

SUPPLEMENTARY INFORMATION

Assessment of the CASP14 Assembly Predictions

Short title: CASP14 Assembly Assessment

Burcu Ozden^{1,2}, Andriy Kryshchak³, Ezgi Karaca^{1,2}

¹ Izmir Biomedicine and Genome Center, 35340, Izmir, Turkey

² Izmir International Biomedicine and Genome Institute, Dokuz Eylul University, 35340, Izmir, Turkey

³Protein Structure Prediction Center, Genome and Biomedical Sciences Facilities, University of California, Davis, California, USA

Correspondence

Ezgi Karaca, email: ezgi.karaca@ibg.edu.tr

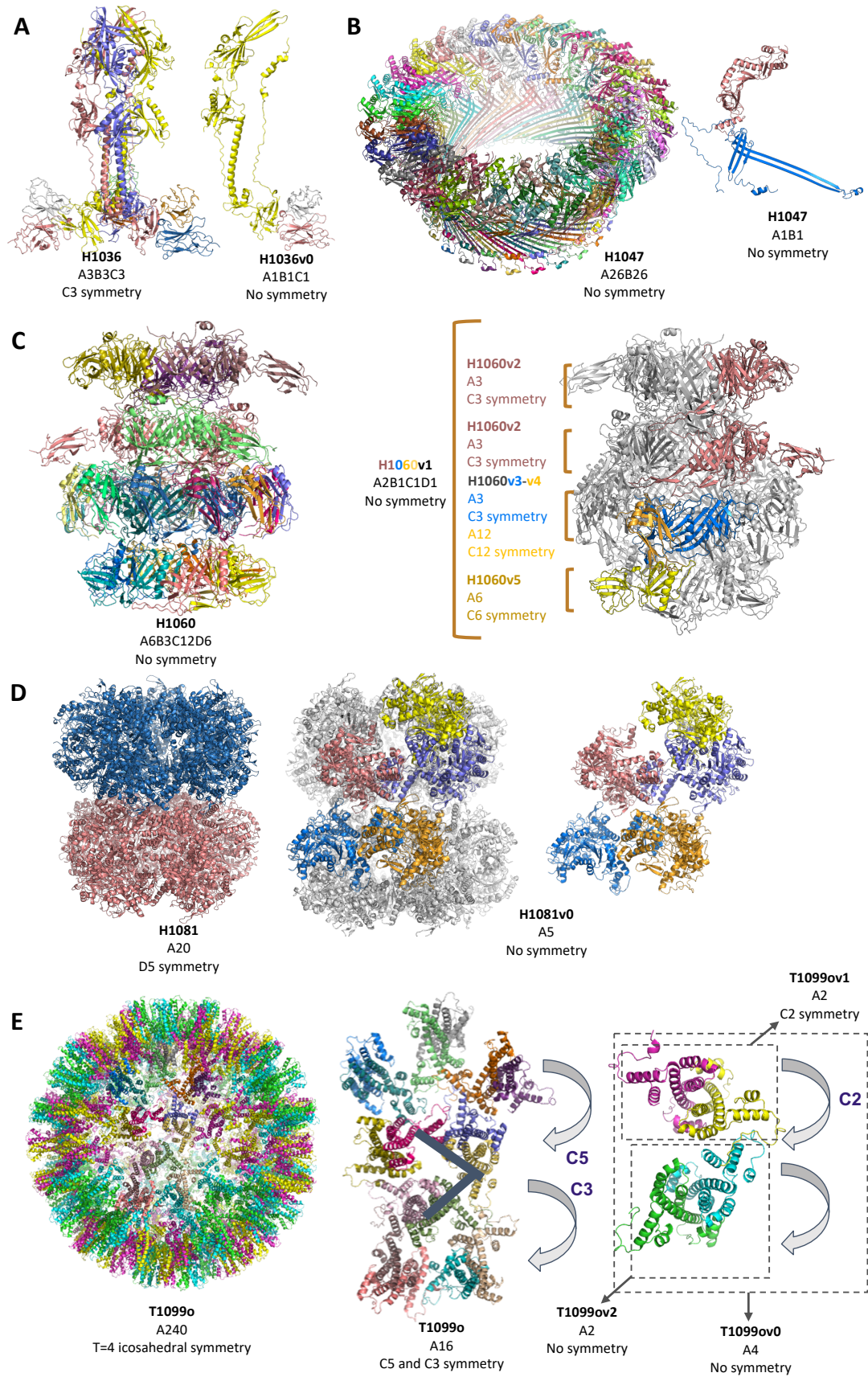


Figure S1. The targets which were assessed in evaluation units are presented with their global stoichiometry and symmetry information. **(A)** H1036 is the monoclonal antibody

