CASP16 meeting, Punta Cana, December 1-4, 2024

Automated prediction of RNA structure using trRosettaRNA2 and template

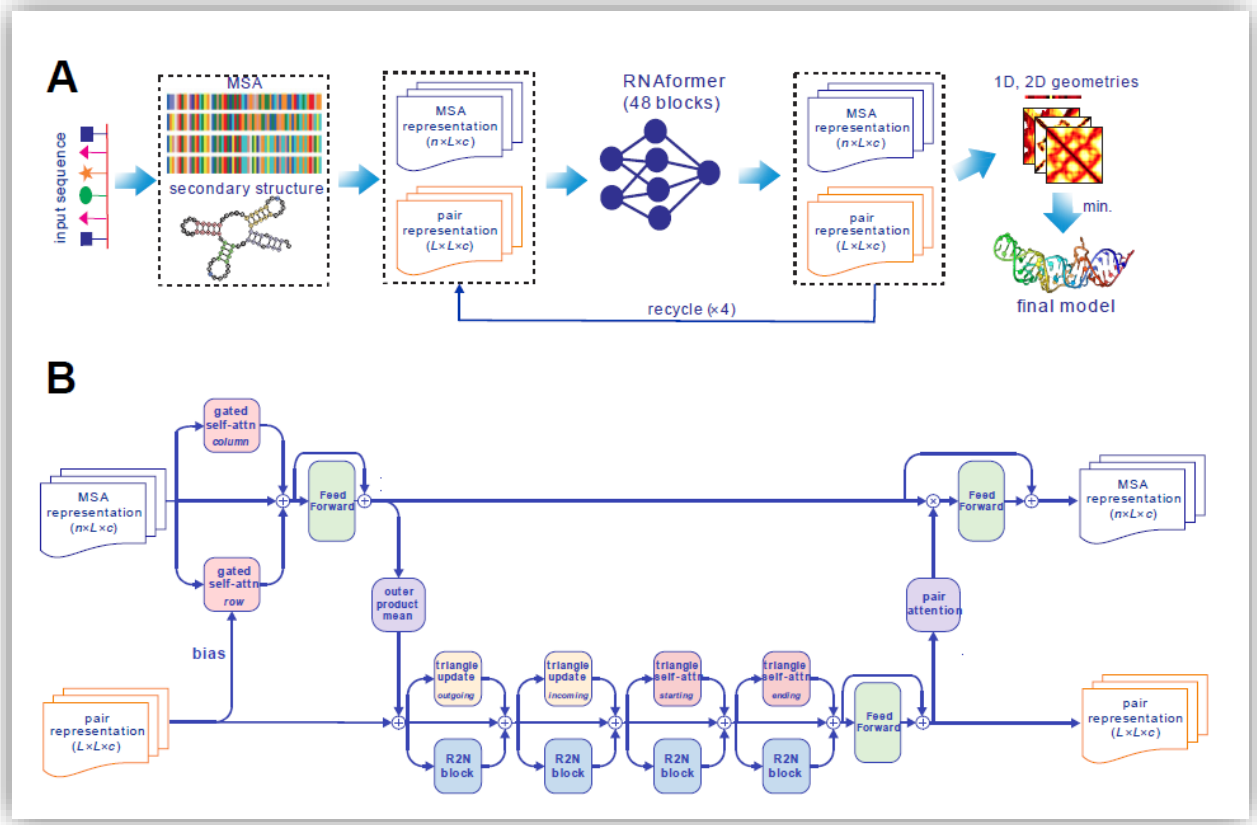
Wenkai Wang

Shandong University

https://yanglab.qd.sdu.edu.cn/

CONTENTS

trRosettaRNA: two-step approach

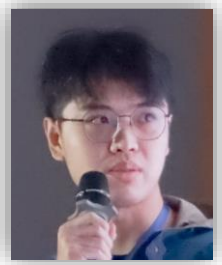


trRosettaRNA: automated prediction of RNA 3D structure with transformer network

W Wang, C Feng, R Han, Z Wang, L Ye, Z Du… - Nature …, 2023 - nature.com

… **trRosettaRNA**, an automated deep learning-based approach to RNA 3D structure prediction. The **trRosettaRNA** … Benchmark tests suggest that **trRosettaRNA** outperforms traditional …

☆ Save Cite Cited by 63 Related articles All 10 versions



Wenkai Wang

Chenjie Feng

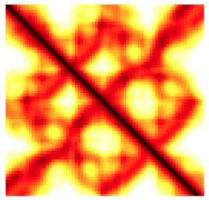
nature communications

trRosettaRNA: automated prediction of RNA 3D structure with transformer network

Wenkai Wang, Chenjie Feng, Renmin Han, Ziyi Wang, Lisha Ye, Zongyang Du, Hong Wei, Fa Zhang Zhenling Peng & Jianyi Yang

