EMA assessment for CASP16

Randy Read, Gabriel Studer and Alisia Fadini

UNIVERSITY OF CAMBRIDGE

Randy J Read
Department of Haematology

CIMR
Molecules
Mechanisms
Medicine

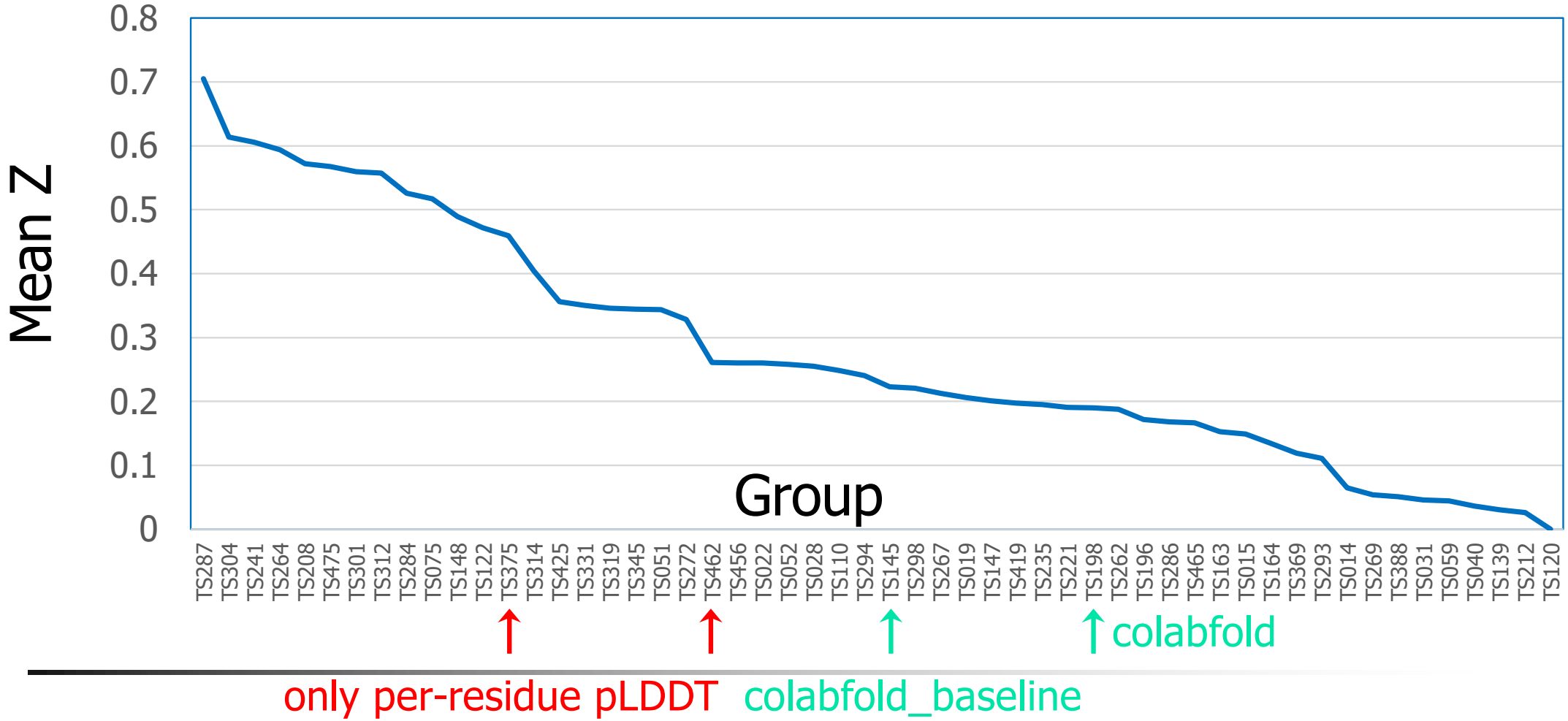
Overview

- Assessment of pLDDT
  - accuracy self-assessment
- QMODE1/2: Gabriel Studer
- QMODE3: Alisia Fadini