bound to the virus with an A3B3C3 stoichiometry. Its evaluation unit H1036v0 has A1B1C1 stoichiometry, where the viral chain is colored with yellow in the evaluation unit. **(B)** H1047 is a A26B26 bacterial flagellar complex. It was assessed as dimer. The stoichiometry of the whole complex was not provided to the predictors. **(C)** H1060 is the bacteriophage T5 tail complex, which is composed of five stacked rings, made of four different monomers. These four rings separated as evaluation units H1060v2-v5. The vertical slice of each ring forms the H1060v1 evaluation unit. **(D)** H1081 is an arginine decarboxylase with A20 stoichiometry, composed of two stacked decamers, arranged in D5 symmetry. So, H1080v0 evaluation unit contains five monomers, which includes all the unique interfaces, repeating four times within the complex. **(E)** T1099o is the viral capsid of the duck hepatitis B virus with a T=4 icosahedral symmetry, composed of 240 subunits. The minimum subunit T1099v0 includes a C5 and C3 symmetry interface. It is separated into two dimers to cover all interfaces. T1099v1 has C2 symmetry, while T1099v2 does not have any symmetry.

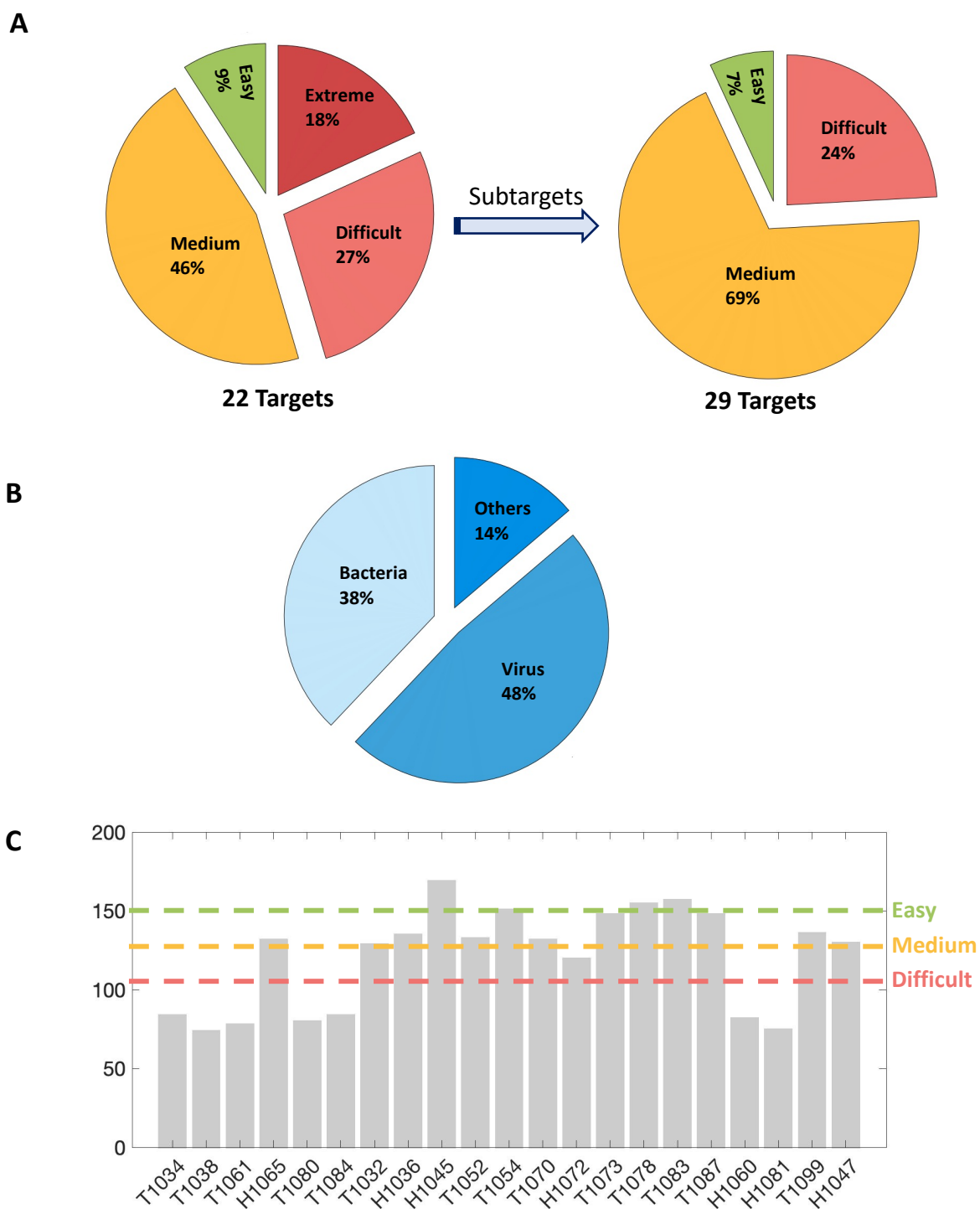


Figure S2. (A) The statistics of unique complexes (left) and the updated target classification, including the evaluation units (right). The targets, which were too large to assess with standard CASP metrics are labeled as extreme. **(B)** The taxonomic distributions of 29 targets (the complete set, including the evaluation units). **(C)** The number of models submitted for a given target. The average models submitted regarding each target difficulty are marked with dashed lines (green: easy, orange: medium, red: difficult)

