Table S5: Clusters of dinucleotide binding sites in the OR-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.	Same Pfam	π type	Cluster size	Cluster members (pdb:chains:nts)	Pre-SCOP annotations
0	0	0	2	2	1wpuACG3A4 1i6uACU25A26	missing and Ribosomal protein S8
18	0	0	2	2	1fxlABU1U2 1fxlABU7U8	RNA-binding domain
23	6	0	2	2	1fxlABU6U7 1sdsADU10G11	RNA-binding domain and L30e-like
32	0	0	1	2	1urnAPU8G9 1urnAPA11C12	RNA-binding domain
34	7	0	2	2	1urnAPC10A11 2a8vADC1C2	RNA-binding domain and Rho termination factor
36	0	0	2	3	2anrABU12C13 2py9AFC9C10 1ec6ADU12C13	missing and Eukaryotic type KH-domain (KH-domain type I)
37	0	0	2	3	2anrABC13A14 1ec6ADC13A14 2py9AFC10C11	missing and Eukaryotic type KH-domain (KH-domain type I)
46	0	0	3	2	1n78ACC33U34 2bx2LRU1U2	Aminoacyl-tRNA synthetases and missing
47	0	0	3	2	1n78ACU34C35 1wpuACA4G5	Aminoacyl-tRNA synthetases and missing
49	7	0	2	2	1n78ACC72C73 1qtqABC72C73	Aminoacyl-tRNA synthetases and Ribosomal protein L25-like
50	7	0	1	2	1n78ACC73A74 1qtqABC73A74	Aminoacyl-tRNA synthetases and Ribosomal protein L25-like
68	0	0	2	2	1m8xBDU1G2 1m8xBDU3A4	ARM repeat
69	0	0	2	2	1m8xBDG2U3 1m8xBDA6U7	ARM repeat
72	6	0	3	2	1m8xBDU5A6 1hq1ABA11C12	ARM repeat and Signal peptide-binding domain
78	0	0	2	2	1qtqABC32U33 2q66AXA3A4	Ribosomal protein L25-like and missing
89	0	0	2	2	1ec6ADA11U12 1yvpAGC2C3	Eukaryotic type KH-domain (KH-domain type I) and missing
110	0	0	1	3	2py9AFC11U12 1ec6ADA14C15 2anrABA14C15	missing and Eukaryotic type KH-domain (KH-domain type I)
143	0	0	2	2	1asyARG30U31 2bh2ACG22U23	Nucleic acid-binding proteins and missing
144	2	0	3	2	1asyARU31C32 1c0aABU31C32	Nucleic acid-binding proteins

146	0	0	3	2	1asyARC65C66 1asyARC66A67	Nucleic acid-binding proteins