

Supplementary Data for Dual Graph Partitioning Highlights a Small Group of Pseudoknot-Containing RNA Submotifs

Swati Jain¹, Cigdem Sevim Bayrak¹, Louis Petingi², and Tamar Schlick^{1,3,4,*}

¹Department of Chemistry, New York University, New York, NY, USA

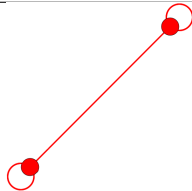
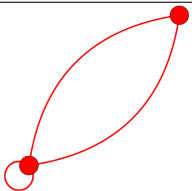
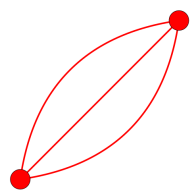
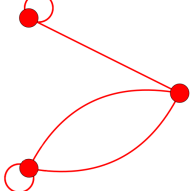
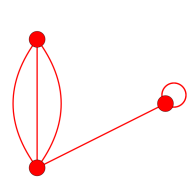
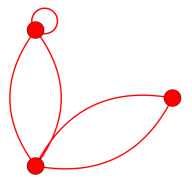
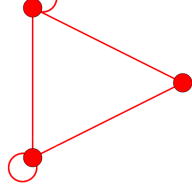
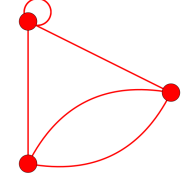
²Computer Science Department, College of Staten Island, City University of New York, New York, NY, USA

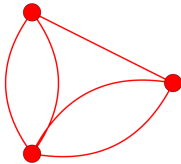
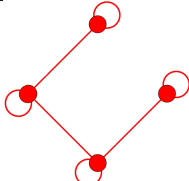
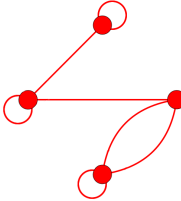
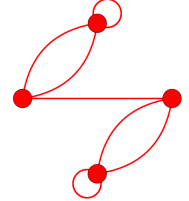
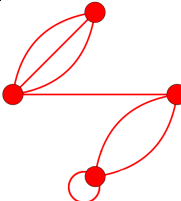
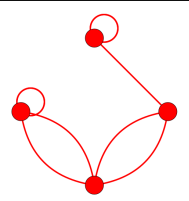
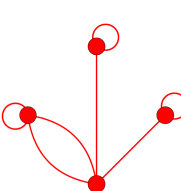
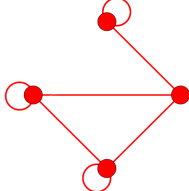
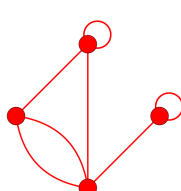
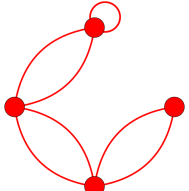
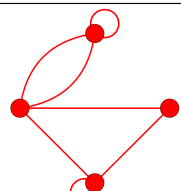
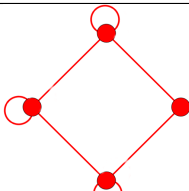
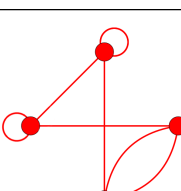
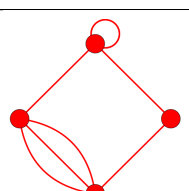
³Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