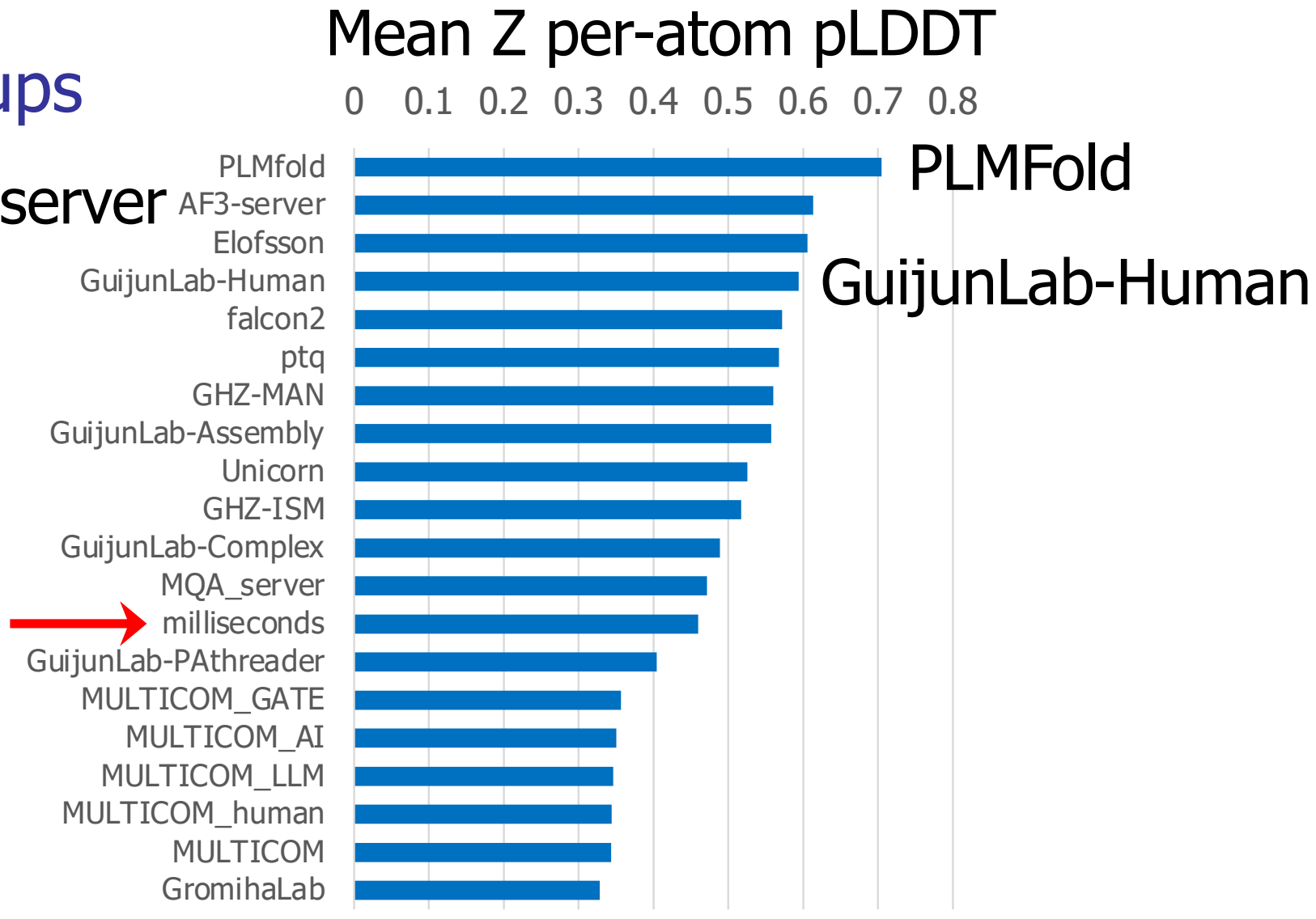
Measuring and ranking pLDDT accuracy

- Ground truth from OpenStructure calculation of per-atom LDDT values (feature added by Gabriel Studer)
- RMSD: actual numerical values matter
  - CASP15 ASE used absolute value of difference
- 93 domain evaluation unit targets from proteins
  - 54 predictors that participated for at least 80% of targets
    - 24 predictors who provided finer-grained pLDDT values

