Table S4: Clusters of dinucleotide binding sites in the AND-set dataset. The first column is the cluster number (according to the enumeration of the binding site that originated the cluster). The second column specifies the maximal SCOP distance between the cluster members. Proteins at distance 3 belong to the same SCOP super family, while proteins at distance 7 have totally different overall folds and no common structural annotation. The distance was measured according the pre-SCOP annotation when available. Distance 0 means that some of the annotations are missing and the distance can not be automatically calculated. Column 3 indicates whether the proteins have similar sequences (more than 25% sequence identity or same Pfam annotation). Column 4 details whether all the cluster dinucleotide pairs are π-stacked (1) or non π-stacked (2). Column 5 indicates the number of binding sites in the cluster and column 6 provides the cluster members and column 7 details their pre-SCOP annotation.

Cl.	Scop dist.	Sim. seq.	π type	Cluster size	Cluster members (pdb:chains:nts)	Pre-SCOP annotations
3	0	1	2	2	1wpuACU6U7 2gztADU13U14	missing
12	0	0	2	3	2gztADU10A11 1wpuACG3A4 2j02HAU607A608	missing
14	0	1	2	2	2gztADG12U13 1wpuACG5U6	missing
21	3	1	2	6	1sdsADG9U10 2hvyDEA31U32 1m90HAA251U252 1rlgACG15U16 2ozbACA11U12 1e7kACA5U6	L30e-like and missing
23	7	0	2	2	1fxlABU1U2 1pvoAGU1C2	RNA-binding domain and Rho termination factor
25	3	1	2	2	1fxlABU3U4 1b7fAPU6U7	RNA-binding domain
28	6	0	2	2	1fxlABU6U7 1sdsADU10G11	RNA-binding domain and L30e-like
29	3	1	2	2	1fxlABU7U8 1b7fAPU10U11	RNA-binding domain
41	0	0	2	3	2anrABU12C13 2py9AFC9C10 1ec6ADU12C13	missing and Eukaryotic type KH-domain (KH-domain type I)
42	0	0	2	3	2anrABC13A14 1ec6ADC13A14 2py9AFC10C11	missing and Eukaryotic type KH-domain (KH-domain type I)
47	0	1	2	2	2b3jAEC7G8 2b3jCGC5G6	missing
52	2	1	3	3	1r3eACC7G8 1k8wABC10G11 2ix1ABA8A9	PUA domain-like and missing
54	7	0	3	2	1n78ACC33U34 1b7fAPU3G4	Aminoacyl-tRNA synthetases and RNA-binding domain
55	0	0	3	3	1n78ACU34C35 1wpuACA4G5 2gztADA11G12	Aminoacyl-tRNA synthetases and missing
58	7	0	1	2	1n78ACC73A74 1qtqABC73A74	Aminoacyl-tRNA synthetases and Ribosomal protein L25-like
59	7	0	2	2	1jbrACA16G17 1m90PAA689U690	Microbial ribonucleases and Ribosomal proteins L15p and L18e
61	0	0	1	2	2bh2ACU8C9 1hq1ABA11C12	missing and Signal peptide-binding domain
67	0	0	2	2	2bh2ACG22U23 2du3ADA35G36	missing

