eLIFE
elifesciences.org

Figures and figure supplements

Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information

Sergey Ovchinnikov, et al.

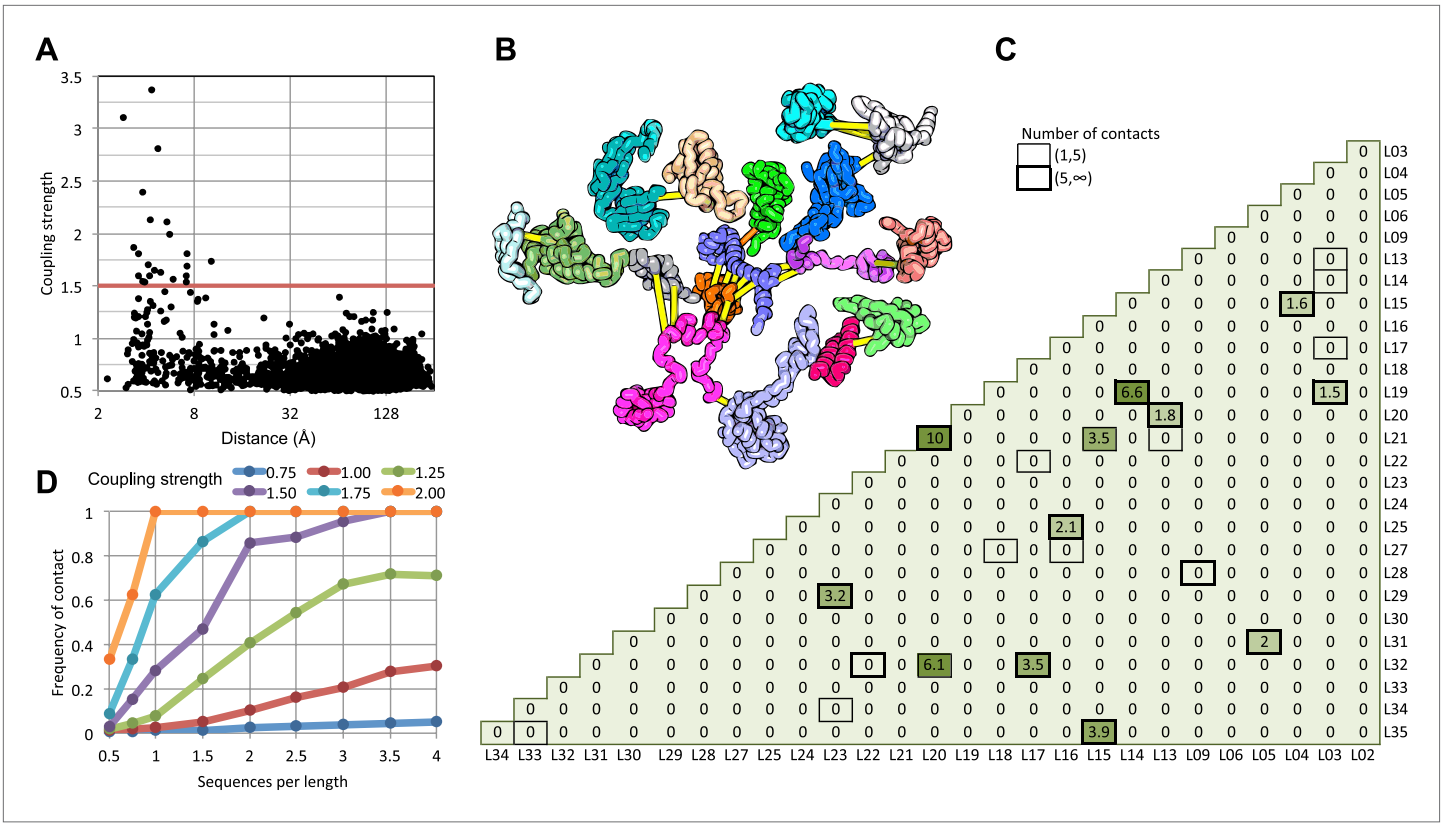

Figure 1. Residue pairs with high normalized coupling strengths are in contact in the 50S ribosomal subunit. (**A**) Coupling strengths and inter-residue distances for each residue pair in the 50S subunit (black dots). Residue pairs with coupling strength greater than 1.5 are nearly always less than 8 Å apart. (**B**) Locations of coevolving (high coupling strength) residue pairs in the protein component of the 50S subunit. The monomers have been pulled apart slightly for clarity. Lines connect residue pairs with coupling strength greater than 1.5; yellow, distance less than 8 Å; orange, distance less than 12 Å. (**C**) Protein pairs with strong inter-residue covariation (colors) make contact in the three-dimensional structure (black boxes). For each protein pair, the sum of the coupling strength greater than 1.5 for each pair of 50S subunit proteins is indicated; black boxes indicate contacts in the crystal structure. (**D**) Dependence of contact prediction accuracy on coupling strength and the number of sequences in the alignments. For each of the indicated coupling strength cutoffs (colors), the frequency of contact in the 50S structure (y axis) was computed for sub alignments with different sequence depths (x axis).
DOI: 10.7554/eLife.02030.003

