Analysis and evaluation of CASP16 protein monomer predictions

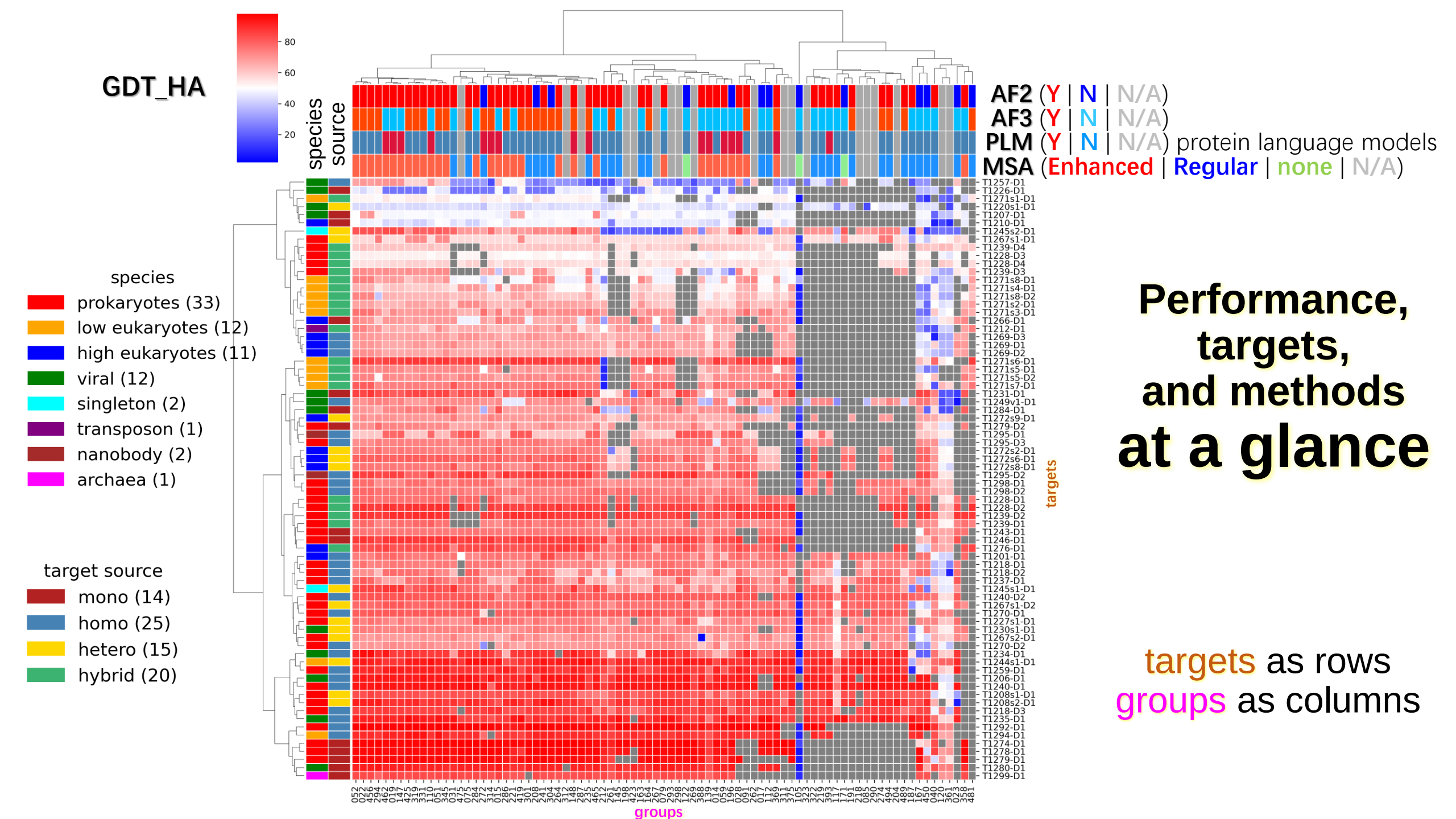
Rongqing Yuan, Jing Zhang, Andriy Kryshtafovych,
Nick V. Grishin, and Qian Cong

University of Texas Southwestern Medical Center

“CASP is an experiment, not a competition”

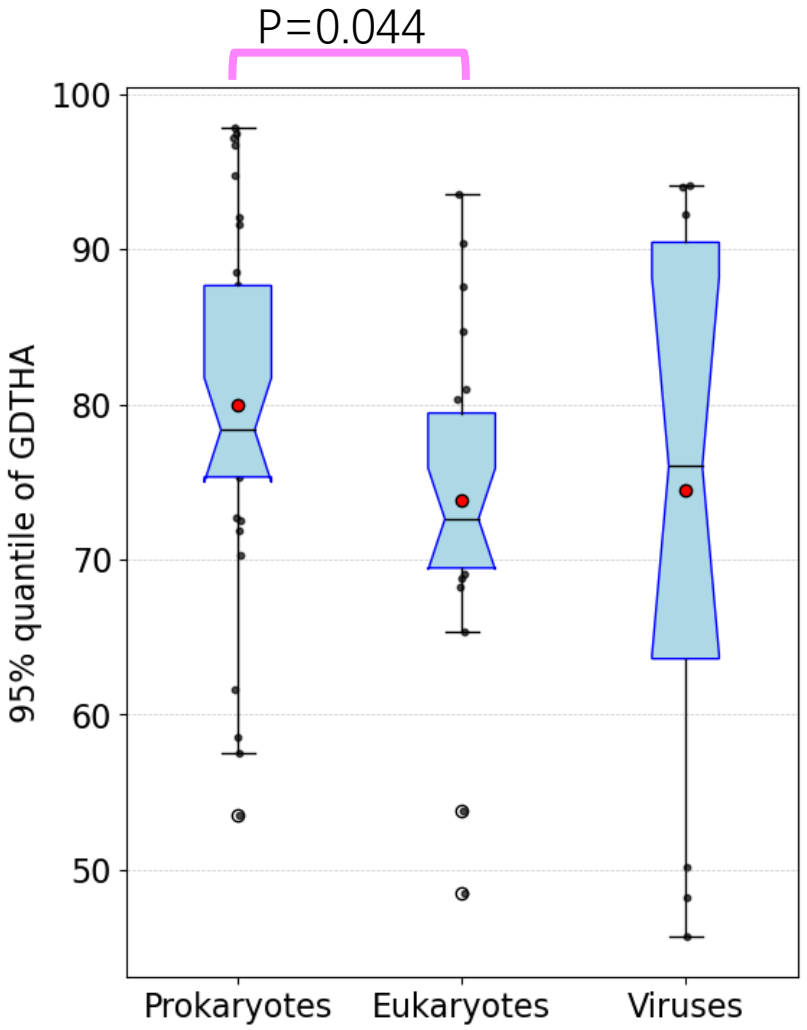
Experiment (insights) first

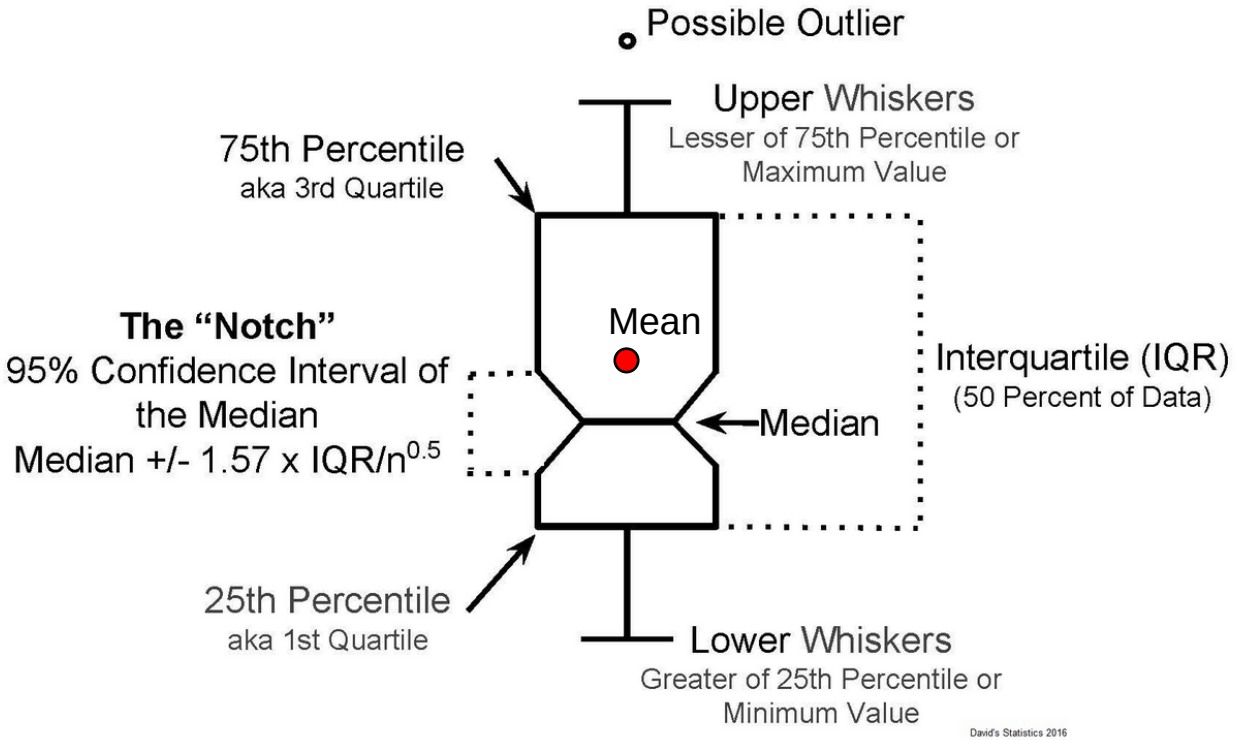
Competition (ranking) second


GDT_HA
80
60
40
20
species
source
AF2 (Y | N | N/A)
AF3 (Y | N | N/A)
PLM (Y | N | N/A) protein language models
MSA (Enhanced | Regular | none | N/A)
species
prokaryotes (33)
low eukaryotes (12)
high eukaryotes (11)
viral (12)
singleton (2)
transposon (1)
nanobody (2)
archaea (1)
target source
mono (14)
homo (25)
hetero (15)
hybrid (20)
targets
groups
Performance, targets, and methods at a glance
targets as rows
groups as columns

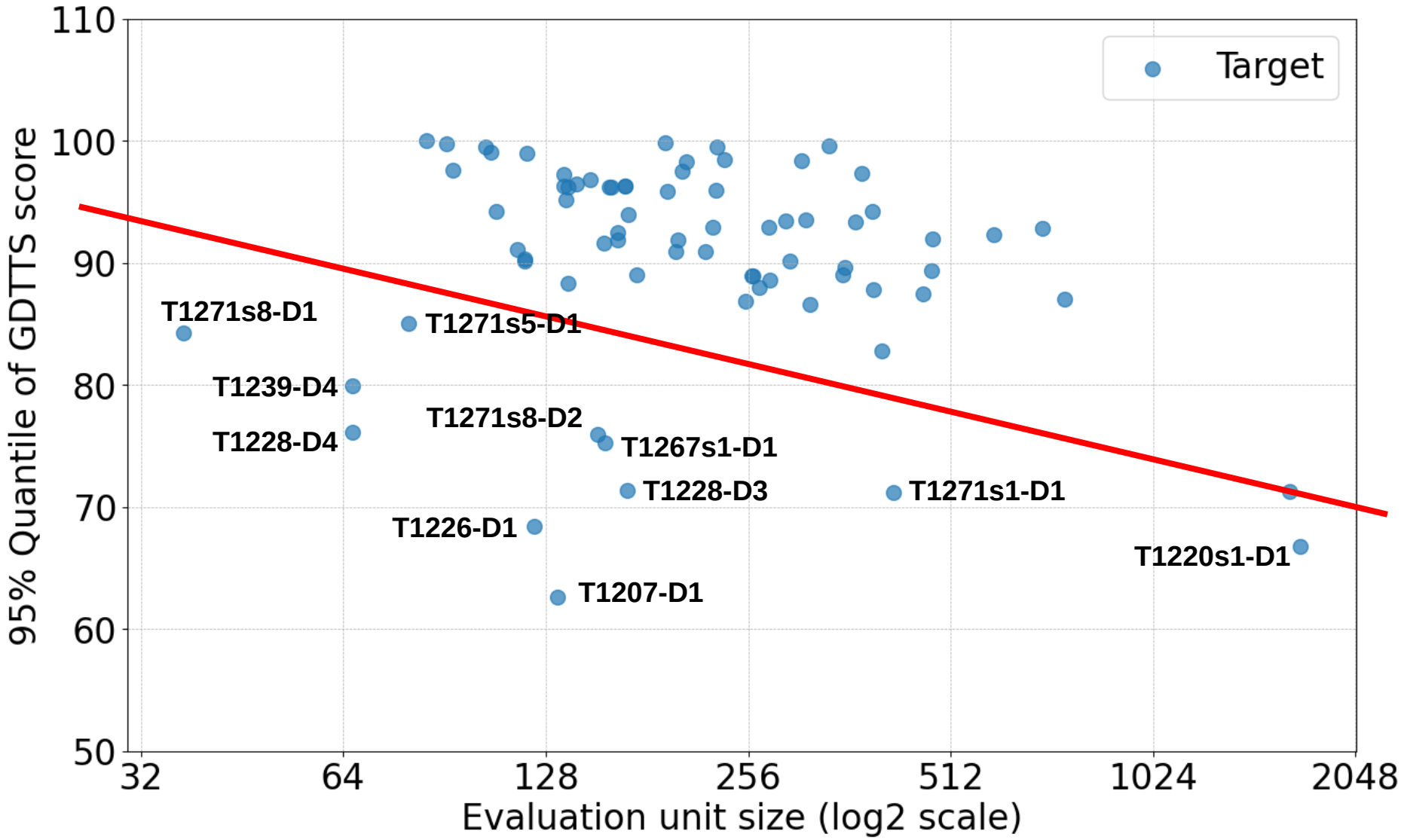
Common failures in monomer modeling

What monomers are still challenging?

