Supplementary Data for Generating Atomic Coordinates from RNA Graphs by Fragment Assembly

Swati Jain[1] and Tamar Schlick[1,2,*]

[1]Department of Chemistry, New York University, New York, NY, USA
[2]Courant Institute of Mathematical Sciences, New York University, NY, USA
[*]Corresponding author: schlick@nyu.edu

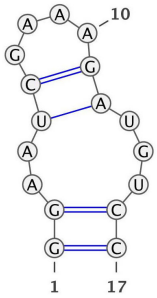
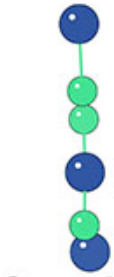
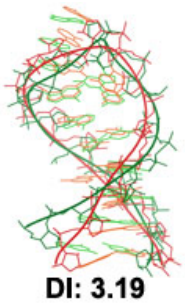
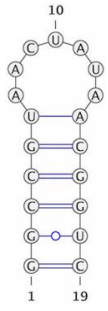
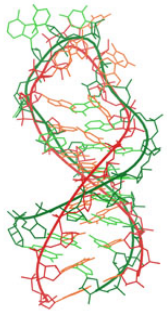
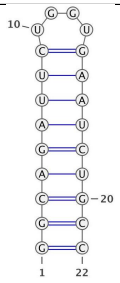
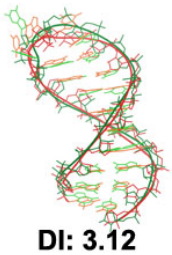
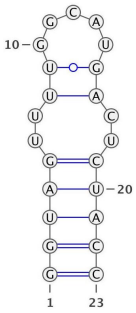
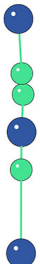
Table S1 : Lowest RMSD and lowest DI atomic models (of the top 20) generated for 50 structures by F-RAG using the fragments returned by RAG-3D search without the additional requirement for same types of loop present in the fragment as the target subgraph.

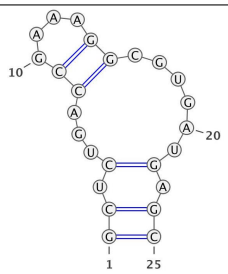
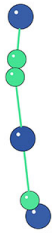
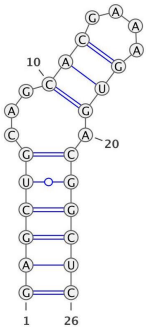
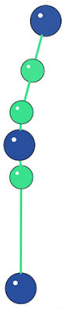
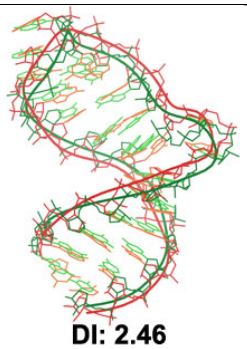
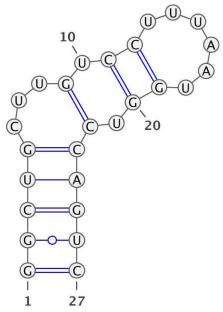
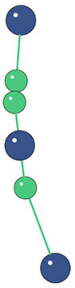
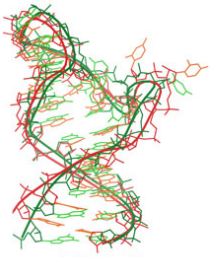
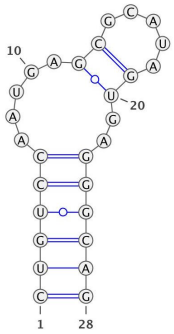
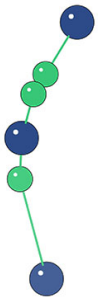
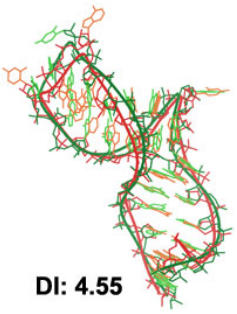
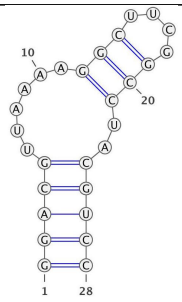
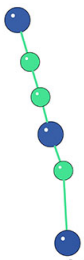
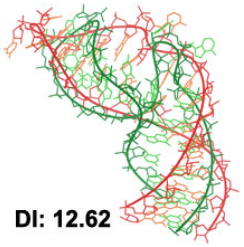
PDB	Lowest RMSD Model (Å)	Lowest DI Model (Å)
2M4W	4.489	5.80
2MEQ	5.857	8.49
2M5U	2.627	3.02
2N7X	3.588	6.07
1RLG	1.236	1.58
2MIS	2.210	2.79
2N0J	2.116	2.62
2NCI	9.916	13.25
3SIU	3.432	4.50
1OOA	5.793	7.90
2IPY	1.577	1.77
2OZB	3.389	3.85
2XEB	5.338	7.03
1MJI	3.246	4.27
2M57	5.063	8.38
4PCJ	1.116	1.23
2HW8	2.109	2.42
2N6S	2.714	3.07
5KQE	9.144	12.19
1I6U	0.828	0.93
1F1T	1.700	1.96
1ZHO	1.763	1.95
2MXL	5.235	6.73
2N6T	4.728	6.58
2N6X	10.285	13.40
5BTM	3.193	4.25
1S03	5.455	6.78
1XJR	7.767	11.16
2MTJ	13.255	21.27
2VPL	6.183	7.92
1U63	5.092	6.19
2PXB	0.687	0.77
2N4L	3.261	4.00
2HGH	3.375	3.99
1DK1	1.008	1.18
1MMS	0.647	0.73
1Y39	0.717	0.81
2N3Q	16.982	25.37
2MQT	6.352	7.94
2N6W	5.998	7.68
1KXK	0.872	1.03
2OIU	N/A	N/A
4LCK	5.162	6.90
1P5O	7.018	10.69
3D2G	1.285	1.52
2HOJ	2.747	3.45
2GDI	3.185	4.08
2GIS	0.737	0.84
1LNG	0.858	0.96
2LKR	20.682	32.91

