

**Supplementary Material**  
**for**

**Comparative analysis of RNA 3D structure prediction methods: towards enhanced modeling of RNA-Ligand interactions**

Chandran Nithin<sup>1,2\*</sup>, Sebastian Kmiecik<sup>2</sup>, Roman Błaszczuk<sup>1</sup>, Julita Nowicka<sup>1</sup>, Irina Tuszyńska<sup>1</sup>

<sup>1</sup>Molecure SA, 02-089 Warsaw, Poland

<sup>2</sup>Laboratory of Computational Biology, Biological and Chemical Research Center, Faculty of Chemistry, University of Warsaw, 02-089 Warsaw, Poland.

\*To whom correspondence should be addressed: [n.chandran@molecure.com](mailto:n.chandran@molecure.com), [n.chandran@cnbc.uw.edu.pl](mailto:n.chandran@cnbc.uw.edu.pl), [i.tuszynska@molecure.com](mailto:i.tuszynska@molecure.com)

**Supplementary Table S1: Dataset of RNA structures for benchmarking.**

PDB ID	Length	Length range	Structure Class*	Type of MWJ†	Class
1AJU	30	1-30	Simple HP		Simple
1ARJ	29	1-30	Simple HP		Simple
1BYJ	27	1-30	Simple HP		Simple
1EHT	33	31-60	Simple HP		Simple
1EI2	25	1-30	Simple HP		Simple
1F1T	38	31-60	Simple HP		Simple
1FMN	35	31-60	Simple HP		Simple
1FYP	27	1-30	Simple HP		Simple
1LVJ	31	31-60	Simple HP		Simple
1NBK	34	31-60	Simple HP		Simple
1NEM	23	1-30	Simple HP		Simple
1QD3	29	1-30	Simple HP		Simple
1RAW	36	31-60	Simple HP		Simple
1TOB	27	1-30	Simple HP		Simple
1Y26	71	61-90	MWJ+PK	3WJ	Difficult
1Y27	68	61-90	MWJ+PK	3WJ	Difficult
2AU4	41	31-60	Simple HP		Simple
2B57	65	61-90	MWJ+PK	4WJ	Difficult
2CKY	77	61-90	MWJ	3WJ	Moderate
2EES	67	61-90	MWJ+PK	4WJ	Difficult
2EET	67	61-90	MWJ+PK	4WJ	Difficult
2EEU	67	61-90	MWJ+PK	4WJ	Difficult
2EEV	67	61-90	MWJ+PK	4WJ	Difficult
2EEW	67	61-90	MWJ+PK	4WJ	Difficult
2G9C	67	61-90	MWJ+PK	4WJ	Difficult
2GDI	80	61-90	Simple HP		Simple
2HOJ	79	61-90	Simple HP		Simple
2HOK	78	61-90	Simple HP		Simple
2HOM	80	61-90	Simple HP		Simple
2HOO	83	61-90	Simple HP		Simple
2HOP	76	61-90	Simple HP		Simple
2JUK	22	1-30	Simple HP		Simple
2KTZ	38	31-60	Simple HP		Simple
2KX8	42	31-60	Simple HP		Simple
2KXM	27	1-30	Simple HP		Simple
2L1V	36	31-60	HP+PK		Moderate
2L94	45	31-60	Simple HP		Simple
2LWK	32	31-60	Simple HP		Simple
2MIY	59	31-60	HP+PK		Moderate
2TOB	20	1-30	Simple HP		Simple
2XO1	65	61-90	MWJ+PK	4WJ	Difficult
3D0U	161	151-180	MWJ+PK	5WJ	Difficult

3DIG	174	151-180	MWJ+PK	5WJ	Difficult
3DS7	67	61-90	MWJ+PK	4WJ	Difficult
3E5E	53	31-60	MWJ	3WJ	Moderate
3F2Q	112	91-120	MWJ+PK	4WJ	Difficult
3FO4	63	61-90	MWJ+PK	4WJ	Difficult
3FO6	67	61-90	MWJ+PK	4WJ	Difficult
3FU2	34	31-60	HP+PK		Moderate
3GCA	33	31-60	HP+PK		Moderate
3GOG	65	61-90	MWJ+PK	4WJ	Difficult
3GX2	94	91-120	MWJ+PK	4WJ	Difficult
3K1V	34	31-60	HP+PK		Moderate
3LA5	71	61-90	MWJ+PK	4WJ	Difficult
3NPN	54	31-60	MWJ+PK	3WJ	Difficult
3NPQ	54	31-60	MWJ+PK	3WJ	Difficult
3OWI	88	61-90	MWJ	3WJ	Moderate
3Q3Z	75	61-90	HP+PK		Moderate
3RKF	67	61-90	MWJ+PK	4WJ	Difficult
3SD3	89	61-90	MWJ+PK	3WJ	Difficult
3SKL	66	61-90	MWJ+PK	4WJ	Difficult
3SKZ	68	61-90	MWJ+PK	4WJ	Difficult
3SUH	101	91-120	MWJ+PK	3WJ	Difficult
4FEJ	67	61-90	MWJ+PK	4WJ	Difficult
4FEL	67	61-90	MWJ+PK	4WJ	Difficult
4FEN	67	61-90	MWJ+PK	4WJ	Difficult
4FEO	67	61-90	MWJ+PK	4WJ	Difficult
4FEP	67	61-90	MWJ+PK	4WJ	Difficult
4FRG	84	61-90	MWJ+PK	4WJ	Difficult
4FRN	102	91-120	MWJ+PK	5WJ	Difficult
4GMA	210	181-210	MWJ+PK	4WJ+3WJ	Difficult
4GXY	172	151-180	MWJ+PK	4WJ+3WJ	Difficult
4JF2	77	61-90	HP+PK		Moderate
4L81	96	91-120	MWJ+PK	4WJ	Difficult
4LVV	89	61-90	MWJ+PK	3WJ	Difficult
4LX5	71	61-90	MWJ+PK	4WJ	Difficult
4QK8	122	121-150	MWJ+PK	3WJ	Difficult
4QK9	124	121-150	MWJ+PK	3WJ	Difficult
4QLM	125	121-150	MWJ+PK	3WJ	Difficult
4RZD	101	91-120	MWJ+PK	3WJ	Difficult
4TZX	71	61-90	MWJ+PK	4WJ	Difficult
4XW7	64	61-90	Simple HP		Simple
4Y1M	107	91-120	MWJ	4WJ	Moderate
4YAZ	84	61-90	MWJ+PK	3WJ	Difficult
4YB0	83	61-90	MWJ+PK	3WJ	Difficult
4ZNP	73	61-90	HP+PK		Moderate
5BTP	75	61-90	HP+PK		Moderate

5D5L	77	61-90	HP+PK		Moderate
5KPY	71	61-90	MWJ+PK	3WJ	Difficult
5LWJ	34	31-60	HP+PK		Moderate
5NEF	18	1-30	Simple HP		Simple
5NWQ	41	31-60	HP+PK		Moderate
5NZ6	41	31-60	HP+PK		Moderate
5OB3	69	61-90	Simple HP		Simple
5T83	95	91-120	HP+PK		Moderate
5U3G	85	61-90	MWJ+PK	3WJ	Difficult
5V3F	31	31-60	Simple HP		Simple
6C63	36	31-60	Simple HP		Simple
6C65	36	31-60	Simple HP		Simple
6CC3	101	91-120	MWJ	4WJ	Moderate
6CK4	117	91-120	MWJ	4WJ	Moderate
6CK5	117	91-120	MWJ+PK	4WJ	Difficult
6E81	36	31-60	HP+G4		Difficult
6E82	36	31-60	HP+G4		Difficult
6E8S	38	31-60	HP+G4		Difficult
6E8T	36	31-60	HP+G4		Difficult
6E8U	37	31-60	HP+G4		Difficult
6GZK	48	31-60	MWJ	3WJ	Moderate
6HAG	43	31-60	HP+PK		Moderate
6N5K	125	121-150	MWJ+PK	3WJ	Difficult
6N5N	125	121-150	MWJ+PK	3WJ	Difficult
6N5O	126	121-150	MWJ+PK	3WJ	Difficult
6N5P	127	121-150	MWJ+PK	3WJ	Difficult
6N5Q	128	121-150	MWJ+PK	3WJ	Difficult
6N5S	123	121-150	MWJ+PK	3WJ	Difficult
6N5T	126	121-150	MWJ+PK	3WJ	Difficult
6P2H	69	61-90	MWJ+PK	4WJ	Difficult
6SVS	77	61-90	HP+PK		Moderate
6VMY	148	121-150	MWJ+PK	4WJ+3WJ	Difficult
6XB7	41	31-60	Simple HP		Simple
6XKN	101	91-120	MWJ+PK	3WJ	Difficult
6XKO	101	91-120	MWJ+PK	3WJ	Difficult
7D7W	51	31-60	HP+PK		Moderate
7EOG	48	31-60	Simple HP		Simple
7KVT	83	61-90	MWJ+PK	3WJ	Difficult
7MKT	23	1-30	G4		Difficult
7MKY	66	61-90	HP+PK		Moderate
7OA3	51	31-60	Simple HP		Simple
7RWR	38	31-60	Simple HP		Simple
7WI9	50	31-60	Simple HP		Simple
8D2A	33	31-60	Simple HP		Simple
8EYU	49	31-60	HP+G4		Difficult

