

Analysis and evaluation of CASP16 protein monomer predictions

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

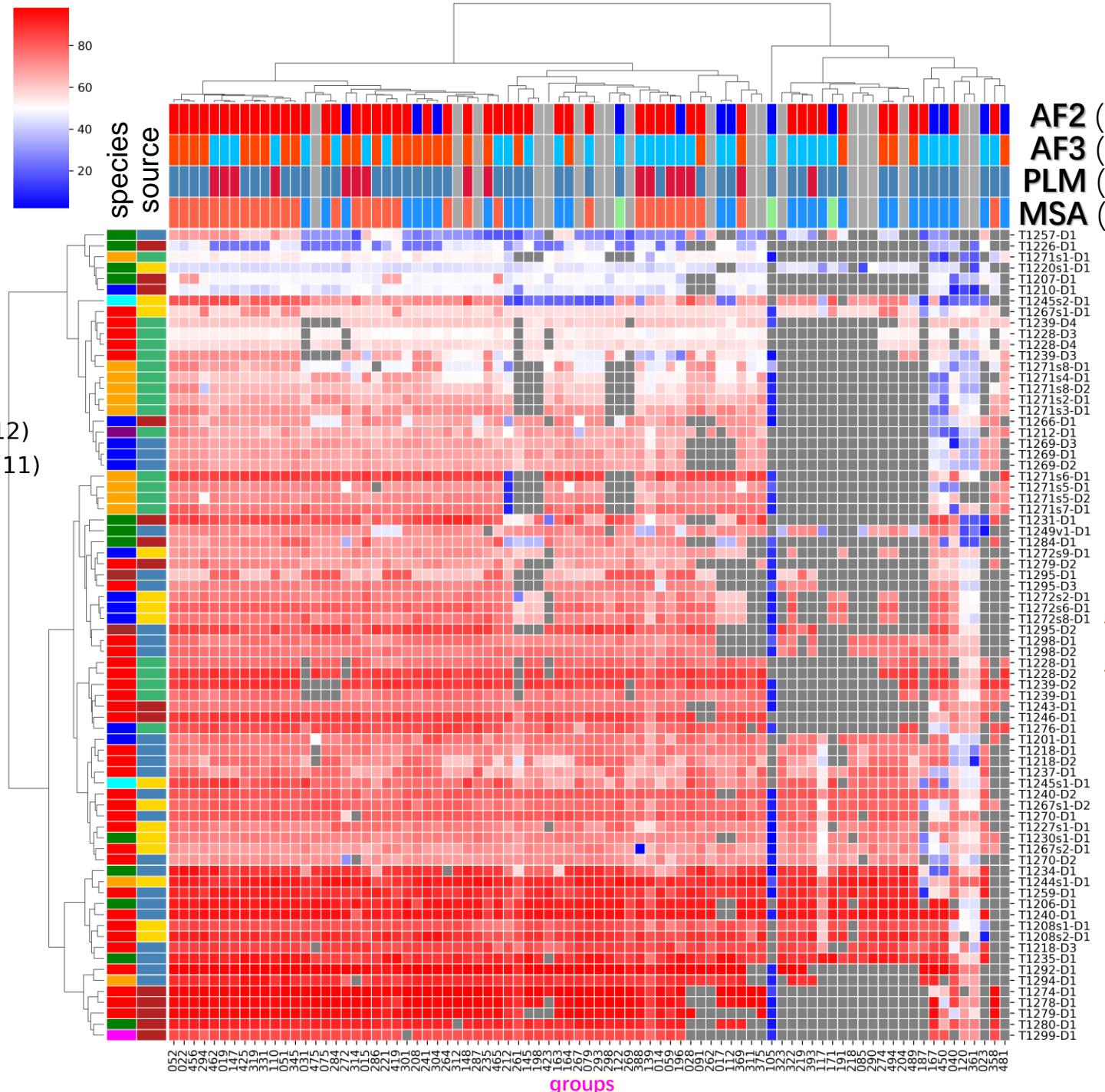
University of Texas Southwestern Medical Center

“CASP is an experiment, not a competition”

Experiment (**insights**) first

Competition (**ranking**) second

GDT_HA

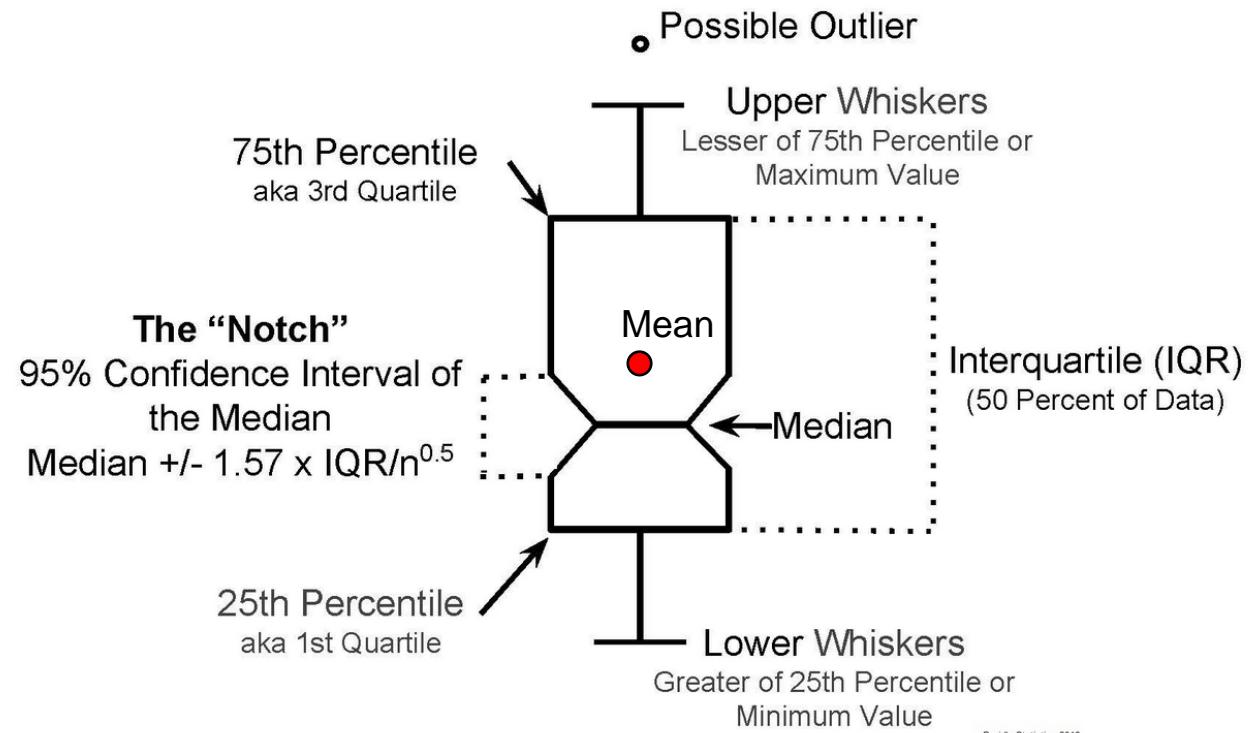
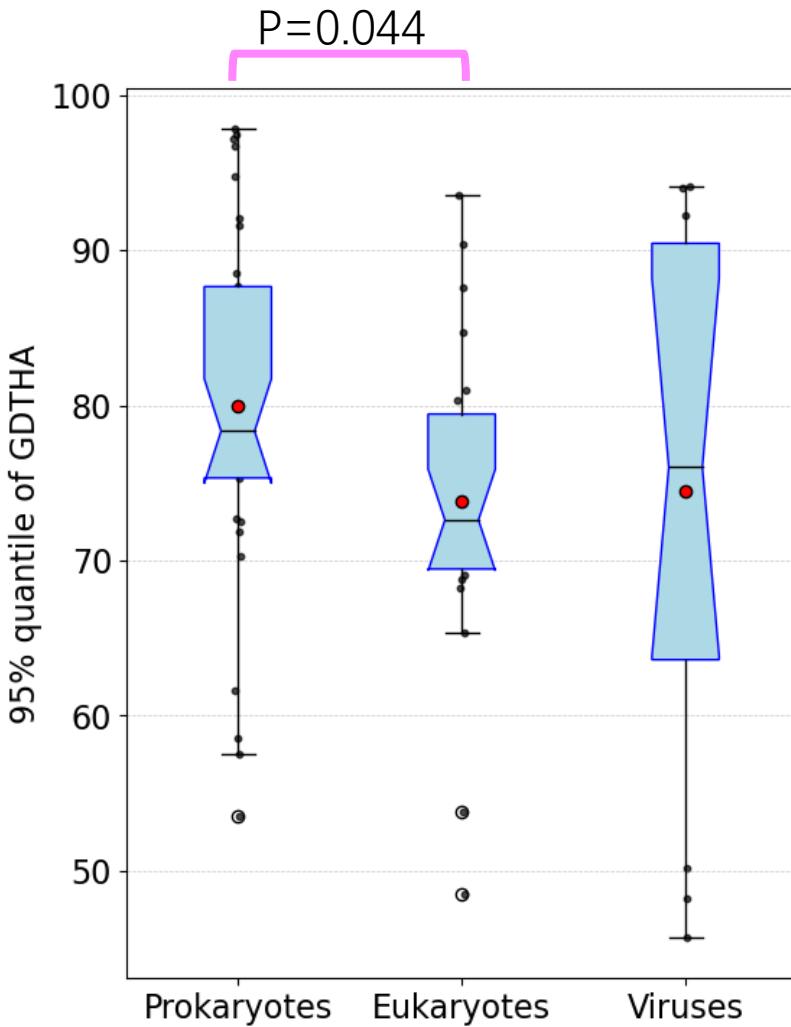