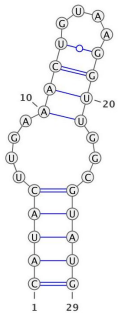
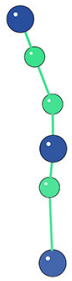
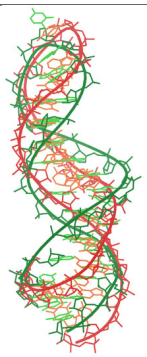
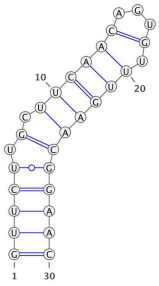
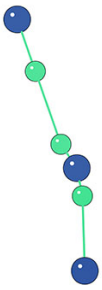
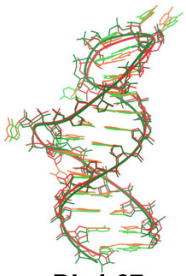
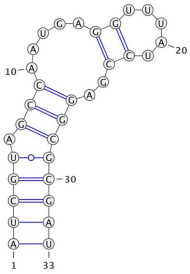
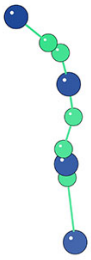
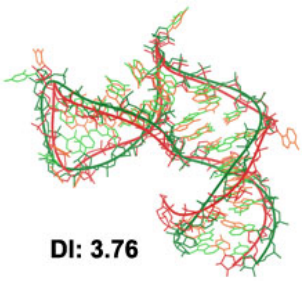
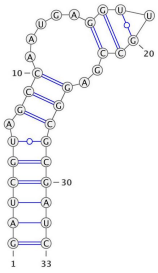
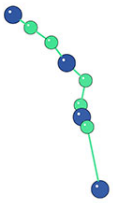
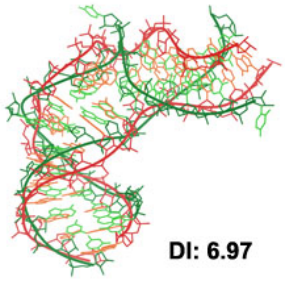
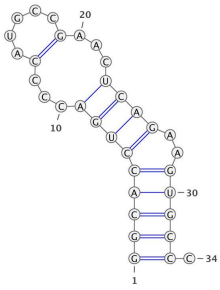
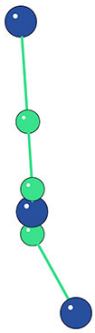
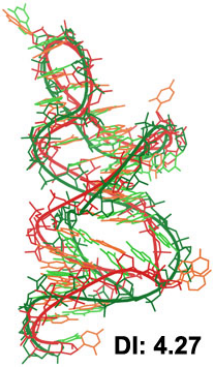
Table S2 : Comparison metrics for the top scoring atomic model (with the highest number of residues) generated by F-RAG.

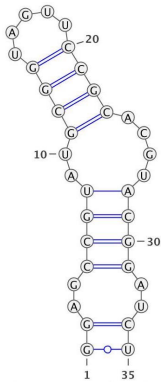
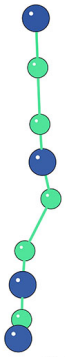
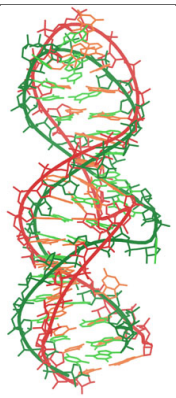
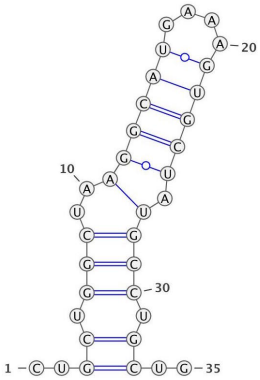
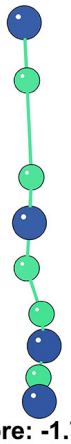
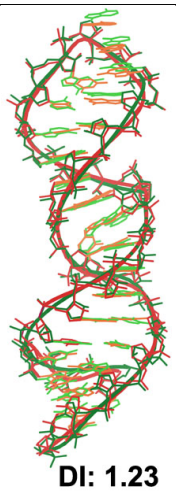
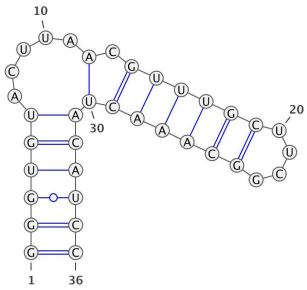
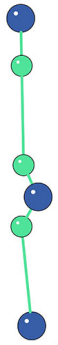
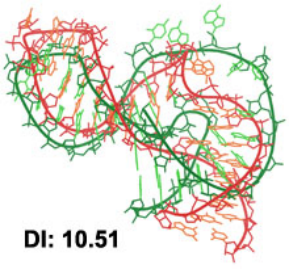
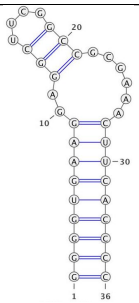
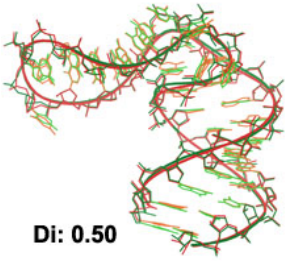
PDB	RMSD(Å)	PPV	STY	INF	DI(Å)	Clashscore
2M4W	4.038	0.75	0.71	0.73	5.55	10.83
2MEQ	6.849	0.73	0.73	0.73	9.34	3.27
2M5U	3.480	0.83	0.91	0.87	4.00	2.83
2N7X	3.625	0.69	0.46	0.56	6.46	1.36
1RLG	1.034	0.90	0.68	0.78	1.32	2.47
2MIS	5.150	0.69	0.80	0.74	6.93	2.36
2N0J	2.170	0.82	0.69	0.75	2.88	2.33
2NCI	11.412	0.73	0.95	0.83	13.70	0.00
3SIU	3.546	0.81	0.75	0.78	4.56	4.40
1OOA	5.005	0.70	0.66	0.68	7.38	10.73
2IPY	5.191	1.00	0.70	0.83	6.22	4.18
2OZB	3.579	0.83	0.94	0.88	4.07	4.70
2XEB	6.014	0.74	0.84	0.79	7.62	1.87
1MJI	3.246	0.87	0.67	0.76	4.27	13.72
2M57	5.416	0.62	0.68	0.65	8.37	14.20
4PCJ	1.116	0.93	0.89	0.91	1.23	0.89
2HW8	2.857	0.94	0.80	0.87	3.28	0.00
2N6S	14.538	0.73	0.54	0.63	23.18	38.36
5KQE	9.835	0.87	0.70	0.78	12.60	0.87
1I6U	3.258	0.90	0.78	0.84	3.90	11.74
1F1T	5.547	0.73	0.61	0.67	8.29	0.81
1ZHO	1.763	0.95	0.86	0.90	1.95	0.81
2MXL	8.419	0.78	0.60	0.68	12.37	20.50
2N6T	6.502	0.65	0.68	0.67	9.75	4.49
2N6X	12.616	0.62	0.88	0.74	16.99	3.64
5BTM	3.237	0.83	0.56	0.68	4.76	2.20
1S03	5.638	0.91	0.72	0.81	6.96	9.89
1XJR	8.622	0.80	0.59	0.69	12.55	3.27
2MTJ	13.255	0.71	0.54	0.62	21.27	3.52
2VPL	0.428	0.98	0.81	0.89	0.48	3.86
1U63	3.408	0.83	0.73	0.78	4.37	1.26
2PXB	5.367	0.90	0.62	0.75	7.15	4.40
2N4L	6.448	0.91	0.72	0.81	7.97	5.90
2HGH	4.563	0.91	0.61	0.74	6.15	1.71
1DK1	2.060	0.85	0.73	0.79	2.61	8.13
1MMS	2.337	0.95	0.71	0.82	2.86	6.43
1Y39	2.036	0.95	0.76	0.85	2.39	5.90
2N3Q	17.061	0.54	0.75	0.64	26.78	5.57
2MQT	12.310	0.80	0.68	0.74	16.65	5.50
2N6W	14.221	0.79	0.72	0.75	18.86	4.17
1KXK	1.800	0.90	0.73	0.81	2.23	3.99
2OIU	14.400	0.95	0.81	0.88	16.43	0.87
4LCK	1.336	0.93	0.76	0.84	1.59	3.72
1P5O	2.176	0.94	0.78	0.86	2.54	8.09
3D2G	2.131	0.90	0.69	0.79	2.70	6.43
2HOJ	2.747	0.92	0.69	0.80	3.45	4.71
2GDI	5.411	0.90	0.60	0.73	7.38	6.98
2GIS	0.929	0.92	0.82	0.87	1.07	5.25
1LNG	2.928	0.91	0.84	0.88	3.34	5.42
2LKR	20.741	0.65	0.61	0.63	32.92	7.75

