

RNA-Puzzles advances and perspectives

Zhichao Miao



Results

Puzzle	Name	Length (nt)	Best RMSD(Å)	DI all	INF all	INF wc	INF nwc	INF stacking	Clash Score
1	Thymidylate synthase mRNA	23	3.4	3.657	0.934	0.953	-	0.924	0
2	Self-assembling RNA square	15+10	2.3	2.846	0.808	0.875	0	0.798	11.76
3	Glycine Riboswitch	84	7.241	9.819	0.737	0.857	0	0.731	0
4	SAM-I riboswitch aptamer	126	3.257	3.472	0.938	0.989	0.667	0.935	2.22
5	Group I intron Gir1	188	9.152	12.019	0.761	0.906	0.334	0.751	6.79
6	Adenosylcobalamin riboswitch	168	11.699	16.151	0.724	0.885	0.316	0.702	23.48
7	Varkud satellite ribozyme	185	20.370	27.245	0.748	0.895	0.105	0.726	10.68
8	SAM-I/IV riboswitch	96	4.801	5.651	0.850	0.985	0.577	0.816	13.8
9	5-hydroxytryptophan aptamer		5.863	8.433	0.695	0.872	0.405	0.679	18.09
10	T-box complex	96+75	6.803	8.365	0.813	0.946	0.700	0.786	11.09
	T-box	96	6.021	7.478	0.805	0.906	0.624	0.804	10.73
	tRNA	75	2.505	2.767	0.905	0.972	0.913	0.883	1.45
11	7SK		5.224	6.456	0.809	0.953	0.000	0.821	0.550
12	ydaO riboswitch	108	10.061	14.453	0.696	0.861	0.000	0.663	12.61
13	ZMP riboswitch	60	5.410	7.032	0.769	0.905	0.258	0.755	10.85
14	L-glutamine riboswitch (Free state)	61	6.513	8.276	0.787	0.884	0.354	0.805	10.66
	L-glutamine riboswitch (Bound state)	61	5.879	7.926	0.742	0.918	0.577	0.701	11.18
15	hammerhead	68	7.121	10.410	0.684	0.728	0.224	0.710	94.72
16	TBA								
17	Pistol ribozyme	62	7.134	10.222	0.698	0.858	0.000	0.701	7.04
18	Zika virus	71	3.740	4.710	0.794	0.962	0.577	0.738	0
19	Twister sister ribozyme	62	5.517	7.480	0.738	0.909	0.000	0.703	18.97
20	TBA								
21	Guanidine III Riboswitch	41	4.506	7.702	0.585	0.957	0.250	0.482	13.53

