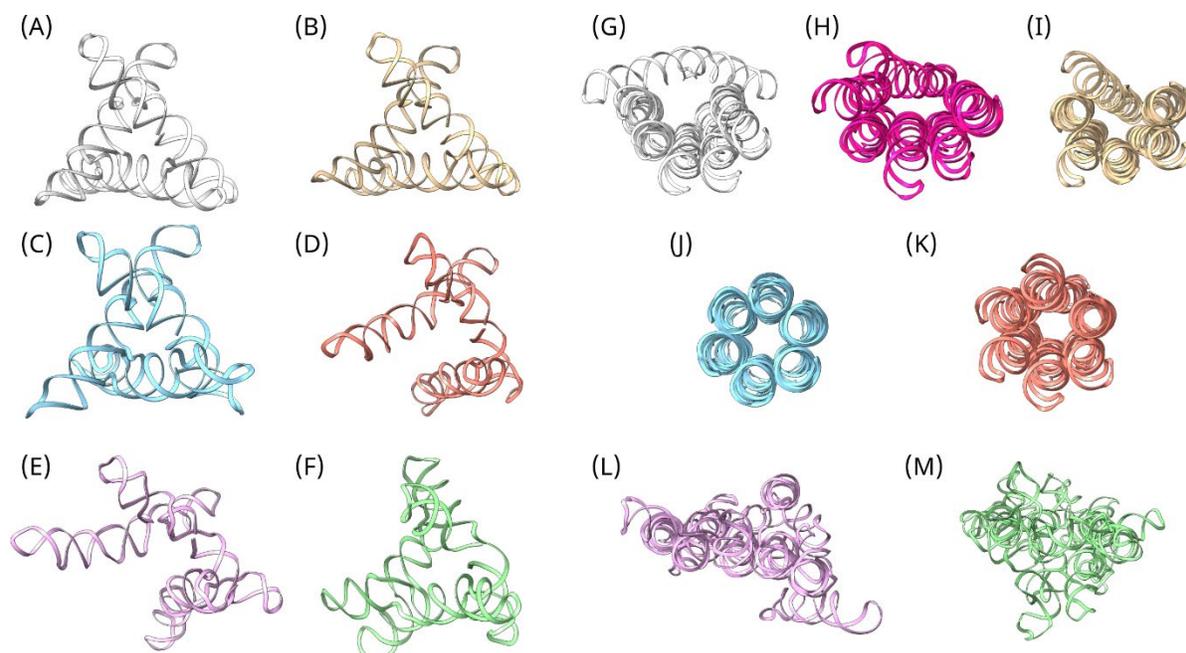


Figure 3. Comparing ribbon models of the experimentally determined structures (A, G, H) to the best predictions from the top 5 groups for CASP:R1128 (A-F, left) and CASP:R1138 (G-M, right). (B) R1128TS232 Aichemy_RNA2, (C) R1128TS287 Chen, (D) R1128TS147 SHT, (E) R1128TS227 GinobiFold, (F) R1127TS125 UltraFold_Server, (I) R1138TS232 Aichemy_RNA2, (J) R1138TS287 Chen, (K) R1138TS081 RNApolis, (L) R1138TS128 GeneSilico, (M) R1138TS227 GinobiFold.

As a final comment, in our designs we frequently used the HIV DIS kissing loop (KL), based on the crystal structures from E. Ennifar and P. Dumas⁵¹. From our highest resolution cryo-EM map we were able to determine that the KL is more compact than in the crystal structure, resulting in a twist defect that is compounded by the number of KLs incorporated. The main difference between our structure and the crystal structure is that the unpaired adenines in our model stack within the helix, while the adenines from the crystal structure are bulged out. Although we designed our structures to have straight helices throughout, this twist defect resulted in an inherent strain throughout our origami and bending of the helices to accommodate this. It appears that the CASP predictors also used this crystal structure to seed their predictions as most of the KLs in the predictions are bulged out. Perhaps consequently, the predictions have much straighter helices than we observe in our empirical structures.

In conclusion, the secondary structures of our synthetic sequences were successfully predicted by most participants. Tertiary structure proved more challenging, especially with long range pseudoknots, 4-way junctions and with motifs that differ in our cryo-EM maps compared to crystal structures. However, at least one group was consistently able to accurately predict the approximate 3D structure. Finally, the inherent complicating factor with RNA is flexibility, which presents challenges not only to prediction but to assessment of predictions. Each structure we submit as a target is the result of averaging thousands of slightly different conformations of the same overall structure using cryo-EM single particle analysis methods. Although the CASP submissions were scored against a single PDB model built into our best resolved reconstruction, we know from 3D variability analysis of our cryo-EM data sets that there exists a dynamic range

of conformations^{45,46}. We anticipate that incorporation of structural dynamics and co-transcriptional folding pathways will be the next major hurdle for RNA structure prediction.

2.4 Coronavirus 5' stem loop 5 (SL5) domain (CASP: R1149 and R1156) Provided by Rachael Kretsch, Lily Xu, Ivan N. Zheludev, Kaiming Zhang, Rhiju Das, and Wah Chiu

Coronaviruses have a highly structured 5' region with several 'stem-loop' (SL) elements; stem-loop 5 (SL5) was predicted to fold into a four-way junction in most SARS-related betacoronaviruses^{52,53}. For some coronaviruses, experimental data from multiple labs, including covariance analysis and chemical mapping, support this secondary structure^{54–58}. The tertiary organization of this junction and the degree of its structural conservation is unknown. In fact, previous computational modeling of the SARS-CoV-2 SL5 suggested that this domain might not have a well-defined tertiary structure⁵⁸. We were pleased to resolve defined tertiary structures for SARS-CoV-2 (R1149) and BtCoV-HKU (R1156) SL5 domains by cryo-EM. These were well-suited for evaluating 3D structure prediction because these RNA folds can be simplified to a handful of elements while also introducing conformational heterogeneity that the RNA and macromolecule modeling communities are increasingly interested in.