8FB3	34	31-60	HP+PK		Moderate
8HB8	55	31-60	HP+PK		Moderate
8HBA	56	31-60	HP+PK		Moderate
8I43	19	1-30	Simple HP		Simple
8I44	19	1-30	Simple HP		Simple
8I45	19	1-30	Simple HP		Simple
8I46	19	1-30	Simple HP		Simple

\*HP, PK, MWJ and G4 stand for hairpins, pseudoknots, multi-way junctions and G-quadruplexes, respectively.

†3WJ, 4WJ and 5WJ stands for three-way, four-way and five-way junctions, respectively.

**Supplementary Table S2: Blind datasets 1 (B1) and 2 (B2) of RNA structures for benchmarking.**

PDB ID	Length	Length range	Structure Class*	Type of MWJ <sup>†</sup>	Class	Datasets <sup>‡</sup>
7EOG	48	31-60	Simple HP		Simple	B2
7MKT	23	1-30	G4		Difficult	B1, B2
7RWR	38	31-60	Simple HP		Simple	B2
7V9E	68	31-60	MWJ	3WJ	Difficult	B2
7WI9	50	31-60	Simple HP		Simple	B1, B2
8EYU	49	31-60	HP+G4		Difficult	B1, B2
8F0N	49	31-60	HP+G4		Difficult	B2
8FB3	34	31-60	HP+PK		Moderate	B1, B2
8FZA	30	1-30	HP+PK		Moderate	B1, B2
8HB8	55	31-60	HP+PK		Moderate	B1, B2
8I43	19	1-30	Simple HP		Simple	B1, B2

\*HP, PK, MWJ and G4 stand for hairpins, pseudoknots, multi-way junctions and G-quadruplexes, respectively.

<sup>†</sup>3WJ stands for three-way junctions, respectively.

<sup>‡</sup>Indicates the presence of the PDB ID in the blind sets B1 and B2.

**Supplementary Table S3: Run time for selected RNAs of different lengths.**

PDB ID	Length	No of CPUs	Run Time (in seconds)*
RhoFold			
1AJU	30	32	1784.433
2MIY	59	32	4052.317
4LVV	89	32	6434.864
6CK4	117	32	6938.916
6VMY	148	32	37196.180
3DIG	174	32	8537.104
4GMA	210	32	70812.830
BRIQ			
1AJU	30	1	527.082
2MIY	59	1	2572.98
4LVV	89	1	5681.28
6CK4	117	1	10499.8
6VMY	148	1	21467
3DIG	174	1	26398.6
4GMA	210	1	60281.2
FARFAR2			
1AJU	30	1	4930
2MIY	59	1	17976
4LVV	89	1	54300
6CK4	117	1	57470
6VMY	148	1	145241
3DIG	174	1	131948
4GMA	210	1	194990
FARFAR2_ss			
1AJU	30	1	17,695
2MIY	59	1	24,576
4LVV	89	1	91274
6CK4	117	1	40029
6VMY	148	1	162503
3DIG	174	1	200086
4GMA	210	1	199317
SimRNA			
1AJU	30	80	57257.210
2MIY	59	80	133566.612
4LVV	89	80	213615.481
6CK4	117	80	276932.503
6VMY	148	80	391330.501
3DIG	174	80	445124.257
4GMA	210	80	543073.527
SimRNA_ss			
1AJU	30	80	60383.902
2MIY	59	80	140421.915

4LVV	89	80	251122.690
6CK4	117	80	304637.303
6VMY	148	80	394999.842
3DIG	174	80	505577.885
4GMA	210	80	574832.954

\*The running time information is missing for DeepFoldRNA and Vfold2 simulations are missing as we couldn't estimate it from the log files for these programs.



**Supplementary Table S4: Correlation of Interface RMSD with RMSD Values and TM Scores.**

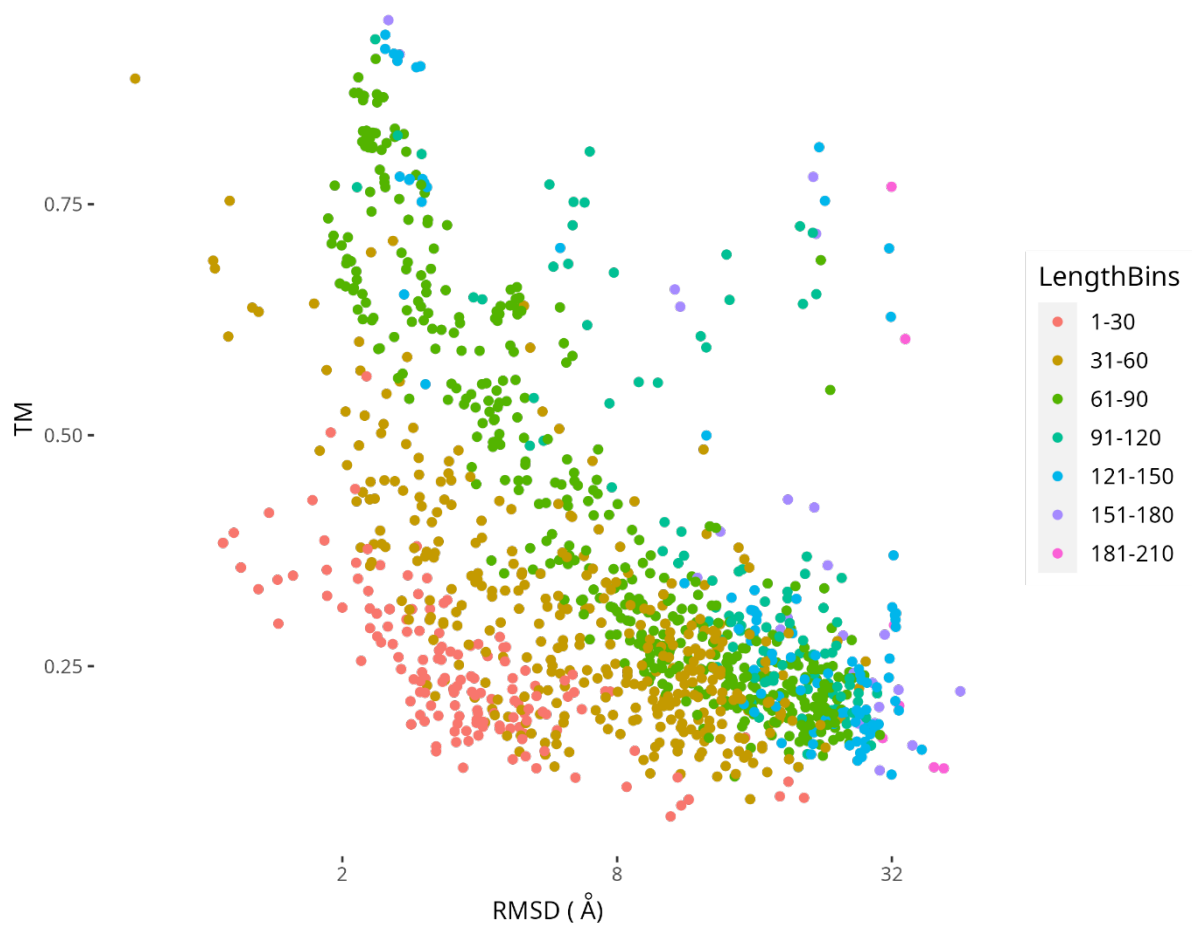
Method	RMSD		TM*	
	Correlation‡	p-value‡	Correlation‡	p-value‡
I-RMSD (entire dataset)				
DeepFoldRNA	0.80	2.16E-32	-0.33	2.82E-03
RhoFold	0.78	6.73E-29	-0.06	5.70E-01
BRIQ	0.80	5.03E-32	-0.35	1.82E-03
FARFAR2	0.82	3.53E-35	-0.39	3.95E-04
FARFAR2_ss	0.81	6.35E-34	-0.64	3.11E-10
SimRNA	0.87	5.70E-44	-0.54	2.60E-07
SimRNA_ss	0.78	8.92E-30	-0.42	1.08E-04
Vfold2	0.92	5.43E-35	-0.82	1.17E-11
Vfold2_ss	0.92	3.31E-50	-0.84	3.62E-18
I-RMSD (<=2.5 Å)†				
DeepFoldRNA	-0.17	1.80E-01	0.33	3.06E-02
RhoFold	0.03	8.69E-01	0.22	3.34E-01
BRIQ	-0.77	2.50E-02	—	—
FARFAR2	0.49	3.29E-01	—	—
FARFAR2_ss	-0.27	1.98E-01	0.33	3.44E-01
SimRNA	-0.37	2.37E-01	—	—
SimRNA_ss	-0.19	5.41E-01	—	—
Vfold2	-0.43	9.14E-03	0.23	2.90E-01
Vfold2_ss	0.62	7.74E-06	-0.54	8.48E-02

\*The TM values for RNAs with length  $\geq 60$  nucleotides are included in the calculations.