Nature Communications **14**, Article number: 7266 (2023) | Cite this article

18k Accesses | **27** Citations | **49** Altmetric | Metrics

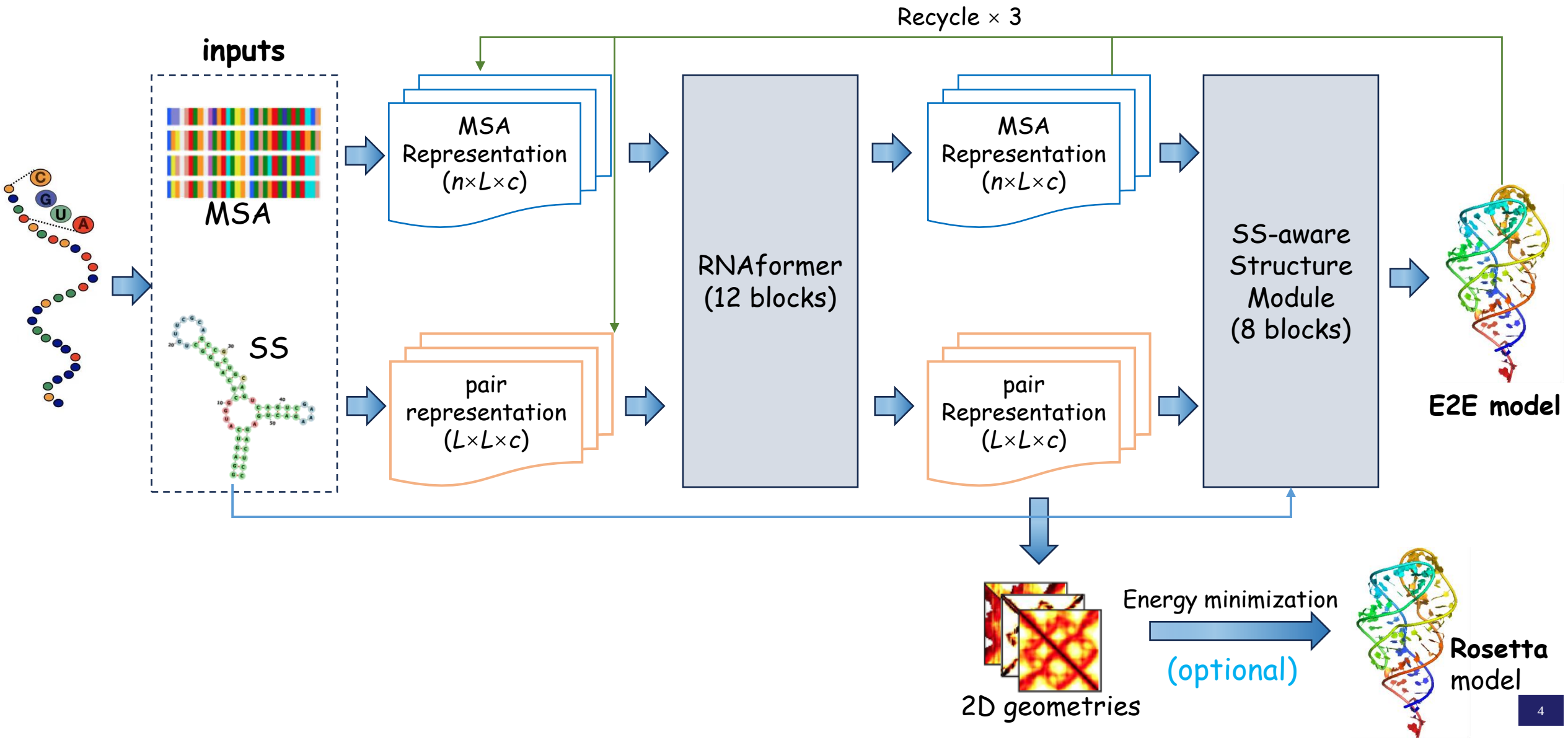


trRosettaRNA

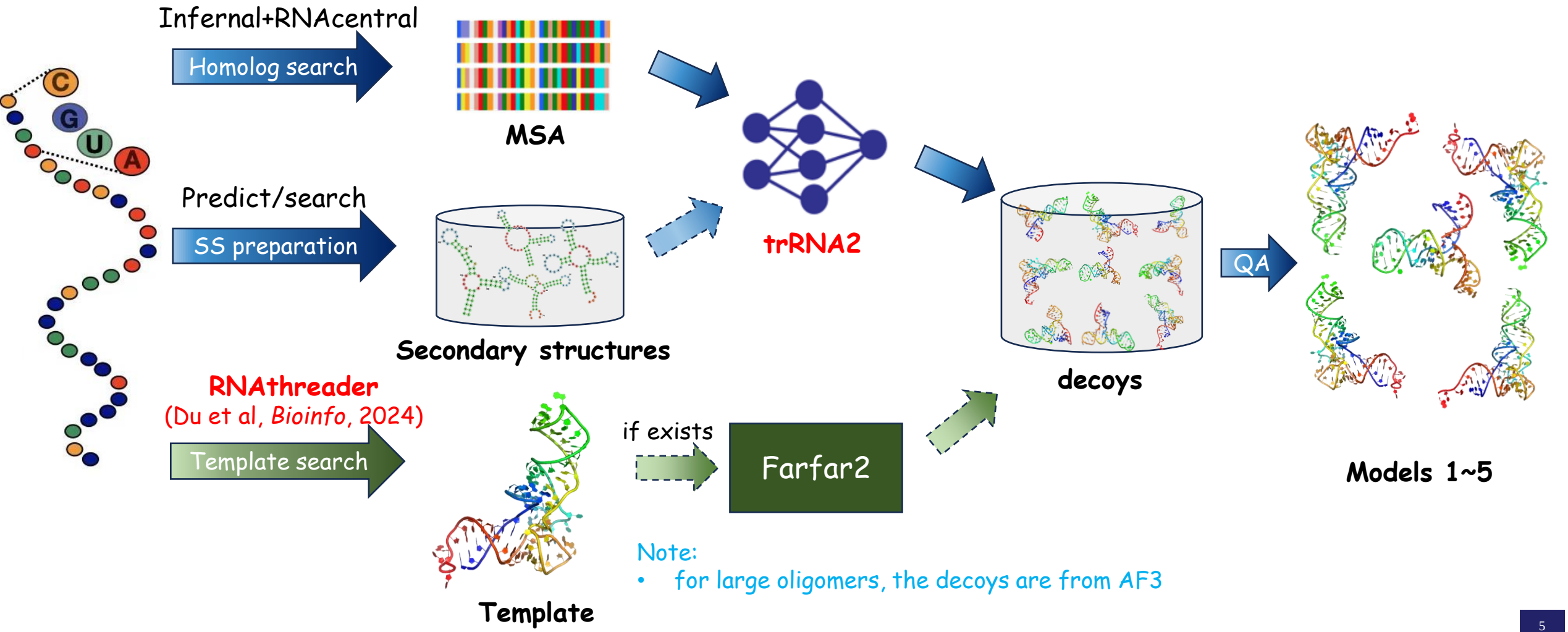
RNA 3D structure prediction by transformer network

https://yanglab.qd.sdu.edu.cn/trRosettaRNA/

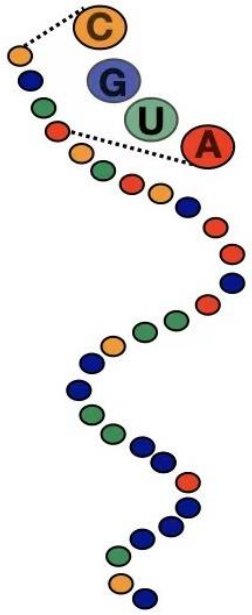
trRosettaRNA2: end-to-end approach



Yang-Server for RNA monomers in CASP16



Secondary structure preparation

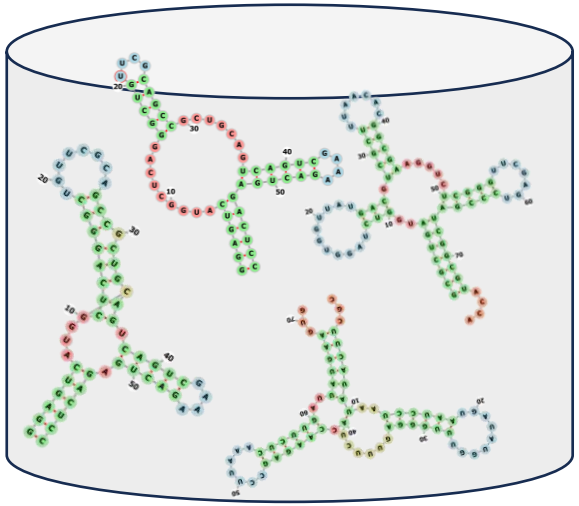


trRNA-SS, SPOT-RNA, EternaFold

SS prediction

RNAthreader (Du et al, *Bioinfo*, 2024), R2DT

Template search



Secondary structures

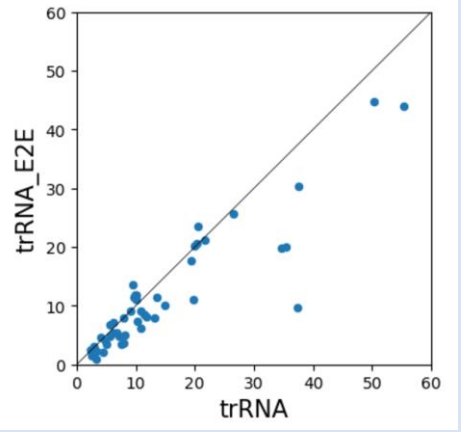
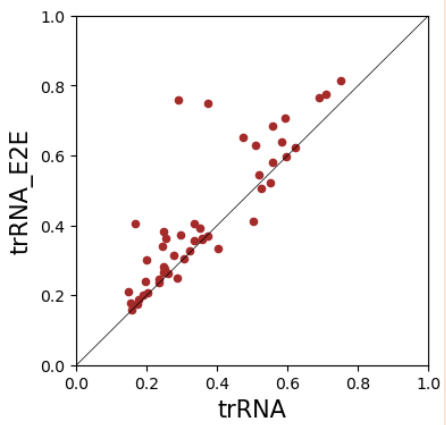
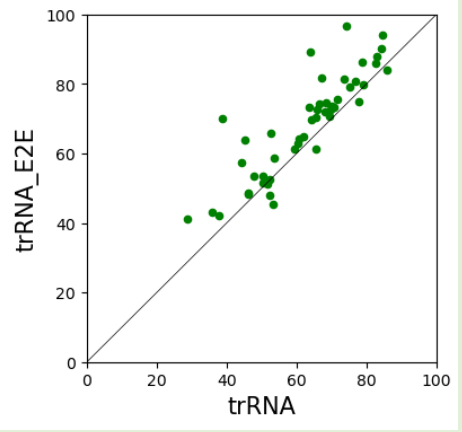
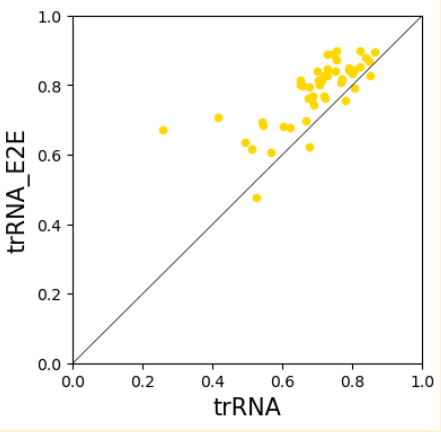
CONTENTS