From multidimensional chemical mapping⁵⁹, medium-resolution cryo-EM maps⁶⁰ and heterogeneity analysis^{60,61} we obtained one map for the SARS-CoV-2 SL5 domain and four maps for BtCoV-HKU5 SL5 domain, after data analysis suggested flexibility in SL5a (**Table 1**). We generated ten models for each map (ten and forty models respectively) to represent our experimental uncertainty, due to their medium resolution nature. Overall, we were pleasantly surprised to find that some predicted models were superimposable on experimental models, and achieved reasonable accuracy, by global metrics such as GDT-TS, including submissions from GeneSilico (TS128) and DeepFoldRNA (TS110) highlighted here (**Figure 4F,H**). For BtCoV-HKU5, we were glad to have conveyed an experimental structure ensemble, because the top model from GeneSilico, R1156TS128_5, was an excellent fit an intermediate conformations but would not have been an excellent fit to our highest resolution map, which captured the highest bend angle of SL5a (**Figure 4H**). Keeping the resolution and flexibility in mind, we enumerated features, and investigated how well the top models modeled these features as well as why some models predicted these features but did not score well globally.

For both the SARS-CoV-2 and BtCoV-HKU5 SL5 domains, we observed that (1) the junction was tight, i.e., it had 4 closing base-pairs without any unpaired bases in the junction; (2) the outermost SL5-stem was stacked on SL5c, and SL5a on SL5b. Focusing on the region of interest, the junction, 73 SARS-CoV-2 models and 46 BtCoV-HKU5 models recovered correct base-pairing at the junction (**Figure 4A,C**), an expected level of accuracy given information in the literature, but, we saw some groups modeled a looser junction with unpaired bases. The second challenge was deciding if the four stems were coaxially stacked at the junction, and if so, what the coaxial stacking pattern was. 43 models for SARS-CoV-2 and 15 models for BtCoV-HKU5 (56% and 33% respectively, of submissions predicting a tight junction) correctly stacked the bases

at the junction (**Figure 4A,C**). These observations suggest that predicting coaxial stacking is still a challenge, despite past literature reporting higher accuracies⁶². Interestingly, the top model by GDT-TS for SARS-CoV-2, DeepFoldRNA's R1149TS110_2, exhibited incorrect base-pairing and stacking at the junction indicated correctly predicting these features is not a prerequisite to obtain overall topology (**Figure 4G**).

The third observation for both SL5 domains was that (3) the pairs of coaxially stacked helices were at a $\sim 90^\circ$ angle with antiparallel strands. This proved to be the most challenging task, even with a lenient angle criterion of $-90^\circ \pm 30^\circ$. Only seven models for SARS-CoV-2 and two models for BtCoV-HKU5 passed this criterion (21% and 13% respectively, of submissions that passed our previous two criteria). For SARS-CoV-2, the predicted models exhibited a wide range of angles with both parallel and antiparallel conformations proposed and no clear preferred orientation amongst the models (**Figure 4B**). For BtCoV-HKU5, we also saw a wide range of angles proposed, with a slight preference for models closer to the parallel orientation than the experimental models (**Figure 4D**). R1156TS287_5 from the Chen group is an example of a model that was accurate except for the angle between the helices; it is in an antiparallel orientation, causing this model to be topologically inaccurate (**Figure 4H**). The prediction of junction angles seems to be a challenge, but models like GeneSilico's R1149TS128_1 and R1156TS128_5 show that it is possible (**Figure 4F,H**).

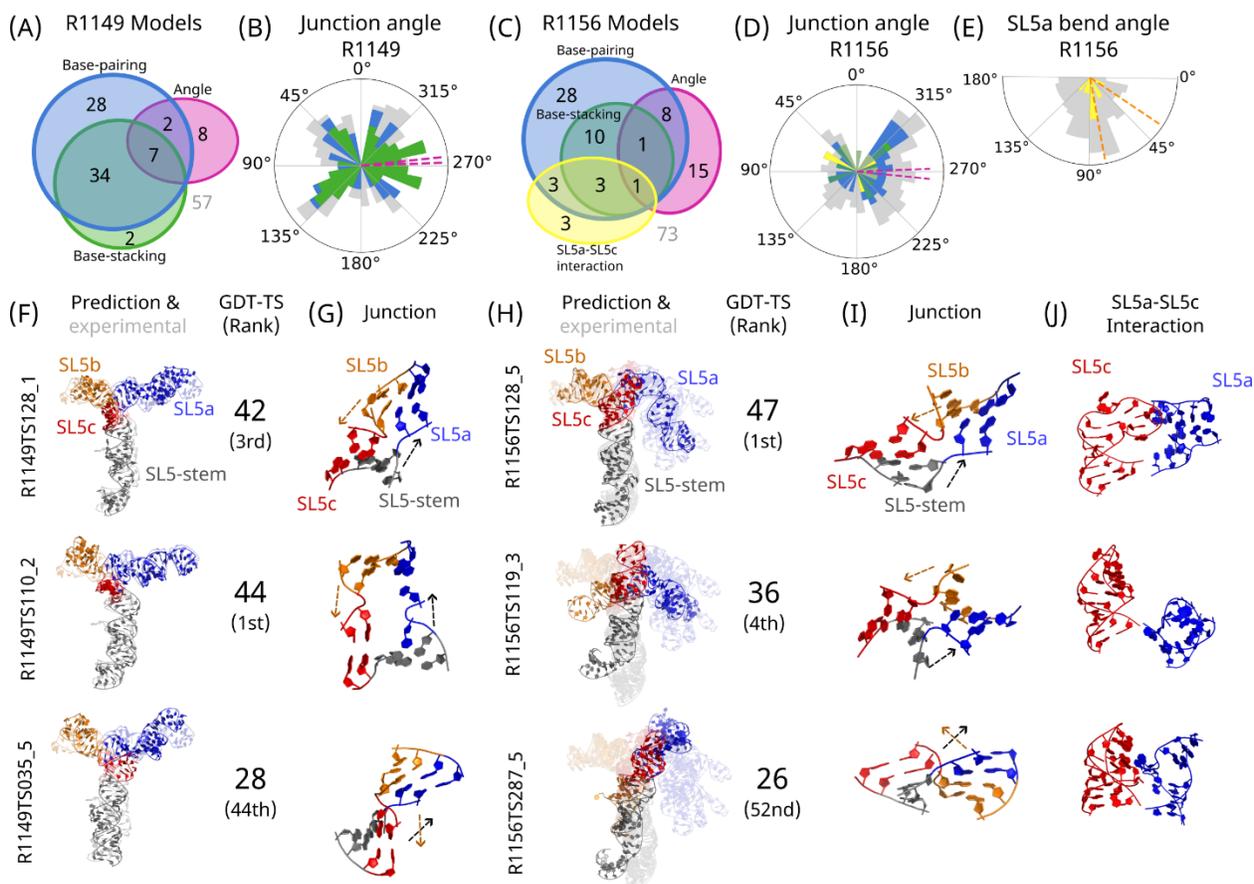


Figure 4: Categorization of all R1149 (SARS-CoV-2 SL5 domain) submitted models (A) and all R1156 (BtCoV-HKU5 SL5 domain) submitted models (C) by features they correctly predict. In the Venn diagram

