
1 PRE-PROCESSING STEP

The predicted interaction can be critical to the accuracy of the 3D modeling. Based on Table 2, the predicted 3D structures using the resolved interactions are generally more accurately than those using the predicted interactions. One of the major reasons is that the stacking interactions with different facing conformations were treated as one type of stacking when we trained the neural networks. As a result, the provided stacking facing conformation that might not always lead to the correct prediction. Therefore, we have introduced a pre-processing step that establishes a serial of

knowledge-based rules to modify the predicted interactions before using them in the 3D modeling step.

Fig. 1 show examples of four uncommon base stackings that cause uncommon shapes in the modeled 3D structures. The pre-processing step solves the problem by removing or replacing certain interactions and results in a more common backbone spin. More specifically, we introduced rules including removing contiguous s53; removing s55 or s33 within 3 intervals; removing contiguous interactions in cis or trans conformation and replacing contiguous s33 with s35. The proposed rules help to improve the result of 3D prediction generally (see Fig. 6 in the main text).

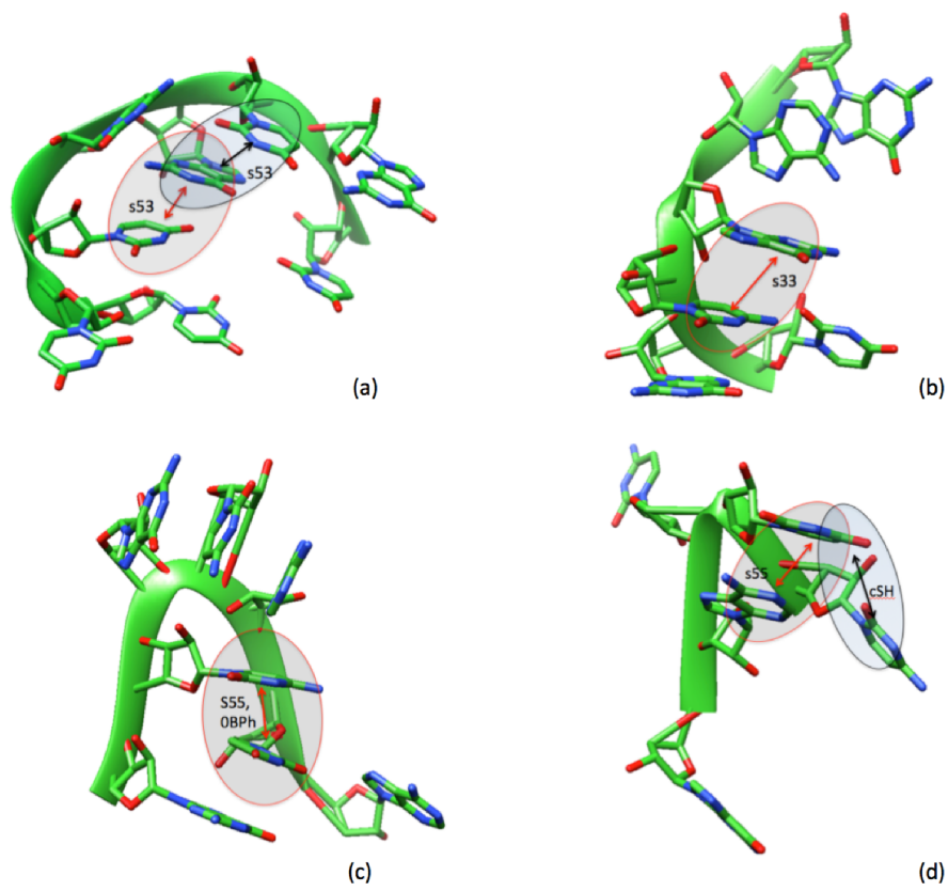


Fig. 1. Examples of uncommon shapes of RNA backbone spins. Stacking interactions other than s35 between neighboring nucleotides will force the backbone twist in a way that its torsion angles are far off their average values.

Table 1. RMSD and DI values of the prediction results by BkTree3D on the benchmark set used in the survey (Laing and Schlick, 2010). The number of canonical base pairs (CPBs) and number of non-canonical interactions (NCIs) are listed. The columns labeled STY, PPV, MCC and EdgeDiff were from our recent results (Ding *et al.*, 2015), where the predicted interactions were used as inputs to BkTree3D. The data of the 7 RNAs displayed with the bold font are not used in the training of the neural networks in the interaction prediction stage.