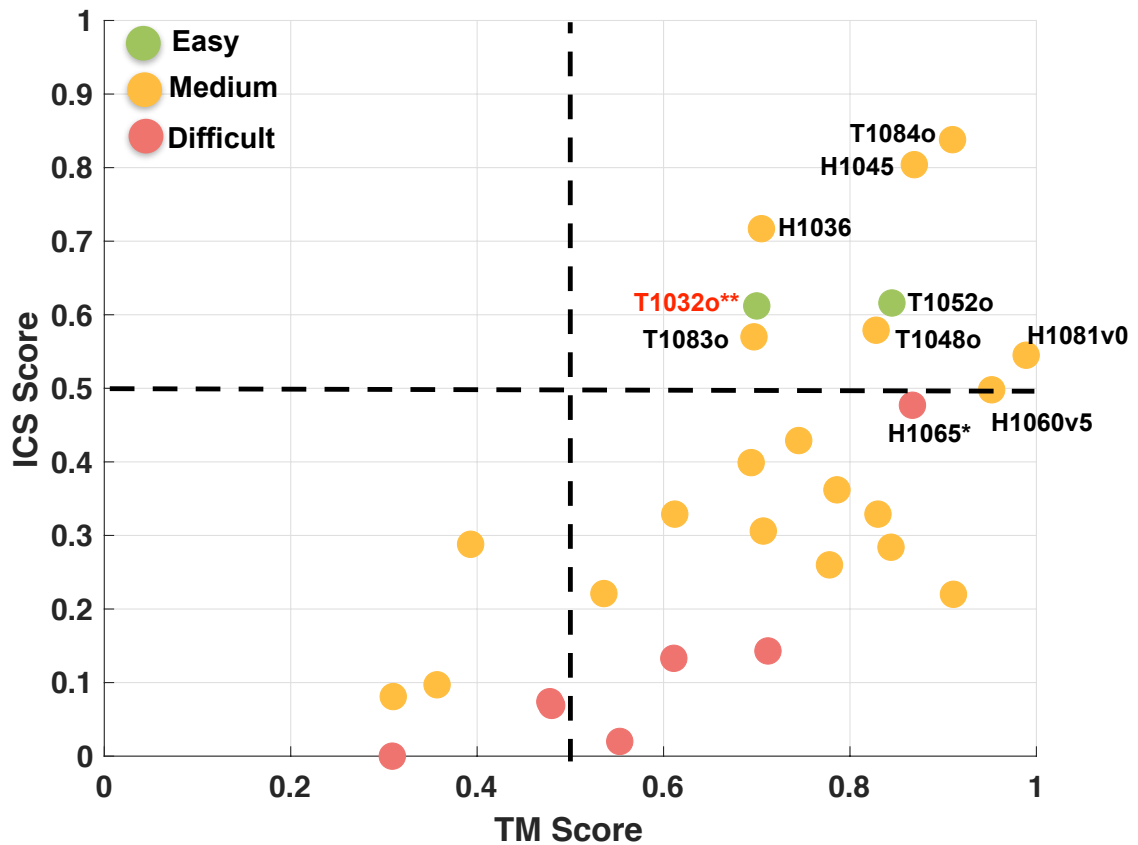


Figure S3. The ICS vs. TM distributions of the best models with respect to the target difficulty (green: easy, orange: medium, red: difficult). The successful target ids are denoted in black. T1032** has its ICS, IPS and TM scores higher than 0.5. Though, the IDDT of this target is 0.41. Therefore, this model was considered unsuccessful. H1065* was categorized as successful, since, while being a difficult target, its ICS score (0.48) can be rounded to 0.5.