76	0	0	2	2	1yvpACC7G8 1il2ACC66A67	missing and Nucleic acid-binding proteins
78	0	1	2	3	1yvpAGC2C3 2i91ADU11C12 1yvpAGC3G4	missing
81	0	0	3	3	1m8xACU2G3 1m8xACU4A5 1m8xACU6A7	ARM repeat
82	0	0	2	3	1m8xACG3U4 1m8xACA7U8 1m8xACA5U6	ARM repeat
94	7	1	2	6	1m5oCBC40A41 1sj3PRC53A54 1urnAPC10A11 1a9nBQC11A12 2a8vADC1C2 2a8vBEC1C2	RNA-binding domain and Rho termination factor
97	1	1	1	3	1sj3PRA49U50 1m5oCBA36U37 1urnAPA6U7	RNA-binding domain
99	3	1	1	4	1sj3PRU51G52 1urnAPU8G9 1a9nBQU9G10 1m5oCBU38G39	RNA-binding domain
100	3	1	2	5	1sj3PRG52C53 1m5oCBG39C40 1urnAPG9C10 1a9nBQG10C11 2g4bABU4U5	RNA-binding domain and missing
103	0	0	2	2	1qtqABC32U33 2q66AXA3A4	Ribosomal protein L25-like and missing
105	7	0	2	2	1qtqABC72C73 1n78ACC72C73	Ribosomal protein L25-like and Aminoacyl-tRNA synthetases
108	0	0	2	2	1wsuAEG11U12 2pjpABG12U13	missing
120	3	1	2	3	1a9nBQU8U9 1sj3PRU50U51 1urnAPU7U8	RNA-binding domain
124	7	1	1	5	1a9nBQA12G13 1m5oCBA41C42 1urnAPA11C12 1sj3PRA54C55 1u0bBAG33C34	RNA-binding domain and Aminoacyl-tRNA synthetases
134	2	1	3	3	1c0aABU31C32 1il2ACU31C32 1asyARU31C32	Nucleic acid-binding proteins
137	1	1	1	2	1c0aABC66C67 1il2ACC65C66	Nucleic acid-binding proteins
157	0	0	2	2	2g4bABU5U6 1b7fAPU9U10	missing and RNA-binding domain
159	1	1	2	6	2iznARU9A10 1zdjARC5A6 7msfARC8A9 6msfARC8A9 2bs0ARA6A7 2b2dCSA1A2	missing and RNA bacteriophage capsid protein
164	0	0	1	3	2py9AFC11U12 1ec6ADA14C15 2anrABA14C15	missing and Eukaryotic type KH-domain (KH-domain type I)
168	7	0	1	2	1cvjAMA4A5 1knzAWG2A3	RNA-binding domain and NSP3 homodimer
175	0	1	1	2	1g1xBEA13A14 2j02OAA694A695	S15/NS1 RNA-binding domain and missing
178	0	1	1	2	1g1xCEC5G6 2j02RAC686G687	Ribosomal protein S18 and missing
182	2	0	2	2	1il2ACG30U31 1asyARG30U31	Nucleic acid-binding proteins
184	1	1	1	2	1il2ACG64C65 1c0aABG65C66	Nucleic acid-binding proteins
194	3	1	2	2	1b7fAPG4U5 1cvjAMA8A9	RNA-binding domain

195	3	1	2	2	1b7fAPU5U6 1fxlABU2U3	RNA-binding domain
198	3	1	2	2	1b7fAPU8U9 1fxlABA5U6	RNA-binding domain
202	3	1	2	4	2ozbACA10A11 1e7kACA4A5 1m90HAA250A251 2hvyDEC30A31	missing and L30e-like
208	2	1	2	2	1i6uACU25A26 1s03HAA13A14	Ribosomal protein S8
235	0	0	2	2	2j017AG116G117 1jidABG11G12	missing and SRP19
242	0	0	2	2	2j01FAA293G294 1m90EAA316U317	missing and Ribosomal protein L4
280	0	1	1	2	2j02RAG684C685 1g1xCEG3C4	missing and Ribosomal protein S18
281	0	1	3	2	2j02RAC685C686 1g1xCEC4C5	missing and Ribosomal protein S18
299	0	1	2	2	1m90CAA862A863 2j01DAA741A742	Translation proteins SH3-like domain and missing
302	0	1	2	2	1m90CAA1834C1835 2j01DAA1685U1686	Translation proteins SH3-like domain and missing
354	7	0	2	2	1q2rAEC8U9 1u0bBAC73A74	tRNA-guanine transglycosylase and Aminoacyl-tRNA synthetases
362	0	0	3	2	1asyARC65C66 1asyARC66A67	Nucleic acid-binding proteins