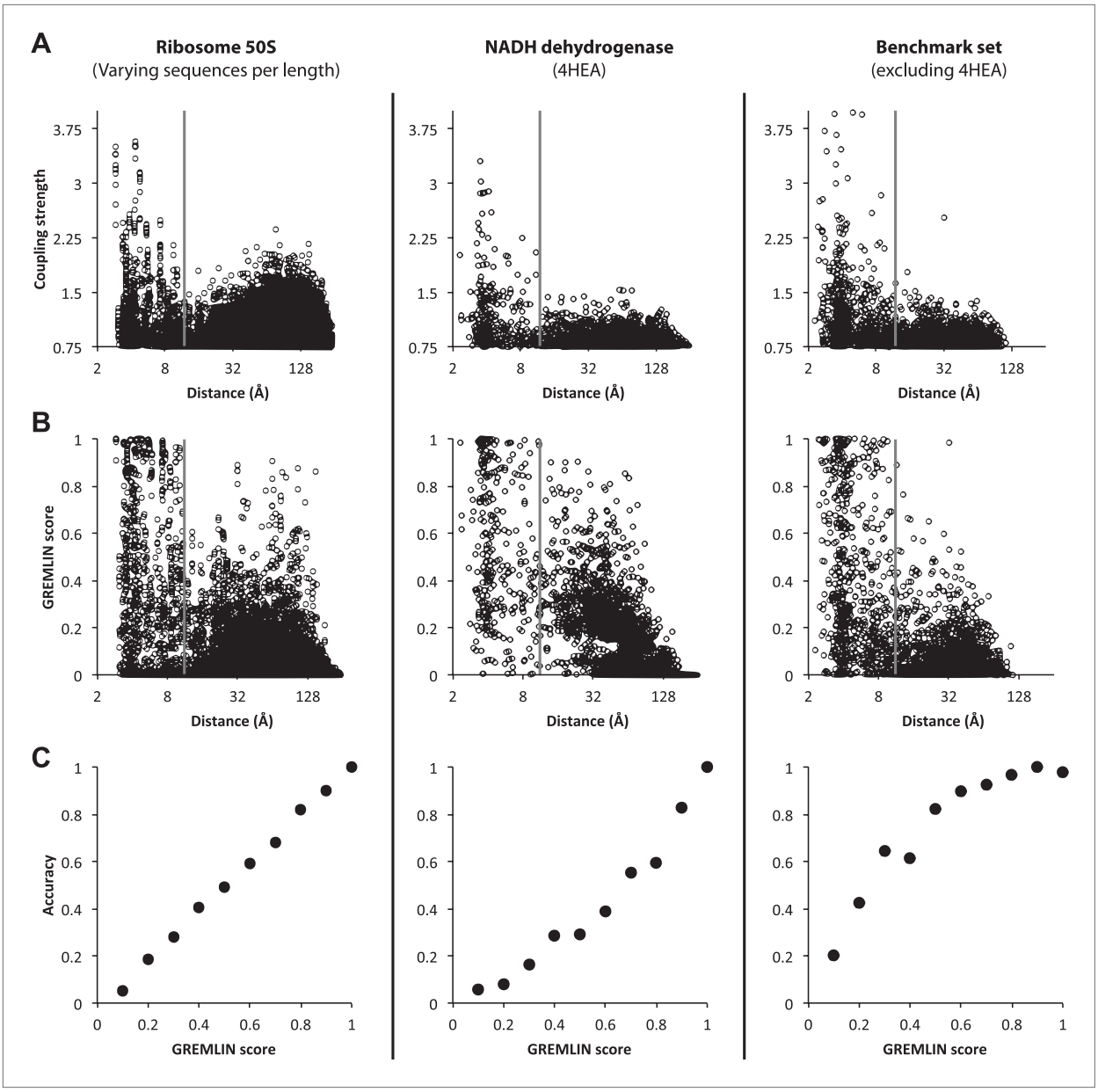

Figure 1—figure supplement 1. Determining GREMLIN scores from normalized coupling strengths.

