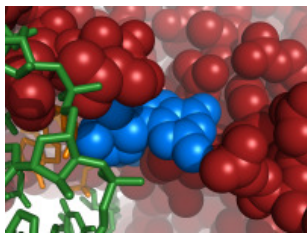


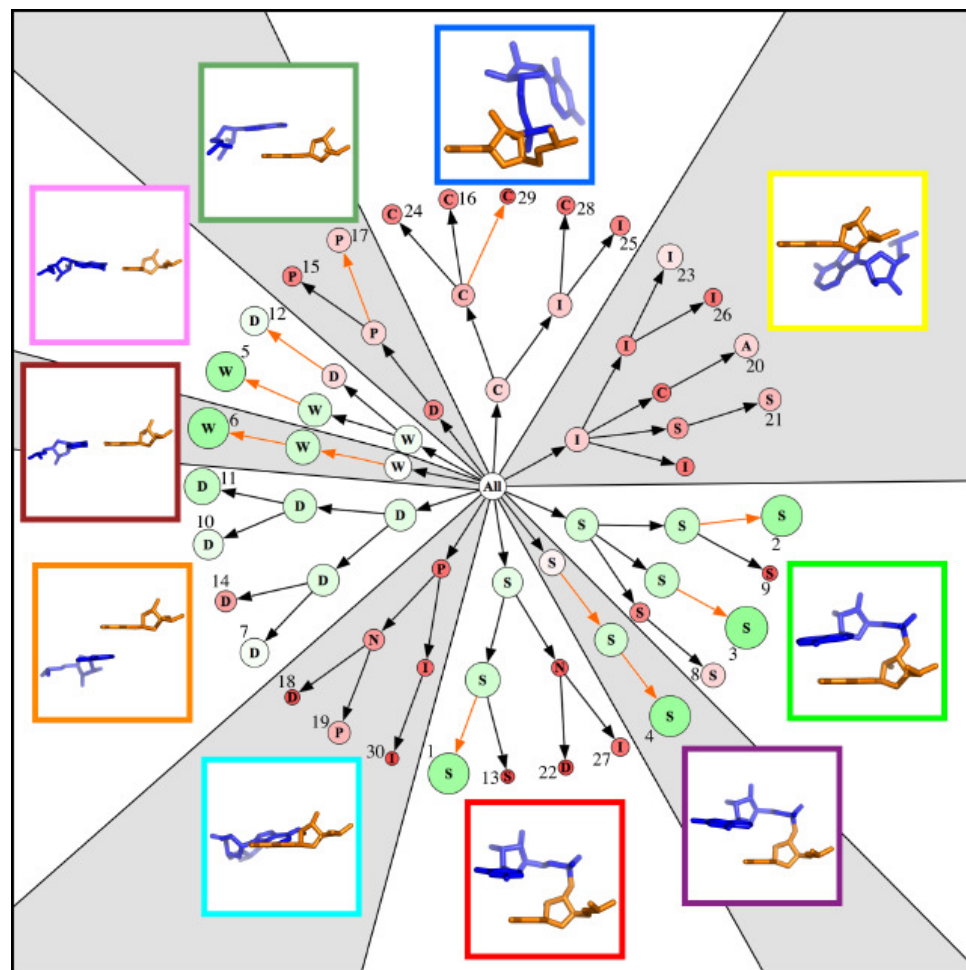
Michael T. Sykes


sykes@scripps.edu
[Home](#)
[Research](#)
[Publications](#)
[Nucleotide Doublets](#)


Nucleotide Doublets

The rapidly increasing wealth of structural information on RNA and knowledge of its varying roles in biology have facilitated the study of RNA structure using computational methods. Here we present a new method to describe RNA structure based on nucleotide doublets, where a doublet is any two nucleotides in a structure. We restrict our search to doublets which are close together in space, but not necessarily in sequence, and obtain doublet libraries of various sizes by clustering a large set of doublets taken from a data set of high resolution RNA structures. We demonstrate that these libraries are able to both capture structural features present in RNA and fit local RNA structure with a high level of accuracy. Libraries ranging in size from 10 to 100 doublets are examined, and a detailed analysis shows that a library with as few as 30 doublets is sufficient to capture the most common structural features, while larger libraries would be more appropriate for accurate modeling. We anticipate many uses for these libraries, from annotation to structure refinement and prediction.

