STRUCTURAL BIOLOGY

Volume 74 (2018)

Supporting information for article:

PRISM-EM: template interface-based modelling of multi-protein complexes guided by cryo-electron microscopy density maps. Corrigendum

Guray Kuzu, Ozlem Keskin, Ruth Nussinov and Attila Gursoy

Table S1 Scores obtained and templates used in homology modeling

Protein	PDB ID	C-score	Templates
Antibody m36	-	1.47	3ux9B, 3u2sA, 2kh2B, 1igtB, 1yc8A, 1dlfH
Acetylcholine-binding protein type 1	4aodA	1.46	3sq6A, 2qc1B, 4aq5B, 2qc1A, 4aoeA
Heavy chain of antibody VRC-PG04	3se9H	0.96	1dfbA, 3mlrH, 1y0lH, 3se8A, 4hsH, 2r56H
Light chain of antibody VRC-PG04	3se9L	0.94	3se8L
Lidless Mm-cpn	3iyfA	1.86	1q3qA, 4aolH, 1a6dA, 3j1bA
Conjugal transfer protein TrwB	1e9rA	1.53	4ag5A
Circadian clock protein KaiC	3dvlA	-1.79	2dr3D, 2ztsC, 1tf7A

Table S2 Evaluation of interfaces predicted by *Multifit*

PDB ID	PDB, interface	Model	Model, interface	IS-score	Result
3cre	AB	3cre-model	AB	0.07	incorrect
	BC		BC	0.07	incorrect
4a6j	AC	4a6j-3mer-model	AB	0.10	incorrect
			BC	0.23	correct
3se9	GH	3se9-model	GH	-	-
	HL		HL	0.12	acceptable
1e9r	AB	1e9r-model	AB	0.08	incorrect
			BC	0.08	incorrect
			CD	0.08	incorrect
			DE	0.08	incorrect
			EF	0.08	incorrect
			FA	0.08	incorrect
3dvl	AB	3dvl-model	AB	0.36	correct
			BC	0.36	correct
			CD	0.36	correct
			DE	0.36	correct
			EF	0.36	correct
			FA	0.36	correct

Table S3 Evaluation of interfaces predicted by *Segger*

PDB ID	PDB, interface	Model	Model, interface	IS-score	Result
4aod	AB	4aod-model	AB	0.08	incorrect
			BC	0.08	incorrect
			CD	0.08	incorrect
			DE	0.08	incorrect
			EA	0.08	incorrect
3iyf	AB	3iyf-model	AB	-	-
			BC	-	-
			CD	0.33	correct
			DE	0.34	correct
			EF	-	-
			FG	-	-
			GH	-	-
			HA	-	-

Table S4 Evaluation of interfaces in the *Multifit* benchmark set

For the cases 1z5sBD predicted by our method and 1z5sAD predicted by *Multifit*, the minimum interface residue number was decreased to 10 (default value: 15) in IS-score calcultaions.

PDB ID	Interface	Our predictions		*Multifit* predictions	
		IS-score	Result	IS-score	Result
1qu9	AB	0.84	correct	0.73	correct
	AC	0.84	correct	0.73	correct
	BC	0.79	correct	0.73	correct
1urz	AB	0.76	correct	0.85	correct
	AC	0.69	correct	0.80	correct
	BC	0.70	correct	0.83	correct
1z5s	AB	0.27	correct	-	
	AC	0.60	correct	0.11	incorrect
	AD	0.88	correct	0.04	incorrect
	BD	0.34	correct	-	
1e6v	AB	0.84	correct	0.15	acceptable
	AC	0.84	correct	0.23	correct
	AD	0.89	correct	0.34	correct
	AE	0.85	correct	0.09	incorrect
	BC	0.90	correct	0.34	correct
	BD	0.86	correct	0.55	correct
	BE	0.82	correct	-	
	DE	0.87	correct	0.09	incorrect
	DF	0.85	correct	-	
	EF	0.91	correct	-	
1mty	BC	0.84	correct	0.74	correct
	BD	0.92	correct	0.66	correct
	BG	0.82	correct	0.49	correct
	CE	0.89	correct	0.49	correct
	CH	0.82	correct	0.46	correct
	DE	0.85	correct	0.28	correct
	DG	0.86	correct	0.39	correct
	EH	0.89	correct	0.43	correct
1tyq	AB	0.48	correct	-	
	AD	0.84	correct	0.21	correct
	AE	0.76	correct	-	
	BF	0.86	correct	-	
	BG	0.55	correct	-	
	CF	0.77	correct	-	
	CG	0.64	correct	-	
	DF	0.87	correct	0.10	incorrect
	FG	0.72	correct	-	
1oel	AB	0.78	correct	0.51	correct