DOI: 10.7554/eLife.02030.004

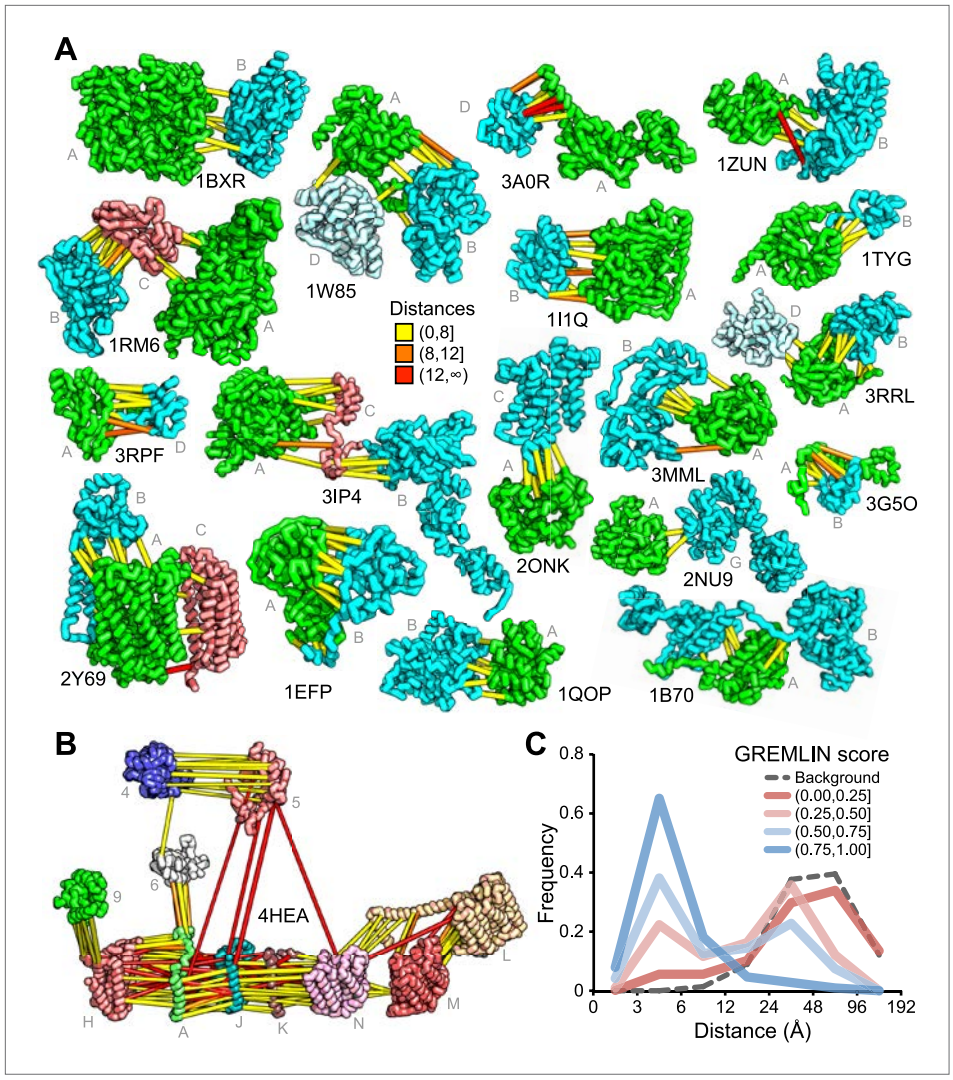

Figure 2. Residue covariation in complexes with known structures. (**A**) Residue-pairs across protein chains with high GREMLIN scores almost always make contact across protein interfaces in experimentally determined complex structures. All contacts with GREMLIN scores greater than 0.6 are shown; the structures are pulled apart for clarity. Labels are according to chains in the PDB structure. (**B**) Complex I of the electron transport chain has an unusually large number of highly co-varying inter residue pairs not in contact in the crystal structure of 4HEA; these contacts may be formed in different state of the complex. Residue pairs within 8 Å are in yellow, between 8 Å and 12 Å in orange, and greater than 12 Å, in red. Distances are the minimal distances between any side chain heavy atom. Labels are according to chains in 4HEA. (**C**) Dependence of inter-residue distance distributions on GREMLIN score. All residue–residue pairs between subunits in the benchmark set were grouped into four bins based on their GREMLIN score (colors), and the distribution of residue–residue distances (x axis) within each bin computed from the three-dimensional structures. See ***Figure 2—source data 1*** for the table of all the interfaces used in the calculation.

DOI: 10.7554/eLife.02030.005

YIAM_YIAN (3.9L)		
21_F	246_F	1.00
91_I	25_L	1.00
42_R	52_D	0.97
14_A	268_L	0.93
87_L	26_L	0.89
59_D	273_K	0.89
17_S	245_A	0.87
28_Y	234_L	0.84
43_Y	48_V	0.82
10_A	268_L	0.79
64_Q	17_G	0.76
56_A	276_S	0.75
9_L	272_A	0.75
46_V	52_D	0.72
13_L	271_A	0.67
50_F	51_A	0.66
95_L	47_L	0.62
46_V	51_A	0.60
10_A	269_I	0.60
YIAN_YIAO (3.6L)		
168_S	196_Y	1.00
296_E	62_K	0.89
166_S	221_E	0.72
237_P	197_T	0.59
YIAM_YIAO (5.0L)		
33_S	186_N	0.56
PFLA_PFLB (0.7L)		
66_F	650_T	0.92
59_E	643_K	0.67
62_T	570_N	0.54
70_S	688_G	0.27
MDTP_MDTN (2.4L)		
431_A	146_E	1.00
223_S	146_E	0.96
227_H	142_P	0.90
231_A	150_S	0.74
431_A	148_F	0.45
435_R	142_P	0.44
METI_METQ (2.0L)		
80_R	193_D	0.99
80_R	192_D	0.91
87_I	182_L	0.70
87_I	66_F	0.56
174_Q	67_N	0.50
87_I	75_A	0.43
UMUC_UMUD (1.0L)		
415_S	38_I	1.00
421_V	54_V	0.86
132_H	76_S	0.85
404_R	33_Y	0.72
183_D	127_V	0.66
78_A	34_V	0.60
IF1_SECY (1.4L)		
32_V	343_A	0.89

