

Supplemental Material

VfoldMCPX: Predicting multistrand RNA complexes

Sicheng ZHANG¹⁺, Yi CHENG¹⁺, Peixuan GUO² and Shi-Jie CHEN^{1,*}

¹Department of Physics, Department of Biochemistry, and Institute for Data Science and Informatics,

University of Missouri, Columbia, Missouri 65211, United States

²Center for RNA Nanobiotechnology and Nanomedicine, College of Pharmacy, Dorothy M. Davis Heart and Lung Research Institute, James Comprehensive Cancer Center, The Ohio State University, Columbus,

Ohio 43210, United States

*Author to whom correspondence should be addressed; E-mail: chenshi@missouri.edu

⁺These authors contributed equally to this work.

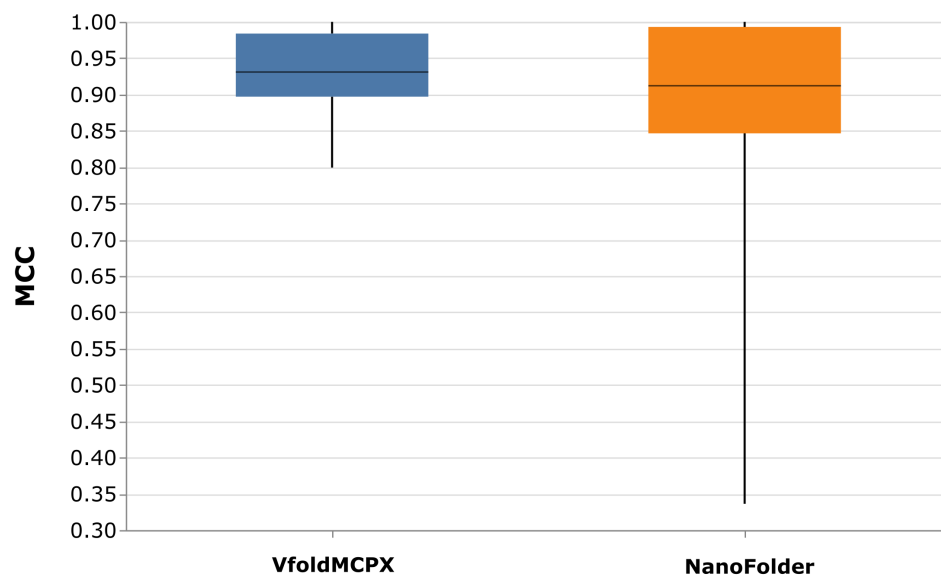


Figure S1: Box-Whisker plot of Matthews correlation coefficients (MCC) for VfoldMCPX (left) and NanoFolder [1] (right). The structures in the test data set (PDB ID shown in Supplemental Table 1) and the plot of NanoFolder are obtained from the published work [1]. The native 2D structures are extracted by RNAPdbec [2, 3] from the PDB database [4]. In the box-whisker plot, the colored boxes depict the first quartile, median and third quartile; the vertical whiskers depict the minimum and maximum scores.