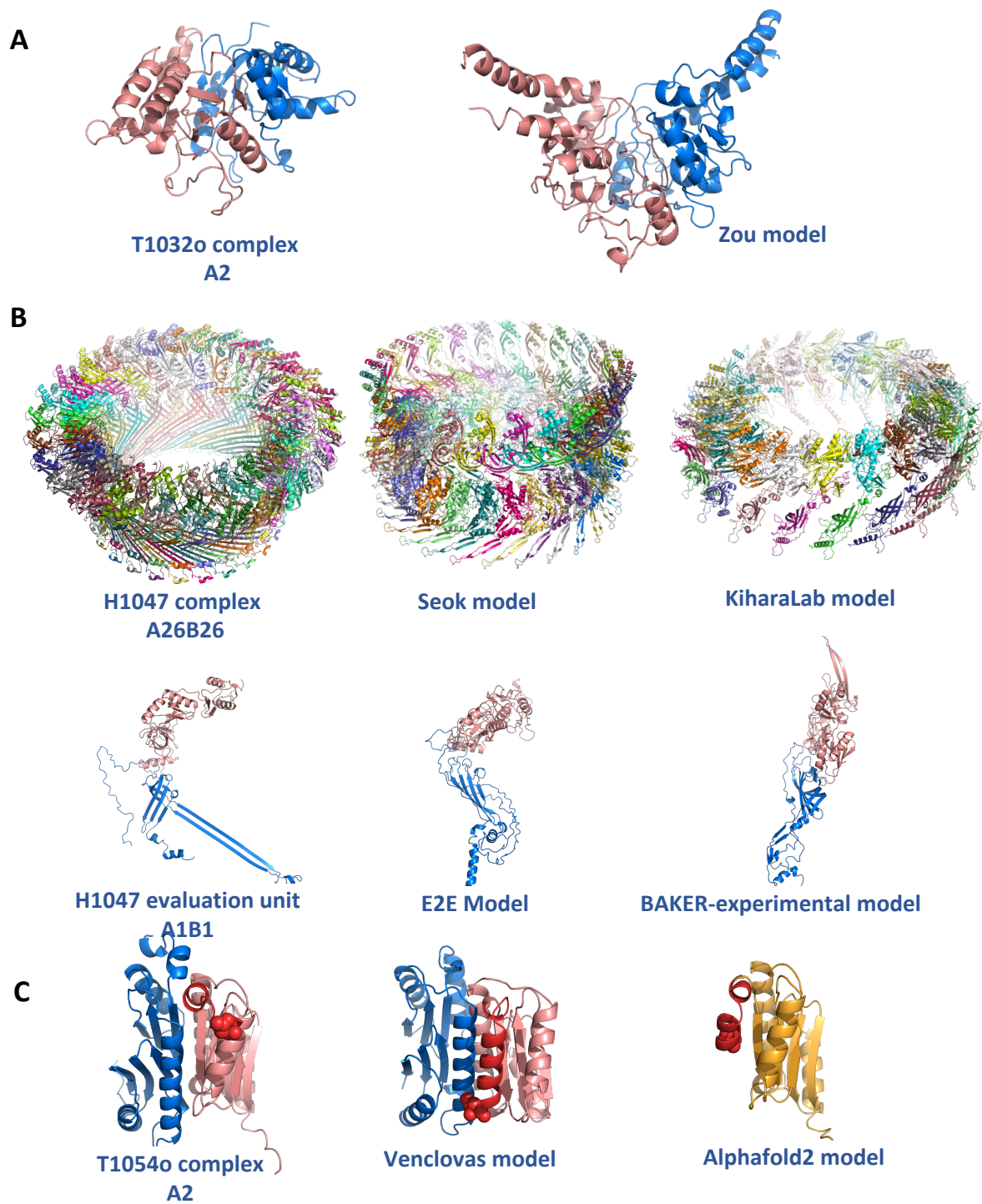


Figure S4. (A) T1032o target and the Zou model, having the best ICS score. (B) First row: H1047 complex and the assemblies submitted with correct stoichiometries (by Seok and KiharaLab). Second row: H1047 evaluation unit as dimer and the models with best TM scores, from E2E and BAKER-experimental. (C) The T1054o complex, together with the best assembly model submitted by Venclovas and the best tertiary structure prediction submitted by AlphaFold2. The first 20 amino acids, which blocks the actual interactions surface is shown in red spheres.

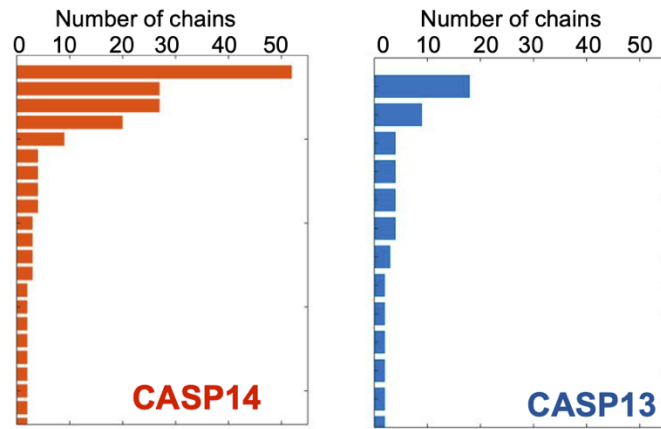


Figure S5. The frequencies of number of subunits present in the CASP14 and CASP13 assembly targets.