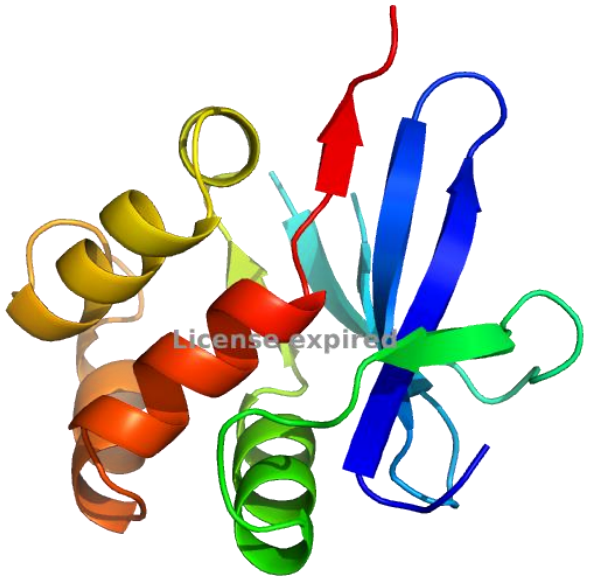




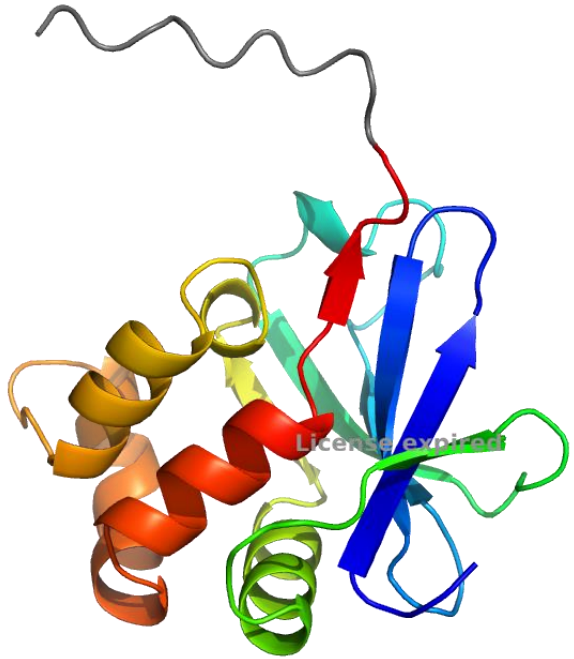
Poorly predicted monomers



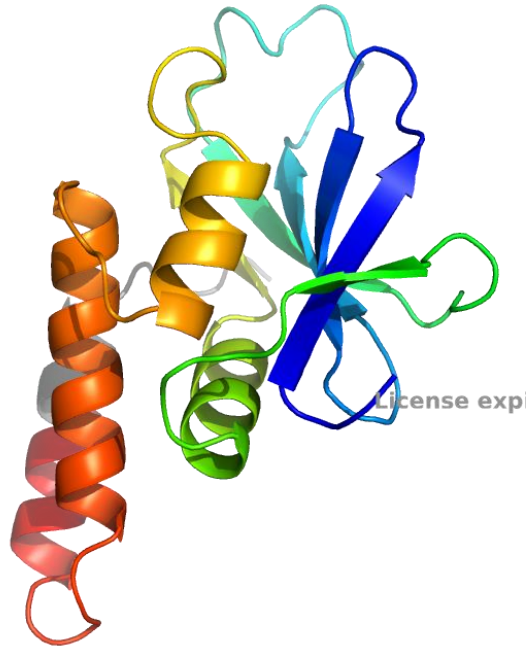
Poor performance by most groups on T1207-D1



Target

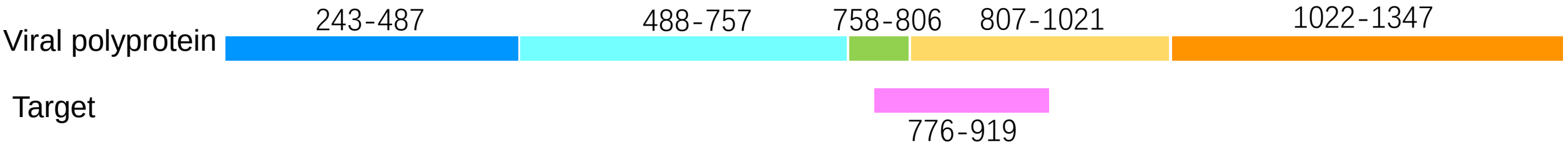


A winning model
from the Wallner group

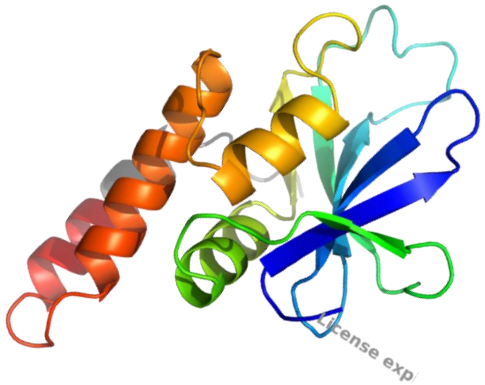


Most common model

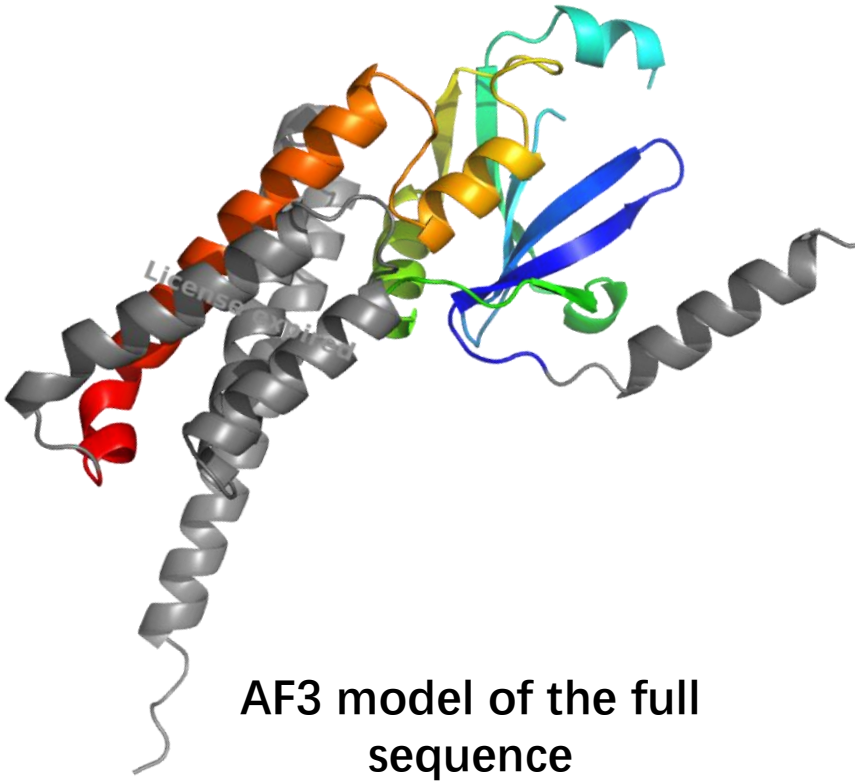
Challenge 1: alternative confirmation after truncation?



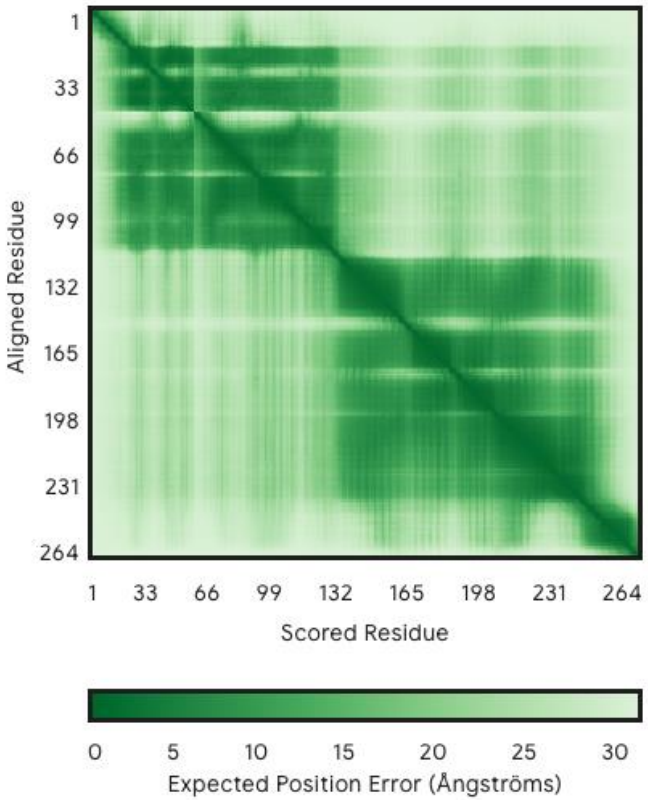
Target



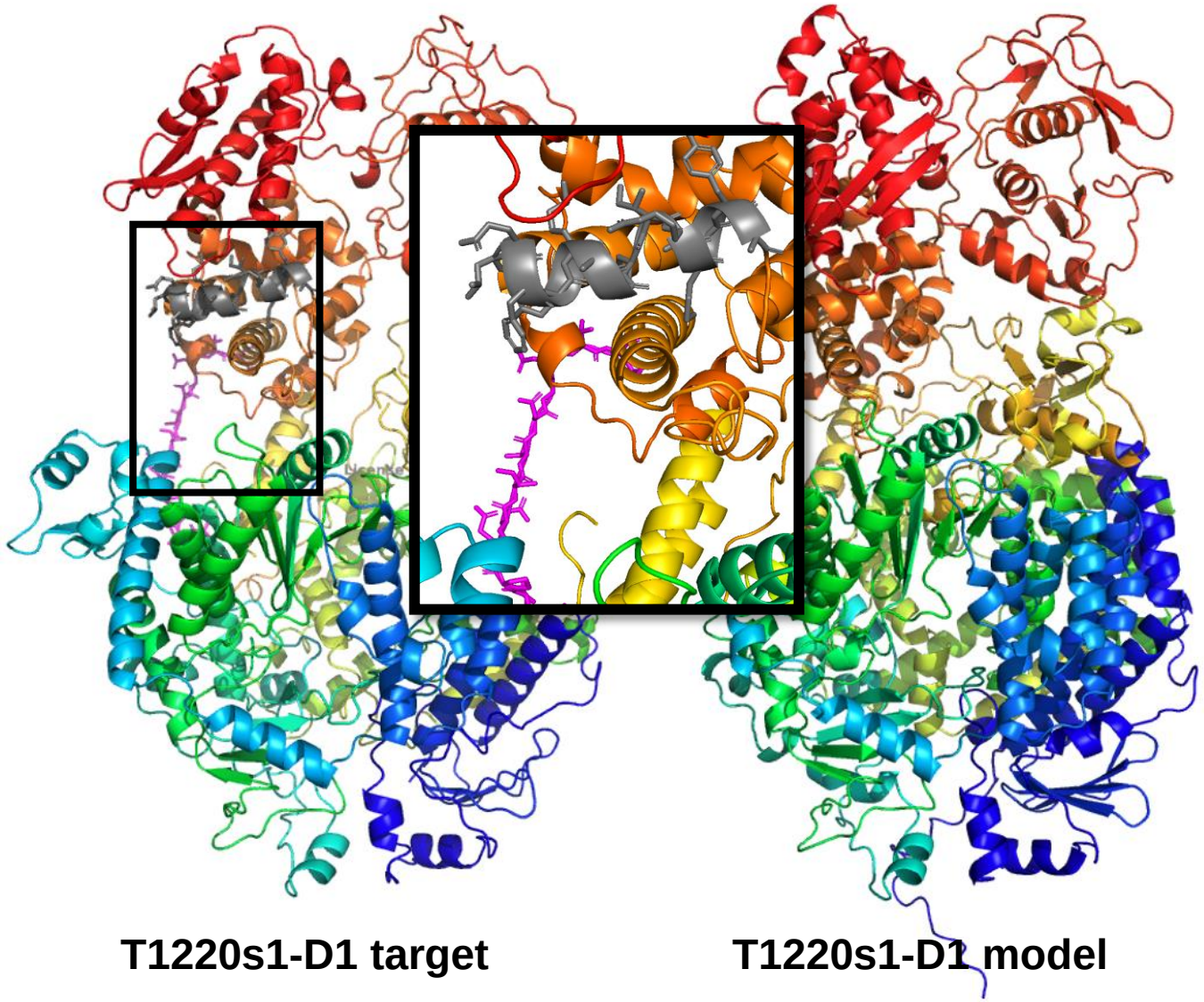
Common model



AF3 model of the full sequence



Challenge 2: poorly resolved experimental structures



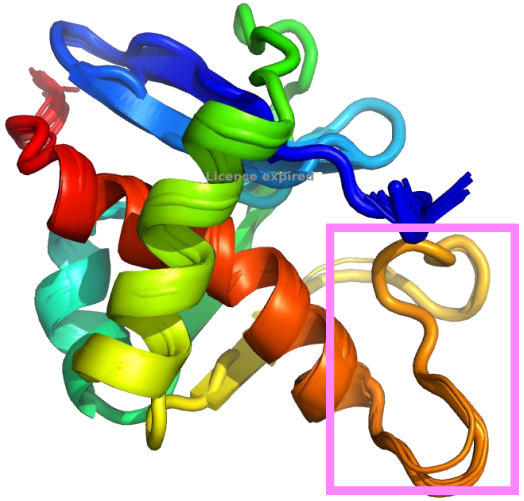
T1220s1-D1 target

T1220s1-D1 model

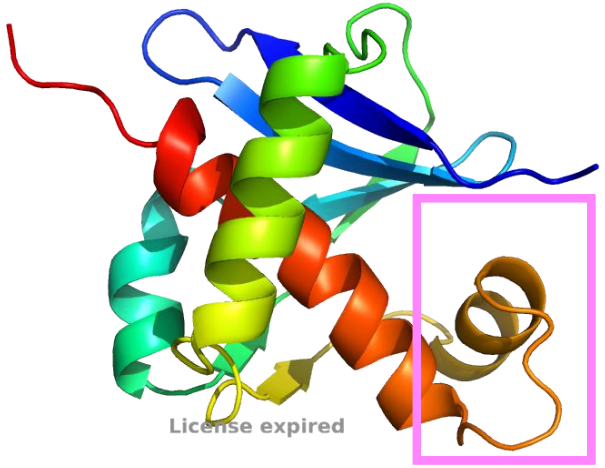
Clashscore, all atoms:	13.03		58th percentile* (N=1784, all resolutions)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	569	39.49%	Goal: <0.3%
Favored rotamers	648	44.97%	Goal: >98%
Ramachandran outliers	9	0.55%	Goal: <0.05%
Ramachandran favored	1493	90.76%	Goal: >98%
Rama distribution Z-s			abs(Z score) < 2
MolProbity score^	3.36		12th ercentile* (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å			0
Bad bonds:	0 / 13316	0.00%	Goal: 0%
Bad angles:	0 / 18102	0.00%	Goal: <0.1%
Cis Prolines:	0 / 87	0.00%	Expected: ≤1 per chain, or ≤5%
Twisted Peptides:	3 / 1648	0.18%	Goal: 0
CaBLAM outliers	98	6.0%	Goal: <1.0%
CA Geometry outliers	21	1.28%	Goal: <0.5%