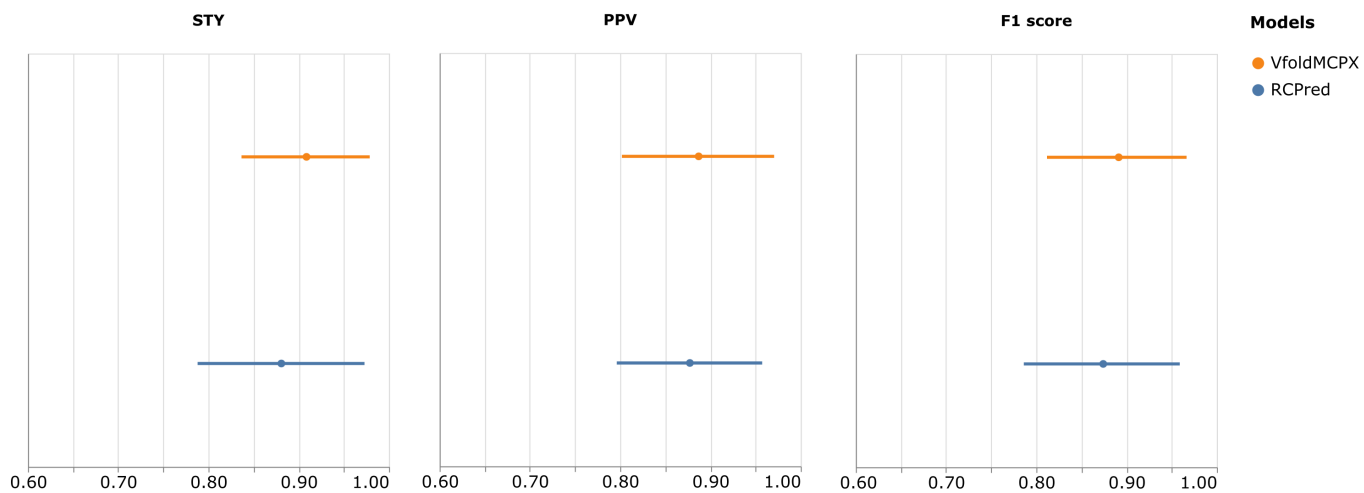


Figure S2: Comparison of the Positive Predictive Value (PPV), Sensitivity (STY), and F1 score between the predicted structures from VfoldMCPX and RCPred [5]. The structures in the test data set (PDB ID shown in Supplemental Table 2) and the predicted structures from RCPred are obtained from the published work [5]. The native 2D structures are extracted by RNApdbec [2, 3] from the PDB database [4]. Mean values are labeled and depicted as dots. Error bars are depicted using confidence intervals at 95% confidence level.

