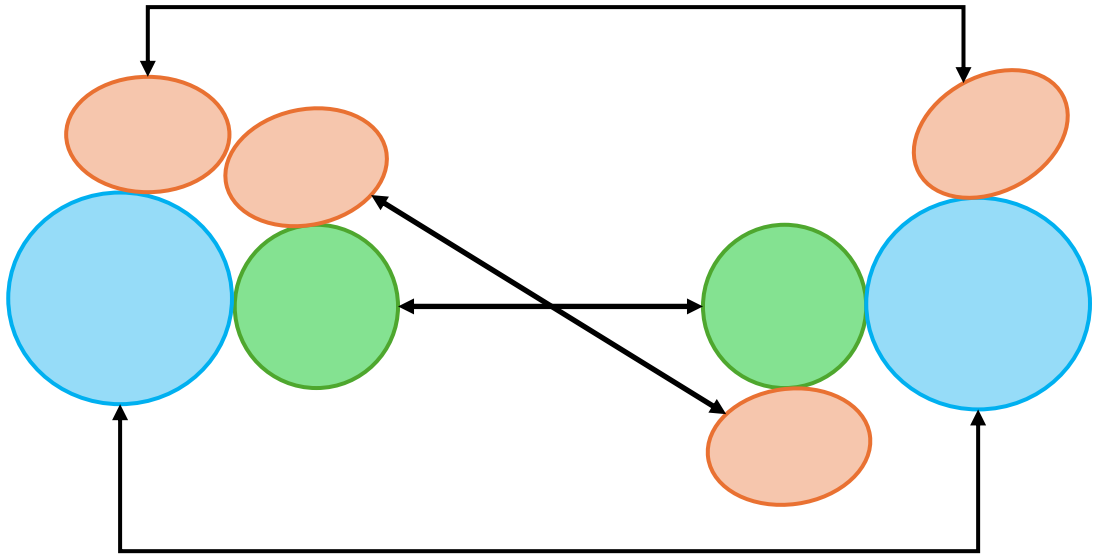
Analysis and evaluation of CASP16 protein oligomer predictions

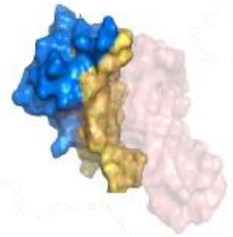
Rongqing Yuan, Jing Zhang, Andriy Kryshtafovych, Gabriel Studer, Nick V. Grishin, and Qian Cong

University of Texas Southwestern Medical Center

The classic oligomer scoring routine in CASP



Interface/Local:



DockQ: interface size-weighted average

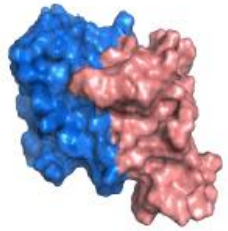
$$\mathrm{IPS}(M,T) = J_C(M,T) = \frac{|M_{\text{i-Res}} \cap T_{\text{i-Res}}|}{|M_{\text{i-Res}} \cup T_{\text{i-Res}}|}$$

Assembly/Global:

$$\mathrm{ICS}(M,T) = 2 \cdot \frac{P(M_{\text{cnt}},T_{\text{cnt}}) \times R(M_{\text{cnt}},T_{\text{cnt}})}{P(M_{\text{cnt}},T_{\text{cnt}}) + R(M_{\text{cnt}},T_{\text{cnt}})}$$

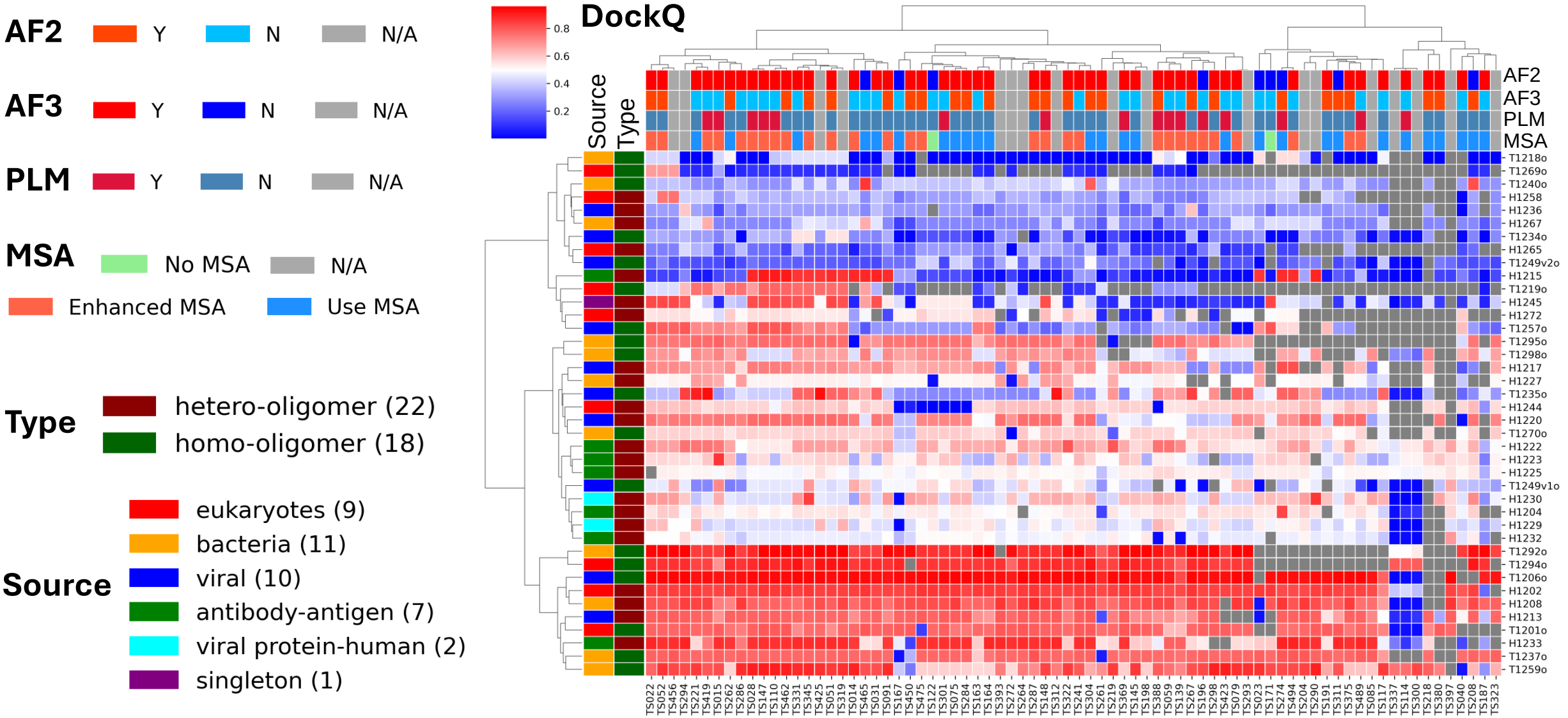
lDDT

TM score



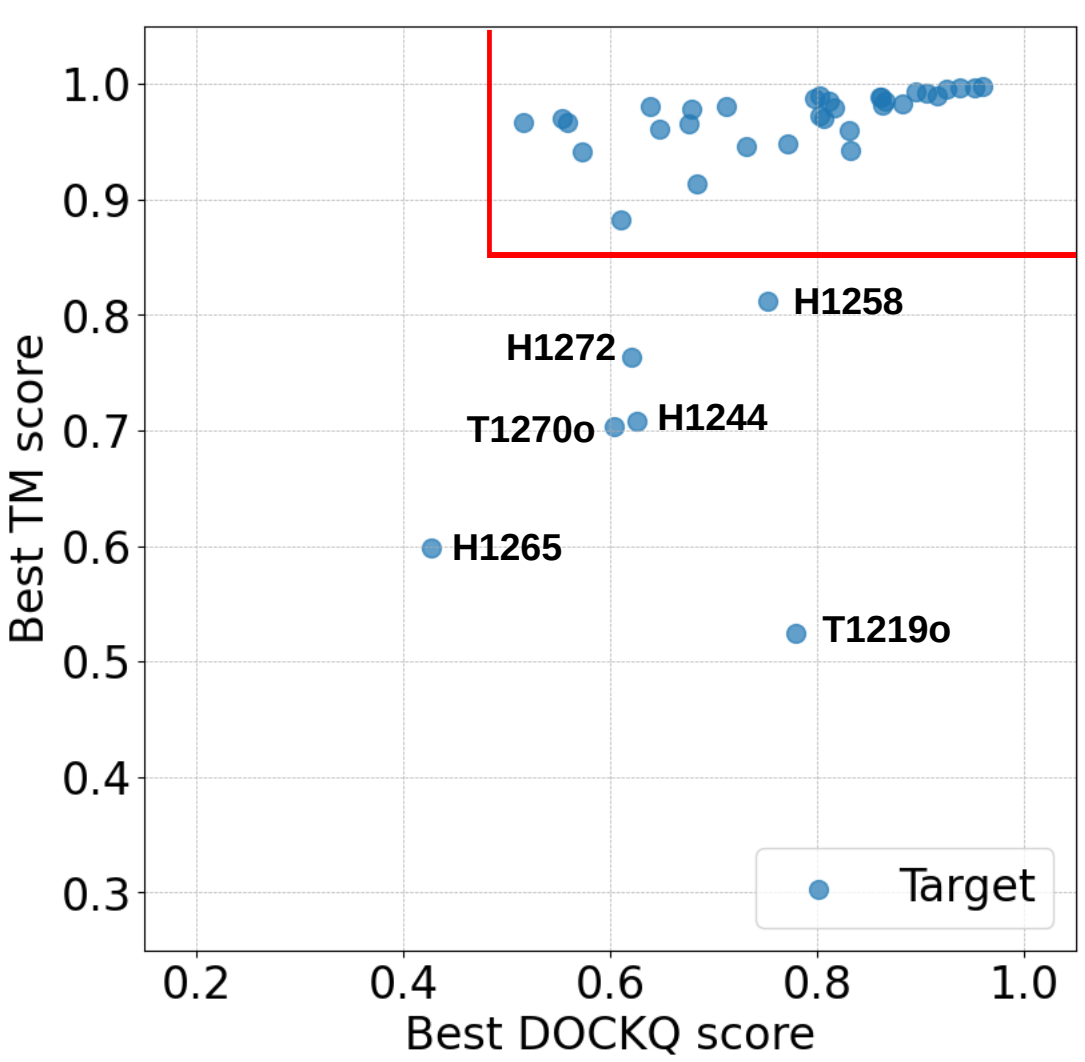
$$\text{QS-best}(M,T) = \frac{|M_{\text{cnt}} \cap T_{\text{cnt}}|}{\max(M_{\text{cnt}},T_{\text{cnt}})}$$

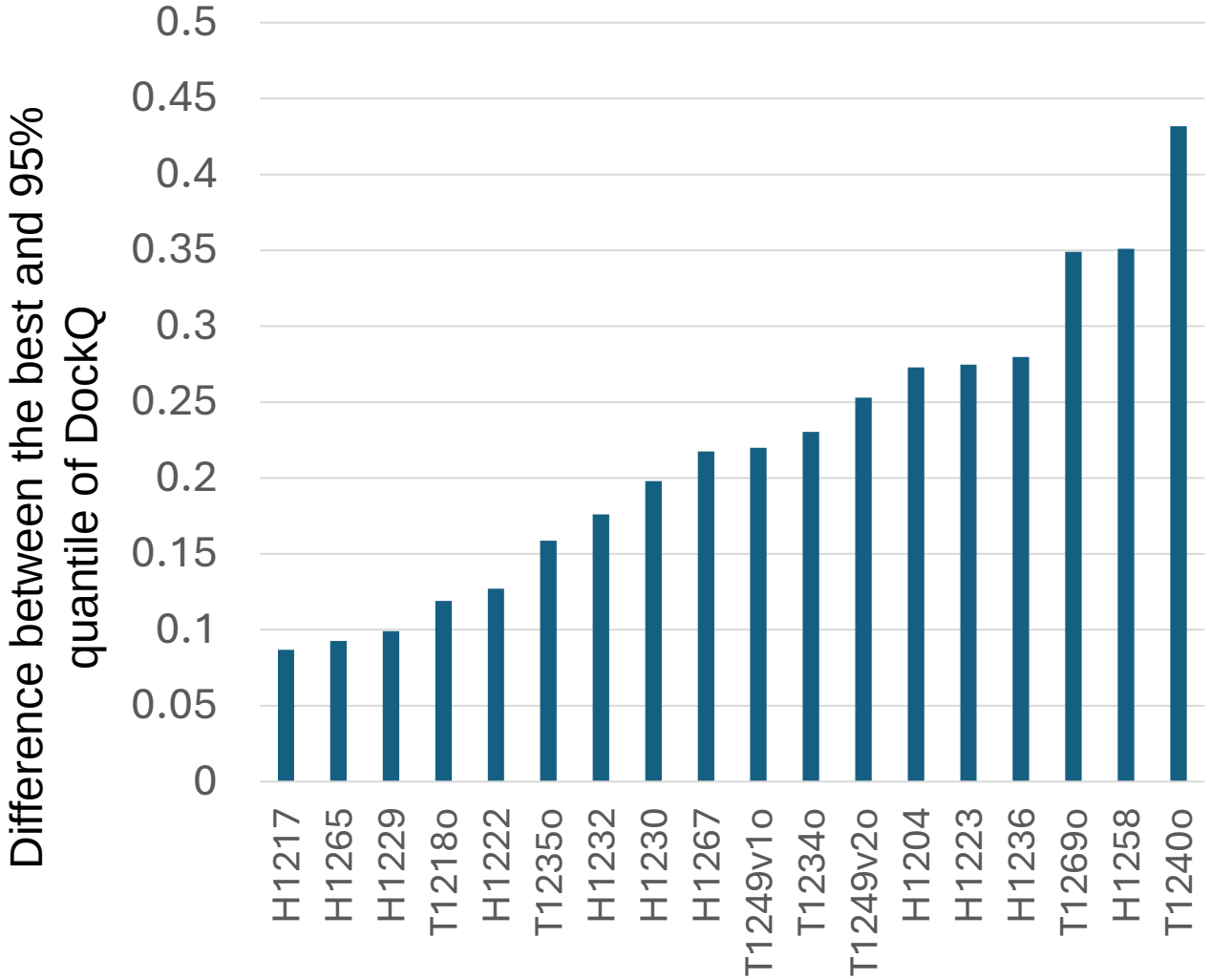
An overview of oligomer targets and groups in CASP16



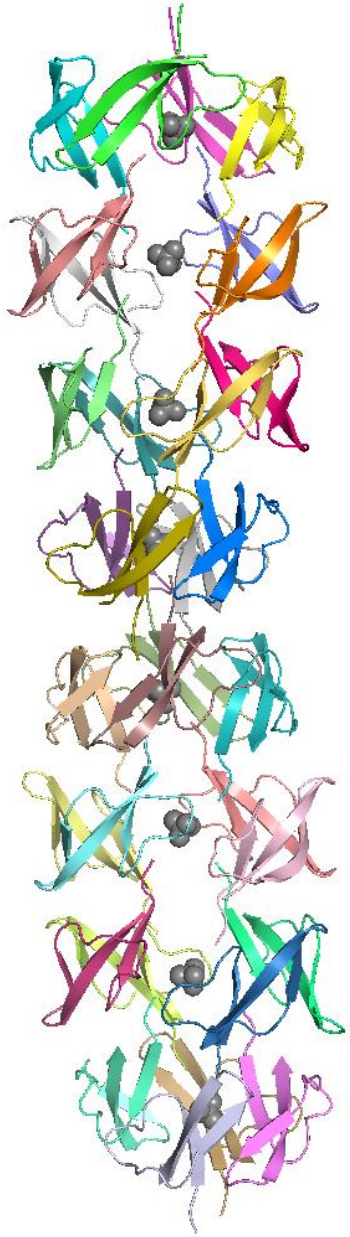
What targets are still challenging for the community?

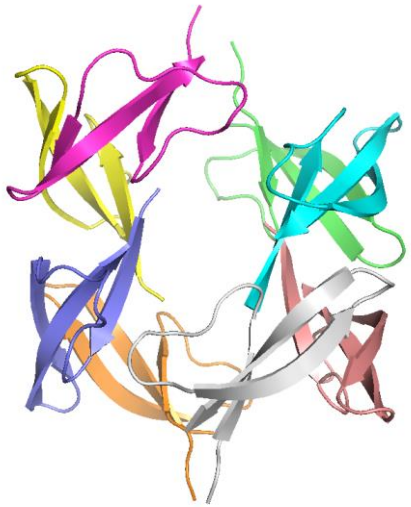
What targets are still challenging for the community?