Michael T. Sykes and Michael Levitt. **Describing RNA structure by libraries of clustered nucleotide doublets.** *J. Molecular Biology*, **351**, pp. 26-38, doi:[10.1016/j.jmb.2005.06.024](https://doi.org/10.1016/j.jmb.2005.06.024) (2005) [PDF](#)



Download Libraries

We have made available for download our Nucleotide Doublet Libraries. Library sizes from 10 to 100 are available as tar/gzipped files. The sequence of each library doublet is the original sequence, but all library doublets have been renumbered. Doublets which are connected in chain are renumbered with the first residue "1" and the second residue "2". Doublets which are not connected in chain are renumbered with the first residue "1" and the second residue "3".

Terms of Use: The doublet libraries are available free of charge for personal or Academic use. Any publications which make use of these libraries must cite the reference given above (Sykes and Levitt, *JMB* 2005). For Corporate or Industrial applications please contact Michael T. Sykes.

- [Size 10 Library](#) - Information File: [HTML/Text](#)
- [Size 20 Library](#) - Information File: [HTML/Text](#)
- [Size 30 Library](#) - Information File: [HTML/Text](#)
- [Size 40 Library](#) - Information File: [HTML/Text](#)
- [Size 50 Library](#) - Information File: [HTML/Text](#)
- [Size 100 Library](#) - Information File: [HTML/Text](#)
- Information files include PDB ID, Residue and Chain information as well as cluster size and <RMSD> to the center.
- Key to Annotations:
 - W = Watson-Crick base-pair
 - P = Non-canonical base-pair
 - S = Stacked Interaction
 - D = Diagonal Interaction
 - I = Tertiary Packing Interaction
 - C = Connected in chain, no easily definable structure
 - N = Not connected in chain, no easily definable structure
 - A = Platform, similar to an A-platform

Dblt	PDB ID	Res1	Chain1	Seq1	Res2	Chain2	Seq1	Size	AvgRMSD	
	Annot									
1	1jj2	2130	0	C	2131	0	G	1717	0.83	S
2	lhq1	174	B	G	175	B	A	2802	0.86	S
3	lm5o	15	A	C	16	A	C	2386	0.88	S
4	lrxb	5	A	G	12	B	C	1821	0.92	W
5	1f27	5	A	C	18	A	G	2055	1.06	W
6	lsdr	10	A	G	16	B	A	2426	1.32	D
7	1fjg	888	A	G	908	A	A	986	2.05	D
8	1fjg	142	A	G	196	A	A	801	2.21	P
9	1jj2	1027	0	G	1032	0	A	1458	2.23	I
10	1jj2	480	0	C	481	0	U	1506	2.43	C

Dblt	PDB ID	Res1	Chain1	Seq1	Res2	Chain2	Seq1	Size	AvgRMSD	
	Annot									
1	1jj2	2268	0	C	2269	0	C	1470	0.47	S
2	ln78	540	C	G	541	C	C	1543	0.51	S
3	lqcu	6	A	G	7	A	G	1532	0.53	S
4	1jj2	2130	0	C	2131	0	G	1430	0.55	S
5	1jj2	2335	0	C	2349	0	G	1360	0.61	W
6	lrxb	5	A	G	12	B	C	1460	0.67	W
7	lgid	140	A	A	141	A	G	486	0.76	S
8	lhq1	134	B	C	175	B	A	1245	0.97	D
9	lhq1	131	B	G	176	B	G	775	1.09	D
10	1jj2	1586	0	G	1610	0	G	1402	1.09	D
11	1jj2	2053	0	G	2054	0	A	562	1.70	S
12	1jj2	784	0	A	1459	0	A	420	1.75	I
13	1jj2	2525	0	G	2526	0	C	408	1.86	C
14	1jj2	1805	0	G	2875	0	A	509	1.89	I
15	lgid	150	A	G	153	A	A	755	1.90	P
16	lmzp	23	B	G	44	B	A	541	1.97	N
17	1jj2	2895	0	C	2897	0	C	336	2.01	N
18	lgid	154	A	C	252	A	A	715	2.11	I
19	1fjg	989	A	C	1016	A	A	374	2.13	I
20	leuy	934	B	C	935	B	U	635	2.17	C