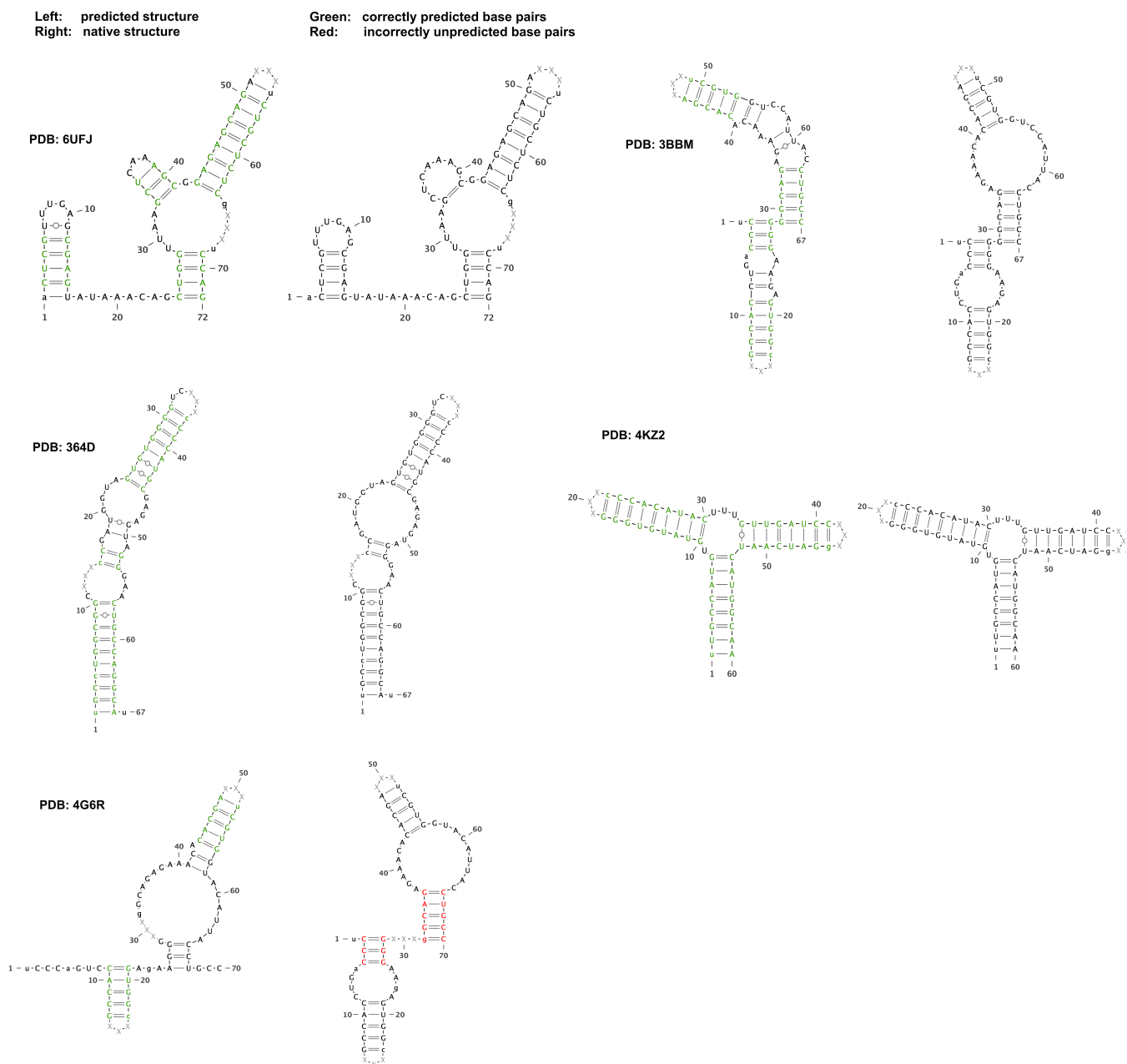


Figure S3: Predicted structures from VfoldMCPX (left) and native structures from the 3D crystal structures (right) for 5 structures with three or more strands (Listed in Table 1). The native structures are extracted by RNAppdbee [2, 3] from the PDB database [4]. PDB IDs are labeled. Correctly predicted base pairs are colored by green in the predicted structures. Incorrectly unpredicted base pairs are colored by red in the native structures.

