RNA fold prediction by Monte Carlo in graph space and the statistical mechanics of tertiary interactions

ETHAN N. H. PHAN[1] AND CHI H. MAK[2]

[1]Department of Chemistry, University of Southern California, Los Angeles, California 90089, USA

[2]Departments of Chemistry and Quantitative and Computational Biology, and Center of Applied Mathematical Sciences, University of Southern California, Los Angeles, California 90089, USA

Corresponding author: ethanpha@usc.edu, cmak@usc.edu

TEST SETS

Rzs secondary-only test set

PDB	length	# native	Vienna		ASMC			Vienna		ASMC	
	(nt)	contacts	predicted	match	predicted	match	μ	TPR	mFPR	TPR	mFPR
1KXK	70	23	24	19	24	19	0	0.83	0.21	0.83	0.21
1U9S	161	51	50	31	48	27	0.7	0.61	0.38	0.53	0.44
2KXM	27	10	8	8	5	5	0	0.8	0	0.5	0
2LU0	49	17	16	15	14	13	0.2	0.88	0.06	0.76	0.07
2MI0	22	7	7	6	7	6	0	0.86	0.14	0.86	0.14
2MIS	26	10	9	9	9	9	0	0.9	0	0.9	0
2MTJ	47	17	15	14	17	14	0	0.82	0.07	0.82	0.18
2N3Q	62	22	19	8	21	20	0	0.36	0.58	0.91	0.05
2N3R	62	22	19	8	21	20	0	0.36	0.58	0.91	0.05
2OEU	66	19	19	16	18	14	0.3	0.84	0.16	0.74	0.22
2OIU	71	26	25	21	26	14	0	0.81	0.16	0.54	0.46
2QUS	69	24	21	20	21	20	0	0.83	0.05	0.83	0.05
2R8S	159	48	53	36	53	33	1	0.75	0.32	0.69	0.38
359D	44	13	13	12	13	12	0.1	0.92	0.08	0.92	0.08
3BBM	67	19	23	12	20	15	0	0.63	0.48	0.79	0.25
3D2V	77	23	18	12	18	12	1	0.52	0.33	0.52	0.33
3E5C	53	17	15	10	15	10	0.2	0.59	0.33	0.59	0.33
3F2X	112	33	32	23	34	12	0.5	0.7	0.28	0.36	0.65
3GS5	64	21	20	13	17	13	1	0.62	0.35	0.62	0.24
3OXE	88	29	26	24	27	21	0	0.83	0.08	0.72	0.22
3PDR	161	107	52	38	48	40	0.6	0.36	0.27	0.37	0.17
4GXY	172	53	57	35	59	38	0.6	0.66	0.39	0.72	0.36
4R4V	186	66	63	57	57	42	1	0.86	0.1	0.64	0.26
4RUM	92	32	30	15	25	20	0.7	0.47	0.5	0.62	0.2
4Y1M	107	22	37	17	31	15	0.5	0.77	0.54	0.68	0.52
4YAZ	84	26	23	15	21	17	0.6	0.58	0.35	0.65	0.19
5DH6	68	17	22	15	22	15	0.2	0.88	0.32	0.88	0.32
5LYS	57	22	22	13	21	21	0	0.59	0.41	0.95	0
5NDH	16	5	6	5	6	5	0	1	0.17	1	0.17
5T83	89	26	27	0	23	14	1	0	1	0.54	0.39
5U3G	85	26	25	20	26	22	0.6	0.77	0.2	0.85	0.15
5U6Z	68	17	22	15	22	15	0	0.88	0.32	0.88	0.32
5UZ6	32	14	12	12	12	12	0	0.86	0	0.86	0
6AZ4	42	15	15	15	13	11	0	1	0	0.73	0.15
6C27	47	18	15	12	16	15	0	0.67	0.2	0.83	0.06
6CB3	101	33	34	32	32	30	0.2	0.97	0.06	0.91	0.06
6CK5	117	38	37	30	37	17	0	0.79	0.19	0.45	0.54
6EZ0	27	7	11	7	11	7	0	1	0.36	1	0.36
6HC5	18	6	7	6	7	6	0	1	0.14	1	0.14
6JQ5	163	52	56	49	54	30	0.1	0.94	0.12	0.58	0.44
6N2V	99	37	34	31	34	32	0.5	0.84	0.09	0.86	0.06
7EAG	41	11	12	6	12	4	0	0.55	0.5	0.36	0.67
7ELQ	45	15	14	12	14	11	0.1	0.8	0.14	0.73	0.21
7MLW	128	44	44	30	39	23	1	0.68	0.32	0.52	0.41
7Q80	68	21	19	17	19	17	0.1	0.81	0.11	0.81	0.11
7TZS	80	23	20	10	20	7	0.1	0.43	0.5	0.3	0.65
							Mean:	0.731	0.260	0.719	0.245

Table S-01.