(not to scale), areas are labeled with the number of models that correctly predict the features whose circles overlaps in that area: base-pairing (blue) and base-stacking (green) at junction, angle between SL5-stem-SL5c and SL5a-SL5b (pink), and presence of SL5a-SL5c interaction (yellow, R1156 only). Angle between SL5-stem-SL5c and SL5a-SL5b for R1149 (B) and R1156 (D) colored by categories in (A) and (C) with the experimental structure range marked in pink; 0° is parallel orientation, 180° is an antiparallel, direction of rotation is defined from the view of F,H as moving SL5b clockwise. (E) For R1156 models, the bend angle of SL5a at the internal loop as measured by the angle between residues 24-27, 64-95 and residues 28-59. Three example models for R1149 (F-G) and R1156 (H-J) with SL5-stem in gray, SL5a in blue, SL5b in orange, and SL5c in red. (F,H) The predicted structures (dark) and the cryo-EM models (translucent) and GDT-TS score with rank over all models. (G,I) The predicted model's 4-way junction with arrows showing 5' to 3' direction, and (J) the SL5a-SL5c interaction.

Finally, we observed additional features in the BtCoV-HKU5 domain, (4) the apical loop of SL5c interacted with the internal loop of SL5a, and (5) SL5a bends at the internal loop with a continuous angular range spanning ~30-80°. Among the 15 models that exhibited correct junction stacking and base-pairing, only four modeled an interaction between SL5a and SL5c (defined as these two regions being within 3.5 Å), with only one, GeneSilico's R1156TS1128_5, correctly modeling the junction orientation (**Figure 4C**). While another top model by GDT-TS, R1156TS119_3 from the Kihara lab, did not predict this interaction, it was able to obtain the correct helical orientation and a bend in SL5a (**Figure 4H,J**). In contrast, R1156TS287_5 from the Chen group was able to model the SL5a-SL5c interaction and a bend in SL5a, yet, modeled a very different helical orientation outside the range of conformations captured by cryo-EM (**Figure 4H,J**). In general, all of the models predicting an interaction between SL5a and SL5c also predicted a bend in SL5a, with many falling within the experimental bend range (**Figure 4E**). Interestingly, a model for the SARS-CoV-2 SL5 domain, R1149TS035_5 from Manifold-E, included a SL5a-SL5c interaction and a helical bend in SL5a that was not observed by cryo-EM for that domain (**Figure 4F**). These models have motivated us to further investigate the relationship between the SL5a-SL5c interaction and the SL5a bend as well as improvements in experimental resolution for more precise experimental description of this interaction.

In summary, the prediction community successfully predicted the global topology, junction geometry, and other interactions of the two coronavirus SL5 domain targets. However, the wide range and uniform distribution of helical orientations in the models, which was not observed experimentally, (**Figure 4B,D**) suggests that selecting accurate models may be difficult. Despite the generally impressive performance of certain groups such as Ai chemistry_RNA2 (TS232) and GeneSilico (TS128), groups submitted a variety of topologies for both SL5 targets and the best predictions were not consistently ranked as model 1 of their 5 submissions. There is significant potential for improvement in both experimental determination and prediction, particularly for increasing accuracy and detail at junctions and other tertiary interactions that are critical for drug discovery efforts^{63,64}. A final challenge is predicting the ensemble of conformations, for example, to predict whether the range of SL5a bend angles are small, as in the SARS-CoV-2 domain, or larger, as in the BtCoV-HKU5 domain.

2.5 RNA-protein complex of RsmZ and RsmA (CASP: R1189 and R1190, PDB: 7YR7 and 7YR6). Provided by Bingnan Luo, Janusz Bujnicki, and Zhaoming Su

Pseudomonas aeruginosa (*P. aeruginosa*) is an opportunistic pathogen that infects hospitalized immunocompromised patients with high mortality rate⁶⁵. The acute and chronic virulence of *P. aeruginosa* could be regulated by type III and type VI secretion systems^{66,67}, biofilm formation⁶⁸, and quorum sensing, a cell density-based intercellular communication network⁶⁹. The repressor of secondary metabolite (Rsm) protein, RsmA, has been reported as a global regulator of gene expression related to acute and chronic virulence at both transcriptional and post-transcriptional levels^{70–74}. RsmZ is a small noncoding RNA that can bind to RsmA and modulate RsmA regulation⁷⁵. Previous studies showed that RsmA can form a homodimer to recognize two separate GGA binding sites^{76–79}, but the molecular mechanism of the full-length RsmZ sequestration of RsmA and regulation of *P. aeruginosa* virulence remains unknown.