	BC	0.79	correct	0.56	correct
	CD	0.66	correct	0.52	correct
	DE	0.88	correct	0.51	correct
	EF	0.80	correct	0.54	correct
	FG	0.75	correct	0.52	correct
	GA	0.80	correct	0.51	correct
1gru	AB	0.80	correct	0.35	correct
	BC	0.81	correct	0.34	correct
	CD	0.78	correct	0.33	correct
	DE	0.79	correct	0.33	correct
	EF	0.71	correct	0.33	correct
	FG	0.75	correct	0.33	correct
	GA	0.49	correct	0.35	correct

Table S5 Predictions of HADDOCK-EM set

Protein structures, templates and Fiberdock energies of the predictions shown in Table 6 are given. In predictions using bound structures, interfaces of bound structures are also added in the template set. Complexes could not be predicted successfully using query unbound proteins were modeled using alternative structures with 100% and 95% sequence similarity (first alternative structures with 100% sequence similarity were used; if predictions were still not successful, then alternative structures with 95% sequence similarity were used). Alternative structures are the representives of proteins (identified by sequence similarity) with RMSD of less than 2 Å calculated by Multiprot

docking query proteins				
Complex	**Template**	**Input Proteins**		**Fiberdock energy**
1avx	1avwAB	1qquA	1ba7B	-56.06
2oul	3e1zAB	2nnrA	3bpfA	-83.71
1ay7	1x1yCF	1rghB	1a19B	-33.96
1ahw	1uj3BC	1fgnHL	1tfhA	-3.45
7cei	2jb0AB	1unkD	1m08B	-30.88
2oob	2denAB	2ooaA	1yj1A	-21.99
2fd6	3bt2HU	2fatHL	1ywhA	-41.86
1ak4	1m9eBC	2cplA	4j93A	-30.5
1b6c	3h9rAB	1iasA	1d6oA	-72.92
1bgx	1bgxHT	1ay1HL	1taqA	-53.59
1r6q	1lzwAB	2w9rA	1r6cX	-7.91
1m10	1u0nAD	1auq	1m0zB	-14.02
1acb	3n10AB	2cgaB	1egl	-4.92
1jk9	1pu0CD	1qupA	2jcwA	-29.9
1bkd	1nvvRS	1ctqA	2ii0A	-11.2
1jmo	1tbrHR	2cn0HL	1jmjA	> 0
docking query proteins and their alternatives				
Complex	**Template**	**Input Proteins**		**Fiberdock energy**
1avx	1avwAB	2a31A	1avwB	-69.52
2oul	3e1zAB	2nnrA	3bpfA	-83.71
1ay7	1b27AD	1lniA	3da7C	-40.49
1ahw	1uj3BC	1fgnHL	1tfhA	-3.45
7cei	2jb0AB	1unkD	1m08B	-30.88
2oob	2denAB	2ooaA	1yj1A	-21.99
2fd6	3bt2HU	2fatHL	3bt2U	-41.73
1ak4	1m9eBC	2cplA	4j93A	-30.5
1b6c	3h9rAB	1iasA	1d6oA	-72.92
1bgx	1bgxHT	1ay1HL	1taqA	-53.59
1r6q	1mbuAC	1mg9B	1mbvB	-37.95
1m10	1u0nAD	1fnsA	4c2aB	-18.45
1acb	3n10AB	2cgaB	1egl	-4.92
1jk9	1pu0CD	1qupA	2jcwA	-29.9
1acb	3rdzAC	1gl1A	4b1tB	-65.43