trRosettaRNA2 vs trRosettaRNA

Two benchmark datasets:

- **TS36**: 36 RNAs released after 2022-01
- **CASP15**: 12 RNAs from CASP15

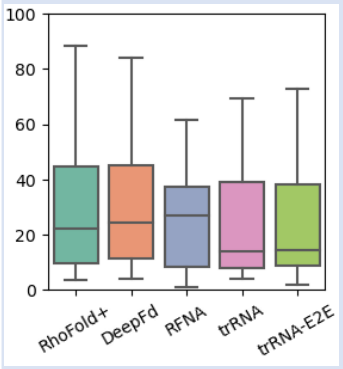
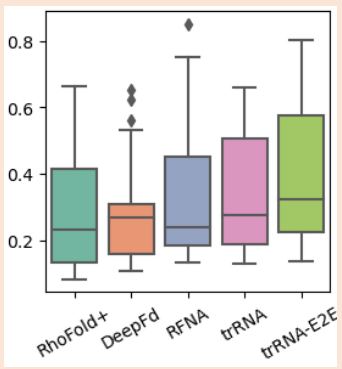
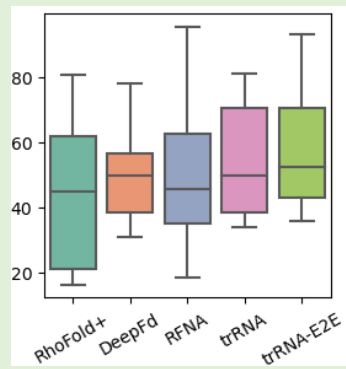
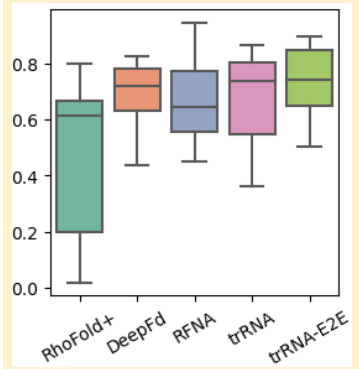
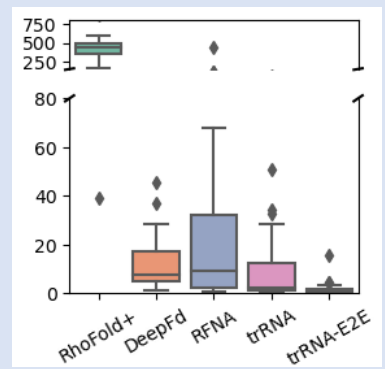
Note:
- with same inputs (MSA/SS)
- training set early than 2022-01

Metrics		RMSD(Å) ↓		TM-score ↑		lDDT ↑		INF ↑	
		TS36	CASP15	TS36	CASP15	TS36	CASP15	TS36	CASP15
Average value	trRNA	9.7	26.0	0.401	0.243	0.659	0.516	0.719	0.617
	trRNA2	**7.8**	**20.8**	**0.458**	**0.292**	**0.716**	**0.579**	**0.790**	**0.751**
Head-to-head comparison									