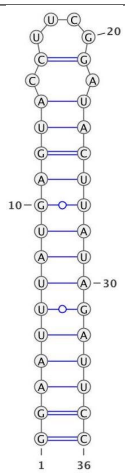
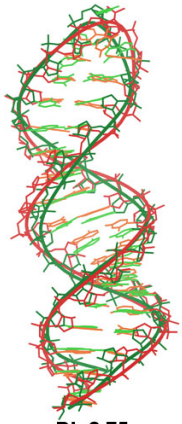
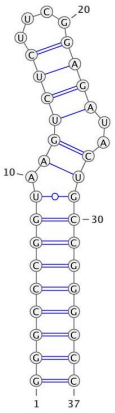
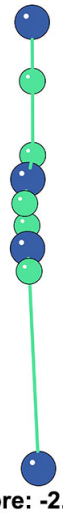
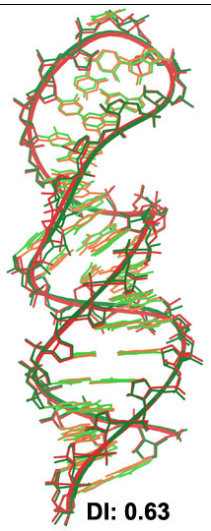
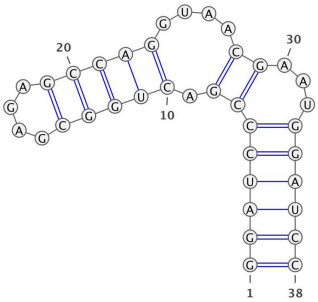
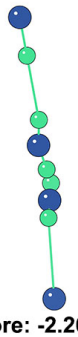
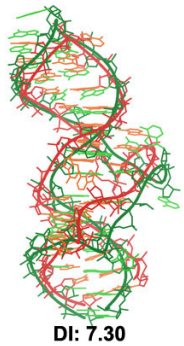
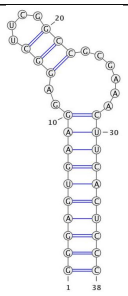
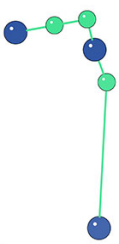
Table S3 : **Images of F-RAG atomic models:** the secondary structure, RAGTOP predicted candidate graph, and the lowest deformation index (DI) atomic model generated by F-RAG are shown for 50 RNA structures. The loop vertices are colored blue and helical vertices are colored sea green in the candidate 3D graphs. The reference model is colored red and the lowest DI atomic model generated by F-RAG is superimposed in green.

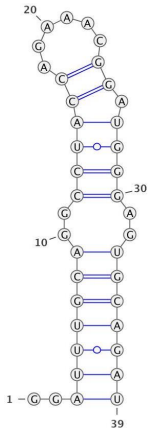
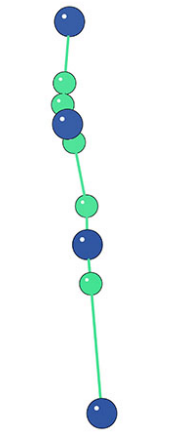
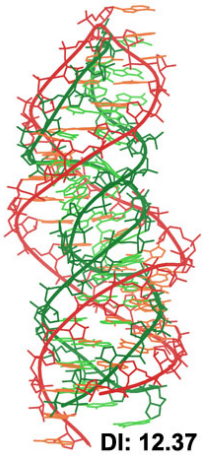
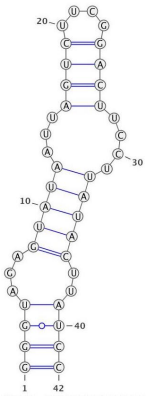
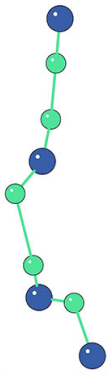
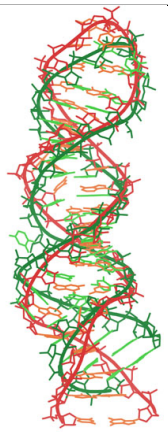
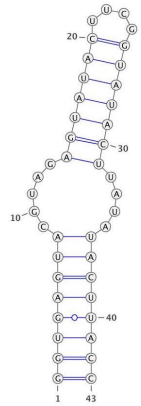
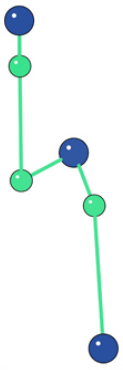
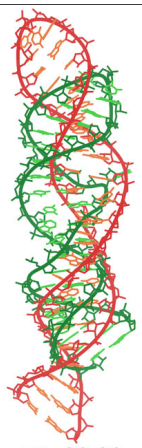
PDB	2D Structure	Candidate graph	Lowest DI atomic model
2M4W		 Score: -0.553	 DI: 3.19
2MEQ		 Score: 0.932	 DI: 8.49
2M5U		 Score: 3.20	 DI: 3.12
2N7X		 Score: -0.784	 DI: 4.68

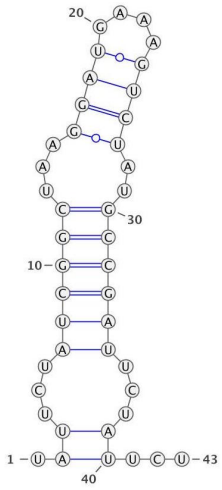
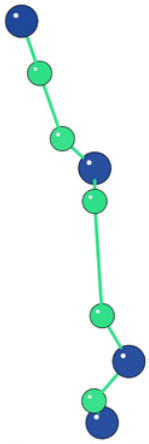
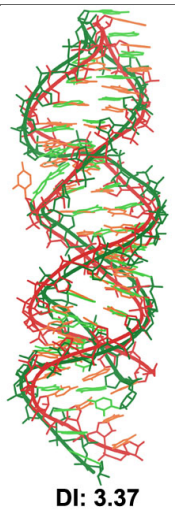
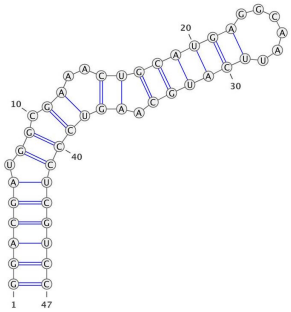
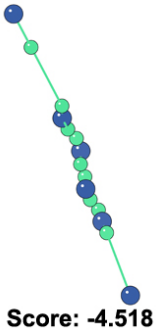
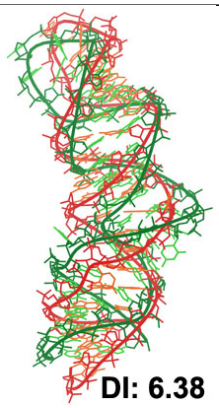
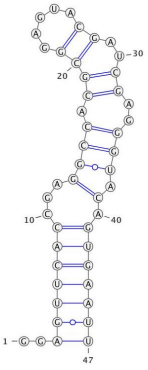
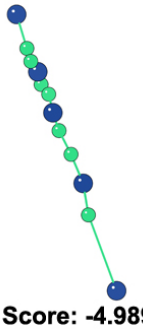
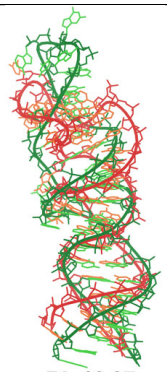
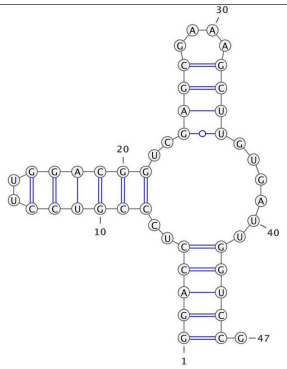
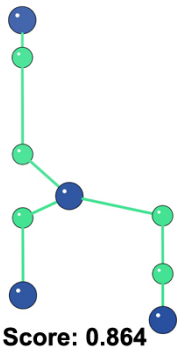
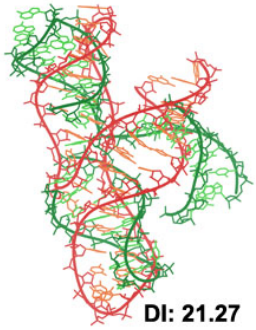
1RLG		 Score: -0.988	 DI: 1.25
2MIS		 Score: -1.532	 DI: 2.46
2N0J		 Score: -0.011	 DI: 2.86
3SIU		 Score: -0.055	 DI: 4.55
2NCI		 Score: -1.373	 DI: 12.62