‡The correlation coefficients and the statistical significance (p-values from a two-tailed t-test).

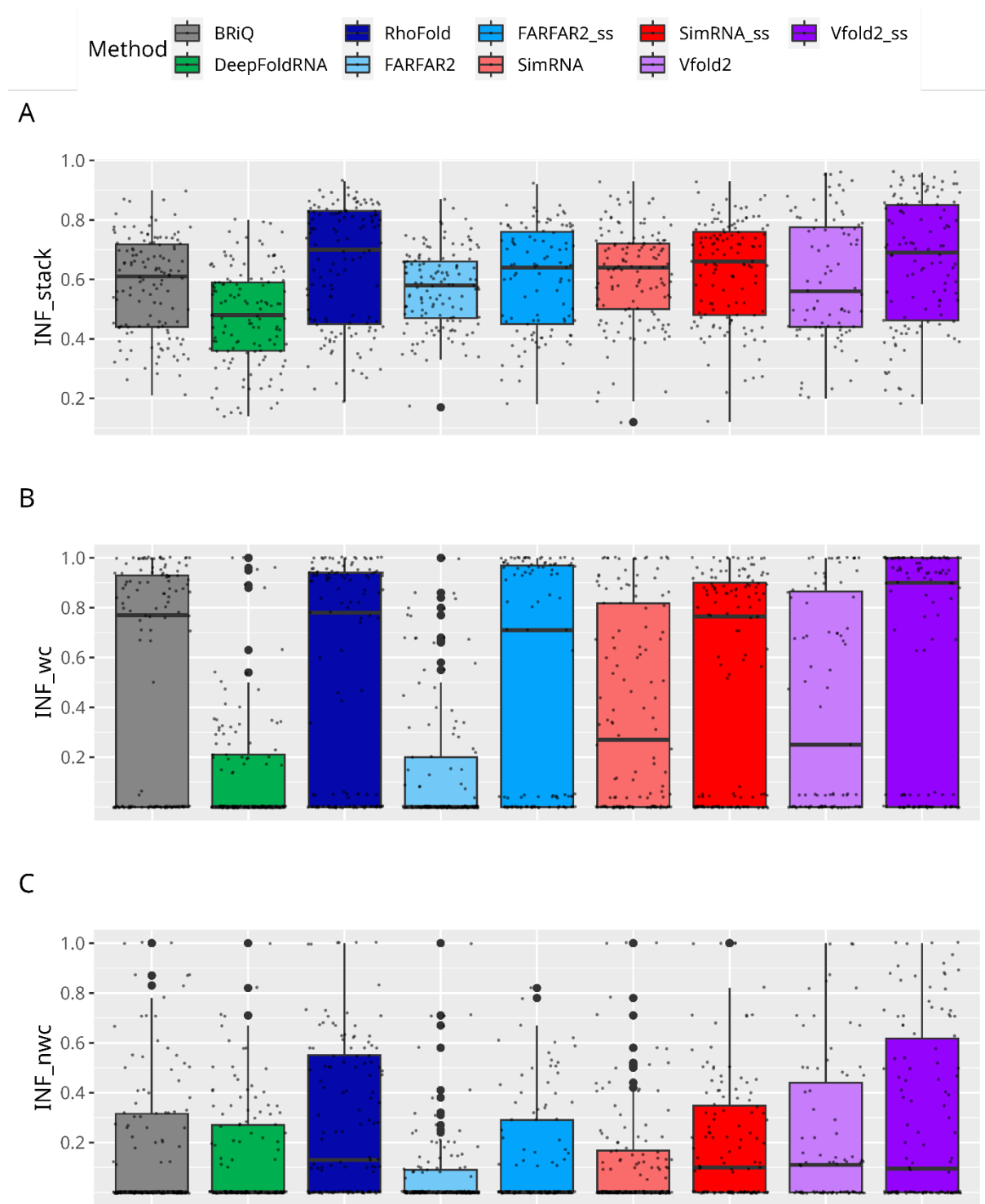
†The values were not calculated when the number of points were insufficient to calculate the correlations and are denoted by '—' in the table.

### Supplementary Figure S1:



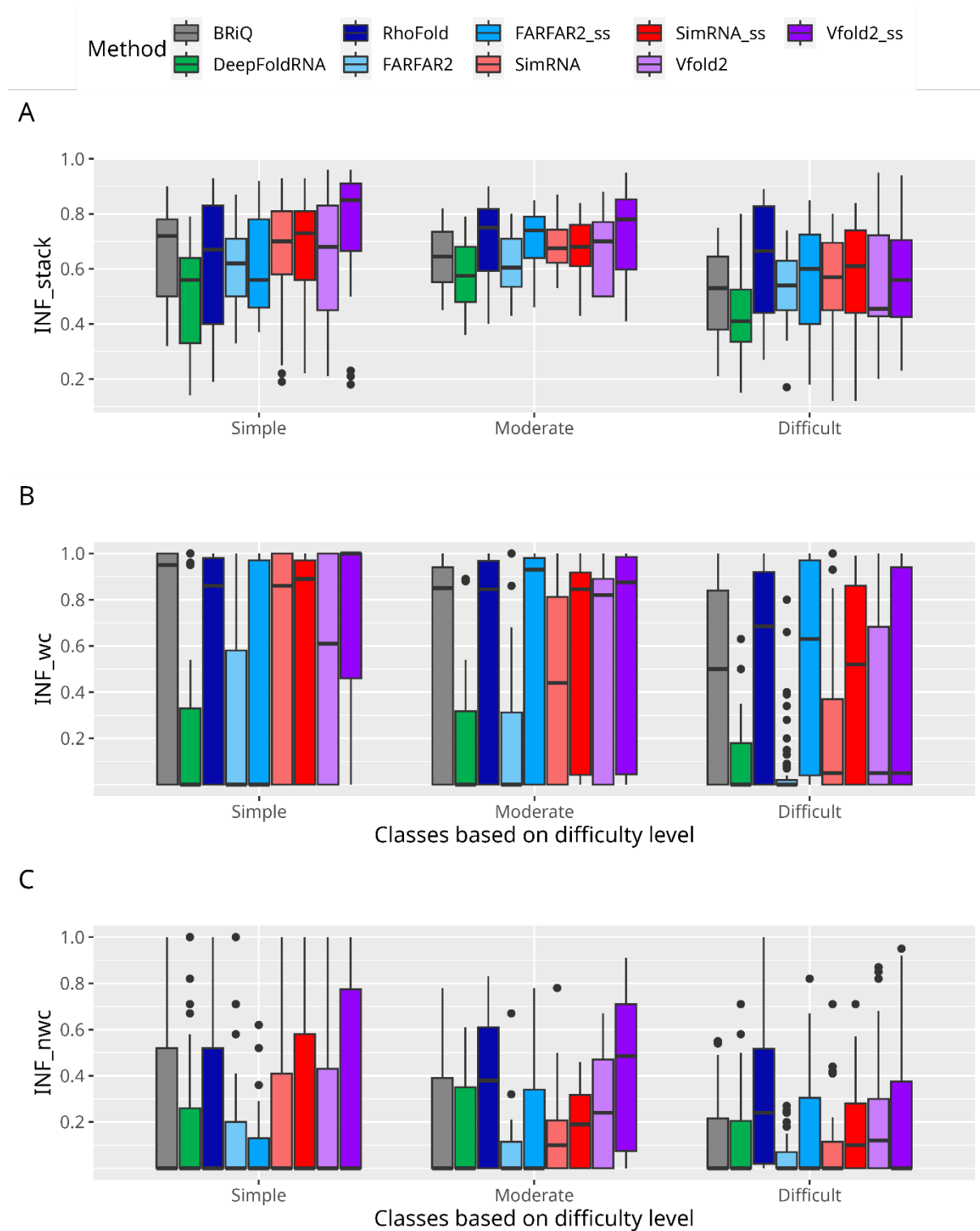
**Figure S1: RMSD vs TM score.** The points colored based on length of RNA with each color representing a 30-nucleotide length bin. The TM values for RNA molecules up to a length of 60 nucleotides are unreliable.