Dblt	PDB ID	Res1	Chain1	Seq1	Res2	Chain2	Seq1	Size	AvgRMSD	
	Annot									
1	1jj2	2268	0	C	2269	0	C	1363	0.39	S
2	ln78	540	C	G	541	C	C	1351	0.42	S
3	lqcu	6	A	G	7	A	G	1465	0.52	S
4	1jj2	2130	0	C	2131	0	G	1416	0.53	S
5	1jj2	2335	0	C	2349	0	G	1314	0.56	W
6	lrxb	5	A	G	12	B	C	1354	0.57	W
7	lm5o	20	A	C	3	B	A	658	0.63	D
8	1lmg	219	B	A	220	B	A	505	0.65	S
9	1fjg	1188	A	A	1189	A	C	202	0.75	S
10	lduq	110	A	A	120	B	G	762	0.84	D
11	lhr2	234	A	G	241	A	U	1073	0.89	D
12	lhq1	131	B	G	176	B	G	712	0.98	D
13	lcsl	77	B	U	78	B	U	171	1.37	S
14	1jj2	1310	0	U	1345	0	A	382	1.50	D
15	lgid	224	A	U	248	A	A	309	1.65	P
16	lehz	7	A	U	8	A	U	314	1.69	C
17	lgid	150	A	G	153	A	A	489	1.71	P
18	1fjg	46	A	G	366	A	C	192	1.74	D
19	ljbr	12	C	A	19	F	G	434	1.76	P

20	1fjg	115	A	G	116	A	A	504	1.76	A
21	1kd5	4	A	C	5	A	A	443	1.79	S
22	1yfg	8	A	U	13	A	C	178	1.84	D
23	1jj2	1071	0	G	1292	0	G	541	1.84	I
24	1jj2	1651	0	C	1652	0	C	313	1.85	C
25	1jj2	1807	0	U	2884	0	G	314	1.85	I
26	1jj2	151	0	A	441	0	A	280	1.90	I
27	1fjg	292	A	G	608	A	A	270	1.91	I
28	1fjg	884	A	U	885	A	G	274	1.92	C
29	1euy	934	B	C	935	B	U	208	1.97	C
30	1glx	666	D	G	732	E	C	167	2.35	I