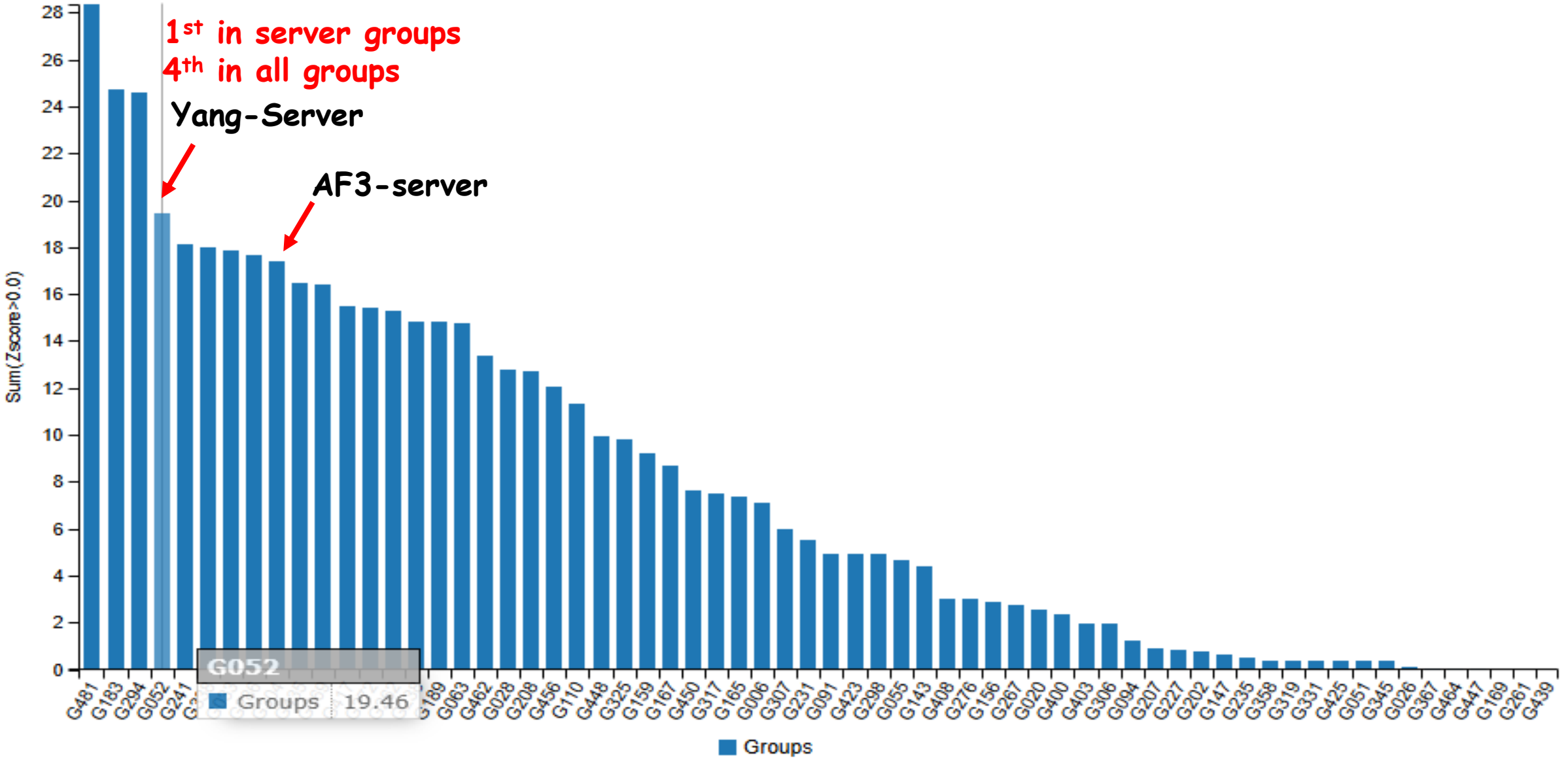
Comparison with other methods on CASP16

Yang-Multimer (TS456): for blind benchmark

Note: with same inputs
- MSA: infernal+RNAcentral
- SS: trRNA-SS

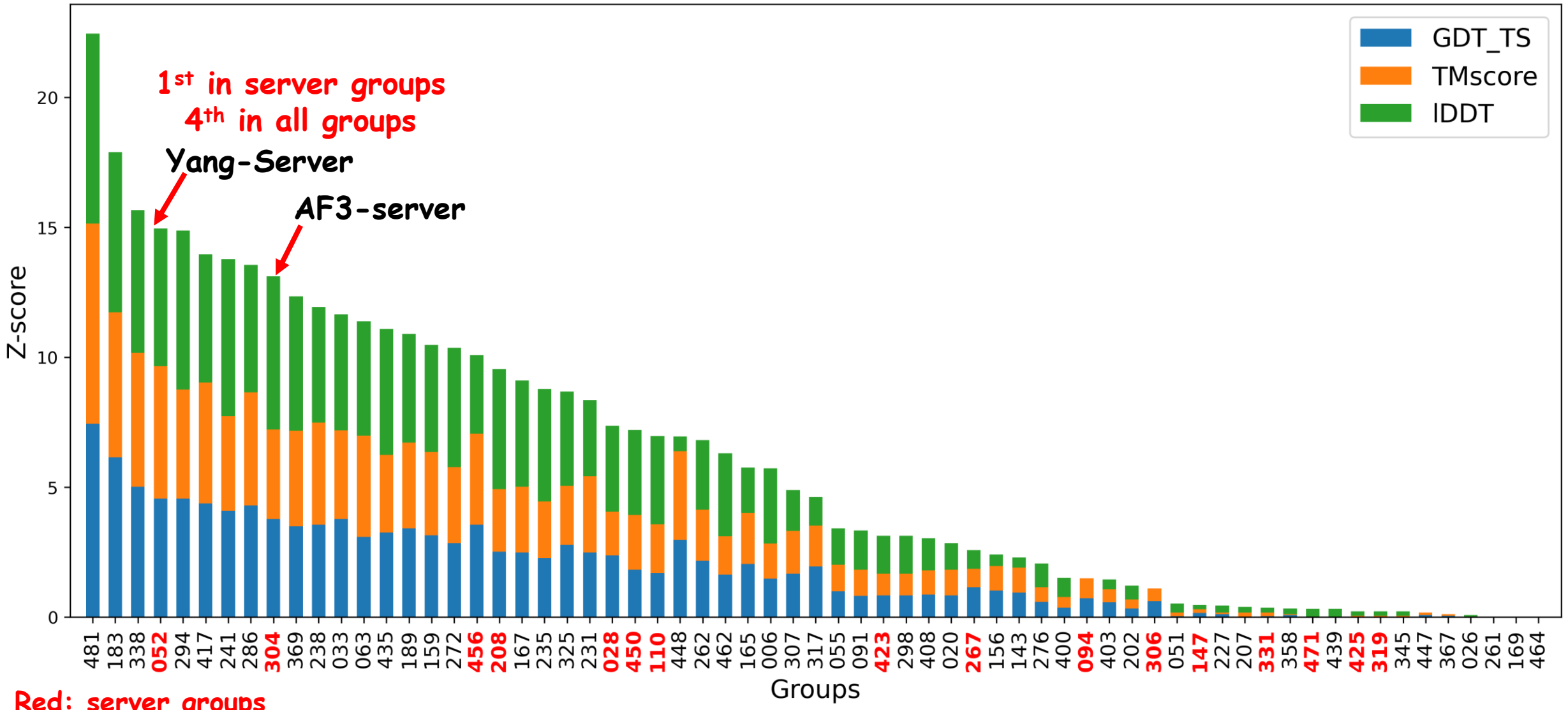
Model names	Methods	RMSD(Å) ↓	TM-score ↑	lDDT ↑	INF ↑	Clashscore ↓
TS456_5	RhoFold	50.7	0.273	43.7	0.438	234.4
Run on local	RhoFold+	37.7	0.287	45.5	0.470	418.1
TS456_4	DeepFoldRNA	30.3	0.279	49.9	0.694	12.8
TS456_3	RFNA	24.6	0.329	50.9	0.661	40.8
TS456_2	trRNA	26.0	0.333	54.6	0.678	9.3
TS456_1	trRNA2	**24.5**	**0.393**	**57.6**	**0.737**	**2.0**
Metric Distribution						

Official ranking on CASP16 (36 targets)



Ranking on single-phase targets

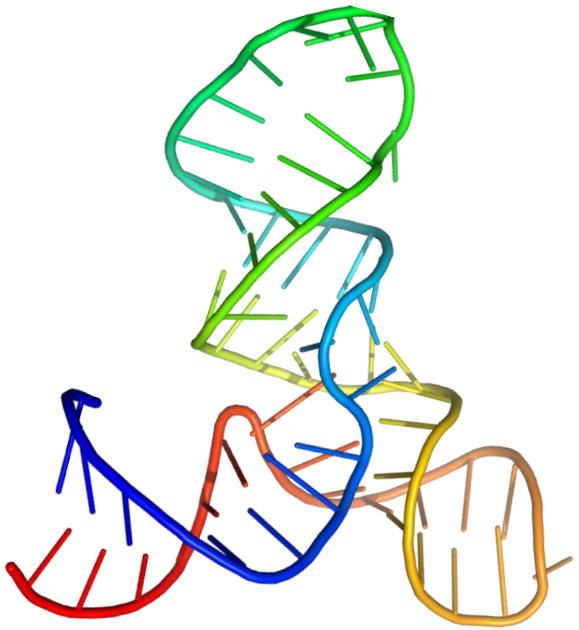
(25 targets out of 36)