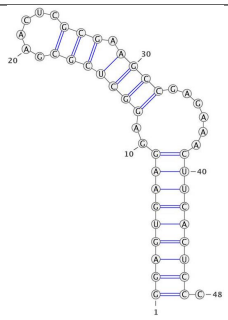
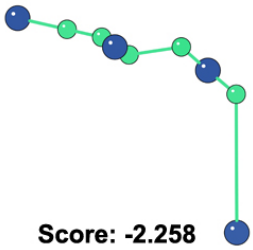
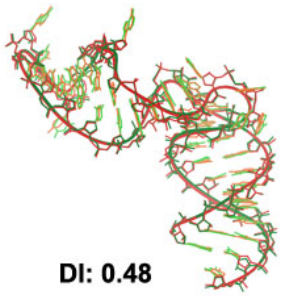
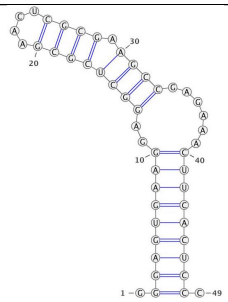
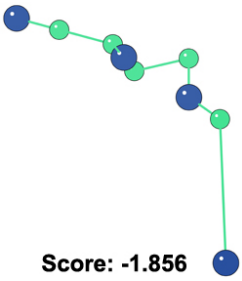
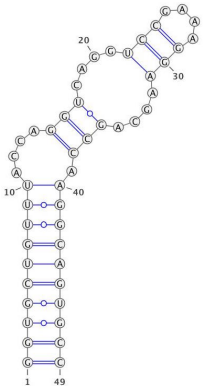
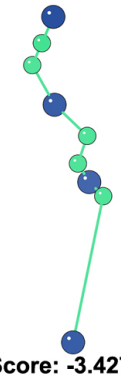
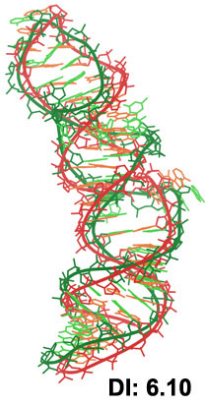
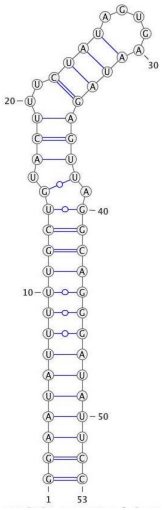
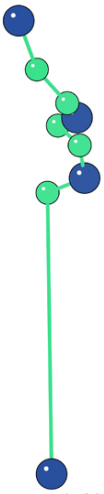
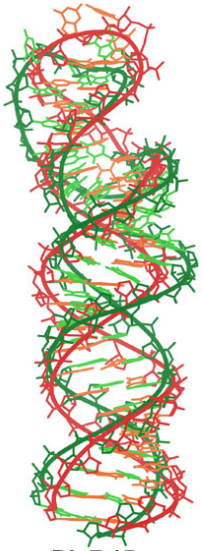
1OOA		 Score: -1.581	 DI: 7.22
2IPY		 Score: -1.586	 DI: 1.67
2OZB		 Score: -1.713	 DI: 3.76
2XEB		 Score: -2.521	 DI: 6.97
1MJI		 Score: 0.805	 DI: 4.27

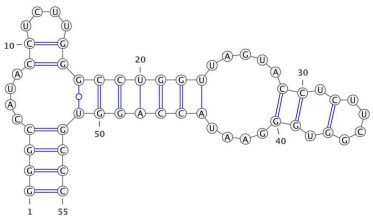
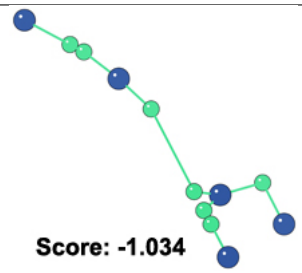
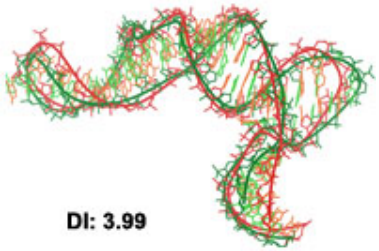
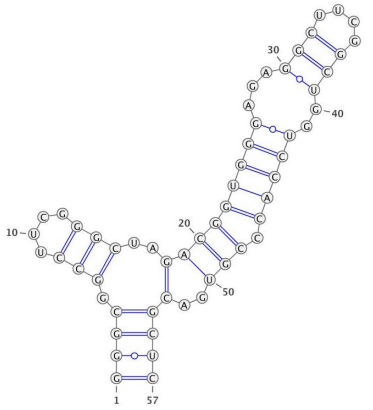
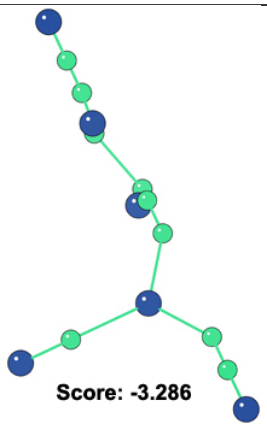
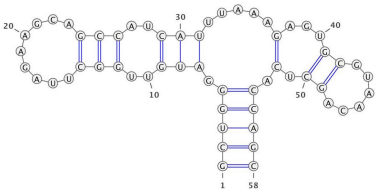
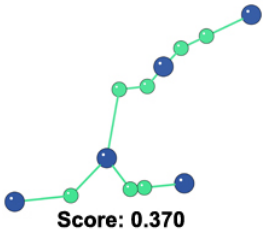
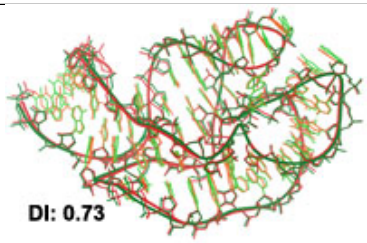
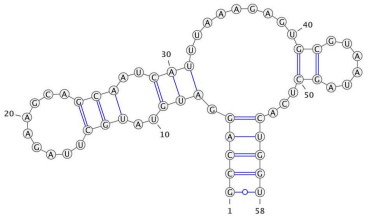
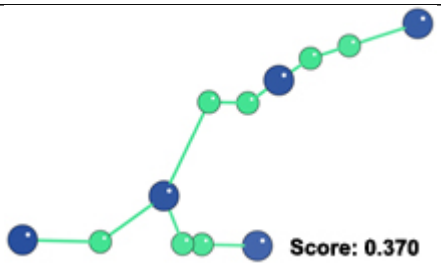
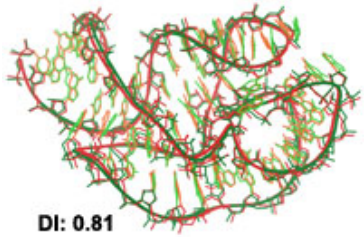
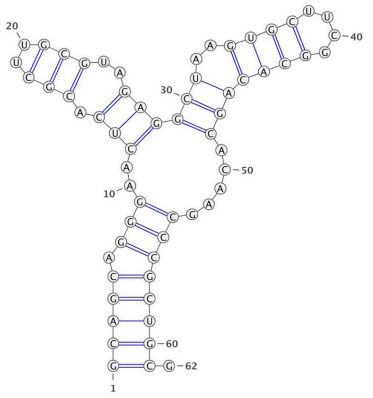
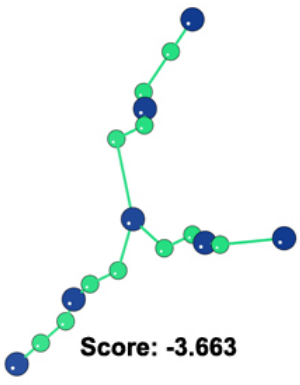
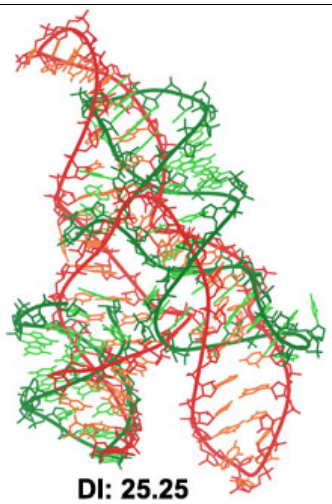
2M57		 Score: -2.810	 DI: 7.75
4PCJ		 Score: -1.727	 DI: 1.23
5KQE		 Score: -1.121	 DI: 10.51
2HW8		 Score: -1.799	 Di: 0.50

