RNA Structure: CASP perspective

Rhiju Das, Rachael Kretsch, Phillip Pham, Ramya Rangan

12 December, 2022
CASP15 conference
Antalya, Turkey

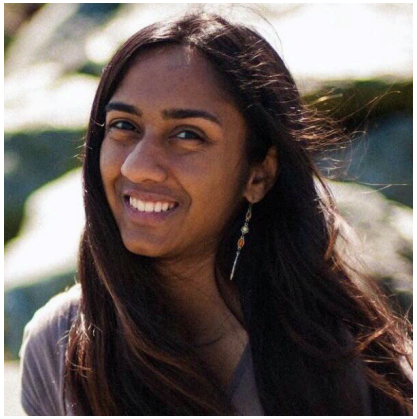
Stanford
MEDICINE

hhmi
Howard Hughes
Medical Institute

@rdaslab

Rachael Kretsch*

Phillip Pham



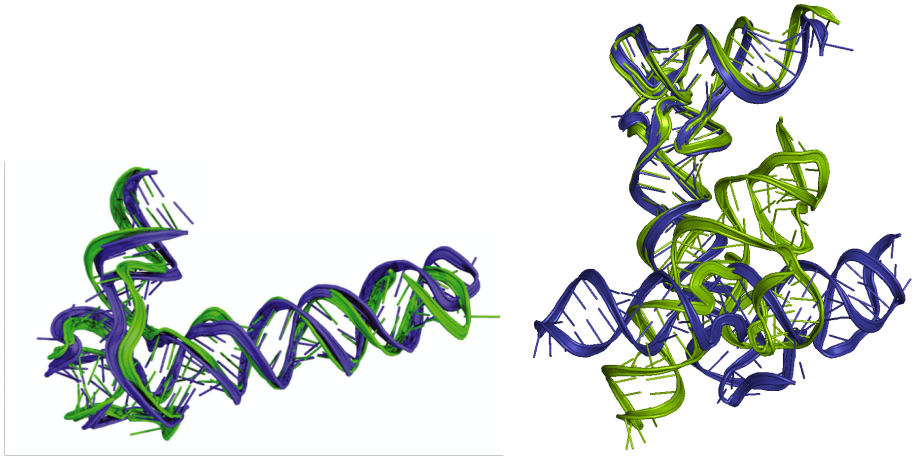
Ramya Rangan

We’re predictors, experimentalists, and (now) assessors.