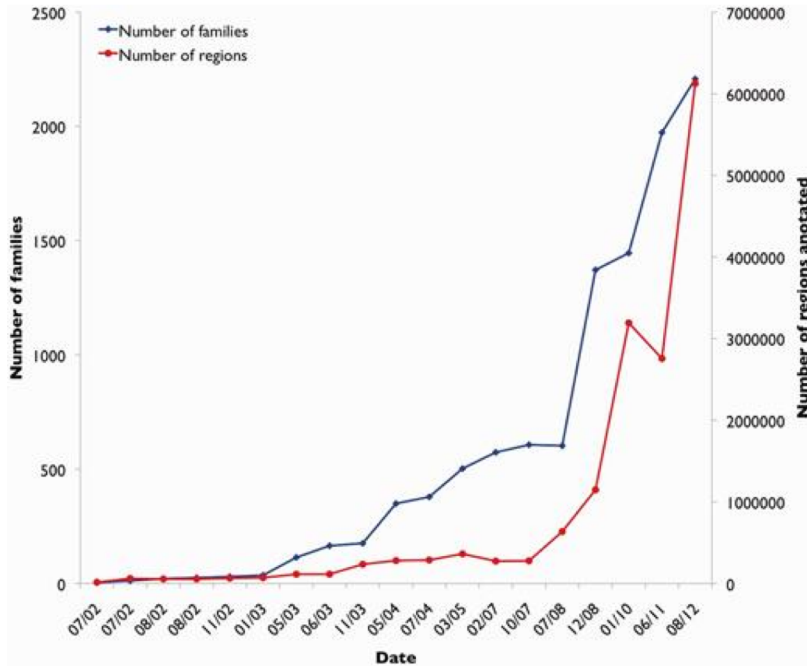
RNA-Puzzles toolkit

<http://rnabench.tk>

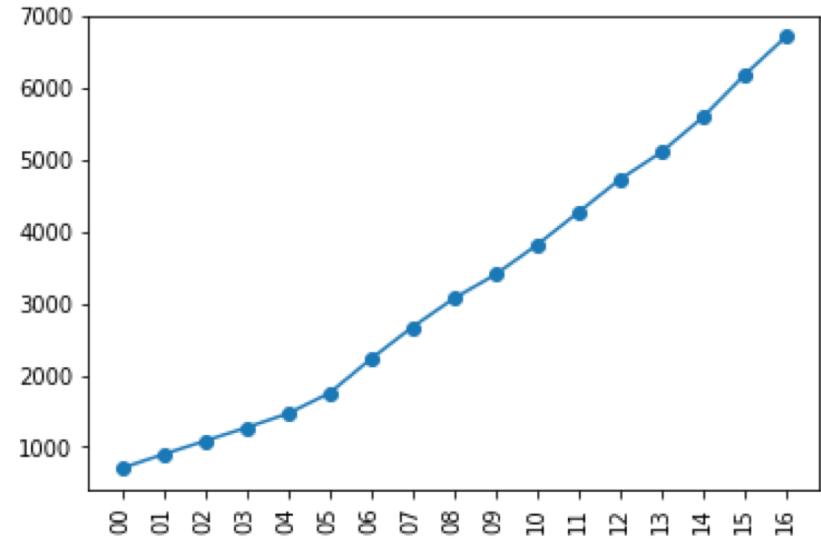
Metrics	Structure	Summary	Annotation
RMSD	All atoms	Atom-wise deviation based on rigid superimposition	General metric
LCS-score	Nucleotide centers	Largest set of Ca fall within a cutoff referred to experimental structure	new
MCQscore	Dihedrals	Dihedral angle deviations	RNA oriented
INF(INFwc, INFnwc, INFstacking)	Residue interactions	Matthews correlation coefficient to measure the accuracy in the prediction of the key interaction networks	
DI/DP	All atoms	Average distance deviations based on residue pair superimpositions	
MolProbity	All atoms	Geometrical reasonability of bonds, angles and contact distances	General metric
P-value	All atoms	Prediction quality based on structure length and RMSD	RNA oriented

Why RNA structure?