Dblt	PDB ID Annot	Res1	Chain1	Seq1	Res2	Chain2	Seq1	Size	AvgRMSD	
1	1jj2	2268	0	C	2269	0	C	1049	0.28	S
2	1mwl	11	A	C	12	A	G	1115	0.43	S
3	1n78	525	C	C	526	C	G	280	0.43	S
4	397d	43	B	G	44	B	G	1262	0.44	S
5	1n78	540	C	G	541	C	C	1472	0.44	S
6	1jj2	838	0	C	839	0	C	177	0.47	S
7	1rxb	5	A	G	12	B	C	1325	0.54	W
8	1jj2	2335	0	C	2349	0	G	1301	0.54	W
9	1f27	25	B	A	26	B	A	579	0.57	S
10	1k8w	403	B	C	422	B	C	535	0.65	D
11	1fjg	989	A	C	990	A	C	207	0.67	S
12	1jj2	300	0	C	352	0	A	795	0.71	D
13	439d	83	A	G	94	B	C	1265	0.96	D
14	1drz	111	B	G	171	B	A	702	0.96	D
15	1jbr	8	C	C	9	C	A	64	1.04	C
16	1fjg	1482	A	G	1483	A	A	316	1.29	S
17	1msy	2657	A	A	2664	A	G	194	1.30	P
18	1jj2	316	0	A	317	0	A	168	1.32	C
19	1fjg	1418	A	A	1483	A	A	174	1.37	D
20	1fjg	606	A	G	632	A	A	372	1.54	P
21	1jj2	2076	0	U	2077	0	C	190	1.56	C
22	1jj2	784	0	A	1459	0	A	295	1.56	I
23	1jj2	2580	0	G	2600	0	A	272	1.57	P
24	1jj2	1188	0	A	1210	0	G	136	1.58	I
25	1jj2	1351	0	G	1353	0	C	164	1.58	N
26	1gtf	131	W	G	132	W	A	226	1.58	C
27	1jj2	783	0	C	1459	0	A	245	1.60	I
28	1fjg	1249	A	C	1288	A	A	269	1.61	P
29	1jj2	1009	0	U	1010	0	C	182	1.65	A
30	1l8v	139	A	A	163	A	G	155	1.66	D
31	1fjg	1013	A	G	1014	A	A	407	1.69	C
32	1qf6	8	B	U	13	B	C	141	1.69	D
33	1fjg	989	A	C	1016	A	A	228	1.75	I
34	1jj2	2648	0	U	2649	0	A	358	1.78	C
35	1jj2	1994	0	A	2594	0	C	292	1.80	I
36	1jj2	1859	0	A	1879	0	U	246	1.82	I
37	1jj2	631	0	A	632	0	A	190	1.95	C
38	1h4s	20	T	U	21	T	A	447	1.97	C
39	1jj2	1704	0	G	2736	0	U	80	2.04	I
40	1jj2	912	0	A	1292	0	G	83	2.27	I

Dblt	PDB ID Annot	Res1	Chain1	Seq1	Res2	Chain2	Seq1	Size	AvgRMSD
------	-----------------	------	--------	------	------	--------	------	------	---------

1	1c0a	666	B	C	667	B	C	890	0.27	S
2	1n78	540	C	G	541	C	C	1306	0.37	S
3	1jj2	2651	0	C	2652	0	U	365	0.40	S
4	1mwl	11	A	C	12	A	G	1121	0.43	S
5	397d	43	B	G	44	B	G	1263	0.44	S
6	1jj2	1764	0	C	1765	0	G	284	0.47	S
7	1k8w	403	B	C	422	B	C	398	0.50	D
8	1rxb	5	A	G	12	B	C	1297	0.51	W
9	1jj2	1902	0	G	1936	0	C	646	0.52	D
10	1jj2	2335	0	C	2349	0	G	1285	0.52	W
11	1f27	25	B	A	26	B	A	573	0.56	S
12	1jj2	300	0	C	352	0	A	689	0.57	D
13	1mwl	10	A	C	11	A	C	171	0.66	S
14	1fjg	1188	A	A	1189	A	C	187	0.68	S
15	1qc0	109	C	G	131	D	A	672	0.68	D
16	1hq1	131	B	G	176	B	G	596	0.84	D
17	1jbr	8	C	C	9	C	A	61	1.00	C
18	1jj2	2545	0	U	2614	0	C	168	1.04	D
19	1jj2	827	0	A	854	0	G	216	1.29	P
20	1fjg	1418	A	A	1483	A	A	156	1.33	D
21	1n78	545	C	A	546	C	A	259	1.37	C
22	1l8v	129	A	G	131	A	U	56	1.41	N
23	1fjg	992	A	U	1044	A	A	243	1.41	P
24	1b7f	9	P	U	10	P	U	134	1.42	A
25	1jj2	2580	0	G	2600	0	A	206	1.43	P
26	1jj2	2076	0	U	2077	0	C	185	1.48	C
27	1fjg	1213	A	A	1215	A	G	98	1.49	N
28	1fjg	606	A	G	632	A	A	336	1.50	P
29	1jj2	1807	0	U	2883	0	A	164	1.50	I
30	1c9s	144	W	G	145	W	A	216	1.51	C
31	1jj2	151	0	A	185	0	G	128	1.54	D
32	1l9a	198	B	G	199	B	A	314	1.55	C
33	1jj2	587	0	A	589	0	U	106	1.57	N
34	1jj2	209	0	G	666	0	A	299	1.57	I
35	1yfg	8	A	U	13	A	C	118	1.58	D
36	1m5o	22	B	A	60	B	U	115	1.62	P
37	1jj2	2718	0	C	2720	0	C	110	1.62	N
38	1jj2	1079	0	A	2068	0	G	227	1.64	I
39	1jj2	520	0	A	638	0	C	262	1.64	I
40	1jj2	2895	0	C	2897	0	C	118	1.66	N
41	1c9s	133	W	U	134	W	G	277	1.67	C
42	1jj2	2115	0	U	2632	0	G	275	1.68	I
43	1jj2	2472	0	C	2634	0	G	222	1.72	I
44	1mzp	14	B	U	15	B	A	242	1.75	C
45	1fjg	117	A	G	313	A	A	180	1.78	I
46	1jj2	1808	0	C	2885	0	A	197	1.80	I
47	1b7f	8	P	U	9	P	U	278	1.83	C
48	1fjg	782	A	A	1514	A	C	133	1.94	I
49	1gtf	134	W	U	136	W	G	56	2.04	N
50	1fjg	666	A	G	725	A	G	60	2.22	I