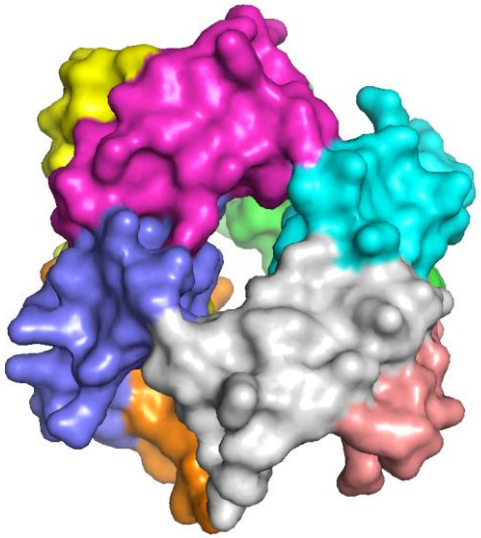


Challenge 1: filament maintained by weak interactions





T1219o target

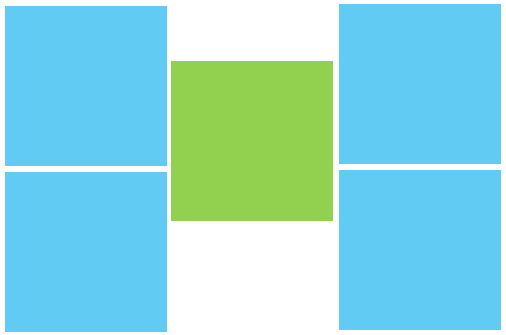


Specified stoichiometry: An

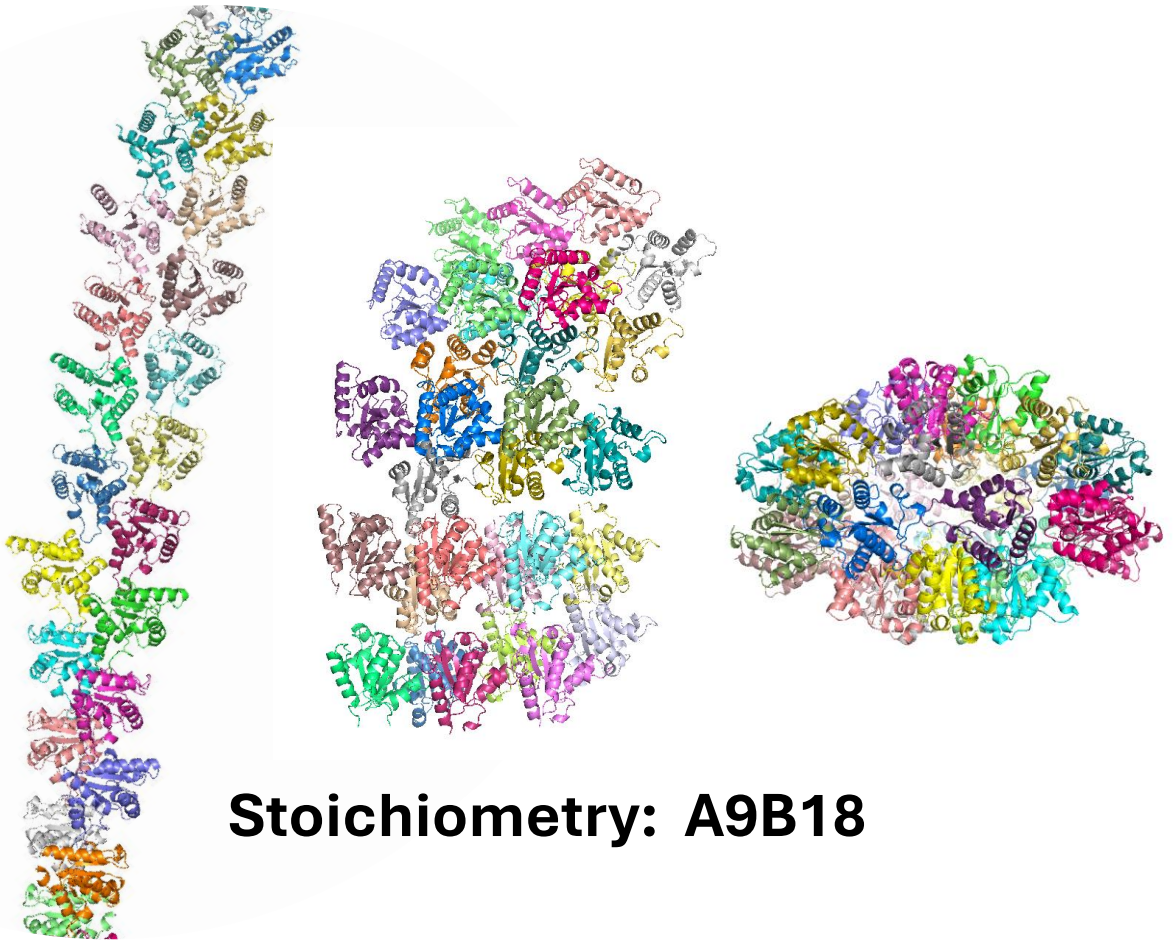
T1219o model

3D structure of the exact protein (PDB: 1zmq)

Challenge 2: multiple possible interfaces between proteins and unusual shape of protein complex



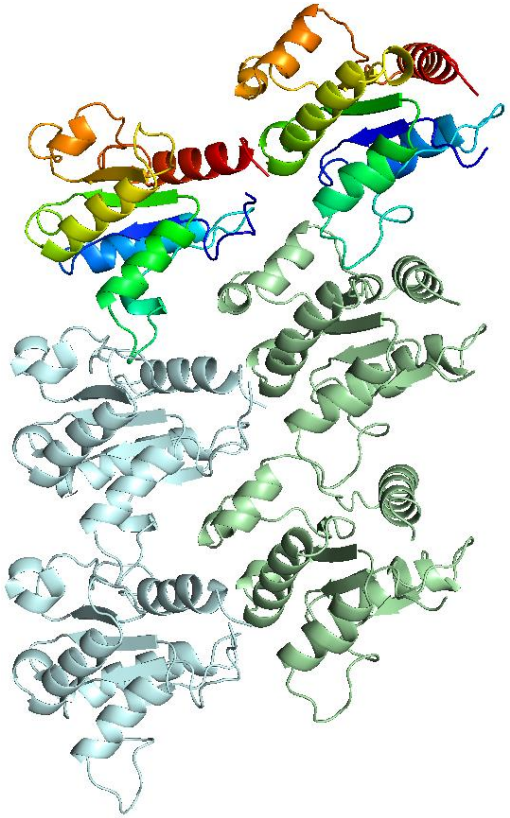
H1265 target



Stoichiometry: A9B18

Examples of models for H1265

Challenge 2: multiple possible interfaces between proteins and unusual shape of protein complex