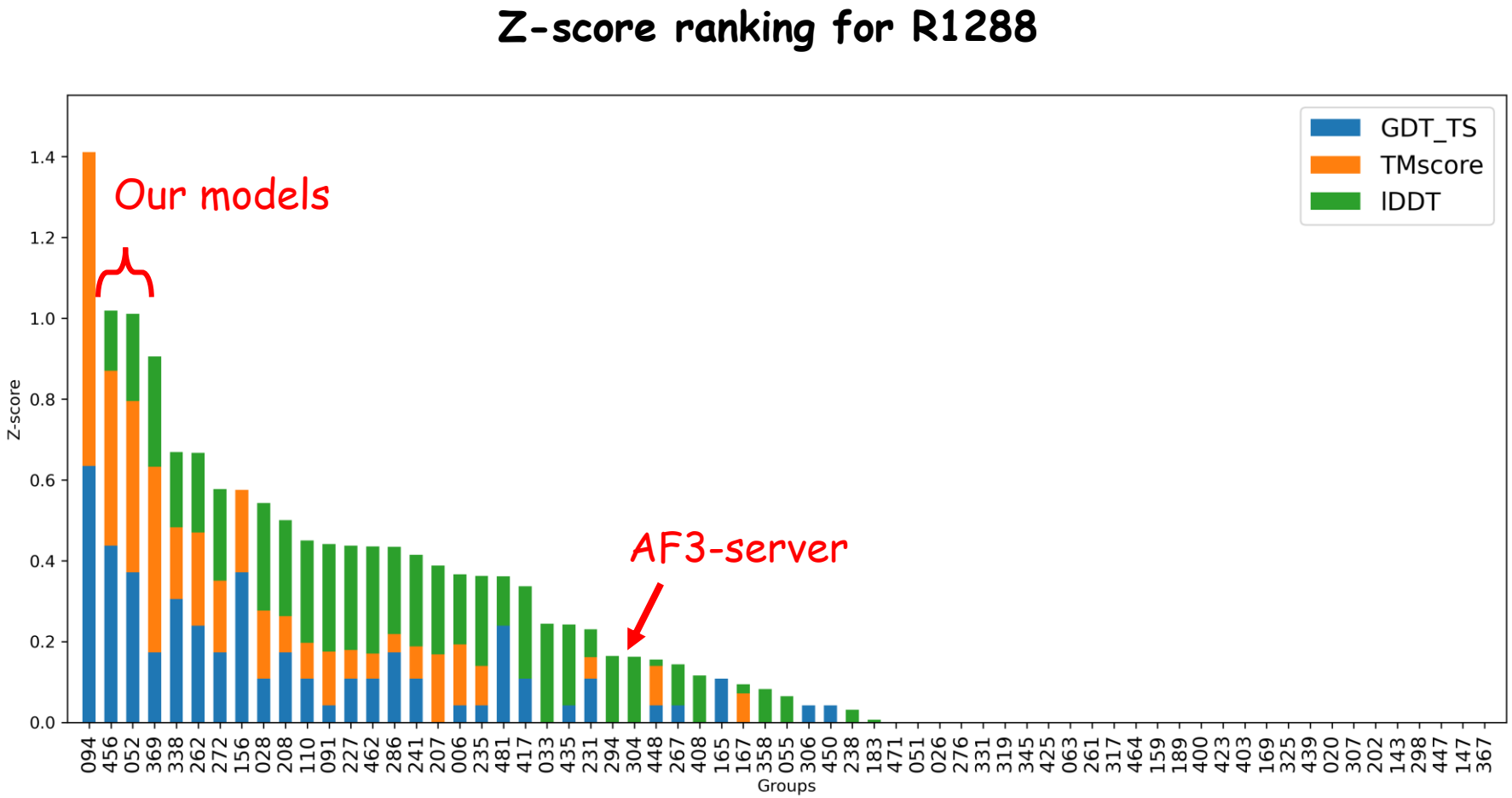


Red: server groups

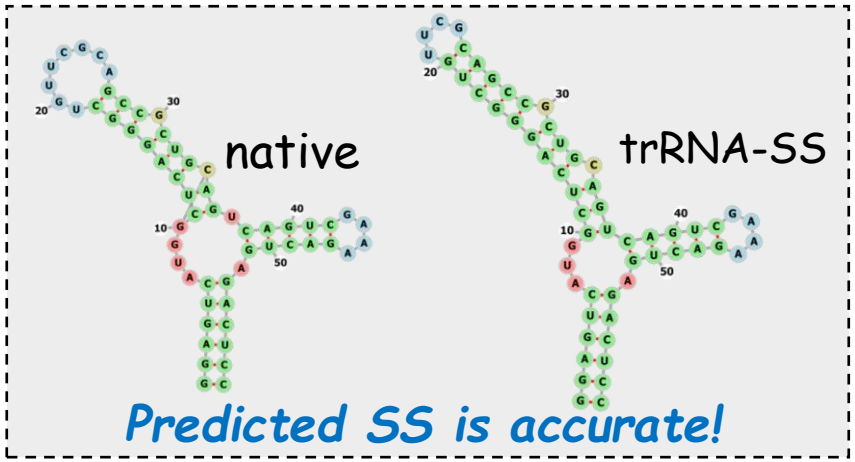
Black: human groups

What went right? R1288





What went right? R1288



- Competitive with AF3 w/o SS
- Further improved by SS restraints
- Has less steric clashes

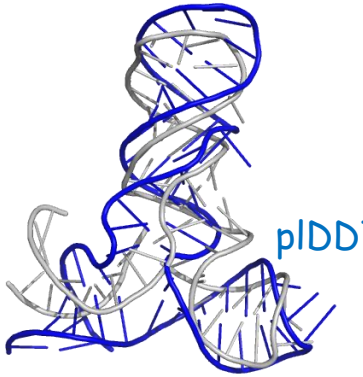
Overall best model (Vfold)