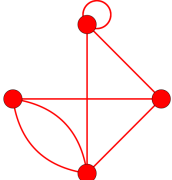
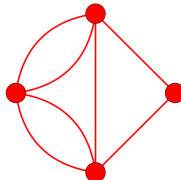
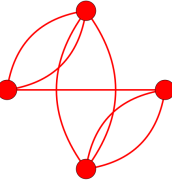
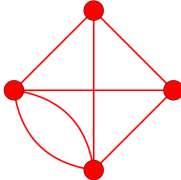
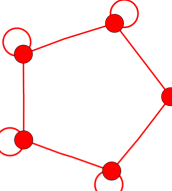
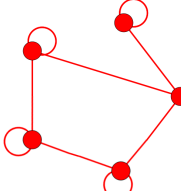
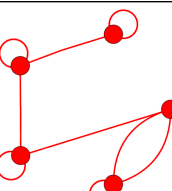
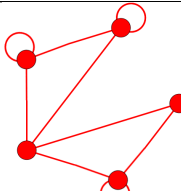
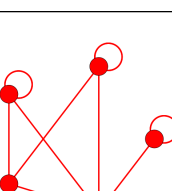
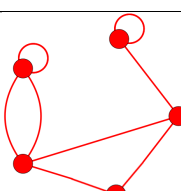
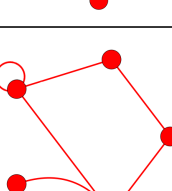
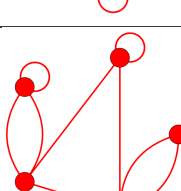
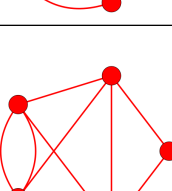
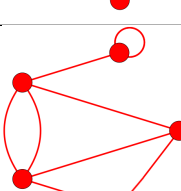
⁴NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, China

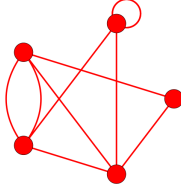
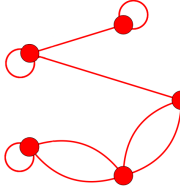
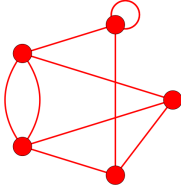
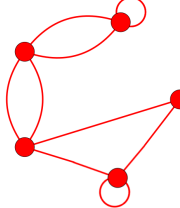
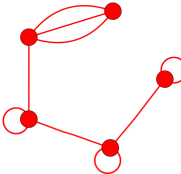
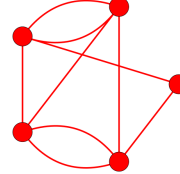
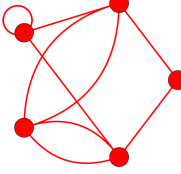
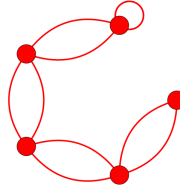
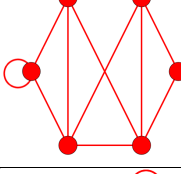
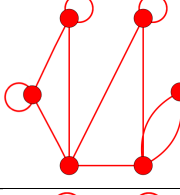
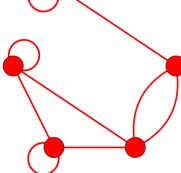
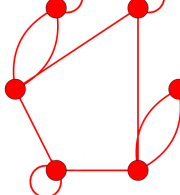
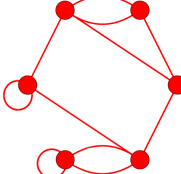
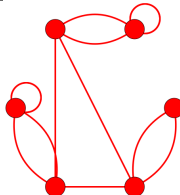
*Corresponding author: schlick@nyu.edu