Rfam 11.0



#RNA structure in PDB



Sequence space vs. Structure space

The unknown Rfam puzzle project

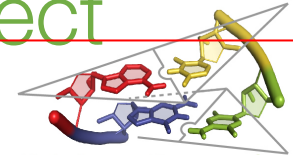
Family	Name	Type	Seed	Full	Species distribution(num_ssp)
RF01745	mscA RNA	cis-reg. (putative riboswitch)	189	351	Marine MG(332), V(10), B(5)
RF01723	TwoAYGGAY	cis-reg	210	904	B(211), Human MG(145), E(10)
RF01699	Cisactidines-1 RNA	sRNA	194	600	B(242), Gut MG(170), E(1)
RF01050	Molybdenum cofactor riboswitch	cis-reg. riboswitch	160	1196	B(933), MG(102), E(8)
RF03093	GOLL RNA	putative riboswitch	35	86	Marine MG(48), B(24), V(5), E(1)
RF01790	srf RNA	putative riboswitch	62	691	B(519), MG(34), E(1)
RF02344	Aspergillus/bacterial ar14	sRNA	119	398	B(263), MG(9), E(3)
RF02003	HNH1 endonuclease-associated RNA and ORF (HEARO) RNA	putative ribozyme	110	529	B(96), MG(72), A(19), E(9), V(1)
RF01709	Flavob-1 RNA	sRNA	201	660	MG(159), B(148), E(19)
RF01719	Phenacetolate-1 RNA motif	sRNA	18	29	MG(28), B(1)
RF00013	ES / Spd3 RNA	sRNA	149	3615	B(2815), MG(600), E(23)
RF00007	U12 minor spliceosomal RNA	sRNA	62	480	E(473)
RF00048	U11 spliceosomal RNA	sRNA	72	671	E(862), MG(3)
RF01748	msaW RNA	cis-reg	45	318	B(274), MG(36), E(3)
RF00031	Selenocysteine insertion sequence 1	cis-reg	61	865	E
RF02221	sRNA-Xcc1	sRNA	74	787	B(531), E(9), B(13)
RF00442	hMC-ynd leader	cis-reg. riboswitch	97	571	B(435), E(8)
RF01696	C4 antisense RNA	antisense	456	3639	B(2222), E(67), V(40)

→ 19/20 target RNA →



Online discussion

→



RNA-Puzzles

3D structure prediction

↓

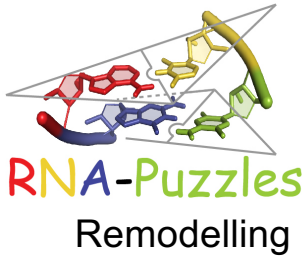
Quality prediction

←



Expert voting

←



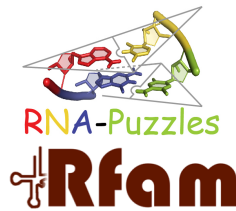
RNA-Puzzles Remodelling

↓



Experimental validation

→



Structure demonstration

→



Prediction web server

Papers to summarize

- RNA-Puzzles Round IV
- RNA-Puzzles toolkit
- Unknown Rfam Puzzles

Perspectives: Structure recruitment

- Adrian Ferre-D'Amare, Jinwei Zhang (NIH)
- Dinshaw Patel, Amy Ren (MSKCC, New-York)
- Thomas Hermann (UCSD)
- David M J Lilley (Dundee)
- Jeffrey S. Kieft (UC denver)
- Barbara L. Golden (Purdue)
- Robert T. Batey (ucolorado boulder)
- Joseph A. Piccinilli (uchicago)



List requirements

All items

Mandatory items

Navigation

- ✓ Instructions
- ✓ Communication
- ✓ Re-upload files
- ✓ Upload summary
- Admin
 - 🔴 Contact information
 - 🔴 Grant information
 - 🔴 **Release status**

Status information

Structure prediction competition ⓘ

Designate your submission as a prediction target if it has sequence identity to available structures below 50% or you expect it to be interesting for prediction otherwise.

Please identify if this entry is part of a target prediction analysis

Do you wish this entry to become a prediction target?: ⓘ N Y

Type of assessment: ⓘ CAPRI CASD-NMR CASP D3R FoldIt GPCR Dock RNA-Puzzles

Release instructions

Specific processing details, hold requests and entry replacement information

Making RNA-Puzzles useful?