Performance,
targets,
and methods
at a glance

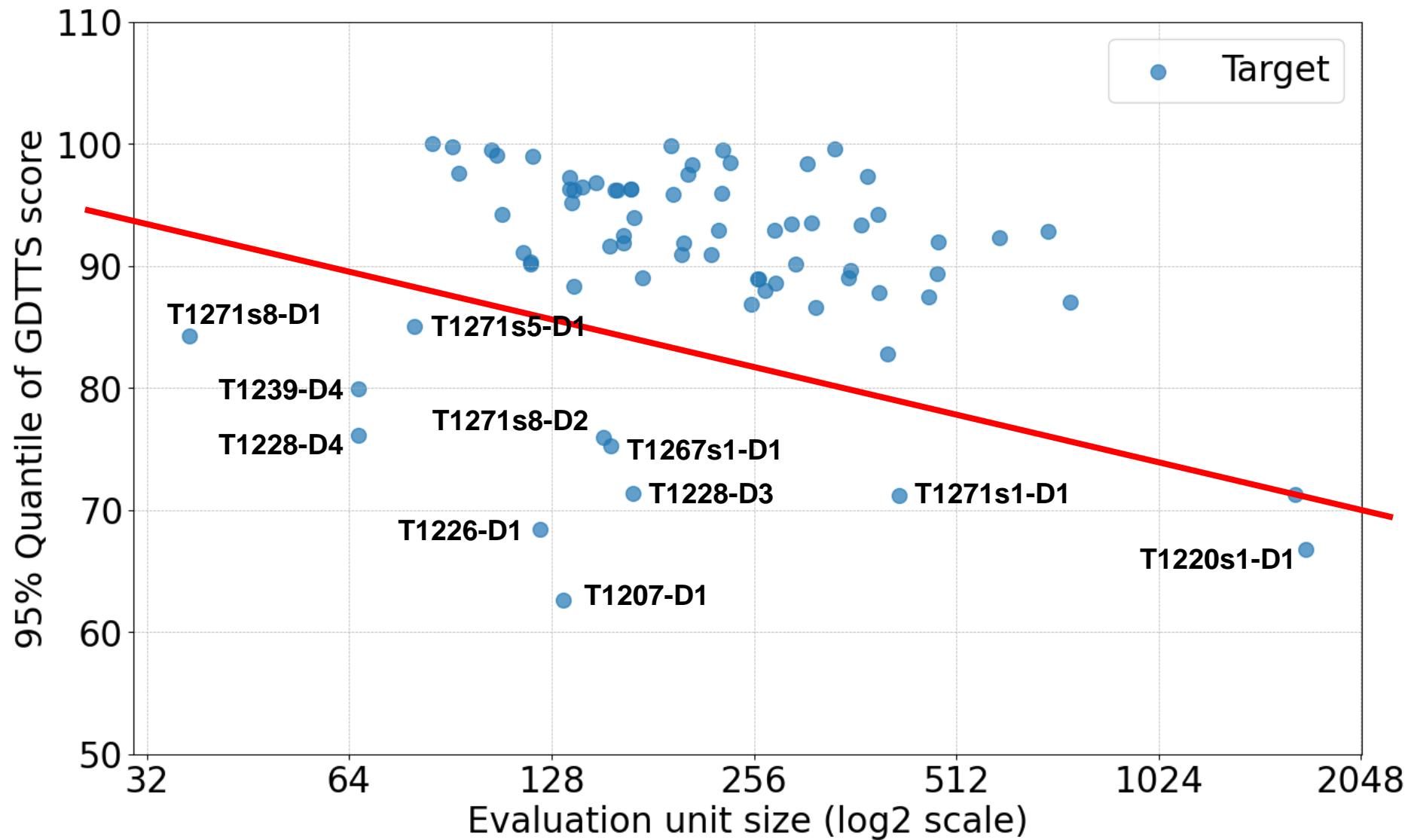
targets as rows
groups as columns

Common failures in monomer modeling

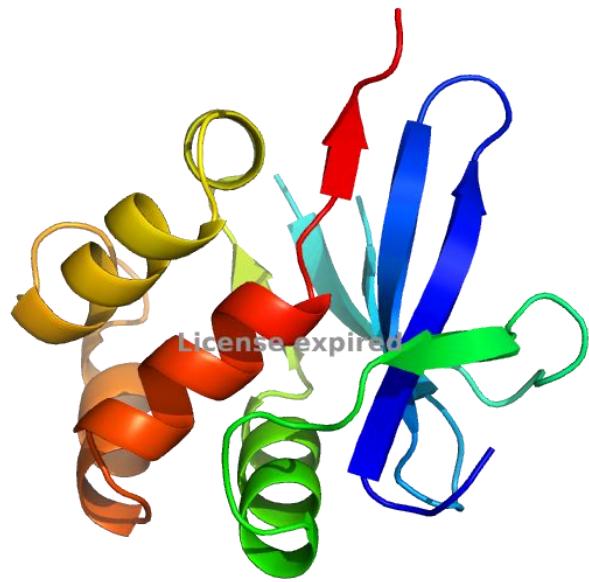
What monomers are still challenging?



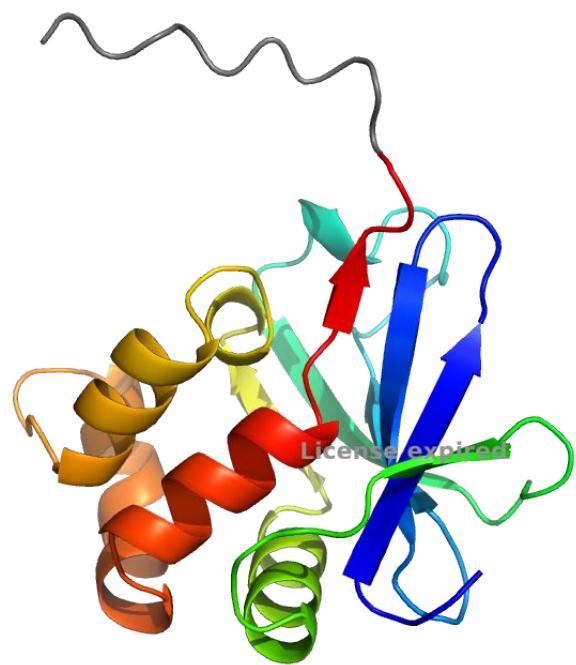
Poorly predicted monomers



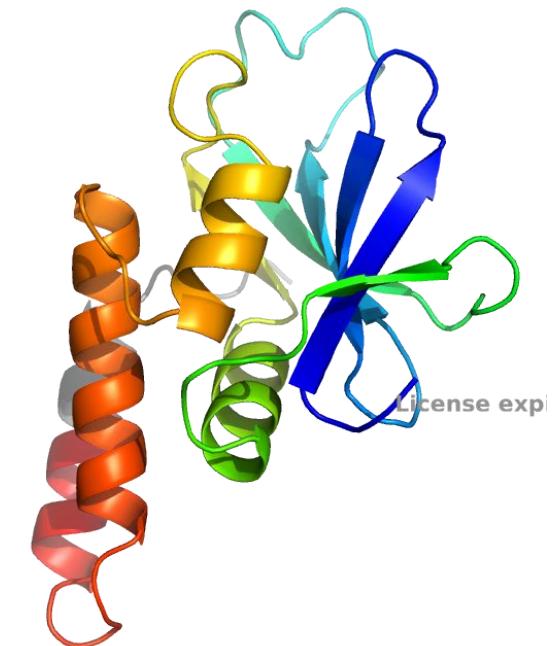
Poor performance by most groups on T1207-D1



Target

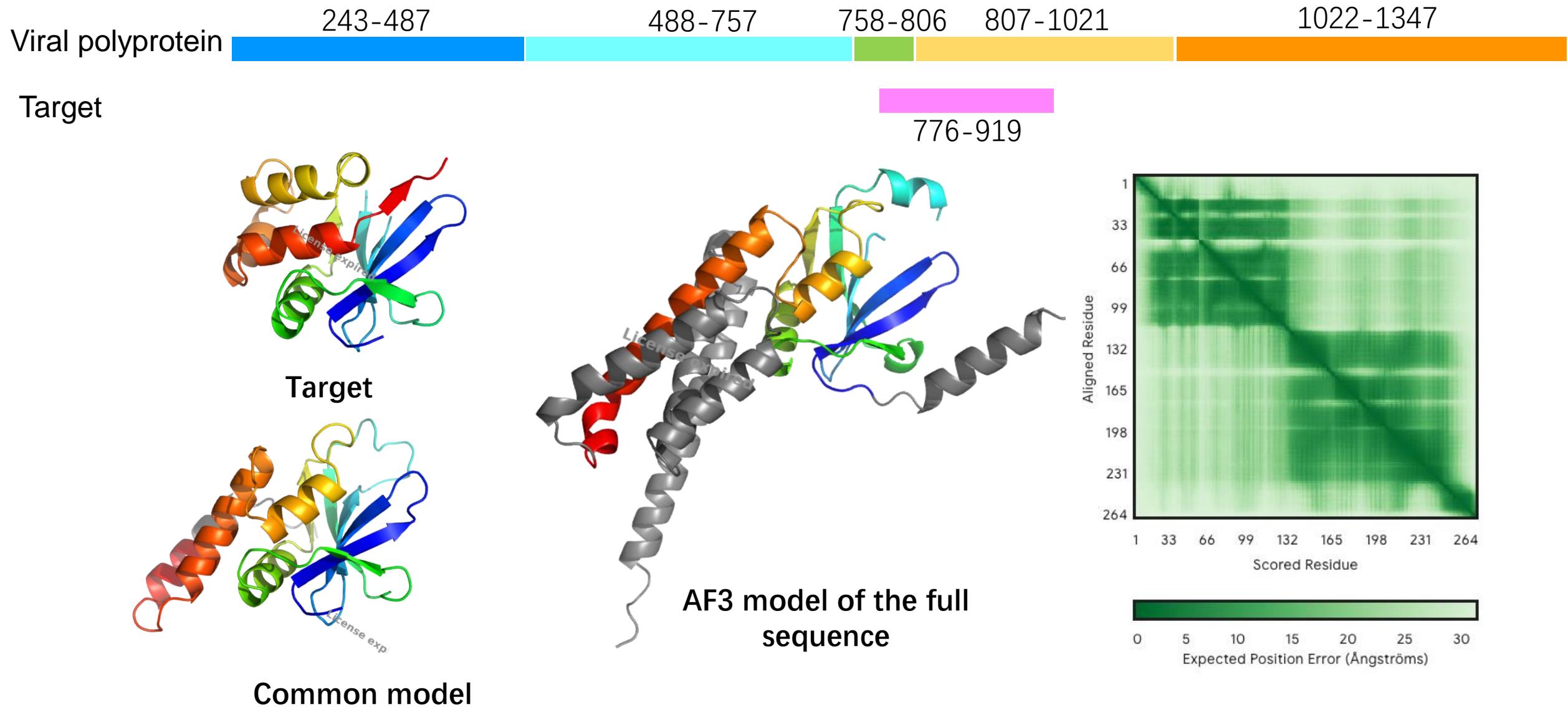


A winning model
from the **Wallner** group

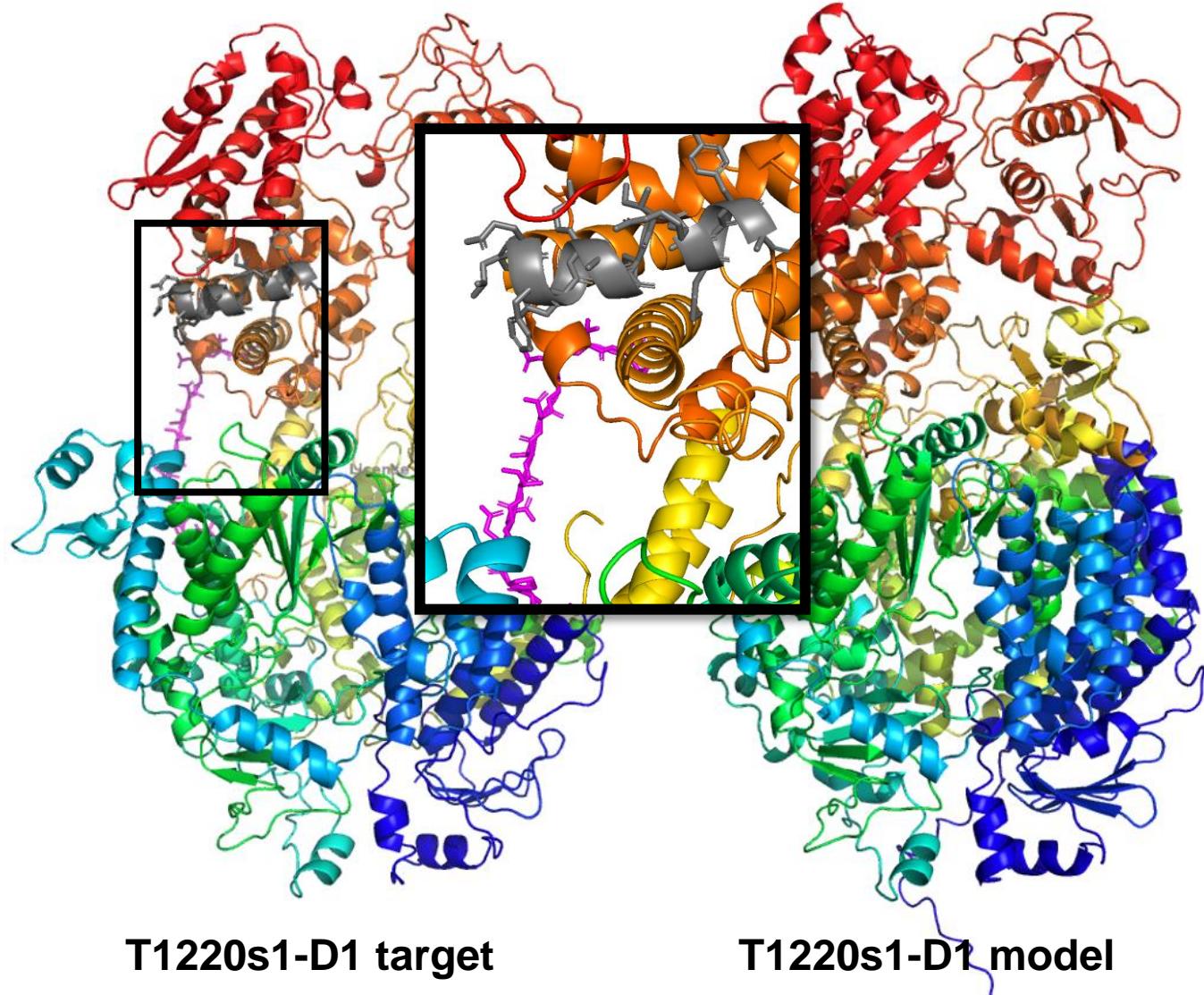


Most common model

Challenge 1: alternative confirmation after truncation?