Thanks: Andriy Kryshtafovych, Krzystof Fidelis, John Moult

An RNA category in CASP15

3D **RNA deep learning** in its nascency



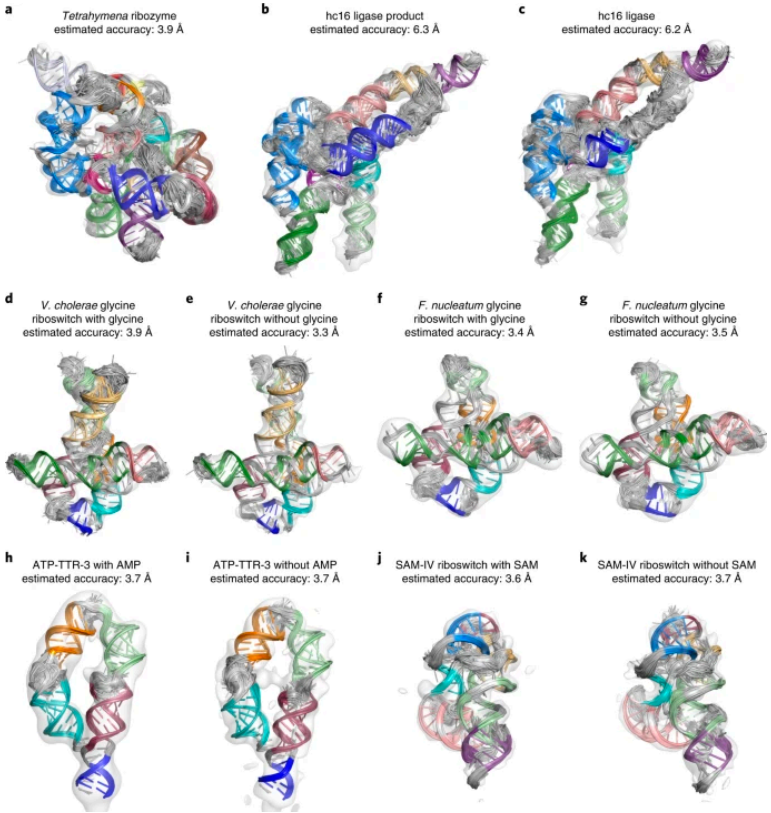
RNA Puzzle 24

RNA Puzzle 28

Townshend, Eismann, Watkins, …, Das, Dror, Science, 2021

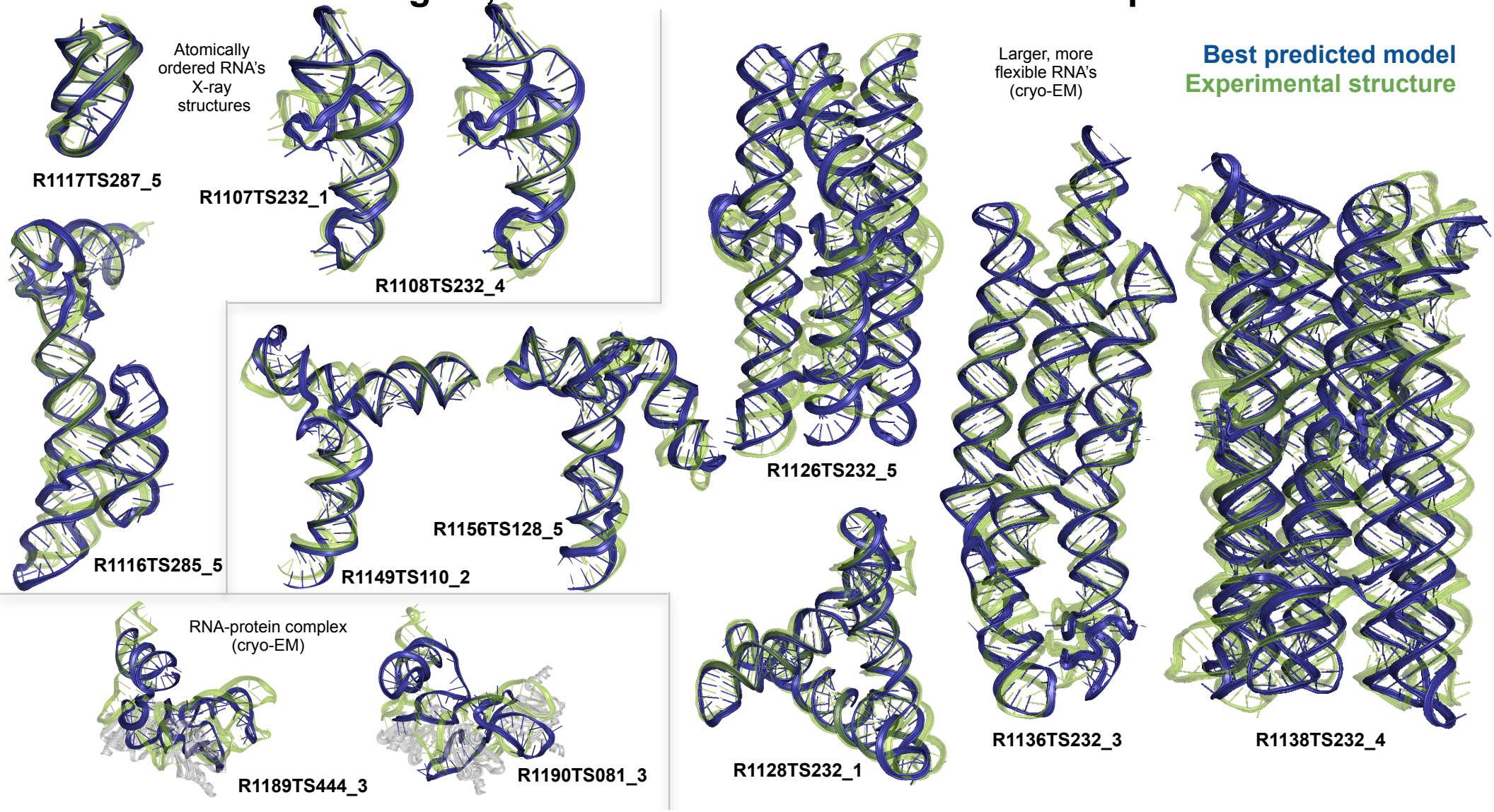
Thanks, Eric Westhof, Chichau Miao and RNA Puzzles community

More targets: increasing throughput of **RNA Cryo-EM**



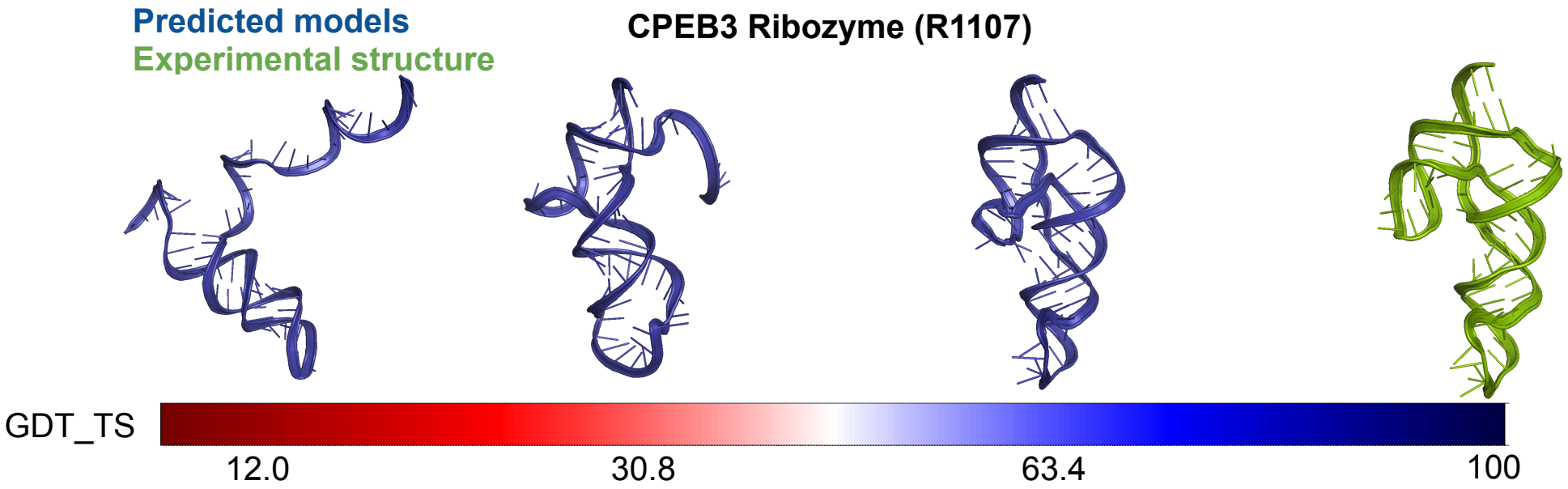
Kappel, Zhang, Su, …, Chiu, Das, Nat. Methods, 2020

CASP15: 12 targets, excellent models from a field of 40 predictors



What do CASP-style quantitative rankings say?

