Supplementary Material

Sense and Simplicity in HADDOCK Scoring: Lessons from CASP-CAPRI

Vangone A., Rodrigues J.P.G.L.M., Xue L., van Zundert G.C.P., Geng C., Kurkcuoglu Z., Nellen M., Narasimhan S., Karaca E., van Dijk M., Melquiond A.S.J., Visscher K., Trellet M., Kastritis P.L., Bonvin A.M.J.J.*

Figure S1. Scatter plots of overall backbone RMSDs from the top-scored model versus HADDOCK score for the 25 target of the CASP-CAPRI round. The HADDOCK scores have been shifted to positive values and are reported in a log scale. Green check marks and dots in the plots indicate the targets for which HADDOCK selected acceptable or better quality models; all the unsuccessful targets are flagged with a red cross.

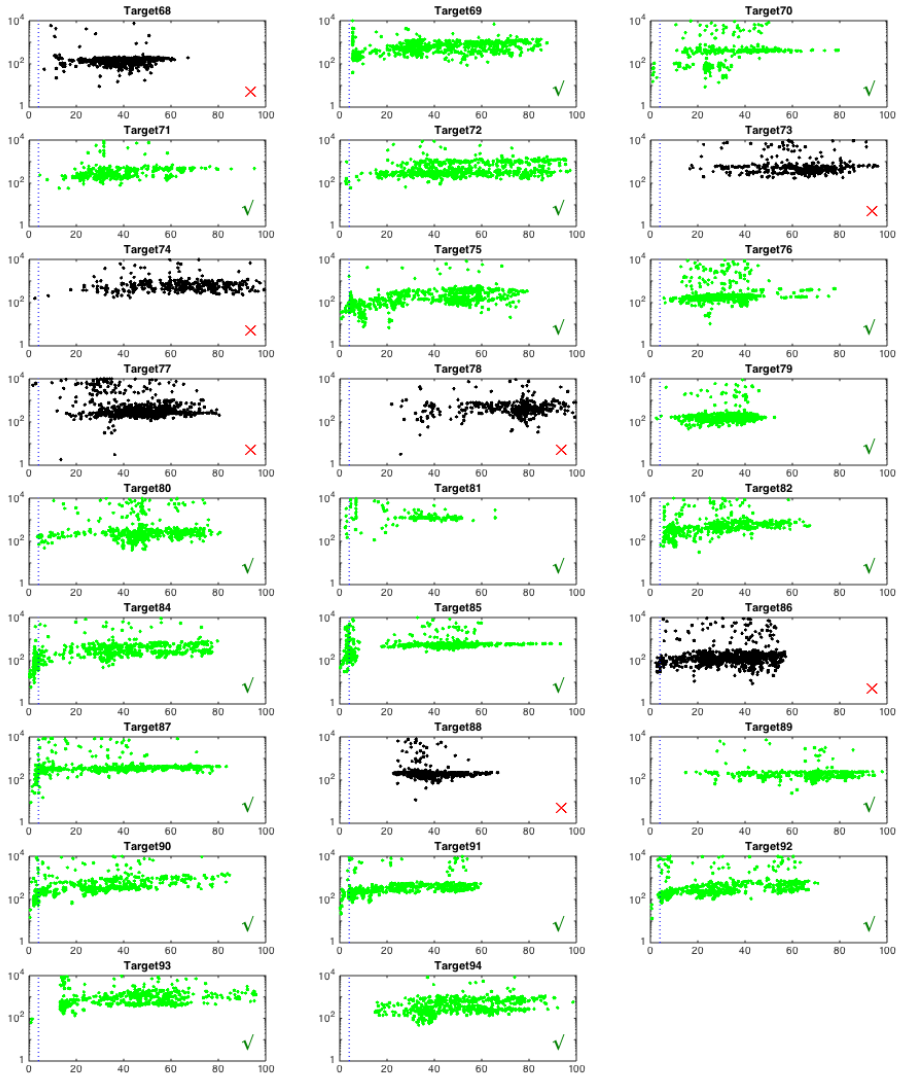

Figure S2. Comparison of th i-RMSD of the best model predicted by HADDOCK (dark grey bars) and overall backbone RMSD of the unbound homology model used as input structure for docking (light gray bars) for the unsuccessful (left) and successful (right) HADDOCK predictions. The dashed line indicates the CAPRI 4.0 Å acceptable i-RMSD cut-off. For heterodimers targets (T81 and T89), the backbone-RMSD of the worst model among the two unbound modeled proteins is reported.