**Supplementary Figure S2:**



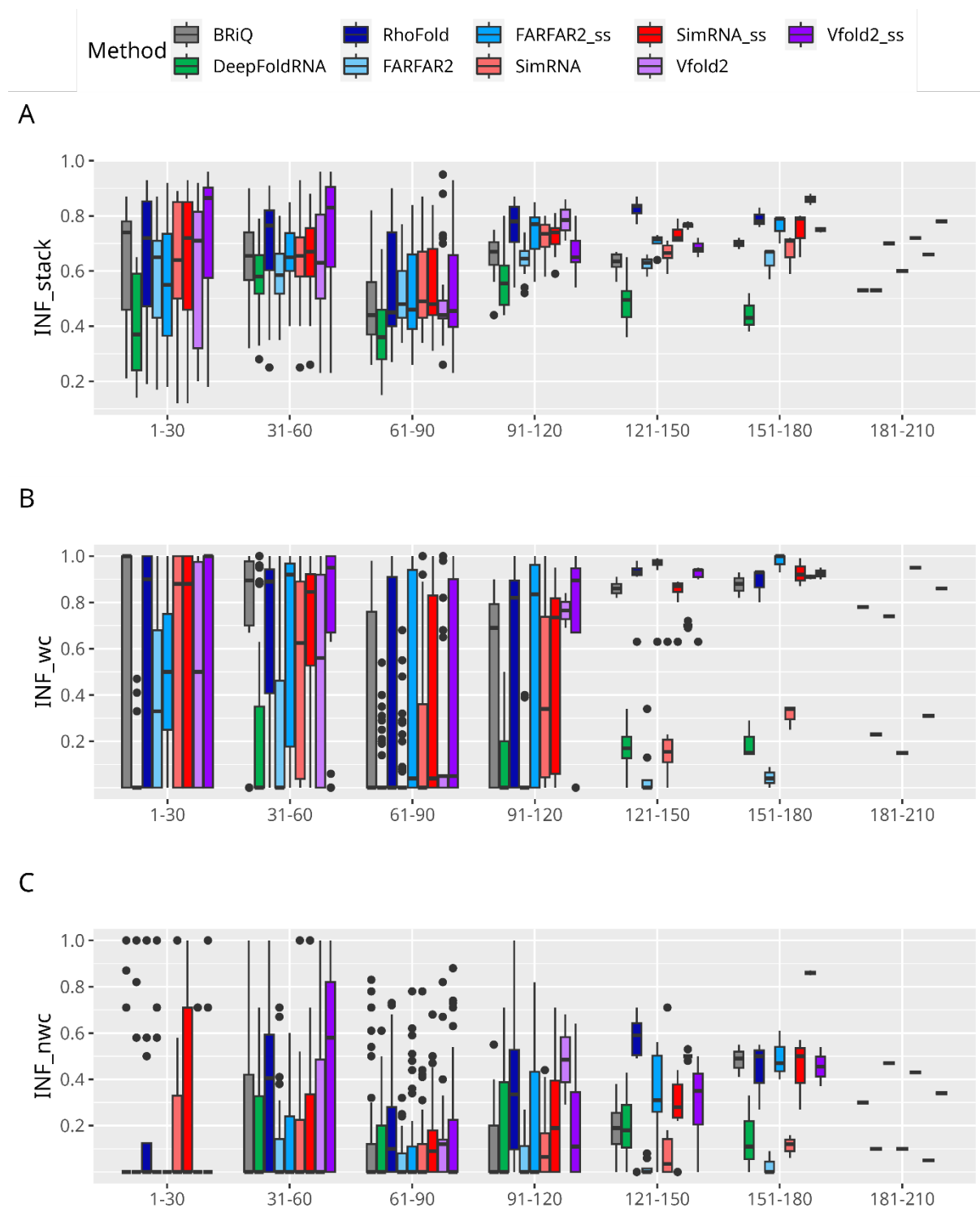
**Figure S2: INF measures for overall dataset:** (A) INF<sub>stack</sub>, (B) INF<sub>wc</sub>, and (C) INF<sub>nwc</sub>. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

### Supplementary Figure S3:



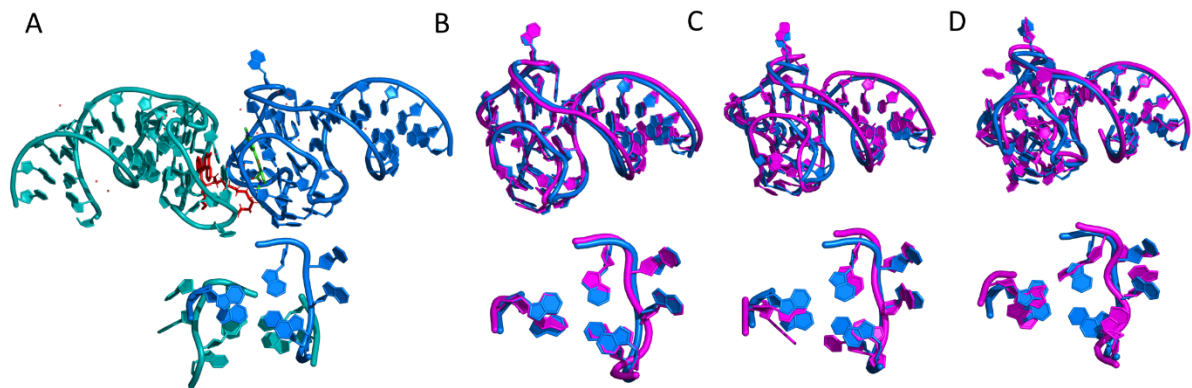
**Figure S3: INF measures across different difficulty classes:** (A) INF<sub>stack</sub>, (B) INF<sub>wc</sub>, and (C) INF<sub>nwc</sub>. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

### Supplementary Figure S4:



**Figure S4: INF measures across different lengths of RNAs.** (A) INF<sub>stack</sub>, (B) INF<sub>wc</sub>, and (C) INF<sub>nwc</sub> are organized into sets of box plots representing 30-nucleotide length bins. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