Piloting GDT for RNA



Thanks, Adam Zemla

CASP15 RNA assessment metric

“Topology” **“Local environment”** **Stereochemical quality**

$$Z_{RNA} = \frac{1}{3}[Z_{TM} + Z_{GDT-TS}] + \frac{1}{8}[Z_{INF} + Z_{lDDT}] + \frac{1}{12}Z_{clash}$$

where Z-score = number of standard deviations from the mean

(after filtering out of poor models with initial Z < -2)

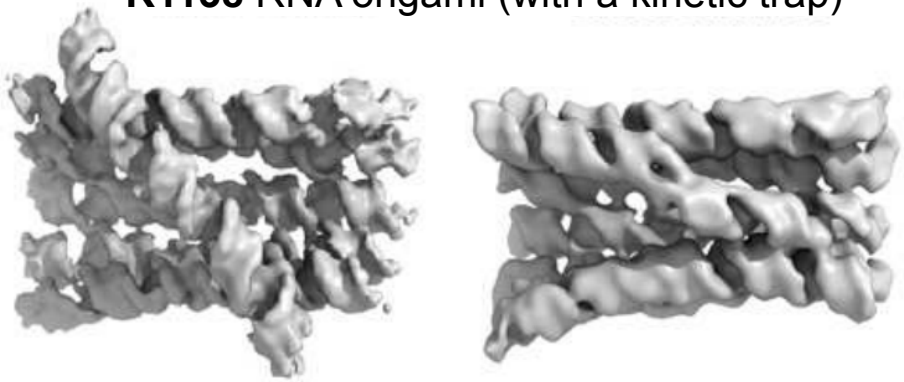
Used in prior CASP topology assessment

Used in (non-CASP) RNA assessment

Thanks: Gabriel Studer, Marcin Magnus, Chengxin Zhang, Marta Szachniuk, Maciej Antczak

How we assessed multi-state RNA's

R1138 RNA origami (with a kinetic trap)



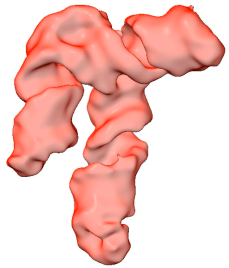
Compare:

all five predictor models

vs.

all available experimental models

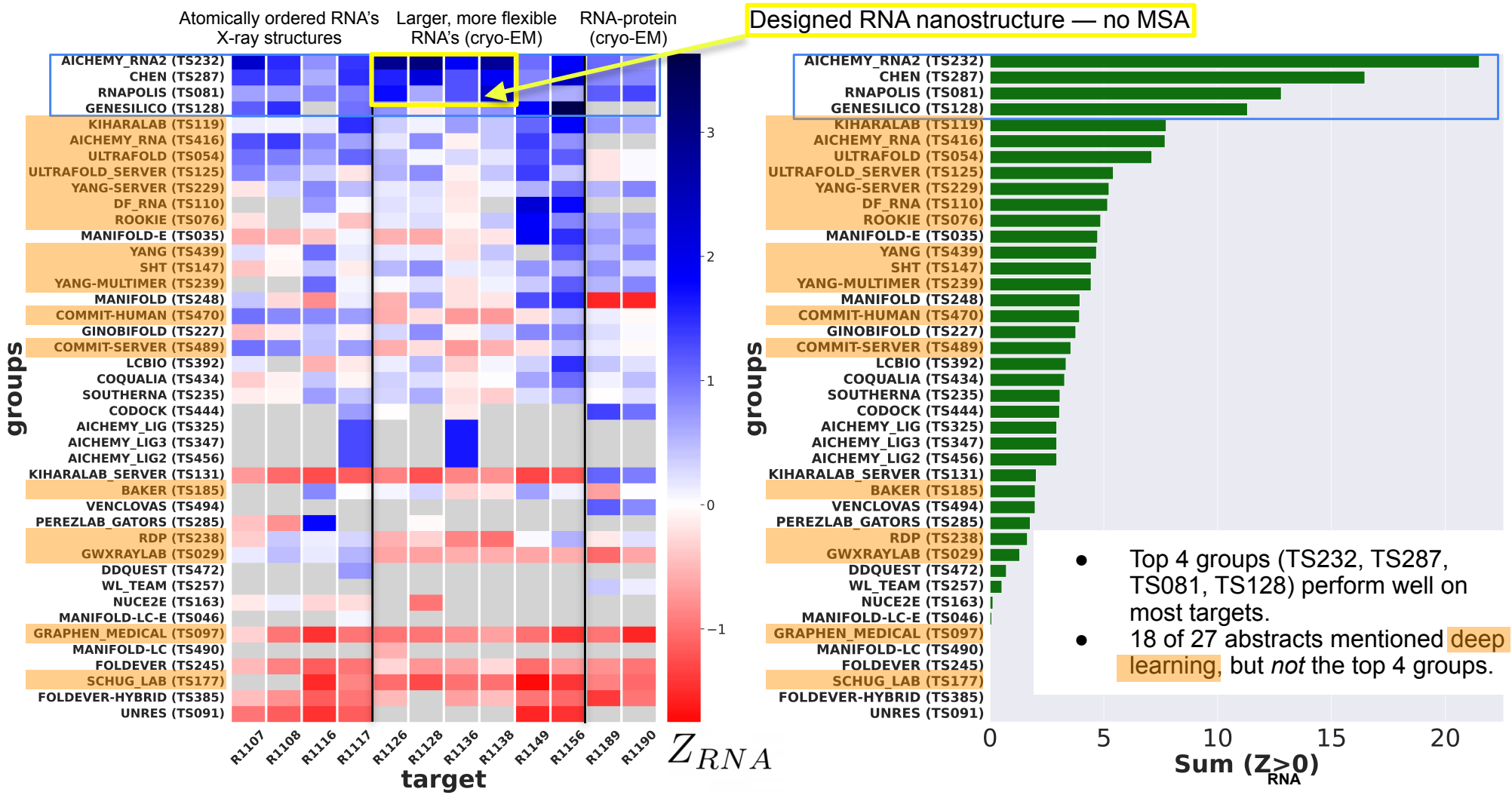
R1156 Bt-CoV-HKU5 SL5 domain



Reward predictor based on *best* score.