YOHJ_YOHK (1.9L)		
64_C	75_Y	1.00
61_N	79_E	1.00
115_S	94_I	1.00
79_G	38_M	1.00
110_V	101_V	0.99
106_V	101_V	0.99
99_S	115_A	0.97
83_M	42_I	0.96
92_Q	116_S	0.93
104_T	184_A	0.89
105_L	161_F	0.84
40_I	146_I	0.84
71_M	34_L	0.83
65_Y	79_E	0.81
113_W	169_M	0.80
16_V	14_I	0.78
74_L	65_L	0.78
82_V	42_I	0.76
101_A	157_L	0.75
39_S	70_V	0.72
115_S	176_A	0.70
106_V	105_T	0.68
115_S	179_L	0.66
100_C	150_C	0.65
18_I	72_A	0.64
109_L	169_M	0.63
110_V	97_I	0.60
110_V	213_I	0.60
CYOC_NUOK (3.8L)		
49_V	80_L	0.96
83_Y	29_L	0.68
75_L	73_A	0.68
65_E	89_N	0.65
101_W	29_L	0.61
42_I	62_Y	0.60
ATP6_ATPF (2.3L)		
74_K	34_E	1.00
149_V	10_Q	0.93
77_T	33_I	0.92
53_G	13_A	0.86
155_L	11_A	0.78
50_V	13_A	0.72
255_I	139_S	0.68
243_I	20_F	0.64
49_S	10_Q	0.64
263_Y	21_C	0.64
239_V	16_L	0.63
111_W	10_Q	0.62
FECI_FECR (3.3L)		
162_E	14_R	0.96
133_Q	57_R	0.95
158_A	18_H	0.61
FTSI_FTSW (2.0L)		
39_L	313_V	0.98
42_V	312_V	0.94
47_V	309_Y	0.80
43_A	309_Y	0.74
34_A	357_L	0.62

FLGB_FLGC (2.4L)		
34_D	13_A	1.00
34_D	107_S	0.99
110_A	117_E	0.98
98_N	11_G	0.98
121_S	129_T	0.96
34_D	103_V	0.96
113_S	108_A	0.89
27_A	114_A	0.88
117_Q	128_K	0.80
111_D	122_V	0.78
113_S	125_M	0.71
27_A	27_N	0.63
131_M	69_E	0.61
23_Q	37_P	0.60
93_P	45_K	0.60
FLIP_FLIQ (2.2L)		
185_I	84_L	0.99
54_I	55_I	0.90
229_V	77_V	0.89
203_V	24_L	0.68
205_M	66_G	0.67
213_P	55_I	0.66
243_F	72_L	0.65
149_L	67_P	0.64
188_T	80_L	0.63
FLGH_FLGI (0.6L)		
52_F	133_V	0.99
82_L	257_S	0.59
FLHB_FLIR (1.1L)		
181_A	188_L	0.98
185_A	192_T	0.64
FLIP_FLIR (1.5L)		
78_L	187_A	0.90
97_T	13_L	0.71
MLAD_MLAE (1.4L)		
14_L	13_G	0.93
14_L	24_G	0.76
8_I	6_L	0.65
58_V	6_L	0.65
65_V	219_F	0.63
CCMC_CCME (0.8L)		
49_Q	104_R	0.95
CCMA_CCMB (1.3L)		
95_E	16_R	0.87
MREB_NIFU (0.6L)		
21_N	57_R	0.97
149_I	108_I	0.72
QMCA_YBBJ (1.5L)		
145_E	115_R	0.91
36_R	113_H	0.72

PTPC1_PTPD (1.2L)		
130_K	58_K	1.00
210_L	189_I	1.00
34_I	122_C	0.98
126_D	28_R	0.98
232_G	180_L	0.93
123_T	27_E	0.86
180_H	258_S	0.85
125_A	32_G	0.85
134_T	53_L	0.72
125_A	61_L	0.66
104_S	71_V	0.64
12_L	223_I	0.60
YADG_YADH (5.4L)		
97_Q	22_R	0.94
49_G	93_E	0.93
106_Y	99_P	0.84
90_N	15_K	0.70
97_Q	18_H	0.68
82_L	97_V	0.67
104_G	14_A	0.64
RBSA_RBSC (5.6L)		
365_S	188_Y	1.00
112_K	188_Y	0.97
301_A	211_L	0.95
93_Q	204_G	0.79
371_E	197_R	0.66
105_E	198_Y	0.64
APAG_PPIC (0.8L)		
68_V	73_E	0.97
DDPB_DDPC (8.7L)		
148_W	268_G	1.00
279_L	231_T	1.00
20_G	135_I	0.99
327_Y	180_L	0.97
327_Y	177_G	0.93
233_E	292_K	0.91
221_R	283_D	0.91
229_E	212_P	0.83
28_I	147_A	0.77
222_Q	217_Q	0.75
13_G	127_A	0.73
268_L	137_L	0.67
126_R	293_A	0.65
143_S	279_N	0.65
140_T	272_L	0.63
24_I	135_I	0.62
233_E	185_Y	0.59
ENGB_NDK (1.0L)		
193_L	60_F	0.87
CLPP_CLPX (1.3L)		
96_F	271_G	0.95
40_E	267_G	0.72