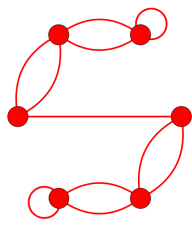
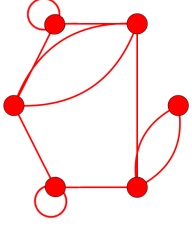
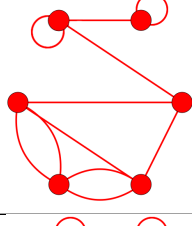
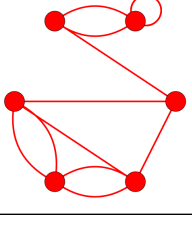
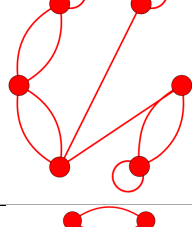
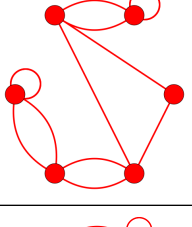
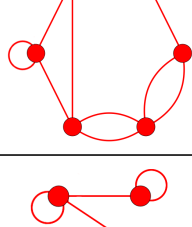
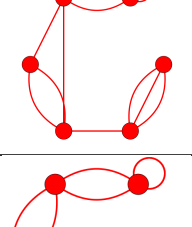
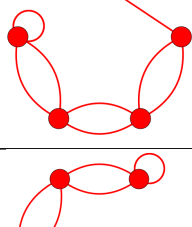
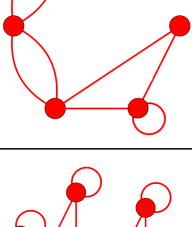
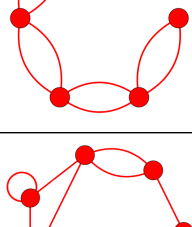
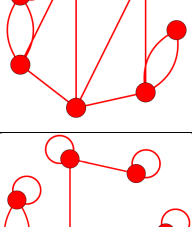
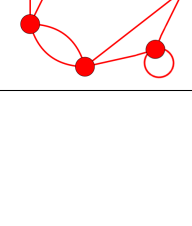
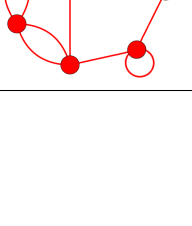
Table S1: 94 dual graph topologies between 2-9 vertices, with IDs and number of structures, for the representative RNA structure dataset

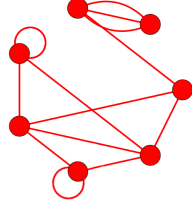
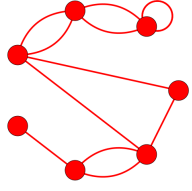
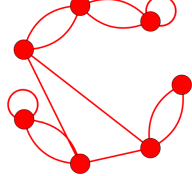
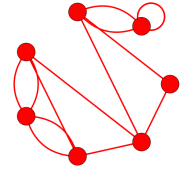
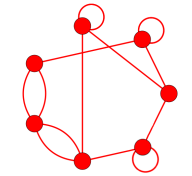
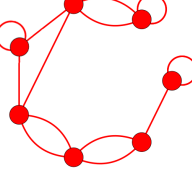
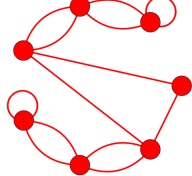
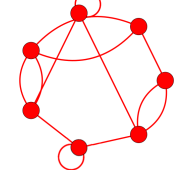
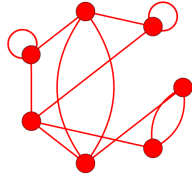
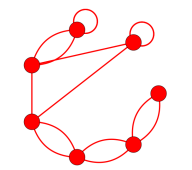
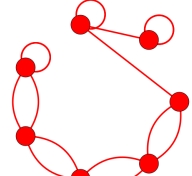
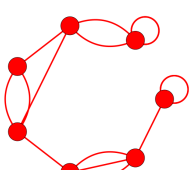
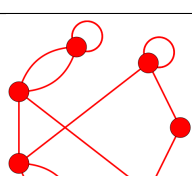
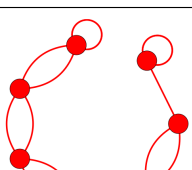
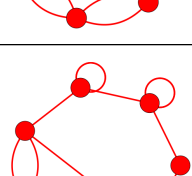
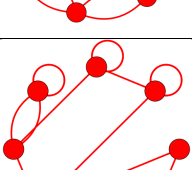
RAG ID	Structures	Topology	RAG ID	Structures	Topology
2_1	6		2_2	269	
2_3	32		3_2	8	
3_3	1		3_4	68	
3_5	23		3_6	7	