1bkd	1nvuRS	4nymR	4us0S	-103.86
1jmo	1tbrHR	2cn0HL	1jmjA	> 0
using bound structure data				
Complex	**Template**	**Input Proteins**		**Fiberdock energy**
1avx	1avwAB	1avxA	1avxB	-113
2oul	3e1zAB	2oulA	2oulB	-124.33
1ay7	1b27AD	1ay7A	1ay7B	-52.36
1ahw	1ahwBC	1ahwAB	1ahwC	-2.07
7cei	2jb0AB	7ceiA	7ceiB	-20.75
2oob	2oobAB	2oobA	2oobB	-18.68
2fd6	3bt2HU	2fd6HL	2fd6U	-43.8
1ak4	1m9eBC	1ak4A	1ak4D	-39.33
1b6c	3h9rAB	1b6cA	1b6cB	-86.9
1bgx	1bgxHT	1bgxHL	1bgxT	-68.11
1r6q	1r6oAC	1r6qA	1r6qC	-71.79
1m10	1u0nAD	1m10A	1m10B	-71.01
1acb	1acbEI	1acbE	1acbI	-89.94
1jk9	1pu0CD	1jk9A	1jk9B	-62.31
1bkd	1nvuRS	1bkdR	1bkdS	-94.31
1jmo	1jmoAH	1jmoHL	1jmoA	-117.94

Table S6 Evaluation of interfaces predicted by *Multifit*

PDB ID	PDB, interface	Model	Model, interface	IS-score	Result
3cre	AB	3cre-model	AB	0.07	incorrect
	BC		BC	0.07	incorrect
4a6j	AC	4a6j-3mer-model	AB	0.10	incorrect
			BC	0.23	correct
3se9	GH	3se9-model	GH	-	-
	HL		HL	0.12	acceptable
1e9r	AB	1e9r-model	AB	0.08	incorrect
			BC	0.08	incorrect
			CD	0.08	incorrect
			DE	0.08	incorrect
			EF	0.08	incorrect
			FA	0.08	incorrect
3dvl	AB	3dvl-model	AB	0.36	correct
			BC	0.36	correct
			CD	0.36	correct
			DE	0.36	correct
			EF	0.36	correct
			FA	0.36	correct

Table S7 Evaluation of interfaces predicted by *Segger*

PDB ID	PDB, interface	Model	Model, interface	IS-score	Result
4aod	AB	4aod-model	AB	0.08	incorrect
			BC	0.08	incorrect
			CD	0.08	incorrect
			DE	0.08	incorrect
			EA	0.08	incorrect
3iyf	AB	3iyf-model	AB	-	-
			BC	-	-
			CD	0.33	correct
			DE	0.34	correct
			EF	-	-
			FG	-	-
			GH	-	-
			HA	-	-

Table S8 Evaluation of interfaces in the *Multifit* benchmark set

For the cases 1z5sBD predicted by our method and 1z5sAD predicted by *Multifit*, the minimum interface residue number was decreased to 10 (default value: 15) in IS-score calcultations.

PDB ID	Interface	Our predictions		*Multifit* predictions	
		IS-score	Result	IS-score	Result
1qu9	AB	0.84	correct	0.73	correct
	AC	0.84	correct	0.73	correct
	BC	0.79	correct	0.73	correct
1urz	AB	0.76	correct	0.85	correct
	AC	0.69	correct	0.80	correct
	BC	0.70	correct	0.83	correct
1z5s	AB	0.27	correct	-	
	AC	0.60	correct	0.11	incorrect
	AD	0.88	correct	0.04	incorrect
	BD	0.34	correct	-	
1e6v	AB	0.84	correct	0.15	acceptable
	AC	0.84	correct	0.23	correct
	AD	0.89	correct	0.34	correct
	AE	0.85	correct	0.09	incorrect
	BC	0.90	correct	0.34	correct
	BD	0.86	correct	0.55	correct
	BE	0.82	correct	-	
	DE	0.87	correct	0.09	incorrect
	DF	0.85	correct	-	
	EF	0.91	correct	-	
1mty	BC	0.84	correct	0.74	correct
	BD	0.92	correct	0.66	correct
	BG	0.82	correct	0.49	correct
	CE	0.89	correct	0.49	correct
	CH	0.82	correct	0.46	correct
	DE	0.85	correct	0.28	correct
	DG	0.86	correct	0.39	correct
	EH	0.89	correct	0.43	correct
1tyq	AB	0.48	correct	-	
	AD	0.84	correct	0.21	correct
	AE	0.76	correct	-	
	BF	0.86	correct	-	
	BG	0.55	correct	-	
	CF	0.77	correct	-	
	CG	0.64	correct	-	
	DF	0.87	correct	0.10	incorrect
	FG	0.72	correct	-	
1oel	AB	0.78	correct	0.51	correct