Thanks: Lisa Kinch, Nick Grishin

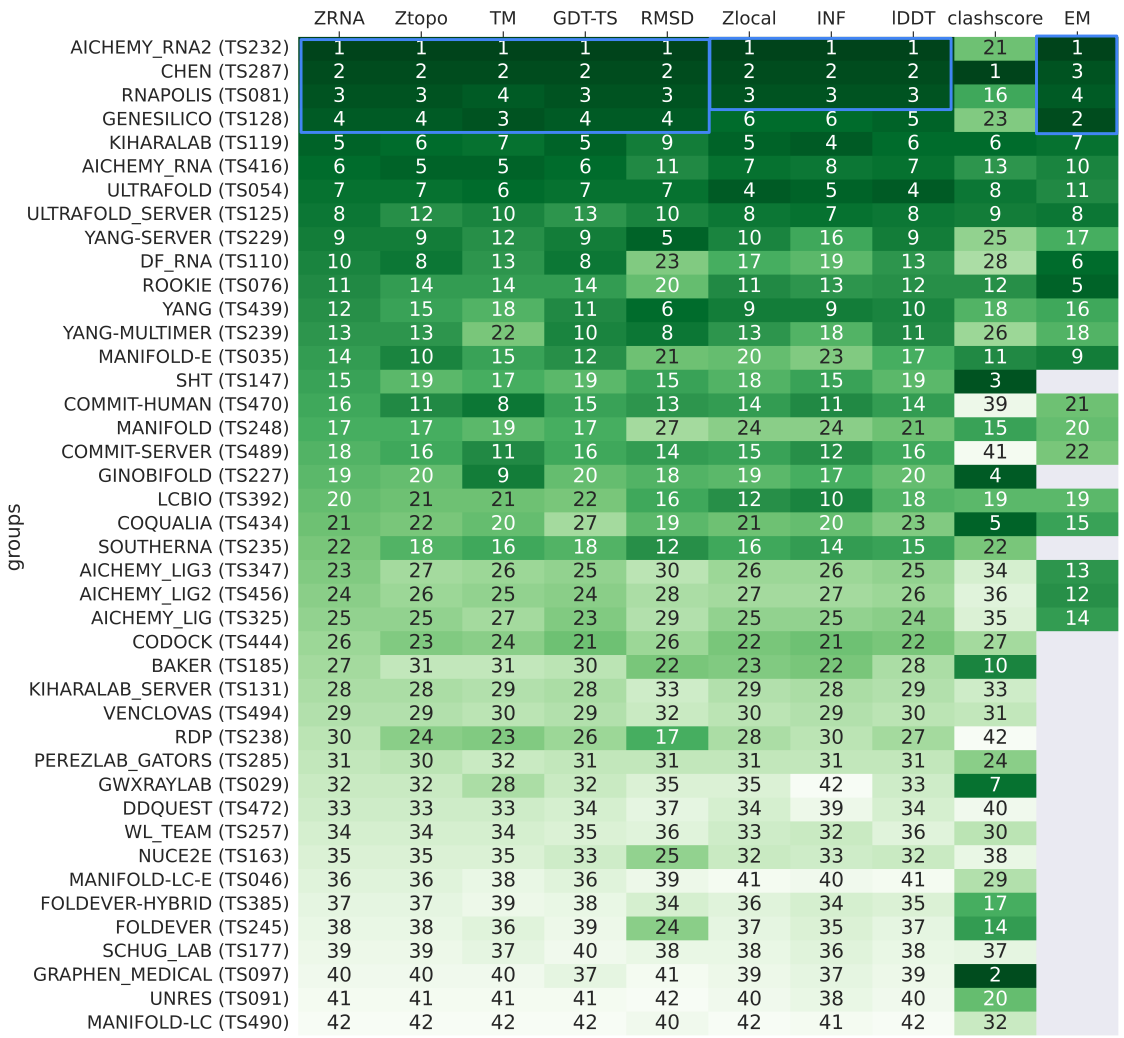
Rankings from CASP15 RNA metric



Let's double-check our rankings

- Topological scores all consistent in ranking top four groups
- In local environment scores, top 3 remain top 3; after that ranking is different from topological scores
- Some top rankers were not the cleanest in terms of clash score, but were also not the large negative outliers
- Comparisons directly to EM maps (to avoid experimental model bias) give same top 4 ranking.

$$Z_{EM} = \frac{1}{5}[Z_{CCmask} + Z_{CCpeaks} + Z_{SMOC} + Z_{MI} + Z_{AI}]$$



groups	ZRNA	Ztopo	TM	GDT-TS	RMSD	Zlocal	INF	lDDT	clashscore	EM
AICHEMY_RNA2 (TS232)	1	1	1	1	1	1	1	1	21	1
CHEN (TS287)	2	2	2	2	2	2	2	2	1	3
RNAPOLIS (TS081)	3	3	4	3	3	3	3	3	16	4
GENESILICO (TS128)	4	4	3	4	4	6	6	5	23	2
KIHARALAB (TS119)	5	6	7	5	9	5	4	6	6	7
AICHEMY_RNA (TS416)	6	5	5	6	11	7	8	7	13	10
ULTRAFOLD (TS054)	7	7	6	7	7	4	5	4	8	11
ULTRAFOLD_SERVER (TS125)	8	12	10	13	10	8	7	8	9	8
YANG-SERVER (TS229)	9	9	12	9	5	10	16	9	25	17
DF_RNA (TS110)	10	8	13	8	23	17	19	13	28	6
ROOKIE (TS076)	11	14	14	14	20	11	13	12	12	5
YANG (TS439)	12	15	18	11	6	9	9	10	18	16
YANG-MULTIMER (TS239)	13	13	22	10	8	13	18	11	26	18
MANIFOLD-E (TS035)	14	10	15	12	21	20	23	17	11	9
SHT (TS147)	15	19	17	19	15	18	15	19	3	
COMMIT-HUMAN (TS470)	16	11	8	15	13	14	11	14	39	21
MANIFOLD (TS248)	17	17	19	17	27	24	24	21	15	20
COMMIT-SERVER (TS489)	18	16	11	16	14	15	12	16	41	22
GINOBIFOLD (TS227)	19	20	9	20	18	19	17	20	4	
LCBIO (TS392)	20	21	21	22	16	12	10	18	19	19
COQUALIA (TS434)	21	22	20	27	19	21	20	23	5	15
SOUTHERNA (TS235)	22	18	16	18	12	16	14	15	22	
AICHEMY_LIG3 (TS347)	23	27	26	25	30	26	26	25	34	13
AICHEMY_LIG2 (TS456)	24	26	25	24	28	27	27	26	36	12
AICHEMY_LIG (TS325)	25	25	27	23	29	25	25	24	35	14
CODOCK (TS444)	26	23	24	21	26	22	21	22	27	
BAKER (TS185)	27	31	31	30	22	23	22	28	10	
KIHARALAB_SERVER (TS131)	28	28	29	28	33	29	28	29	33	
VENCLOVAS (TS494)	29	29	30	29	32	30	29	30	31	
RDP (TS238)	30	24	23	26	17	28	30	27	42	
PEREZLAB_GATORS (TS285)	31	30	32	31	31	31	31	31	24	
GWXRAYLAB (TS029)	32	32	28	32	35	35	42	33	7	
DDQUEST (TS472)	33	33	33	34	37	34	39	34	40	
WL_TEAM (TS257)	34	34	34	35	36	33	32	36	30	
NUCE2E (TS163)	35	35	35	33	25	32	33	32	38	
MANIFOLD-LC-E (TS046)	36	36	38	36	39	41	40	41	29	
FOLDEVER-HYBRID (TS385)	37	37	39	38	34	36	34	35	17	
FOLDEVER (TS245)	38	38	36	39	24	37	35	37	14	
SCHUG_LAB (TS177)	39	39	37	40	38	38	36	38	37	
GRAPHEN_MEDICAL (TS097)	40	40	40	37	41	39	37	39	2	
UNRES (TS091)	41	41	41	41	42	40	38	40	20	
MANIFOLD-LC (TS490)	42	42	42	42	40	42	41	42	32	