Clashscore, all atoms:	2.28		99th percentile* (N=1784, all resolutions)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	33	2.21%	Goal: <0.3%
Favored rotamers	1401	93.84%	Goal: >98%
Ramachandran outliers	25	1.46%	Goal: <0.05%
Ramachandran favored	1563	91.46%	Goal: >98%
Rama distribution Z-score			abs(Z score) < 2
MolProbity score^	1.77		86th ercentile* (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å			0
Bad bonds:	12 / 13800	0.09%	Goal: 0%
Bad angles:	148 / 18756	0.79%	Goal: <0.1%
Cis Prolines:	2 / 91	2.20%	Expected: ≤1 per chain, or ≤5%
Twisted Peptides:	19 / 1710	1.11%	Goal: 0
CaBLAM outliers	28	1.6%	Goal: <1.0%
CA Geometry outliers	19	1.11%	Goal: <0.5%

Challenge 2: poorly resolved experimental structures

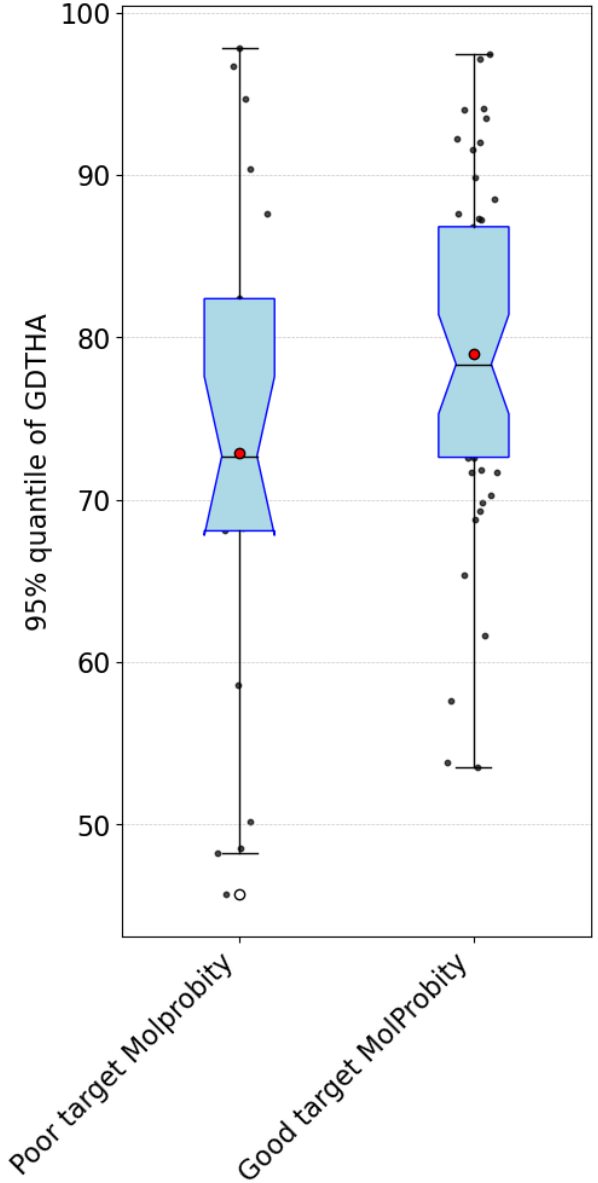


T1226-D1 target

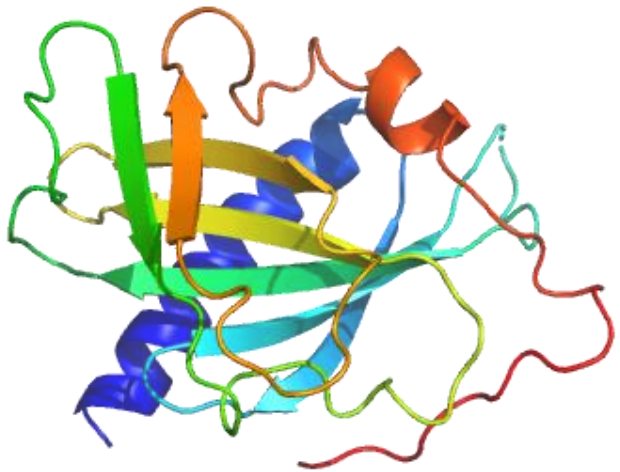


T1226-D1 model

Clashscore, all atoms:	49.14		4th percentile* (N=1784, all resolutions)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	21	21.21%	Goal: <0.3%
Favored rotamers	64	64.65%	Goal: >98%
Ramachandran outliers	1	0.83%	Goal: <0.05%
Ramachandran favored	110	90.91%	Goal: >98%
Rama distribution Z-scor			s(Z score) < 2
MolProbity score^	3.70		5th pe entile* (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å			
Bad bonds:	0 / 955	0.00%	Goal: 0%
Bad angles:	0 / 1291	0.00%	Goal: <0.1%
Cis Prolines:	0 / 3	0.00%	Expected: ≤1 per chain, or ≤5%
CaBLAM outliers	3	2.5%	Goal: <1.0%
CA Geometry outliers	1	0.84%	Goal: <0.5%

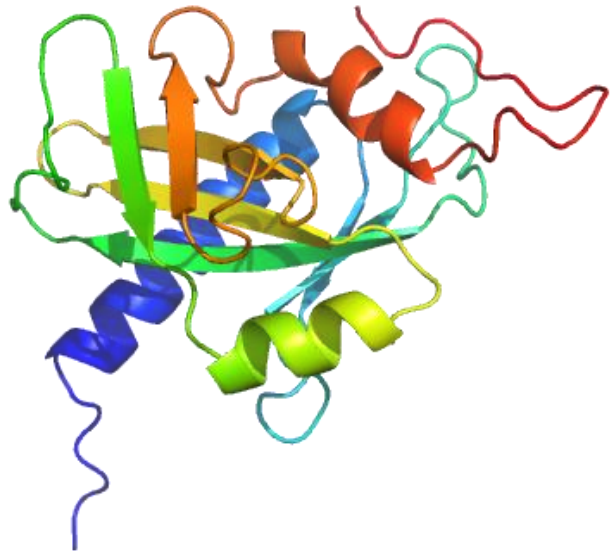


Poor performance by most groups on T1267s1-D1



Target

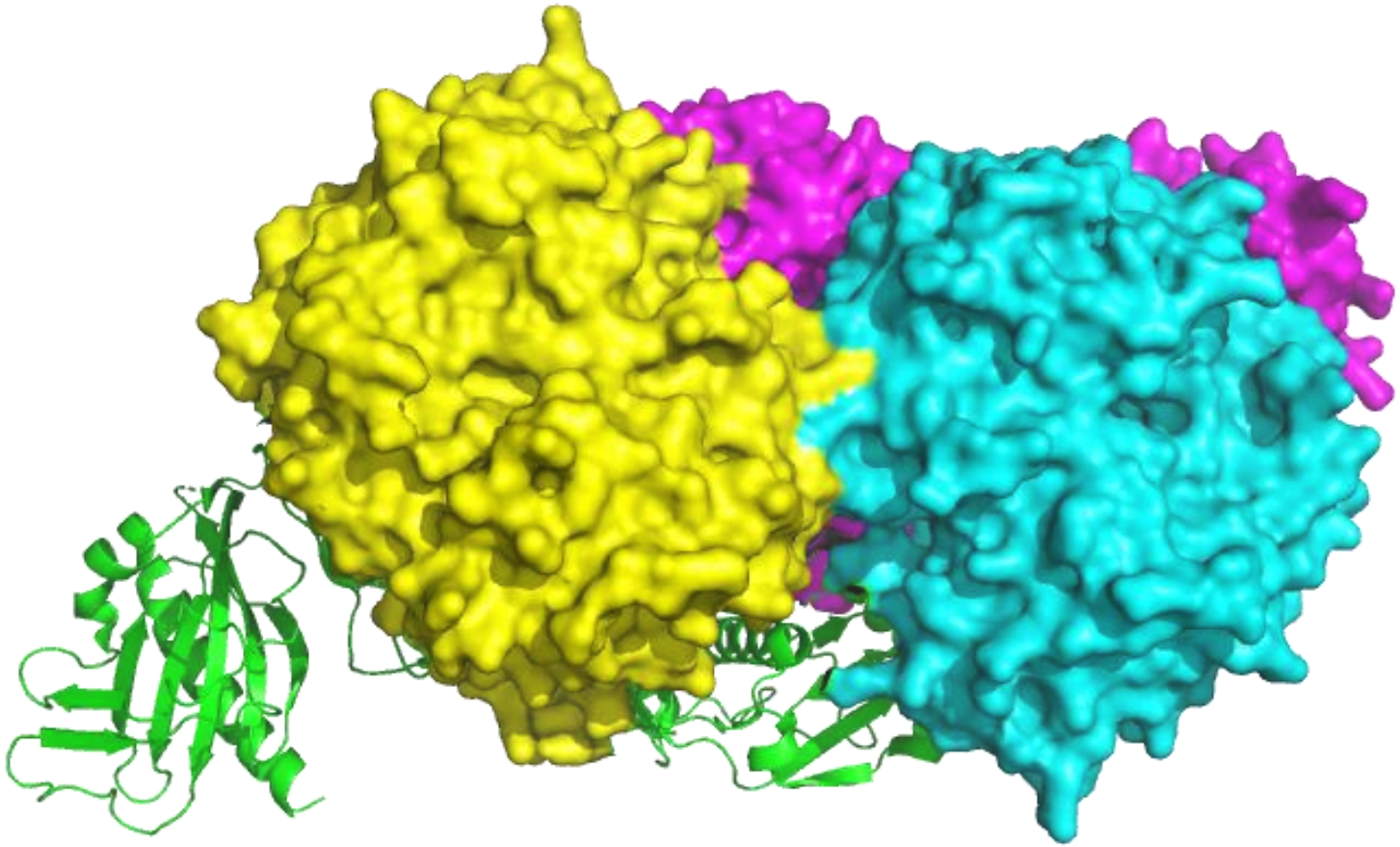
A winning model
from the Kozakov group



Most common model

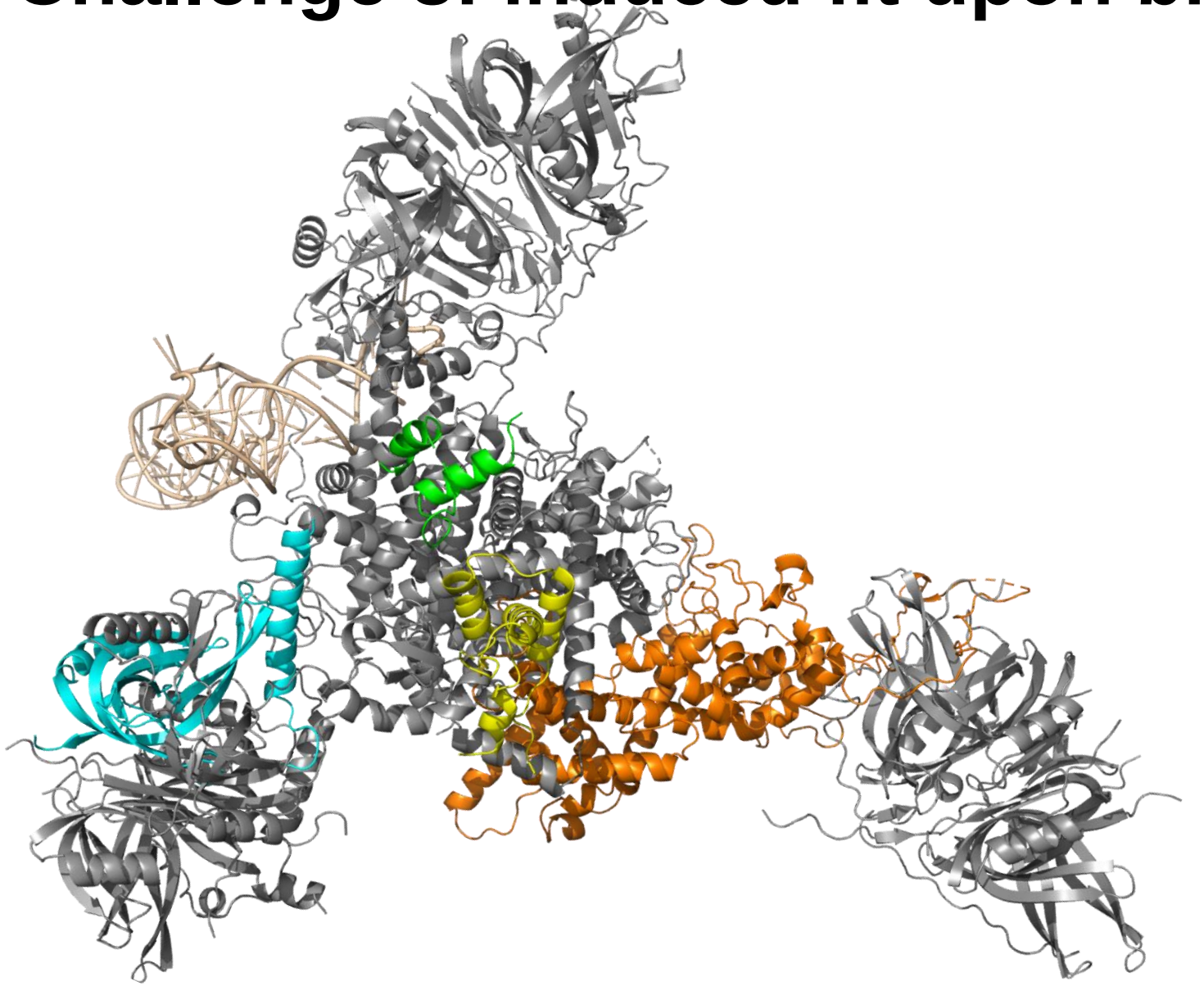
Challenge 3: induced fit upon binding to partners