APPB_APPC (1.2L)		
91_I	474_S	1.00
101_P	68_A	0.99
95_C	481_L	0.95
88_V	473_F	0.94
95_C	477_M	0.92
342_I	98_M	0.90
91_I	477_M	0.77
332_S	90_D	0.70
RNFE_RNFG (1.8L)		
188_L	160_V	0.90
85_M	27_N	0.90
67_I	12_L	0.89
146_G	171_T	0.66
NRFC_NRFD (1.2L)		
130_Y	92_S	0.93
129_Q	76_T	0.79
124_L	87_H	0.69
154_F	55_T	0.63
TOLQ_TOLR (4.9L)		
149_I	27_V	0.91
177_A	27_V	0.91
177_A	25_L	0.90
177_A	26_L	0.84
139_Y	19_V	0.83
142_L	20_P	0.77
179_A	24_V	0.76
167_V	27_V	0.76
179_A	26_L	0.73
177_A	18_I	0.69
133_V	102_N	0.67
177_A	19_V	0.67
156_L	19_V	0.66
142_L	22_L	0.66
200_N	11_D	0.63
134_G	19_V	0.59
ATKA_ATKC (0.7L)		
190_I	30_G	1.00
448_P	91_N	1.00
249_A	175_V	0.99
186_A	26_T	0.99
183_L	26_T	0.84
249_A	167_L	0.79
11_T	170_Y	0.78
198_F	24_L	0.63
238_A	83_G	0.62
244_F	55_I	0.60
FEOA_FEOB (0.9L)		
41_R	378_F	0.94
YEJA_YEJB (3.1L)		
466_D	325_R	0.90

Figure 3. Predicted residue–residue interactions across protein interfaces of unknown structure. Strongly co-evolving residue pairs for complexes without known structure that had at least one prediction with GREMLIN score greater than or equal to 0.85. Each row shows the residue pairs, their sequence identity and the GREMLIN score. Structure models for complexes highlighted in red are shown in ***Figure 5***. Full dataset is provided with the deposited data.
DOI: 10.7554/eLife.02030.007

A

Name	Complex	Chain 1	Chain 2	iRMSD	Fnat
Cytochrome c oxidase CO1/CO3	2y69AC	3dtuA	2y69C	0.4	0.99
3-oxoadipate-CoA transferase	3rrlAB	1k6dA	3rrlB	0.4	0.99
4-hydroxybenzoyl-CoA reductase	1rm6BC	1dgjA	1rm6C	0.6	0.93
Dihydroorotate dehydrogenase	1ep3AB	1dorA	1ep3B	0.9	0.89
Pyruvate dehydrogenase	1w85AB	1w85A	1ik6A	1.0	0.91
PTS-dependent dihydroxyacetone kinase	3pnlAB	3pnkA	2btdA	1.0	0.89
ABC transporter	2onkAC	2it1A	2onkC	1.4	0.88
Tryptophan synthase	1qopAB	2ekcA	1v8zA	1.6	0.83
Anthranilate synthase	1i1qAB	1k0eA	1i1qB	1.8	0.98
Succinyl-CoA synthetase	2nu9AG	1oi7A	2nu9G	1.9	0.83
Histidine kinase/response regulator	3a0rAD	3a0rA	3a10A	2.0	0.84
NADH dehydrogenase ND6/NDA	4hea6A	3ias6	4heaA	2.3	0.77
GatCAB	3ip4AC	2gi3A	3ip4C	3.2	0.78
Sulfate adenylyltransferase	1zunAB	1zunA	1jnyA	3.6	0.77
Phenylalanyl-tRNA synthetase	1b70AB	3tupA	1b70B	4.3	0.73
Thiazole synthase/sulfur carrier	1tygAB	1xm3A	1tygB	4.8	0.42
Allophanate hydrolase	3mmlAB	3mmlA	2phcB	12.1	0.72
F1-ATP synthase gamma/epsilon	3oaaHG	1bsnA	3oaaG	16.5	0.50