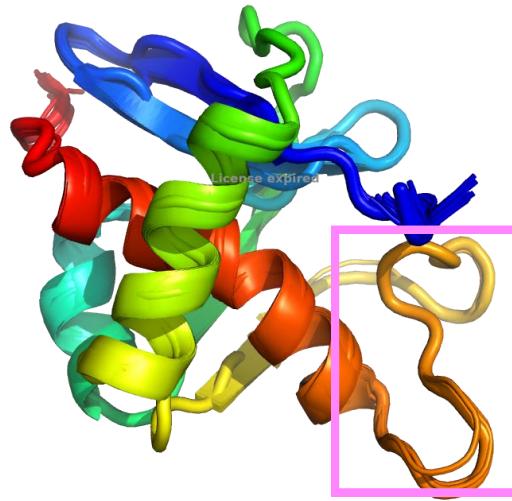
Challenge 2: poorly resolved experimental structures



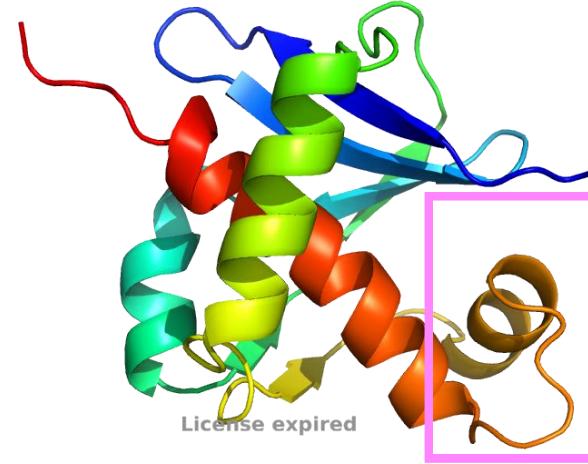
Clashscore, all atoms:	13.03	58 th percentile * (N=1784, all resolutions)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Poor rotamers	569	39.49% Goal: <0.3%
Favored rotamers	648	44.97% Goal: >98%
Ramachandran outliers	9	0.55% Goal: <0.05%
Ramachandran favored	1493	90.76% Goal: >98%
Rama distribution Z-score	1.82	99 th percentile * (N=1784, abs(Z score) < 2)
MolProbity score^	3.36	12 th percentile * (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å	2	0.00% Goal: 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	99 th 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	1.82	99 th percentile * (N=1784, abs(Z score) < 2)
MolProbity score^	1.77	86 th percentile * (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å	20	1.24% Goal: 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

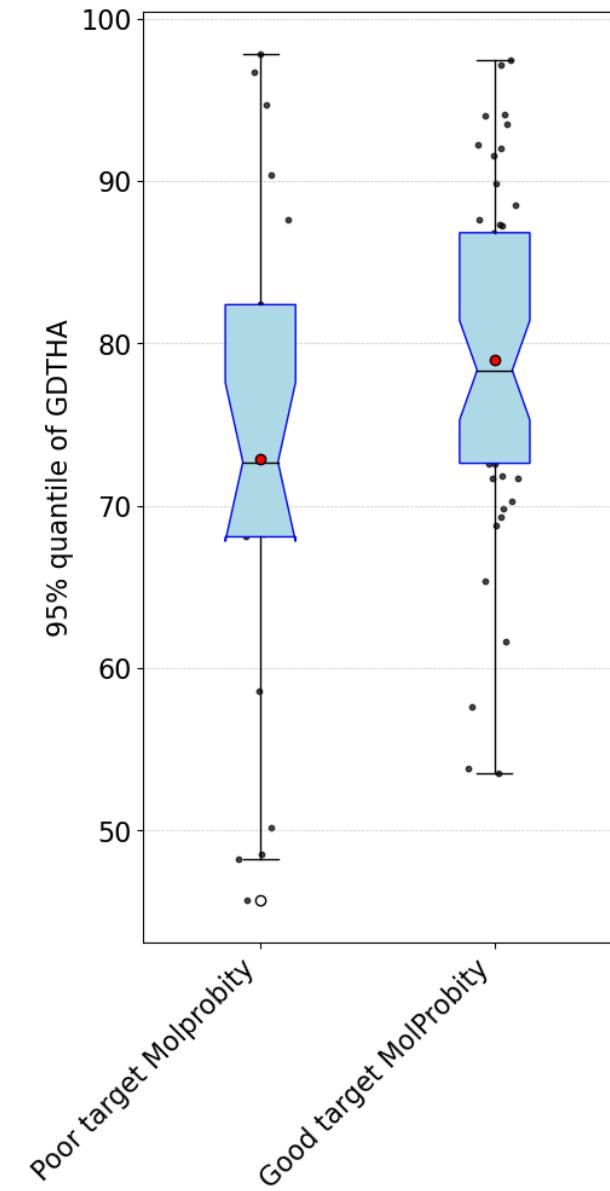


T1226-D1 target

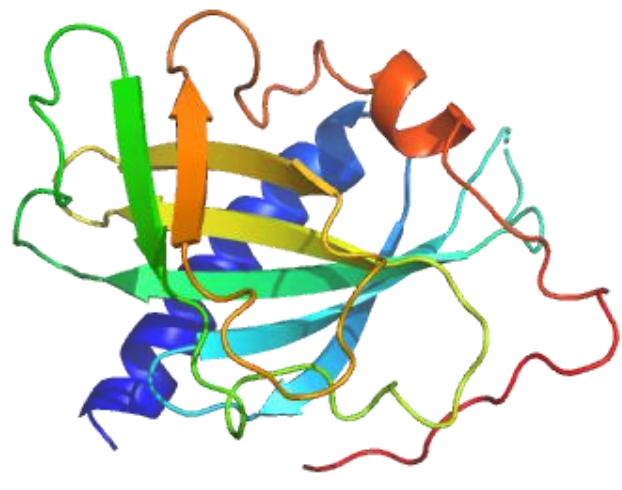


T1226-D1 model

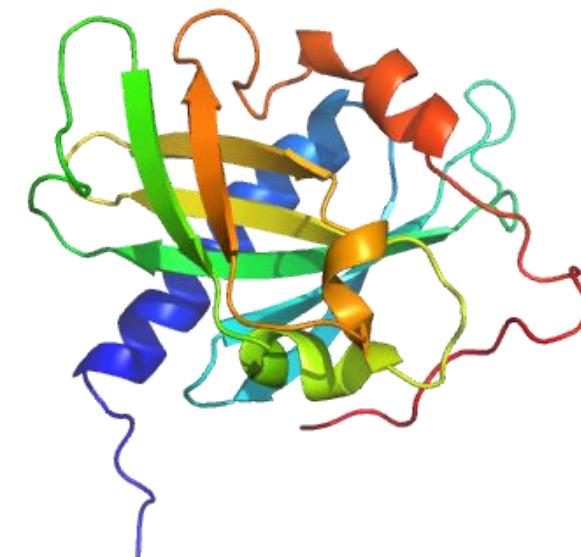
Clashscore, all atoms:	49.14	4 th percentile * (N=1784, all resolutions)
Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.		
Poor rotamers	21	21.21% Goal: <0.3%
Favored rotamers	64	64.65% Goal: >98%
Ramachandran outliers	1	0.83% Goal: <0.05%
Ramachandran favored	110	90.91% Goal: >98%
Rama distribution Z-score	4.43 ± 0.30	Goal: abs(Z score) < 2
MolProbity score [^]	3.70	5 th percentile * (N=27675, 0Å - 99Å)
Cβ deviations >0.25Å	0 / 955	0.00% Goal: 0
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%



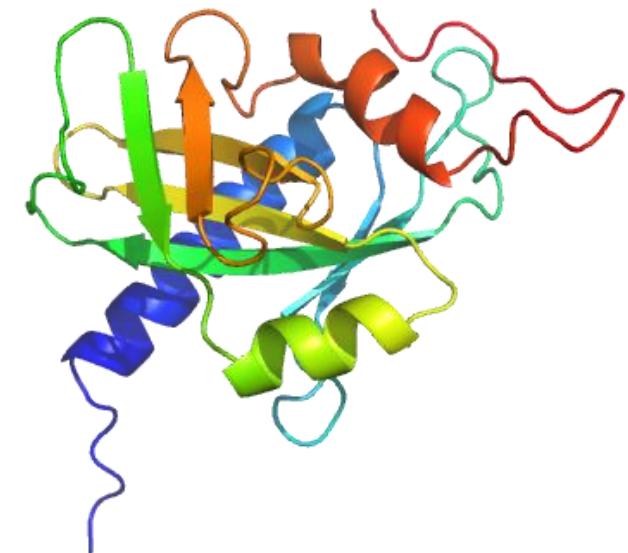
Poor performance by most groups on T1267s1-D1