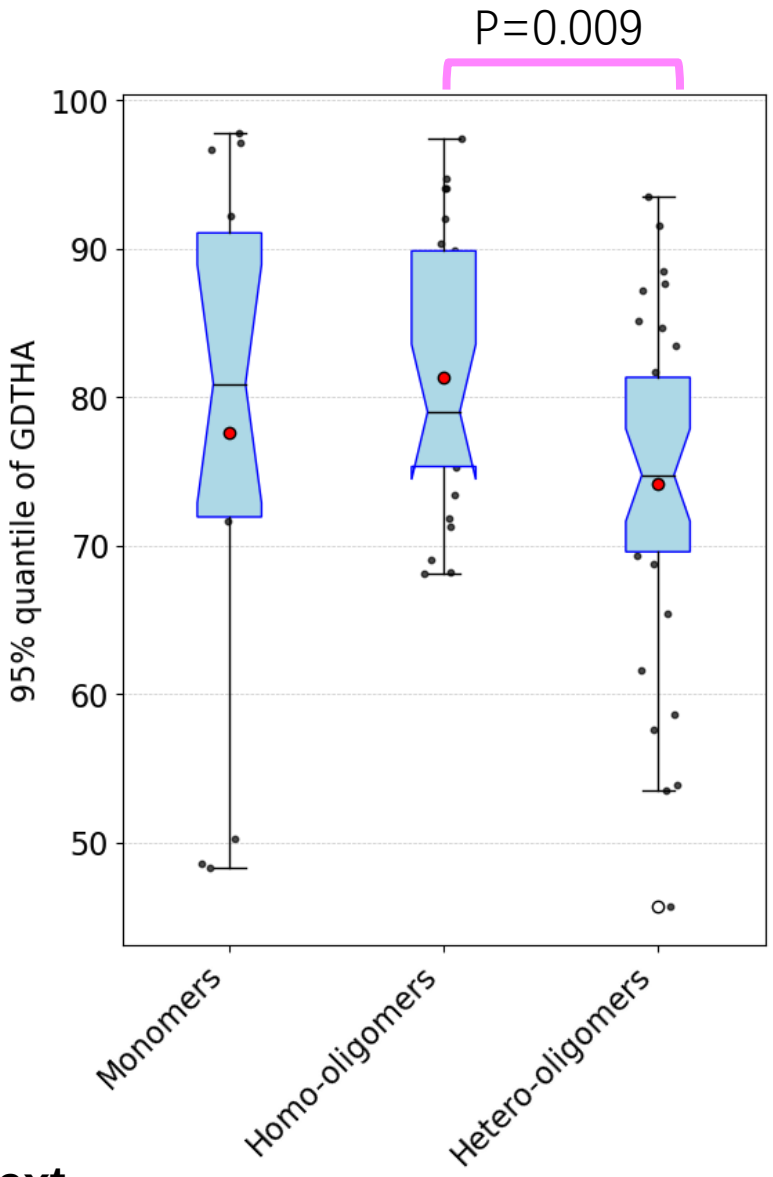
T1267s1-D1 target



T1267s1-D1 in its context

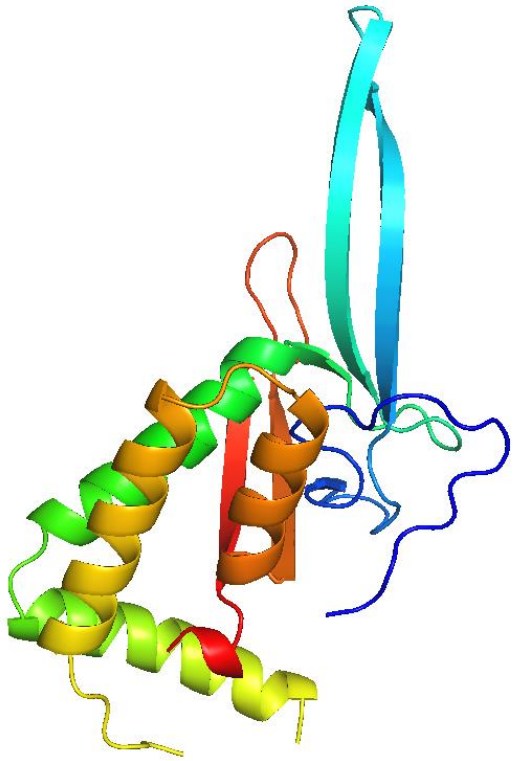
Challenge 3: induced fit upon binding to partners





T1271s1-D1, **T1271s5-D1**, **T1271s8-D1**, and **T1271s8-D2** in their context

Challenge 4: irregularities in secondary structures

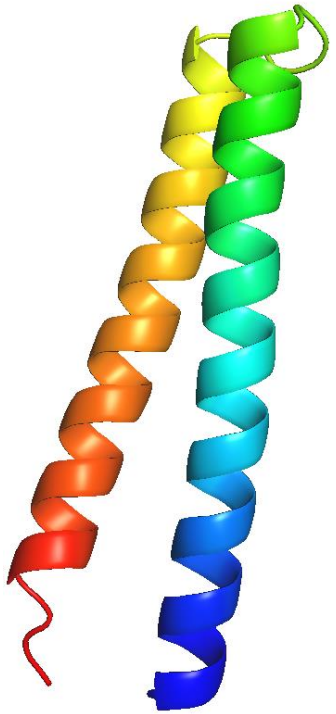


T1228-D3 target

T1228-D3 representative model

T1228-D3 in its context

Challenge 4: irregularities in secondary structures



T1228-D4 target

T1228-D4 representative model

T1228-D4 in its context

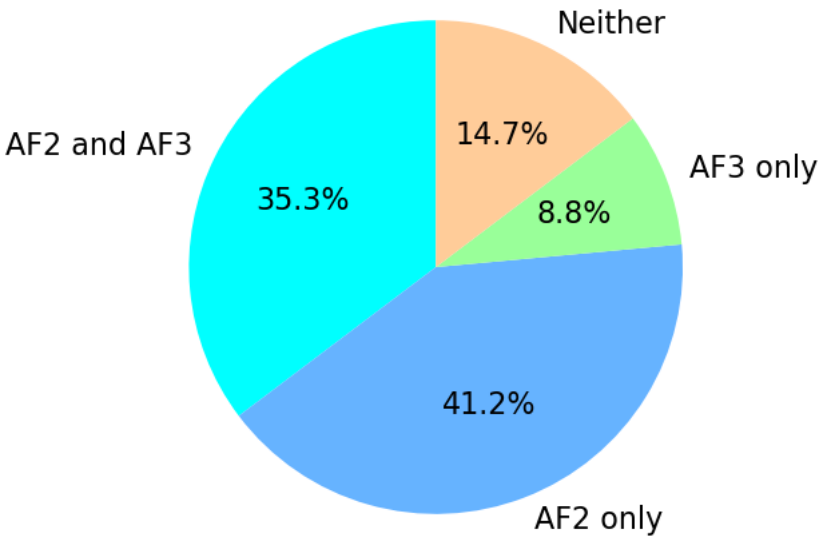
T1228 target has good MolProbity scores

Clashscore, all atoms:	1.46		99th percentile* (N=1784, all resolutions)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Poor rotamers	2	1.23%	Goal: <0.3%
Favored rotamers	149	91.41%	Goal: >98%
Ramachandran outliers	0	0.00%	Goal: <0.05%
Ramachandran favored	152	91.57%	Goal: >98%
Rama distribution Z-score	-2.92 ± 0.55		Goal: abs(Z score) < 2
MolProbity score^	1.45		96th percentile* (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å	1	0.60%	Goal: 0
Bad bonds:	0 / 1385	0.00%	Goal: 0%
Bad angles:	11 / 1864	0.59%	Goal: <0.1%
Cis Prolines:	0 / 6	0.00%	Expected: ≤1 per chain, or ≤5%
Twisted Peptides:	1 / 168	0.60%	Goal: 0
CaBLAM outliers	7	4.3%	Goal: <1.0%
CA Geometry outliers	3	1.85%	Goal: <0.5%

Winning strategies for monomer modeling

Common strategies used by predictors

68 groups provided abstract

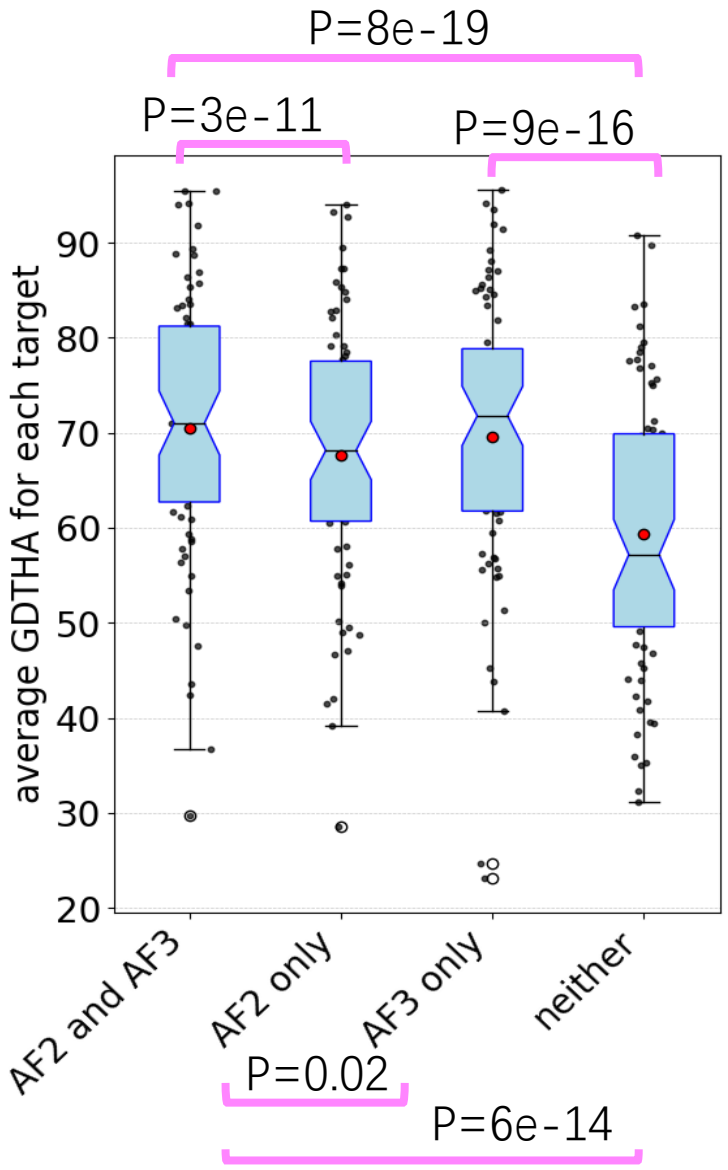


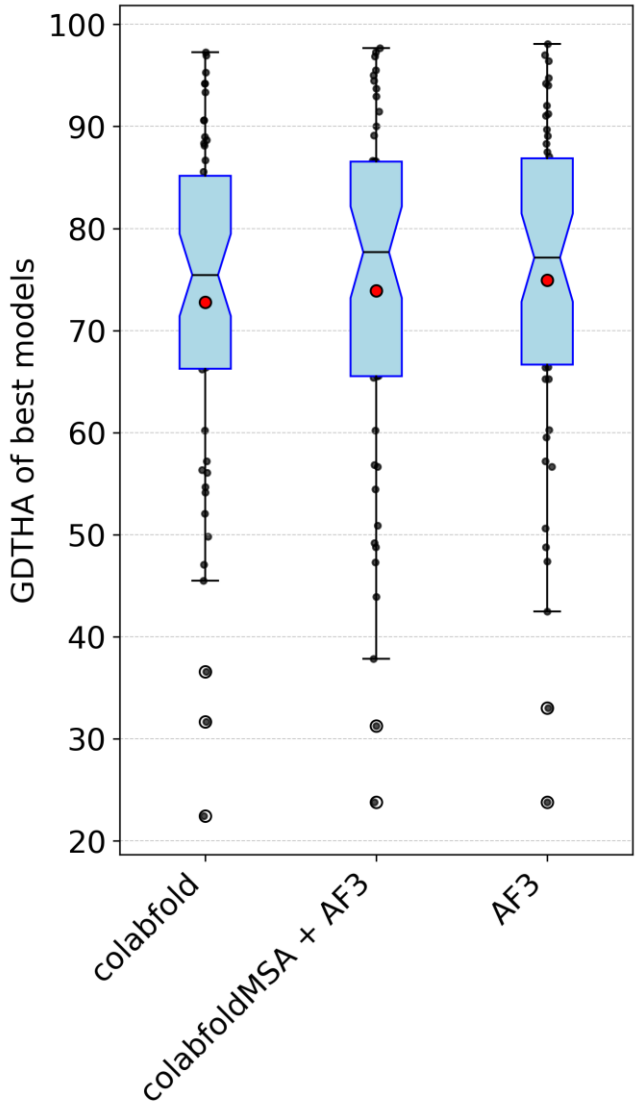
Group Name	Group ID	Type	AF2	AF3	MSA	LM	Group Name	Group ID	Type	AF2	AF3	MSA	LM
Cool-PSP	TS014	H	Y	N	S	N	XGroup-server	TS219	S	Y	N	Y	N
PEZYFoldings	TS015	H	Y	N	S	Y	CSSB_FAKER	TS221	H	Y	N	S	N
Seder2024hard	TS017	H	N	N	Y	N	isyslab-hust	TS235	H	Y	Y	Y	Y
Zheng-Server	TS019	S	Y	N	S	Y	elofsson	TS241	H	Y	Y	Y	N
Yang	TS022	H	Y	Y	S	N	UNRES	TS261	H	Y	Y	Y	N
FTBiot0119	TS023	H	N	N	Y	N	GuijunLab-Human	TS264	H	Y	Y	S	N
NKRNA-s	TS028	S	Y	N	S	Y	GromihaLab	TS272	H	N	Y	Y	Y
MassiveFold	TS031	H	Y	N	Y	N	Unicorn	TS284	S	Y	N	S	N
DELCLAB	TS040	H	Y	N	Y	N	CSSB_experimental	TS286	H	Y	Y	S	N
Yang-Server	TS052	S	Y	Y	S	N	KiharaLab	TS294	H	Y	Y	S	N
DeepFold	TS059	H	Y	N	S	Y	ARC	TS300	S	Y	N	Y	Y
GHZ-ISM	TS075	S	Y	Y	Y	N	GHZ-MAN	TS301	H	Y	Y	Y	N
MRAFold	TS079	S	Y	N	S	N	AF3-server	TS304	S	N	Y	Y	N
Huang-HUST	TS091	H	Y	Y	S	N	GuijunLab-PAthreader	TS314	S	Y	Y	S	Y
PFSC-PFVM	TS105	H	N	N	N	N	MULTICOM_LLM	TS319	S	Y	Y	S	N
MIEnsembles-Server	TS110	S	Y	N	S	Y	XGroup	TS322	H	Y	N	Y	N
Seder2024easy	TS112	H	N	N	Y	N	MULTICOM_AI	TS331	S	Y	Y	S	N
COAST	TS114	S	Y	N	Y	Y	APOLLO	TS337	S	Y	N	Y	Y
Vakser	TS117	H	Y	N	Y	N	GeneSilico	TS338	H	N	Y	Y	N
MQA_server	TS122	S	N	N	N	N	MULTICOM_human	TS345	H	Y	Y	S	N
DeepFold-refine	TS139	H	Y	N	S	Y	PerezLab_Gators	TS358	H	Y	N	S	N
colabfold_baseline	TS145	S	Y	N	Y	N	Bhattacharya	TS369	H	Y	Y	S	N
Zheng-Multimer	TS147	S	Y	N	S	Y	DeepFold-server	TS388	S	Y	N	S	Y
Guijunlab-Complex	TS148	S	Y	Y	S	Y	GuijunLab-QA	TS393	H	Y	N	Y	Y
MultiFOLD2	TS163	S	Y	N	Y	N	CSSB-Human	TS419	H	Y	Y	S	N
McGuffin	TS164	H	Y	Y	Y	N	MULTICOM_GATE	TS425	S	Y	Y	S	N
OpenComplex	TS167	H	N	N	Y	N	OpenComplex_Server	TS450	S	N	N	Y	N
ChaePred	TS171	H	N	N	N	N	Yang-Multimer	TS456	S	Y	Y	S	N
Ayush	TS187	H	Y	N	Y	N	Zheng	TS462	H	Y	N	S	Y
LCBio	TS189	H	N	Y	Y	N	Wallner	TS465	H	Y	N	S	N
Schneidman	TS191	H	Y	Y	Y	N	Pcons	TS471	S	N	N	Y	N
HYU_MLLAB	TS196	H	N	N	S	Y	Vfold	TS481	H	N	Y	Y	N
falcon2	TS208	S	N	Y	Y	N	Fernandez-Recio	TS489	H	Y	Y	Y	N
PIEFold_human	TS212	H	Y	N	Y	N	ClusPro	TS494	H	Y	Y	Y	N