Thanks: Maya Topf, Tom Mulvaney, Andriy Kryshtafovych

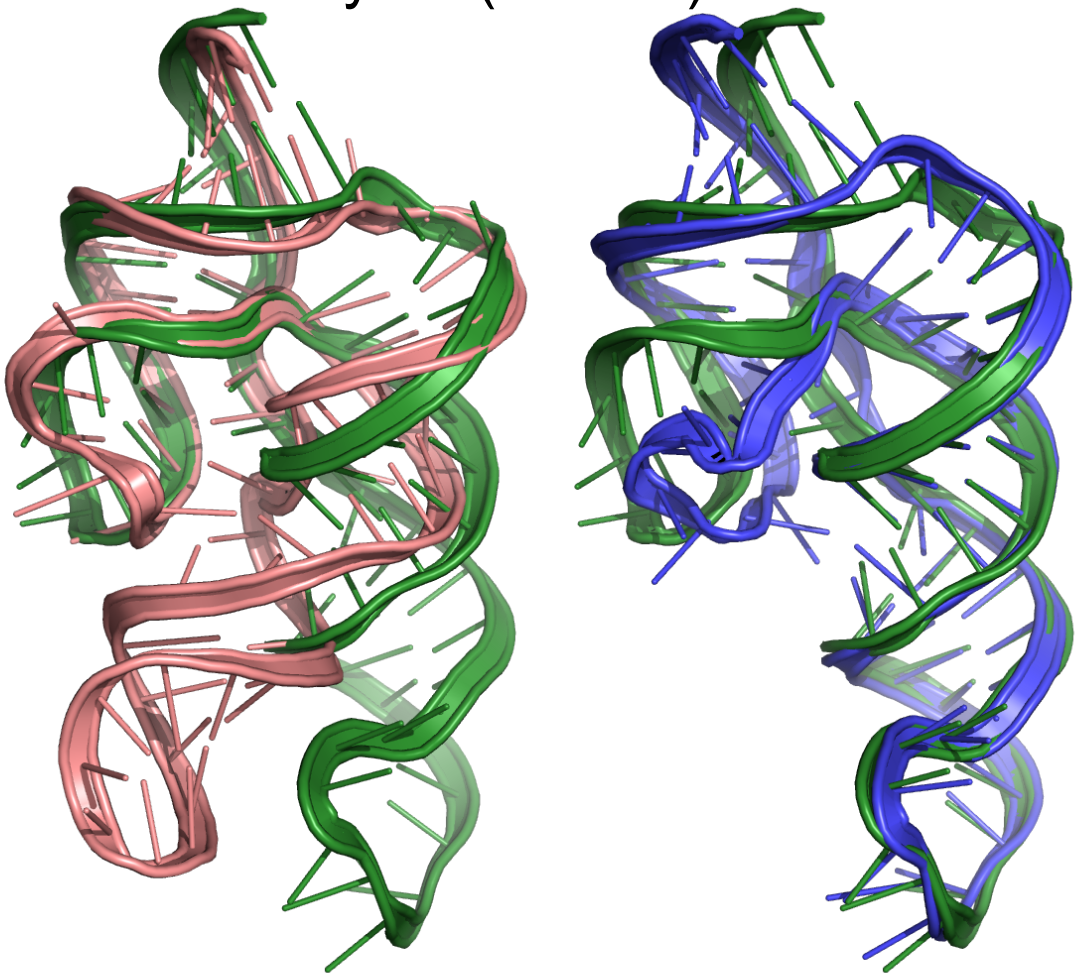
Three highlights and one challenge case

R1107 CPEB3 ribozyme (human)

Rosetta template-based models 1
(Prepared by Ramya in 2019)

Crystal structure **(of dimer)**

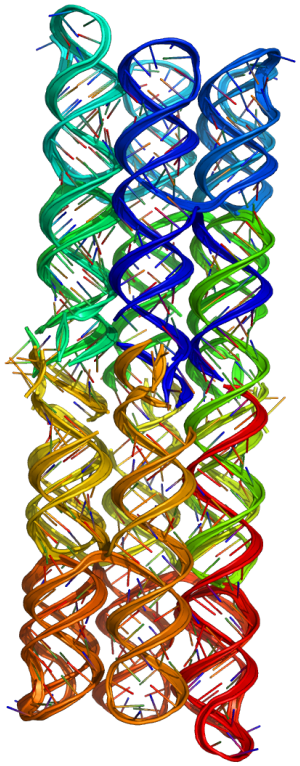
AIchemy-RNA2 (TS232) Model 1



Przytula-Mally et al., bioRxiv (2022)

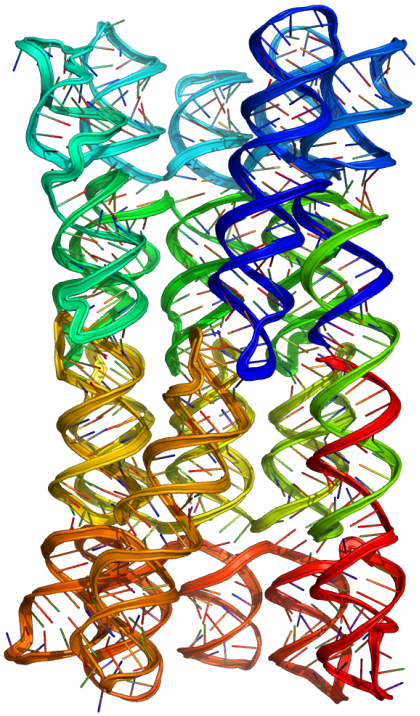
Thanks: Masquida, Sigel groups; Eric Westhof