PDB ID	Length	# CBPs	# NCIs	STY	PPV	MCC	RMSD	DI	EdgeDiff	Structure complexity
2F8K	16	6	14	85	80	0.828	2.911	3.515	12/12	Hairpin
2AB4	20	6	20	100	90	0.953	2.753	2.887	12/12	Hairpin
361D	20	5	17	100	73	0.859	2.525	0.936	10/10	Hairpin
2ANN	23	3	24	100	96	0.979	4.351	4.440	12/12	Hairpin
1RLG	25	5	22	95	60	0.756	2.630	3.475	15/15	Hairpin, internal loop
2QUX	25	9	22	90	86	0.889	3.162	3.556	15/15	Hairpin
387D	26	4	23	95	78	0.866	6.069	7.000	10/10	Hairpin
1MSY	27	6	39	97	97	0.974	1.304	1.338	23/23	Hairpin
1L2X	28	8	34	97	94	0.956	3.534	3.694	23/22	Pseudoknot
2AP5	28	8	29	79	65	0.721	6.231	8.631	23/21	Pseudoknot
1JID	29	8	31	87	69	0.776	2.346	3.021	20/20	Hairpin, internal loop
1OOA	29	8	29	68	60	0.646	4.165	6.442	15/15	Hairpin, internal loop
430D	29	6	37	89	80	0.847	4.335	5.116	21/21	Hairpin, internal loop
3SNP	30	12	31	93	82	0.880	1.198	1.360	22/22	Hairpin, internal loop
2OZB	33	10	33	93	83	0.887	3.497	3.941	26/26	Hairpin, internal loop
1MJI	34	10	44	75	86	0.807	1.758	2.178	29/28	Hairpin, internal loop
1ET4	35	8	40	77	86	0.816	4.565	5.588	28/25	Pseudoknot
2HW8	36	12	44	90	85	0.879	6.673	7.586	30/30	Hairpin, internal loop
1I6U	37	15	47	91	89	0.905	3.095	3.418	28/28	Hairpin, internal loop
1F1T	38	10	38	78	66	0.725	4.258	5.869	29/29	Hairpin, internal loop
1ZHO	38	13	46	91	85	0.884	3.264	3.689	31/31	Hairpin, internal loop
1S03	47	18	53	88	82	0.855	3.485	4.075	37/37	Hairpin, internal loop
1XJR	47	15	55	89	94	0.916	2.494	2.721	35/35	Hairpin, internal loop
1U63	49	17	50	90	70	0.795	2.087	2.623	36/36	Hairpin, internal loop
2PXB	49	16	66	100	98	0.992	1.140	1.148	38/38	Hairpin, internal loop
3E5C	53	21	65	90	80	0.856	2.042	2.384	40/39	3-way junction (r: riboswitch)
1MZP	55	17	73	78	86	0.821	11.23	13.675	44/38	Hairpin internal
1DK1	57	24	65	95	88	0.919	1.029	1.119	46/45	3-way junction
1MMS	58	20	86	77	89	0.834	2.032	2.435	57/45	3-way junction
3EGZ	65	23	72	77	74	0.762	6.433	8.441	39/35	3-way junction (r)
2QUS	69	26	81	70	76	0.731	9.864	13.488	48/42	Pseudoknot, 3-way junction
1KXK	70	28	87	90	90	0.908	3.230	3.555	54/54	Hairpin, internal loop
2DU3	71	27	75	82	77	0.800	3.751	4.686	50/43	4-way junction (tRNA)
2OIU	71	29	84	91	89	0.905	5.629	6.213	51/51	3-way junction (r)
1SJ4	73	19	83	73	82	0.778	13.99	17.973	53/38	Pseudoknot, 4-way junction
1P5O	77	29	86	95	79	0.871	6.826	7.834	51/51	Hairpin, internal loop
3D2G	77	28	103	79	87	0.833	4.552	5.462	70/60	3-way junction (r)
2HOJ	79	27	100	85	87	0.863	4.024	4.662	64/58	3-way junction (r)
2GDI	80	32	100	82	81	0.815	5.176	6.343	63/56	3-way junction (r)
2GIS	94	36	125	84	80	0.828	14.87	17.951	79/67	Pseudoknot, 4-way junction (r)
1LNG	97	38	124	86	83	0.849	10.86	12.786	76/70	3-way junction (SRP)
1MFQ	128	49	164	79	80	0.800	24.58	30.723	101/92	3-way junction (SRP)

Table 2. RMSD values from the results of three independent tests using BkTree3D on the benchmark set used in the survey (Laing and Schlick, 2010). The resolved interactions from RNA 3D Atlas were used as inputs to BkTree3D to produce the results in RMSD_r column. For RMSD_p column, the predicted interactions were preprocessed (see Section X) before inputting to BkTree3D. In addition to the preprocessing of the predicted interactions, the geometry candidates of a testing RNA were removed from GDB before GDB was used by BkTree3D for the results in RMSD_g column.