Targets in the Rzs test set with only secondary interactions in their native structures. PDB code, sequence length and the number of base pair contacts in the PDB structure are shown in the first three columns. The number of contacts

predicted by ViennaRNA and how many of those match the native structure are given in columns 4 and 5. Columns 6 and 7 show the same for ASMC, and the lowest value of μ at which the ensemble could be equilibrated in ASMC is given in the column labeled μ. The number of predicted contacts that match native as a fraction of all native contacts are given for Vienna and ASMC in the columns labeled TPR. The number of overpredicted contacts by Vienna and ASMC as a fraction of all predicted contacts are given for Vienna and ASMC in the columns labeled mFPR. The sample means of TPR and mFPR are given on the last row.

Rzs kissing loop test set

PDB	length	# native	Vienna		ASMC			Vienna		ASMC	
	(nt)	contacts	predicted	match	predicted	match	μ	TPR	mFPR	TPR	mFPR
1Y26	71	26	21	21	25	24	0	0.81	0	0.92	0.04
3D0U	161	61	55	47	52	48	0.5	0.77	0.15	0.79	0.08
3DIL	174	67	63	56	59	54	0.4	0.84	0.11	0.81	0.08
3IVN	70	26	25	21	23	21	0.2	0.81	0.16	0.81	0.09
3LA5	71	25	21	20	21	20	0.3	0.8	0.05	0.8	0.05
3RKF	67	25	20	20	22	20	0	0.8	0	0.8	0.09
3SKI	68	25	22	21	24	20	0	0.84	0.05	0.8	0.17
4FEN	67	23	24	21	21	18	0.1	0.91	0.12	0.78	0.14
4FRG	84	26	27	7	23	18	0.4	0.27	0.74	0.69	0.22
4MGN	164	48	53	32	46	26	0.6	0.67	0.4	0.54	0.43
4XNR	71	26	21	21	25	24	0	0.81	0	0.92	0.04
5C45	113	34	32	7	31	4	0.1	0.21	0.78	0.12	0.87
5FJC	95	31	27	21	26	21	1	0.68	0.22	0.68	0.19
5NDI	43	16	17	2	16	14	0	0.12	0.88	0.88	0.12
5SWD	71	24	21	14	25	18	0	0.58	0.33	0.75	0.28
6DN2	112	33	32	16	34	7	0.5	0.48	0.5	0.21	0.79
6E1S	33	7	5	5	6	6	0.3	0.71	0	0.86	0
6VMY	148	47	43	28	41	27	2	0.6	0.35	0.57	0.34
							Mean:	0.650	0.269	0.707	0.223

Table S-02.

Targets in the Rzs test set with kissing loop interactions in their native structures. See Table S-01 for definitions of each data column.

Rzs pseudoknot test set

PDB	length	# native	Vienna		ASMC			Vienna		ASMC	
	(nt)	contacts	predicted	match	predicted	match	μ	TPR	mFPR	TPR	mFPR
2MIY	59	19	19	18	18	18	0.2	0.95	0.05	0.95	0
2QWY	52	16	10	4	9	8	0	0.25	0.6	0.5	0.11
2Z75	143	48	48	28	52	44	0.1	0.58	0.42	0.92	0.15
3K1V	34	9	5	5	9	9	0	0.56	0	1	0
3NPQ	54	14	19	10	15	10	0	0.71	0.47	0.71	0.33
3Q3Z	77	28	28	0	22	12	0.1	0	1	0.43	0.45
4ENC	52	15	16	15	16	15	0	1	0.06	1	0.06
4FRN	102	33	30	14	22	18	1	0.42	0.53	0.55	0.18
4JF2	77	25	29	24	29	24	0	0.96	0.17	0.96	0.17
4KQY	119	40	40	35	41	38	0.1	0.88	0.12	0.95	0.07
4LVW	89	33	34	25	26	20	0.6	0.76	0.26	0.61	0.23
4OJI	54	19	14	13	12	12	0.1	0.68	0.07	0.63	0
4OQU	97	33	33	32	32	22	0	0.97	0.03	0.67	0.31
4QJD	71	30	18	11	20	20	0.1	0.37	0.39	0.67	0