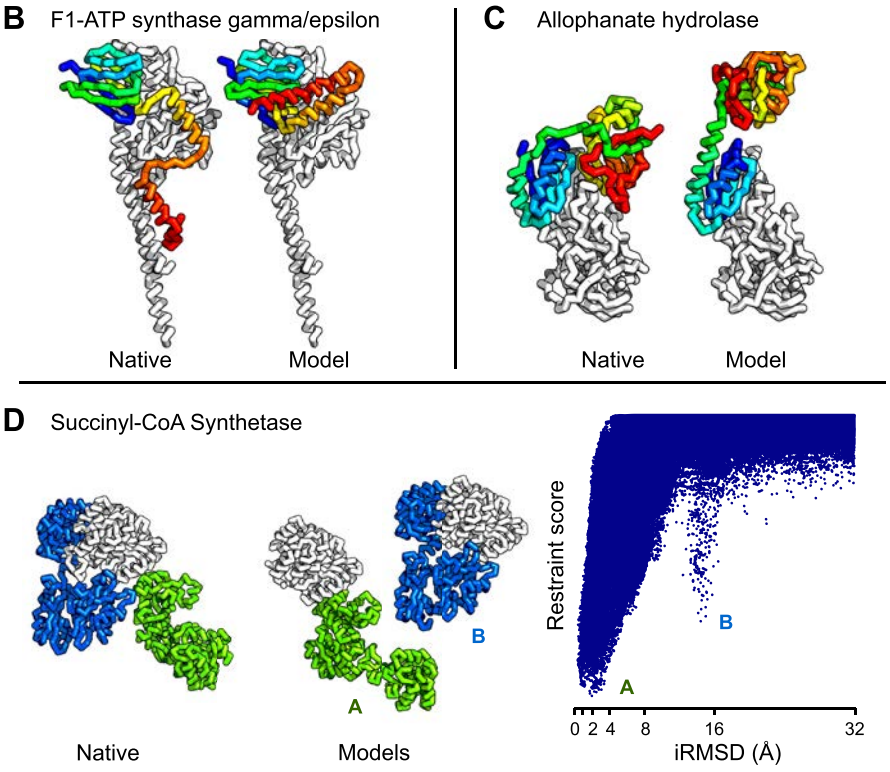

Figure 4. Contact guided protein–protein docking on a benchmark set of 18 protein complexes. (**A**) Structure models for each complex were generated by docking structures of its constituents, at least one of which (blue) was not from the structure of the complex guided by coevolution derived distance restraints. The interface C-alpha RMSD (iRMSD) of the structural model with the lowest energy to the experimentally determined structure and the fraction of native contacts are shown. Structure models for cases in red are shown in **B** and **C** and **D**. (**B** and **C**) Comparison between native and docked structure for the two largest failures in the benchmark: the large iRMSD is due to large conformational changes in the monomers upon docking but the interface is still modeled correctly in the region not involved in conformational change. (**D**) Multiple minima in the docking landscape (right) correspond to distinct interfaces in the complex (left).