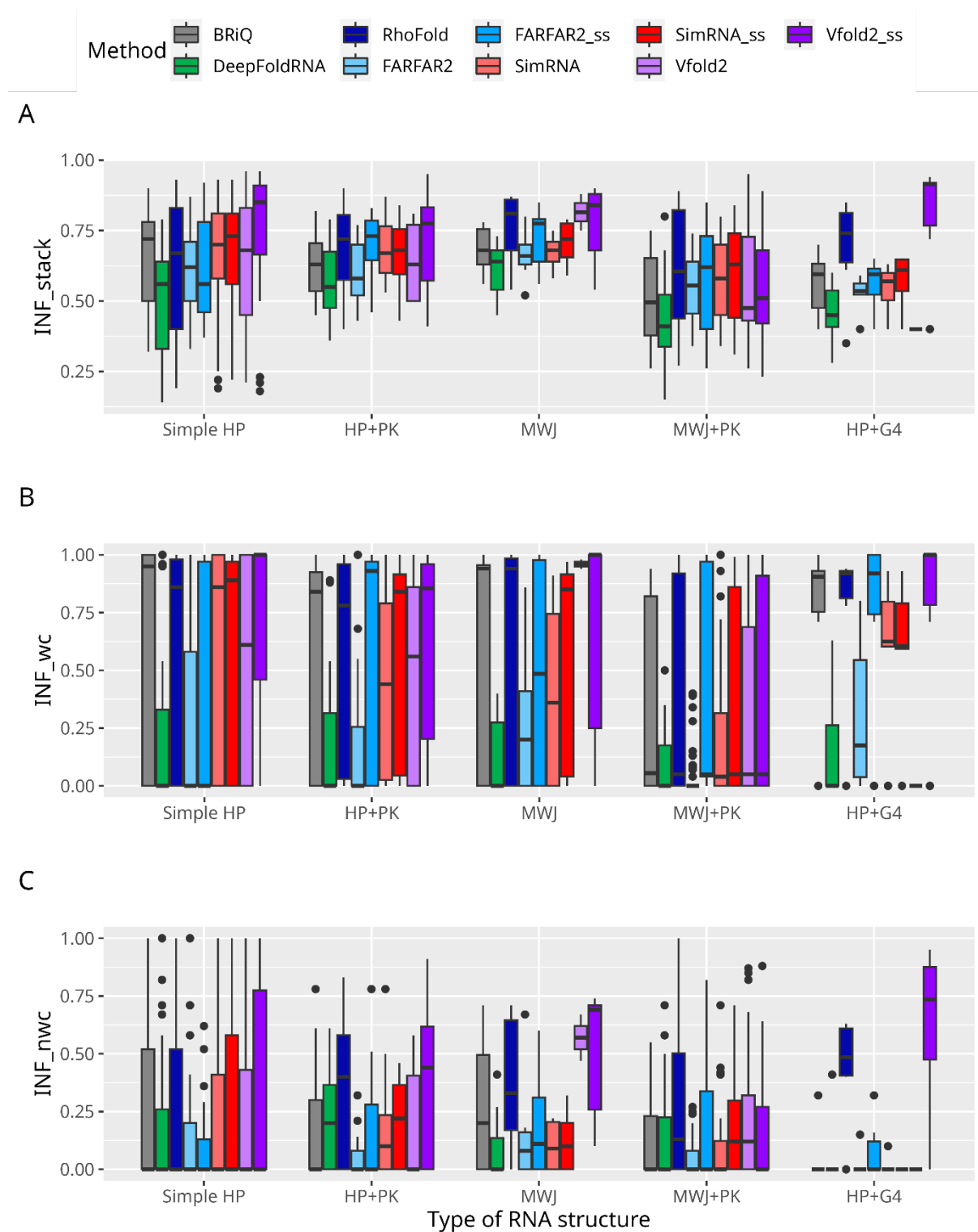
### Supplementary Figure S5:



**Figure S5: Example of modeling the G4 element within a hairpin (HP) structure.**

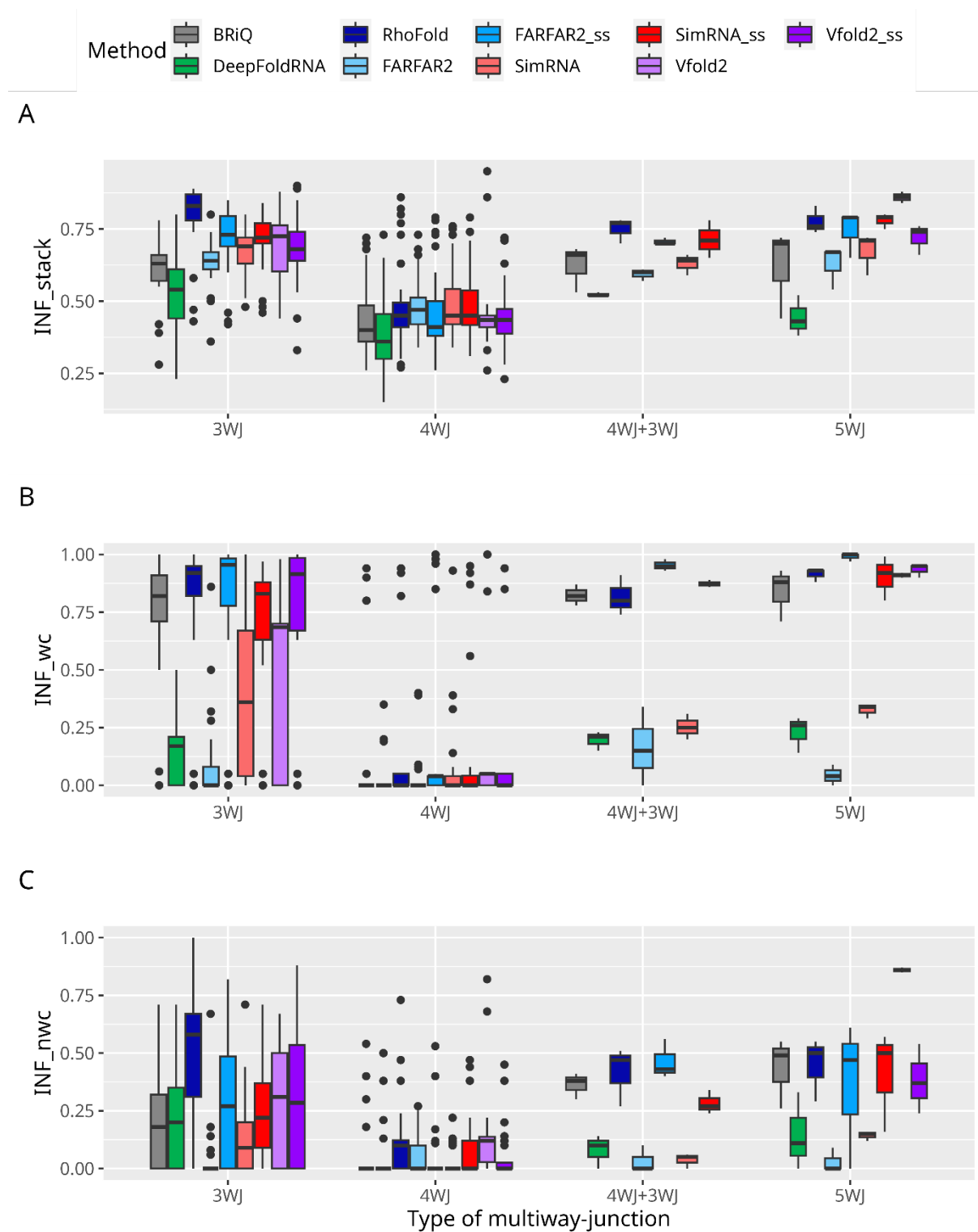
(A) The XRD structure of iMango-III aptamer features two chains in the asymmetric unit colored in teal and marine (PDB ID: 6E8S). The biologically relevant assembly includes only one RNA chain with a length of 38 nucleotides and was modeled with all methods. Only three methods were able to model the G4 part of the structure and their superpositions to native structures are shown here. For these models, each panel displays two images: the upper image showing the superpositions of models (magenta) to the native structure (marine), and the lower image highlighting the superpositions of corresponding ligand binding regions. (B) Vfold2 fails to accurately model this structure, however, the Vfold2\_ss accurately predicts the RNA (RMSD = 1.05 Å, INF\_all = 0.93, INF\_stack = 0.92, INF\_wc = 1, and INF\_nwc = 0.92). The ligand binding region which involves the G4 is well-modeled by Vfold2\_ss, showing minimal deviation from the reference structure (I-RMSD = 0.99 Å). (C) The RhoFold model is less accurate than Vfold2\_ss (RMSD = 2.03 Å, INF\_all = 0.77, INF\_stack = 0.83, INF\_wc = 0.94, and INF\_nwc = 0.42). The local binding site also shows larger deviation than Vfold2\_ss (I-RMSD = 2.14 Å). (D) The DeepFoldRNA model exhibits certain limitations in accurately capturing the structure's complexity (RMSD = 2.77 Å, INF\_all = 0.58, INF\_stack = 0.36, INF\_wc = 0.43, and INF\_nwc = 0.35). The deviation in binding site prediction is highest among the three models (I-RMSD = 3.35 Å). This underscores the complexities inherent in accurately predicting RNA structures with G4 elements. While Vfold2\_ss demonstrates notable success, both RhoFold and DeepFoldRNA exhibit varying degrees of challenges, particularly in capturing the intricate non-Watson-Crick interactions and local binding site conformations.