R1138 Six-helix bundle RNA origami

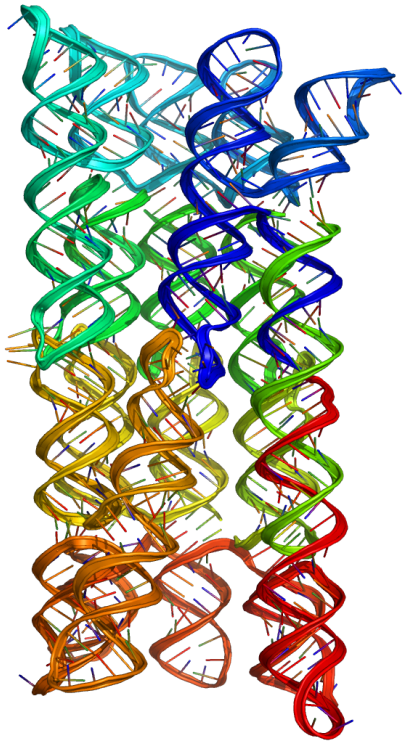


Design

TM-score 0.623

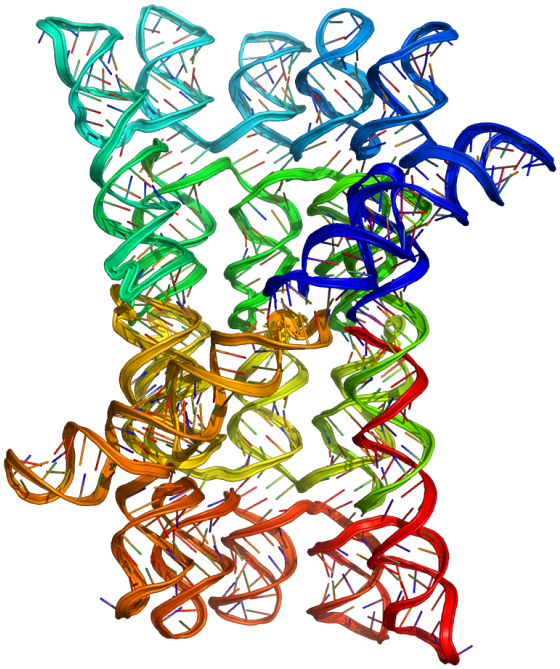


Cryo-EM structure



Alchemy-RNA2 (TS232) Model 4

TM-score 0.800

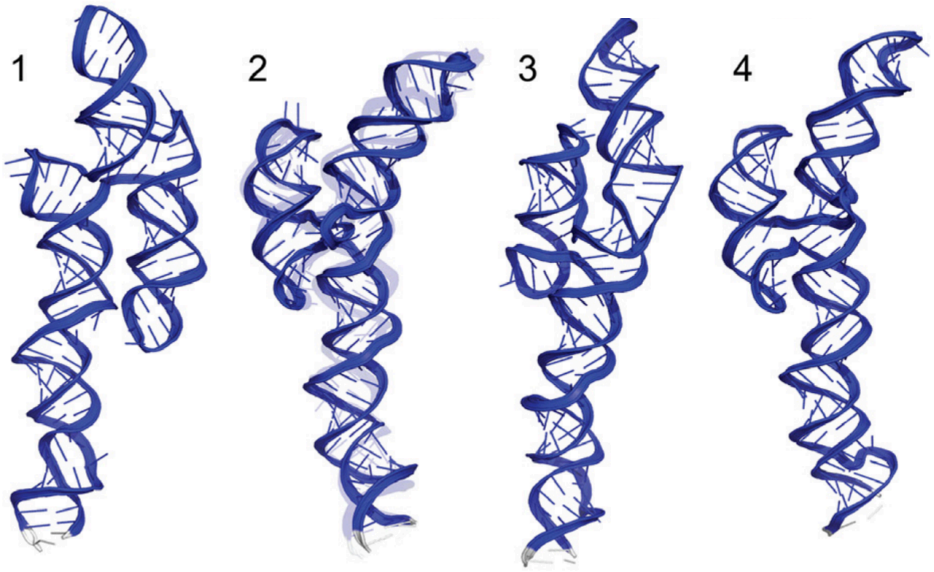


Another resolved Cryo-EM structure

[no CASP models were within TM-score of 0.63]

Thanks: Ebbe Anderson and colleagues

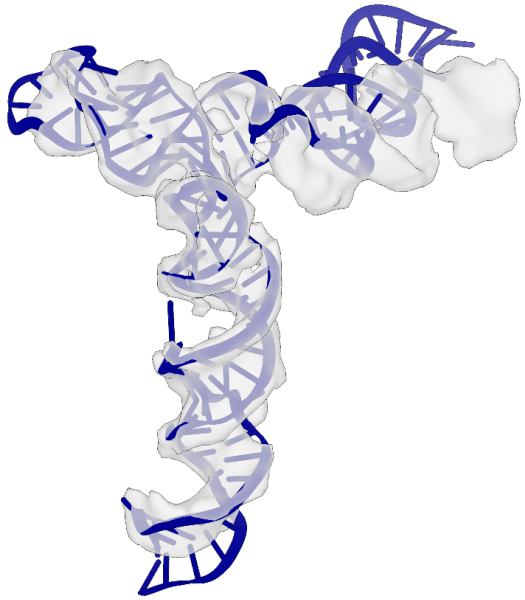
R1149 SARS-CoV-2 SL5 domain



Prior Rosetta FARFAR2 modeling suggested
no well-defined 3D structure for this sequence

Rangan et al., NAR (2021)

GeneSilico (TS128) Model 1



The domain *is* resolvable by cryo-EM

A “T-shape” not an “H-shape”