Dblt	PDB ID	Res1	Chain1	Seq1	Res2	Chain2	Seq1	Size	AvgRMSD
1	1jj2	1987	0	C	1988	0	C	343	0.20
2	1msy	2649	A	C	2650	A	U	201	0.21
3	1m5o	29	B	C	30	B	C	328	0.22
4	1jj2	1083	0	C	1084	0	C	87	0.27
5	259d	5	A	G	6	A	G	444	0.28
6	1fjg	70	A	G	73	A	C	665	0.29

7	1jj2	2593	0	C	2594	0	C	128	0.30
8	1hr2	120	A	U	121	A	C	82	0.30
9	1n78	523	C	G	524	C	A	402	0.32
10	1jj2	352	0	A	353	0	G	264	0.34
11	280d	20	B	G	21	B	C	649	0.37
12	1jj2	1462	0	C	1481	0	G	831	0.37
13	1fjg	654	A	G	655	A	A	240	0.37
14	1jj2	2130	0	C	2131	0	G	1179	0.40
15	483d	2650	A	U	2651	A	C	35	0.42
16	1k8w	406	B	C	418	B	G	270	0.42
17	1mwl	13	A	G	32	B	C	578	0.43
18	1jj2	2560	0	C	2561	0	C	44	0.44
19	1fjg	909	A	A	910	A	C	102	0.44
20	1ffv	3	T	G	70	T	C	705	0.48
21	1fjg	593	A	G	647	A	C	277	0.48
22	1jj2	838	0	C	839	0	C	131	0.49
23	1hr2	191	A	G	192	A	A	257	0.51
24	1fjg	74	A	C	97	A	G	654	0.51
25	397d	28	A	G	38	B	U	393	0.53
26	1k8w	403	B	C	422	B	C	499	0.54
27	1d4r	4	A	G	26	B	A	582	0.56
28	413d	6	A	U	7	A	U	21	0.62
29	1jj2	2599	0	A	2600	0	A	131	0.63
30	1jj2	2776	0	A	2777	0	G	112	0.64
31	1hq1	131	B	G	176	B	G	513	0.68
32	1fjg	95	A	U	96	A	G	178	0.69
33	1jj2	694	0	A	695	0	C	38	0.75
34	405d	6	A	C	27	B	A	173	0.83
35	1jj2	783	0	C	1459	0	A	65	0.86
36	1jj2	1391	0	G	1392	0	A	55	0.87
37	1n78	526	C	G	544	C	A	50	0.90
38	1jbr	8	C	C	9	C	A	58	0.94
39	1gid	150	A	G	152	A	A	111	0.94
40	1jj2	38	9	A	39	9	U	64	0.94
41	1f7y	12	B	G	13	B	G	24	1.01
42	1jj2	2770	0	G	2805	0	A	138	1.04
43	1c9s	136	W	G	137	W	A	57	1.09
44	1fjg	1482	A	G	1483	A	A	233	1.12
45	1fjg	992	A	U	1044	A	A	158	1.15
46	1jj2	827	0	A	854	0	G	173	1.15
47	1fjg	68	A	G	101	A	A	57	1.16
48	1fjg	1418	A	A	1483	A	A	126	1.16
49	1jj2	520	0	A	638	0	C	92	1.17
50	1qf6	32	B	U	33	B	U	35	1.19
51	1fjg	1432	A	G	1468	A	A	208	1.20
52	420d	6	A	G	7	A	U	16	1.20
53	1fjg	909	A	A	1413	A	A	64	1.22
54	1jj2	2580	0	G	2600	0	A	134	1.24
55	1fjg	297	A	G	298	A	A	175	1.24
56	1jj2	1554	0	U	1631	0	A	120	1.29
57	1fjg	1213	A	A	1215	A	G	74	1.31
58	1jj2	1493	0	A	1513	0	C	102	1.31
59	1h4s	8	T	U	13	T	C	66	1.33
60	1jj2	1787	0	C	2875	0	A	77	1.35
61	1fjg	1297	A	C	1298	A	C	101	1.37
62	1jj2	2076	0	U	2077	0	C	160	1.38
63	118v	124	A	C	126	A	G	53	1.40
64	1d4r	13	A	G	16	B	G	69	1.40
65	1hnh	70	A	U	140	A	A	104	1.41