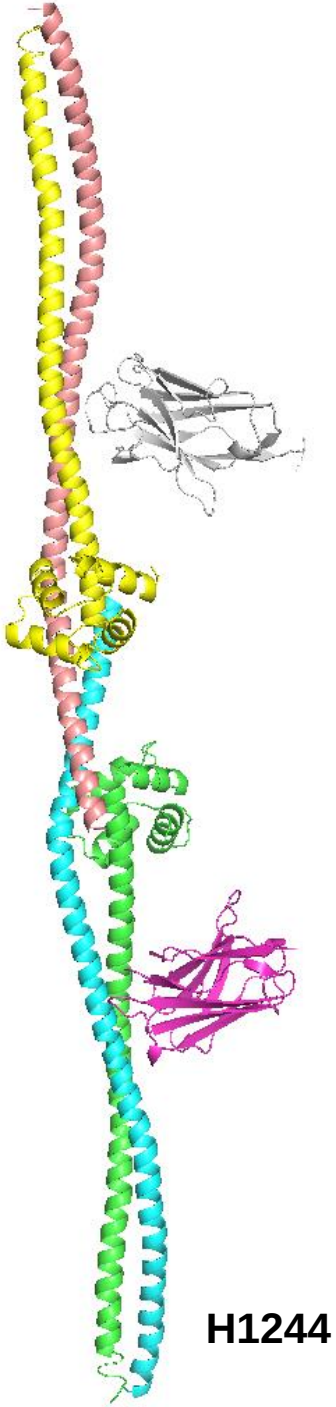


The shape is correct, but the interfaces are incorrect

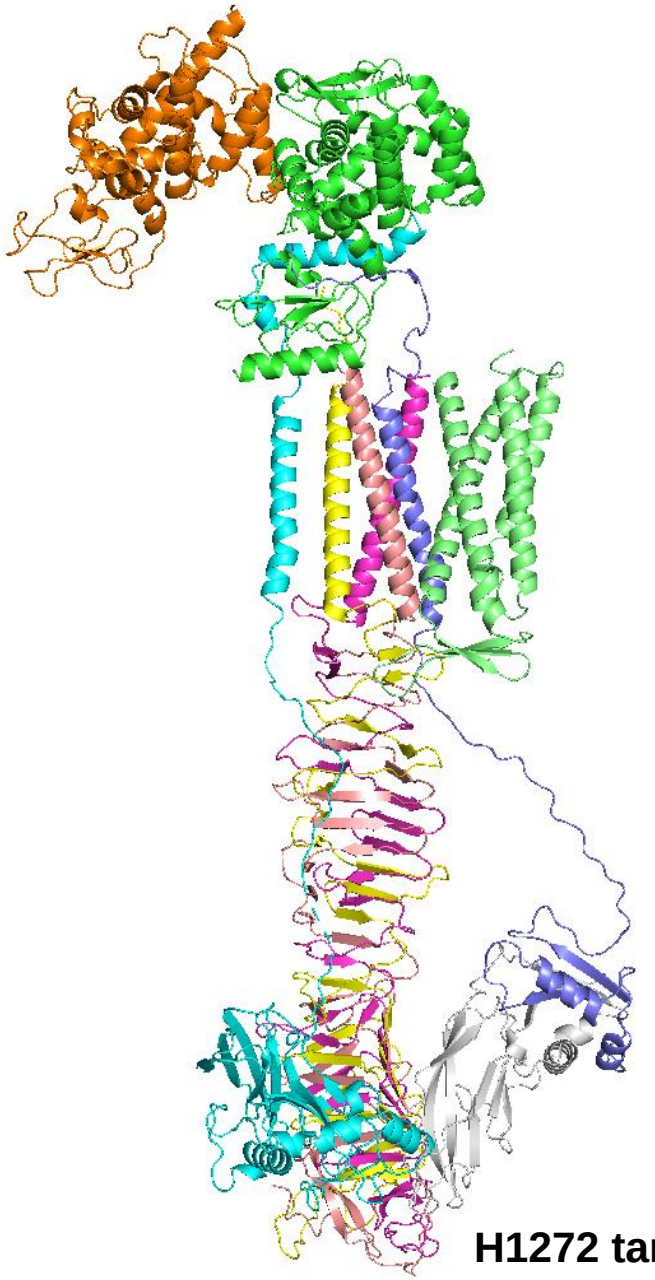
H1265 target

A winning model from the Kihara group

Other challenging targets

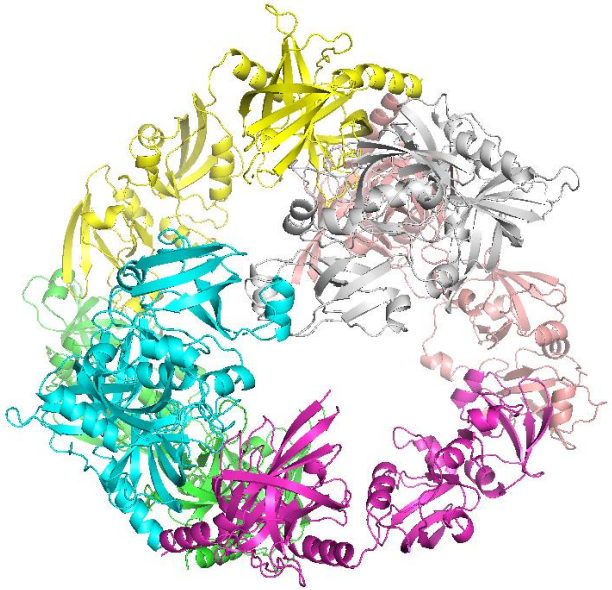


H1244 target



H1258 target

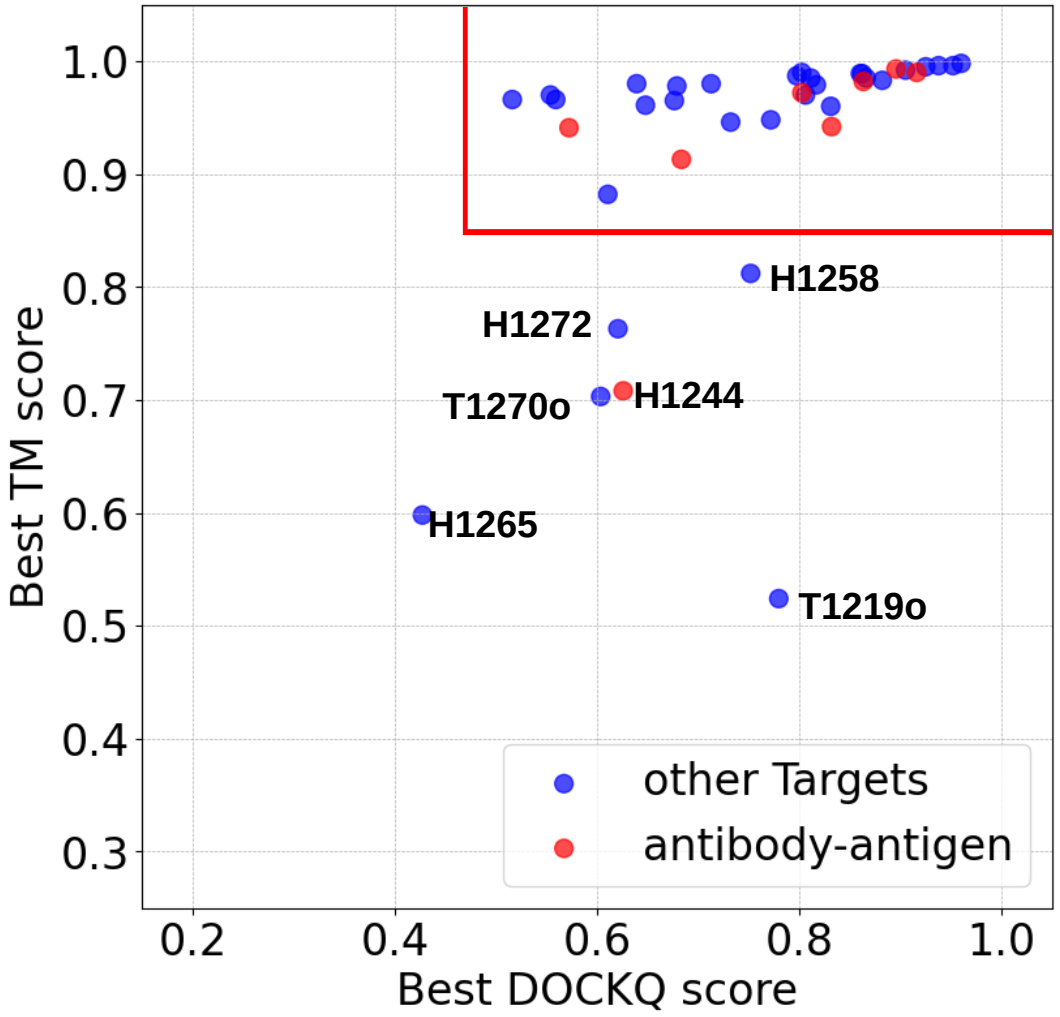
H1272 target

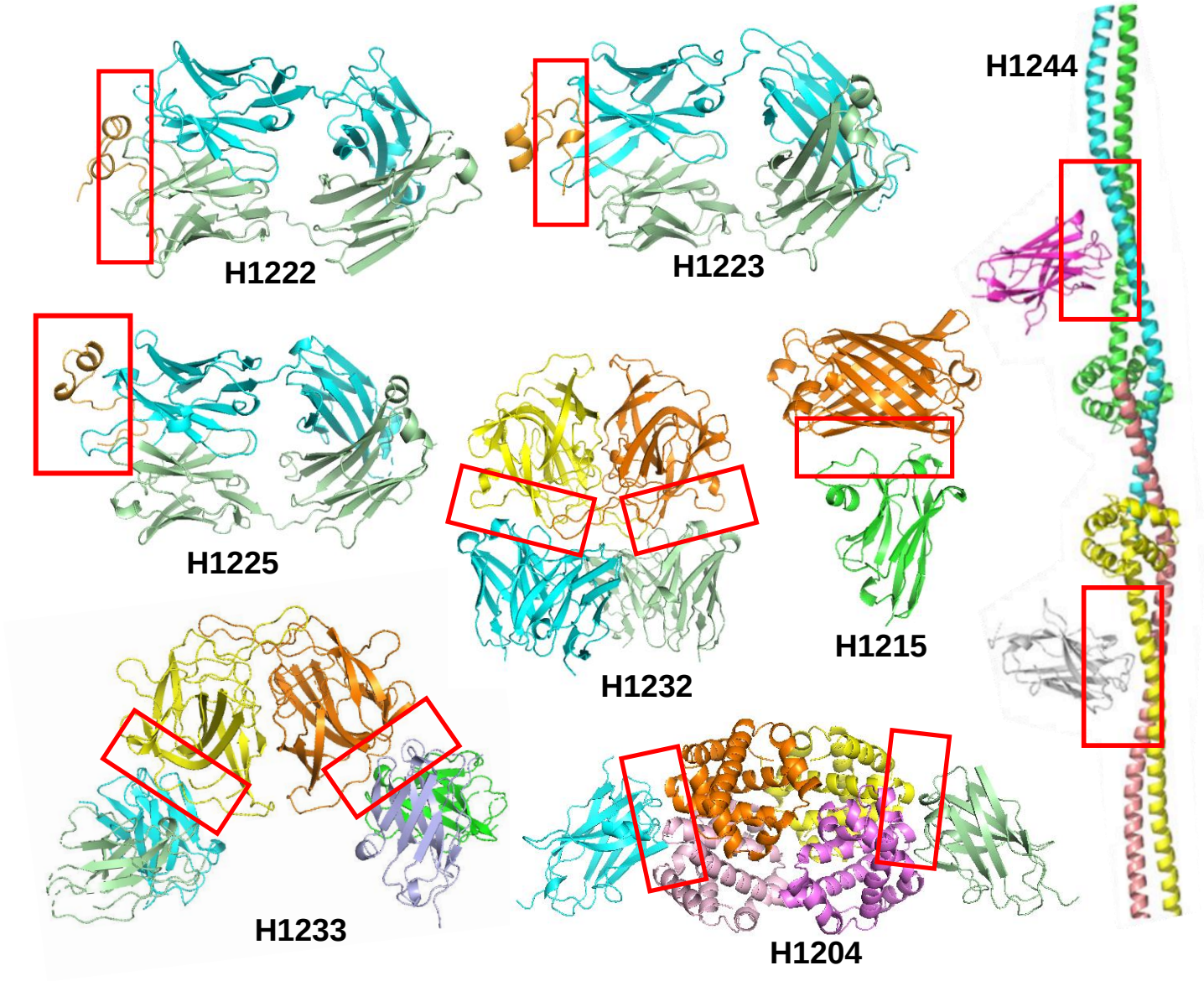


T1270o target

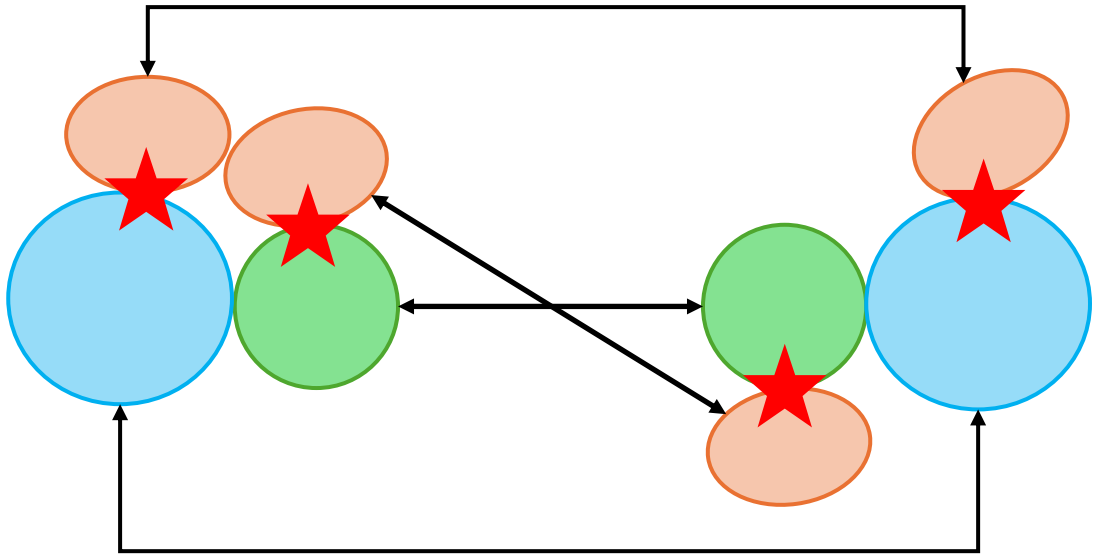
Are we making progress?

What about antibody-antigen interactions ?

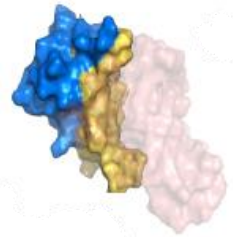




A specialized scoring routine for antibody-antigen targets



Interface/Local:



DockQ: interface size-weighted average

$$\mathrm{IPS}(M,T) = J_C(M,T) = \frac{|M_{\text{i-Res}} \cap T_{\text{i-Res}}|}{|M_{\text{i-Res}} \cup T_{\text{i-Res}}|}$$

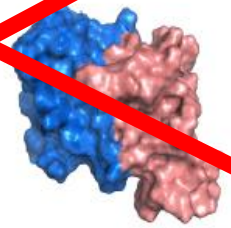
$$\mathrm{ICS}(M,T) = 2 \cdot \frac{P(M_{\text{cnt}},T_{\text{cnt}}) \times R(M_{\text{cnt}},T_{\text{cnt}})}{P(M_{\text{cnt}},T_{\text{cnt}}) + R(M_{\text{cnt}},T_{\text{cnt}})}$$

$$\text{QS-best}(M,T) = \frac{|M_{\text{cnt}} \cap T_{\text{cnt}}|}{\max(M_{\text{cnt}},T_{\text{cnt}})}$$

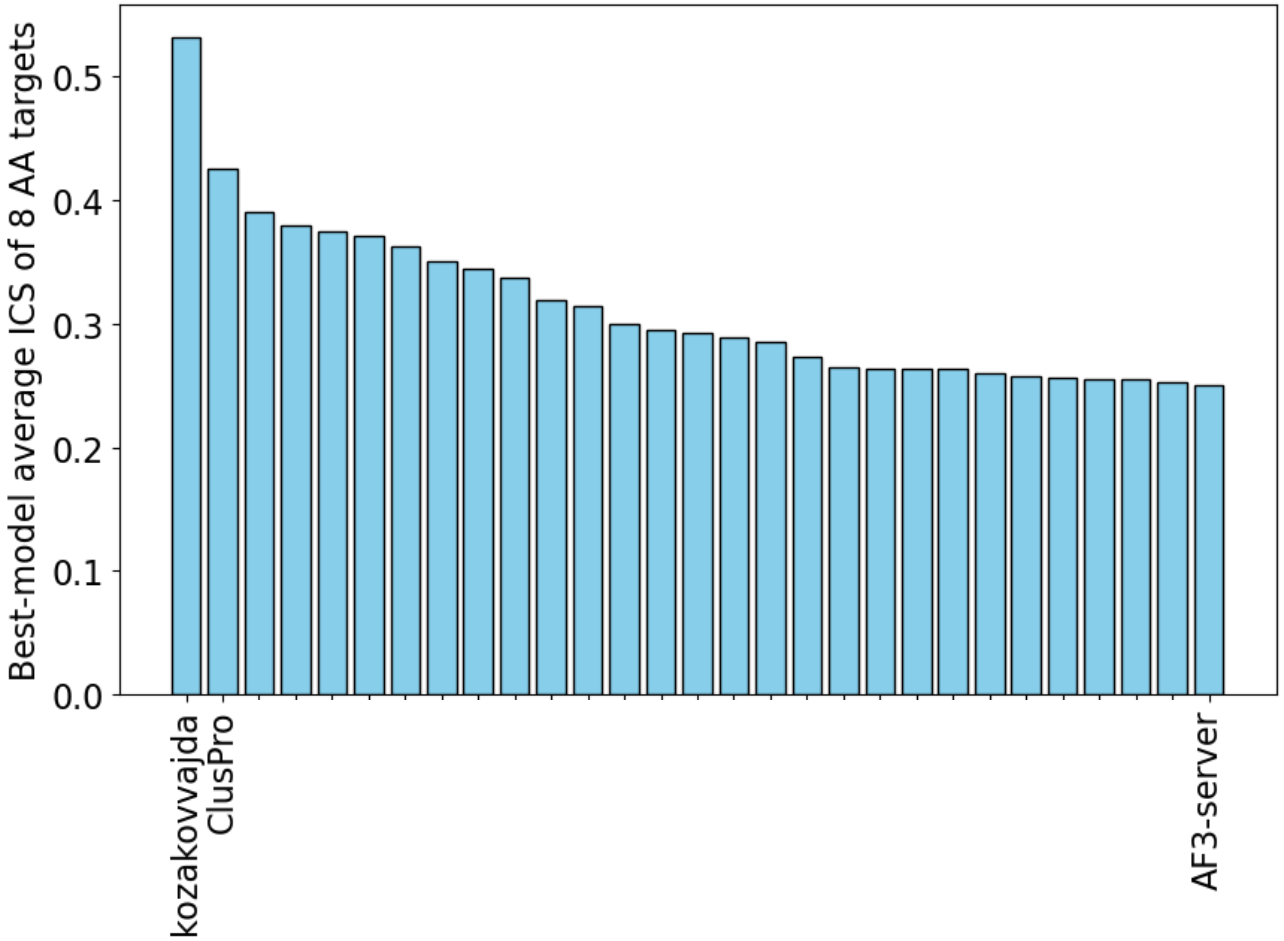
Assembly/Global:

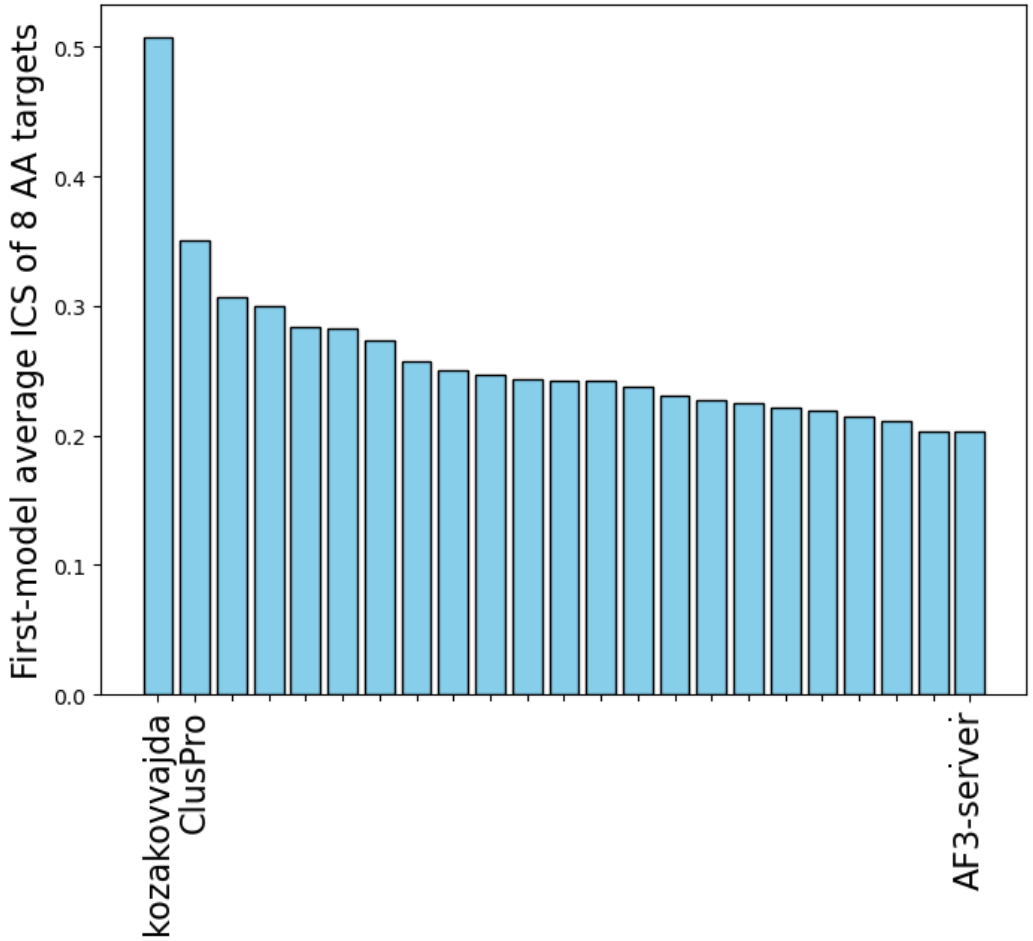
lDDT

TM-score

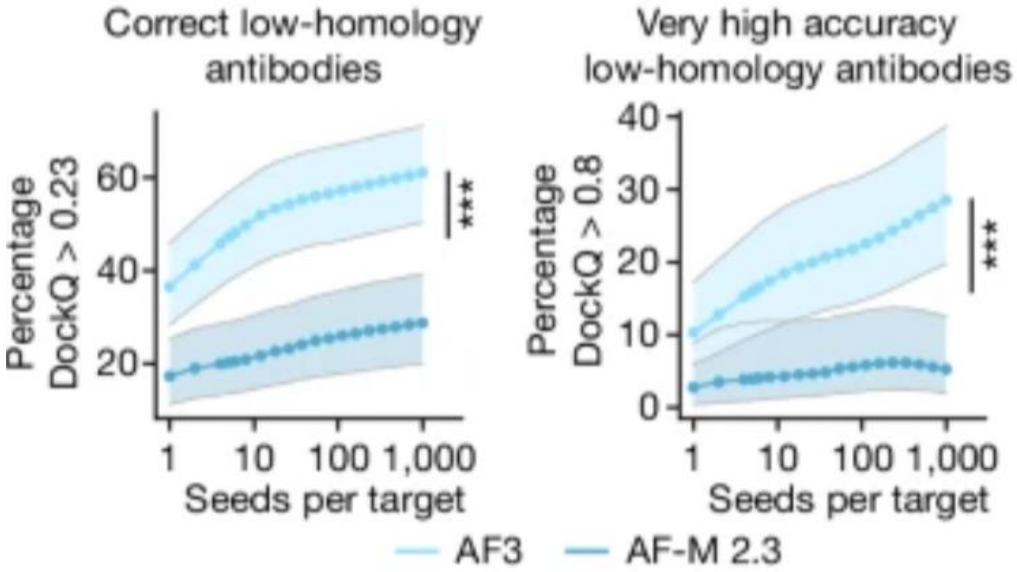


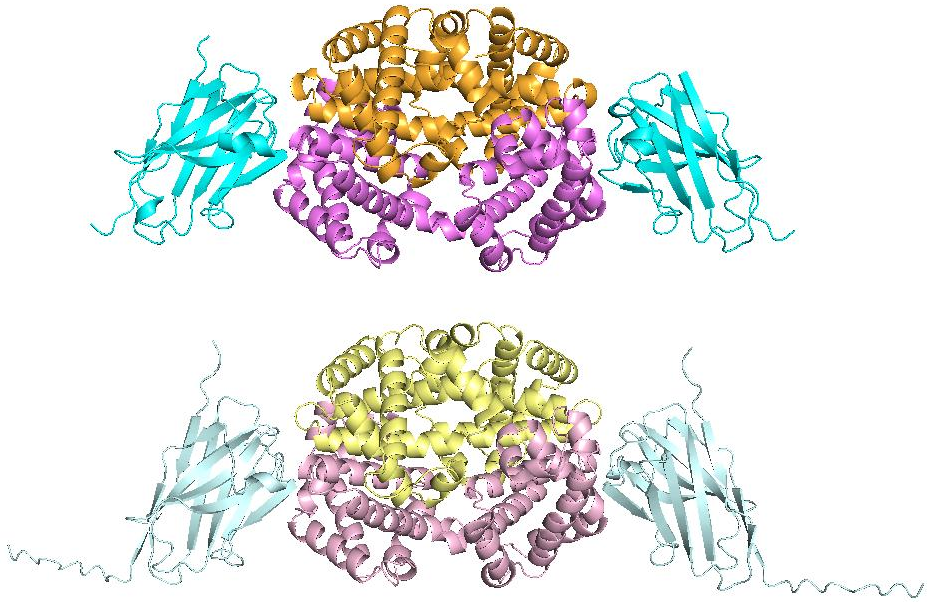
The Kozakov group is outstanding in antibody-antigen (AA) targets





Can people remarkably outperform AF3?