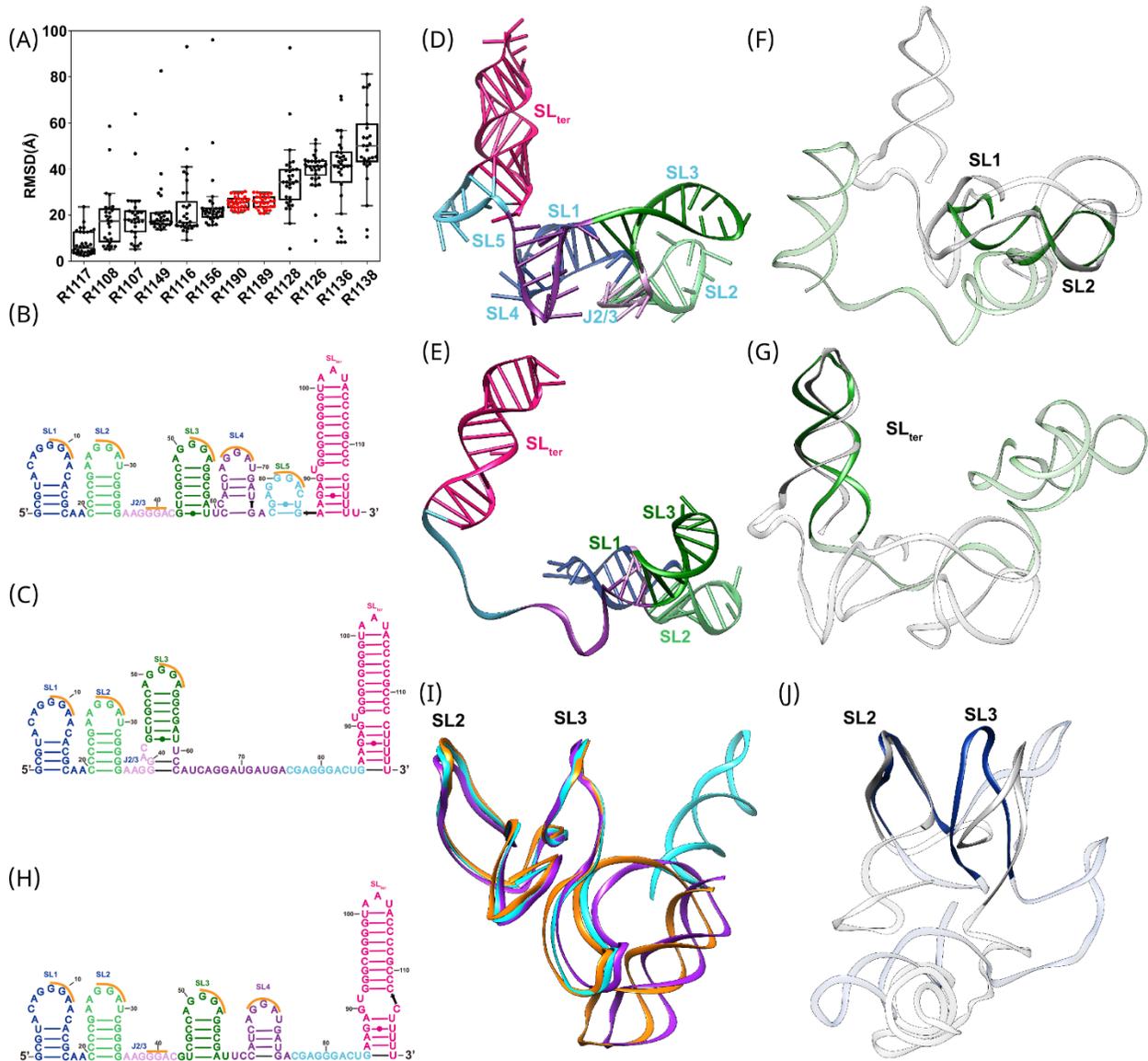


Figure 5. (A) Summary of the overall average RMSD of all predictions in each of the 12 RNA targets. (B) The experimental secondary structure of RsmZ with protein binding site GGA marked in orange. (C) The secondary structure of RsmZ predicted by Yang-server with protein binding site GGA marked in orange. (D) Cryo-EM model of RsmZ colored the same as the secondary structure. (E) Predicted model of RsmZ by Yang-server colored the same as the secondary structure. (F) Superposition of the cryo-EM (gray) and Yang-server predicted (green) RsmZ structures aligned on the stacked SL1-SL2. (G) Superposition of the cryo-EM (gray) and Yang-server predicted (green) RsmZ structures aligned on the longest SL_{ter}. (H) Secondary structure of top-ranked models by TM-score and GDT_TS of target R1189. (I) Superposition of all top-ranked models by TM-score and GDT_TS of target R1189 aligned on the protein binding site SL2 and SL3, with cyan from Venclovas, orange from CoDock, magenta from Kiharalab_Server. (J) Superposition of the RsmZ cryo-EM structure (gray) and a representative RsmZ model predicted by RNAPolis (blue) aligned on SL2 and SL3.

We obtained the RsmZ-A complex by incubation of *in vitro* transcribed full-length RsmZ with recombinantly expressed RsmA with a molar ratio of 1 to 4. The cryo-EM structure of RsmZ in complex with three RsmA homodimers (RsmZ-A₃) at 3.80 Å resolution (PDB: 7YR7 CASP:R1189 **Table 1**) showed that RsmZ comprises six consecutive stem-loops (SL1-SL5 and a terminator stem-loop SL_{ter}) with six GGA binding sites in the loop regions of SL1-SL5 and a single-stranded junction (J2/3) grouped into three pairs, SL1 and SL5, SL2 and SL3, J2/3 and SL4⁸⁰. In addition, we observed another conformation of RsmZ in complex with two RsmA homodimers (RsmZ-A₂) at 4.60 Å resolution (PDB: 7YR6 CASP:RT1190 **Table 1**), with the binding site between SL2 and SL3 unoccupied and a subtle change of 6.5° in SL_{ter}⁸⁰.

In CASP15, RNA and ribonucleoprotein (RNP) molecules were introduced into structural prediction and assessment for the first time, with both RNA and protein sequences and the binding stoichiometry provided. The best r.m.s. deviations (RMSD) of all predictions of each of the 12 RNA targets were all better than 10 Å, except for the RNA targets from the RsmZ-A₂ and RsmZ-A₃ complexes (R1189 and R1190) (**Figure 5A**). All predictions of the RNP targets had RMSD worse than 15 Å.