Target

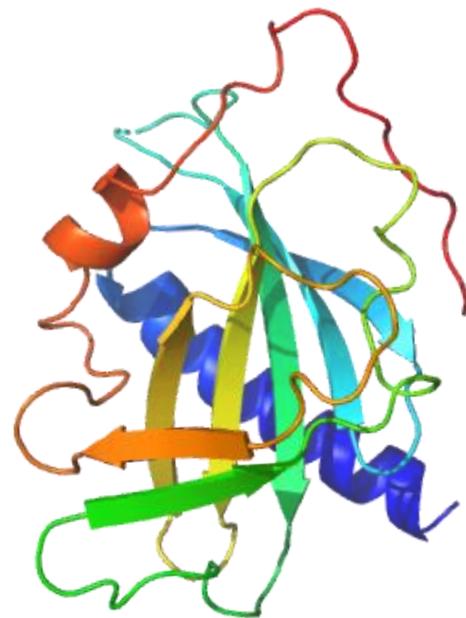


A winning model
from the **Kozakov** group

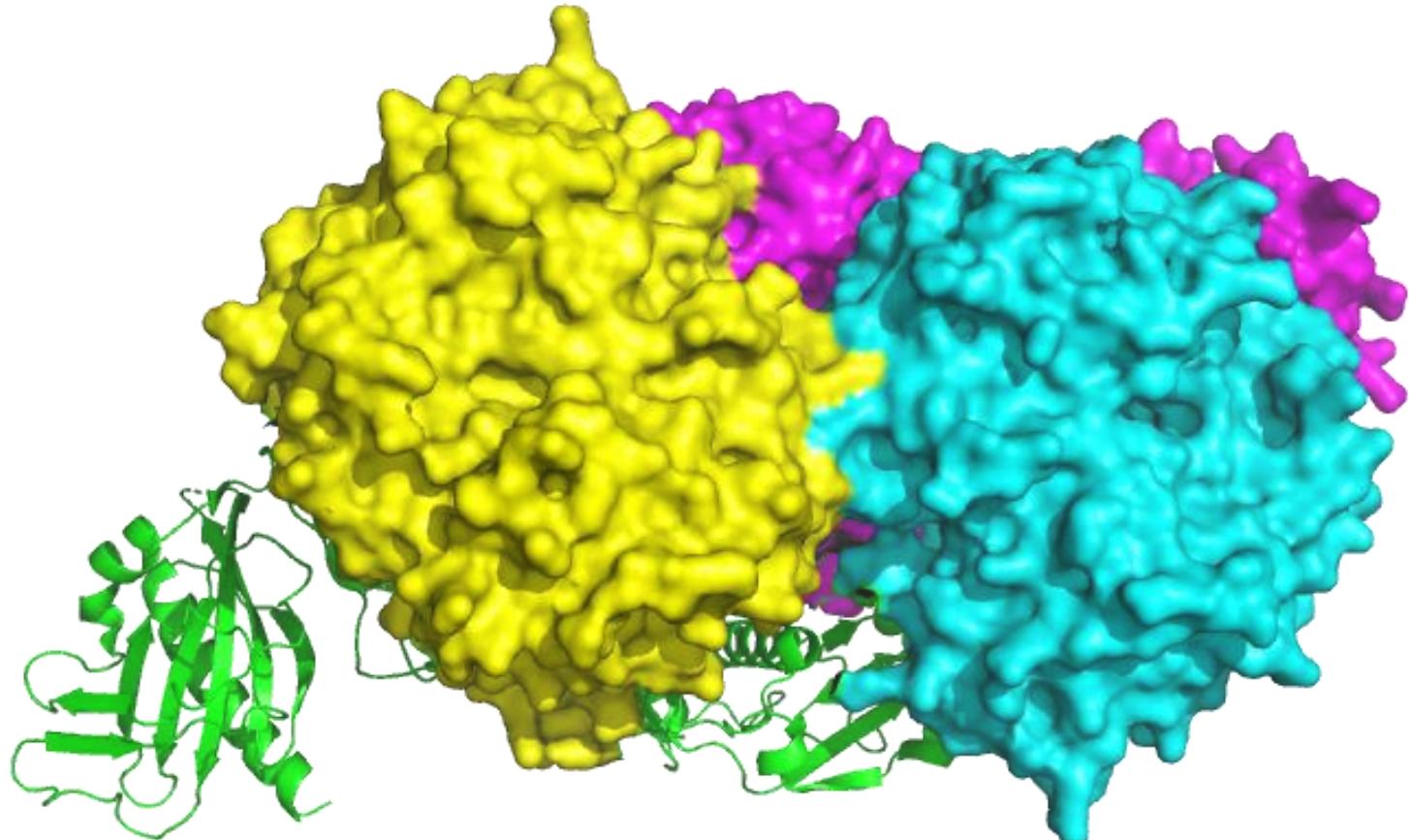


Most common model

Challenge 3: induced fit upon binding to partners

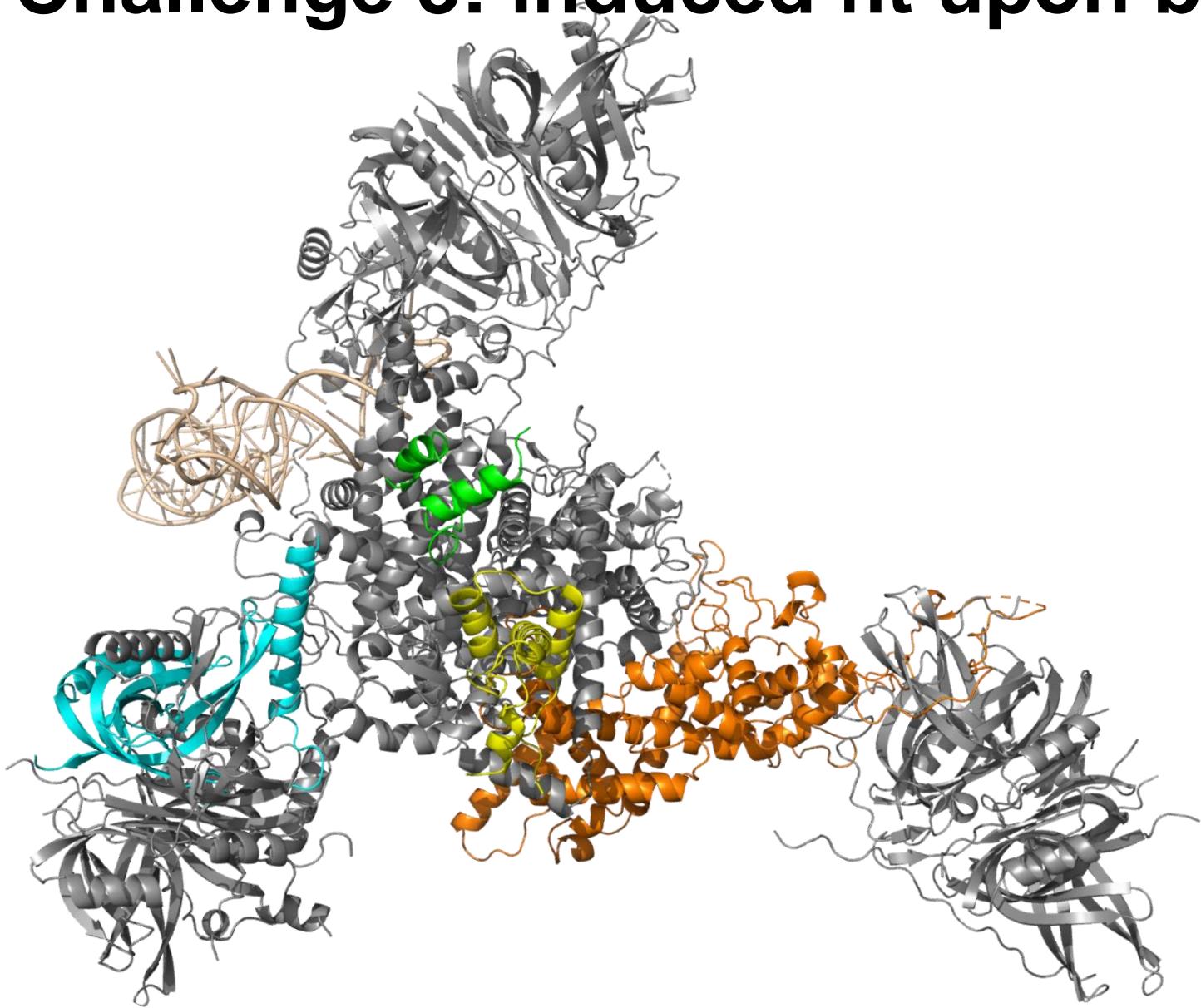


T1267s1-D1 target

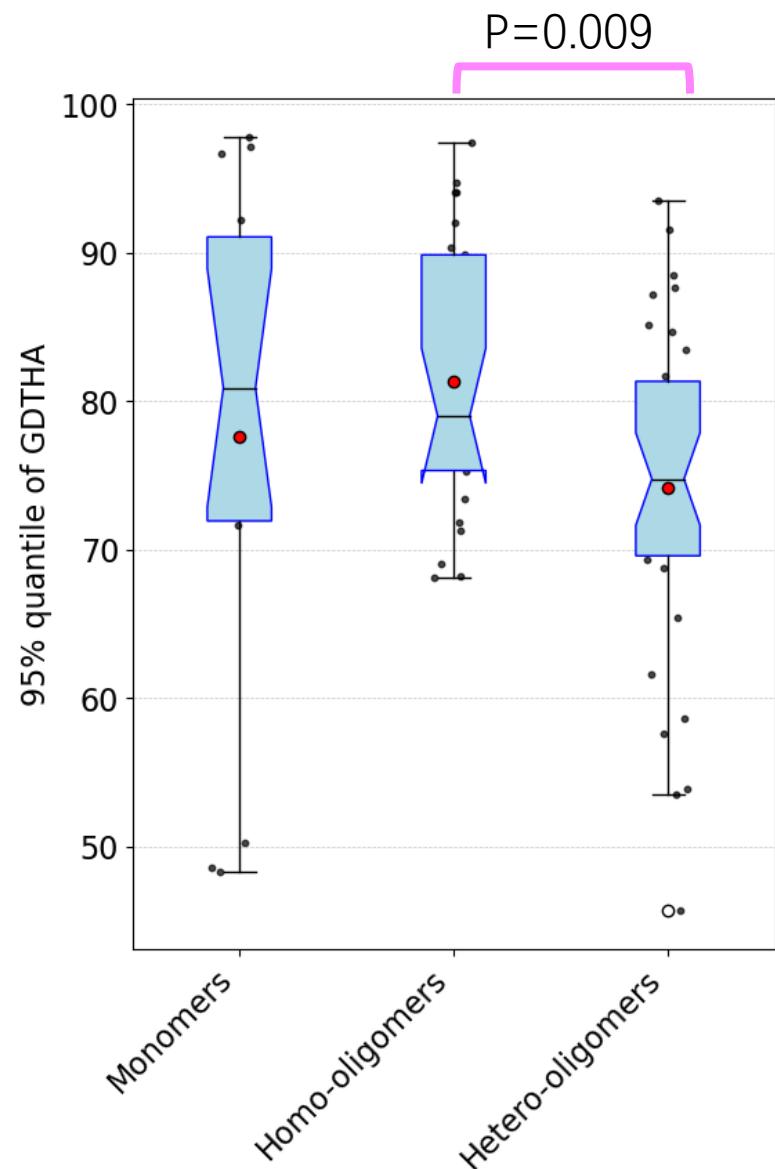


T1267s1-D1 in its context

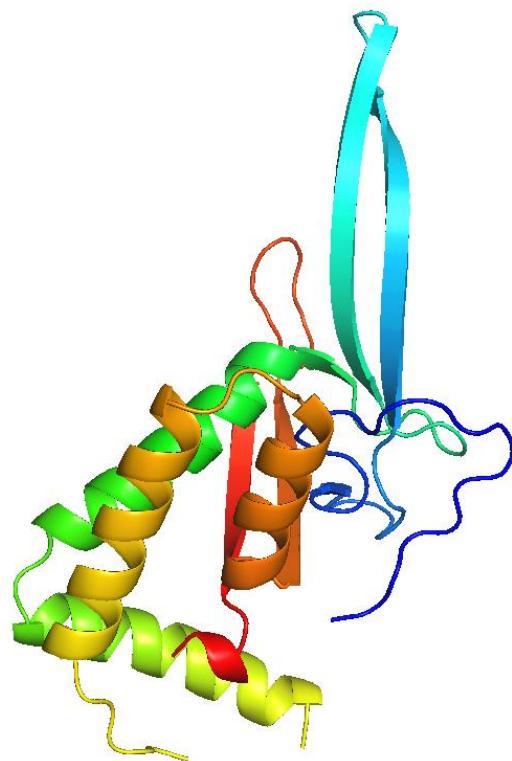
Challenge 3: induced fit upon binding to partners



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



Challenge 4: irregularities in secondary structures

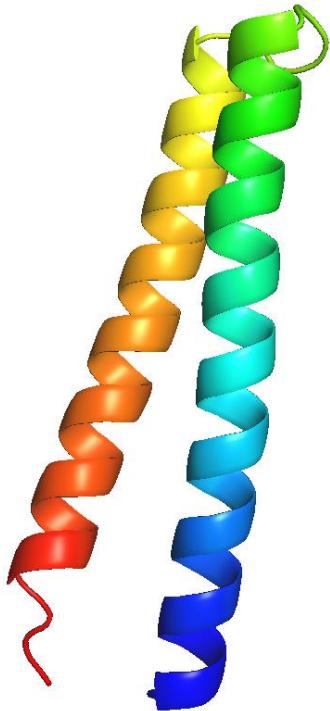


T1228-D3 target

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

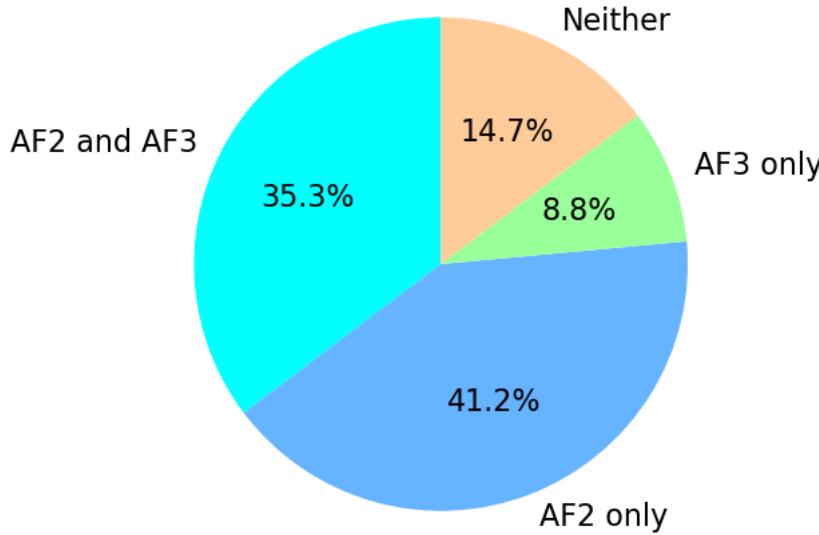
T1228 target has good MolProbity scores

Clashscore, all atoms:	1.46	99 th 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	96 th 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	CSB_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
MIEensembles-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	CSB-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
Schniedman	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