3_8	11		4_1	3	
4_2	2		4_4	2	
4_5	1		4_8	3	
4_9	1		4_10	1	
4_12	5		4_14	21	
4_16	16		4_19	105	
4_20	1		4_21	1	

4_25	5		4_27	14	
4_29	1		4_30	1	
5_2	15		5_4	1	
5_6	3		5_9	1	
5_11	1		5_20	2	
5_21	3		5_32	12	
5_35	1		5_45	1	

5_46	1		5_48	1	
5_52	2		5_62	6	
5_64	1		5_68	1	
5_94	1		5_96	10	
6_36	1		6_57	1	
6_70	1		6_91	2	
6_101	1		6_109	1	

6_198	1		6_226	1	
6_235	1		6_236	1	
6_257	2		6_263	13	
6_281	1		6_303	1	
6_311	1		6_372	2	
6_460	2		7_222	1	
7_729	2		7_934	1	

7_1082	1		7_1091	1	
7_1119	2		7_1141	1	
7_1219	1		7_1235	1	
7_1311	8		7_1329	1	
7_1341	2		7_1714	3	
7_1928	1		7_1949	1	
7_2019	4		7_2244	1	
7_2329	1		8_721	1	

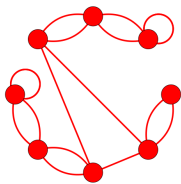
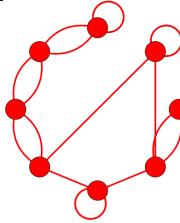
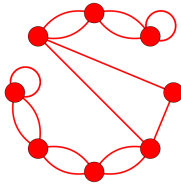
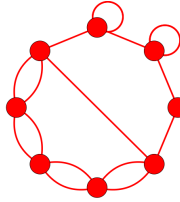
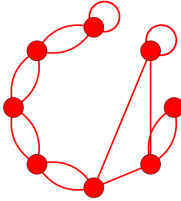
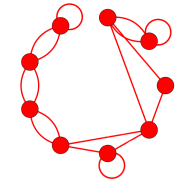
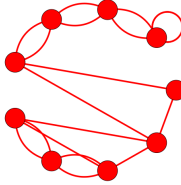
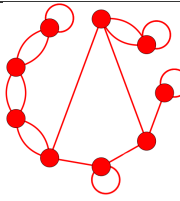
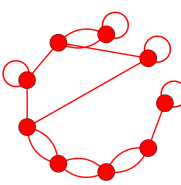
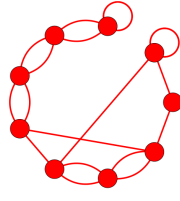
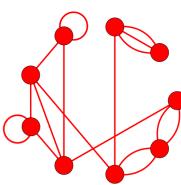
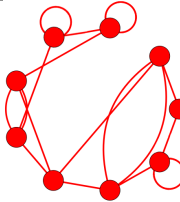
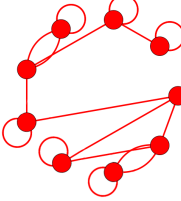
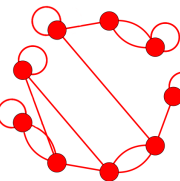
8_5491	2		8_7748	1	
8_8221	1		8_9902	1	
8_10647	1		9_18405	1	
9_19203	1		9_20569	1	
9_20790	1		9_35189	1	
9_38596	1		9_38597	3	
9_38598	1		9_38599	1	

Table S2: Number of structures in the representative dataset that have dual graphs with ≥ 10 vertices

Vertices	Structures	Vertices	Structures	Vertices	Structures	Vertices	Structures
10	9	11	4	12	2	13	6
14	5	15	3	17	2	18	1
20	1	23	1	24	1	29	1
31	1	38	1	48	1	50	1
54	1	55	1	56	1	61	1
71	1	72	1	75	1	76	2
77	3	78	2	79	2	80	2
81	1	83	2	84	1	85	1
86	1	87	6	88	1	90	1
96	2	98	2	100	4	101	1
128	1	141	1	145	1	147	1
151	1	156	1	157	2	158	3
161	1	162	1	163	2	166	1
169	1	171	1	174	3	176	1
184	1	186	1	187	1	192	1
199	1						

Table S3: Dual graph topologies that were removed from the updated library of existing dual graph topologies due to reasons as listed.