	BC	0.79	correct	0.56	correct
	CD	0.66	correct	0.52	correct
	DE	0.88	correct	0.51	correct
	EF	0.80	correct	0.54	correct
	FG	0.75	correct	0.52	correct
	GA	0.80	correct	0.51	correct
1gru	AB	0.80	correct	0.35	correct
	BC	0.81	correct	0.34	correct
	CD	0.78	correct	0.33	correct
	DE	0.79	correct	0.33	correct
	EF	0.71	correct	0.33	correct
	FG	0.75	correct	0.33	correct
	GA	0.49	correct	0.35	correct

Table S9 Correlations calculated for each subunit in the *Multifit* set

Protein	**Multifit**	**Our method**	**Protein**	**Multifit**	**Our method**
7catA	0.90	0.89	1mtyG	0.88	0.89
7catB	0.90	0.89	1mtyH	0.89	0.89
1gteA	0.91	0.88	1tyqA	0.87	0.93
1gteB	0.90	0.89	1tyqB	0.88	0.92
1qu9A	0.91	0.92	1tyqC	0.81	0.95
1qu9B	0.91	0.92	1tyqD	0.85	0.92
1qu9C	0.91	0.91	1tyqE	0.89	0.94
1urzA	0.92	0.92	1tyqF	0.82	0.90
1urzB	0.92	0.92	1tyqG	0.66	0.89
1urzC	0.92	0.92	1oelA	0.94	0.94
1z5sA	0.92	0.93	1oelB	0.94	0.94
1z5sB	0.71	0.89	1oelC	0.94	0.94
1z5sC	0.93	0.94	1oelD	0.93	0.94
1z5sD	0.63	0.79	1oelE	0.94	0.94
1e6vA	0.90	0.90	1oelF	0.94	0.94
1e6vB	0.92	0.91	1oelG	0.93	0.94
1e6vC	0.91	0.85	1gruA	0.88	0.87
1e6vD	0.92	0.90	1gruB	0.88	0.87
1e6vE	0.89	0.90	1gruC	0.88	0.87
1e6vF	0.91	0.85	1gruD	0.88	0.87
1mtyB	0.92	0.92	1gruE	0.88	0.87
1mtyC	0.92	0.92	1gruF	0.88	0.87
1mtyD	0.94	0.94	1gurG	0.88	0.87
1mtyE	0.94	0.94			

Table S10 RMSD calculated for each subunit in the *Multifit* set

Protein	*Multifit*	Our method	Protein	*Multifit*	Our method
7catA	0.00	0.00	1mtyG	3.57	0.91
7catB	5.86	0.74	1mtyH	5.15	1.92
1gteA	0.00	0.79	1tyqA	0.00	0.02
1gteB	5.36	0.04	1tyqB	6.70	2.65
1qu9A	0.00	0.00	1tyqC	6.99	2.66
1qu9B	1.86	0.89	1tyqD	5.23	0.68
1qu9C	1.67	0.83	1tyqE	56.30	1.02
1urzA	0.00	0.00	1tyqF	31.88	2.05
1urzB	1.81	1.31	1tyqG	78.82	2.48
1urzC	1.75	1.62	1oelA	0.00	0.00
1z5sA	0.00	0.02	1oelB	2.60	0.80
1z5sB	44.17	3.60	1oelC	3.16	1.59
1z5sC	20.86	1.79	1oelD	3.67	2.47
1z5sD	44.85	1.12	1oelE	3.71	2.42
1e6vA	0.00	0.19	1oelF	3.41	2.06
1e6vB	6.82	1.09	1oelG	2.43	1.44
1e6vC	4.79	1.21	1gruA	0.00	0.04
1e6vD	5.09	0.93	1gruB	3.35	0.85
1e6vE	14.20	1.11	1gruC	4.93	1.85
1e6vF	72.83	1.64	1gruD	6.31	1.96
1mtyB	3.08	0.76	1gruE	6.63	1.89
1mtyC	4.38	1.20	1gruF	5.88	1.61
1mtyD	0.00	0.00	1gurG	4.13	1.96
1mtyE	3.33	1.43			