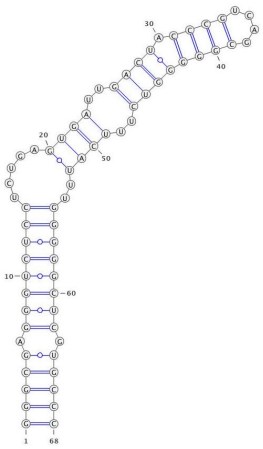
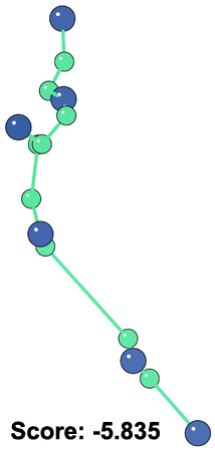
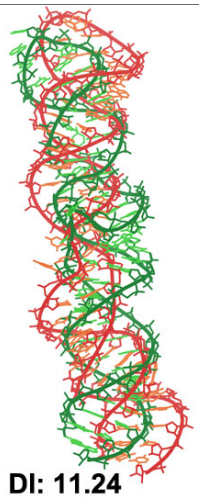
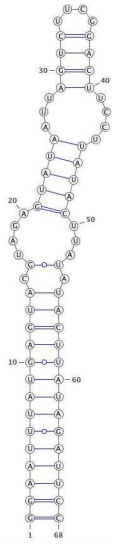
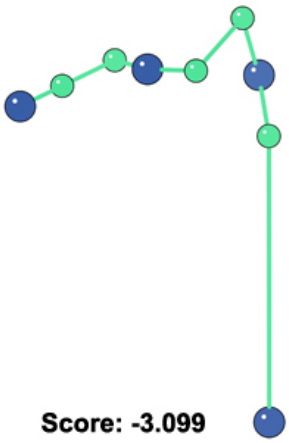
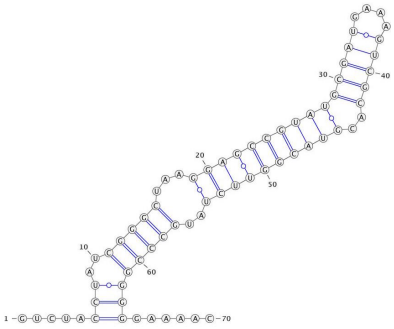
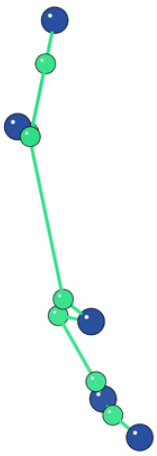
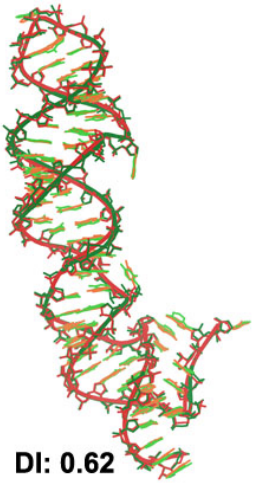
2N6S		 Score: 6.852	 DI: 2.75
1I6U		 Score: -2.867	 DI: 0.63
1F1T		 Score: -2.265	 DI: 7.30
1ZHO		 Score: -2.343	 DI: 1.21