At least 31 (45%) groups worked on better MSAs
At least 18 (24%) groups used PLMs (none is purely PLM-based)

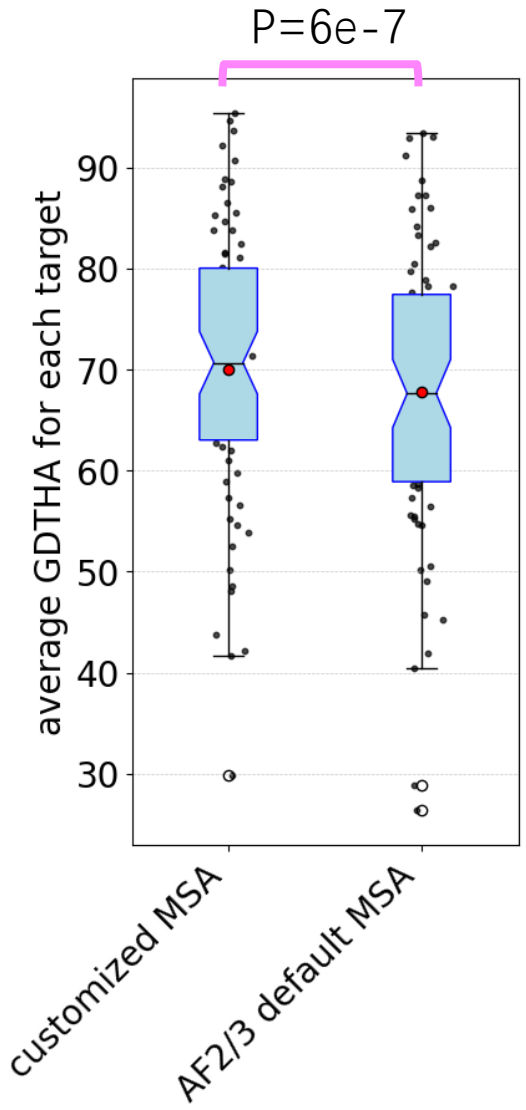
Type: H/S - human/server
MSA: Y/S/N - yes/special/no

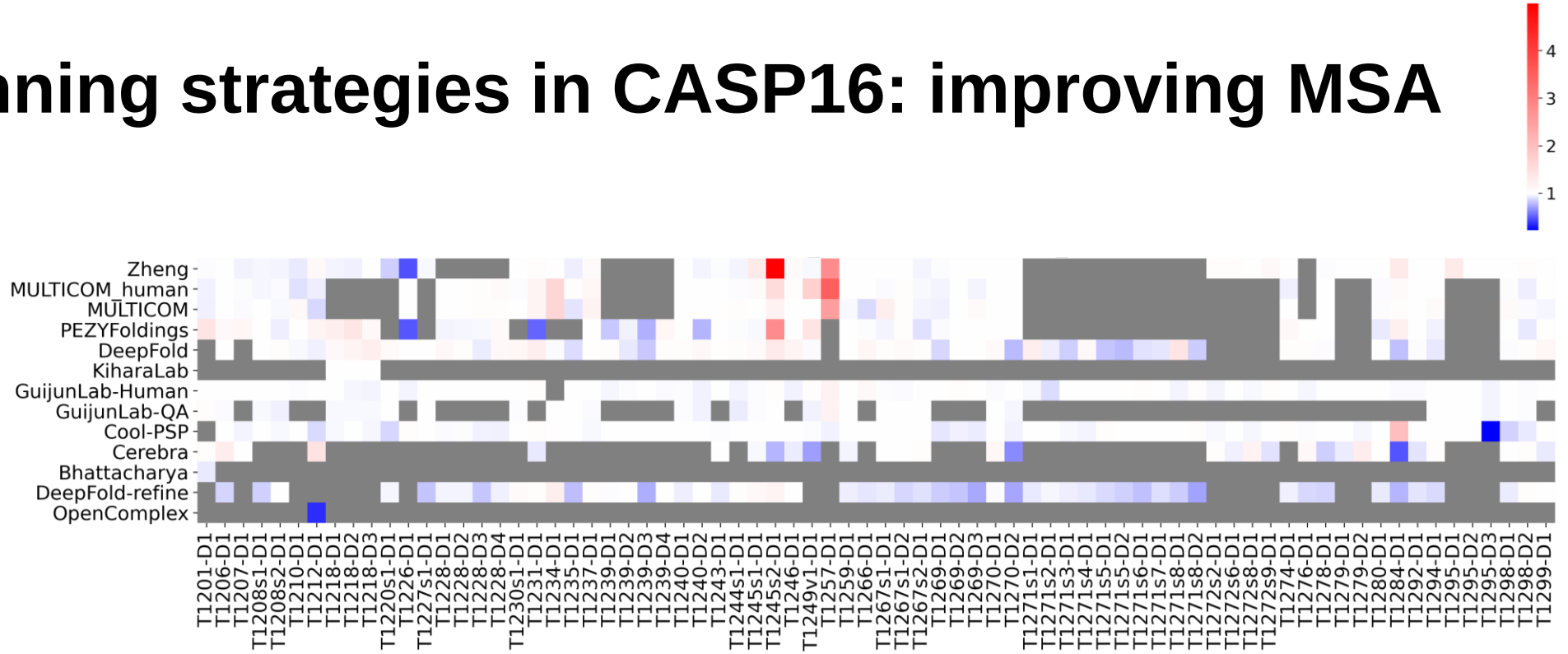
General winning strategies in CASP16: using AF3





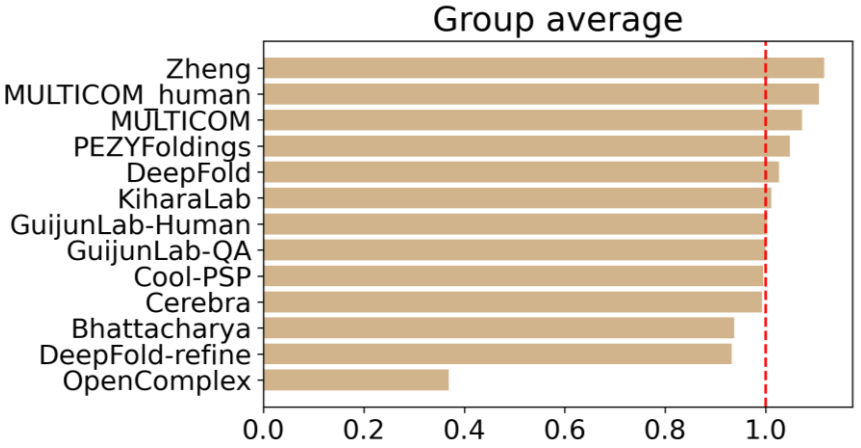
General winning strategies in CASP16: improving MSA





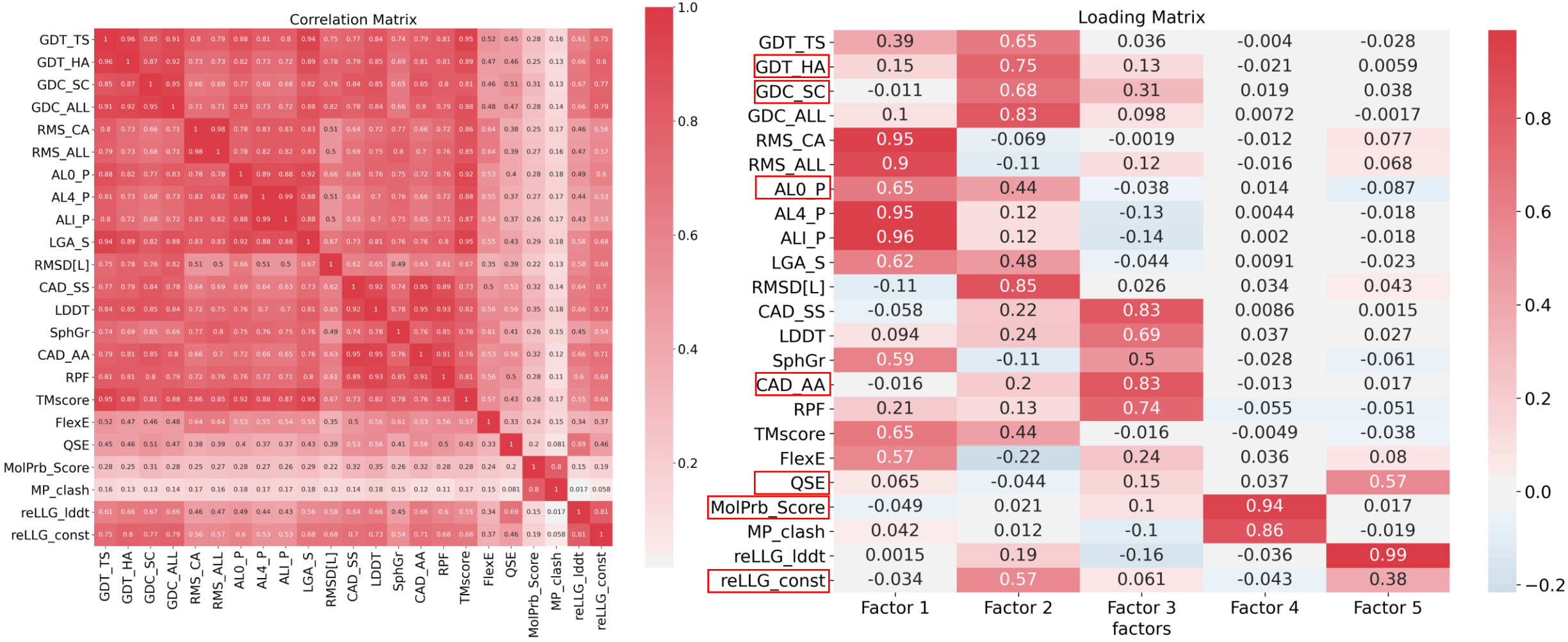
Model 1 GDT_HA divided by model 6 GDT_HA

The same method, but different MSAs



Ranking

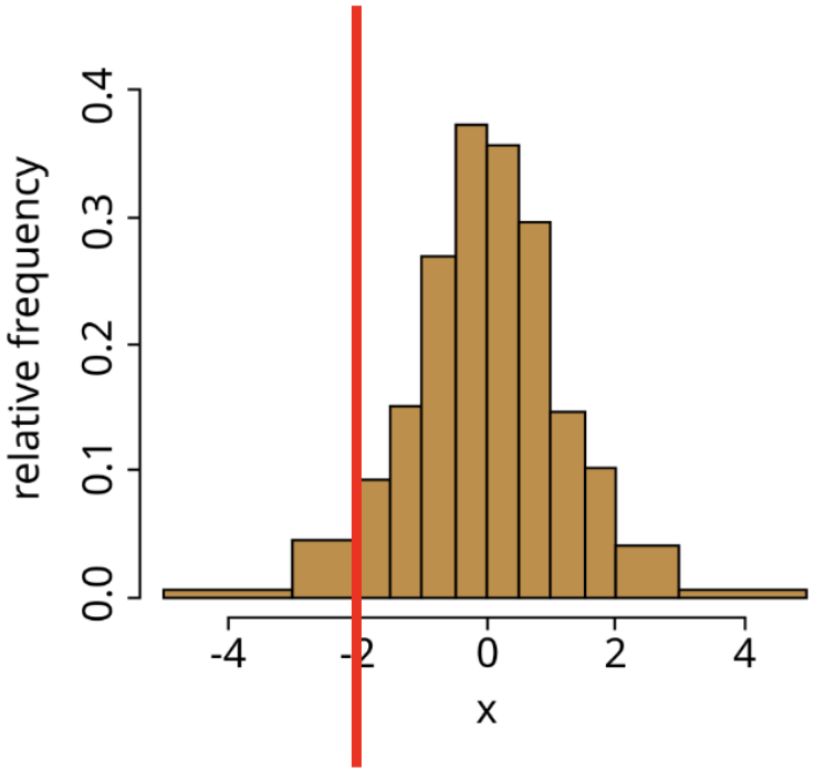
Comparison and factorization of scores



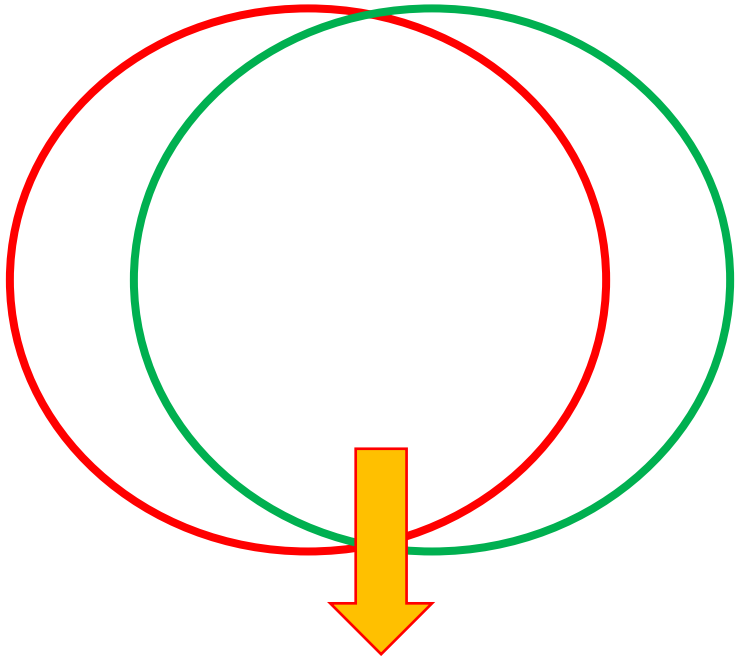
Final empirical scoring formula

$$\mathbf{Z(monomer)} = \sum_{\text{target}} \Bigg(\sum_{\text{EU}} w_{\text{EU}} \Big(\frac{1}{16} \times \big(Z(\text{GDC\_SC}) + Z(\text{AL0\_P}) + Z(\text{lDDT}) + Z(\text{MolProbity} - \text{Score}) \big) + \frac{1}{8} \times \big(Z(\text{SphGr}) + Z(\text{CAD}_{\text{AA}}) \big) + \frac{1}{6} \times \big(Z(\text{GDT\_HA}) + Z(\text{QSE}) + Z(\text{reLLG\_const}) \big) \Big) \Bigg), w_{\text{EU}} = \ \mathbf{1/EUs\ in\ this\ target}$$