RMSD=5.1 Å
Clashscore=14.5

AF3 best model

RMSD=8.5 Å
Clashscore=13.9

trRNA2 models

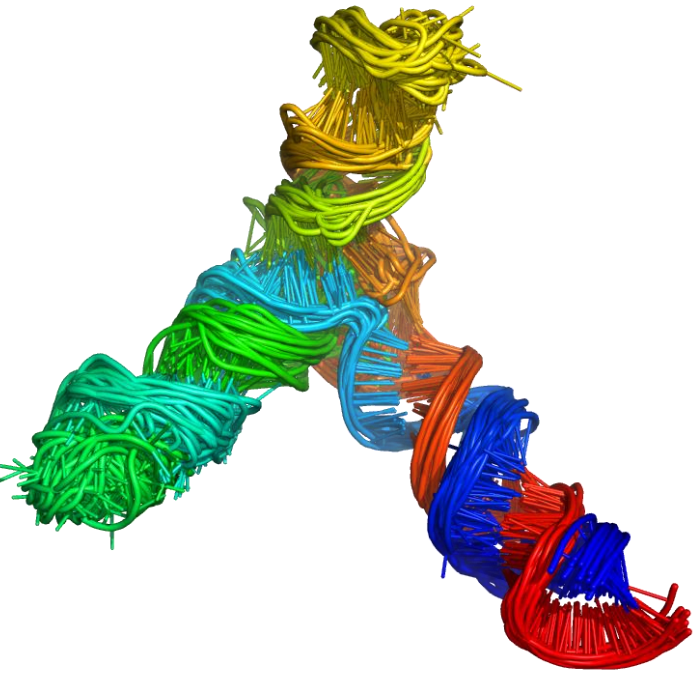


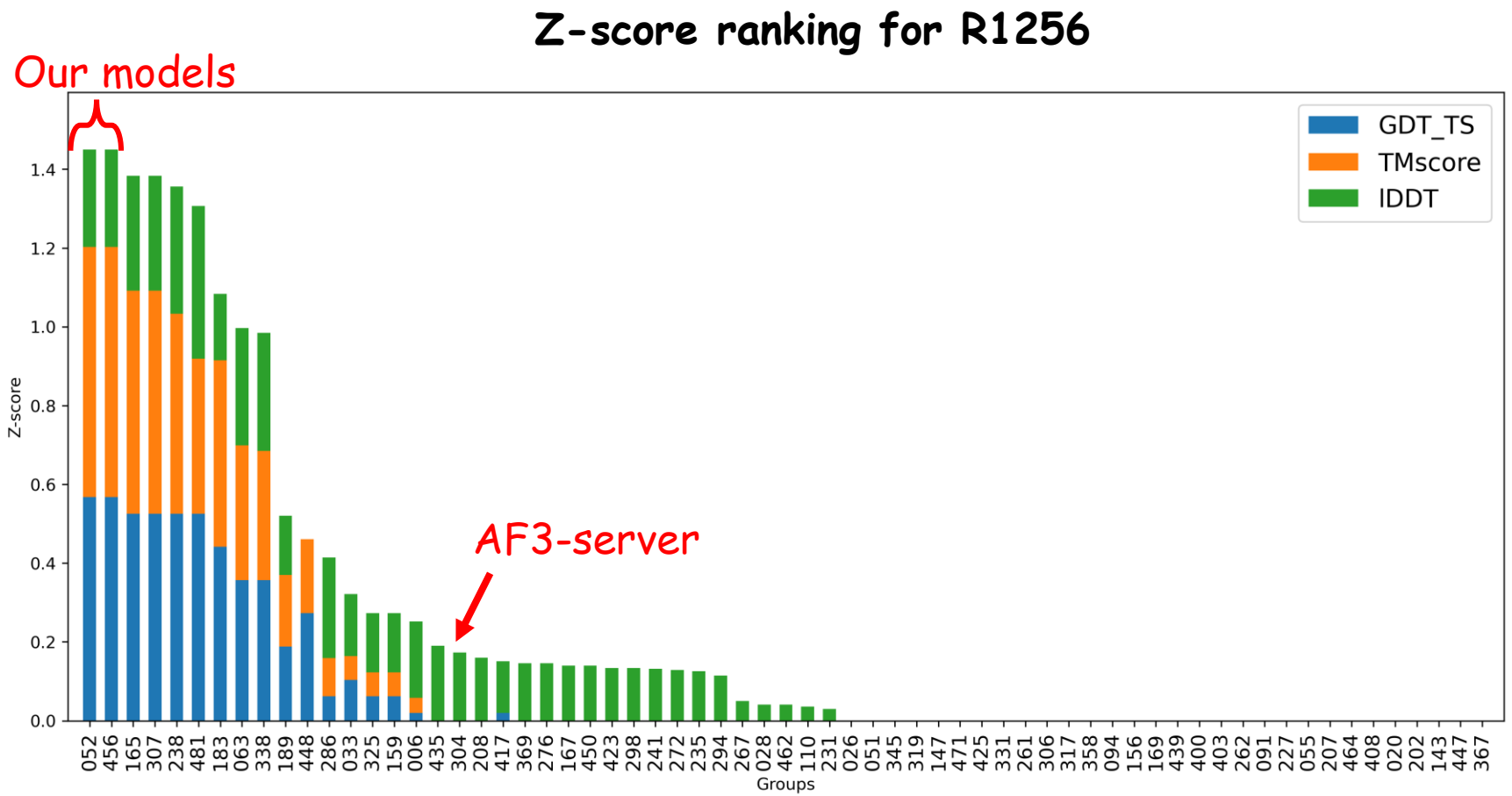
w/o SS
RMSD=8.9 Å
Clashscore=2.1

w/ predicted SS
RMSD=6.4 Å
Clashscore=2.1

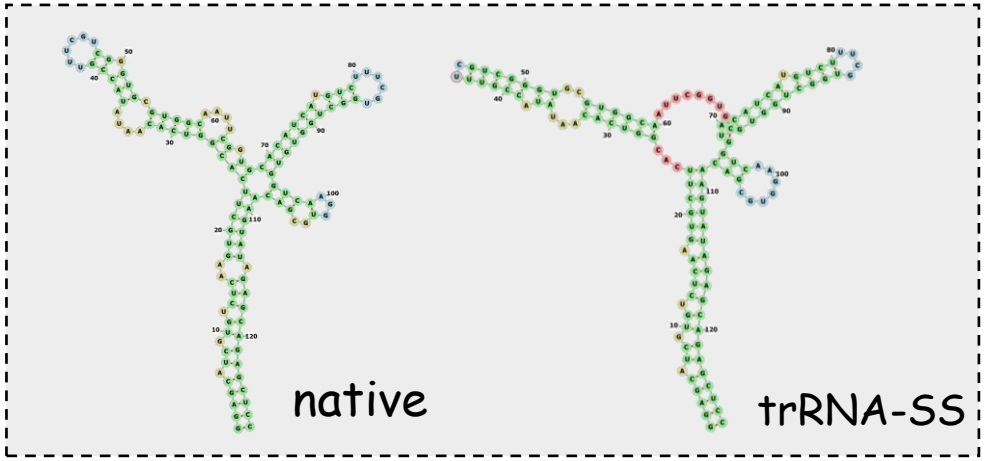
Gray: native

What went right? R1256





What went right? R1256

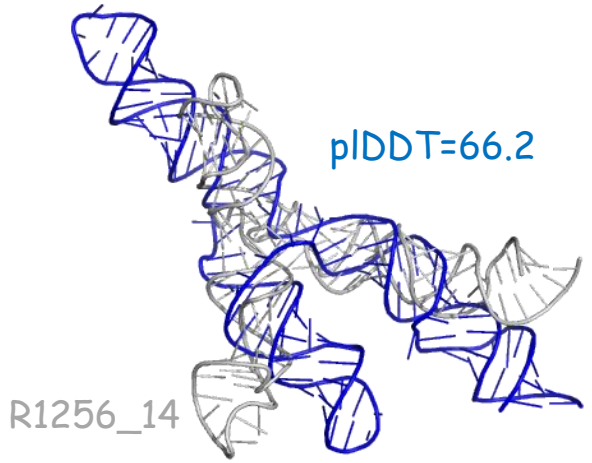


- Competitive with AF3 w/o SS
- Further improved by SS restraints
- Has less steric clashes