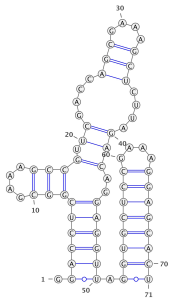
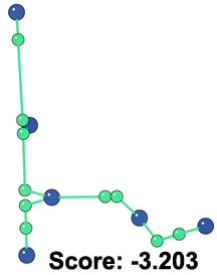
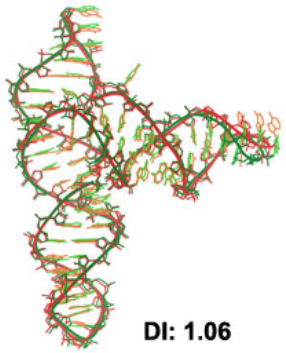
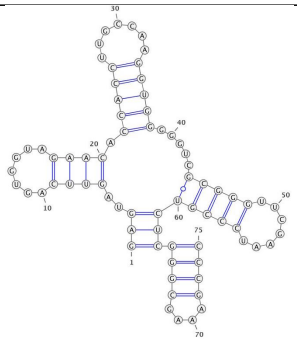
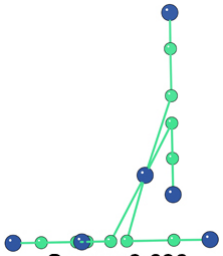
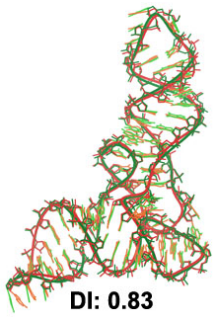
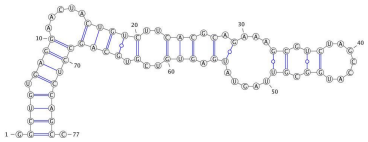
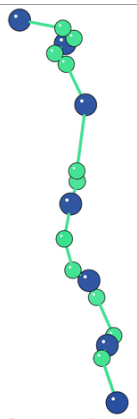
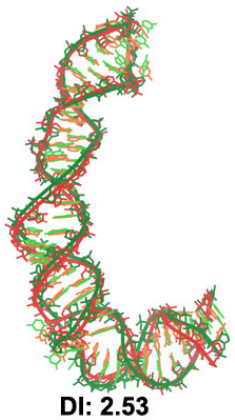
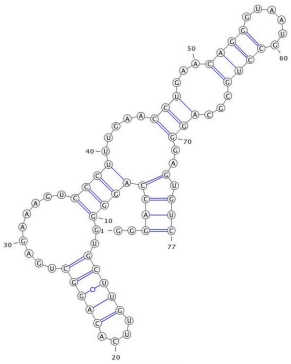
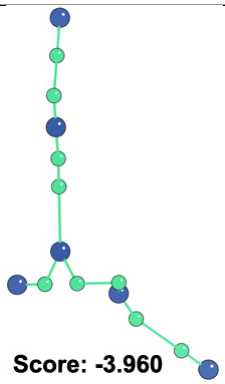
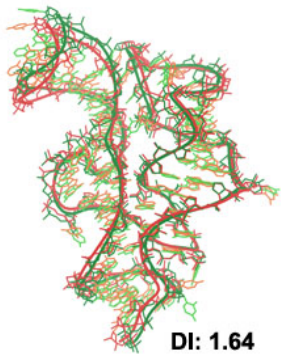
2MXL		 Score: -2.827	 DI: 12.37
2N6T		 Score: -3.132	 DI: 7.72
2N6X		 Score: -1.081	 DI: 16.20

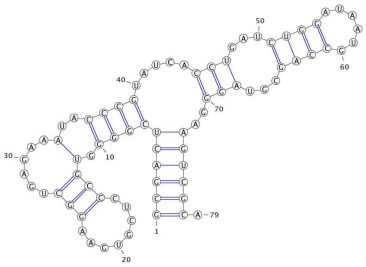
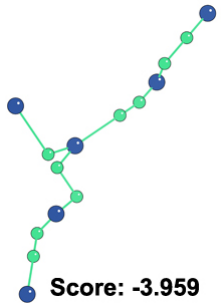
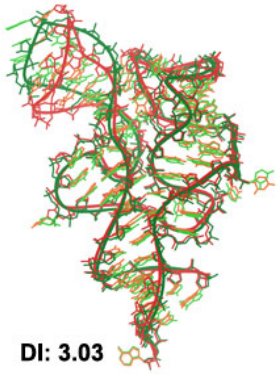
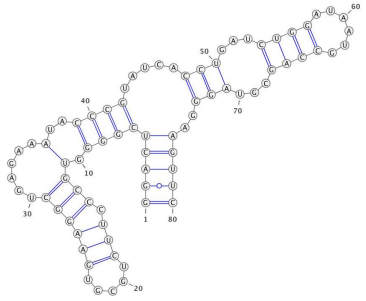
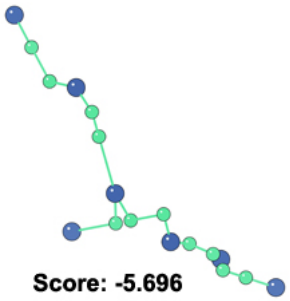
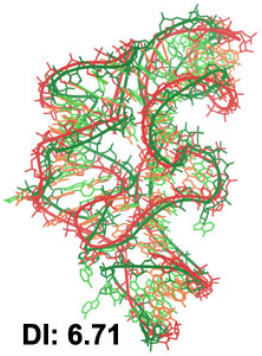
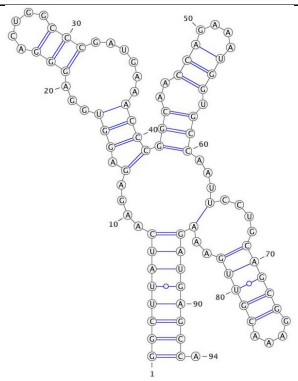
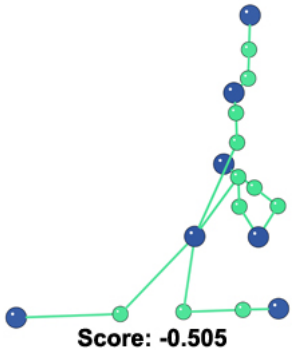
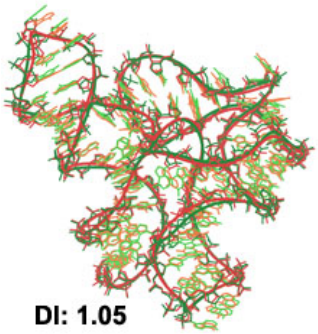
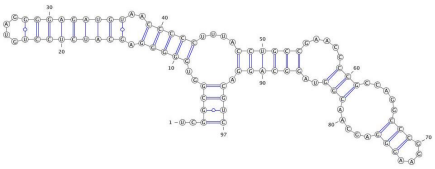
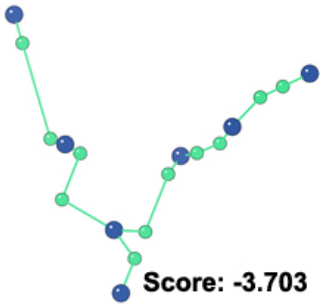
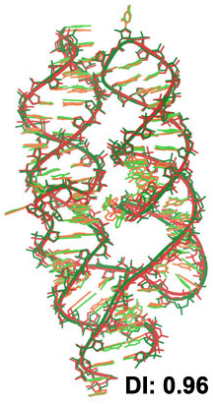
5BTM		 Score: -3.409	 DI: 3.37
1S03		 Score: -4.518	 DI: 6.38
1XJR		 Score: -4.989	 Di: 10.97
2MTJ		 Score: 0.864	 DI: 21.27

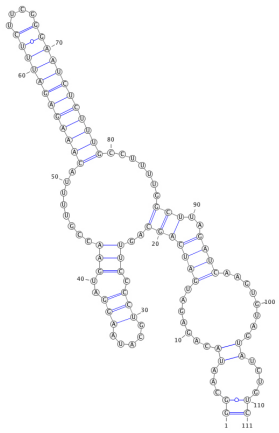
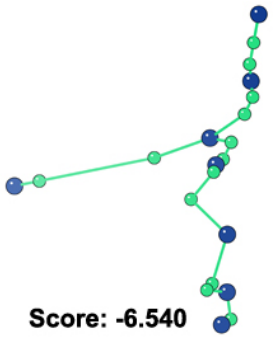
2VPL		 Score: -2.258	 DI: 0.48
1U63		 Score: -1.856	 DI:4.37
2PXB		 Score: -3.427	 DI: 6.10
2N4L		 Score: -3.197	 Di: 7.15

2HGH			
1DK1			
1MMS			
1Y39			
2N3Q			

2MQT			
2N6W			
1KXK			

2OIU		 Score: -3.203	 DI: 1.06
4LCK		 Score: 3.693	 DI: 0.83
1P5O		 Score: -8.516	 DI: 2.53
3D2G		 Score: -3.960	 DI: 1.64