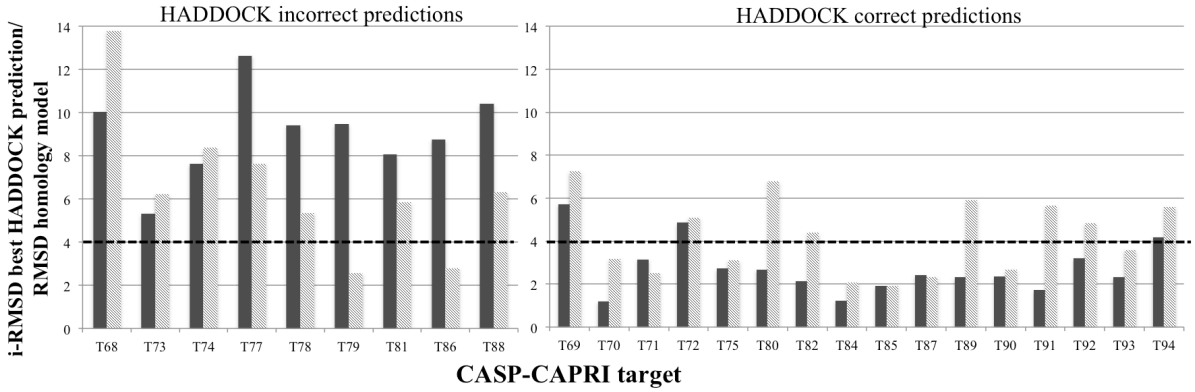

Table S1. Modeling and prediction approach used by HADDOCK during CASP-CAPRI server prediction

Most of our predictions started from a model built with either MODELLER (v9.12) [1] using HHpred [2] for template search, or I-TASSER [3], or an ensemble of models generated using both methods. The target/template alignment was optimized, if necessary, either with NEEDLE [4] in case of high sequence identity, or using manually-curated multiple sequence alignments. In one particular case, T77, we created a model of the monomer using a flexible multibody docking approach [5] due to the disposition of the individual domains. Typically, an ensemble of the ten best monomer models was used as starting point for HADDOCK. Since all targets were homo-oligomeric structures, non-crystallography symmetry and regular symmetry restraints were used [5] in combination with center-of-mass restraints (ab initio mode of HADDOCK). In cases where homologous complexes were identified, Ca-Ca distance restraints were applied, derived either manually or determined by PS-HOMPPI [6]. A number of targets were also directly modeled as a multimer in MODELLER and only refined in explicit solvent in HADDOCK. The models for submission were distributed over the clusters ranked by HADDOCK. A summary of the modelling methodology and restraints used is provided in the table.

Target ID CAPRI	CASP	Alignment	Template	Modelling	Docking	Restraints
T68	T0759	HHpred	1LM5, 1LM7	Modeller	HADDOCK	Ca-Ca + Symmetry
T69	T0764	Hhpred + Needle	1QLW	Modeller	HADDOCK	Ca-Ca + Symmetry
T70	T0765	-	-	I-Tasser	HADDOCK	Symmetry + CM
T71	T0768	HHpred	4GT6, 2Z80	Modeller	HADDOCK	Symmetry + CM
T72	T0770	MSAProbs	3EJN, 3FDH, 3MX3	Modeller	HADDOCK	Symmetry + CM
T73	T0772	HHpred	3hbk	Modeller	HADDOCK	Symmetry + CM
T74	T0774	HHpred	4k4k	Modeller	HADDOCK	Symmetry + CM
T75	T0776	Hhpred + Needle	4jhl	Modeller	HADDOCK	Symmetry + AIRs (Homology)
T77	T0780	HHpred	2kq1	Modeller + HADDOCK	HADDOCK	Symmetry + CM
T78	T0781	HHpred	1gup	Modeller	HADDOCK	Symmetry + CM
T79	T0782	HHpred	3rco	Modeller	HADDOCK	Symmetry + CM
T80	T0786	Hhpred + Needle	1mdo	Modeller	HADDOCK (Refinement)	Symmetry + Ca-Ca (PS-HOMPPI)
T81	T0797-T0798	Hhpred + Needle	2gzd	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T82	T0801	HHpred	3eo8 / 4dn2	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T84	T0811	HHpred	1b9b	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T85	T0813	HHpred	3ggg	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T86	T0815	HHpred	4mjd	Modeller	HADDOCK	Symmetry + CM
T87	T0819	HHpred	3get	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry + Ca-Ca (PS-HOMPPI)
T88	T0825	HHpred + Needle	2ymu	As above. Mutated	HADDOCK	Symmetry + CM
T89	T0840					
	T0841	HHpred + Needle	1shy / 4fww	Modeller	HADDOCK	Symmetry + CM
T90	T0843	HHpred	4cl3	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T91	T0847	HHpred	1bry	Modeller/I-Tasser	HADDOCK	Symmetry + Ca-Ca (PS-HOMPPI)
T92	T0849	Delta-Blast	1lbk	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T93	T0851	HHpred	2y0c	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry
T94	T0852	HHpred	3gff	Modeller (Multimer)	HADDOCK (Refinement)	Symmetry