AF3 best model

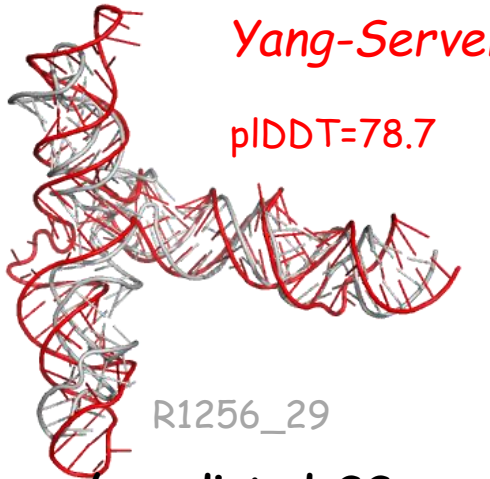
RMSD=19.7 Å
Clashscore=12.1

trRNA2 models



w/o SS
RMSD=19.9 Å
Clashscore=0.7

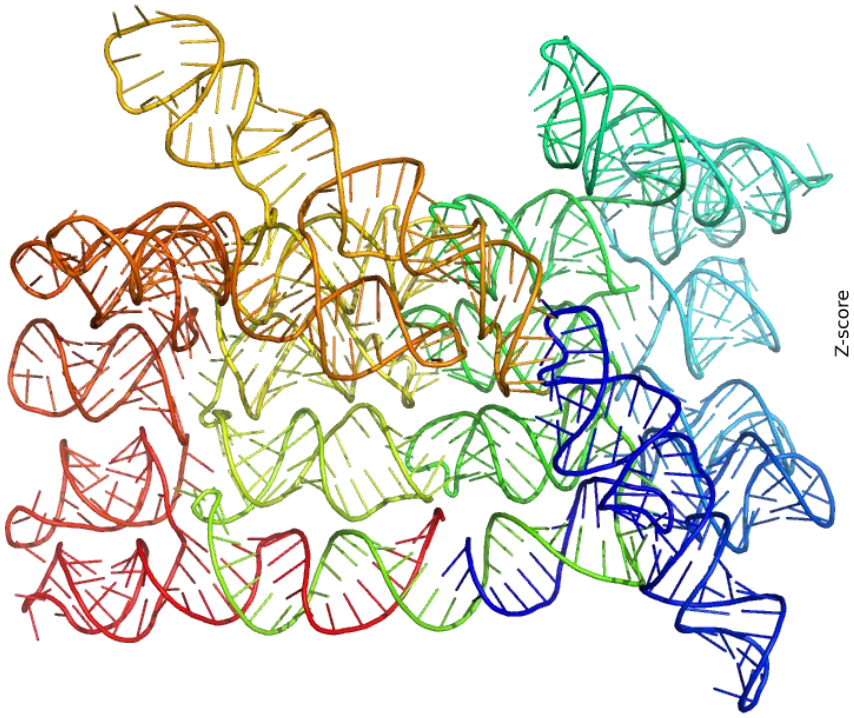
Yang-Server

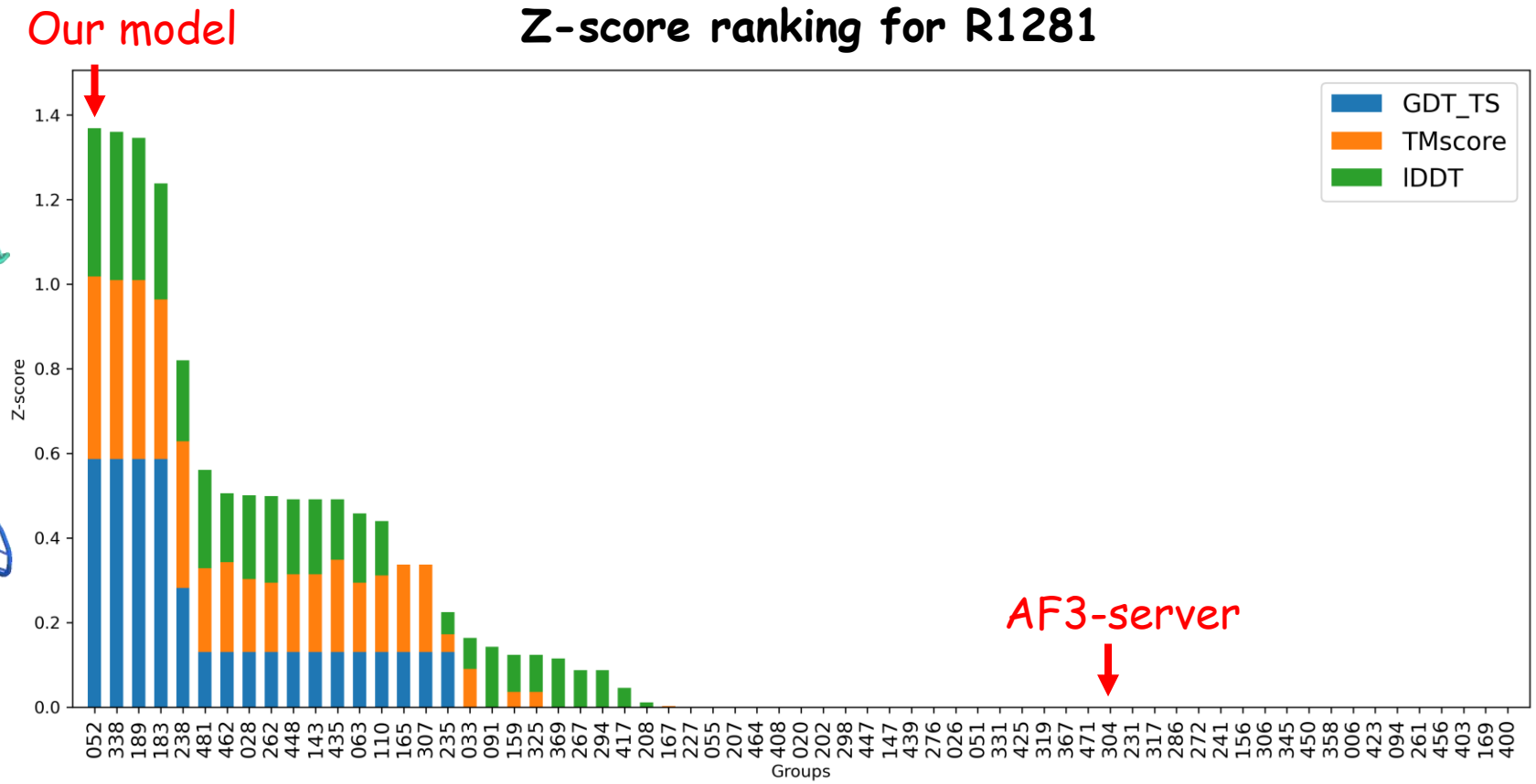


w/ predicted SS
RMSD=7.5 Å
Clashscore=2.0

Gray: native

What went right? R1281

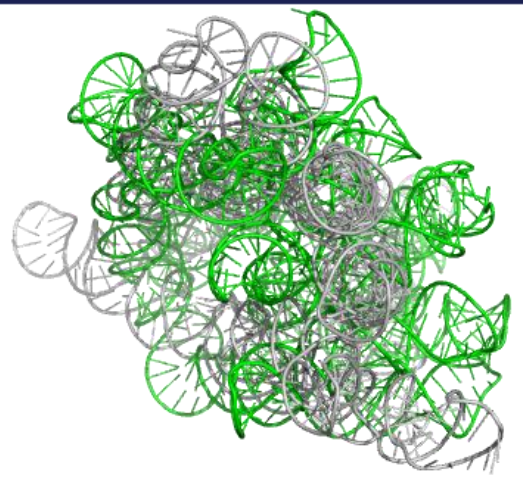




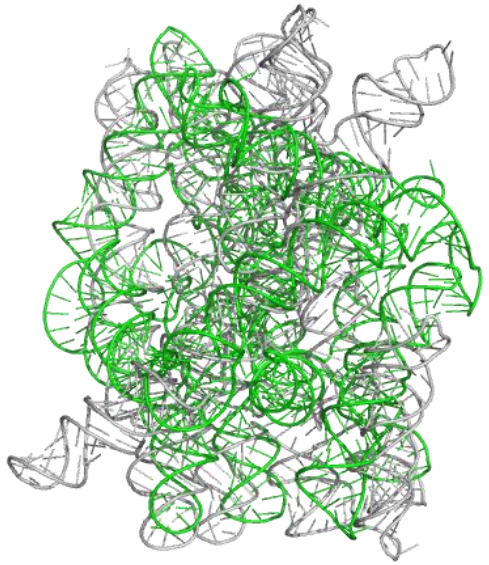
What went right? R1281

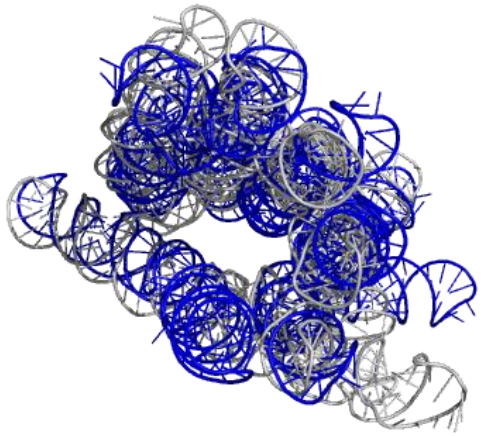
- template is helpful

Detected template:
7PTK; 7PTL
(by RNAthreader)