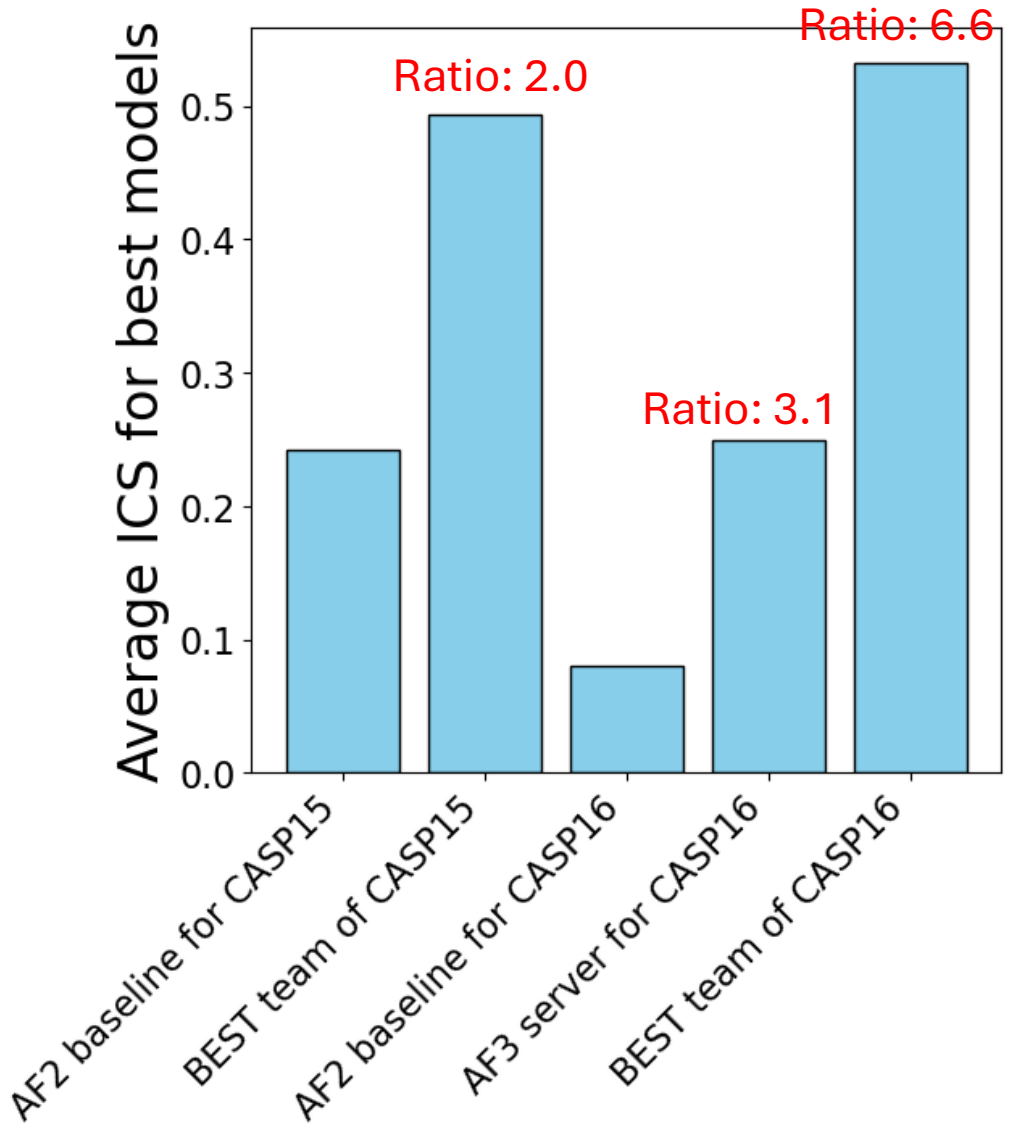


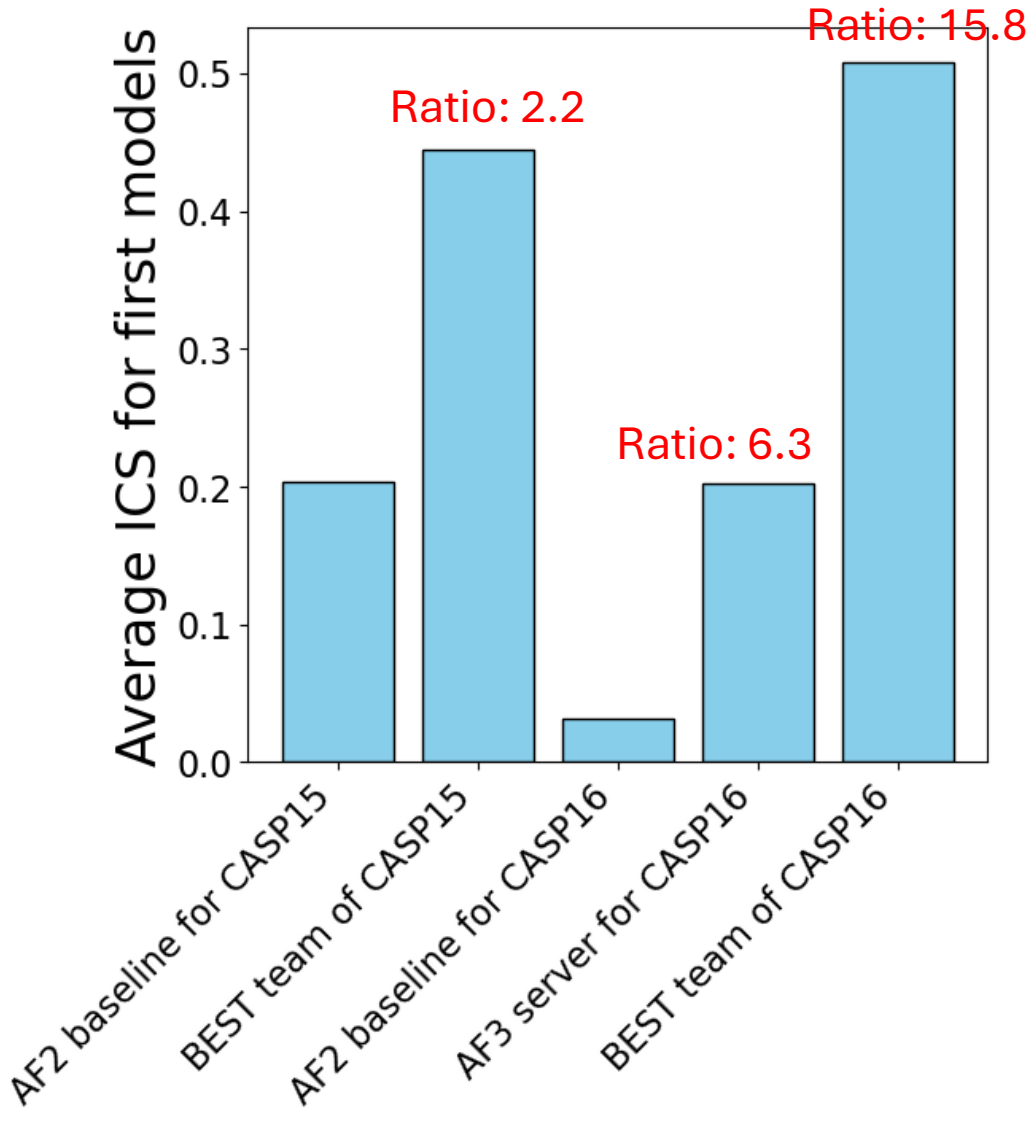


H1204 and a winning model from the Kozakov group

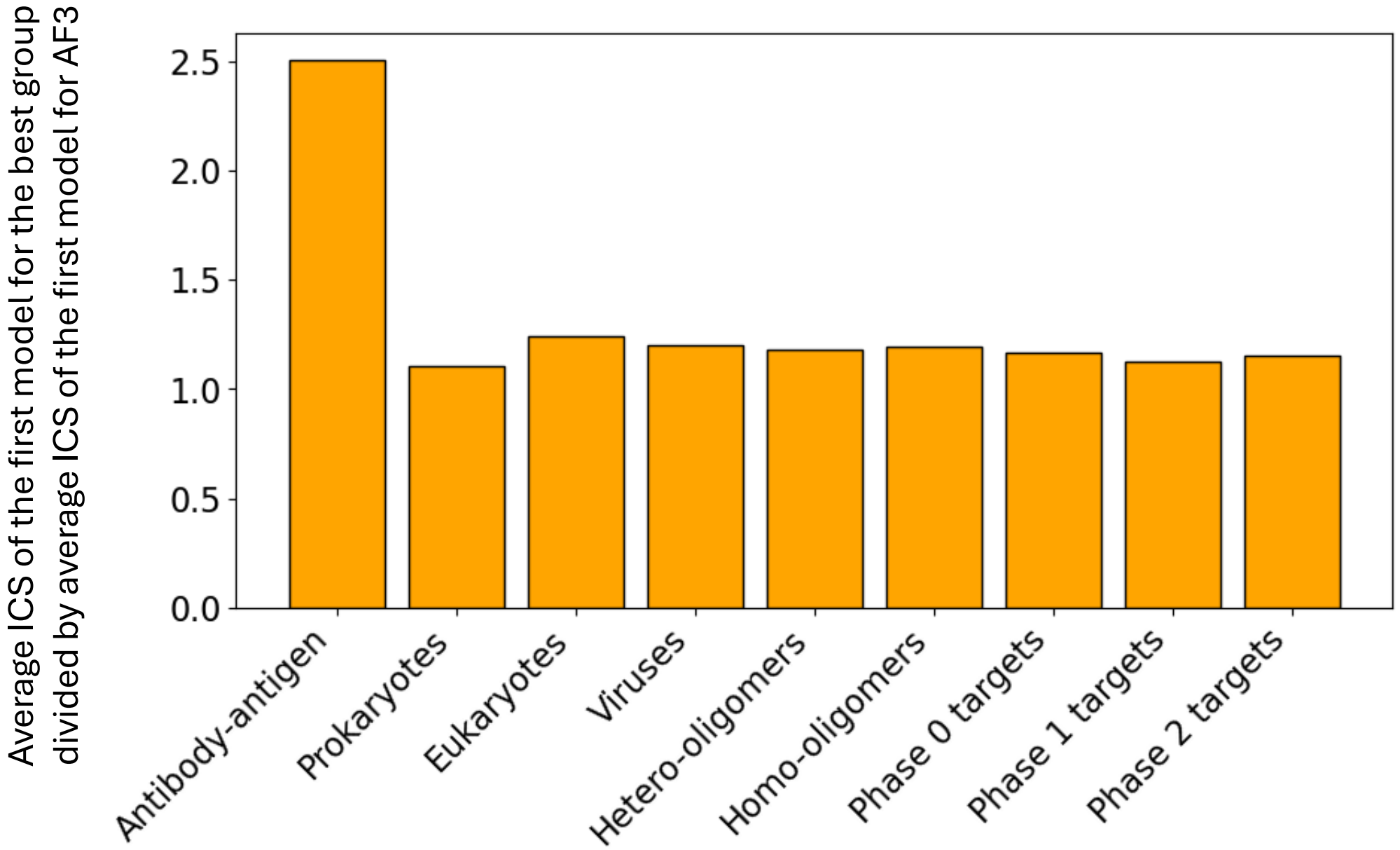
	H1204	H1215	H1222	H1223	H1225	H1232	H1233	H1244	DockQ > 0.23	DockQ > 0.8
Kozakov	0.8755	0.88	0.334	0.7575	0.158	0.0485	0.8995	0.0048	**62.5%**	**37.5%**
AF3	0.0192	0.195	0.503	0.1	0.098	0.0245	0.8845	0.0172	**25%**	**12.5%**

Are we making any progress in antibody targets?

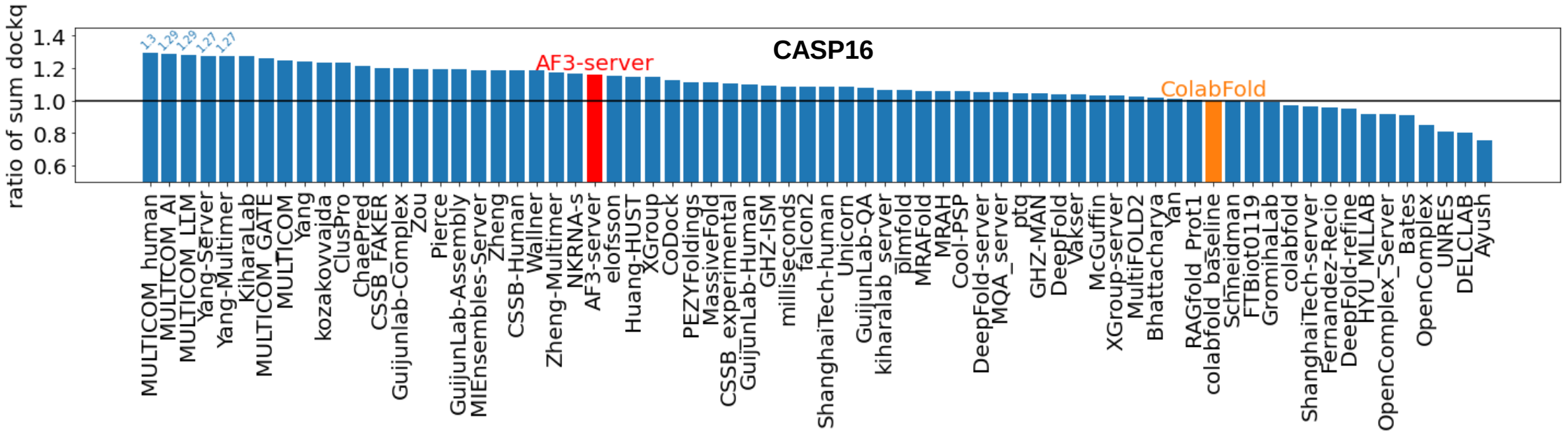


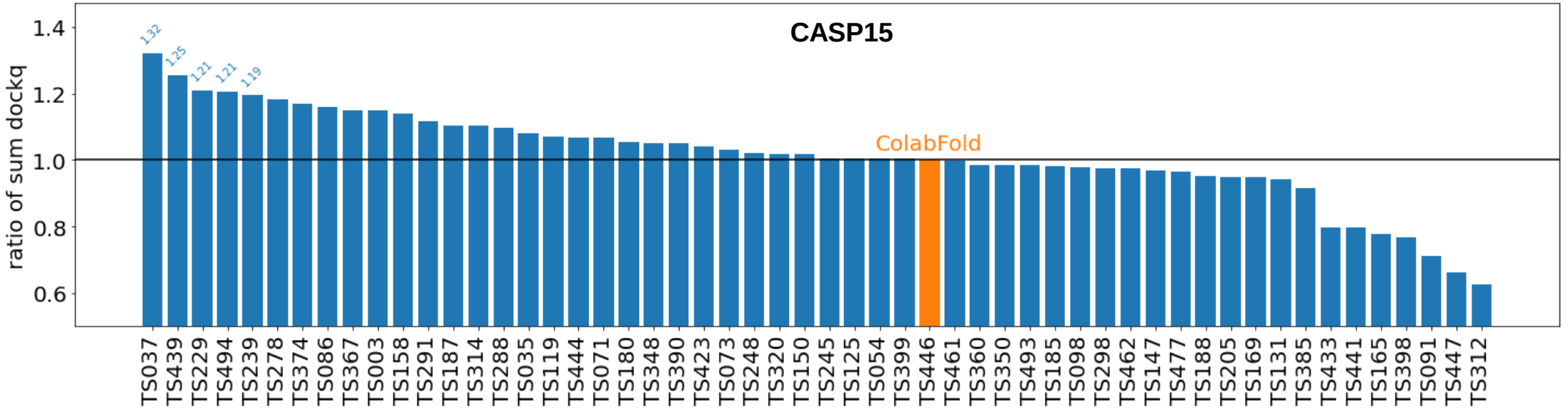


In what other aspects do people outperform AF3?