DOI: 10.7554/eLife.02030.008

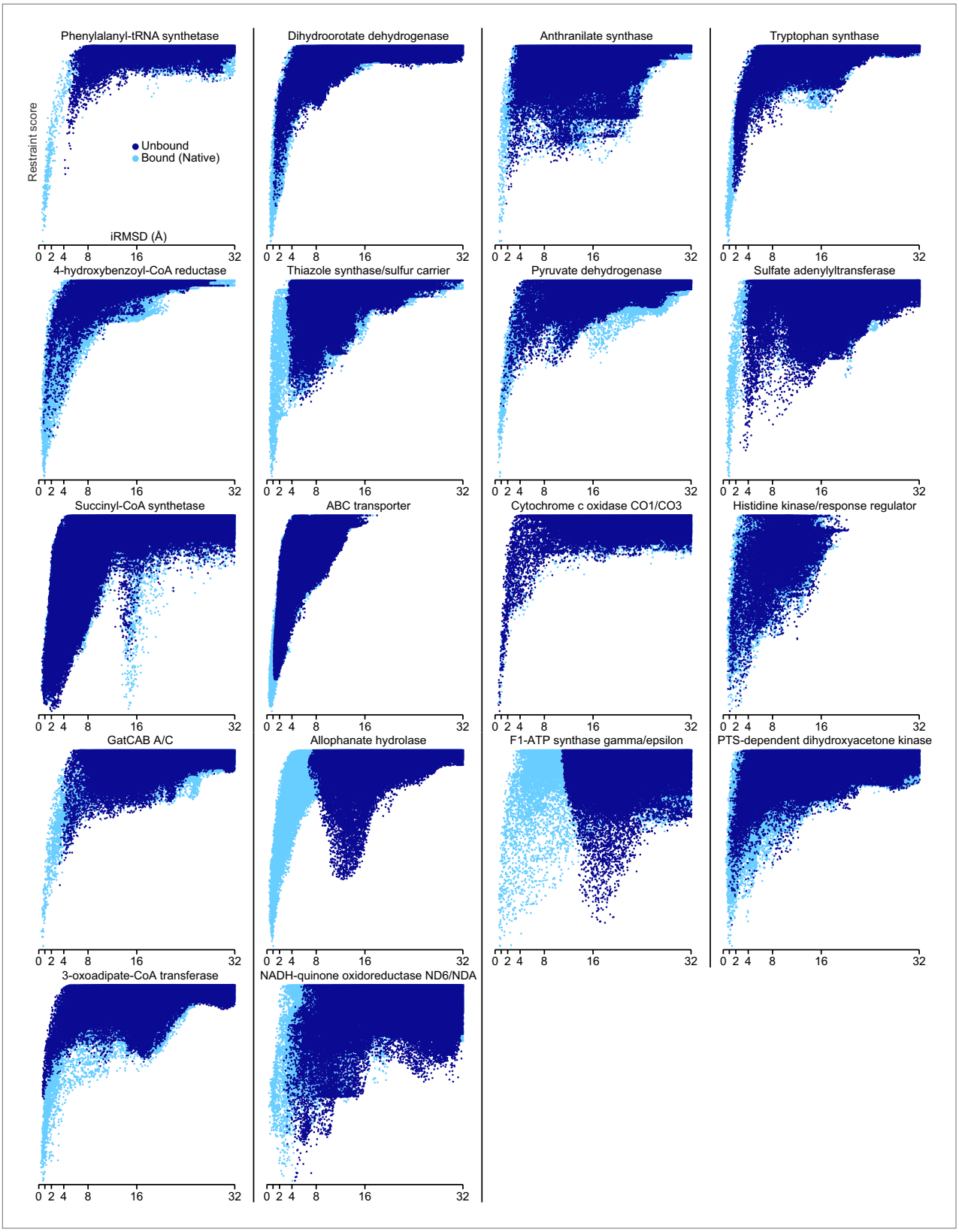

Figure 4—figure supplement 1. Docking landscapes showing iRMSD (x-axis) vs GREMLIN restraint score (y-axis).
DOI: 10.7554/eLife.02030.010