The top-ranked results by RMSD were generated by the Yang groups with an RMSD of 16.3 Å compared to RsmZ-A₃ (R1189TS229_3, R1189TS239_3, R1189TS439_3) and an RMSD of 16.0 Å compared to RsmZ-A₂ (R1190TS229_3, R1190TS239_3, R1190TS439_3), respectively. The difference that we first noticed was in RNA secondary structure (**Figure 5B-C**). While SL1, SL2, SL3 and SL_{ter} were accurately predicted, these predictions missed J2/3, SL4 and SL5. When comparing three-dimensional (3D) architectures, it was very challenging to align the entire RsmZ RNA structure (**Figure 5D-E**). Instead, we assessed the prediction results by aligning on either the stacking SL1-SL2 (nts 1-35, **Figure 5F**) or the longest SL_{ter} (nts 86-118, **Figure 5G**). In both assessments, we observed drastic deviations in the rest of the RNA structure.

A previous study on the RsmZ-E complex structure based on nuclear magnetic resonance and electron paramagnetic resonance from *P. fluorescens* revealed similar 3D architecture of a protein binding site consisting of SL2 and SL3 compared to our RsmZ-A complex structures from *P. aeruginosa*⁸⁰. Intriguingly, top-ranked TM-score and GDT_TS predictions from CoDock (R1189TS444_3), Venclovas (R1189TS494_3), Kiharalab_Server (R1189TS131_2) and

RNApolis (R1189TS081_1) for target R1189 generated a more accurate secondary structure with the inclusion of SL4 (**Figure 5H**), and successfully retained the architecture of protein binding site SL2 and SL3, albeit the RMSD values were slightly worse than the Yang groups predictions and the rest of protein binding sites were not accurately predicted (**Figure 5I-J**). In RNP prediction, the predicted RNA structures are generally closer to the experimental result when protein binding sites are taken into consideration. However, it seems that accurate predictions of all protein binding sites remain challenging, resulting in drastic differences of the overall RNA fold in the predicted structures compared to the experimental structures. Improved protein binding prediction will likely enable more accurate RNP structure prediction.

In conclusion, compared to protein 3D structure predictions, the current RNP 3D structure predictions are rather inaccurate and far from practical use in providing accurate structural information, which is likely caused by the paucity of experimentally determined RNP structures. Although machine learning algorithms have been reported to improve prediction accuracy using a small training dataset⁸¹, this improvement was not demonstrated in CASP15, and general applications and utilization of deep learning algorithms on RNP 3D structure prediction remain challenging since a much larger RNP 3D structure data set is required. This might be eventually overcome by continued advancement in RNP 3D structure determination and development in deep learning algorithms for smaller training datasets.

3 Conclusions

Here we provide insight into the functional and structural relevance of nine of the RNA targets of CASP15 from the perspective of the scientists who determined the experimental tertiary structures. These analyses complement the CASP assessors' comments^{TBD} with function-focused analysis, deeper focus on structural regions of importance, and comments on the utility of the current predictions for practical application in RNA structure research.

There were a few groups that consistently impressed. Ai chemistry_RNA2 was highlighted for their models of the CPEB3 ribozyme and RNA origami nanostructures; the Chen group submitted a very accurate model for the preQ₁ riboswitch aptamer; the GeneSilico group had especially high accuracy for the coronavirus SL5 structures; and while all groups were challenged by the RNA-protein complexes, several groups predicted accurate secondary structure. Topologically similar structures were obtained for all but the RNA-protein complexes. We will now summarize challenges from the experimentalists' perspectives to stimulate further advances.

While global topologies were predicted, there was a desire for improved accuracy in prediction of local interactions. The CPEB3 ribozyme analysis emphasized improvement in loop conformation prediction. The case of the preQ₁ riboswitch aptamer calls for increased accuracy in binding pocket prediction vital for informing the gene regulation of this element. The coronavirus structures showed inaccuracies in junction geometries.

The difficulty in predicting changes in RNA tertiary structure based on structure determination technique or condition also remains a challenge. For example, the design and prediction of the RNA origami structures had systematic inaccuracies because of the use of a kissing-loop structure from X-ray crystallography that in-fact was compacted in the solvated cryo-EM structure. With the largest RNA origami, no groups predicted close to an early kinetically trapped state, showing a gap in predicting structures along folding pathways. Further, in the case of the CPEB3 ribozymes, it was speculated that some of the errors made in modeling were because of not accounting for the dimeric state.

A final challenge, which has clearly not been overcome, was modeling RNA-protein complexes. Although some predictions generated quite accurate secondary structure and protein binding sites, the predicted 3D architectures of the RNA remain topologically different from the experimentally determined structure. Enabling the prediction of such complexes requires the combined expertise of protein, RNA, and multimer prediction groups as well as structure determination groups to provide the data that is currently lacking.

Despite room for improvement, particularly relative to the accuracies now enjoyed in the protein world, RNA predictions and experiments may be synergistic at this early stage. For example, the models can be used directly in the experimental structure determination. In the case of the preQ₁ riboswitch, the models allowed the structure determination from experimental X-ray data by molecular replacement. Elsewhere in this issue, the utility of models for molecular replacement and refinement into cryo-EM maps for all targets is discussed in more detail^{TBD}. Furthermore, the range of models submitted sparked questions about the limitations of traditional comparisons against one native structure. It is noted in both the RNA origami and the coronavirus sections that models not fitting the highest resolution experimental structure are not necessarily inaccurate, but may take on another state within the ensemble. As a future challenge, experimentalists, assessors, and predictors should emphasize analysis of RNA structures in solution or flash frozen from solution, with incorporation of structural dynamics and other heterogeneities like folding pathways that can now be captured by cryo-EM.

Overall, the RNA CASP15 experiment highlighted the utility of RNA tertiary structure predictors, but also the areas of improvement for predictors to support broader benefits. In its first iteration in CASP, the RNA structure community as a whole widely participated including the six groups that provided a total of 12 new RNA structures in the short three-month prediction season. Increased participation of experimentalists and predictors will continue to improve RNA tertiary structure prediction and its practical applications.