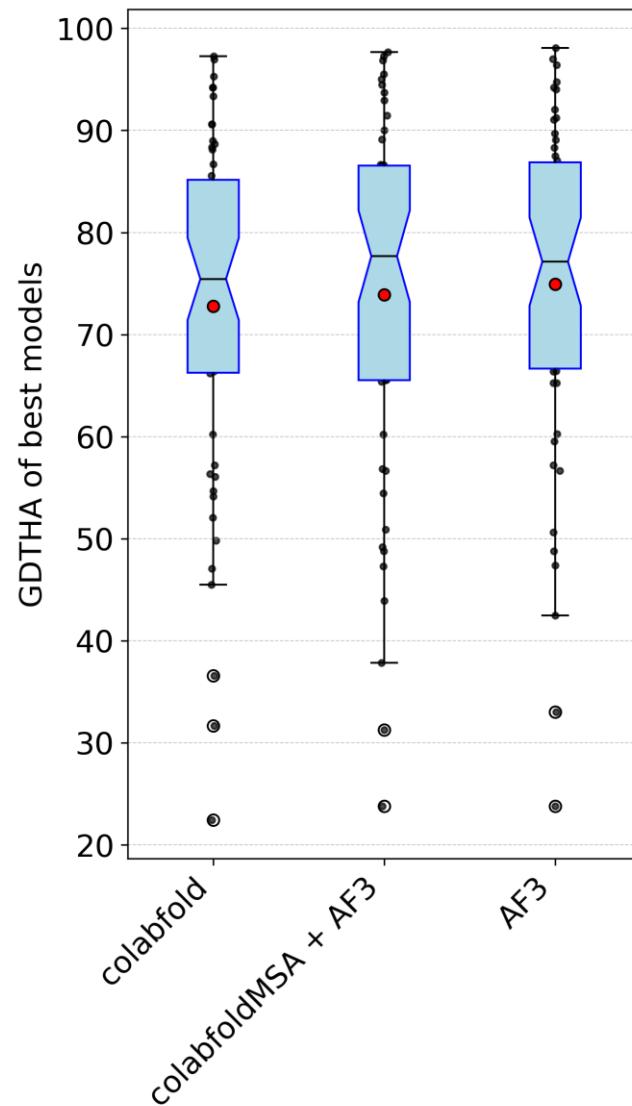
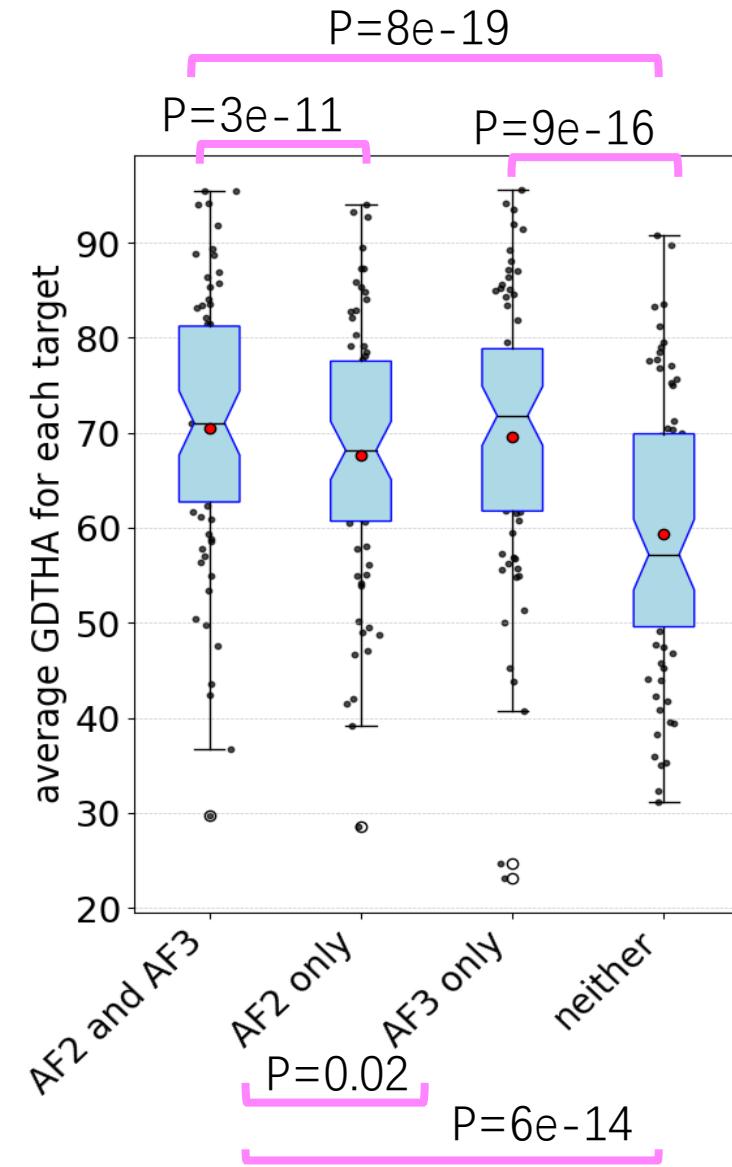
At least 31 (45%) groups worked on better MSAs

At least 18 (24%) groups used PLMs (none is purely PLM-based)

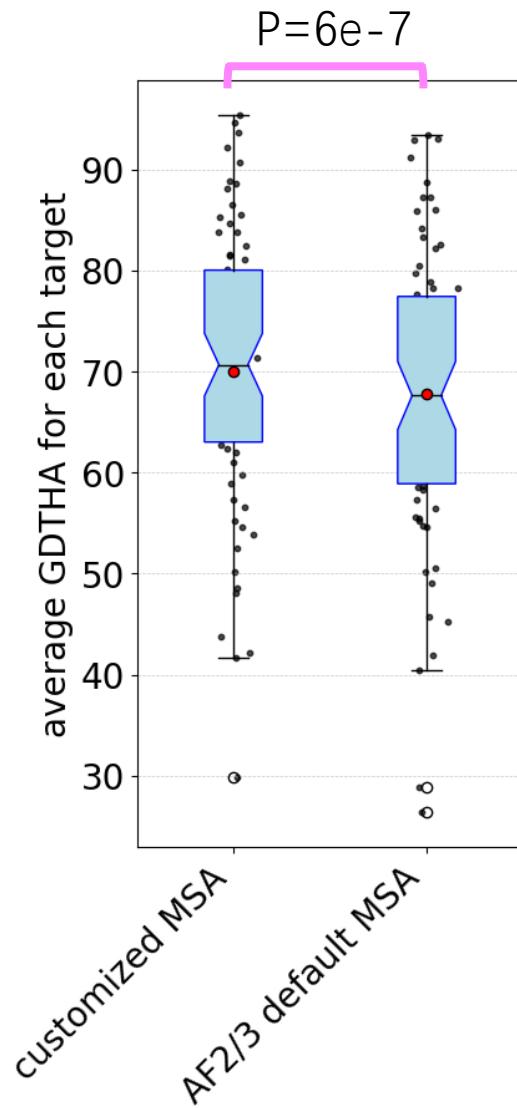
Type: H/S - human/server

MSA: Y/S/N - yes/special/no

General winning strategies in CASP16: using AF3

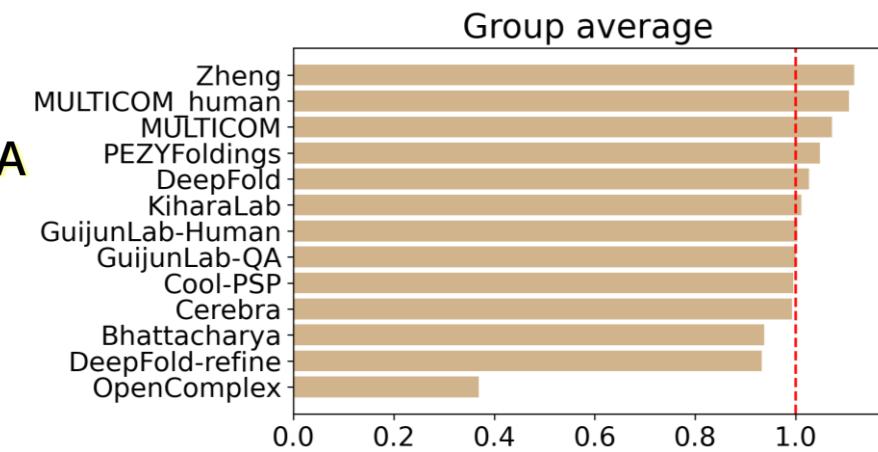
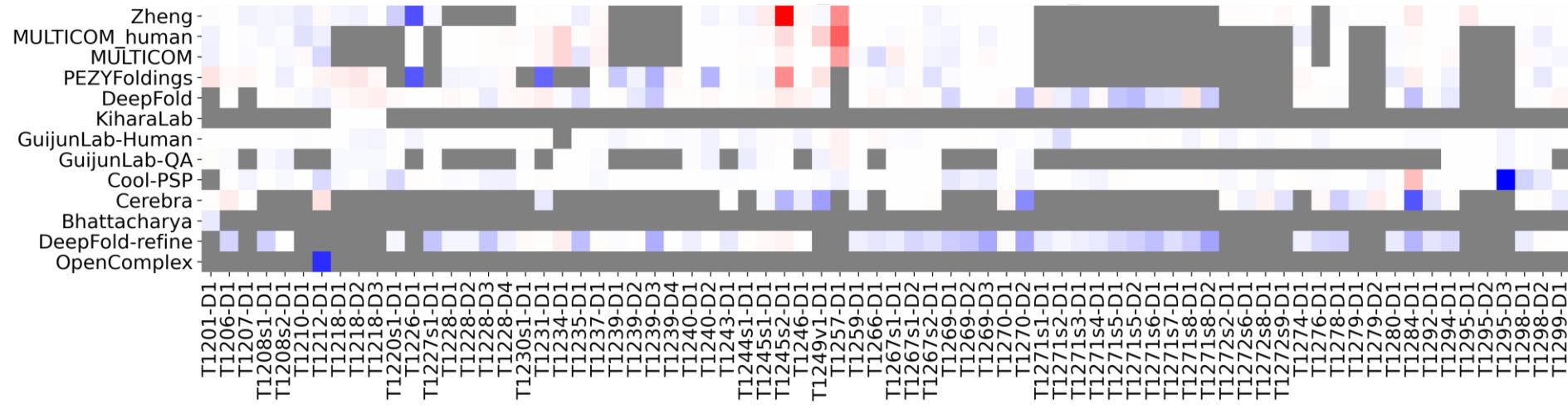


General winning strategies in CASP16: improving MSA



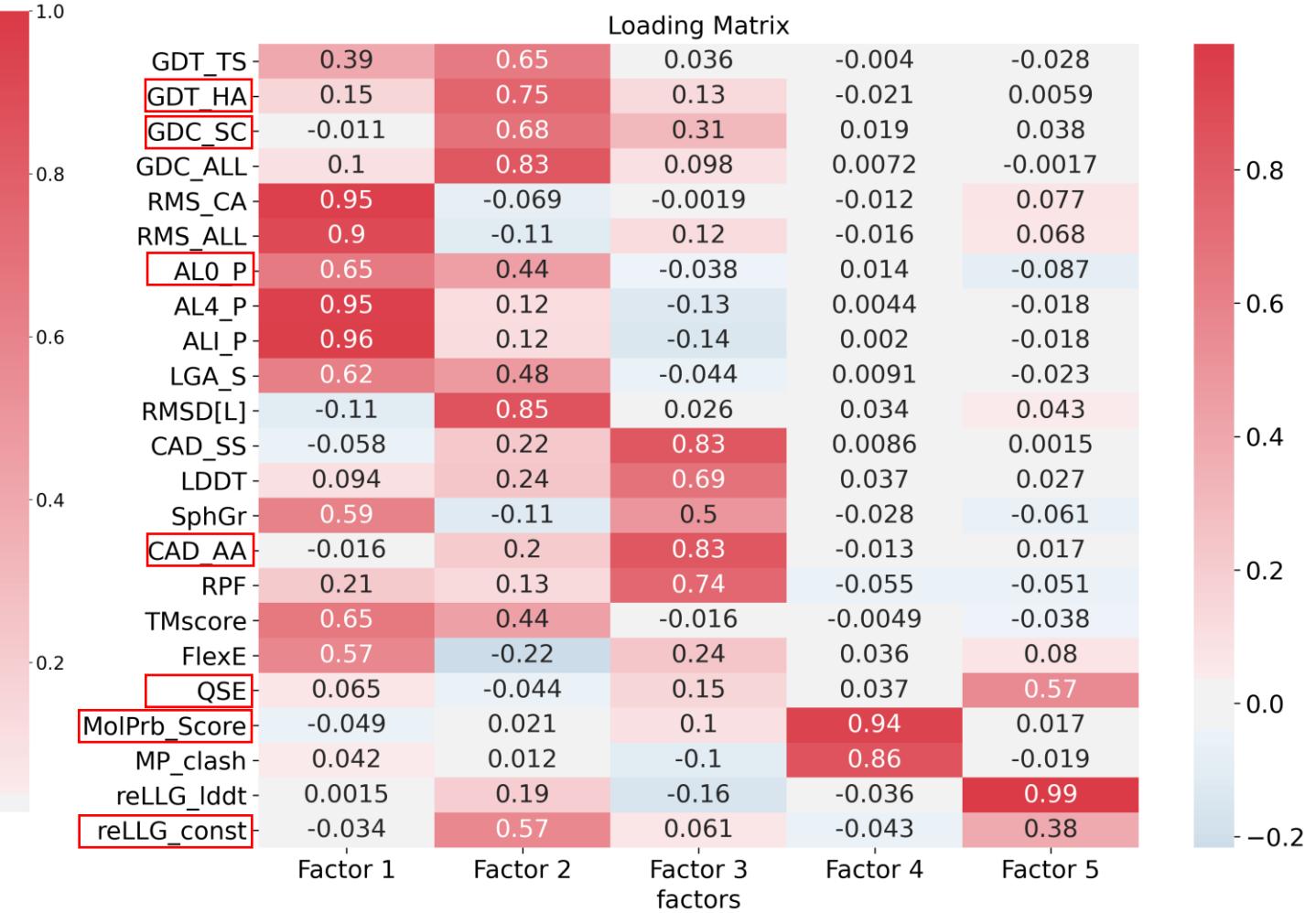
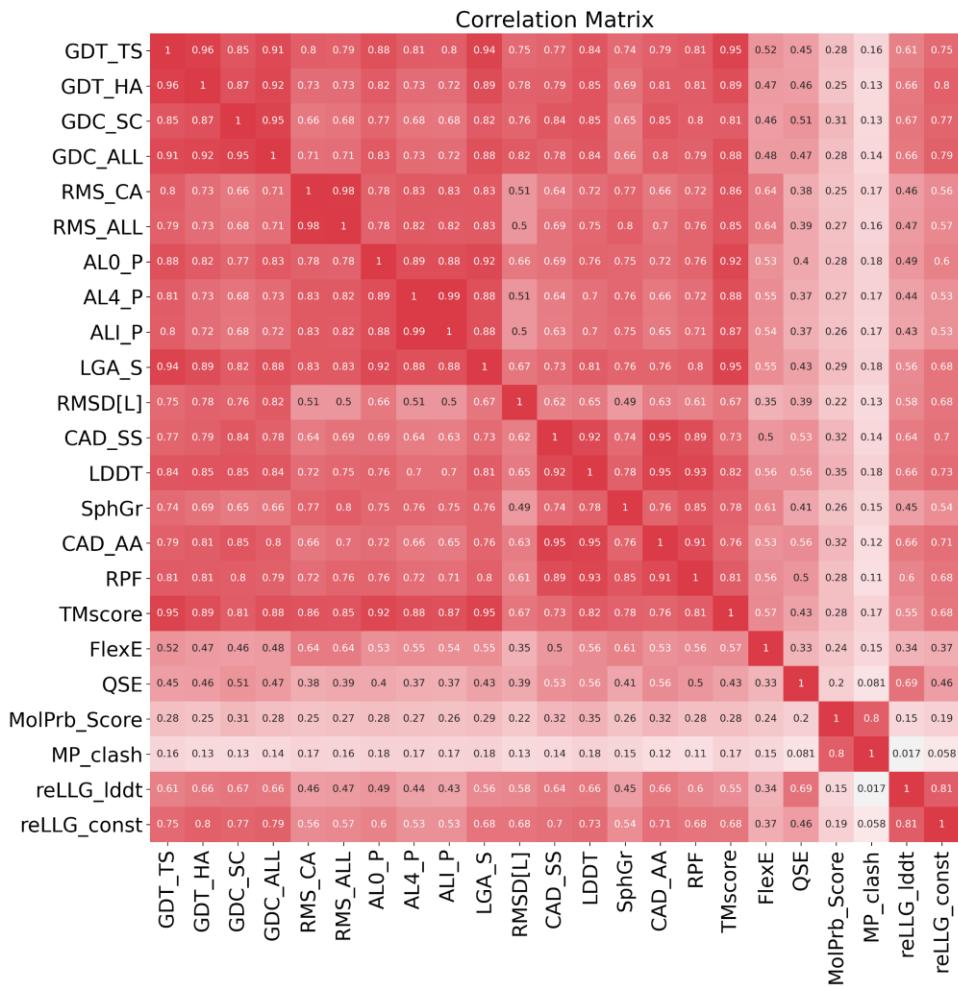
Model 1 GDT_HA divided by model 6 GDT_HA