4QK9	124	39	37	5	40	22	0	0.13	0.86	0.56	0.45
4QLM	110	37	42	26	35	28	0.2	0.7	0.38	0.76	0.2
4RGE	59	20	19	13	13	12	0	0.65	0.32	0.6	0.08
5BTP	75	21	24	20	20	19	0	0.95	0.17	0.9	0.05
5D5L	77	25	29	24	29	24	0	0.96	0.17	0.96	0.17
5KH8	47	14	14	13	14	13	0	0.93	0.07	0.93	0.07
5NWQ	41	11	15	6	11	10	0	0.55	0.6	0.91	0.09
6FZ0	49	12	10	2	9	4	0	0.17	0.8	0.33	0.56
6HAG	43	11	12	8	10	8	0	0.73	0.33	0.73	0.2
6N5P	127	39	42	29	43	16	0.4	0.74	0.31	0.41	0.63
6QN3	100	37	29	10	37	32	0	0.27	0.66	0.86	0.14
6XKO	96	31	30	22	30	23	1	0.71	0.27	0.74	0.23
6YL5	35	6	13	6	9	5	0.3	1	0.54	0.83	0.44
						Mean:		0.651	0.357	0.743	0.199

Table S-03.

Targets in the Rzs test set with H-type pseudoknot interactions in their native structures. See Table S-01 for definitions of each data column.

Rfam test set

PDB	Rfam ID	Length (nt)	# native contacts	Vienna		ASMC		μ	Vienna		ASMC	
				pred	match	pred	match		TPR	mFPR	TPR	mFPR
1KXK	RF00029	70	23	24	19	24	19	0	0.83	0.21	0.83	0.21
1M5K	RF04190	113	41	40	30	37	30	0	0.73	0.25	0.73	0.19
1N8X	RF00175	36	15	14	14	14	14	0	0.93	0	0.93	0
1NBS	RF00011	150	43	49	33	48	19	0.4	0.77	0.33	0.44	0.6
1P6V	RF00023	24	7	7	7	7	7	0	1	0	1	0
1U9S	RF00010	161	51	48	39	48	27	0.7	0.76	0.19	0.53	0.44
1XJR	RF00164	47	17	15	14	13	9	0.2	0.82	0.07	0.53	0.31
1Z2J	RF00480	45	19	19	19	19	19	0	1	0	1	0
2KE6	RF00207	48	19	19	18	19	17	0	0.95	0.05	0.89	0.11
2L1F	RF00374	131	52	48	25	49	33	0	0.48	0.48	0.63	0.33
2L3J	RF02095	71	32	30	30	27	27	0	0.94	0	0.84	0
2LC8	RF01073	56	19	18	18	11	7	1	0.95	0	0.37	0.36
2MF0	RF00166	72	18	20	18	20	18	0.5	1	0.1	1	0.1
2MIY	RF01054	59	19	19	18	18	18	0.2	0.95	0.05	0.95	0
2N1Q	RF01381	155	60	51	40	50	37	0.5	0.67	0.22	0.62	0.26
2NBX	RF00210	108	44	40	36	41	36	0	0.82	0.1	0.82	0.12
2QUS	RF00008	69	24	20	20	21	20	0	0.83	0	0.83	0.05
2V3C	RF00169	96	34	31	23	33	26	0.7	0.68	0.26	0.76	0.21
2Z75	RF00234	143	48	47	37	52	44	0.1	0.77	0.21	0.92	0.15
3D2V	RF00059	77	23	18	12	18	12	1	0.52	0.33	0.52	0.33
3DIL	RF00168	174	67	62	56	59	54	0.4	0.84	0.1	0.81	0.08
3F2X	RF00050	112	33	30	22	34	12	0.5	0.67	0.27	0.36	0.65
3NDB	RF01857	136	50	48	46	46	35	0.2	0.92	0.04	0.7	0.24
3OXE	RF00504	88	29	25	24	27	21	0	0.83	0.04	0.72	0.22
3PDR	RF00380	161	107	52	40	48	40	0.6	0.37	0.23	0.37	0.17
3Q3Z	RF01786	77	28	27	12	22	12	0.1	0.43	0.56	0.43	0.45
3SN2	RF02253	29	12	11	11	11	11	0	0.92	0	0.92	0
3SNP	RF00037	29	12	11	11	11	11	0	0.92	0	0.92	0
4FEN	RF00167	67	23	24	21	21	18	0.1	0.91	0.12	0.78	0.14
4FRG	RF01689	84	26	27	7	23	18	0.4	0.27	0.74	0.69	0.22
4LVW	RF01831	89	33	34	25	26	20	0.6	0.76	0.26	0.61	0.23
4OQU	RF01725	97	33	33	32	32	22	0	0.97	0.03	0.67	0.31
4PQV	RF01415	68	21	22	9	19	9	0.7	0.43	0.59	0.43	0.53