Trimming off lower-end tails and imputing missing values



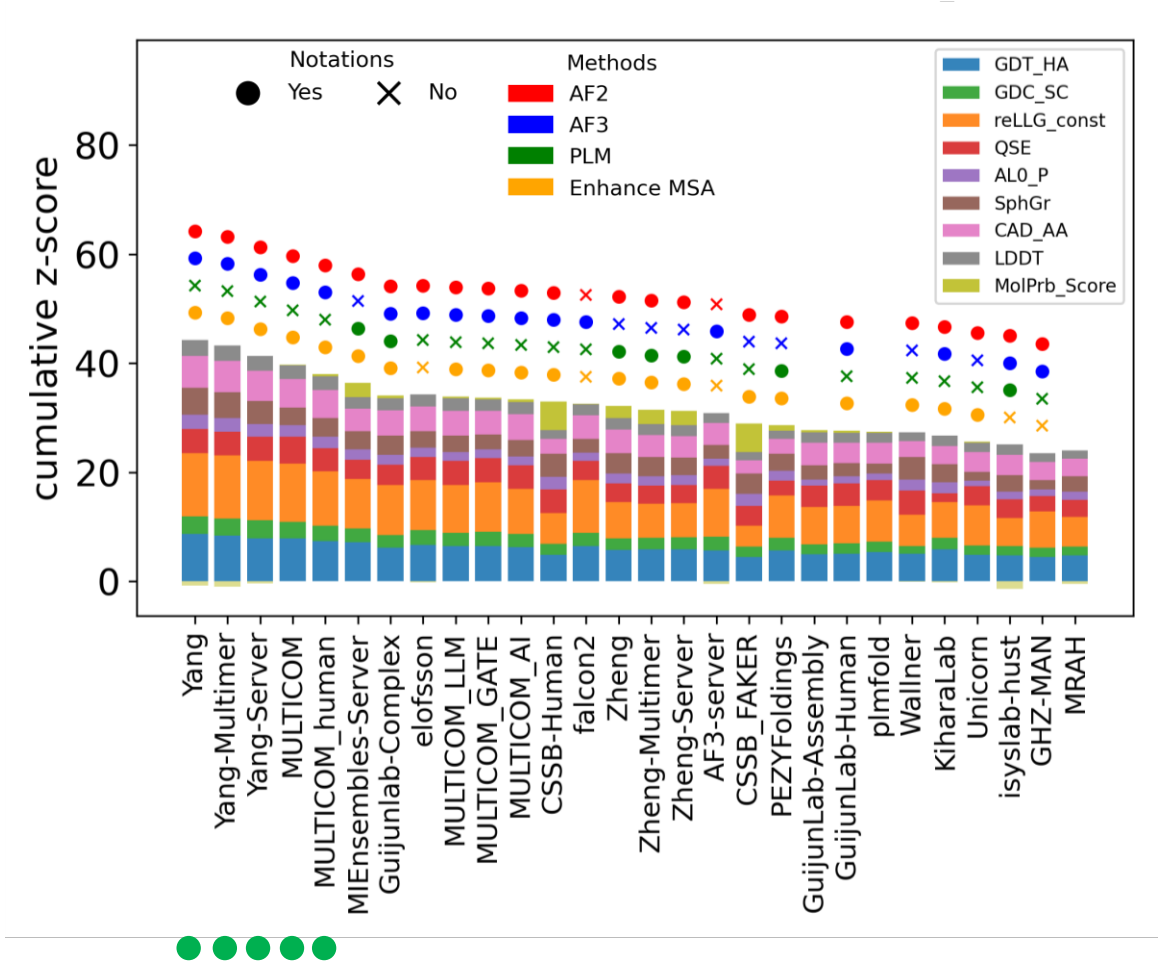
Head-to-head tests

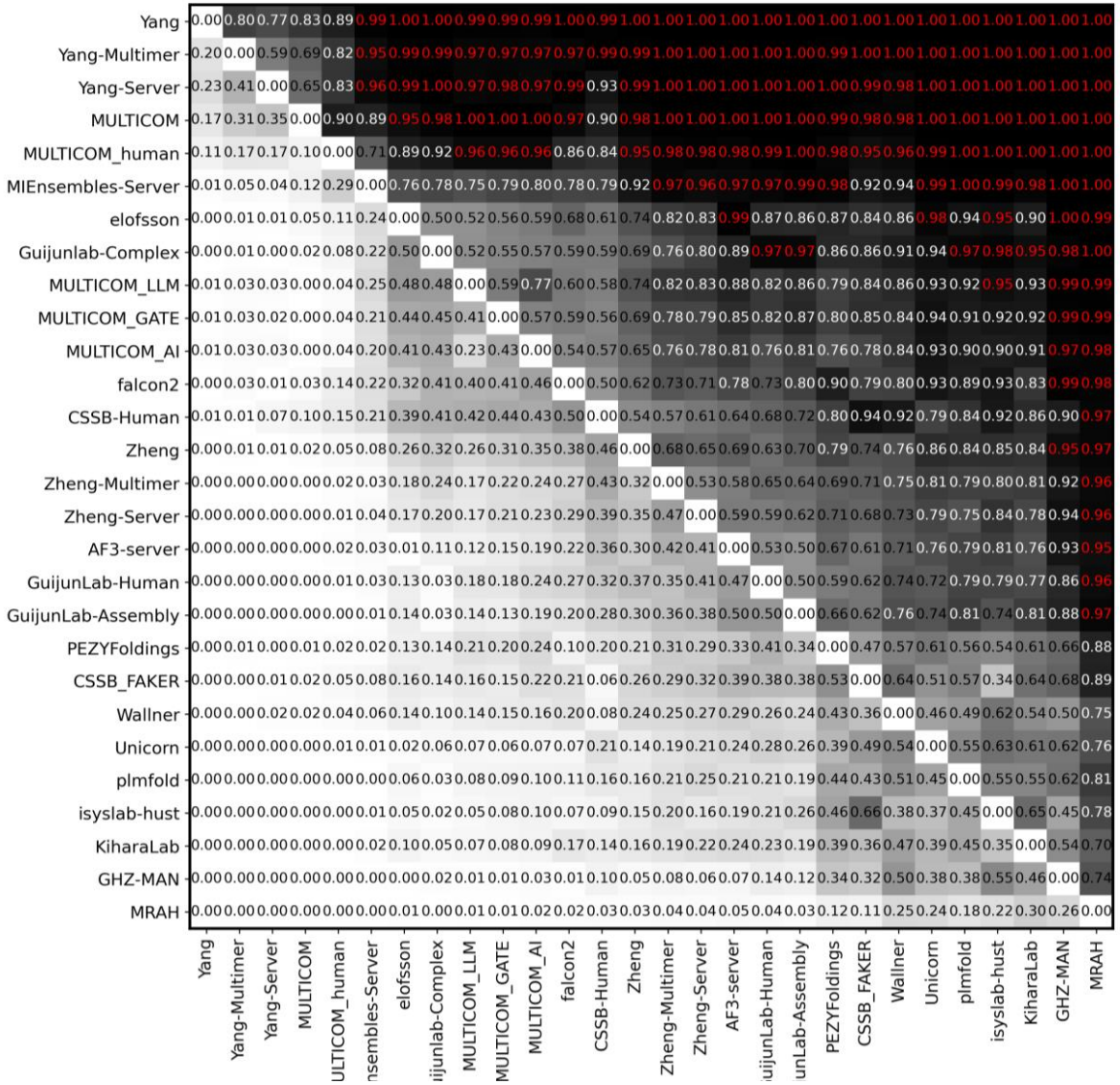


Bootstrap

Morale: want to win, please, please, predict all targets

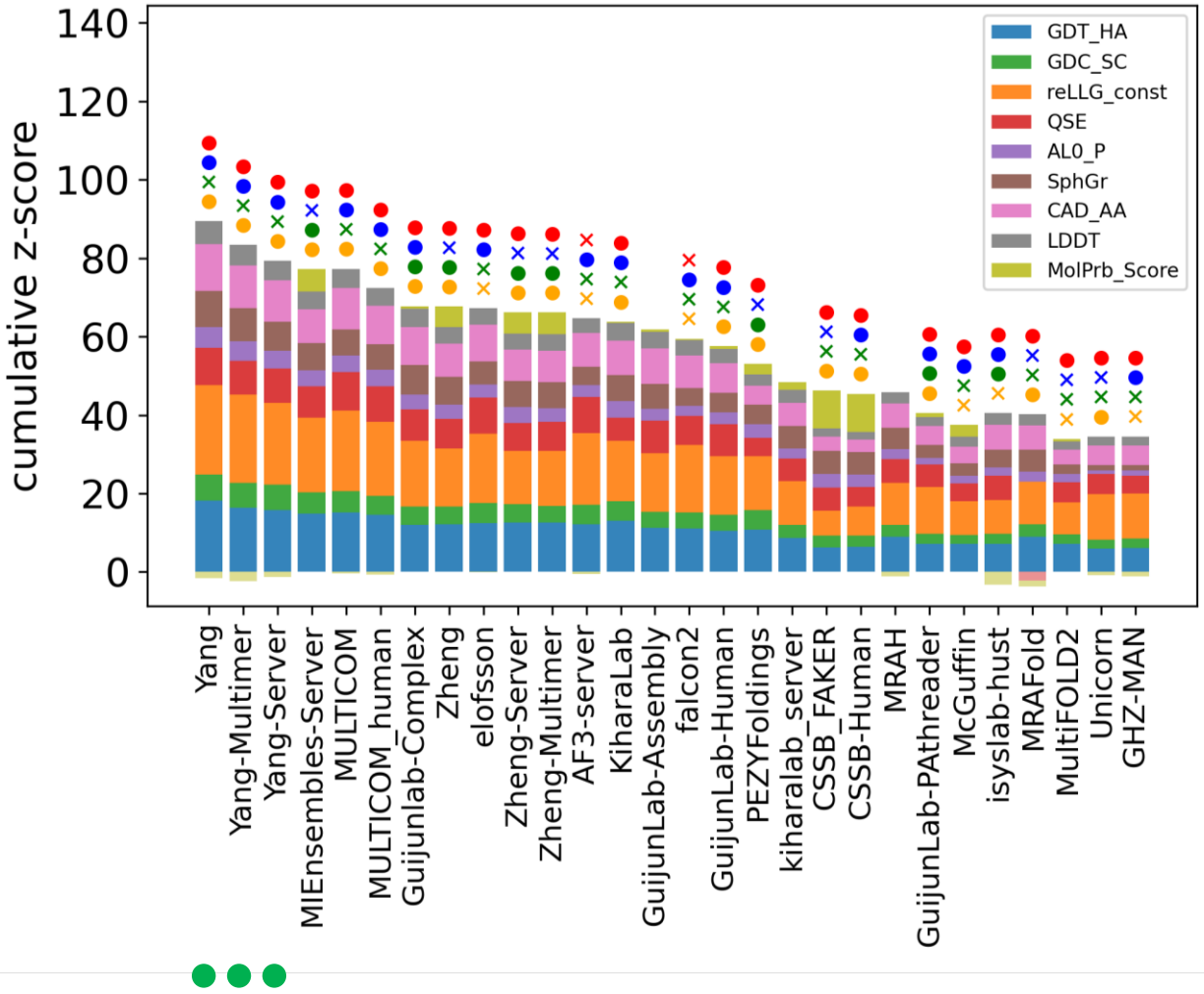
Ranking by best models for phase 1 targets