66	1gtf	131	W	G	132	W	A	172	1.41
67	1jj2	1079	0	A	2068	0	G	165	1.42
68	1jj2	2507	0	G	2510	0	C	71	1.43
69	1jj2	1807	0	U	2883	0	A	141	1.43
70	1jj2	1237	0	U	1238	0	C	48	1.45
71	1jj2	784	0	A	1459	0	A	244	1.46
72	1cx0	130	B	C	167	B	U	28	1.46
73	1jj2	204	0	A	436	0	A	85	1.49
74	1jj2	521	0	A	1365	0	C	119	1.50
75	1f7u	948	B	U	959	B	G	26	1.50
76	1dul	143	B	A	144	B	G	125	1.52
77	1jj2	706	0	G	721	0	A	47	1.52
78	1jj2	2380	0	A	2381	0	C	124	1.53
79	1ffv	21	T	U	21A	T	A	135	1.55
80	1fjg	758	A	G	821	A	G	23	1.55
81	1jj2	2718	0	C	2720	0	C	96	1.55
82	1qf6	18	B	G	58	B	A	74	1.55
83	1jj2	2895	0	C	2897	0	C	88	1.56
84	1jj2	2555	0	C	2577	0	A	41	1.56
85	1fjg	27	A	G	297	A	G	27	1.58
86	1jj2	2729	0	C	2892	0	G	41	1.59
87	1jj2	1544	0	U	1619	0	G	130	1.60
88	1jj2	2674	0	G	2811	0	A	101	1.62
89	1jj2	2009	0	G	2014	0	G	96	1.63
90	1jj2	2557	0	U	2684	0	A	87	1.64
91	1fjg	989	A	C	1017	A	G	199	1.64
92	1jj2	519	0	A	639	0	A	98	1.65
93	1qtq	944	B	C	945	B	A	107	1.65
94	1fjg	561	A	U	562	A	C	171	1.66
95	1jj2	410	0	A	411	0	A	124	1.68
96	1jj2	1387	0	G	1392	0	A	75	1.69
97	1jj2	869	0	G	886	0	A	42	1.72
98	1fjg	663	A	A	835	A	U	95	1.72
99	1jj2	241	0	A	269	0	G	86	1.73
100	1jj2	1116	0	U	1139	0	U	72	1.86