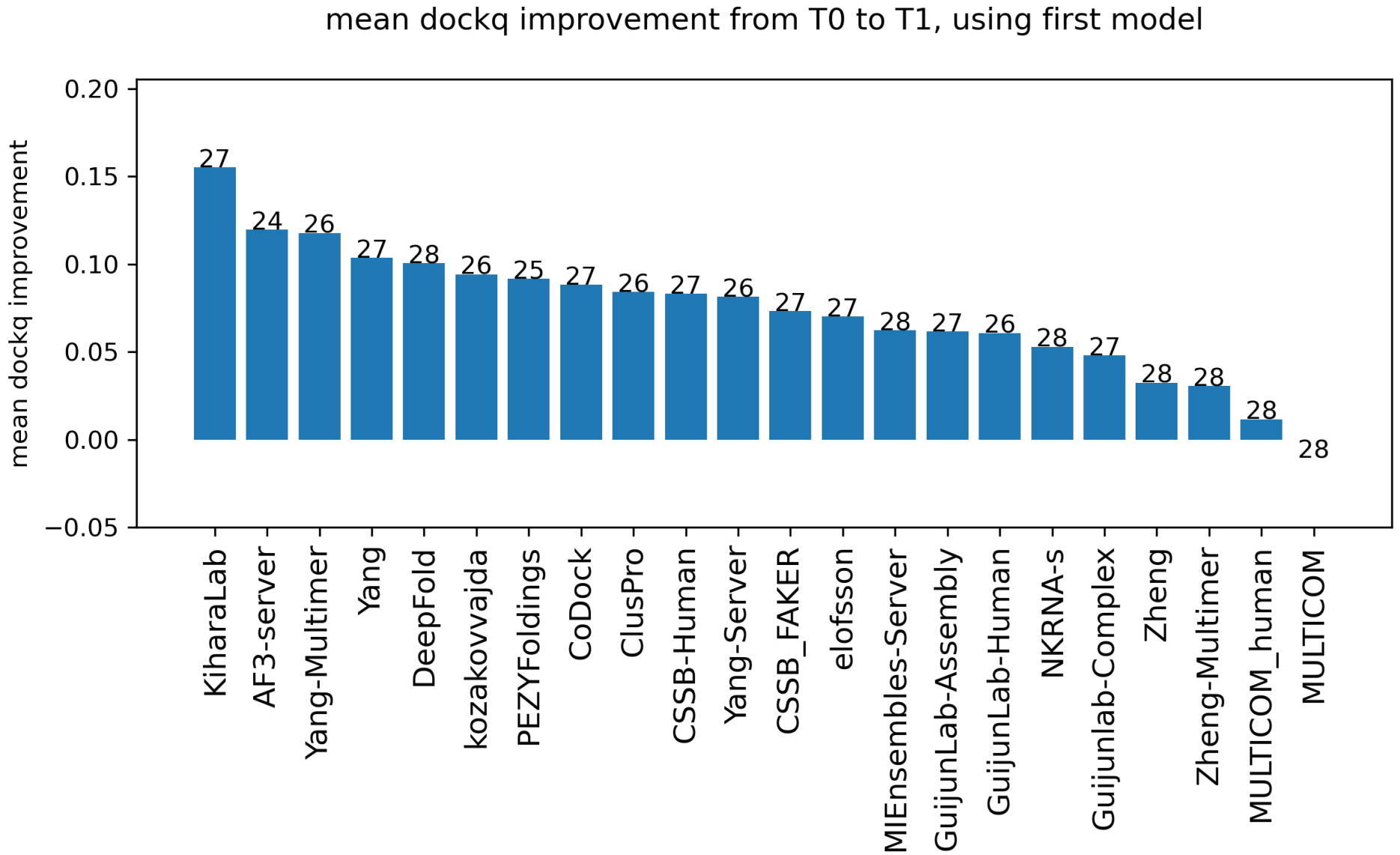
There is no visible progress in other targets



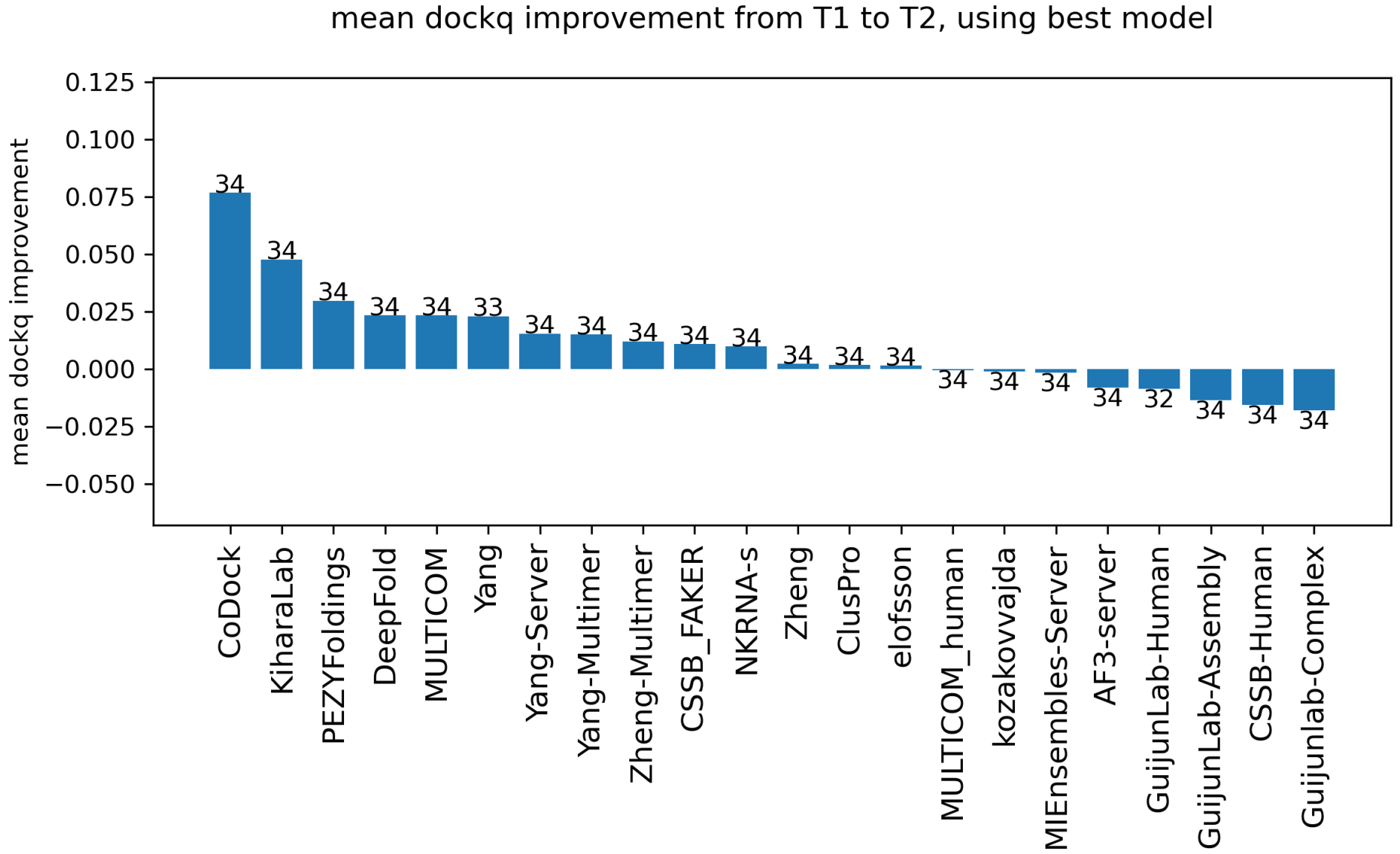


Additional experiments of this CASP

T1 vs T0: knowing the stoichiometry help all top 25% groups in predicting oligomer structures

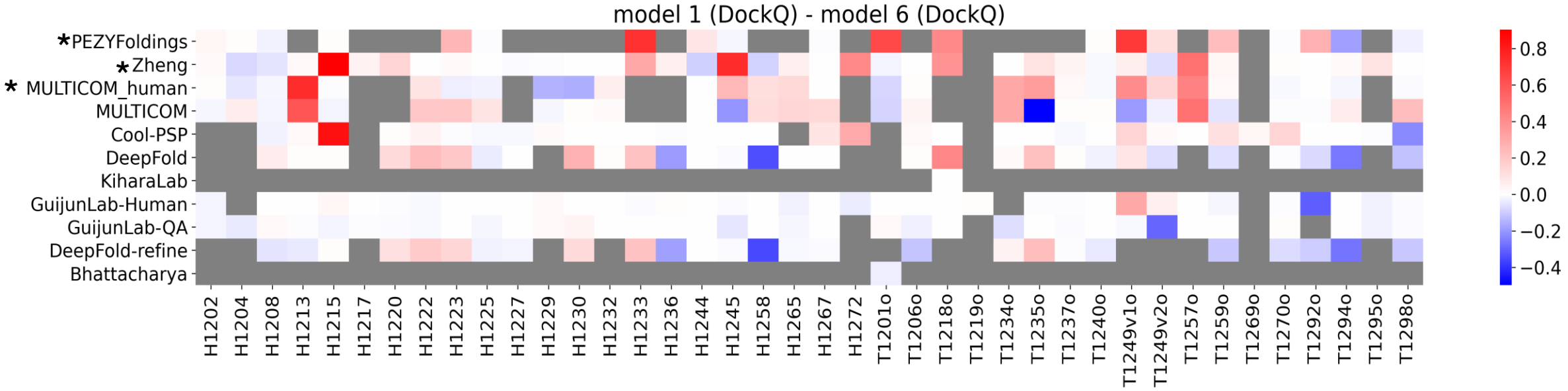


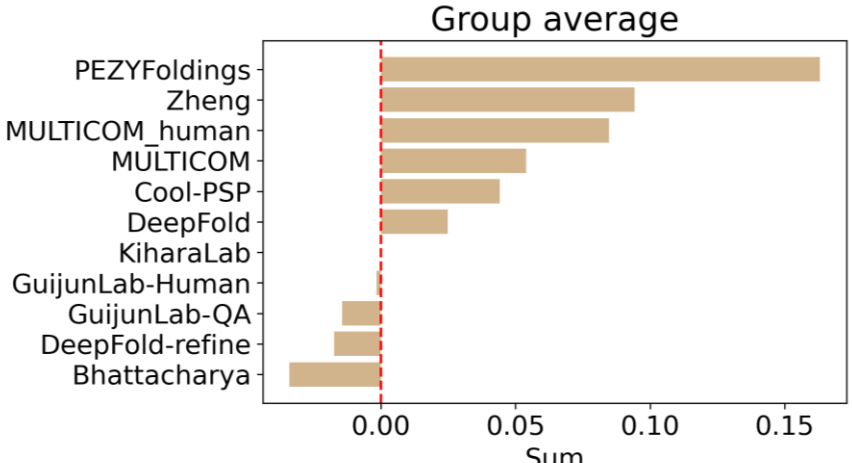
T2 vs T1: massive models are helpful for most of top 25% groups in predicting oligomer structures



Comparison of model 1 with model 6 of the same group reveals MSA-building ability