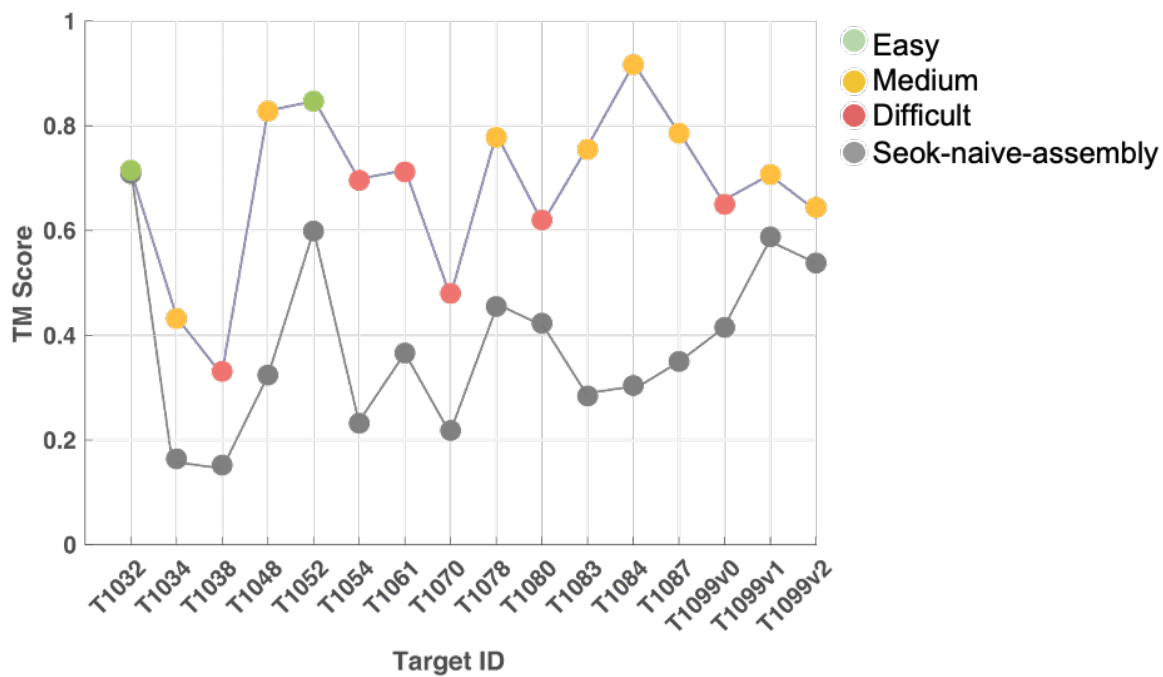


Figure S6. The TM scores of best homomeric models submitted compared with the TM scores of Seok-naïve-assembly models. The Seok-naïve-assembly are represented in gray circles, the rest are colored according to the difficulty level of the target (easy: green, medium: orange, difficult: red).

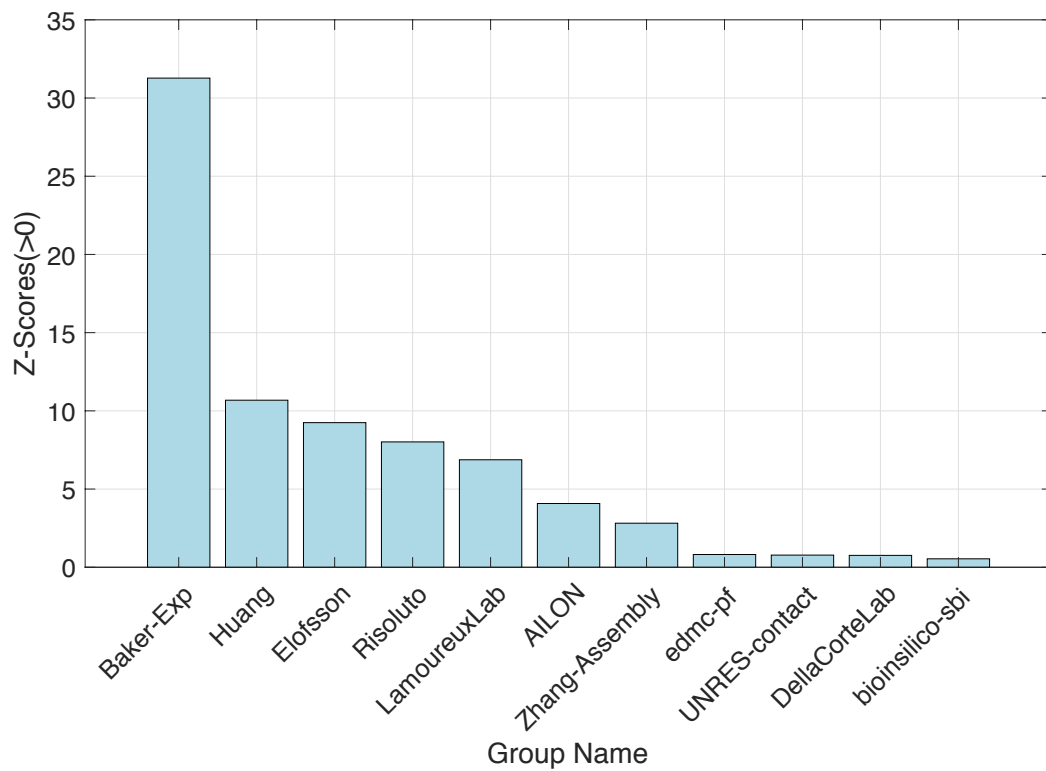


Figure S7. Z-scores-based ranking of the groups, which indicated the use of contact predictions in their CASP14 abstracts.

Table S1. The ranking of CASP groups, according to their best submitted models over 29 evaluation units (as published on the CASP14 server). ** denotes a missing target for that particular group, as the related submitted models could not pass the CASP validation system.