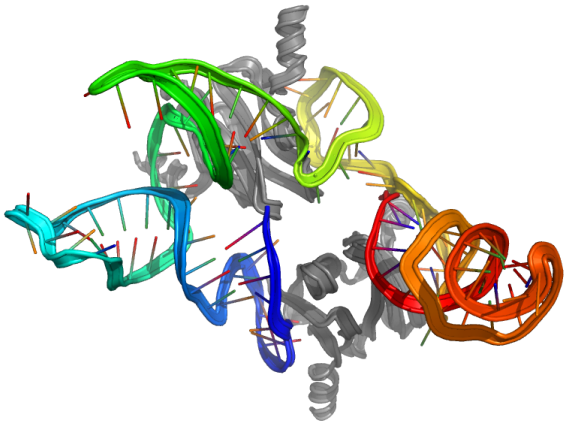
Fold captured by some CASP predictors

Similar for R1156 (more flexible bat CoV homolog)

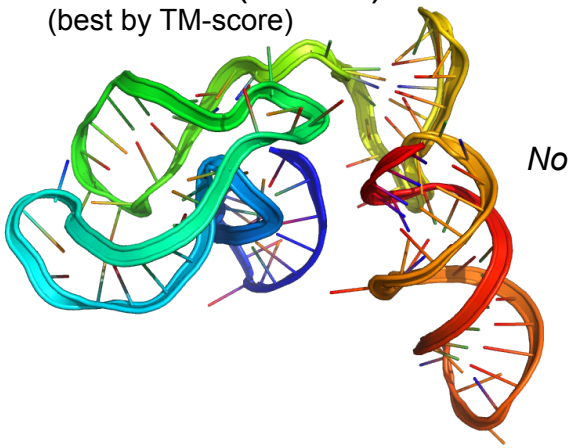
Thanks: Rachael, Wah Chiu and collaborators

R1190 CsrA RNA-protein complex

Cryo-EM structure



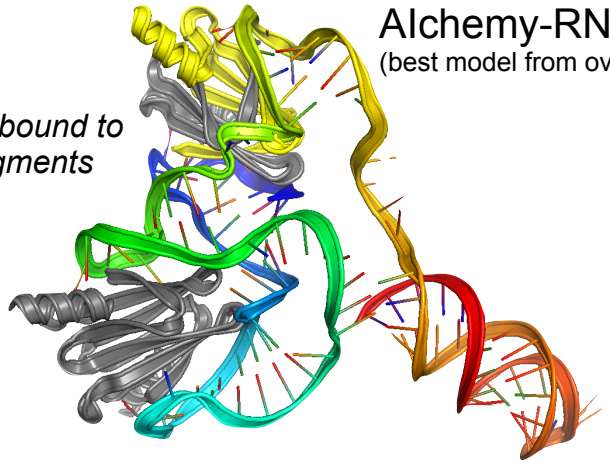
RNAPolis (TS081) Model 5
(best by TM-score)



No proteins submitted

Alchemy-RNA2 (TS232) Model 5
(best model from overall top RNA group)

Protein dimers bound to wrong RNA segments



Similar problems predicting R1189 (same RNA with an additional protein dimer)

Thanks, Zhaoming Su

In an absolute sense, how did CASP15 RNA modelers do?

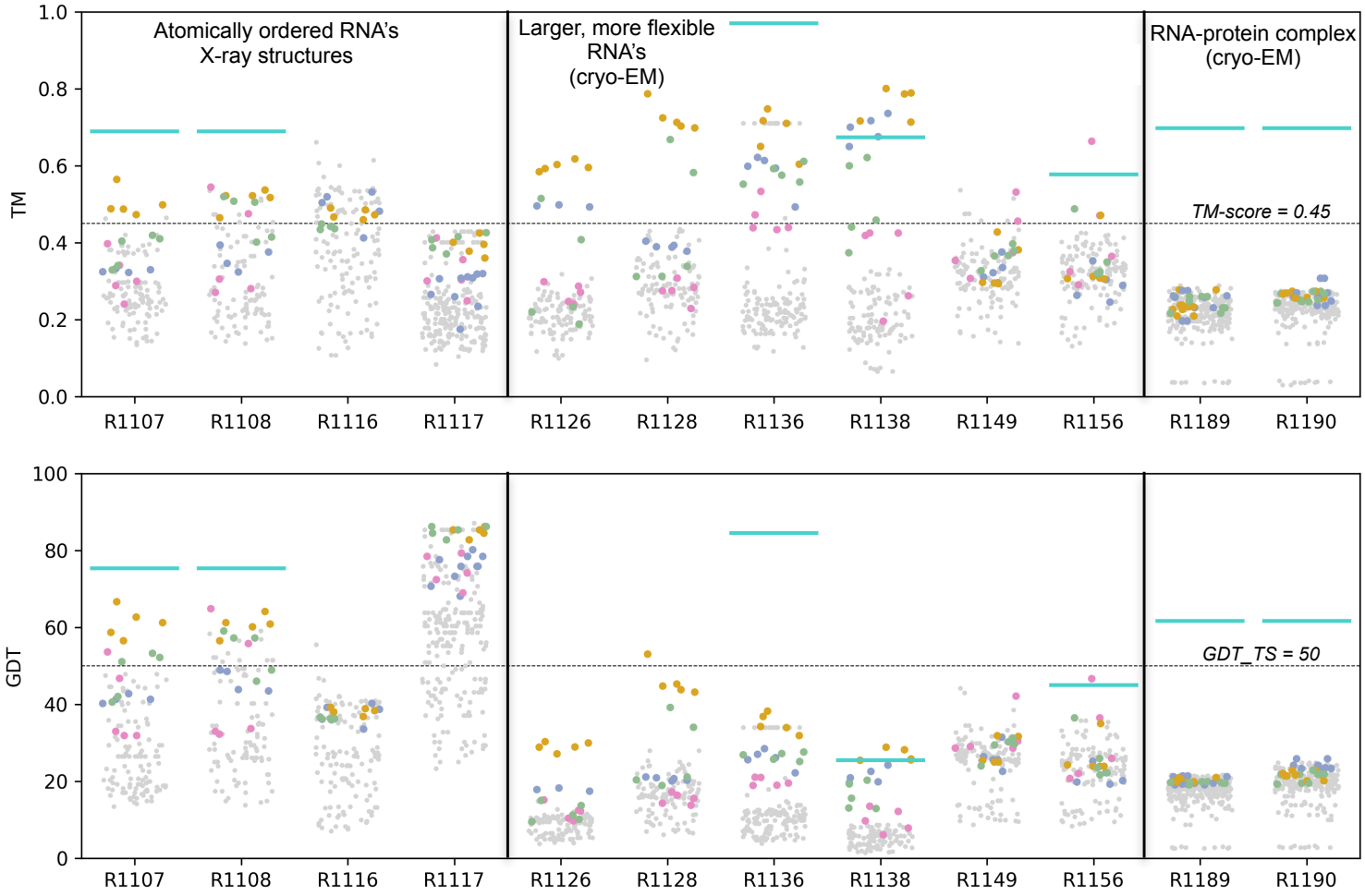
TM-score > 0.45 means ‘matching template’ (for RNA)