## Supplementary Figure S6:



**Figure S6: INF measures across different structural classes:** (A) INF<sub>stack</sub>, (B) INF<sub>wc</sub>, and (C) INF<sub>nwc</sub>. HP, PK, MWJ and G4 stand for hairpins, pseudoknots, multi-way junctions and G-quadruplexes, respectively. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

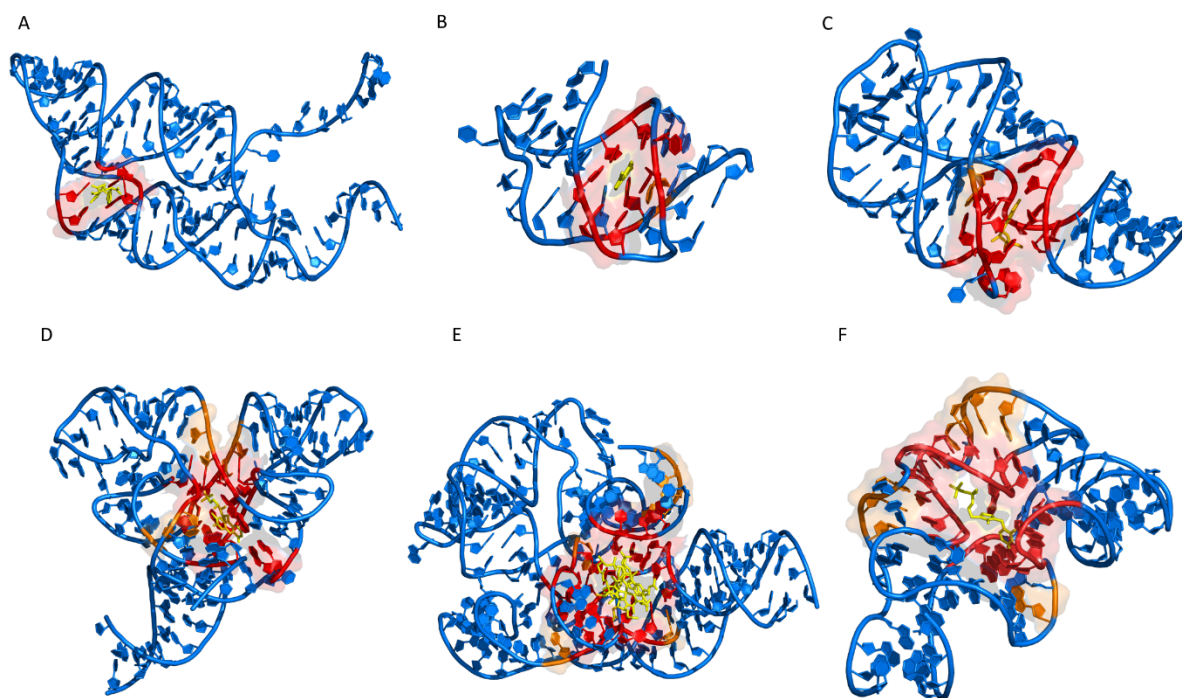
## Supplementary Figure S7:



**Figure S7: INF measures across different multiway junctions:** (A) INF\_stack, (B) INF\_wc, and (C) INF\_nwc. 3WJ, 4WJ and 5WJ stands for three-way, four-way and five-way junctions, respectively. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

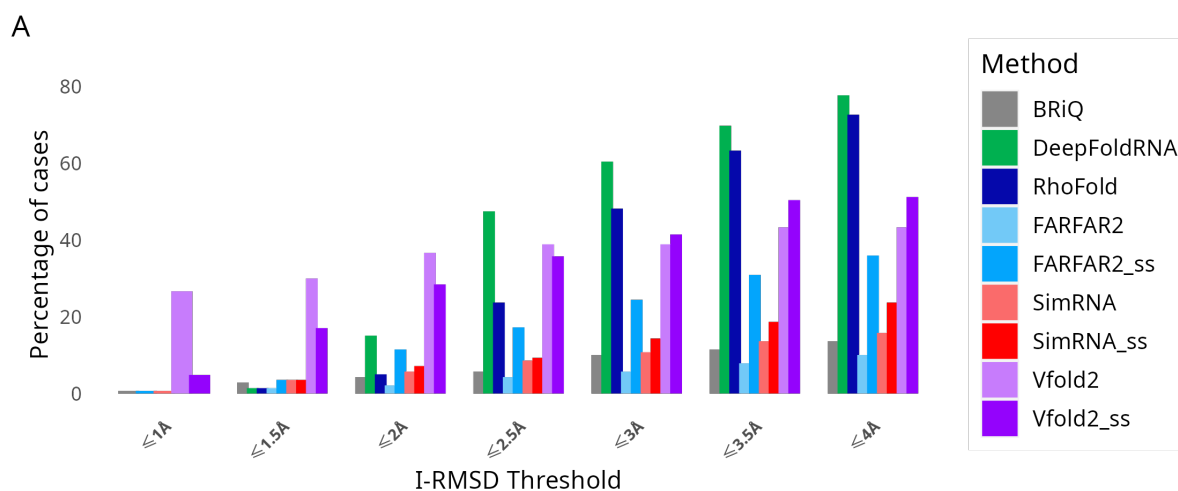


## Supplementary Figure S8:



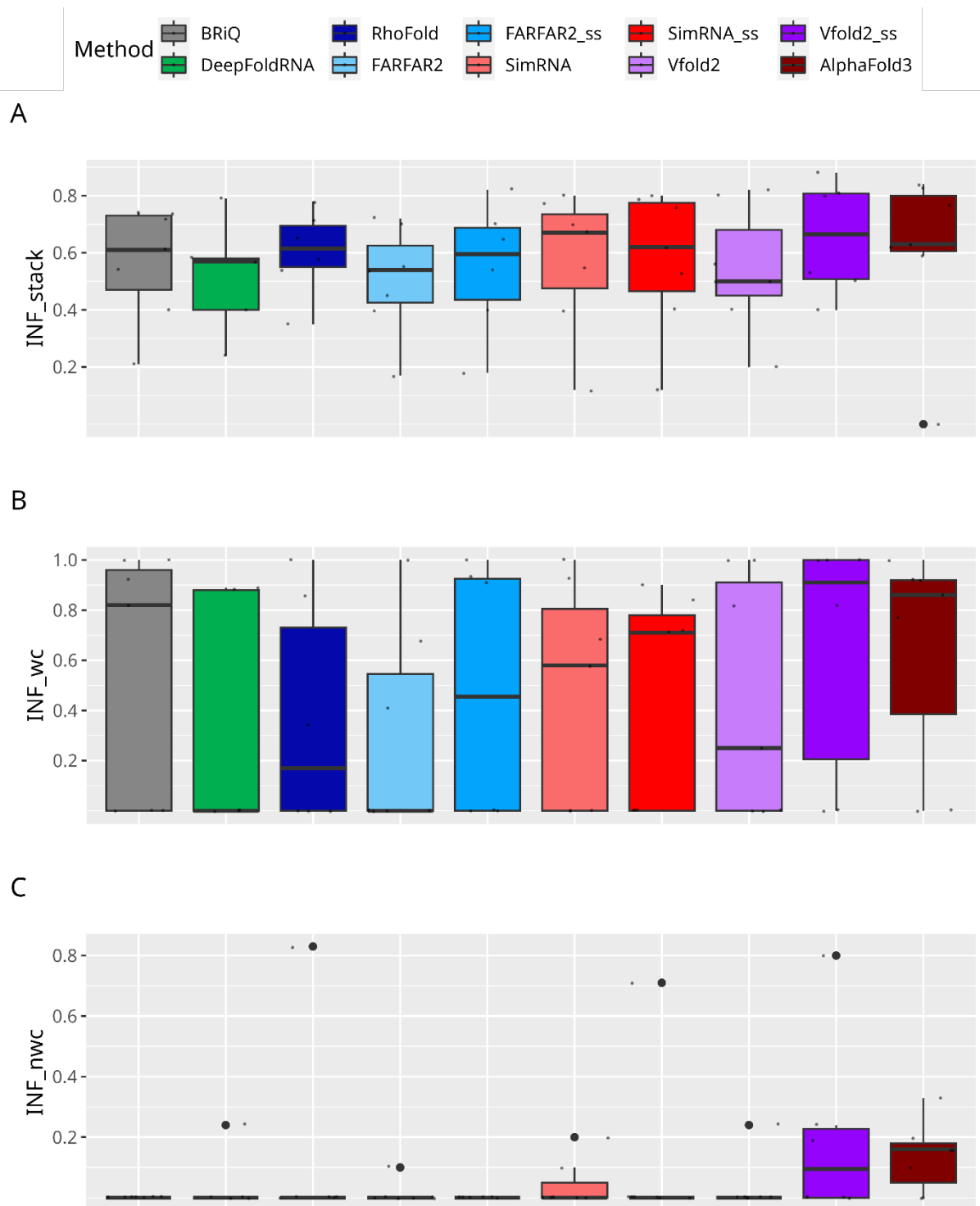
**Figure S8: RNA-Ligand complexes with varying interface sizes.** (A) THF riboswitch RNA bound to 5-formyl-THF (PDB ID: 3SUH), (B) preQ1 riboswitch bound to its precursor metabolite (PDB ID: 3GCA), (C) the 2'-deoxyguanosine (dG) riboswitch bound to guanosine (PDB ID: 3SKZ), (D) Flavin mononucleotide (FMN) riboswitch bound to FMN (PDB ID: 3F2Q), (E) adenosylcobalamin (AdoCbl) riboswitch bound to AdoCbl (PDB ID: 4GXY), and (F) thiamine pyrophosphate (TPP) riboswitch bound to TPP (PDB ID: 2GDI), with binding interfaces of 5, 9, 15, 21, 26, and 27 nucleotides, respectively. Binding residues are colored in red and residues base-pairing with the binding residues are in orange, both shown using both surface and cartoon representations. The remainder of the RNA is visualized exclusively in cartoon representation and colored in marine. Ligands are colored yellow.