Topology	Reason for removal
3_1	Duplicate chains (e.g., PDB ID: 1Q96) and SRP database
3_7	Pseudobase++ database
4_15	Rfam database
4_17	Pseudobase++ database
4_22	Rfam database and Pseudobase++ database
4_28	Pseudobase++ database
5_10	Rfam database
5_22	Pseudobase++ database
5_25	Pseudobase++ database
5_38	Pseudobase++ database
5_39	Pseudobase++ database
5_41	Pseudobase++ database
5_43	Duplicate chains (e.g., PDB ID: 1ML5)
5_51	Pseudobase++ database
5_55	Pseudobase++ database
5_70	Pseudobase++ database
5_77	Pseudobase++ database
5_91	Non-canonical base pair considered (e.g., PDB ID: 1SER)
6_1	Duplicate chains (e.g., PDB ID: 2DEU)
6_7	Duplicate chains (e.g., PDB ID: 1MME)
6_42	Duplicate chains (e.g., PDB ID: 1J2B) and Rfam database
6_66	Pseudobase++ database
6_78	Larger structure in the representative dataset (e.g., PDB ID: 2R8S)
6_80	Pseudobase++ database
6_127	Pseudobase++ database
6_149	Rfam database
7_74	Duplicate chains (e.g., PDB ID: 1UN6)
7_137	Chains separated in the representative dataset (e.g., PDB ID: 1YSH)
7_513	Pseudobase++ database
7_2314	SRP database
7_2315	Pseudobase++ database
8_1199	Duplicate chains (e.g., PDB ID: 1ZHO)
8_2314	Chains separated in the representative dataset (e.g., PDB ID: 2GCV)
9_20589	SRP database
9_28291	Duplicate chains (e.g., PDB ID: 406D)
9_30950	SRP database
9_38133	SRP database

Table S4: Number of subgraph blocks ≥ 10 vertices identified by the dual graph partitioning algorithm in the representative dataset of RNA structures

No. of vertices	No. of occurrences
11	7
12	3
17	1
21	1
23	1
37	1
39	1
40	1
41	5
42	2
43	2
44	2
46	1
47	1
67	1
68	3
70	4
71	1
72	2
73	1
75	1
76	5
90	1

Table S5: PDB IDs and chains for rRNAs of small ribosomal subunits for 10 prokaryotic and 12 eukaryotic species

PDB ID	Chain	Species	Organism	rRNA
Prokaryotes				
4V4N	B2	<i>Methanocaldococcus jannaschii</i>	Archae	16S
5JB3	2	<i>Pyrococcus abyssi</i>	Archae	16S
4V6U	A2	<i>Pyrococcus furiosus</i>	Archae	16S
3J9W	AA	<i>Bacillus subtilis</i>	Bacteria	16S
4V4Q	CA	<i>Escherichia coli</i>	Bacteria	16S
5MYJ	AA	<i>Lactococcus lactis</i>	Bacteria	16S
5O61	BA	<i>Mycobacterium smegmatis</i>	Bacteria	16S
5V93	a	<i>Mycobacterium tuberculosis</i>	Bacteria	16S
5ND8	a	<i>Staphylococcus aureus</i>	Bacteria	16S
1FJG	A	<i>Thermus thermophilus</i>	Bacteria	16S
Eukaryotes				
3JBN	A	<i>Plasmodium falciparum</i>	Protozoa	18S
4BTS	DA	<i>Tetrahymena thermophila</i>	Protozoa	18S
5XYI	2	<i>Trichomonas vaginalis</i>	Protozoa	18S
4V7H	AA	<i>Thermomyces lanuginosus</i>	Fungi	18S
3JAM	2	<i>Kluyveromyces lactis</i>	Yeast	18S
4UER	A	<i>Lachancea kluyveri</i>	Yeast	18S
4V88	A6	<i>Saccharomyces cerevisiae</i>	Yeast	18S
4V7E	Ad	<i>Triticum aestivum</i>	Plant	18S
4V6W	B2	<i>Drosophila melanogaster</i>	Fly	18S
5VYC	i4	<i>Homo sapiens</i>	Mammal	18S
4KZY	i	<i>Oryctolagus cuniculus</i>	Mammal	18S
3J7P	S2	<i>Sus scrofa</i>	Mammal	18S