The same method, but different MSAs



Ranking

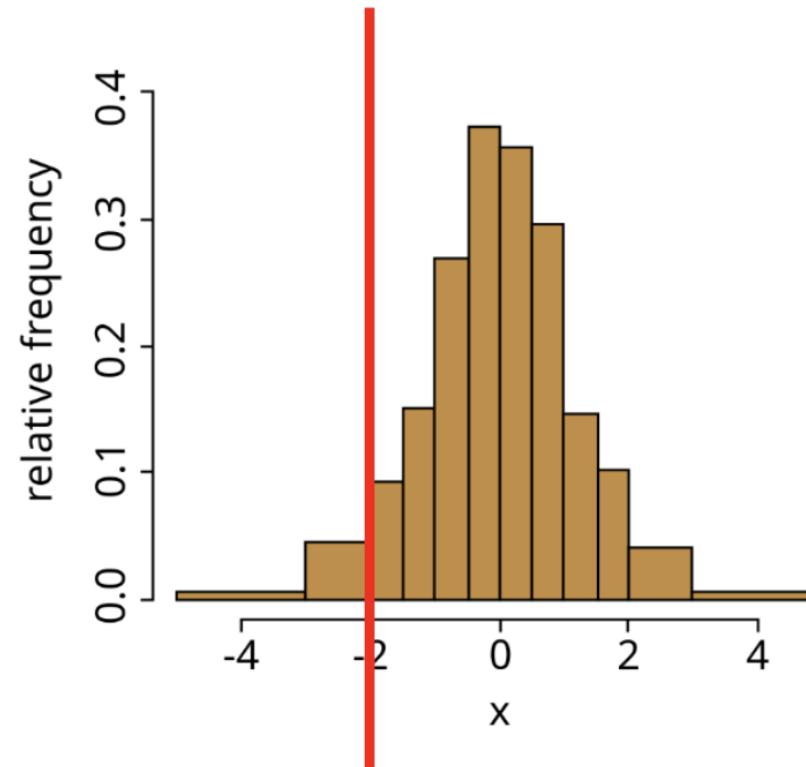
Comparison and factorization of scores



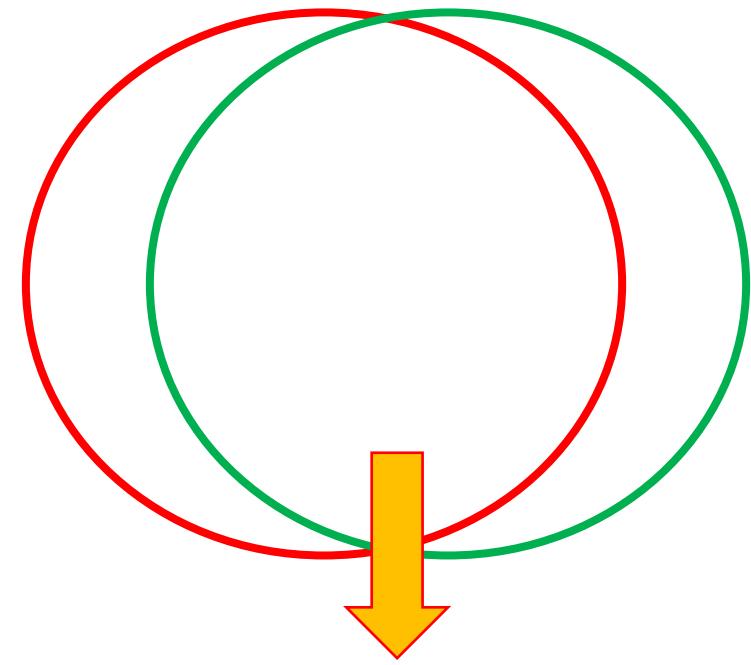
Final empirical scoring formula

$$\begin{aligned} Z(\text{monomer}) = \sum_{\text{target}} \left(\sum_{\text{EU}} w_{\text{EU}} \left(\frac{1}{16} \times (Z(\text{GDC_SC}) + Z(\text{AL0_P}) + Z(\text{IDDT}) + Z(\text{MolProbit} - \text{Score})) \right. \right. \\ \left. \left. + \frac{1}{8} \times (Z(\text{SphGr}) + Z(\text{CAD}_{\text{AA}})) \right. \right. \\ \left. \left. + \frac{1}{6} \times (Z(\text{GDT_HA}) + Z(\text{QSE}) + Z(\text{reLLG_const})) \right) \right), w_{\text{EU}} = 1/\text{EUs in this target} \end{aligned}$$

Trimming off lower-end tails and imputing missing values



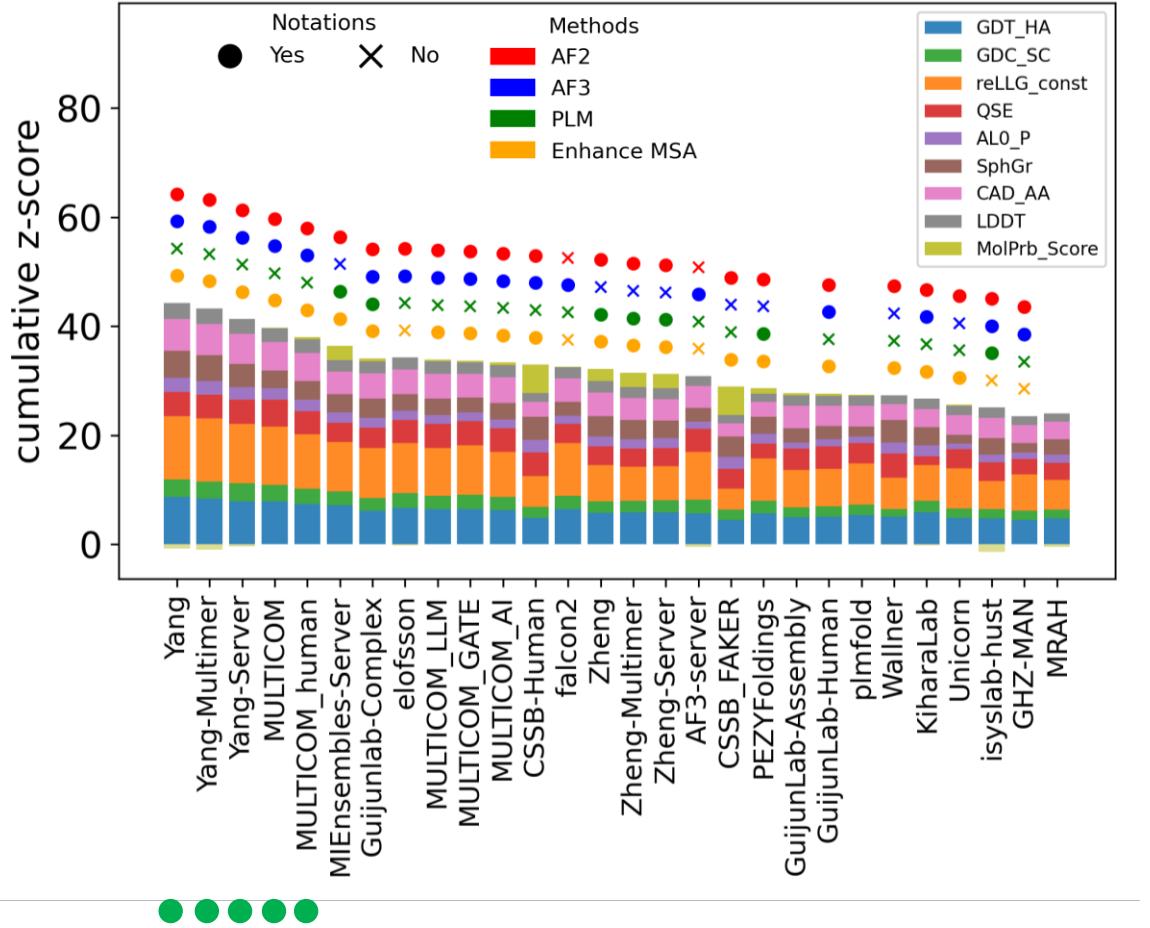
Head-to-head tests



Bootstrap

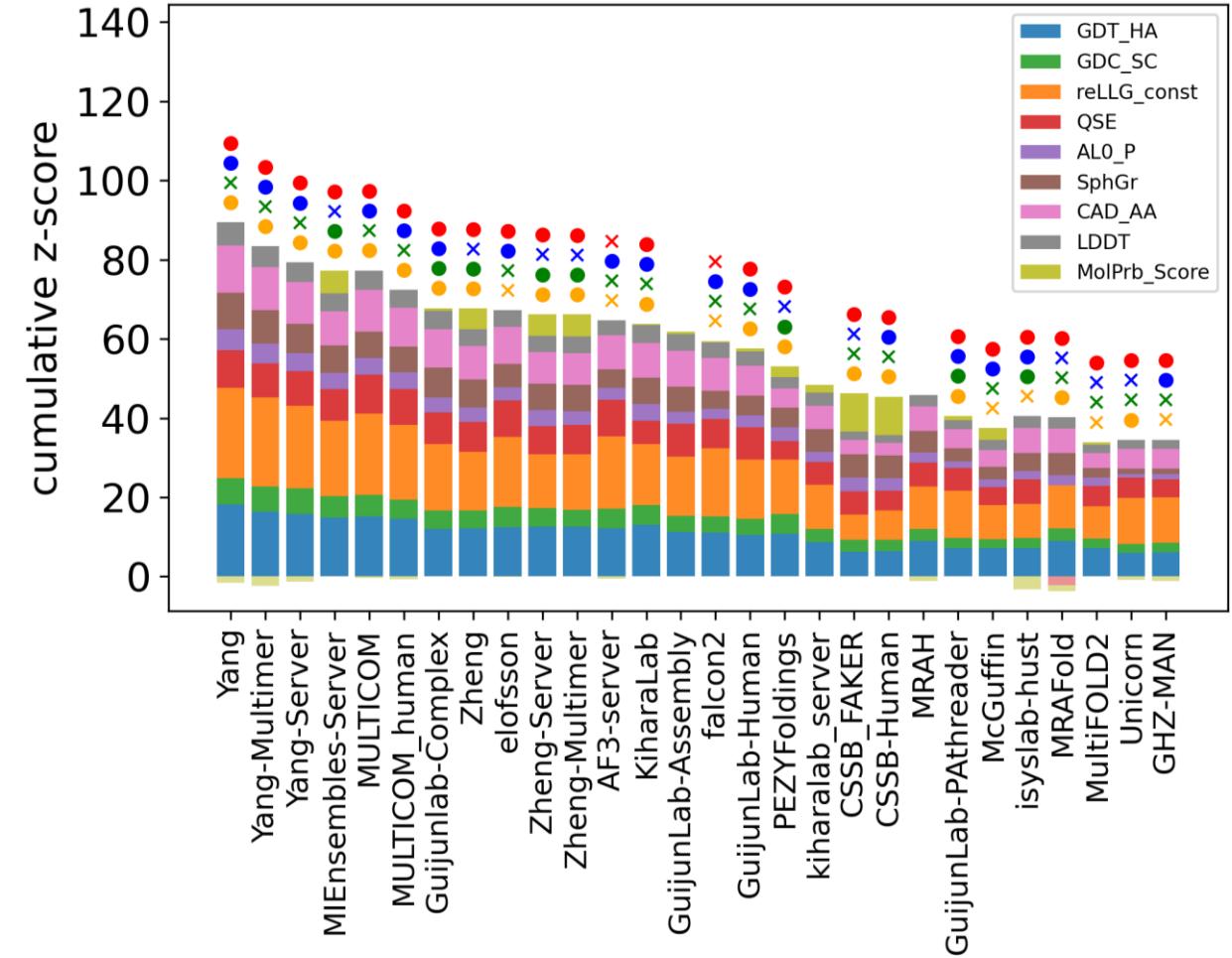
Morale: want to win, please, please, predict all targets