## Supplementary Figure S9:



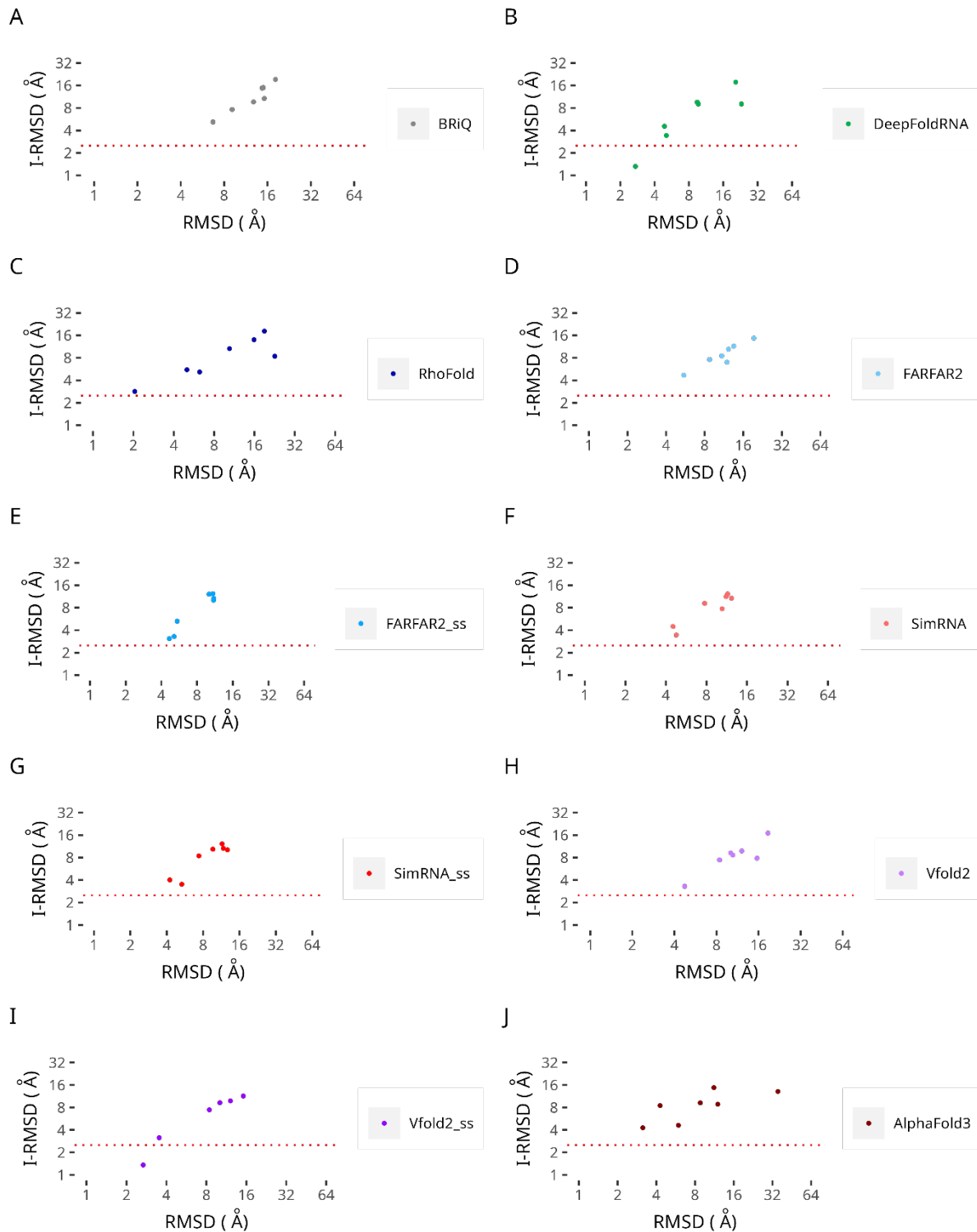
**Figure S9: Interface RMSDs.** (A) Percentage of successful modeling cases achieving an interface RMSD (I-RMSD) at different threshold values. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are depicted in darker shades of the same color. The dotted red line on the scatter plots marks the 2.5 Å I-RMSD.

### Supplementary Figure S10:



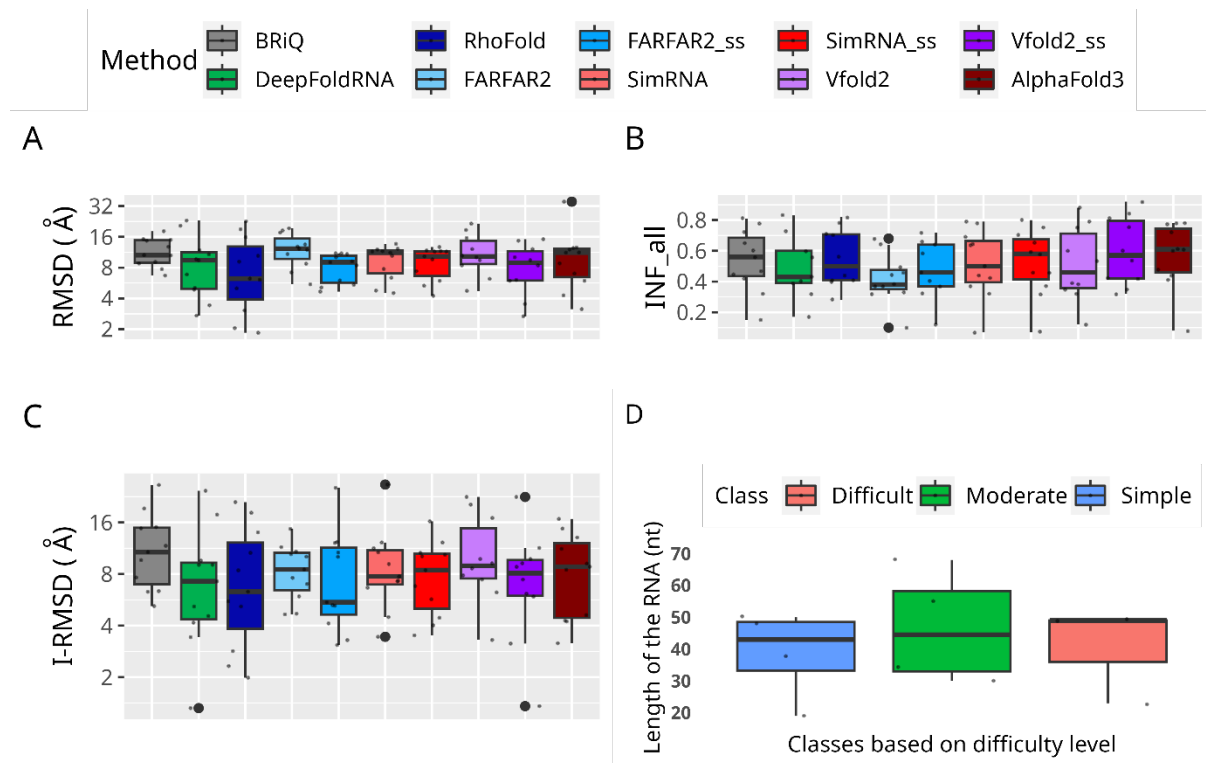
**Figure S10: INF measures for B1 dataset:** (A) INF\_stack, (B) INF\_wc, and (C) INF\_nwc. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