2HOJ		 Score: -3.959	 DI: 3.03
2GDI		 Score: -5.696	 DI: 6.71
2GIS		 Score: -0.505	 DI: 1.05
1LNG		 Score: -3.703	 DI: 0.96

2LKR			

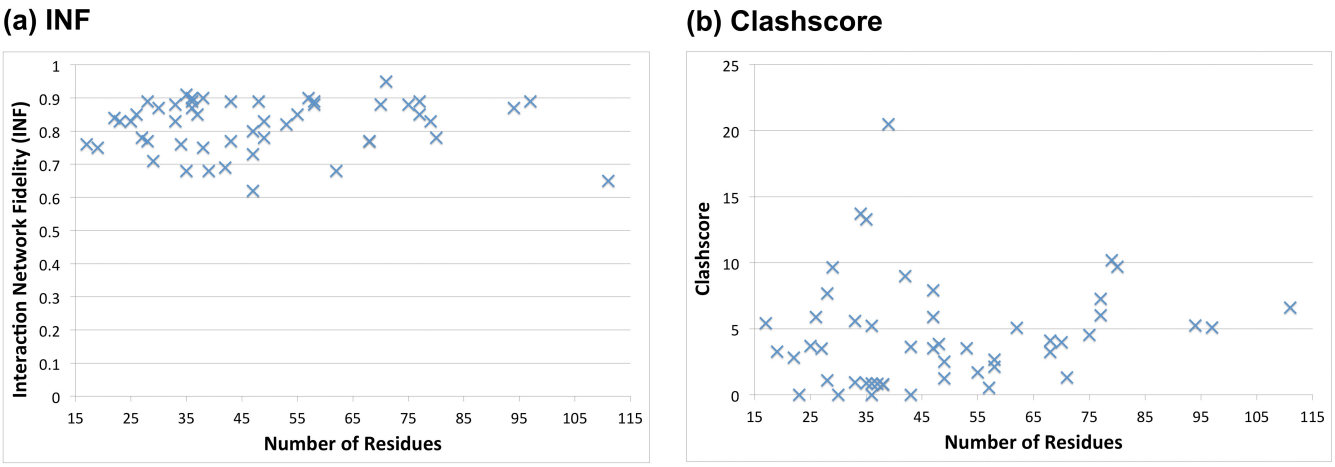


Figure S1 : **Statistics for the lowest DI atomic model generated by F-RAG for 50 RNA structures.** (a) Number of Residues vs Interaction Network Fidelity (INF) (b) Number of residues vs Clashscore. See the main text for definitions of these measures.

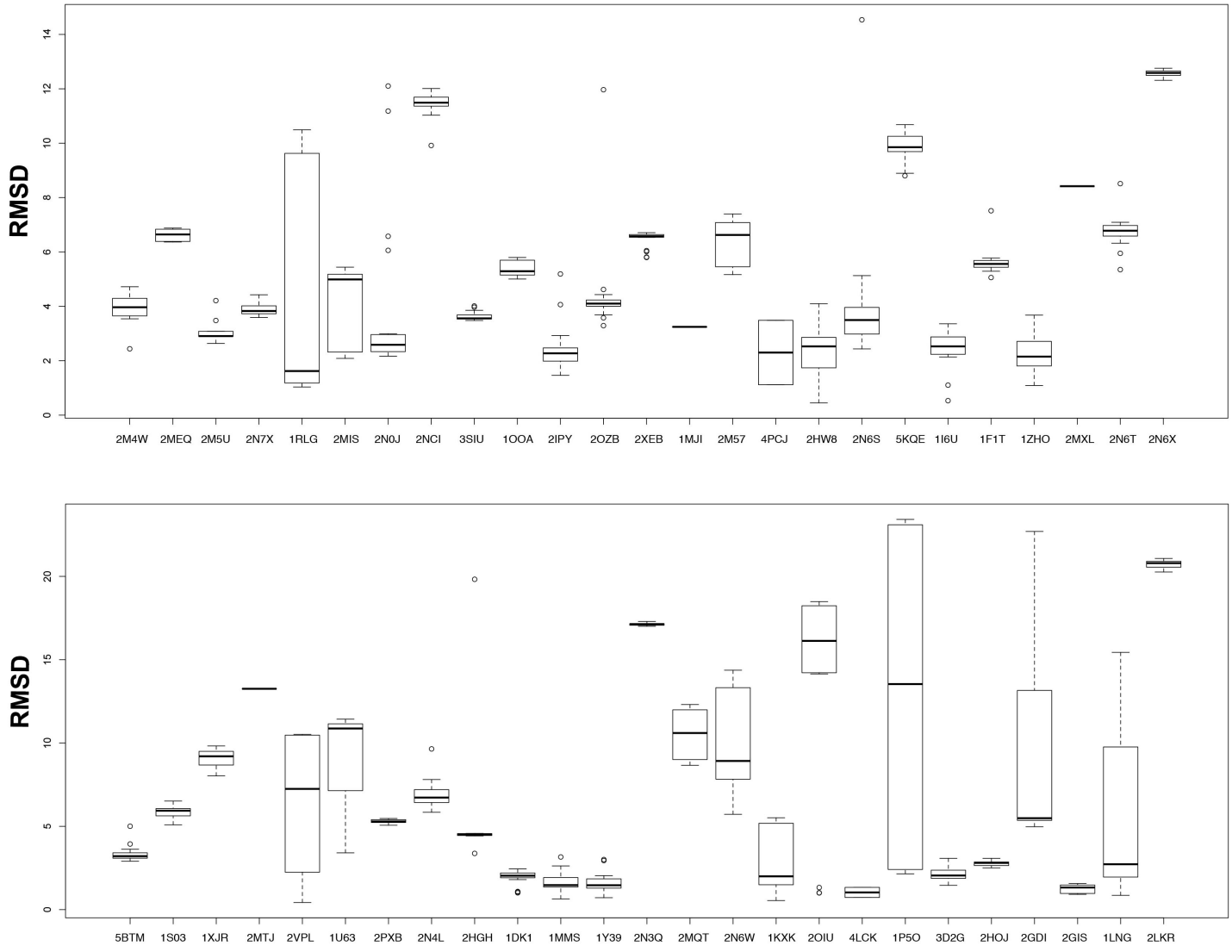


Figure S2 : RMSD of top atomic models generated by F-RAG for 50 RNA structures after geometry optimization with PHENIX.