Acknowledgements

This work was supported by: Stanford BioX (Bowes Graduate Student Fellowship to R.C.K., Interdisciplinary Initiative Program to R.D.), the National Institutes of Health (R35 GM122579 to R.D., R01-GM06132 to J.E.W., T32-GM118283 to G.M.S.), Howard Hughes Medical Institute (to R.D.), the National Science Foundation (GRFP DGE-1656518 to L.X.), Ministry of Science and

Technology of China (2022YFC2303700 and 2021YFA1301900 to Z.S.), the National Natural Science Foundation of China (32222040 and 32070049 to Z.S.), the US National Institute of General Medical Sciences (NIGMS/NIH R01GM100482 to A.K.), the National Science Centre, Poland (NCN, 2017/26/A/NZ1/01083 to J.M.B.), Elon Huntington Hooker Fellowship (to G.M.S.), the Interdisciplinary Thematic Institute IMCBio of the ITI 2021-2028 program at the University of Strasbourg, CNRS and Inserm by IdEx Unistra (ANR-10-IDEX-0002 to B.M.), and under the framework of the French Investments Program for the Future EUR (IMCBio ANR-17-EUR-0023 to B.M.), Independent Research Fund Denmark (9040-00425B to E.S.A.), Novo Nordisk Foundation (NNF21OC0070452 to E.S.A.), the Canadian Natural Sciences and Engineering Research Council (532417 to E.K.S.M.), and the Houston Methodist Research Institute and the Center for RNA Therapeutics (E.K.S.M.). For the work of G.M.S and J.E.W., the use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, is supported by the U.S. Department of Energy, Office of Basic Energy Sciences [DE-AC02-76SF00515]; the SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences [P41 GM103393]. B.M. thanks Eric Westhof for critical reading of Section 2.1. 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 of this article can be made freely available under a CC BY 4.0 license immediately upon publication.

Author contributions

Concept, abstract, introduction, conclusion, editing, and coordination: R.C.K, A.K., and J.M.. Target-specific sections: by authors provided in the sections' titles.

Conflict of Interest

All authors declare that they have no competing interests.

References

- [1] Miao Z, Adamiak RW, Antczak M, Boniecki MJ, Bujnicki J, Chen SJ, et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA*. 2020;26(8):982-995.
- [2] Miao Z, Adamiak RW, Antczak M, Batey RT, Becka AJ, Biesiada M, et al. RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*. 2017;23(5):655-672.
- [3] Miao Z, Adamiak RW, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, et al. RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA

- structures. *RNA*. 2015;21(6):1066-1084.
- [4] Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, et al. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA*. 2012;18(4):610-625.
- [5] Alexander LT, Lepore R, Kryshtafovych A, Adamopoulos A, Alahuhta M, Arvin AM, et al. Target highlights in CASP14: Analysis of models by structure providers. *Proteins*. 2021;89(12):1647-1672.
- [6] Lepore R, Kryshtafovych A, Alahuhta M, Veraszto HA, Bomble YJ, Bufton JC, et al. Target highlights in CASP13: Experimental target structures through the eyes of their authors. *Proteins*. 2019;87(12):1037-1057.
- [7] Kryshtafovych A, Albrecht R, Baslé A, Bule P, Caputo AT, Carvalho AL, et al. Target highlights from the first post-PSI CASP experiment (CASP12, May-August 2016). *Proteins*. 2018;86 Suppl 1(Suppl 1):27-50.
- [8] Kryshtafovych A, Moult J, Baslé A, Burgin A, Craig TK, Edwards RA, et al. Some of the most interesting CASP11 targets through the eyes of their authors. *Proteins*. 2016;84 Suppl 1(Suppl 1):34-50.
- [9] Kryshtafovych A, Moult J, Bales P, Bazan JF, Biasini M, Burgin A, et al. Challenging the state of the art in protein structure prediction: Highlights of experimental target structures for the 10th Critical Assessment of Techniques for Protein Structure Prediction Experiment CASP10. *Proteins*. 2014;82 Suppl 2(0 2):26-42.
- [10] Kryshtafovych A, Moult J, Bartual SG, Bazan JF, Berman H, Casteel DE, et al. Target highlights in CASP9: Experimental target structures for the critical assessment of techniques for protein structure prediction. *Proteins*. 2011;79 Suppl 10(0 10):6-20.
- [11] Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res*. 2003;31(13):3370-3374.
- [12] Mariani V, Biasini M, Barbato A, Schwede T. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics*. 2013;29(21):2722-2728.
- [13] Magnus M, Antczak M, Zok T, Wiedemann J, Lukasiak P, Cao Y, et al. RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. *Nucleic Acids Res*. 2020;48(2):576-588.
- [14] Richter JD. CPEB: a life in translation. *Trends Biochem Sci*. 2007;32(6):279-285.
- [15] Lin CL, Huang YT, Richter JD. Transient CPEB dimerization and translational control. *RNA*. 2012;18(5):1050-1061.
- [16] Drisaldi B, Colnaghi L, Levine A, Huang Y, Snyder AM, Metzger DJ, et al. Cytoplasmic Polyadenylation Element Binding Proteins CPEB1 and CPEB3 Regulate the Translation of FosB and Are Required for Maintaining Addiction-Like Behaviors Induced by Cocaine. *Front Cell Neurosci*. 2020;14:207.
- [17] Ford L, Ling E, Kandel ER, Fioriti L. CPEB3 inhibits translation of mRNA targets by