Ranking by best models for phase 1 targets



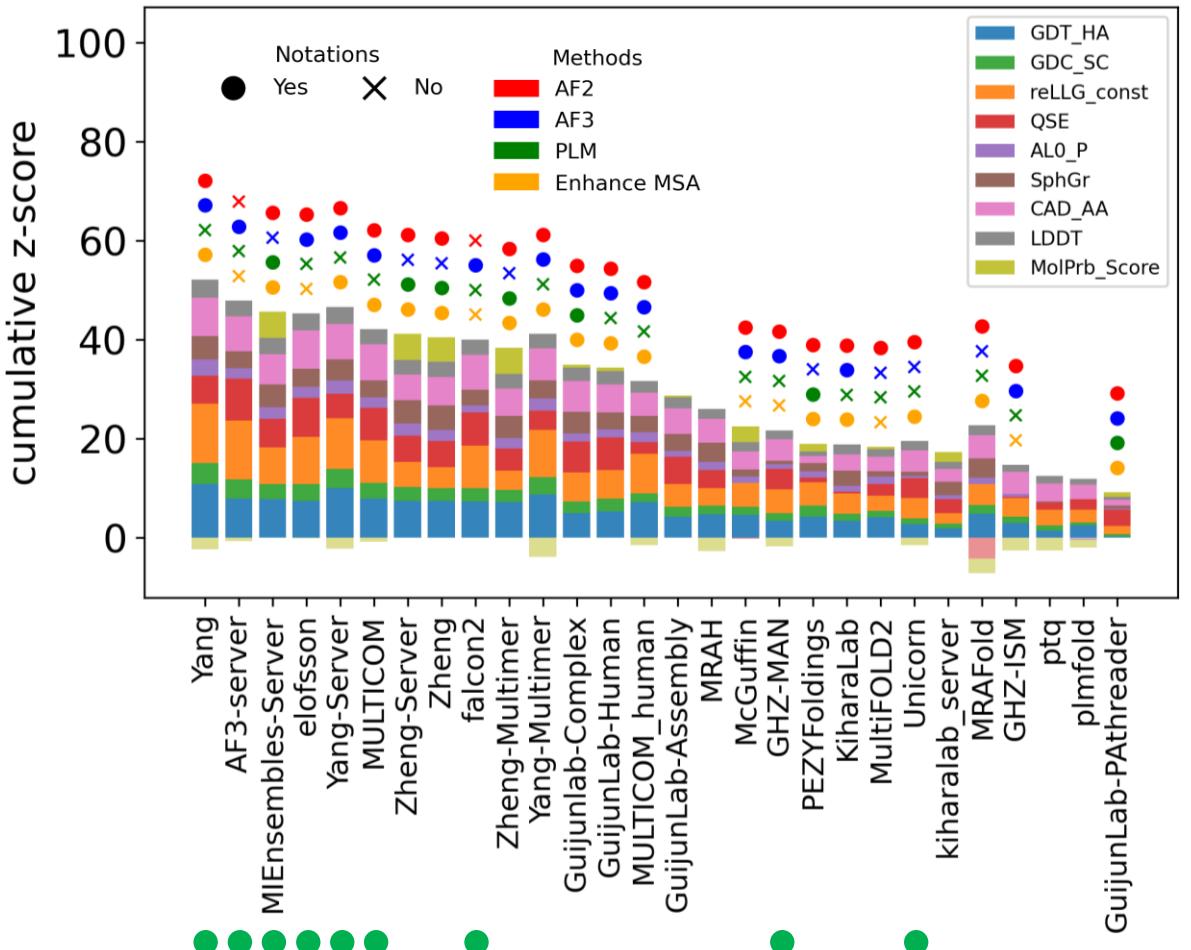
statistical significance (red is significant)

Best models for targets from all phases combined



statistical significance (red is significant)

First models for targets from all phases combined

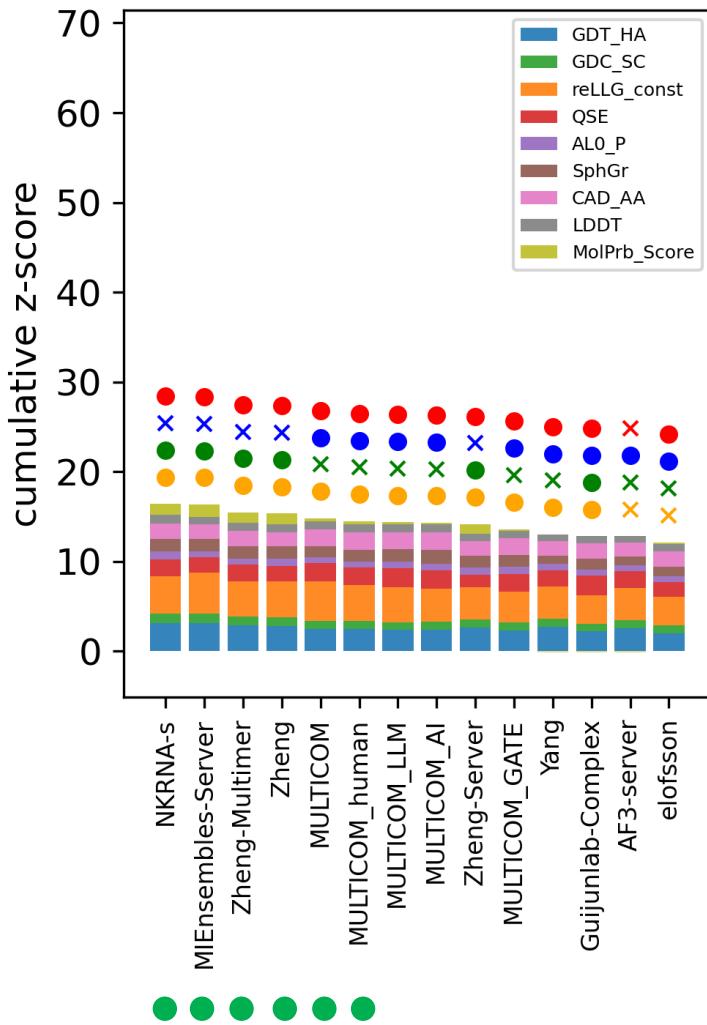


Yang	0.00	0.46	0.68	0.80	0.79	0.91	0.71	0.99	0.93	0.95	0.96	0.80	0.99	0.97	1.00	0.99	0.98	0.96	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00									
Yang-Server	0.54	0.00	0.65	0.82	0.70	0.86	0.74	0.98	0.93	0.94	0.95	0.78	0.99	0.98	1.00	1.00	0.98	0.96	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00									
AF3-server	0.32	0.35	0.00	0.66	0.77	0.86	0.94	0.86	0.87	0.87	0.88	0.98	0.95	0.97	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
MIEnsembles-Server	0.20	0.18	0.34	0.00	0.54	0.69	0.75	0.71	0.78	0.94	0.95	0.85	0.99	0.94	0.97	0.98	0.92	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00									
elofsson	0.21	0.30	0.23	0.46	0.00	0.76	0.85	0.76	0.80	0.78	0.90	0.94	0.98	0.99	0.97	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
falcon2	0.10	0.14	0.14	0.31	0.24	0.00	0.61	0.58	0.42	0.59	0.65	0.73	0.70	0.82	0.92	0.96	0.96	0.94	0.96	0.96	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
GHZ-MAN	0.29	0.26	0.06	0.25	0.15	0.39	0.00	0.57	0.47	0.46	0.66	0.82	0.79	0.71	0.69	0.72	0.94	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00								
Yang-Multimer	0.01	0.02	0.14	0.29	0.24	0.42	0.43	0.00	0.54	0.59	0.61	0.57	0.78	0.83	0.92	0.95	0.86	0.88	0.88	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
MULTICOM	0.07	0.07	0.13	0.22	0.20	0.58	0.53	0.47	0.00	0.49	0.53	0.62	0.72	0.76	0.87	0.90	0.91	0.94	0.92	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00						
Zheng-Server	0.05	0.06	0.13	0.06	0.24	0.41	0.54	0.41	0.51	0.00	0.57	0.64	0.86	0.80	0.87	0.91	0.76	0.92	0.93	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00							
Zheng	0.04	0.05	0.12	0.04	0.22	0.35	0.34	0.39	0.47	0.43	0.00	0.50	0.81	0.72	0.83	0.86	0.75	0.85	0.85	0.98	0.99	0.99	1.00	0.99	0.99	0.99	0.99	1.00	1.00						
Unicorn	0.20	0.22	0.02	0.14	0.07	0.27	0.18	0.43	0.38	0.36	0.50	0.00	0.69	0.63	0.61	0.59	0.84	0.95	1.00	0.92	0.98	0.99	0.98	0.99	0.99	0.99	0.99	0.99	1.00						
Zheng-Multimer	0.01	0.01	0.05	0.01	0.11	0.24	0.21	0.22	0.28	0.14	0.19	0.31	0.00	0.64	0.76	0.79	0.67	0.73	0.76	0.95	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.98	1.00					
GuijunLab-Human	0.03	0.02	0.03	0.06	0.06	0.18	0.29	0.17	0.24	0.20	0.28	0.37	0.36	0.00	0.69	0.77	0.72	0.76	0.78	0.98	0.99	0.97	0.99	0.98	0.99	0.99	0.99	0.99	0.99	1.00					
Guijunlab-Complex	0.00	0.00	0.00	0.00	0.03	0.02	0.08	0.31	0.08	0.13	0.13	0.17	0.39	0.24	0.31	0.00	0.72	0.56	0.81	0.80	0.81	0.96	0.99	0.96	0.98	0.95	0.97	0.98	0.99	0.99	1.00				
GuijunLab-Assembly	0.01	0.00	0.00	0.00	0.01	0.10	0.04	0.28	0.05	0.11	0.09	0.14	0.41	0.21	0.23	0.28	0.00	0.52	0.79	0.82	0.72	0.93	0.98	0.93	0.97	0.95	0.97	0.98	0.99	0.99	1.00				
plmfold	0.02	0.02	0.02	0.08	0.03	0.04	0.07	0.14	0.09	0.24	0.25	0.16	0.34	0.28	0.43	0.48	0.00	0.44	0.48	0.67	0.77	0.91	0.89	0.94	0.92	0.88	0.94	0.94	0.94	0.94					
ptq	0.04	0.04	0.04	0.00	0.02	0.01	0.06	0.00	0.12	0.06	0.08	0.15	0.05	0.27	0.24	0.19	0.21	0.56	0.00	0.53	0.55	0.88	0.90	0.91	0.93	0.90	0.90	0.90	0.90	0.96	0.98				
GHZ-ISM	0.02	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.69	0.87	0.89	0.92	0.91	0.90	0.89	0.90	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	
MULTICOM_human	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.02	0.01	0.02	0.08	0.05	0.13	0.19	0.28	0.33	0.45	0.31	0.00	0.80	0.86	0.88	0.89	0.86	0.89	0.89	0.95	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98
MRAH	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00			
McGuffin	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
PEZYFoldings	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
GuijunLab-PAthreader	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
Kiharalab_server	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
KiharaLab	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
MultiFOLD2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			
MRAFold	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00			