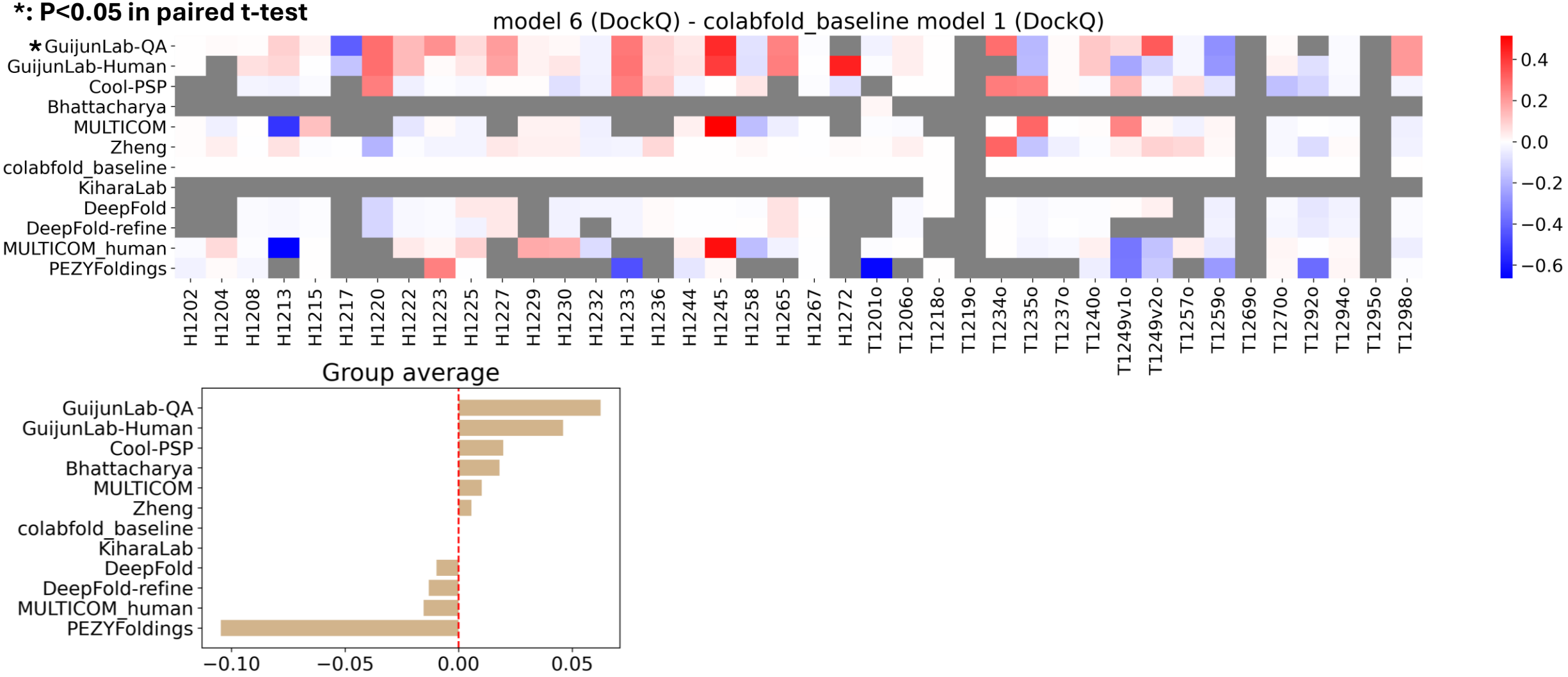
***: P<0.05 in paired t-test**



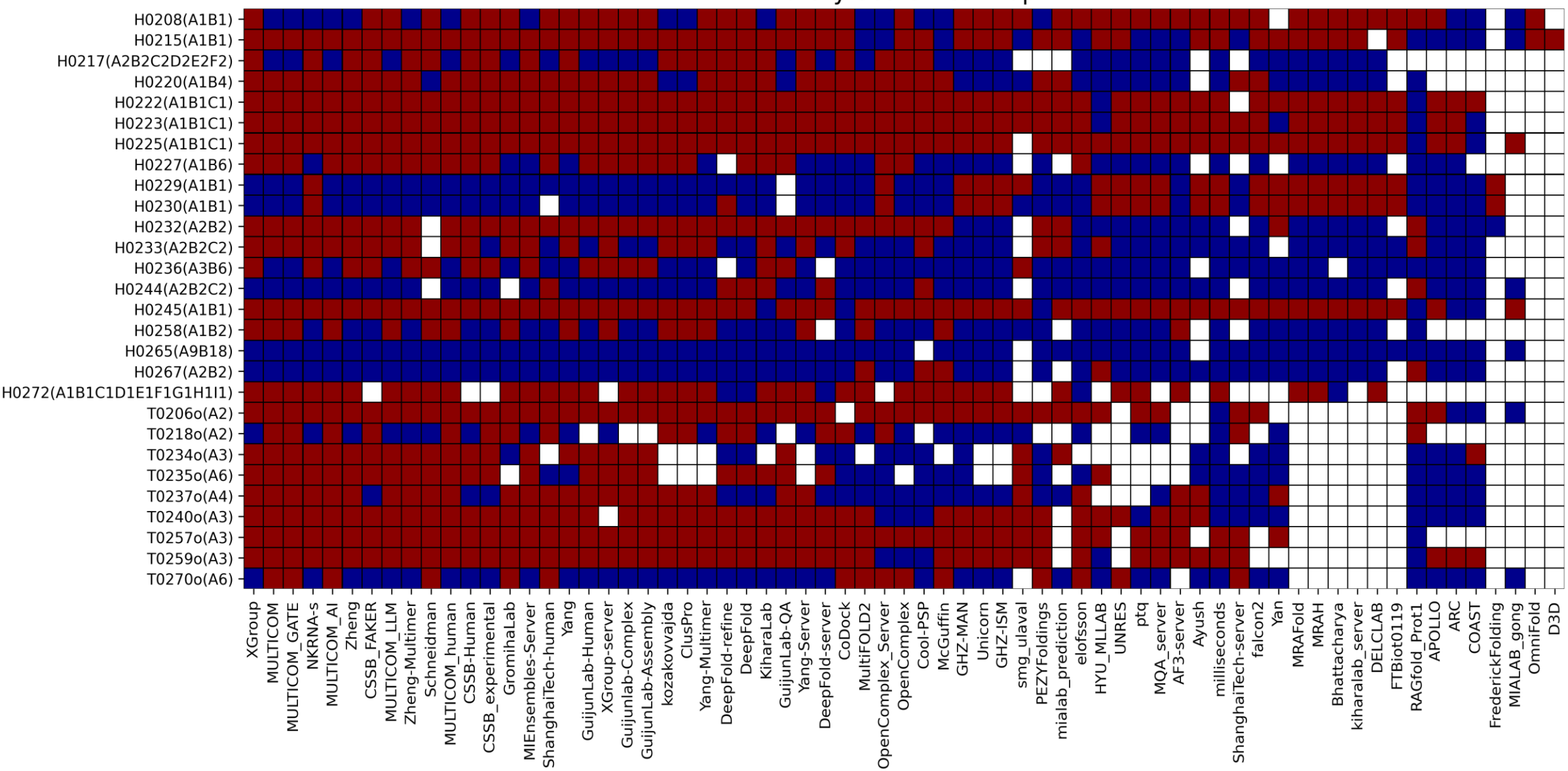


Comparison of model 6 with ColabFold model 1 reveals modeling ability

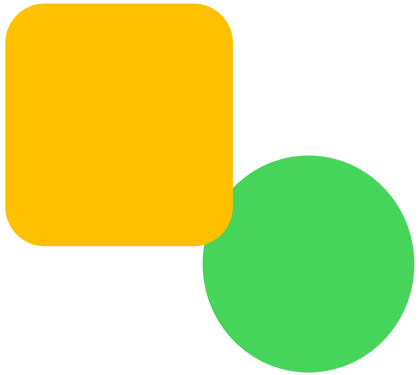
***: P<0.05 in paired t-test**



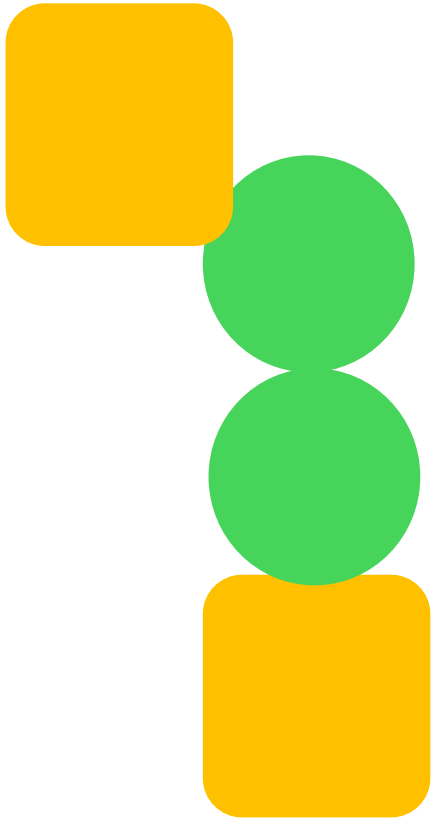
Phase 0 test people's ability to predict stoichiometry



Phase 0, as well as filament targets, challenge our ability to evaluate models

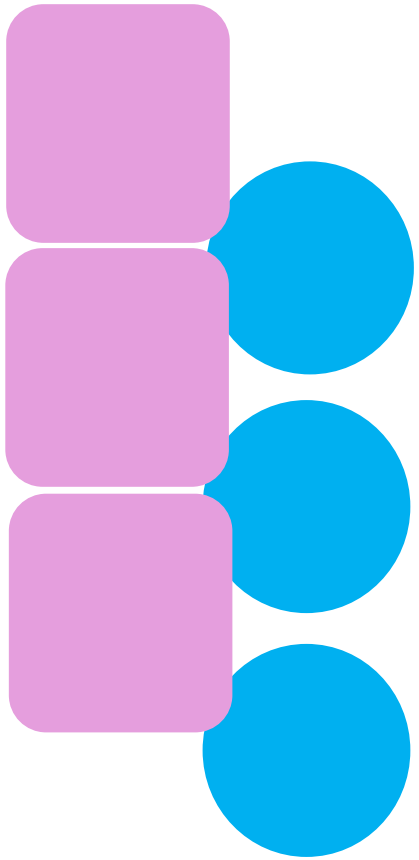


Target

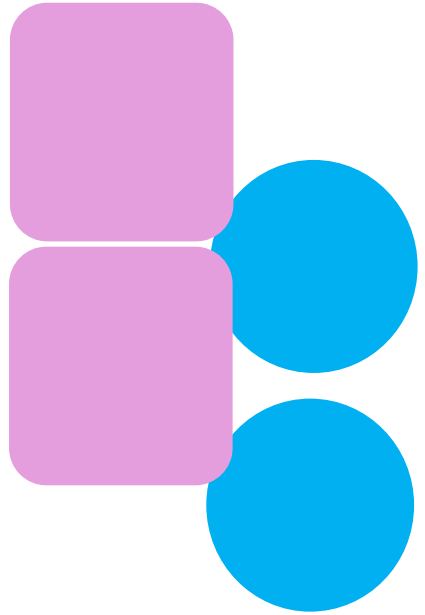


Model

Bias towards



Target

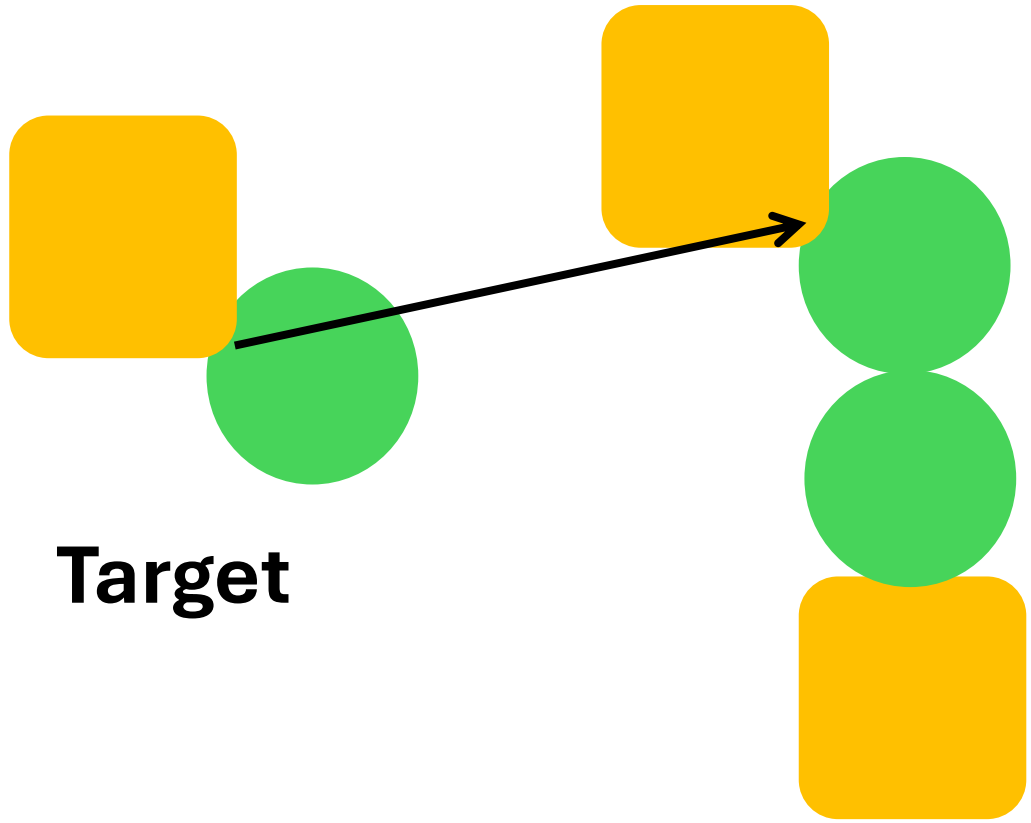


Model

Bias against

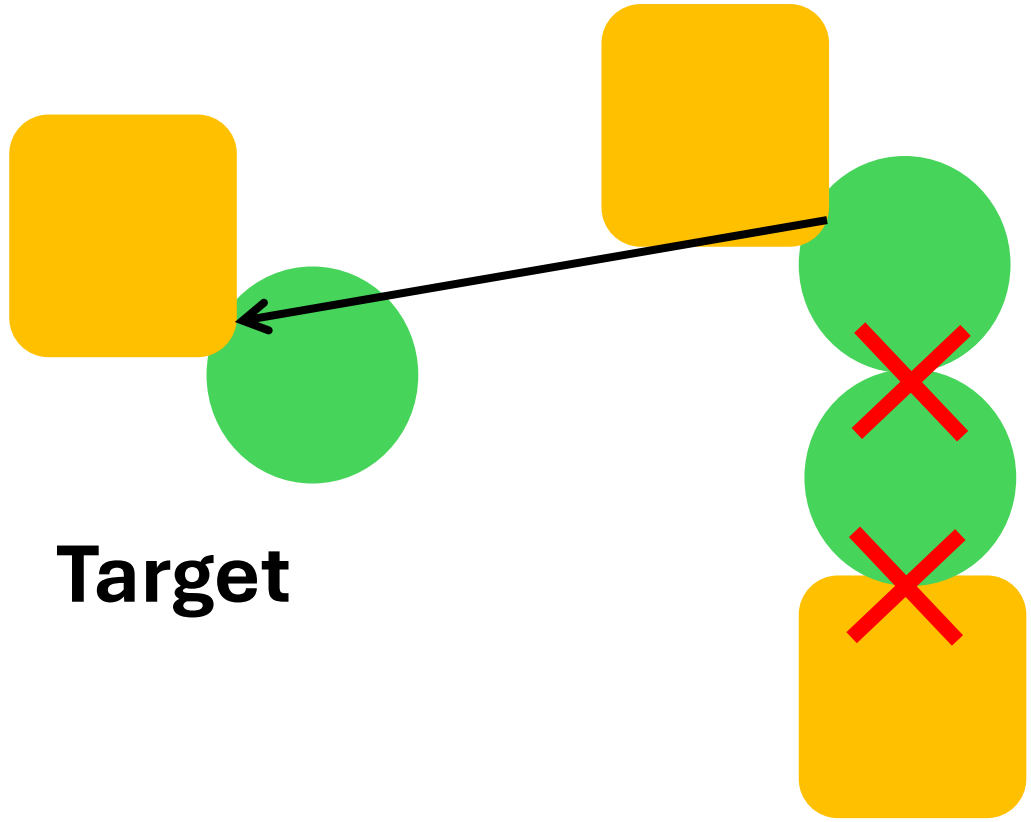
New evaluation routine for targets of unknown stoichiometry