No	Group name	Group no	Number of participated targets	Score (Sum Z>0)	Ranking
1	BAKER-experimental	403	29	33.2402	1
2	Venclovas	29	29	25.9313	2
3	Takeda-Shitaka-Lab	55	29	22.0163	3
4	Seok	193	28**	18.535	4
5	DATE	288	25	17.1605	5
6	Kiharalab	71	28	14.0879	6
7	Bates_BMM	336	29	12.9151	7
8	Huang	298	23	11.8882	8
9	Elofsson	472	22	9.7202	9
10	Ornate-select	375	23	9.3433	10
11	VoroCNN-select	343	28	9.1835	11
12	Risoluto	341	29	8.559	12
13	Vakser	173	25	8.4447	13
14	Seok-assembly	451	24	7.9951	14
15	SBROD-select	62	24	7.5175	15
16	LamoureuxLab	66	22	6.9113	16
17	UNRES-template	18	22	6.3225	17
18	AILON	192	12	4.4694	18
19	McGuffin	220	13	4.24	19
20	Zhang-Assembly	182	11	3.4158	20
21	E2E	376	2	3.1912	21
22	Seok-naive_assembly	491	16	3.0485	22
23	UNRES	360	21	2.5229	23
24	ricardo	467	6	1.8131	24
25	UNRES-contact	96	13	1.6878	25
26	DellaCorteLab	323	13	1.3089	26
27	edmc_pf	387	1	0.8579	27
28	bioinsilico_sbi	196	6	0.8031	28
29	DELCLAB	369	16	0.7464	29
30	xianmingpan	14	5	0.587	30
31	CAO-QA1	217	7	0.5066	31
32	ropius0	254	3	0.4208	32
33	Pharmulator	340	7	0.2172	33
34	Spider	349	6	0.2058	34
35	ict-ams	476	3	0.1943	35
36	MULTICOM-CLUSTER	75	1	0.0521	36
37	Jones-UCL	304	1	0	37
38	BAKER-ROBETTA	278	1	0	37
39	FoldEM	107	1	0	37

Table S2. The ranking of CASP and CAPRI groups, according to their best submitted models over 19 evaluation units (as published on the CASP14 server). The CAPRI groups are marked with *. ** denotes a missing target for that particular group, as the related submitted models could not pass the CASP validation system.

No	Group name	Group no	Number of participated targets	Score (Sum Z>0)	Ranking
1	BAKER-experimental	403	19	18.5925	1
2	Venclovas	29	19	17.3622	2
3	Takeda-Shitaka-Lab	55	19	16.2411	3
4	Seok	193	18**	15.4096	4
5	Zou	177*	19	13.9369	5
6	Kiharalab	71	19	11.7287	6
7	DATE	288	15	9.5991	7
8	CoDock	477*	13**	9.1612	8
9	CAPRI-Shen	103*	10	8.1517	9
10	PierceLab	279*	10	8.124	10
11	Bates_BMM	336	19	7.4713	11
12	Huang	298	13	7.2065	12
13	Elofsson	472	16	7.118	13
14	Fernandez-Recio	099*	13	6.9312	14
15	Seok-assembly	451	14	6.4848	15
16	Vakser	173	15	6.0925	16
17	Ornate-select	375	13	5.856	17
18	Risoluto	341	19	5.5354	18
19	Kozakov-Vajda	221*	12	5.4261	19
20	Kiharalab_Assembly	285*	18	5.18	20
21	LamoureuxLab	66	14	5.1204	21
22	SBROD-select	62	14	4.7462	22
23	VoroCNN-select	343	18	4.6556	23
24	UNRES-template	18	19	4.3384	24
25	CLUSPRO	155*	13	3.4028	25
26	htjcadd	191*	7	3.3123	26
27	McGuffin	220	7	2.9804	27
28	Zhang-Assembly	182	7	2.5437	28
29	Seok-naive_assembly	491	10	1.9448	29
30	UNRES	360	18	1.7575	30
31	bio3d	418*	2	1.6869	31
32	AILON	192	8	1.4871	32
33	ricardo	467	4	1.1947	33
34	UNRES-contact	96	10	1.0485	34
35	MULTICOM-AI	275*	6	0.8746	35
36	edmc_pf	387	1	0.8579	36
37	bioinsilico_sbi	196	6	0.8031	37
38	DELCLAB	369	9	0.6383	38
39	DellaCorteLab	323	9	0.5806	39
40	CAO-QA1	217	3	0.4249	40
41	ropius0	254	3	0.4208	41
42	ict-ams	476	2	0.1943	42
43	xianmingpan	14	2	0.0558	43
44	MULTICOM-CLUSTER	75	1	0.0521	44
45	Pharmulator	340	3	0	45
46	FoldEM	107	1	0	45
47	Spider	349	3	0	45