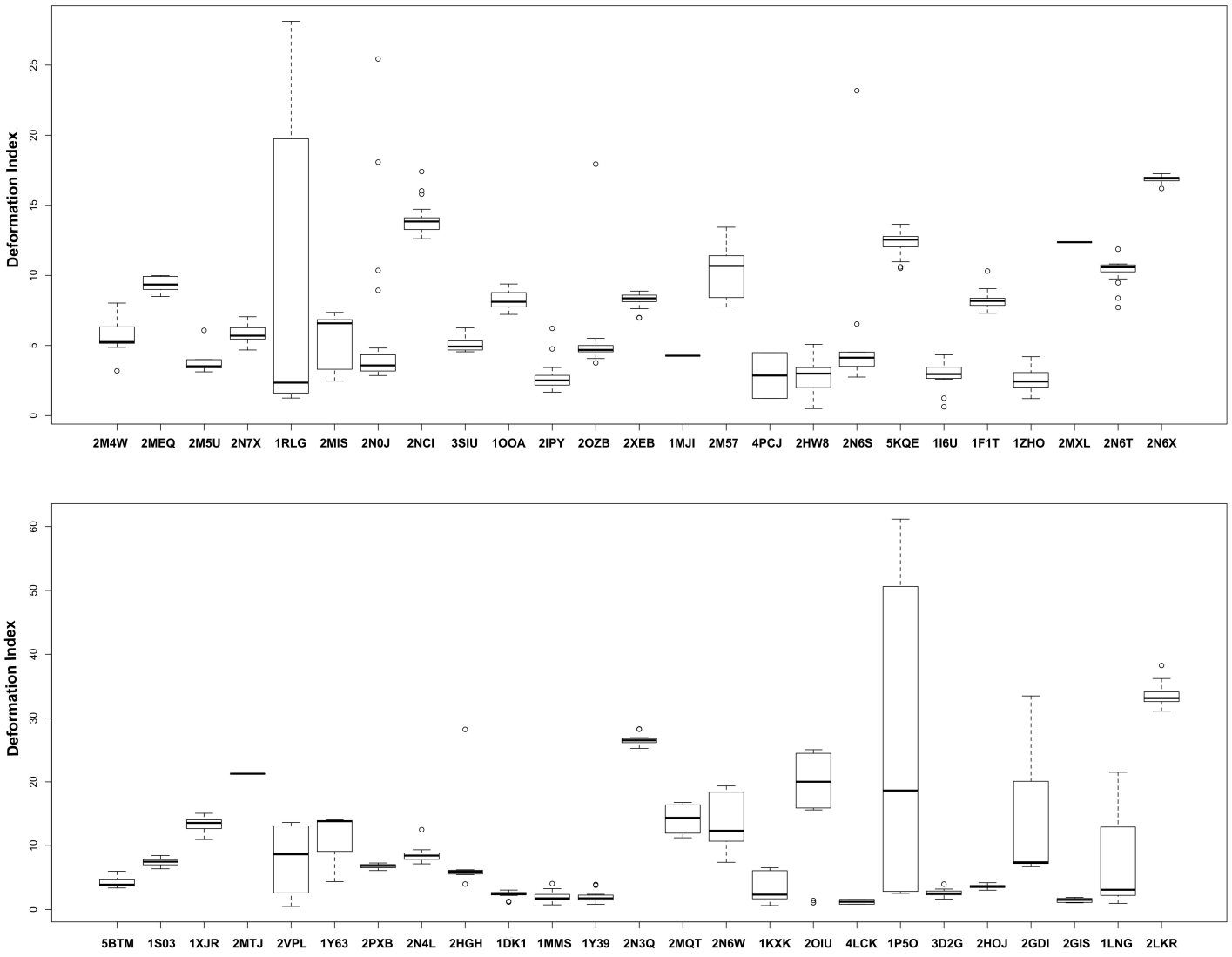


Figure S3 : Deformation Index (DI) values of top atomic models generated by F-RAG for 50 RNA structures after geometry optimization with PHENIX.

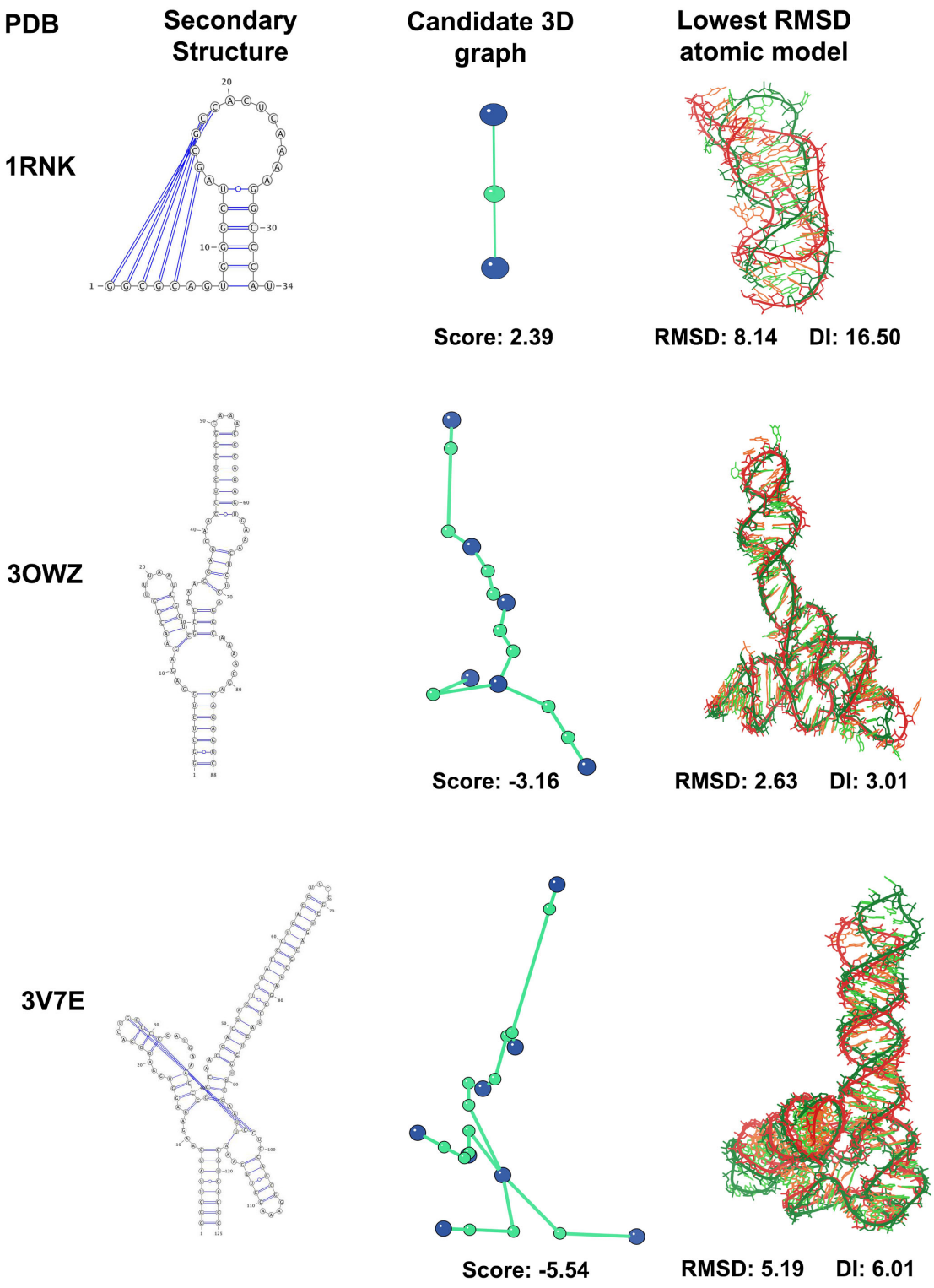


Figure S4 : The seconadry structure, the candidate 3D graph generated by RAGTOP, and the lowest RMSD model generated by F-RAG for three additional RNA structures: small hairpin pseudoknot (PDB ID: 1RNK), glycine riboswitch without a pseudoknot (PDB ID: 3OWZ), SAM-I riboswitch with a pseudoknot (PDB ID: 3V7E). PDB IDs 3OWZ and 3V7E are structures used in the RNA-Puzzles competition. The lowest RMSD atomic model for 1RNK has 8 missing residues from the dangling end. The reference structure is colored in red and the F-RAG generated atomic model is superimposed in green.