Table S6: PDB IDs and chains for rRNAs of large ribosomal subunits for 11 prokaryotic and 11 eukaryotic species

PDB ID	Chain	Species	Organism	rRNA
Prokaryotes				
4V9F	0	<i>Haloarcula marismortui</i>	Archae	23S
4V4N	A1	<i>Methanocaldococcus jannaschii</i>	Archae	23S
4V6U	B1	<i>Pyrococcus furiosus</i>	Archae	23S
3J9W	BA	<i>Bacillus subtilis</i>	Bacteria	23S
4IOA	X	<i>Deinococcus radiodurans</i>	Bacteria	23S
5J7L	DA	<i>Escherichia coli</i>	Bacteria	23S
5MYJ	BA	<i>Lactococcus lactis</i>	Bacteria	23S
5O60	A	<i>Mycobacterium smegmatis</i>	Bacteria	23S
5V7Q	A	<i>Mycobacterium tuberculosis</i>	Bacteria	23S
5HL7	X	<i>Staphylococcus aureus</i>	Bacteria	23S
5FDU	1A	<i>Thermus thermophilus</i>	Bacteria	23S
Eukaryotes				
3J79	AC	<i>Plasmodium falciparum</i>	Protozoa	28S/5.8S
4V8P	F1E2	<i>Tetrahymena thermophila</i>	Protozoa	26S/5.8S
5XY3	14	<i>Trichomonas vaginalis</i>	Protozoa	25S/5.8S
4V7H	B5B4	<i>Thermomyces lanuginosus</i>	Fungi	26S/5.8S
4V91	14	<i>Kluyveromyces lactis</i>	Yeast	25S/5.8S
4V88	A5A8	<i>Saccharomyces cerevisiae</i>	Yeast	25S/5.8S
4V7E	AaAc	<i>Triticum aestivum</i>	Plant	25S/5.8S
4V6W	A5A8	<i>Drosophila melanogaster</i>	Fly	28S/5.8S
4V6X	A5A8	<i>Homo sapiens</i>	Mammal	28S/5.8S
5LZS	58	<i>Oryctolagus cuniculus</i>	Mammal	28S/5.8S
3J7P	58	<i>Sus scrofa</i>	Mammal	28S/5.8S

Table S7: Occurrences of non-separable dual graph blocks between 2-9 vertices in 660 of the 863 RNA 2D structure files without rRNA residues

Subgraph ID	Occurrences	Subgraph ID	Occurrences
2_2	636	4_19	125
2_1	107	3_5	84
2_3	39	4_27	17
5_2	16	3_6	14
3_8	11	4_25	7
5_49	7	6_369	5
5_25	2	5_5	2
5_52	2	7_1341	2
7_729	2	9_38601	2
4_20	1	4_21	1
4_22	1	4_30	1
5_14	1	5_35	1
5_46	1	5_68	1
5_94	1	6_190	1
6_267	1	6_28	1
6_281	1	6_36	1
7_1219	1	7_1329	1
7_2329	1	7_747	1
8_9902	1	9_23267	1
9_38602	1	9_38603	1