S. Gong, C. Zhang, and Y. Zhang, Bioinformatics (2019)

GDT_TS > 50 means ‘correct topology’ (for proteins)

CASP9 (Kinch et al.), CASP10 (Tai et al.), CASP11 (Abriata et al.)…
[CASP topology assessment papers]

GDT and TM summary



- AICHEMY_RNA2 (TS232)
- CHEN (TS287)
- RNAPOLIS (TS081)
- GENESILICO (TS128)
- Other groups

- All 10 RNA-only targets have submissions with *TM-score > 0.45* and/or *GDT_TS > 50*
- Similarity to experimental structure typically worse than similarity between alternative experimentally captured conformations

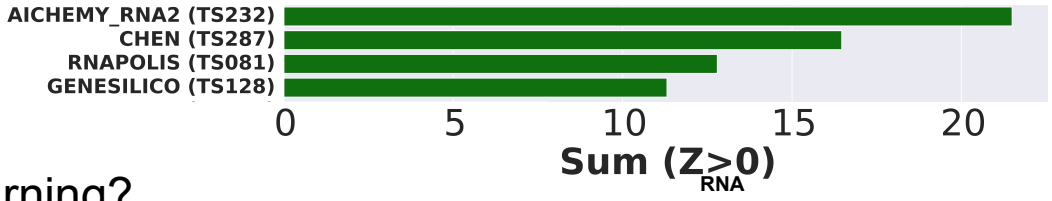
Summary

12 interesting targets (most from cryo-EM)

40 predictors (most new to RNA)

Classic CASP metrics are useful for RNA assessment, even flexible targets

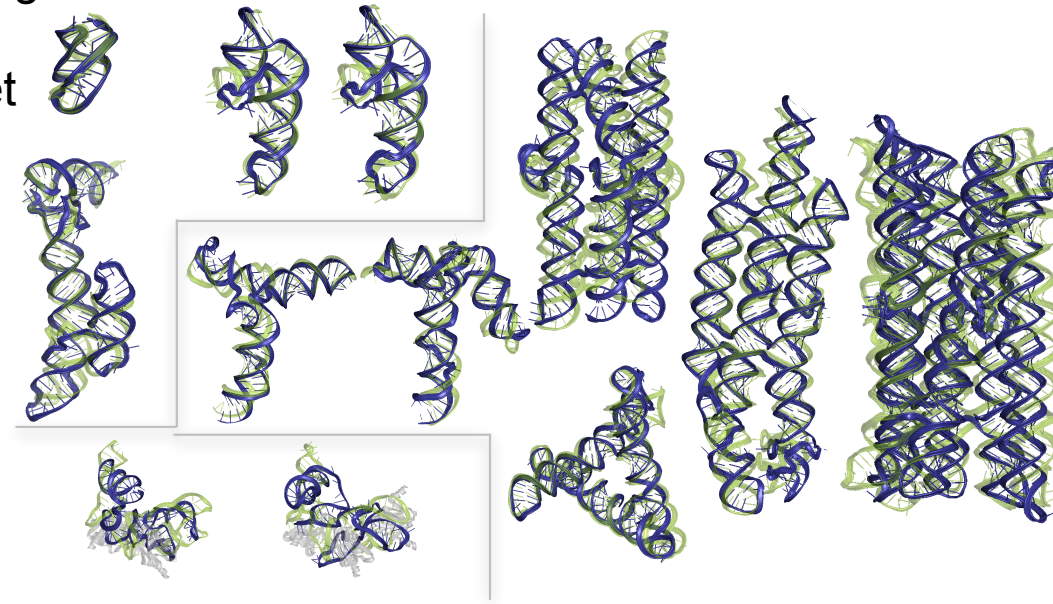
$$Z_{RNA} = \frac{1}{3}[Z_{TM} + Z_{GDT-TS}] + \frac{1}{8}[Z_{INF} + Z_{lDDT}] + \frac{1}{12}Z_{clash}$$



Top four predictors unambiguous. No deep learning?

At least one good model for each RNA-only target

- Clear refinement over template or designed structure
- Achieved in all cases: TM-score > 0.45 or GDT > 50
- Similarity to closest experimental structure typically worse than similarity between alternative experimental structures

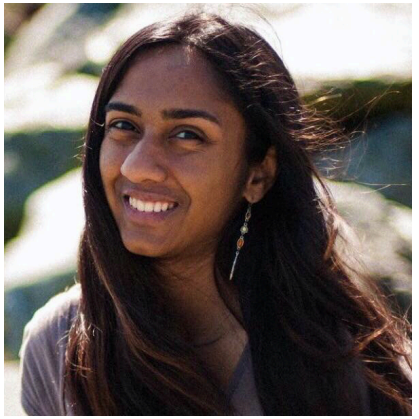


RNA-protein complexes remain challenging*

*And maybe atomically ordered *new* RNA folds?

Thank you!

Eric Westhof, Chichau Miao, Marta Szachniuk, Maciej Antczak, Maya Topf, Tom Mulvaney, Gabriel Studer, Marcin Magnus, Adam Zemla, Chengxin Zhang, Nick Grishin, Lisa Kinch, **Andriy Kryshtafovych, Krzystof Fidelis, John Moult**



Rachael Kretsch Phillip Pham Ramya Rangan

Join us at Stanford! E-mail: rhiju@stanford.edu

Howard Hughes Medical Institute

Stanford Gerald J. Lieberman Fellowship