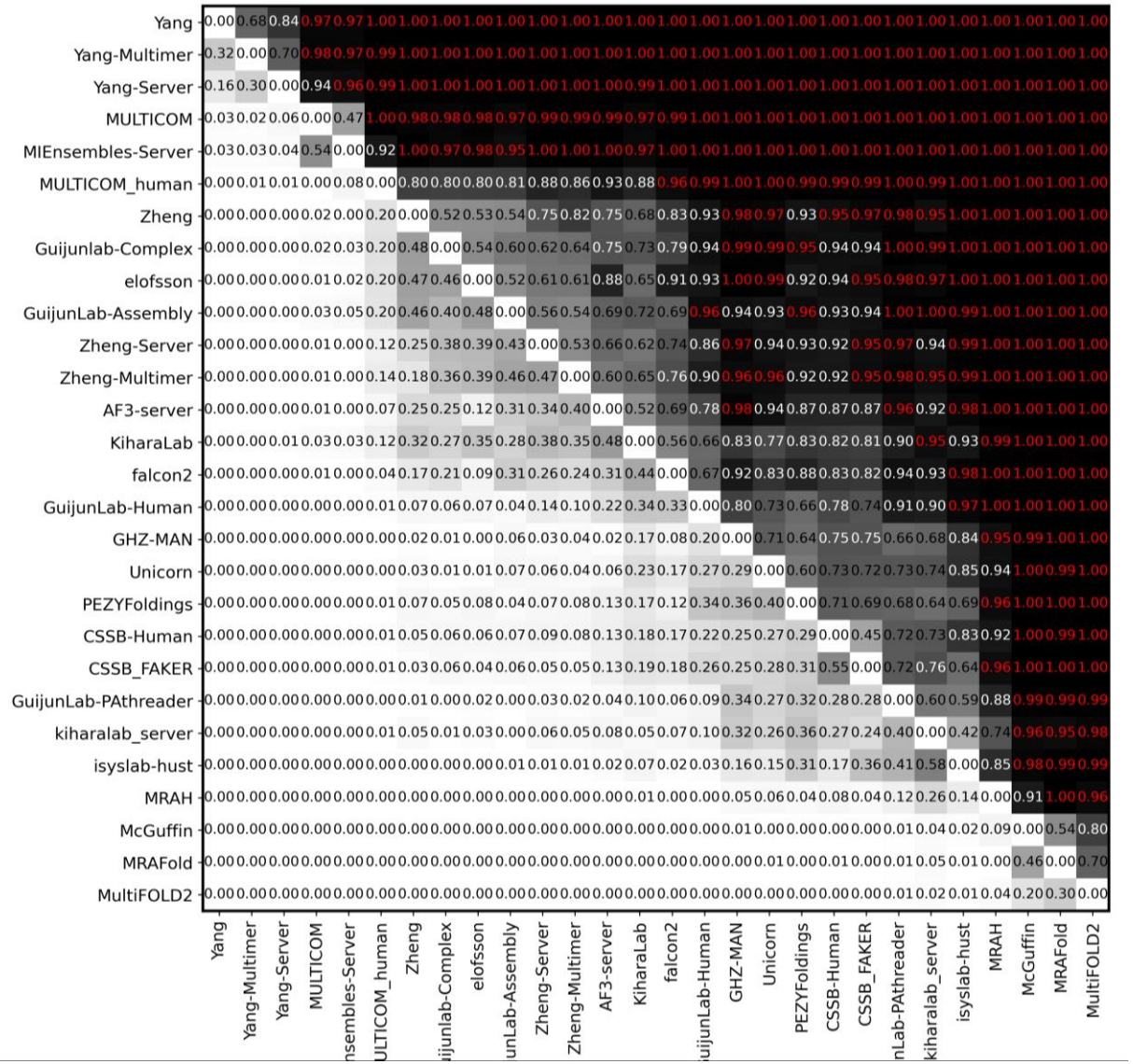




statistical significance (red is significant)

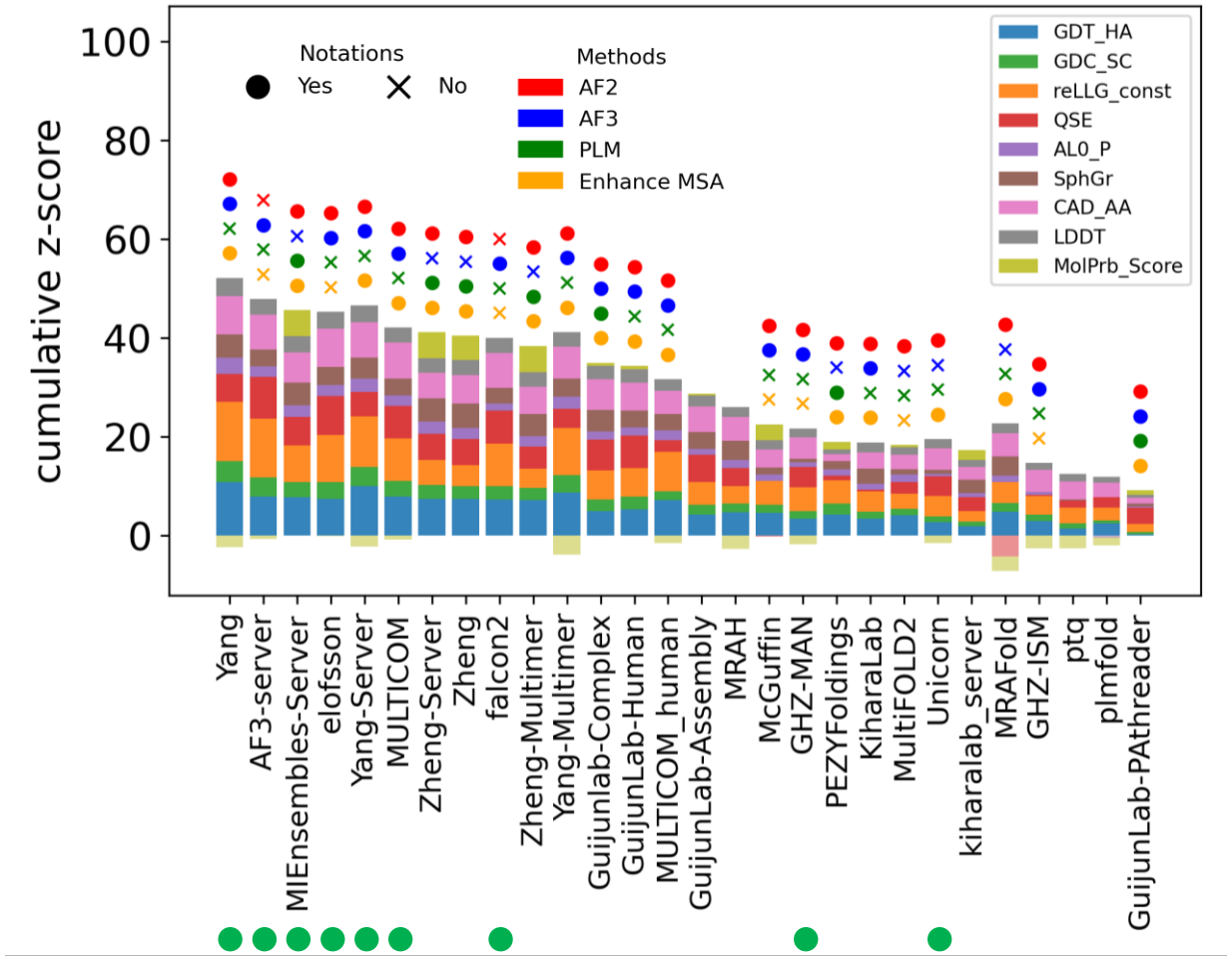
Best models for targets from all phases combined

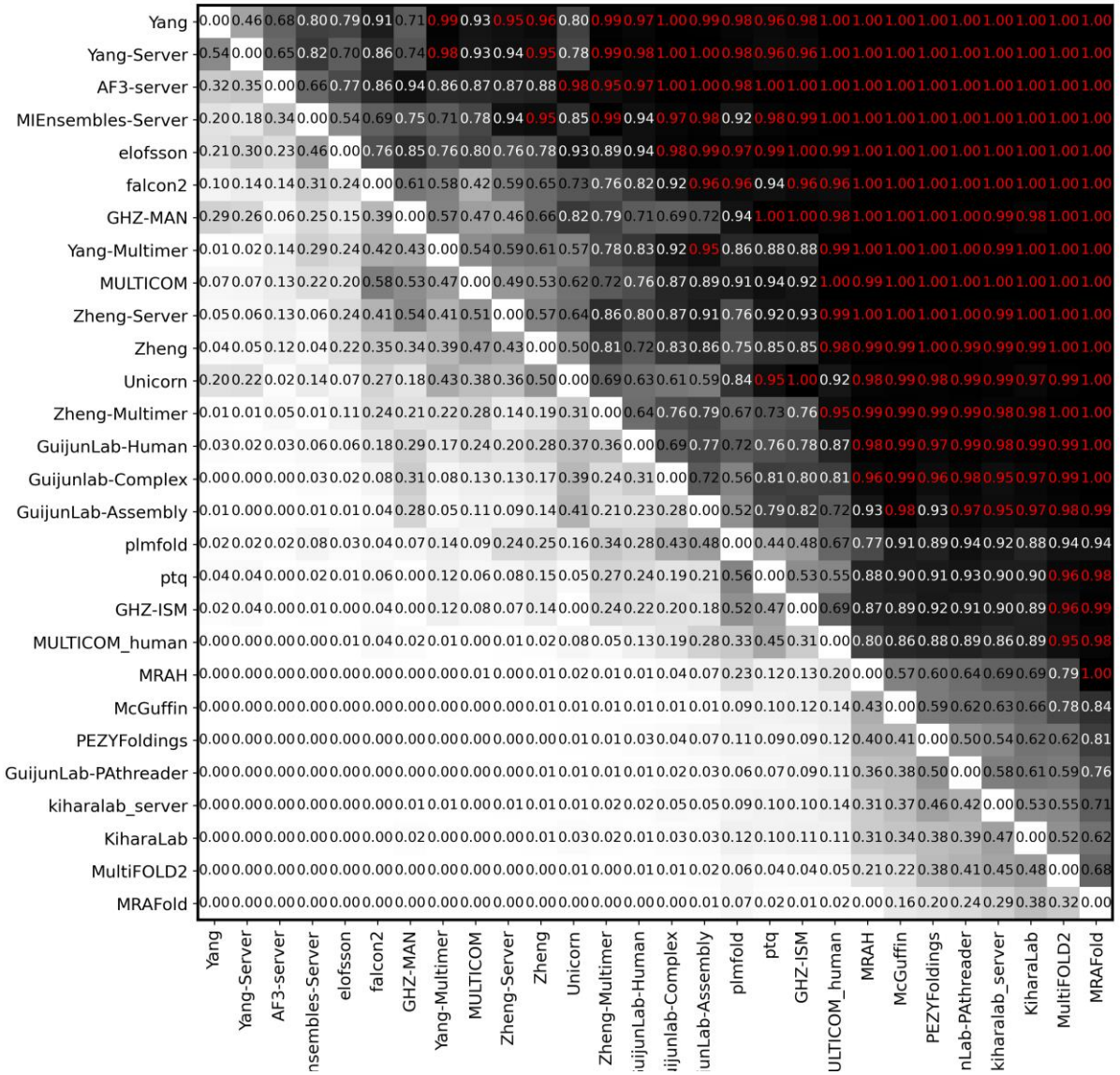




statistical significance (red is significant)

First models for targets from all phases combined

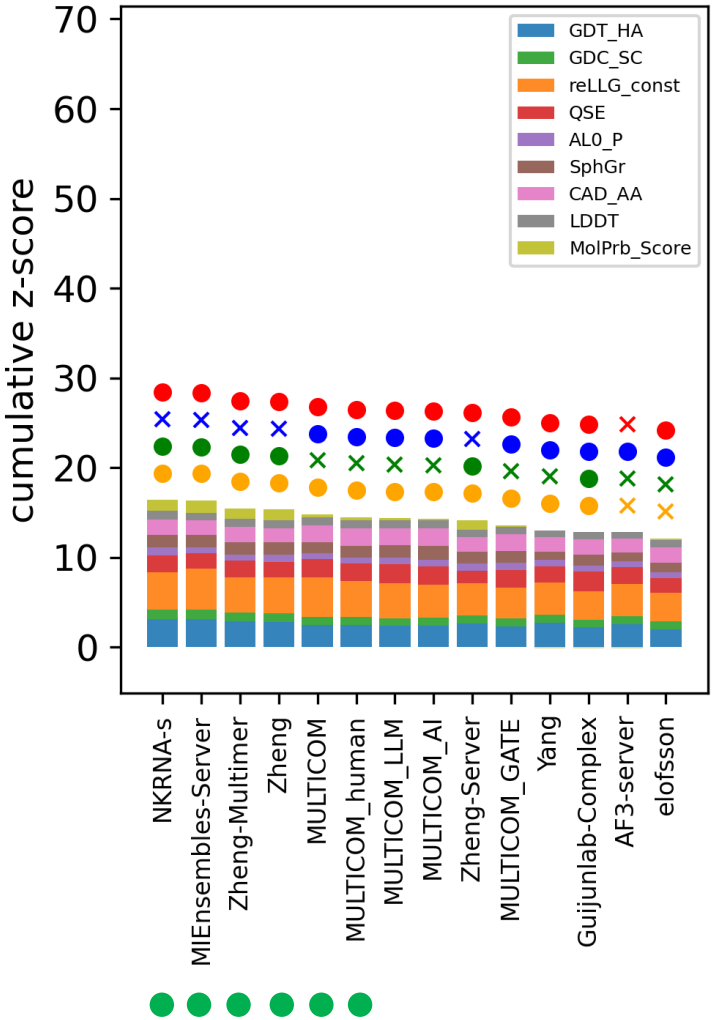




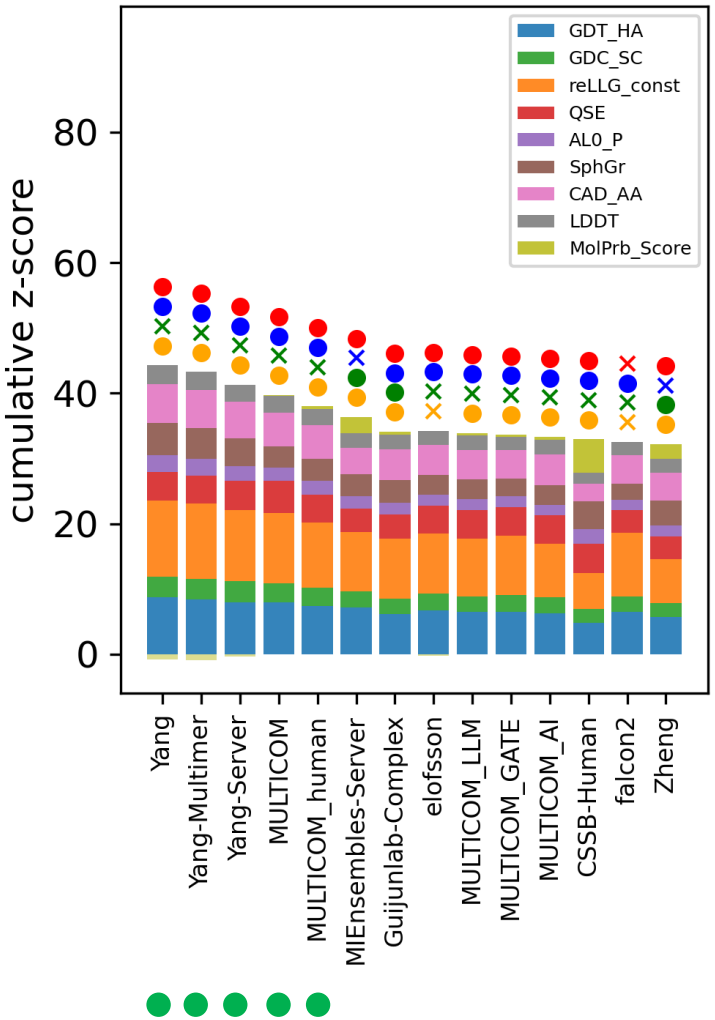
statistical significance (red is significant)