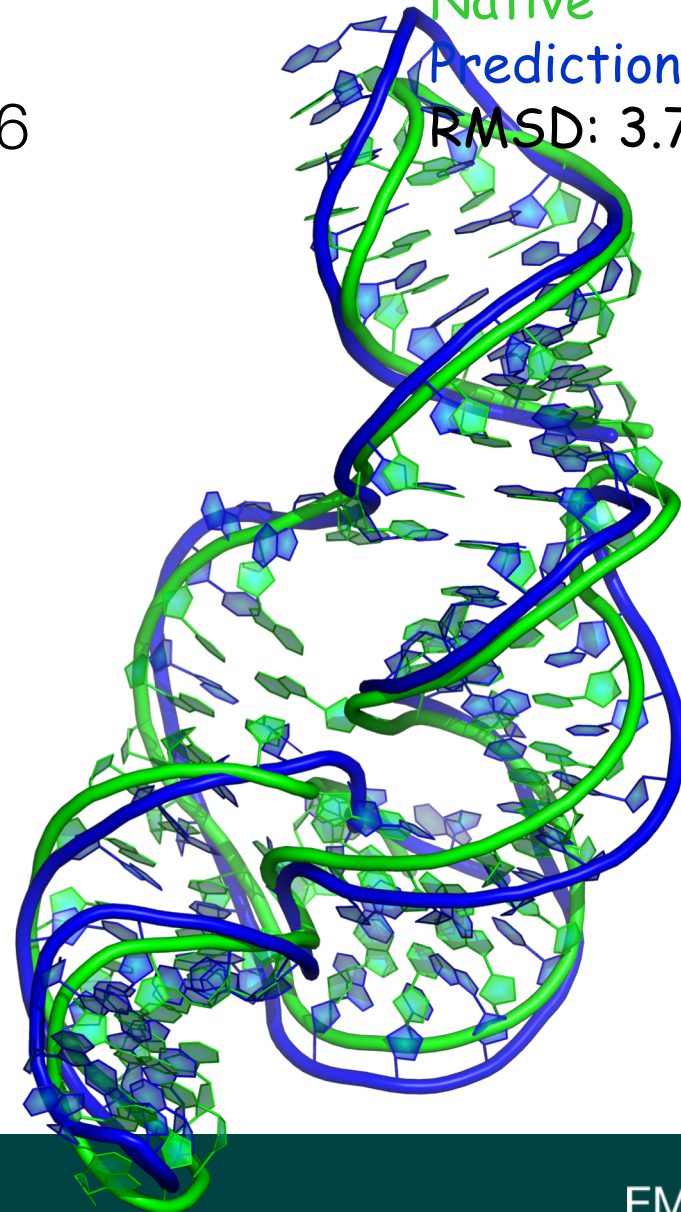
- Zika virus break out in 2016
- Shall we endeavor a prediction puzzle without knowing the solved structure when RNA virus breaks out?

DI all: 4.71
INF all: 0.794
INF wc: 0.962
INF nwc: 0.577
INF stack: 0.738
MCQ: 18.49

Puzzle 18: Zika Virus

Native
Prediction

RMSD: 3.74Å



Web server? Domain name?



RNA-Puzzles

is a collective experiment for blind RNA structure prediction.

Aims of the project

- To determine the capabilities and limitations of current methods of 3D RNA structure prediction based on sequence;
- To find whether and how progress has been made, as well as what has yet to be done to achieve better solutions
- To identify whether there are specific bottlenecks that hold back the field;
- To promote the available methods and guide potential users in the choice of suitable tools for real-world problems;
- To encourage the RNA structure prediction community in their efforts to improve the current tools and to make automated prediction tools available.
- To explore the underlying mechanism of ligand-RNA binding and the conformational changes in riboswitches.

Basic Rules

- Confidential agreement: "I will use the provided sequence data for prediction purposes only. I will not distribute the sequences in any form outside my working group."
- It is strongly recommended that all the predictions formatted in standard **PDB format**. So that all the modellers could have a fair comparison without considering the file format and prediction region. We provide a python script [here](#) for generating standard PDB format files, only the coordinates need to be replaced by the predictors.
- RNA chains should be named in alphabetical order. e.g. A for the first chain, B for the second chain.
- Up to **FIVE** prediction models could be predicted as the final results of one problem for each group. It is also recommended to submit ten models.
- Please put all prediction models together in one PDB file as NMR format (Separate models by lines: "MODEL n" and "ENDMDL").
- For submission, please either use our online [submission](#) system or send us an [email](#).