PDB ID	Length	RMSD_r	RMSD_p	RMSD_g	Structure complexity
2F8K	16	2.02	2.674	2.694	Hairpin
2AB4	20	3.01	2.787	2.571	Hairpin
361D	20	3.99	5.809	5.825	Hairpin
2ANN	23	3.04	1.642	2.649	Hairpin
1RLG	25	4.61	2.169	4.021	Hairpin, internal loop
2QUX	25	1.05	2.156	2.156	Hairpin
387D	26	0.06	6.002	6.002	Hairpin
1MSY	27	1.00	2.49	2.337	Hairpin
1L2X	28	1.45	2.433	3.750	Pseudoknot
2AP5	28	2.85	1.997	2.711	Pseudoknot
1JID	29	1.77	3.249	4.180	Hairpin, internal loop
1OOA	29	0.159	2.262	2.262	Hairpin, internal loop
430D	29	2.26	4.165	4.282	Hairpin, internal loop
3SNP	30	2.132	1.602	2.263	Hairpin, internal loop
2OZB	33	1.365	2.143	2.347	Hairpin, internal loop
1MJI	34	1.105	4.173	4.173	Hairpin, internal loop
1ET4	35	3.816	5.214	5.214	Pseudoknot
2HW8	36	1.574	4.255	4.542	Hairpin, internal loop
1I6U	37	1.569	3.051	3.051	Hairpin, internal loop
1F1T	38	2.501	8.015	8.174	Hairpin, internal loop
1ZHO	38	2.07	4.094	4.328	Hairpin, internal loop
1S03	47	1.512	2.707	2.907	Hairpin, internal loop
1XJR	47	1.355	2.88	2.88	Hairpin, internal loop
1U63	49	1.421	4.991	5.293	Hairpin, internal loop
2PXB	49	0.884	2.506	2.506	Hairpin, internal loop
3E5C	53	7.598	1.716	1.716	3-way junction (r: riboswitch)
1MZP	55	4.862	5.648	5.833	Hairpin internal
1DK1	57	4.944	1.686	1.686	3-way junction
1MMS	58	4.768	2.864	2.973	3-way junction
3EGZ	65	4.893	10.33	10.33	3-way junction (r)
2QUS	69	3.001	11.01	11.13	Pseudoknot, 3-way junction
1KXK	70	2.666	3.261	3.261	Hairpin, internal loop
2DU3	71	0.558	2.993	2.993	4-way junction (tRNA)
2OIU	71	1.21	5.9	5.9	3-way junction (r)
1SJ4	73	7.333	11.18	11.18	Pseudoknot, 4-way junction
1P5O	77	11.47	5.166	5.166	Hairpin, internal loop
3D2G	77	2.461	7.162	7.162	3-way junction (r)
2HOJ	79	8.283	2.564	3.583	3-way junction (r)
2GDI	80	2.227	3.22	3.22	3-way junction (r)
2GIS	94	5.713	14.13	14.34	Pseudoknot, 4-way junction (r)
1LNG	97	19.91	7.959	10	3-way junction (SRP)
1MFQ	128	20.79	21.33	20.79	3-way junction (SRP)

Table 3. RMSD values of 5 top structures predicted by BkTree3D on the benchmark set used in the survey (Laing and Schlick, 2010). The predicted interactions were preprocessed (see Section X) before inputting to BkTree3D. The best and average results were also listed for each RNA.