- localizing them to P bodies. *Proc Natl Acad Sci U S A*. 2019;116(36):18078-18087.
- [18] Stephan JS, Fioriti L, Lamba N, Colnaghi L, Karl K, Derkatch IL, et al. The CPEB3 Protein Is a Functional Prion that Interacts with the Actin Cytoskeleton. *Cell Rep*. 2015;11(11):1772-1785.
- [19] Drisaldi B, Colnaghi L, Fioriti L, Rao N, Myers C, Snyder AM, et al. SUMOylation Is an Inhibitory Constraint that Regulates the Prion-like Aggregation and Activity of CPEB3. *Cell Rep*. 2015;11(11):1694-1702.
- [20] Salehi-Ashtiani K, Lupták A, Litovchick A, Szostak JW. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*. 2006;313(5794):1788-1792.
- [21] Chen CC, Han J, Chinn CA, Li X, Nikan M, Myszkka M, et al. The CPEB3 ribozyme modulates hippocampal-dependent memory. *bioRxiv*. Published online May 5, 2021:2021.01.23.426448. doi:10.1101/2021.01.23.426448
- [22] Ferré-D'Amaré AR. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. *Methods*. 2010;52(2):159-167.
- [23] Przytula-Mally AI, Engilberge S, Johannsen S, Olieric V, Masquida B, Sigel RKO. Anticodon-like loop-mediated dimerization in the crystal structures of HdV-like CPEB3 ribozymes. *bioRxiv*. Published online September 22, 2022:2022.09.22.508989. doi:10.1101/2022.09.22.508989
- [24] Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. A new understanding of the decoding principle on the ribosome. *Nature*. 2012;484(7393):256-259.
- [25] Rozov A, Demeshkina N, Westhof E, Yusupov M, Yusupova G. Structural insights into the translational infidelity mechanism. *Nat Commun*. 2015;6:7251.
- [26] Westhof E, Yusupov M, Yusupova G. Recognition of Watson-Crick base pairs: constraints and limits due to geometric selection and tautomerism. *F1000Prime Rep*. 2014;6:19.
- [27] Ferré-D'Amaré AR, Zhou K, Doudna JA. Crystal structure of a hepatitis delta virus ribozyme. *Nature*. 1998;395(6702):567-574.
- [28] Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. Genetic control by a metabolite binding mRNA. *Chem Biol*. 2002;9(9):1043.
- [29] McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. Riboswitch diversity and distribution. *RNA*. 2017;23(7):995-1011.
- [30] Breaker RR. Riboswitches and Translation Control. *Cold Spring Harb Perspect Biol*. 2018;10(11). doi:10.1101/cshperspect.a032797
- [31] Ariza-Mateos A, Nuthanakanti A, Serganov A. Riboswitch Mechanisms: New Tricks for an Old Dog. *Biochemistry*. 2021;86(8):962-975.
- [32] Babina AM, Lea NE, Meyer MM. In Vivo Behavior of the Tandem Glycine Riboswitch in *Bacillus subtilis*. *MBio*. 2017;8(5). doi:10.1128/mBio.01602-17

- [33] Breaker RR. The Biochemical Landscape of Riboswitch Ligands. *Biochemistry*. 2022;61(3):137-149.
- [34] McCown PJ, Liang JJ, Weinberg Z, Breaker RR. Structural, functional, and taxonomic diversity of three preQ₁ riboswitch classes. *Chem Biol*. 2014;21(7):880-889.
- [35] Schroeder GM, Cavender CE, Blau ME, Jenkins JL, Mathews DH, Wedekind JE. A small RNA that cooperatively senses two stacked metabolites in one pocket for gene control. *Nat Commun*. 2022;13(1):199.
- [36] Jenkins JL, Krucinska J, McCarty RM, Bandarian V, Wedekind JE. Comparison of a preQ₁ riboswitch aptamer in metabolite-bound and free states with implications for gene regulation. *J Biol Chem*. 2011;286(28):24626-24637.
- [37] Spitale RC, Torelli AT, Krucinska J, Bandarian V, Wedekind JE. The structural basis for recognition of the PreQ₀ metabolite by an unusually small riboswitch aptamer domain. *J Biol Chem*. 2009;284(17):11012-11016.
- [38] Kang M, Peterson R, Feigon J. Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol Cell*. 2009;33(6):784-790.
- [39] Klein DJ, Edwards TE, Ferré-D'Amaré AR. Cocrystal structure of a class I preQ₁ riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol*. 2009;16(3):343-344.
- [40] Roth A, Winkler WC, Regulski EE, Lee BWK, Lim J, Jona I, et al. A riboswitch selective for the queuosine precursor preQ₁ contains an unusually small aptamer domain. *Nat Struct Mol Biol*. 2007;14(4):308-317.
- [41] Schroeder GM, Dutta D, Cavender CE, Jenkins JL, Pritchett EM, Baker CD, et al. Analysis of a preQ₁-I riboswitch in effector-free and bound states reveals a metabolite-programmed nucleobase-stacking spine that controls gene regulation. *Nucleic Acids Res*. 2020;48(14):8146-8164.
- [42] Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66(Pt 2):213-221.
- [43] Geary C, Rothmund PWK, Andersen ES. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science*. 2014;345(6198):799-804.
- [44] Geary C, Grossi G, McRae EKS, Rothmund PWK, Andersen ES. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nat Chem*. 2021;13(6):549-558.
- [45] McRae EKS, Rasmussen HØ, Liu J, Bøggild A, Nguyen MTA, Sampedro Vallina N, et al. Structure, folding and flexibility of co-transcriptional RNA origami. *Nat Nanotechnol*. Published online February 27, 2023. doi:10.1038/s41565-023-01321-6
- [46] Sampedro Vallina N, McRae EKS, Hansen BK, Boussebayle A, Andersen ES. RNA origami scaffolds facilitate cryo-EM characterization of a Broccoli-Pepper aptamer FRET pair.

- [47] Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 2011;6:26.
- [48] Fornace ME, Huang J, Newman CT, Porubsky NJ, Pierce MB, Pierce NA. NUPACK: Analysis and Design of Nucleic Acid Structures, Devices, and Systems. *Chemrxiv.* Published online November 10, 2022. doi:10.26434/chemrxiv-2022-xv98l
- [49] Zadeh JN, Steenberg CD, Bois JS, Wolfe BR, Pierce MB, Khan AR, et al. NUPACK: Analysis and design of nucleic acid systems. *J Comput Chem.* 2011;32(1):170-173.
- [50] Sampedro Vallina N, McRae EKS, Geary C, Andersen ES. An RNA Paranemic Crossover Triangle as A 3D Module for Cotranscriptional Nanoassembly. *Small.* 2023;19(13):e2204651.
- [51] Ennifar E, Dumas P. Polymorphism of bulged-out residues in HIV-1 RNA DIS kissing complex and structure comparison with solution studies. *J Mol Biol.* 2006;356(3):771-782.
- [52] Chen SC, Olsthoorn RCL. Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs. *Virology.* 2010;401(1):29-41.
- [53] Rangan R, Zheludev IN, Hagey RJ, Pham EA, Wayment-Steele HK, Glenn JS, et al. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA.* 2020;26(8):937-959.
- [54] Manfredonia I, Nithin C, Ponce-Salvatierra A, Ghosh P, Wirecki TK, Marinus T, et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* 2020;48(22):12436-12452.
- [55] Sun L, Li P, Ju X, Rao J, Huang W, Ren L, et al. In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell.* 2021;184(7):1865-1883.e20.
- [56] Huston NC, Wan H, Strine MS, de Cesaris Araujo Tavares R, Wilen CB, Pyle AM. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol Cell.* 2021;81(3):584-598.e5.
- [57] Iserman C, Roden CA, Boerneke MA, Sealfon RSG, McLaughlin GA, Jungreis I, et al. Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol Cell.* 2020;80(6):1078-1091.e6.
- [58] Rangan R, Watkins AM, Chacon J, Kretsch R, Kladwang W, Zheludev IN, et al. De novo 3D models of SARS-CoV-2 RNA elements from consensus experimental secondary structures. *Nucleic Acids Res.* 2021;49(6):3092-3108.
- [59] Cheng CY, Kladwang W, Yesselman JD, Das R. RNA structure inference through chemical mapping after accidental or intentional mutations. *Proc Natl Acad Sci U S A.* 2017;114(37):9876-9881.
- [60] Kappel K, Zhang K, Su Z, Watkins AM, Kladwang W, Li S, et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods.* 2020;17(7):699-707.