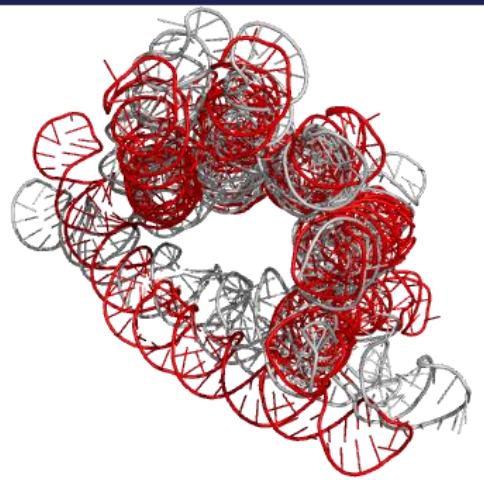


AF3 best model
RMSD=28.0 Å
Clashscore=17.6





trRNA2 model
with template SS
RMSD=18.6 Å
Clashscore=2.5



Yang-Server best model
by *RNAthreader*
RMSD=17.5 Å
Clashscore=6.5

Gray: native

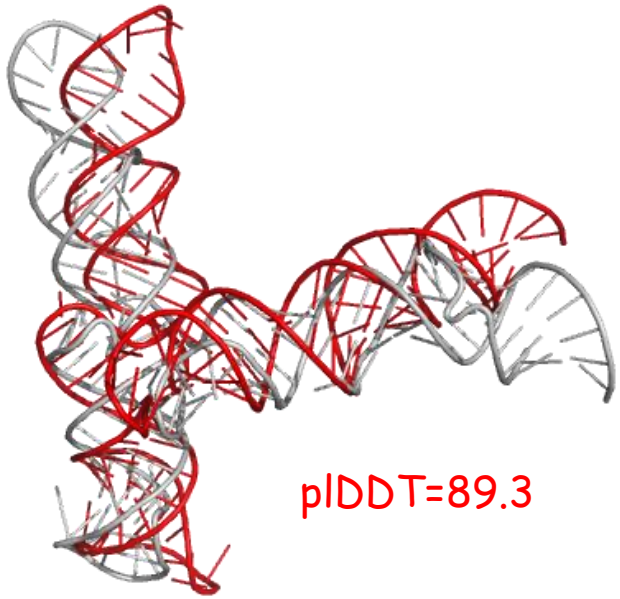
What went wrong? R1255

- QA fails on R1255

trRNA2 models

plDDT=82.6

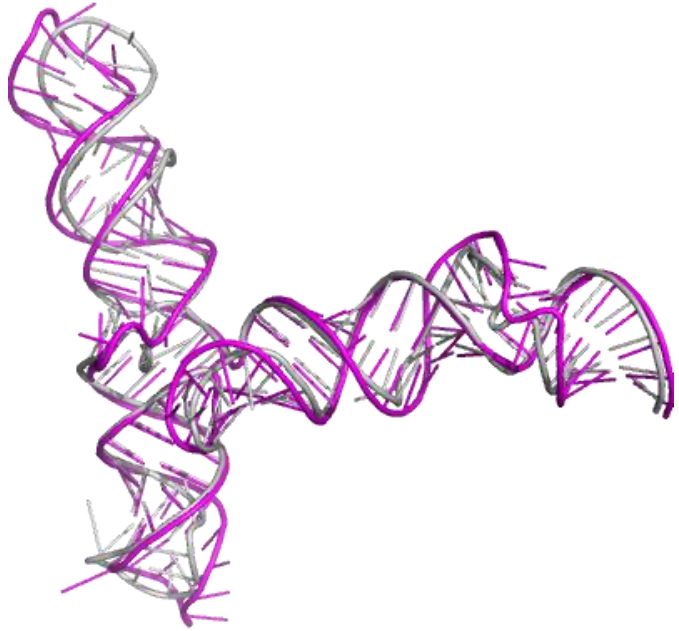
Yang-Multimer model (trRNA-SS)
RMSD=5.3 Å



plDDT=89.3

Yang-Server model (w/o SS)
RMSD=10.3 Å

Overall best model



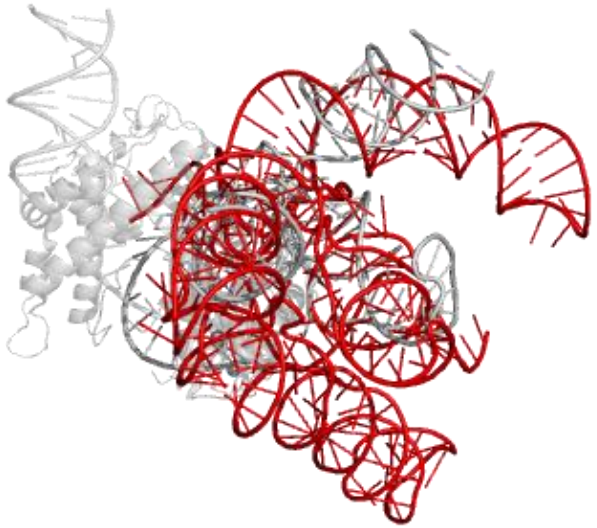
GuangzhouRNA-human
RMSD=3.4 Å

Gray: native

What went wrong? R1212(Prot-NucA)

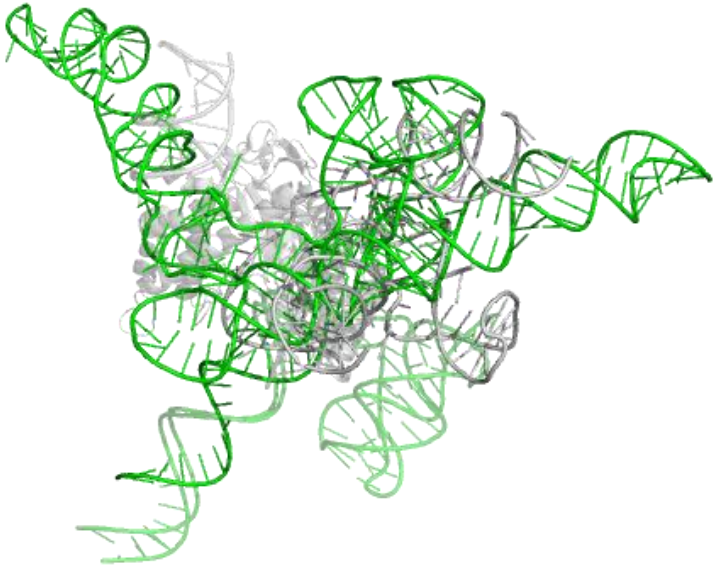
- Protein-RNA complex modeling is still challenging

Yang-Server model

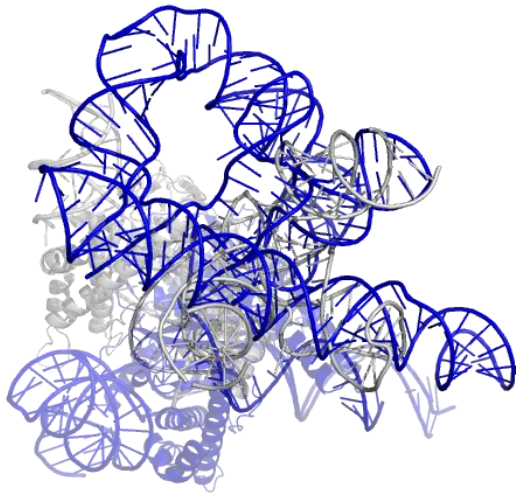


w/o protein
RMSD=25.4 Å

AF3 models



w/o protein
RMSD=25.6 Å



w/ protein
RMSD=27.7 Å

Gray: native

CONTENTS

Conclusion

- trRosettaRNA2 significantly outperforms trRosettaRNA
- SS/template restraints are helpful

Challenges:

- Long RNAs (>500 nt)
- Oligomers with >2 chains
- Protein-RNA complexes

Acknowledgments

organizers, assessors

Jianyi Yang

NSFC