PDB ID	Length	RMSD1	RMSD2	RMSD3	RMSD4	RMSD5	Best	Average	Structure complexity
2F8K	16	2.674	2.644	2.644	3.544	3.544	2.644	3.004	Hairpin
2AB4	20	5.809	4.950	4.975	4.939	5.005	4.939	5.139	Hairpin
361D	20	2.787	2.472	2.466	2.543	2.431	2.431	2.497	Hairpin
2ANN	23	1.642	2.474	2.526	2.367	2.224	2.224	2.448	Hairpin
1RLG	25	2.156	3.938	3.984	3.874	4.205	2.156	3.632	Hairpin, internal loop
2QUX	25	2.169	4.245	3.854	3.996	4.514	3.854	4.126	Hairpin
387D	26	6.002	6.442	6.762	8.708	7.986	6.002	7.180	Hairpin
1MSY	27	2.49	2.133	2.064	2.100	2.087	2.064	2.144	Hairpin
1L2X	28	2.433	3.956	4.817	4.842	4.614	3.750	4.396	Pseudoknot
2AP5	28	1.997	2.711	2.711	2.711	2.711	2.711	2.711	Pseudoknot
1JID	29	3.249	2.029	1.828	2.044	2.039	1.828	2.024	Hairpin, internal loop
1OOA	29	2.262	4.244	4.258	4.256	4.055	4.055	4.219	Hairpin, internal loop
430D	29	1.602	3.277	3.098	5.184	3.285	2.263	3.422	Hairpin, internal loop
3SNP	30	4.165	1.789	2.739	2.739	2.750	1.602	2.324	Hairpin, internal loop
2OZB	33	2.143	3.889	4.013	3.174	3.987	3.037	3.620	Hairpin, internal loop
1MJI	34	4.173	4.196	4.243	3.939	4.060	3.939	4.193	Hairpin, internal loop
1ET4	35	5.214	5.764	5.138	5.264	5.207	5.138	5.362	Pseudoknot
2HW8	36	4.255	3.742	3.667	3.836	3.635	3.579	3.692	Hairpin, internal loop
1I6U	37	3.051	3.432	3.414	3.486	3.273	3.273	3.448	Hairpin, internal loop
1F1T	38	4.094	3.912	3.622	4.039	3.934	3.622	3.912	Hairpin, internal loop
1ZHO	38	8.015	4.925	5.903	5.944	4.429	4.429	5.257	Hairpin, internal loop
1S03	47	2.88	2.670	2.562	2.542	2.401	2.401	2.545	Hairpin, internal loop
1XJR	47	2.707	3.113	3.035	2.626	3.197	2.267	2.848	Hairpin, internal loop
1U63	49	2.506	3.186	3.554	5.375	2.900	2.571	3.517	Hairpin, internal loop
2PXB	49	4.991	4.979	5.181	6.136	6.059	4.979	5.468	Hairpin, internal loop
3E5C	53	1.716	2.041	3.253	2.021	3.024	1.785	2.425	3-way junction (r: riboswitch)
1MZP	55	5.648	7.685	7.976	8.633	9.966	5.731	7.998	Hairpin internal
1DK1	57	1.686	1.994	1.982	2.228	3.092	1.685	2.196	3-way junction
1MMS	58	2.864	3.701	4.439	4.240	4.025	3.483	3.977	3-way junction
3EGZ	65	10.33	11.95	12.46	11.32	11.20	10.33	11.45	3-way junction (r)
2QUS	69	11.01	9.857	15.39	17.66	21.89	15.39	14.83	Pseudoknot, 3-way junction
1KXK	70	3.261	5.089	6.140	5.032	6.574	4.114	5.390	Hairpin, internal loop
2DU3	71	2.993	3.303	2.703	3.187	3.494	2.703	3.136	4-way junction (tRNA)
2OIU	71	5.9	6.832	10.24	6.127	8.771	10.24	7.582	3-way junction (r)
1SJ4	73	11.18	11.44	10.08	10.23	11.58	10.08	11.02	Pseudoknot, 4-way junction
1P5O	77	5.166	5.477	6.657	5.978	5.694	5.466	5.854	Hairpin, internal loop
3D2G	77	7.162	7.939	7.523	10.32	9.456	10.32	8.588	3-way junction (r)
2HOJ	79	2.564	2.750	3.709	3.162	3.461	2.566	3.130	3-way junction (r)
2GDI	80	3.22	3.149	4.142	3.189	4.025	3.112	3.523	3-way junction (r)
2GIS	94	14.13	15.32	17.45	15.80	16.51	14.13	15.84	Pseudoknot, 4-way junction (r)
1LNG	97	7.959	7.804	8.330	7.735	8.801	7.735	8.177	3-way junction (SRP)
1MFQ	128	18.3	24.42	24.92	23.37	24.19	22.37	23.85	3-way junction (SRP)