### Supplementary Figure S11:



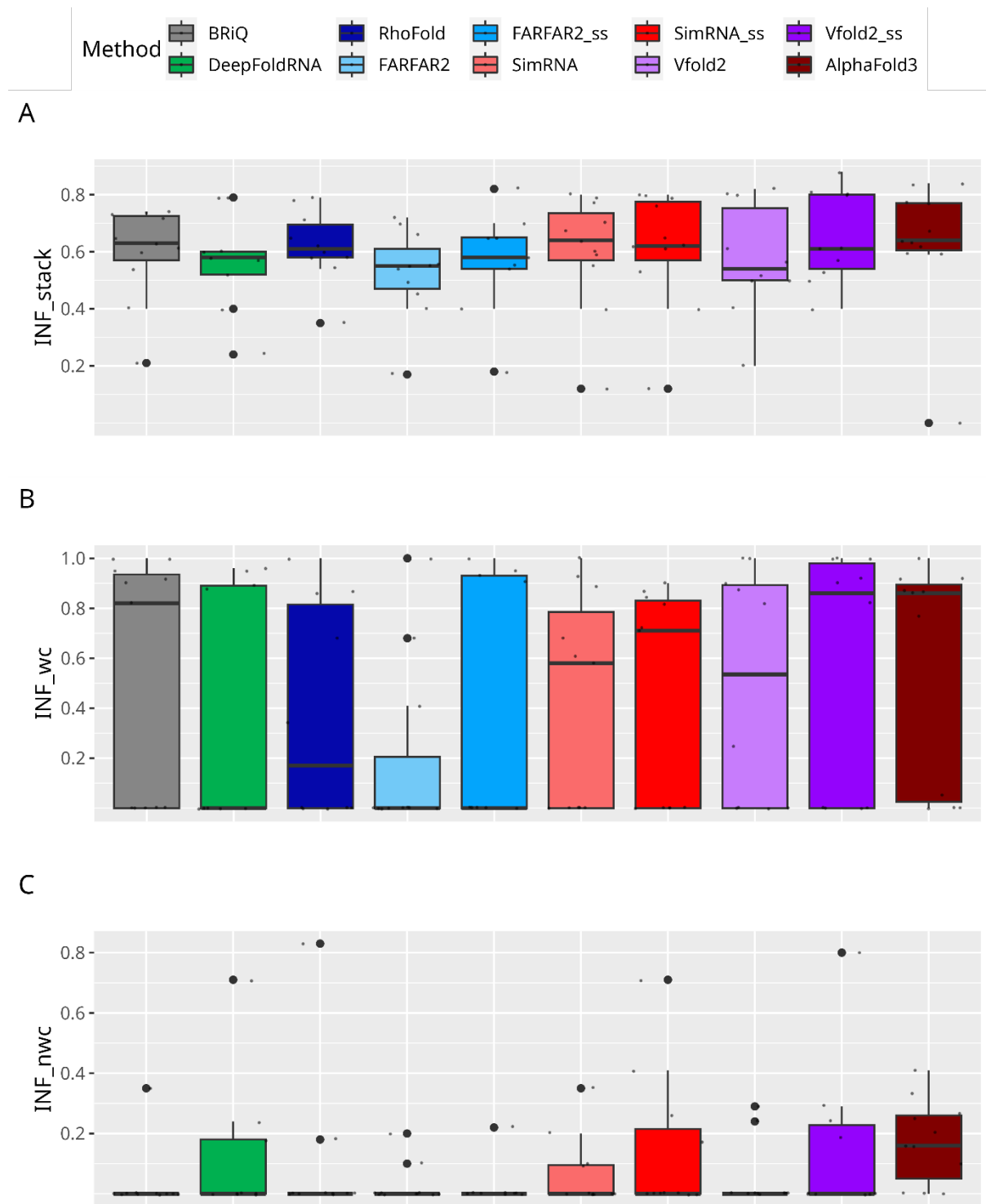
**Figure S11: Plot of I-RMSD vs RMSD for B1 dataset.** Each method is represented by a unique color in panels A-J. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are depicted in darker shades of the same color.

### Supplementary Figure S12:



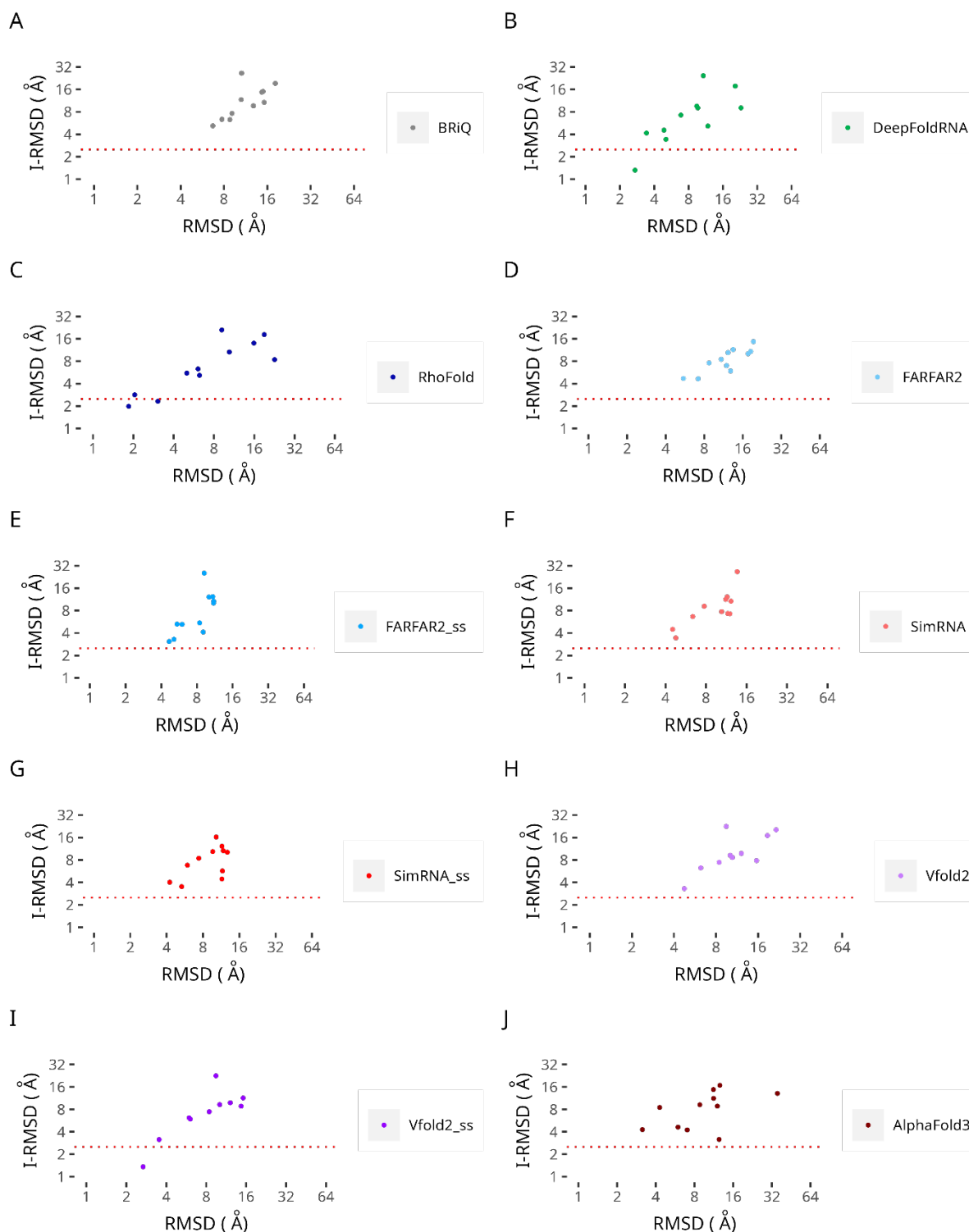
**Figure S12: Performance of 3D modeling methods on blind dataset 2 (B2) containing structures not used in training by the ML methods.** Performance measures for RNA 3D structure prediction methods are displayed across the B2 dataset: (A) RMSD and (B) INF\_all (C)I-RMSD. (D) Length distribution of RNAs on three difficulty classes of modeling - 'simple', 'moderate', and 'difficult'.

**Supplementary Figure S13:**



**Figure S13: INF measures for B2 dataset:** (A) INF<sub>stack</sub>, (B) INF<sub>wc</sub>, and (C) INF<sub>nwc</sub>. Each method is represented by a unique color. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are in darker shades of the same color.

### Supplementary Figure S14:



**Figure S14: Plot of I-RMSD vs RMSD for B2 dataset.** Each method is represented by a unique color in panels A-J. Models generated without secondary structure restraints are shown in lighter shades, while those with restraints are depicted in darker shades of the same color.