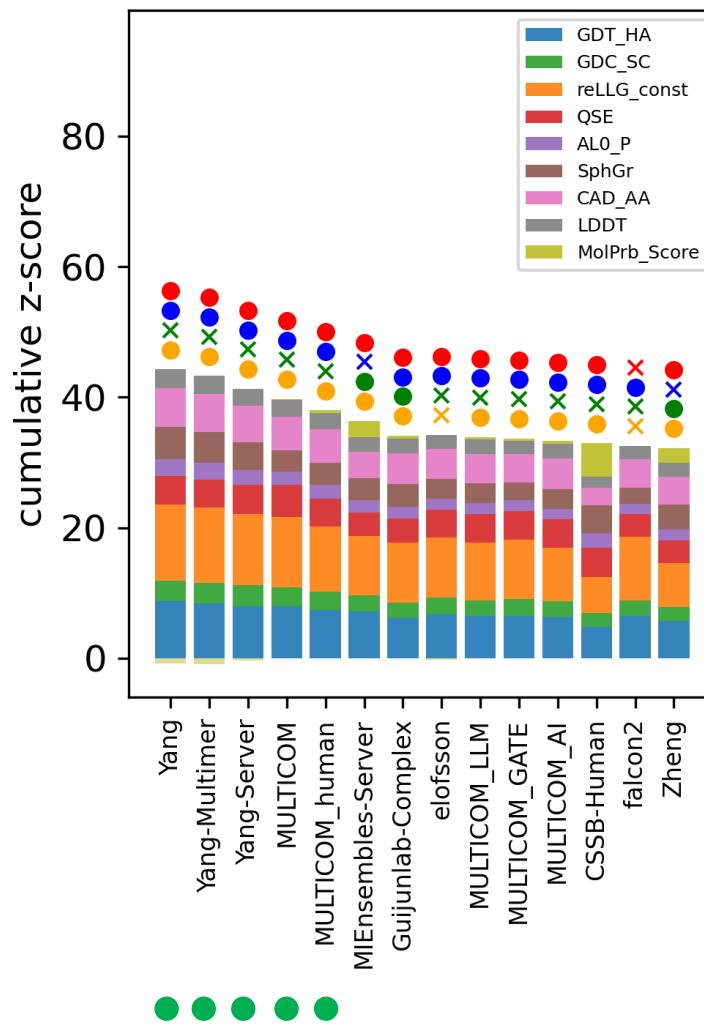
statistical significance (red is significant)

Best models by phase: T0, T1, T2

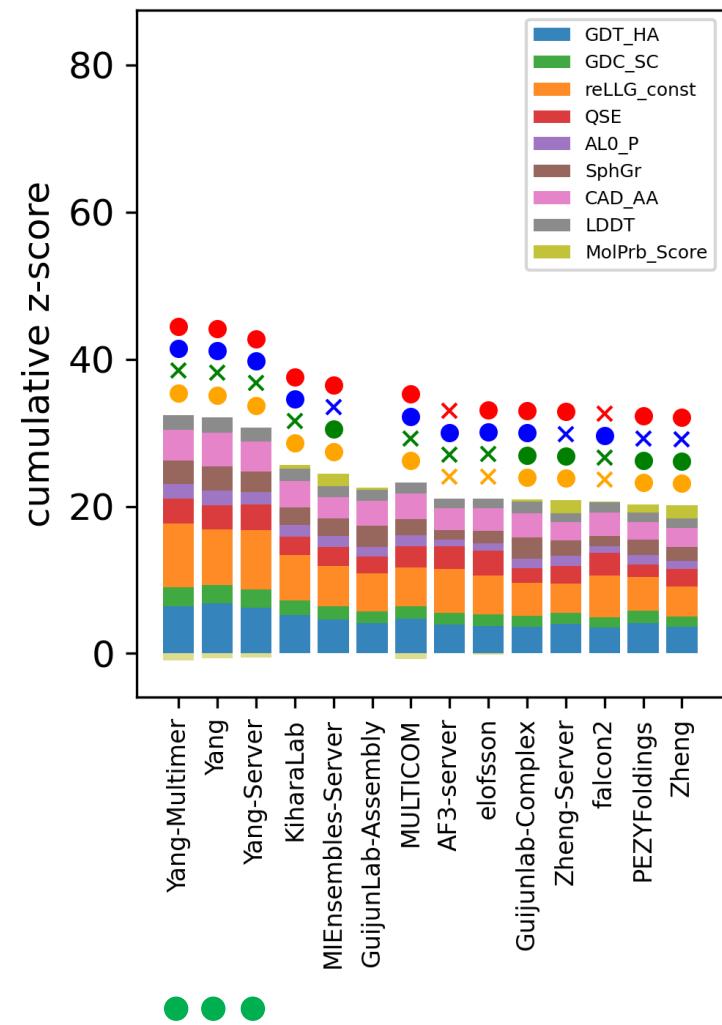
T0



T1

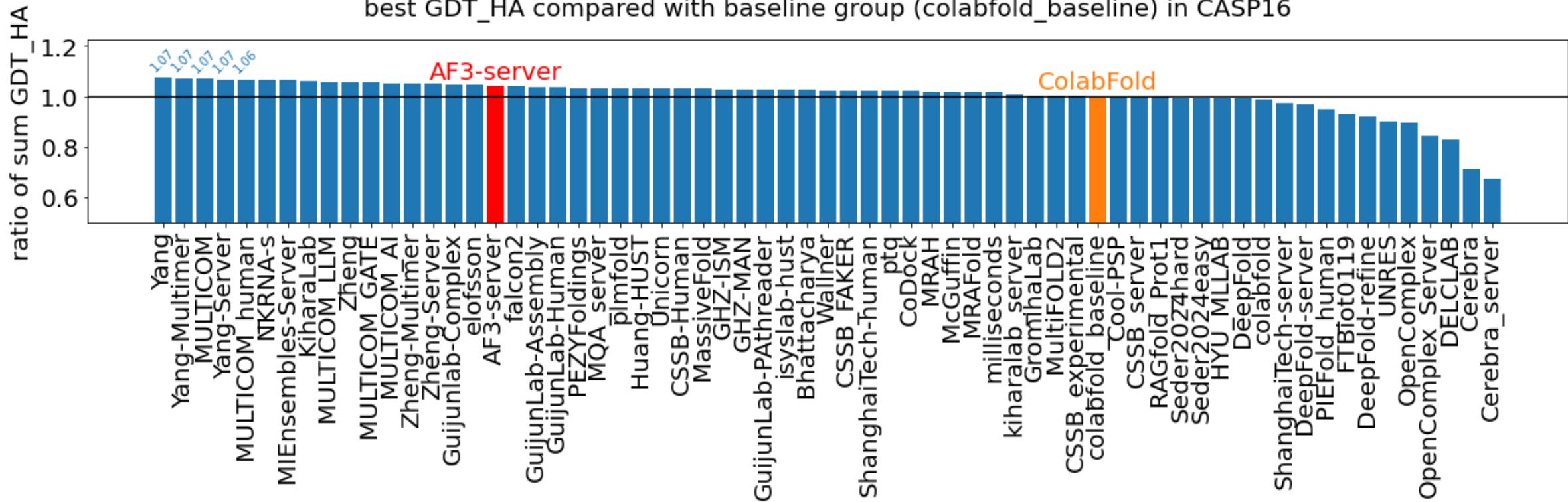


T2

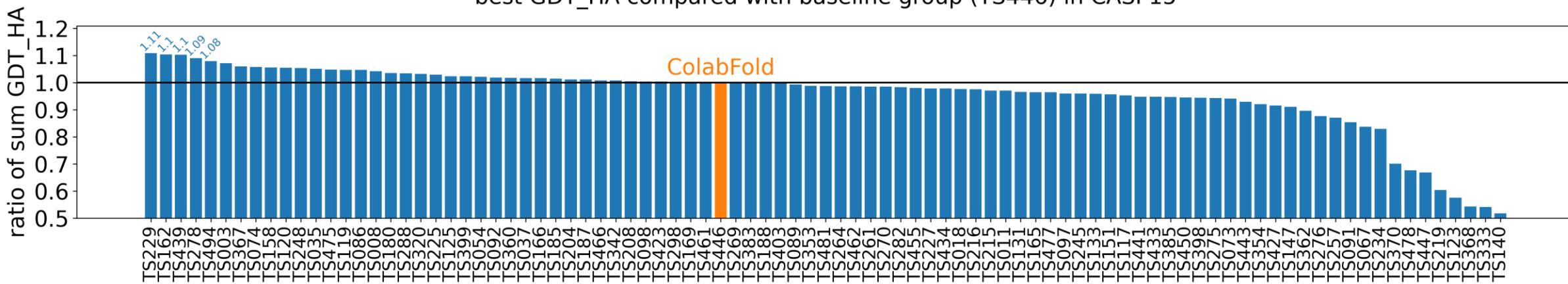


Is there measurable progress in the field?

best GDT_HA compared with baseline group (colabfold_baseline) in CASP16

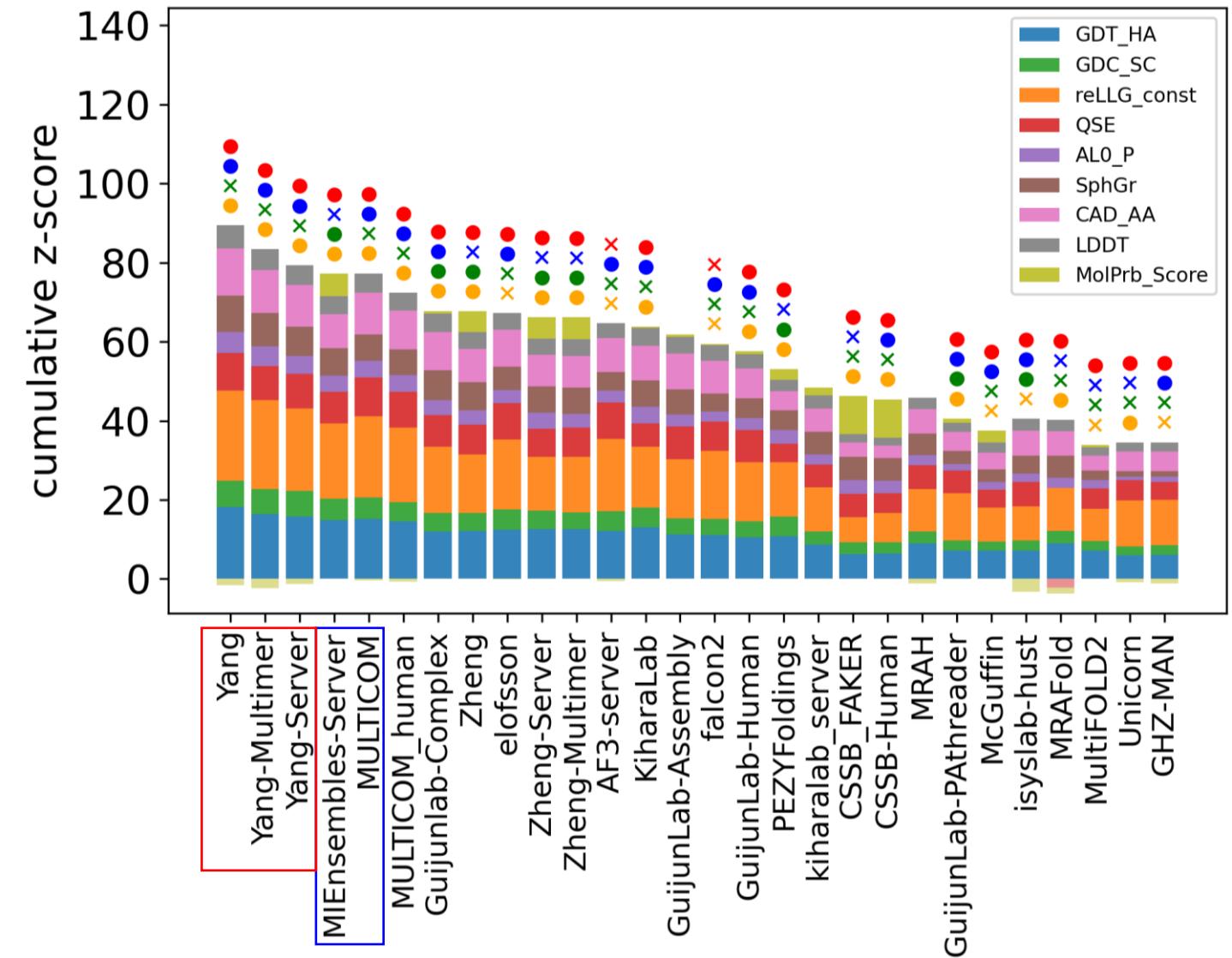


best GDT_HA compared with baseline group (TS446) in CASP15



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!).