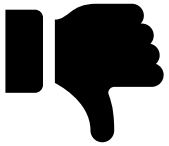
Reciprocal Best Match pair-wise scoring



New evaluation routine for targets of unknown stoichiometry

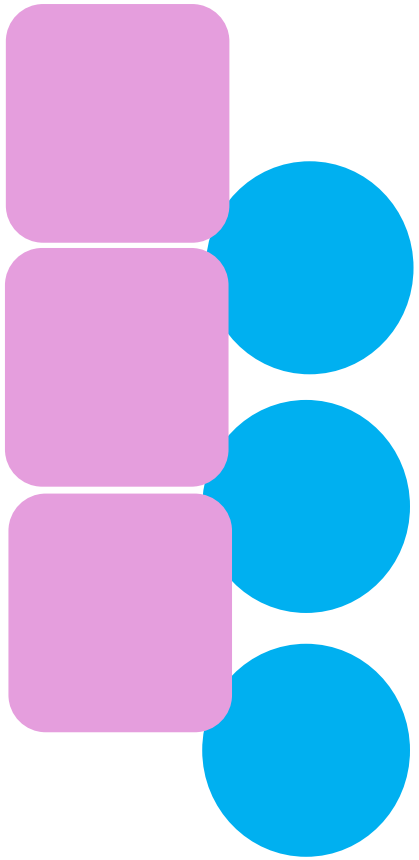
Reciprocal Best Match pair-wise scoring

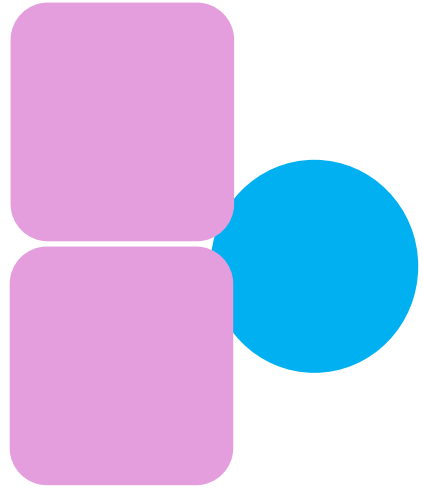




New evaluation routine for targets of unknown stoichiometry

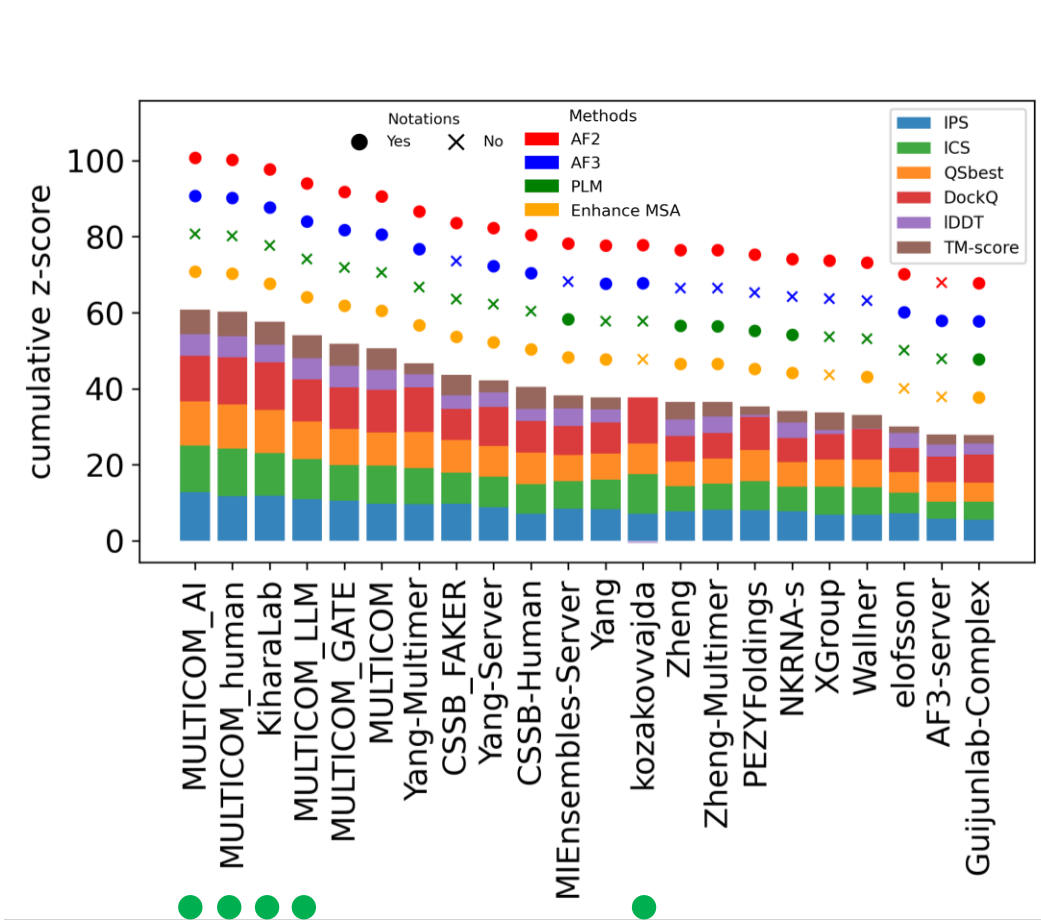
Reciprocal Best Match pair-wise scoring

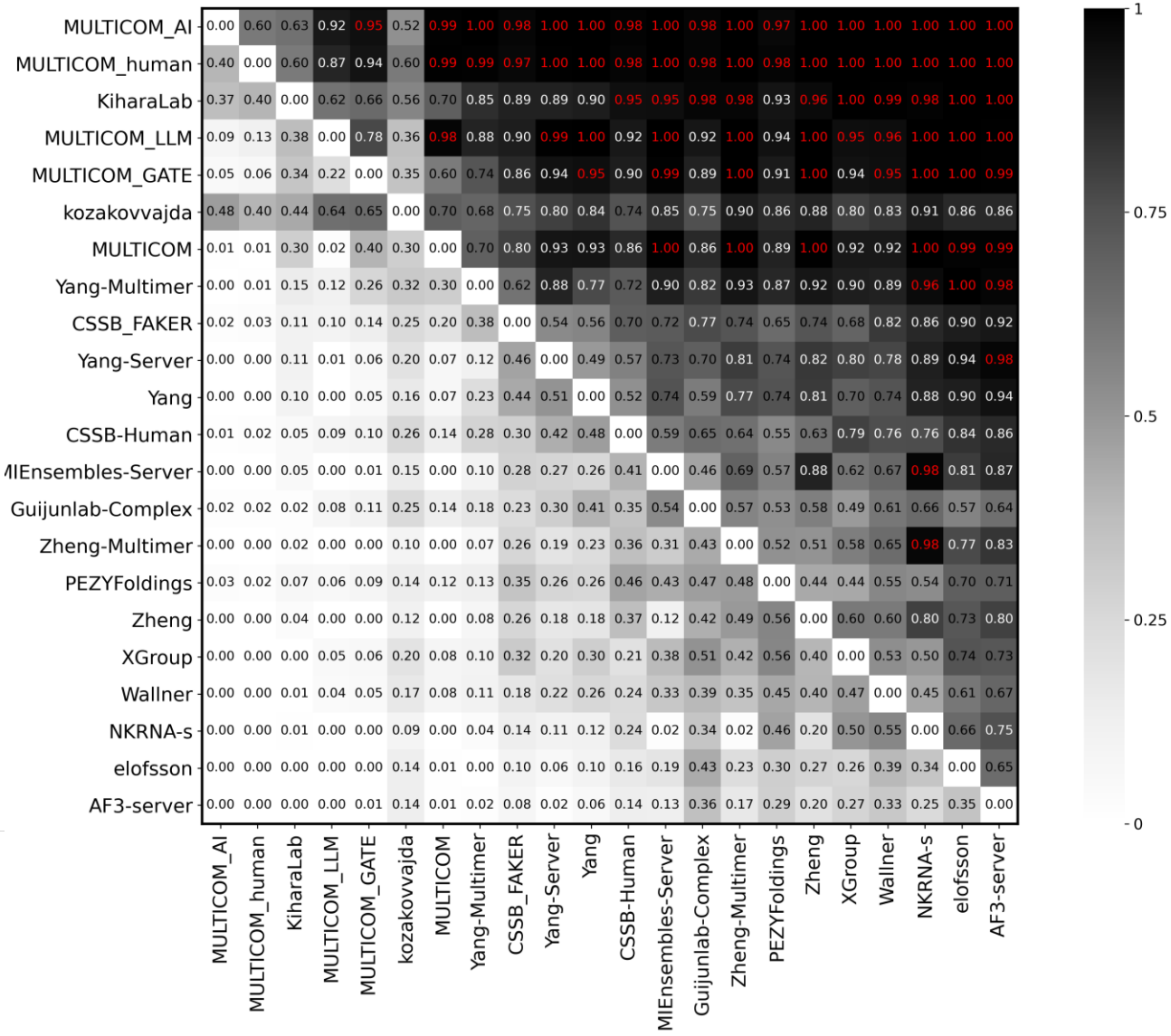




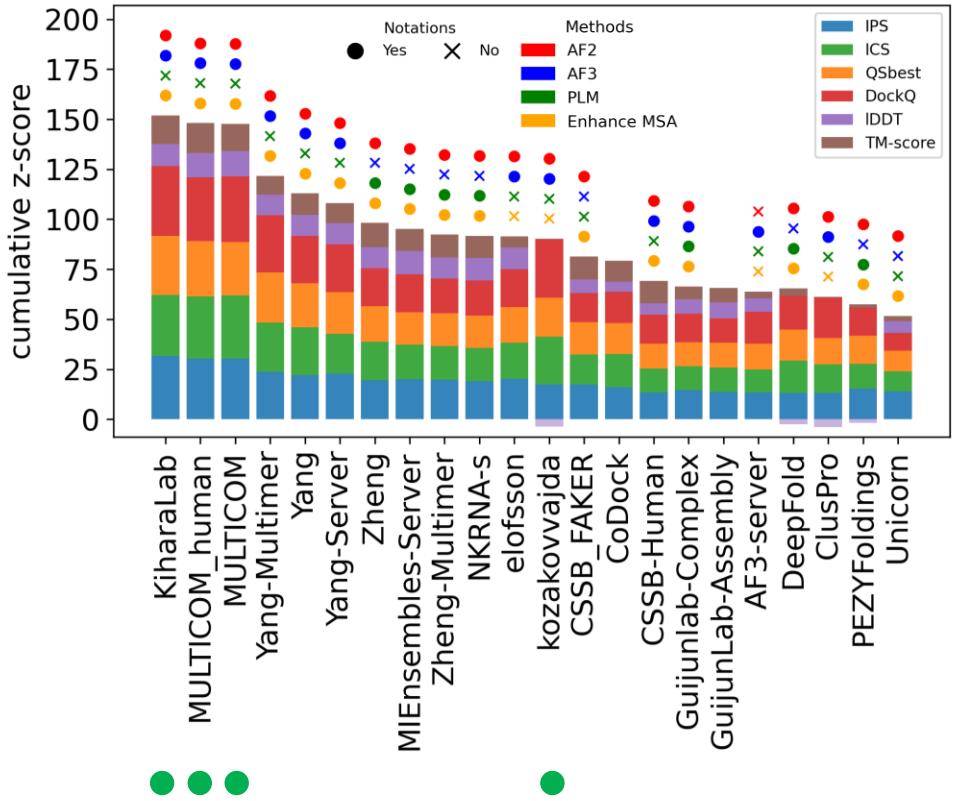
The ranking

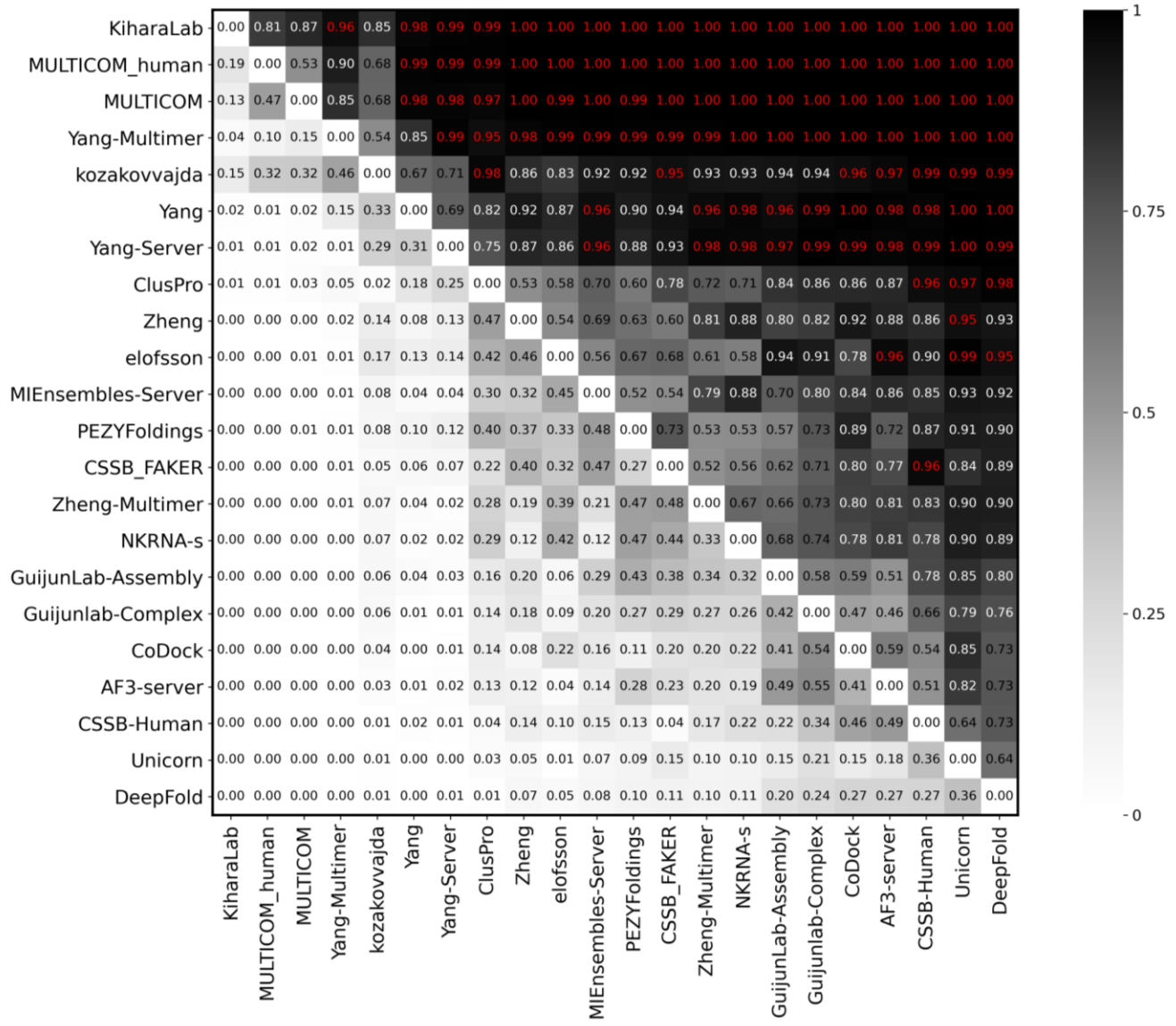
Ranking on Phase 1 best models



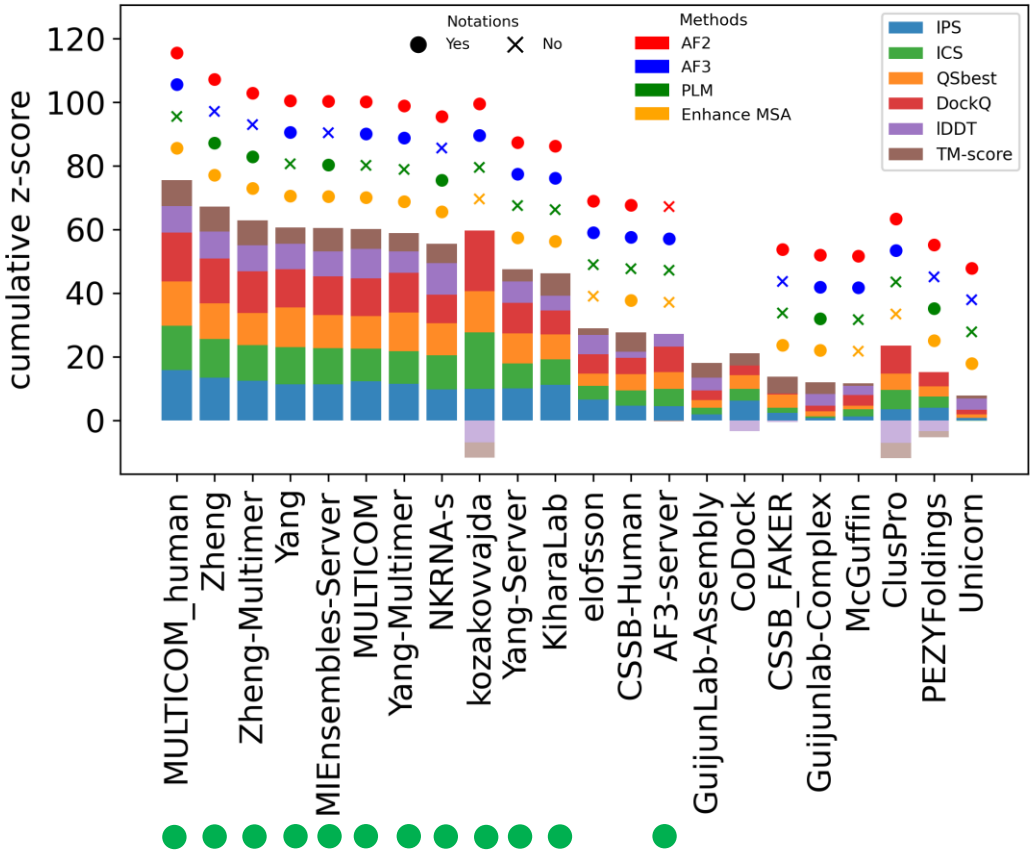


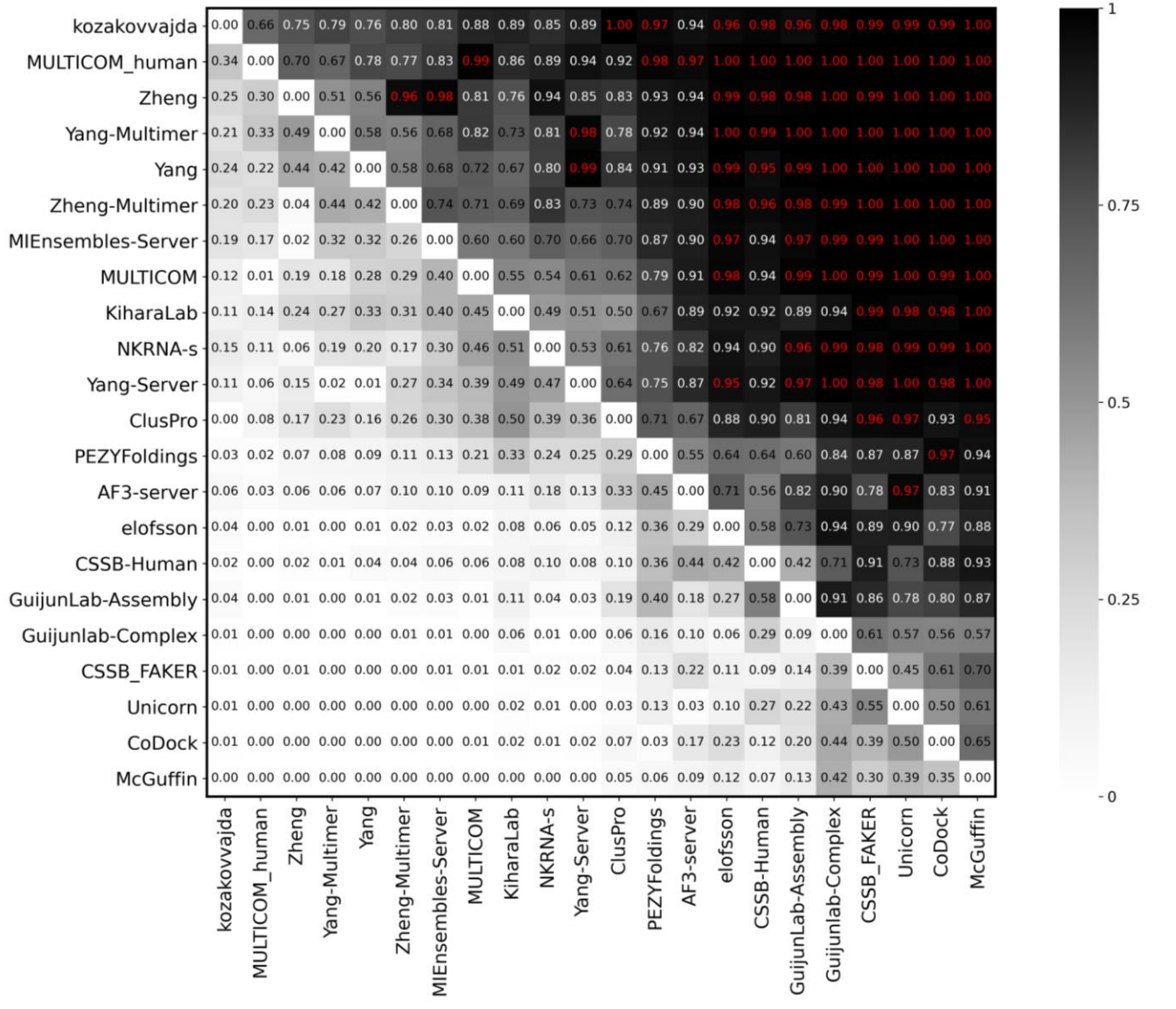
Ranking on best models for targets from all phases