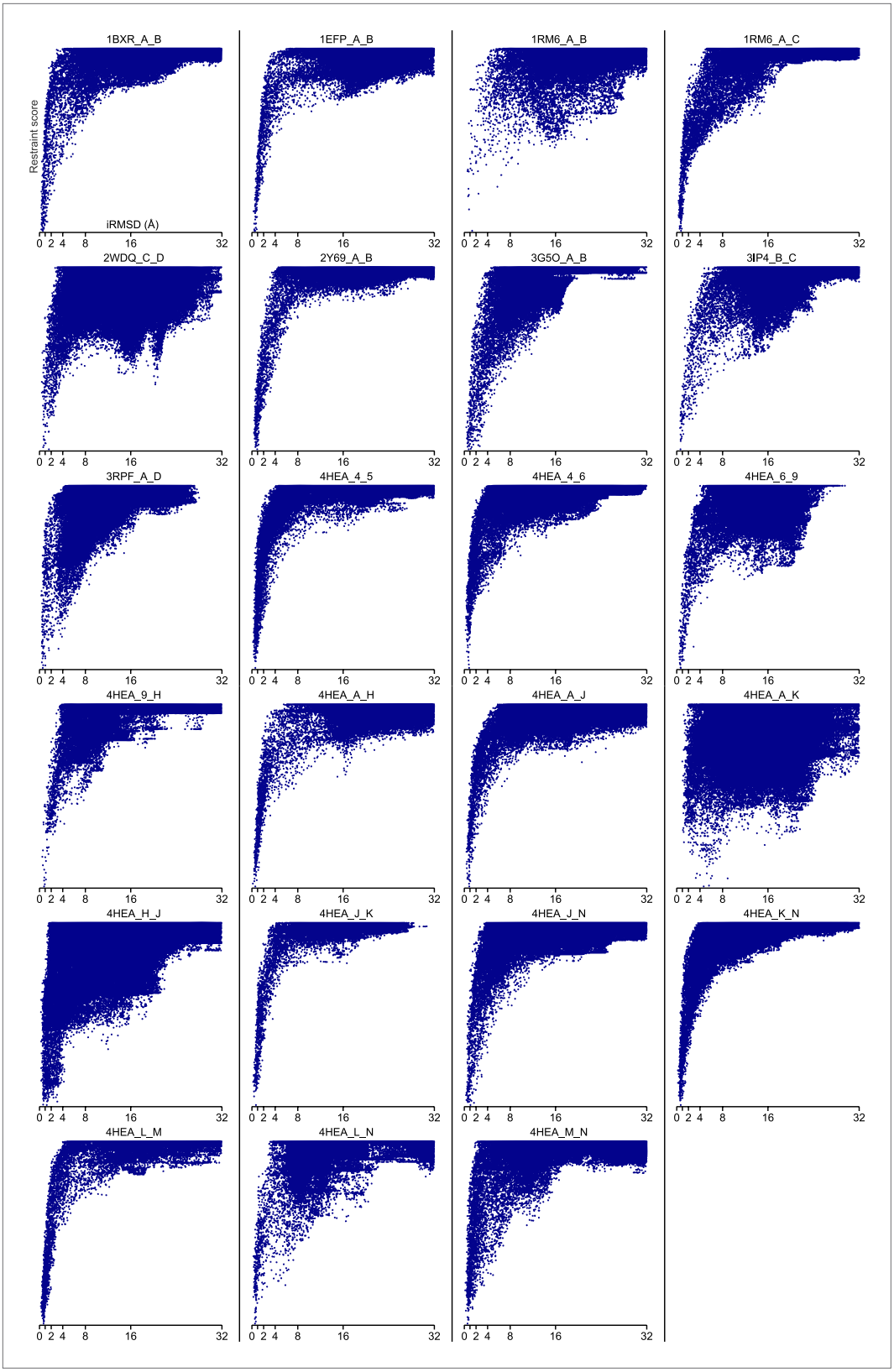

Figure 4—figure supplement 2. Bound set.
DOI: 10.7554/eLife.02030.011

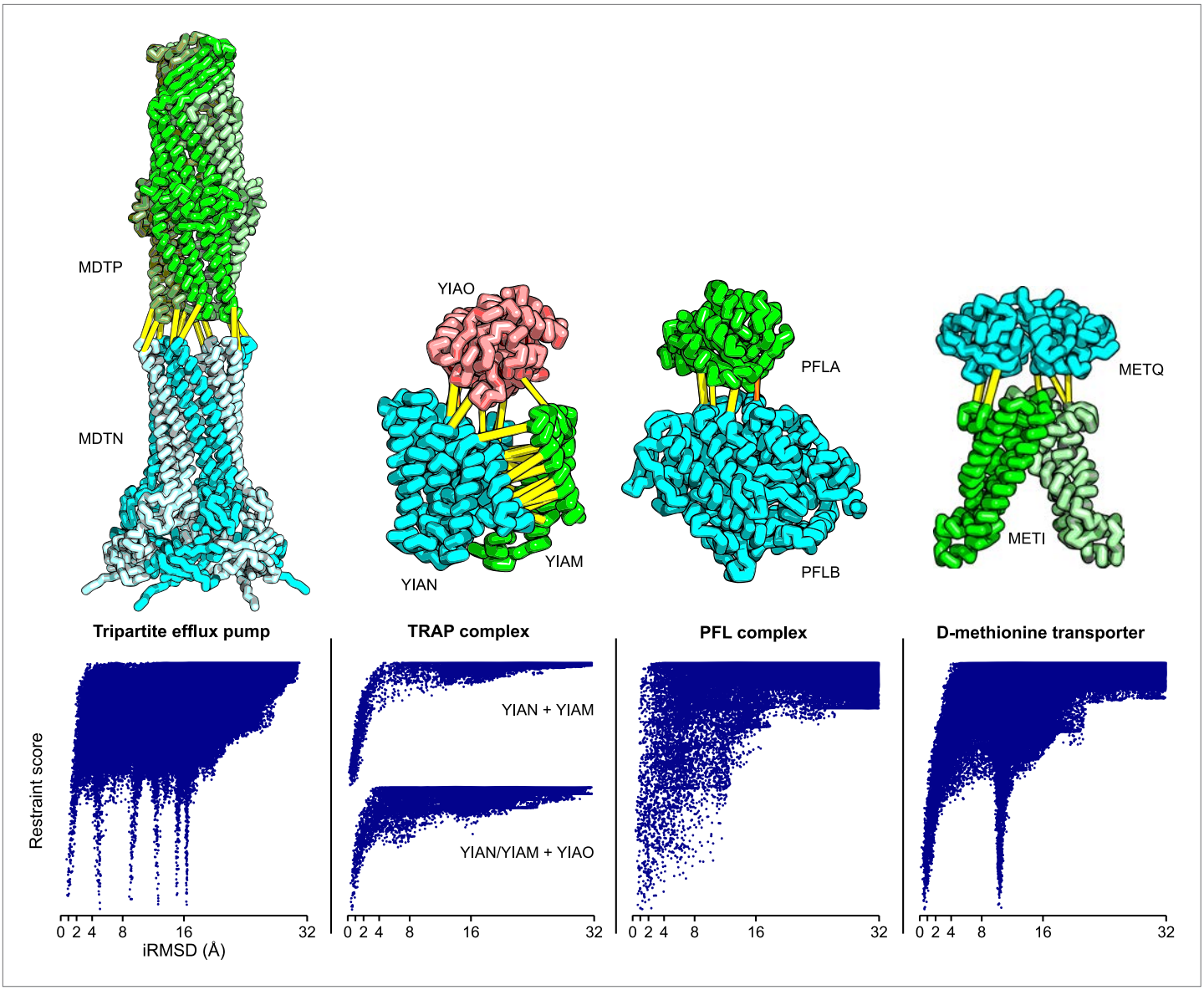

Figure 5. Structure models for complexes with unknown structures. Residue pairs with GREMLIN scores ≥ 0.60 are connected by yellow bars; the structures are pulled apart for clarity. For METQ-METI and PFLA-PFLB GREMLIN scores ≥ 0.3 are shown. For each docking calculation the docking energy landscape is shown, with iRMSD to the selected model on the x-axis. The multiple minima correspond to permutations of the labels on the subunits of the homo-oligomer complex. Predicted structures of each complex are provided with the deposited data.

DOI: 10.7554/eLife.02030.012