- [61] Punjani A, Fleet DJ. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol.* 2021;213(2):107702.
- [62] Laing C, Wen D, Wang JTL, Schlick T. Predicting coaxial helical stacking in RNA junctions. *Nucleic Acids Res.* 2012;40(2):487-498.
- [63] Manigrasso J, Marcia M, De Vivo M. Computer-aided design of RNA-targeted small molecules: A growing need in drug discovery. *Chem.* 2021;7(11):2965-2988.
- [64] Childs-Disney JL, Yang X, Gibaut QMR, Tong Y, Batey RT, Disney MD. Targeting RNA structures with small molecules. *Nat Rev Drug Discov.* 2022;21(10):736-762.
- [65] Moradali MF, Ghods S, Rehm BHA. *Pseudomonas aeruginosa* Lifestyle: A Paradigm for Adaptation, Survival, and Persistence. *Front Cell Infect Microbiol.* 2017;7:39.
- [66] Hauser AR. The type III secretion system of *Pseudomonas aeruginosa*: infection by injection. *Nat Rev Microbiol.* 2009;7(9):654-665.
- [67] Suarez G, Sierra JC, Erova TE, Sha J, Horneman AJ, Chopra AK. A type VI secretion system effector protein, VgrG1, from *Aeromonas hydrophila* that induces host cell toxicity by ADP ribosylation of actin. *J Bacteriol.* 2010;192(1):155-168.
- [68] Høiby N, Ciofu O, Bjarnsholt T. *Pseudomonas aeruginosa* biofilms in cystic fibrosis. *Future Microbiol.* 2010;5(11):1663-1674.
- [69] Lee J, Zhang L. The hierarchy quorum sensing network in *Pseudomonas aeruginosa*. *Protein Cell.* 2015;6(1):26-41.
- [70] Goodman AL, Kulasekara B, Rietsch A, Boyd D, Smith RS, Lory S. A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in *Pseudomonas aeruginosa*. *Dev Cell.* 2004;7(5):745-754.
- [71] Allsopp LP, Wood TE, Howard SA, Maggiorelli F, Nolan LM, Wettstadt S, et al. RsmA and AmrZ orchestrate the assembly of all three type VI secretion systems in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A.* 2017;114(29):7707-7712.
- [72] Irie Y, Starkey M, Edwards AN, Wozniak DJ, Romeo T, Parsek MR. *Pseudomonas aeruginosa* biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA. *Mol Microbiol.* 2010;78(1):158-172.
- [73] Bröms JE, Forslund AL, Forsberg A, Francis MS. PcrH of *Pseudomonas aeruginosa* is essential for secretion and assembly of the type III translocon. *J Infect Dis.* 2003;188(12):1909-1921.
- [74] Casilag F, Lorenz A, Krueger J, Klawonn F, Weiss S, Häussler S. The LasB Elastase of *Pseudomonas aeruginosa* Acts in Concert with Alkaline Protease AprA To Prevent Flagellin-Mediated Immune Recognition. *Infect Immun.* 2016;84(1):162-171.
- [75] Brencic A, McFarland KA, McManus HR, Castang S, Mogno I, Dove SL, et al. The GacS/GacA signal transduction system of *Pseudomonas aeruginosa* acts exclusively through its control over the transcription of the RsmY and RsmZ regulatory small RNAs. *Mol Microbiol.* 2009;73(3):434-445.

- [76] Janssen KH, Diaz MR, Golden M, Graham JW, Sanders W, Wolfgang MC, et al. Functional Analyses of the RsmY and RsmZ Small Noncoding Regulatory RNAs in *Pseudomonas aeruginosa*. *J Bacteriol.* 2018;200(11). doi:10.1128/JB.00736-17
- [77] Duss O, Michel E, Yulikov M, Schubert M, Jeschke G, Allain FHT. Structural basis of the non-coding RNA RsmZ acting as a protein sponge. *Nature.* 2014;509(7502):588-592.
- [78] Schubert M, Lapouge K, Duss O, Oberstrass FC, Jelesarov I, Haas D, et al. Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat Struct Mol Biol.* 2007;14(9):807-813.
- [79] Morris ER, Hall G, Li C, Heeb S, Kulkarni RV, Lovelock L, et al. Structural rearrangement in an RsmA/CsrA ortholog of *Pseudomonas aeruginosa* creates a dimeric RNA-binding protein, RsmN. *Structure.* 2013;21(9):1659-1671.
- [80] Jia X, Pan Z, Yuan Y, Luo B, Luo Y, Mukherjee S, et al. Structural basis of sRNA RsmZ regulation of *Pseudomonas aeruginosa* virulence. *Cell Res.* 2023;33(4):328-330.
- [81] Townshend RJL, Eismann S, Watkins AM, Rangan R, Karelina M, Das R, et al. Geometric deep learning of RNA structure. *Science.* 2021;373(6558):1047-1051.