Call for participants

- We hope more crystallographers could help us in providing more crystal, cryo-EM or NMR structures. Any kind of RNA related

RNA-Puzzles

A CASP-like evaluation of RNA three-dimensional structure prediction



Recent Posts

New RNA-Puzzles Website at Github

Puzzle 21

Puzzle 20

Puzzle 19

Puzzle 18

GitHub Repos

Assessments

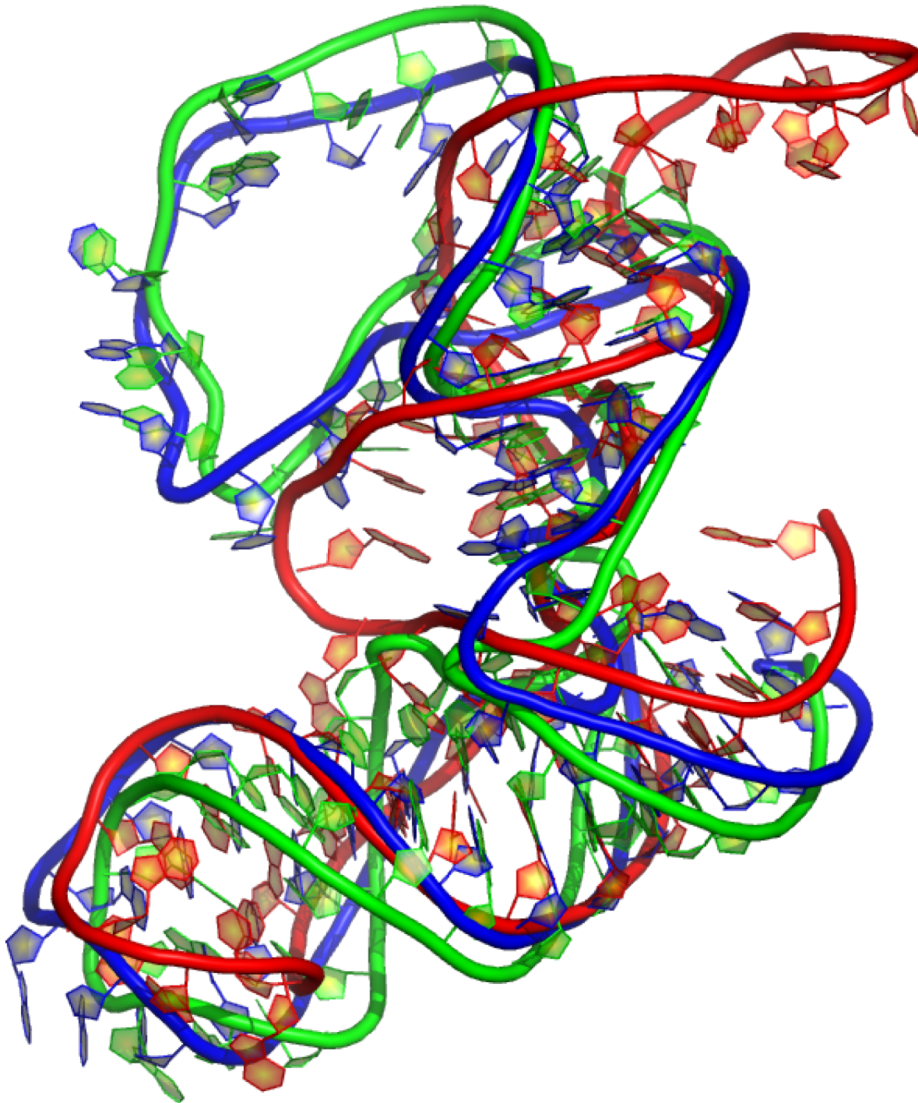
Assessment metrics tools to compare RNA 3D structures.

RNA-Puzzles_dataset

This repository includes all the submitted RNA structures in RNA-Puzzles.



Why WC not enough?



- Puzzle 14 (L-glutamine)
- Native
- w/ non-WC + 98% WC
- w/o non-WC + 100% WC