Table S2.
For each target, i-RMSD, l-RMSD and Fnat values are reported for the best model submitted by HADDOCK and the best model of the remaining of the CAPRI community. Cases for which HADDOCK submitted the best model in terms of i-RMSD, l-RMSD or fnat are highlighted in bold.

Target	HADDOCK			CAPRI community		
	i-RMSD	l-RMSD	fnat	i-RMSD	l-RMSD	fnat
T69.1	2,78	2,40	0,62	2,47	2,18	0,65
T70.2	1,29	**3,19**	0,64	1,20	4,51	0,75
T71.1	4,82	19,88	0,13	3,12	14,29	0,30
T71.3	3,33	8,82	0,33	2,30	7,55	0,63
T72.1	5,19	8,55	**0,63**	4,27	7,29	0,47
T73.1	6,55	16,99	0,26	4,75	14,33	0,29
T73.2	5,52	14,45	**0,75**	4,33	9,79	0,50
T75.1	2,28	4,76	0,49	1,65	3,48	0,55
T79.1	**2,29**	4,43	**0,75**	2,29	3,84	0,63
T79.2	2,36	**4,12**	**0,84**	2,12	4,16	0,80
T79.3	10,64	21,29	0,07	4,66	8,24	0,26
T80.1	1,57	2,68	0,76	1,20	1,49	0,81
T81.1	3,01	7,42	0,31	1,60	2,72	0,79
T82.1	2,18	2,61	0,60	2,03	2,45	0,73
T84.1	1,37	3,09	**0,87**	1,21	2,16	0,86
T85.1	1,96	3,30	0,75	1,21	1,52	0,79
T86.1	5,60	18,03	**0,57**	2,59	5,28	0,57
T86.2	7,11	22,22	0,16	1,04	4,71	1,00
T87.1	2,44	2,68	0,63	1,87	2,40	0,65
T89.1	1,97	4,89	0,68	1,28	4,66	0,79
T90.1	2,35	4,41	0,61	2,01	2,59	0,75
T91.1	1,59	4,54	0,70	1,46	3,40	0,80
T92.1	2,85	3,39	0,50	2,58	3,08	0,52
T93.1	2,02	3,52	0,56	1,50	2,10	0,74
T94.1	3,69	9,72	0,68	3,53	7,46	0,71

References

1. A. Sali and T. L. Blundell. “Comparative protein modelling by satisfaction of spatial restraints”. In: *J. Mol. Biol.* 234.3 (1993),pp. 779–815.

2. A. Hildebrand et al. “Fast and accurate automatic structure prediction with HHpred”. In: *Proteins* 77 Suppl 9 (2009),pp. 128–132.

3. Roy, A. Kucukural, and Y. Zhang. “I-TASSER: a unified platform for automated protein structure and function predic- tion”. In: *Nat Protoc* 5.4 (2010), pp. 725–738.

4. H. McWilliam et al. “Analysis Tool Web Services from the EMBL-EBI”. In: *Nucleic Acids Res.* 41.Web Server issue (2013), pp. 597–600.

5. E. Karaca et al. “Building macromolecular assemblies by information-driven docking: introducing the HADDOCK multi- body docking server”. In: *Mol. Cell Proteomics* 9.8 (2010), pp. 1784–1794.

6. L. C. Xue, D. Dobbs, and V. Honavar. “HomPPI: a class of se- quence homology based protein-protein interface prediction methods”. In: *BMC Bioinformatics* 12 (2011), p. 244.