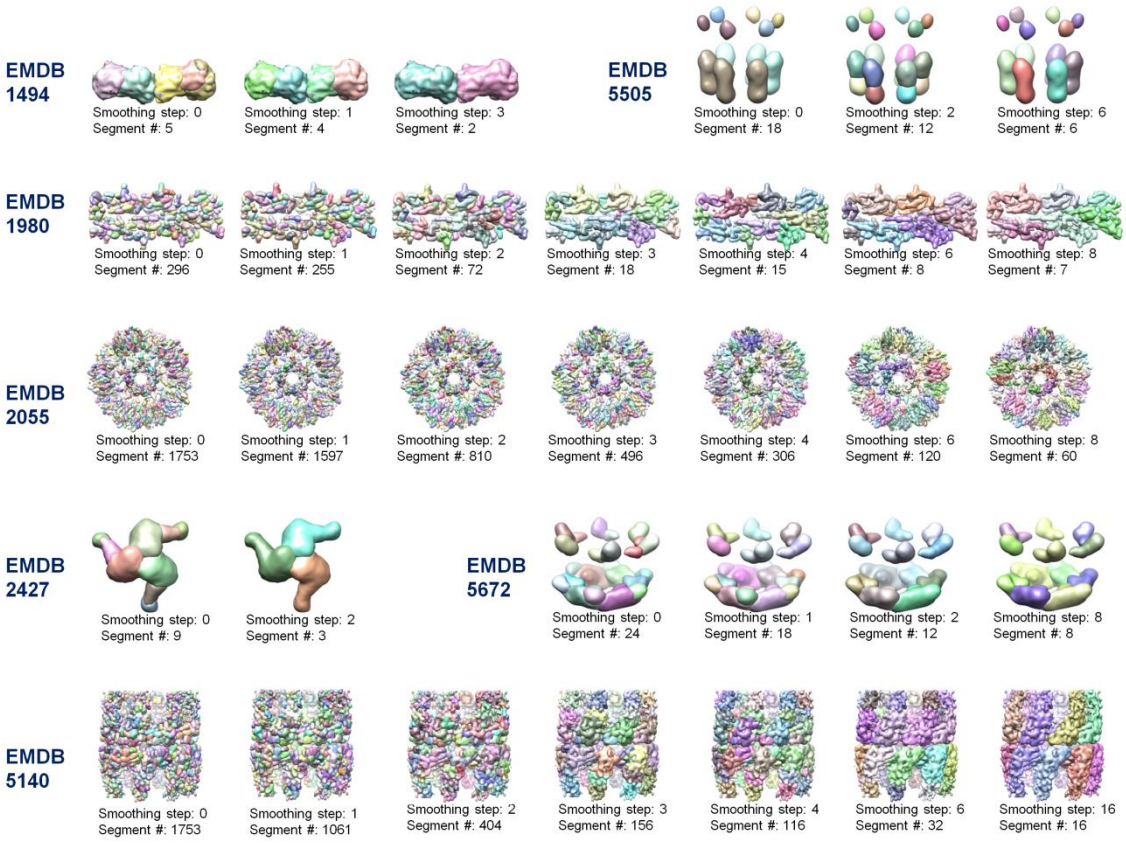

Figure S1 Results of segmentation by *Segger*. *Segger* gives different segmentation for different smoothing step number. Smoothing step and segment numbers are given below each figure. Figures obtained with different smoothing step number but giving the same result are not shown here; for example, smoothing step 2 gave the same result as smoothing step 1. We could obtain proper segmentation for EMDB IDs 2055 and 5140. We could obtain desired number of segments for EMDB IDs 1980, 5505 and 5672, but segmentation was not properly. Using different contour level of density maps did not change the segmentation.