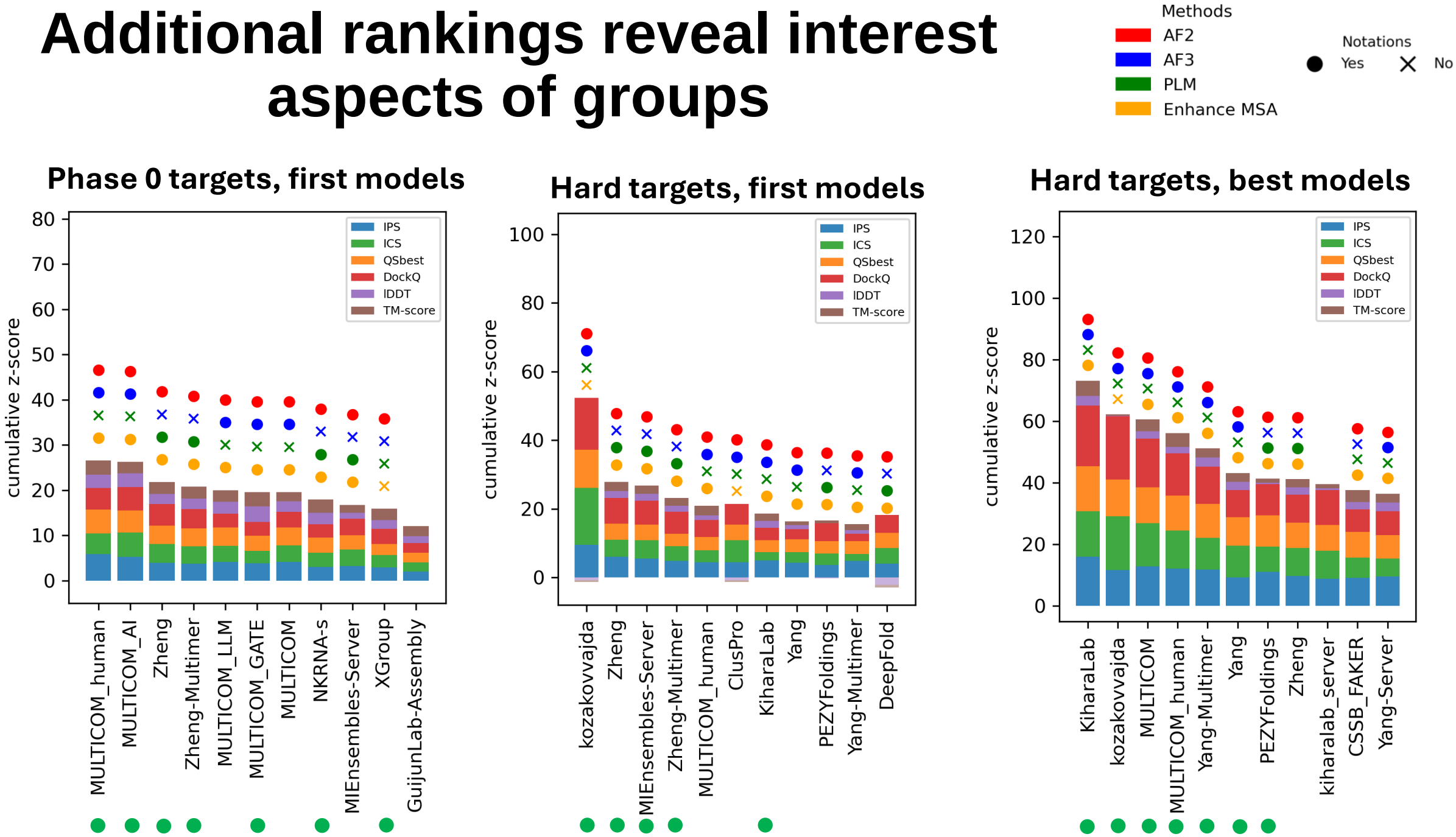


Ranking on first models for targets from all phases

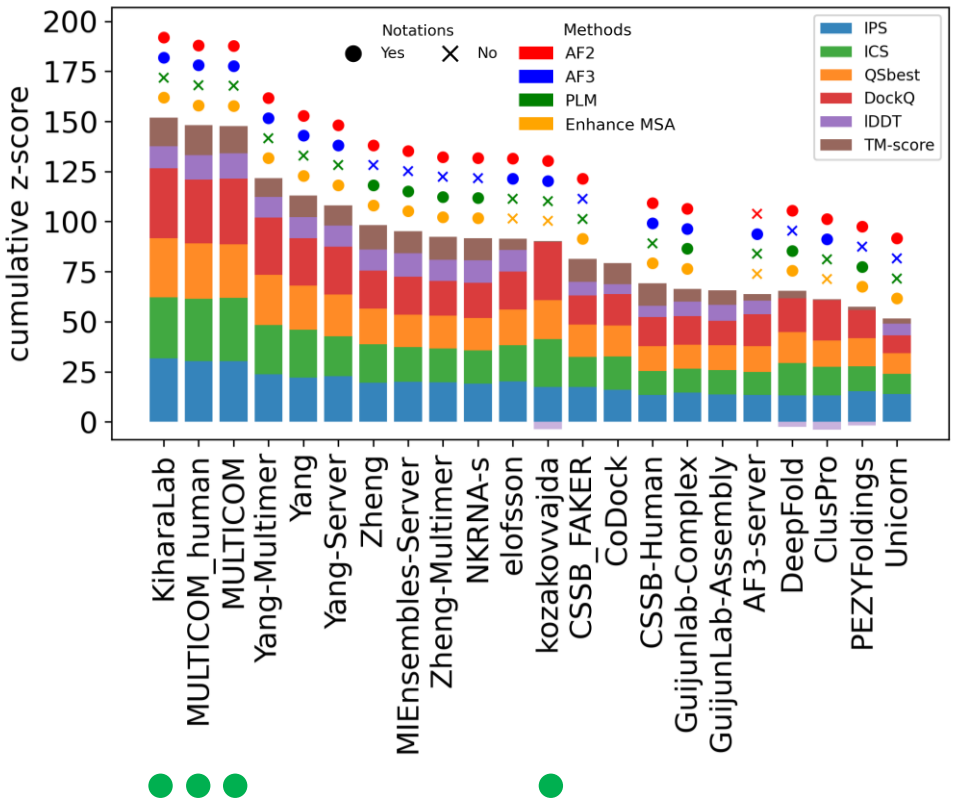


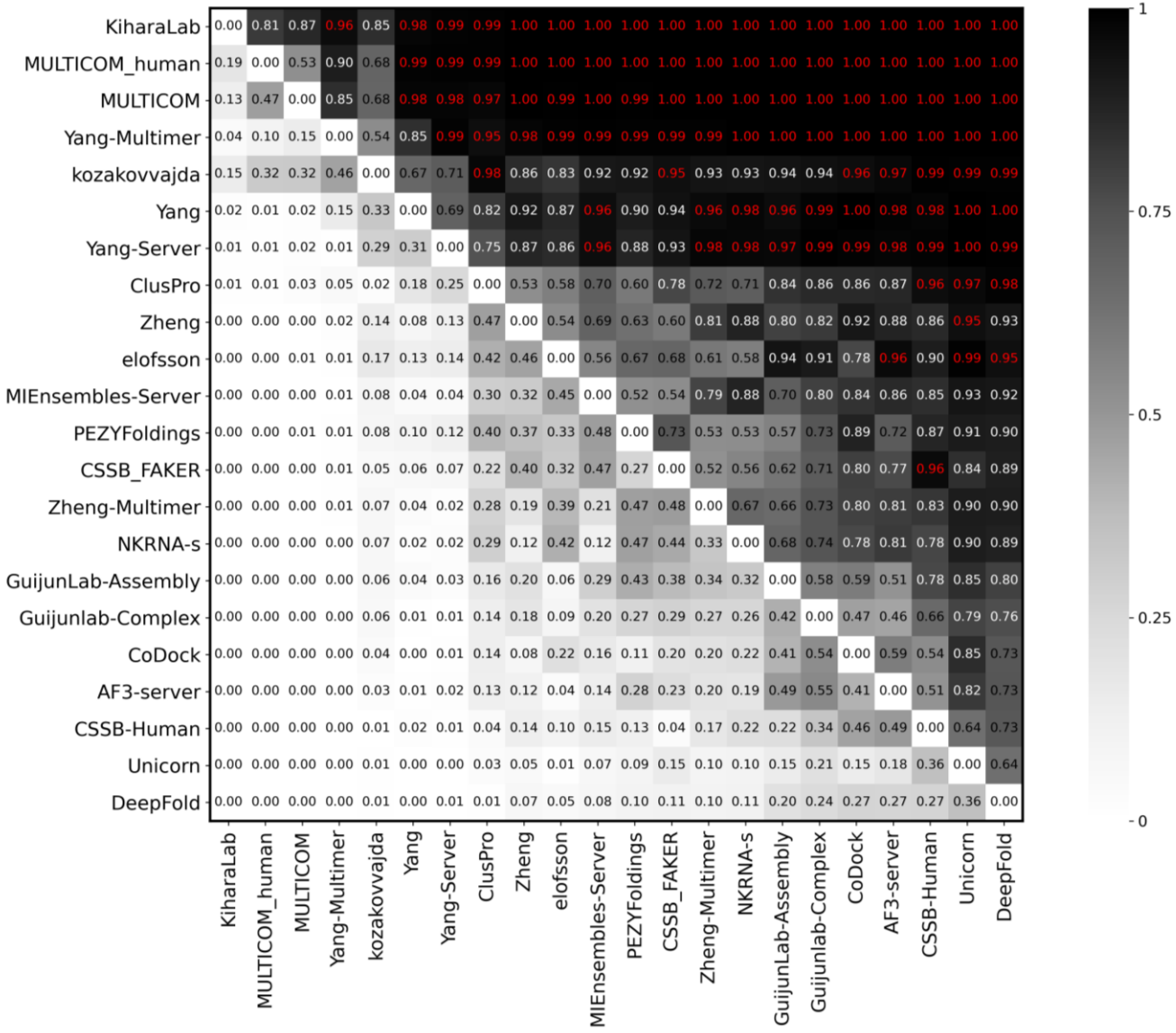


Additional rankings reveal interest aspects of groups

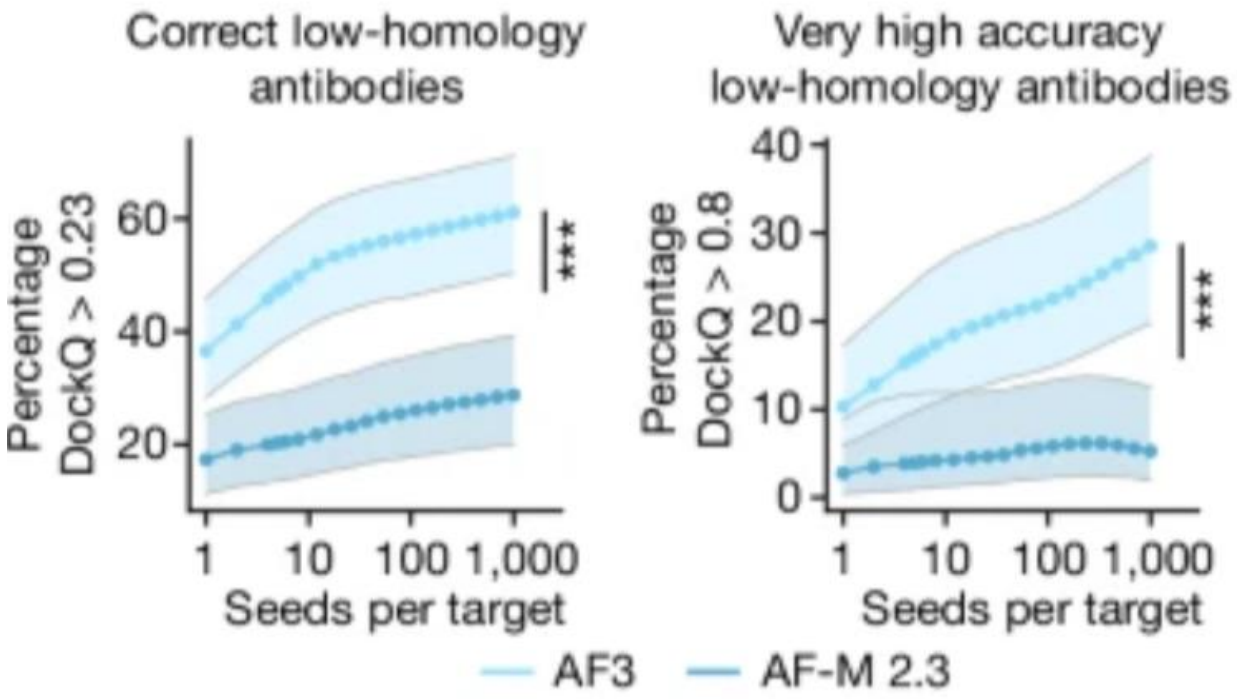


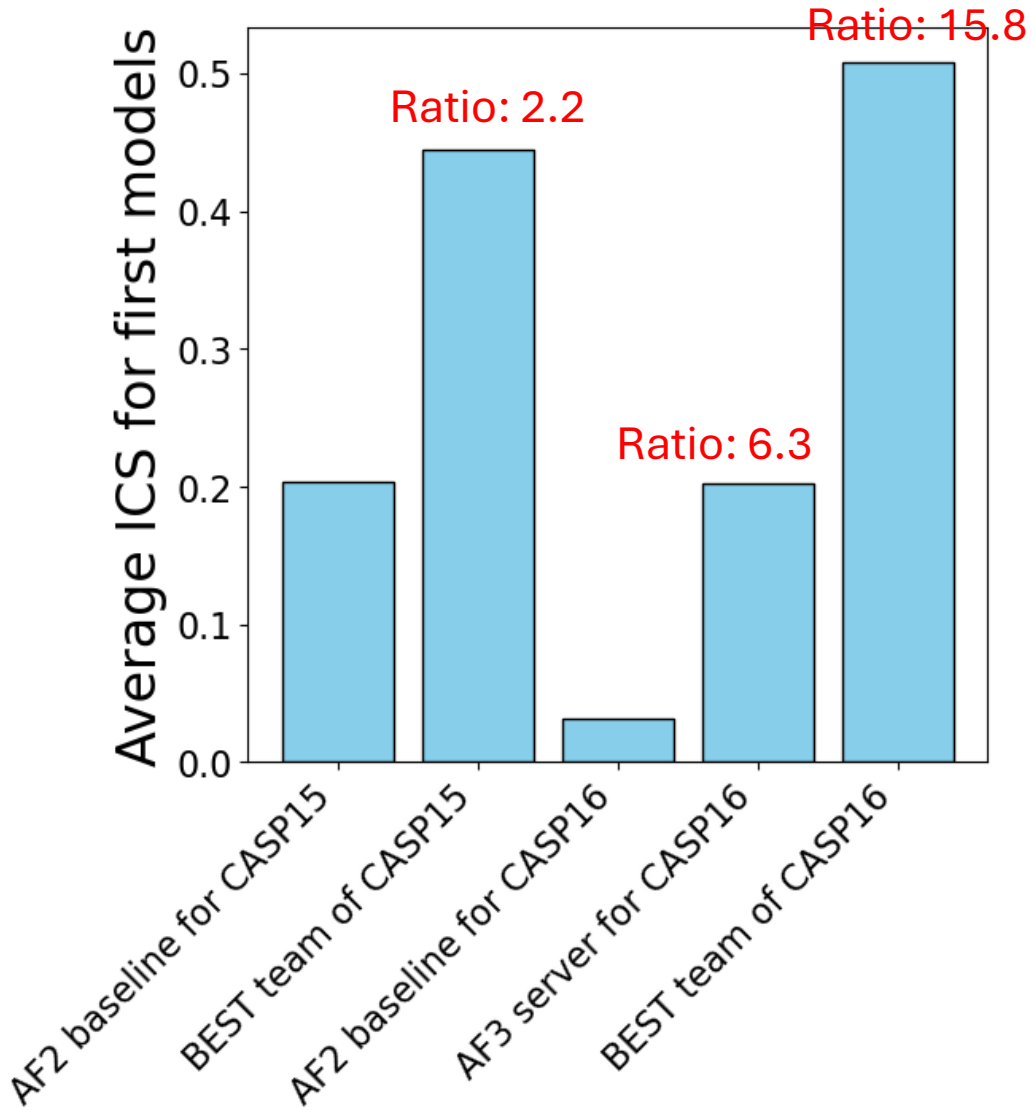
Final “ranking”(with ties) for oligomers
Best models over all phases





Huge progress in antibody-antigen interactions





Final thoughts on oligomer prediction

0. Exciting progress in antigen-antibody interactions. We may want more antibody targets in the future to more robustly evaluate the progress.

1. Protein complex modeling is not "solved": each group gets a subset correctly.

2. We think Phase 0 should be the future of oligomer prediction.

3. Evaluating targets with unknown stoichiometry needs better tools and we provided a start.

4. Weak interactions and multiple alternative interfaces are hard to predict.

5. Unseen flexibility in experimental structures could be an issue.