4QLM	RF00379	110	37	34	32	35	28	0.2	0.86	0.06	0.76	0.2
4RUM	RF02683	92	32	29	28	25	20	0.7	0.88	0.03	0.62	0.2
4V2S	RF00505	57	18	17	17	17	17	0	0.94	0	0.94	0
4WFL	RF01854	107	37	38	19	37	27	0	0.51	0.5	0.73	0.27
4YAZ	RF01051	84	26	19	15	21	17	0.6	0.58	0.21	0.65	0.19
5BTP	RF01750	75	21	20	20	20	19	0	0.95	0	0.9	0.05
5FJC	RF00162	95	31	26	21	26	21	1	0.68	0.19	0.68	0.19
5KH8	RF01734	47	14	14	13	14	13	0	0.93	0.07	0.93	0.07
5LYS	RF00100	57	22	21	21	21	21	0	0.95	0	0.95	0
5NWQ	RF01763	41	11	10	0	11	10	0	0	1	0.91	0.09
5T5A	RF02681	62	24	17	0	20	12	0.3	0	1	0.5	0.4
5T83	RF00442	89	26	27	0	23	14	1	0	1	0.54	0.39
6B19	RF00375	38	19	15	15	15	15	0	0.79	0	0.79	0
6CC1	RF00080	93	66	34	28	31	27	0.1	0.42	0.18	0.41	0.13
6CU1	RF02553	80	28	27	22	29	27	0	0.79	0.19	0.96	0.07
6FZ0	RF01826	49	12	13	0	9	4	0	0	1	0.33	0.56
6HAG	RF01727	43	11	12	8	10	8	0	0.73	0.33	0.73	0.2
6JQ5	RF02678	163	52	56	49	54	30	0.1	0.94	0.12	0.58	0.44
6LXD	RF00455	72	29	21	19	17	15	0.6	0.66	0.1	0.52	0.12
6MWN	RF00228	92	68	32	25	35	25	0	0.37	0.22	0.37	0.29
6OL3	RF00102	111	48	43	43	47	47	0	0.9	0	0.98	0
6QN3	RF01704	100	37	29	10	37	32	0	0.27	0.66	0.86	0.14
6V5C	RF00254	66	29	26	25	27	24	0	0.86	0.04	0.83	0.11
6VMY	RF00174	148	47	42	28	41	27	2	0.6	0.33	0.57	0.34
6WLQ	RF00634	119	49	42	33	40	31	0	0.67	0.21	0.63	0.23
6XKO	RF02680	96	31	34	23	30	23	1	0.74	0.32	0.74	0.23
7D81	RF03013	50	18	17	17	17	17	0.1	0.94	0	0.94	0
7ELP	RF03054	45	36	14	12	14	11	0.1	0.33	0.14	0.31	0.21
7JJU	RF04222	102	34	28	0	28	18	0.3	0	1	0.53	0.36
7KGA	RF00525	90	30	31	26	25	21	0.6	0.87	0.16	0.7	0.16
7LYF	RF02340	139	49	50	39	47	37	10	0.8	0.22	0.76	0.21
7QR3	RF00622	69	51	23	22	22	22	0.2	0.43	0.04	0.43	0
7SAM	RF01084	169	56	57	49	57	49	0.5	0.88	0.14	0.88	0.14
7WIB	RF02977	50	18	16	16	16	16	0.2	0.89	0	0.89	0
8DP3	RF00386	90	39	35	35	35	35	0	0.9	0	0.9	0
8FCS	RF00661	71	30	27	27	21	21	0.2	0.9	0	0.7	0
8GZP	RF00185	68	23	21	10	22	20	0.3	0.43	0.52	0.87	0.09
8SH5	RF02521	88	36	35	29	32	27	0	0.81	0.17	0.75	0.16
							Mean:		0.702	0.225	0.713	0.187

Table S-04.

Targets in the Rfam test set. In addition to the same data columns defined in the caption of Table S-01, the second column provides the Rfam ID of each target.