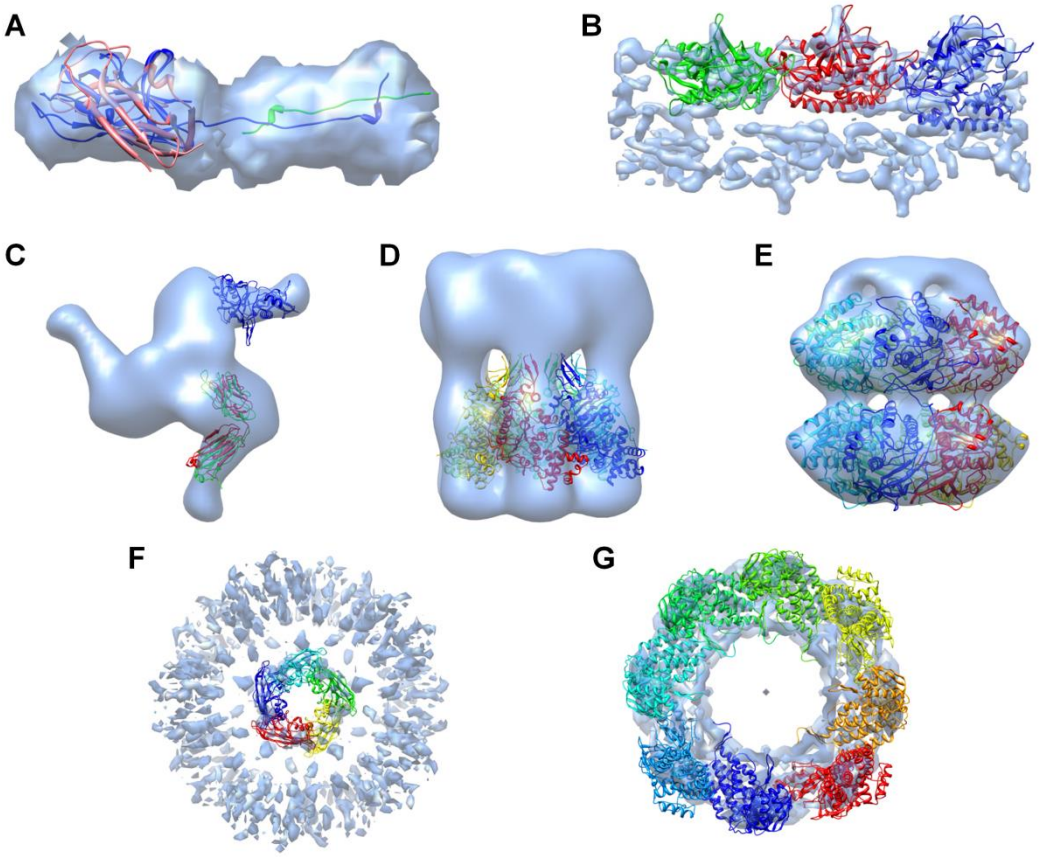

Figure S2 Models obtained by using *Multifit* and *Segger*. Models are fitted into EM density maps. Chains are given in different colors. (A) *Multifit* model of Saf Pilus fitted into EMDB ID 1494. (B) *Multifit* model of ParM filament fitted into EMDB ID 1980. (C) *Multifit* model of HIV-1 gp120/VRC-PG04 complex fitted into EMDB ID 2427. (D) *Multifit* model of conjugal transfer protein TrwB fitted into EMDB ID 5505. (E) *Multifit* model of circadian clock protein KaiC fitted into EMDB ID 5672. (F) *Segger* model of acetylcholine-binding protein type 1 fitted into EMDB ID 2055. (G) *Segger* model of lidless Mm-cpn fitted into EMDB ID 5140.