Best models by phase: T0, T1, T2

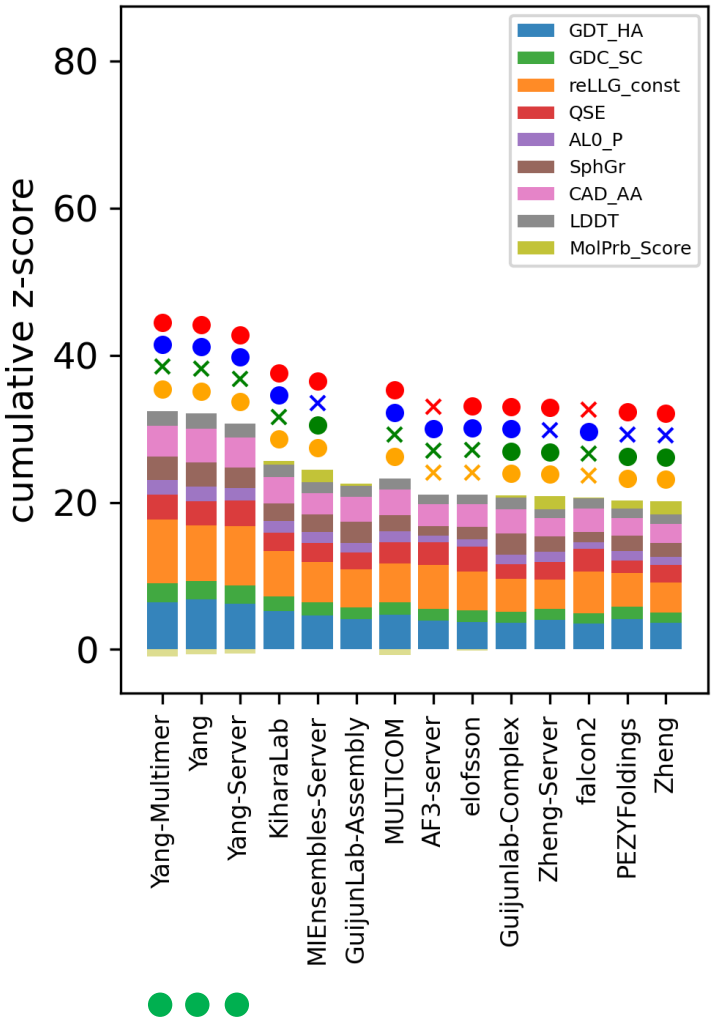
T0



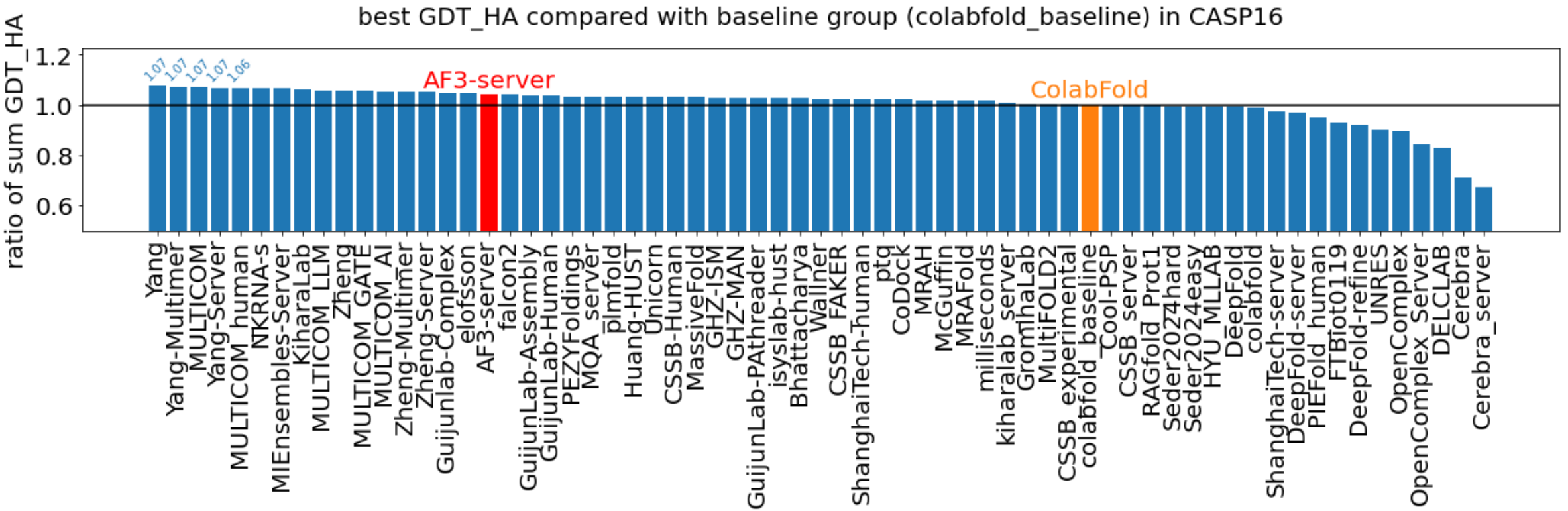
T1

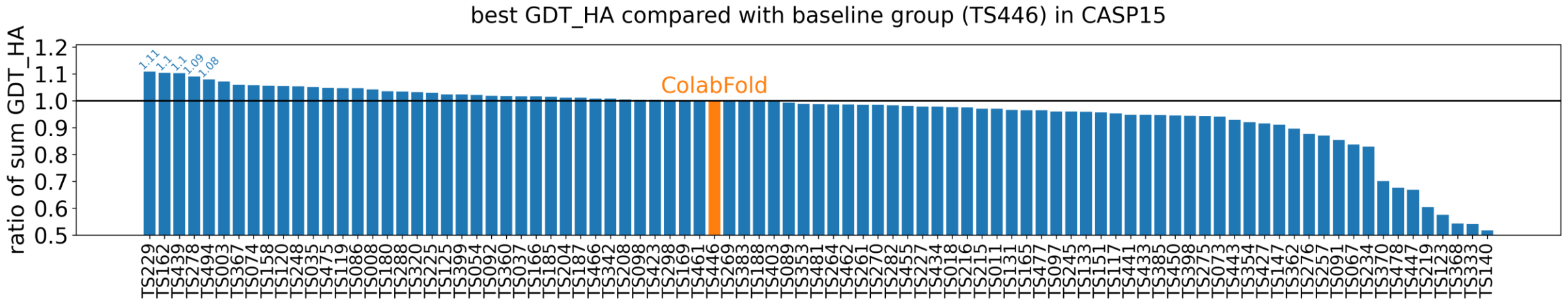


T2

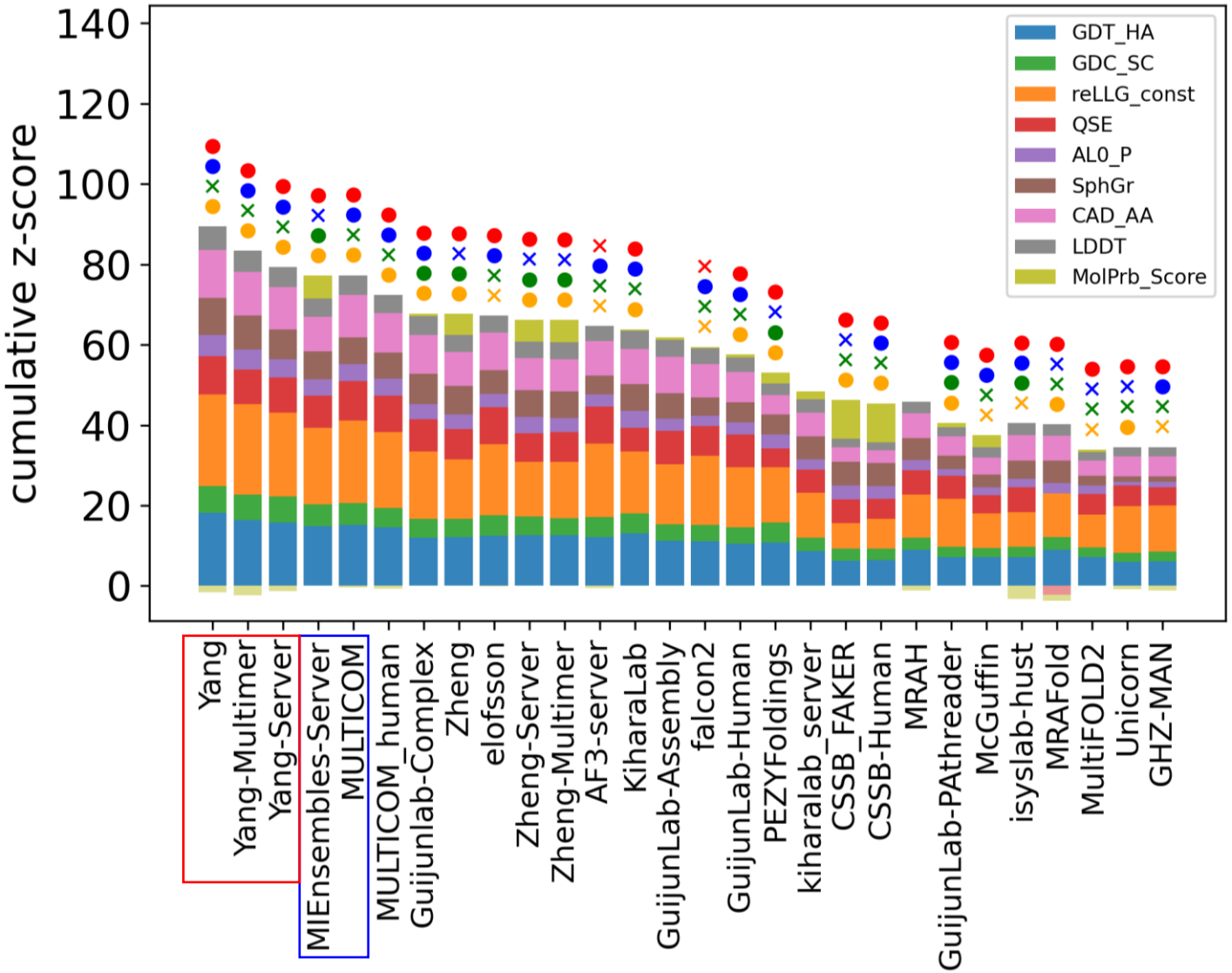


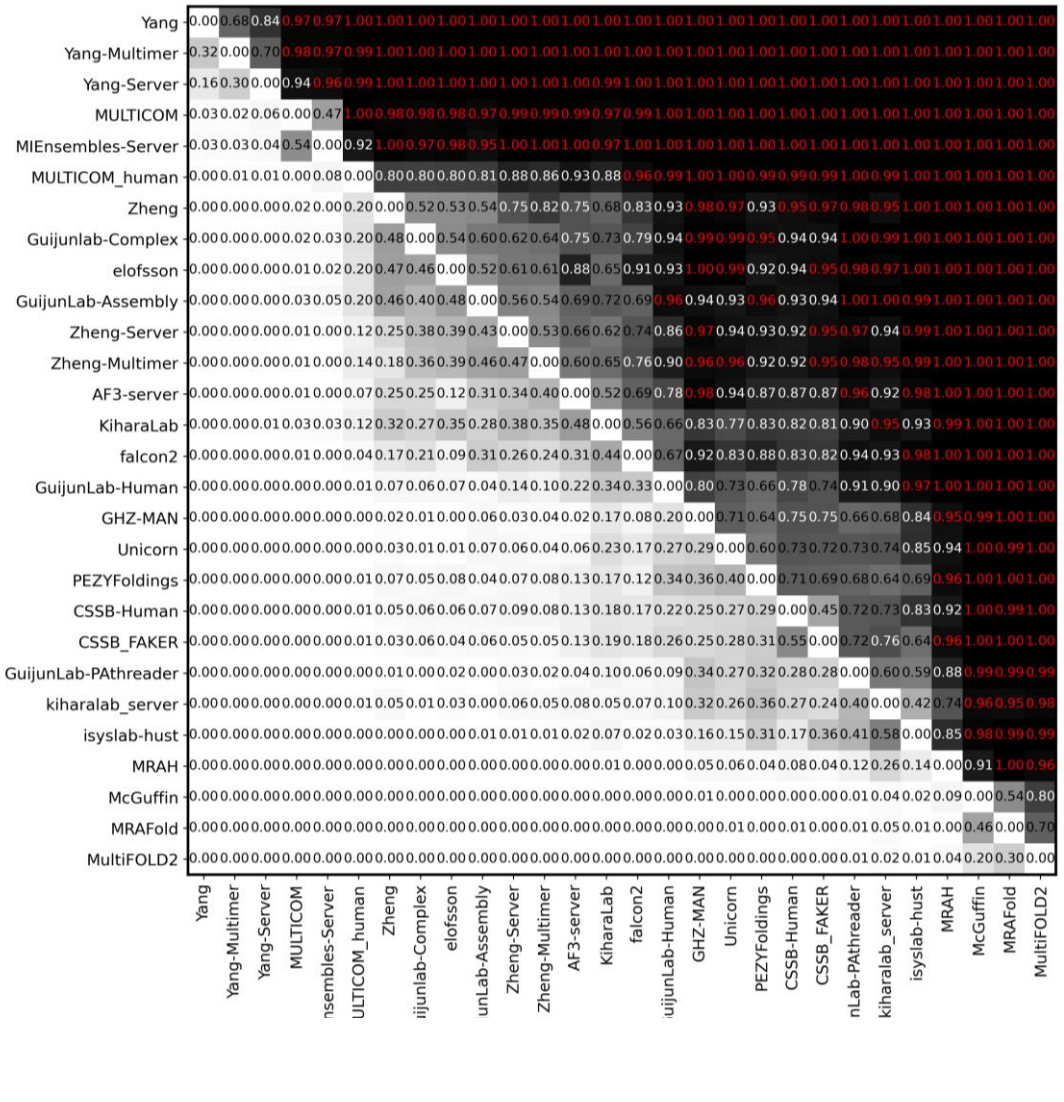
Is there measurable progress in the field?





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





Final thoughts on monomer prediction

0. Single protein "folding problem" is still "solved": not a single fold is predicted incorrectly.

1. The current methods are known to be not sensitive to mutations and truncations. Future CASPs may focus on such cases.

2. Peripheral regions, irregular structures, regions involved in interactions have errors in predictions.

3. Viral and eukaryotic monomers are harder to predict well.

4. Proteins with shallow alignments are predicted worse.

5. Quality of experimental structures could be an issue, and they need to be checked by Mprob.

6. More extensive AF3 use gives better performance.

7. Progress compared to AF3 is measurable, but seemingly incremental (sorry!).