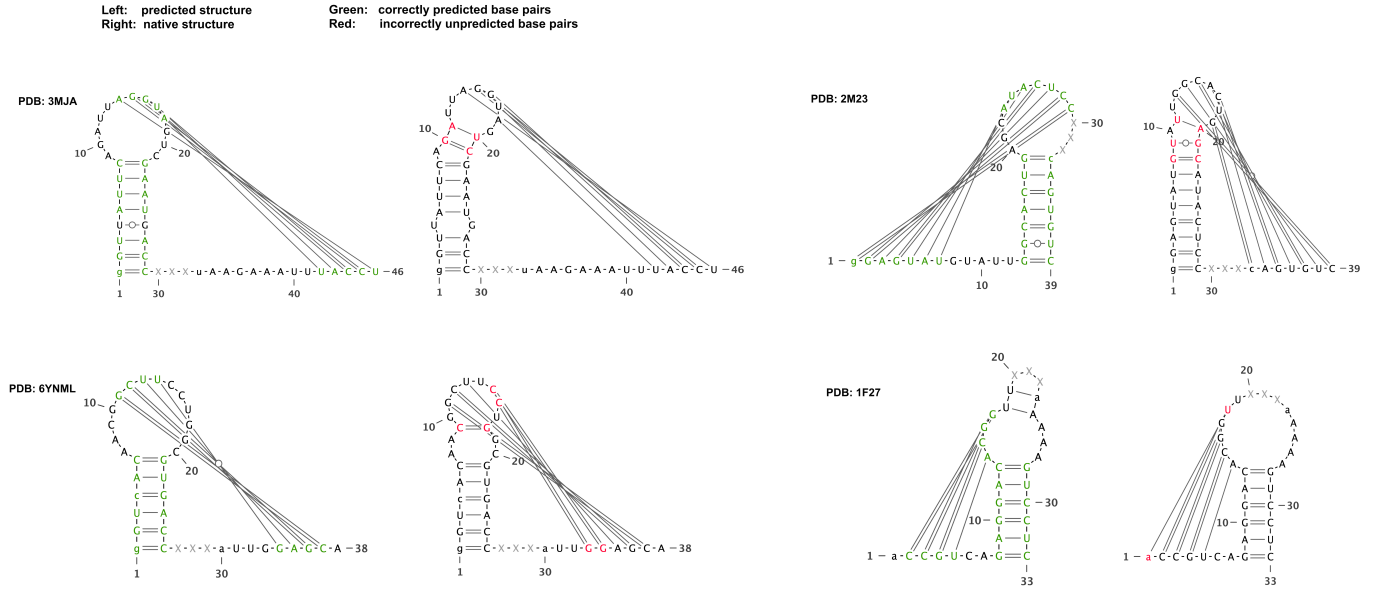


Figure S4: Predicted structures from VfoldMCPX (left) and native structures from the 3D crystal structures (right) for 4 structures with H-type pseudoknots (Listed in Table 1). The native structures are extracted by RNApdbec [2, 3] from the PDB database [4]. PDB IDs are labeled. Correctly predicted base pairs are colored by green in the predicted structures. Incorrectly unpredicted base pairs are colored by red in the native structures.

Table S1: List of PDB structures analyzed for the comparison of Matthews correlation coefficients (MCC) between the predicted structures from VfoldMCPX and NanoFolder (Supplemental Figure S1).

1NTA	2D1A	2G5K	2OE5	2OEU
2PN3	2XEB	3BNP	3CJZ	429D

Table S2: List of PDB structures analyzed for the comparison of PPV, STY, and F_1 score between the predicted structures from VfoldMCPX and RCPred (Supplemental Figure S2).

1DUQ	1FEU	1G1X	1JBR	1M5K
1MME	1NUJ	1QC0	1T0D	1TFW
1TFY	1VTL	1YSH	1YVP	1ZX7
1ZZ5	2GJW	2I91	2NOK	2NUG
2RFK	357D	488D		

Table S3: List of PDB structures analyzed for the comparison of PPV, STY, and F_1 score between the predicted structures from VfoldMCPX, RNAcofold, bifold, and MultiRNAFold (Fig. 3).

6DN2	4TS2	2QUW	5EAO	2OEU
1YLS	4PHY	3BNR	3LOA	2PN3
6AZ4	359D	1NTA	6HU6	3SJ2
1T0E	5T3K	2XEB	2OE5	5UZ6
5KVJ	1ZX7	1CSL	5CNR	397D
1MHK	1DUQ	7A3Y	3CGP	

References

- [1] Bindewald, E.; Afonin, K.; Jaeger, L.; Shapiro, B. A. Multistrand RNA secondary structure prediction and nanostructure design including pseudoknots. *ACS nano* **2011**, 5, 9542–9551.
- [2] Zok, T.; Antczak, M.; Zurkowski, M.; Popenda, M.; Blazewicz, J.; Adamiak, R. W.; Szachniuk, M. RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic acids research* **2018**, 46, W30–W35.
- [3] Antczak, M.; Zok, T.; Popenda, M.; Lukasiak, P.; Adamiak, R. W.; Blazewicz, J.; Szachniuk, M. RNApdbee—a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic acids research* **2014**, 42, W368–W372.
- [4] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic acids research* **2000**, 28, 235–242.
- [5] Legendre, A.; Angel, E.; Tahi, F. RCPred: RNA complex prediction as a constrained maximum weight clique problem. *BMC bioinformatics* **2019**, 20, 53–62.