Ranking of per-atom LDDT prediction by RMSD



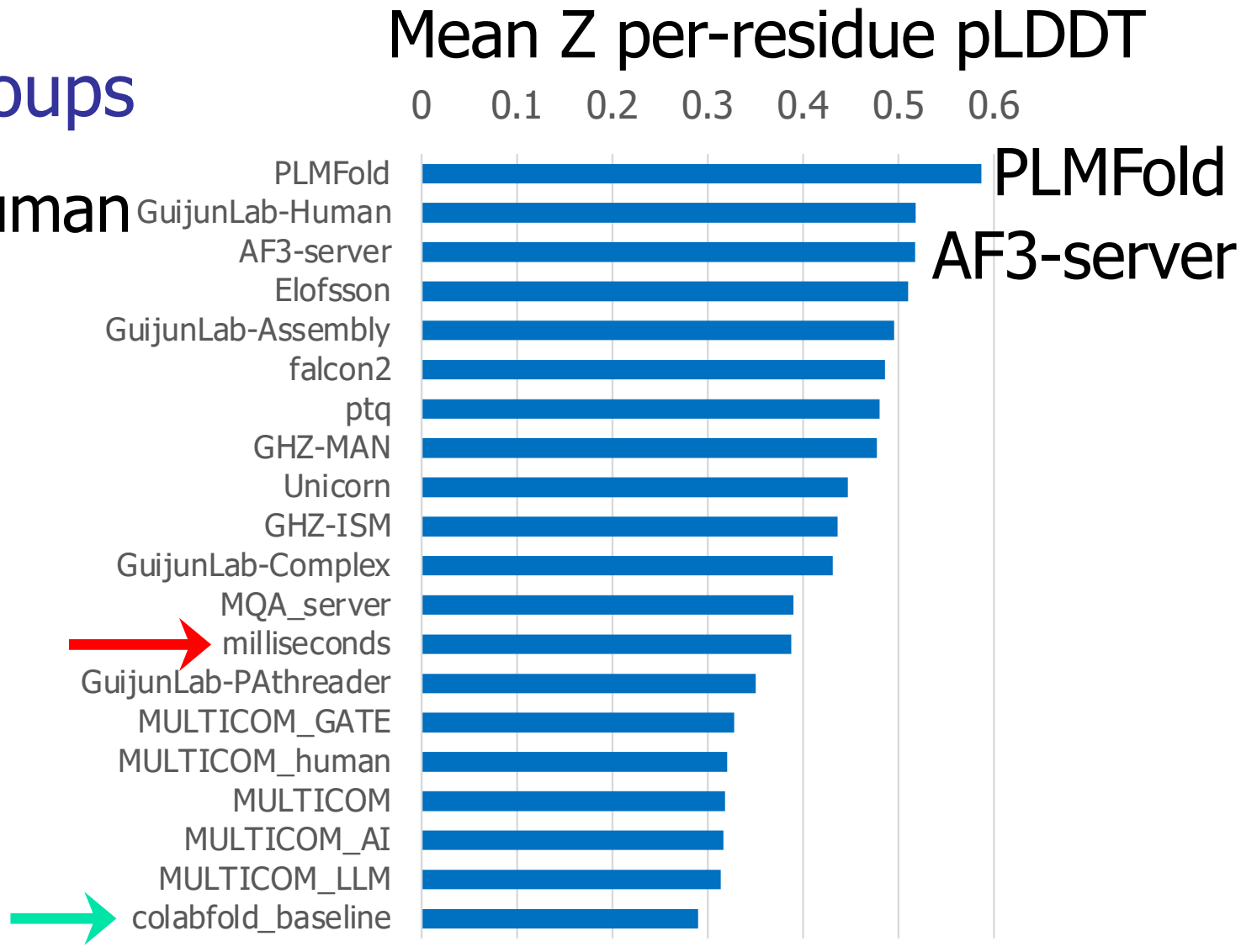
Top 20 groups



AF3-server

Top 20 groups

GuijunLab-Human



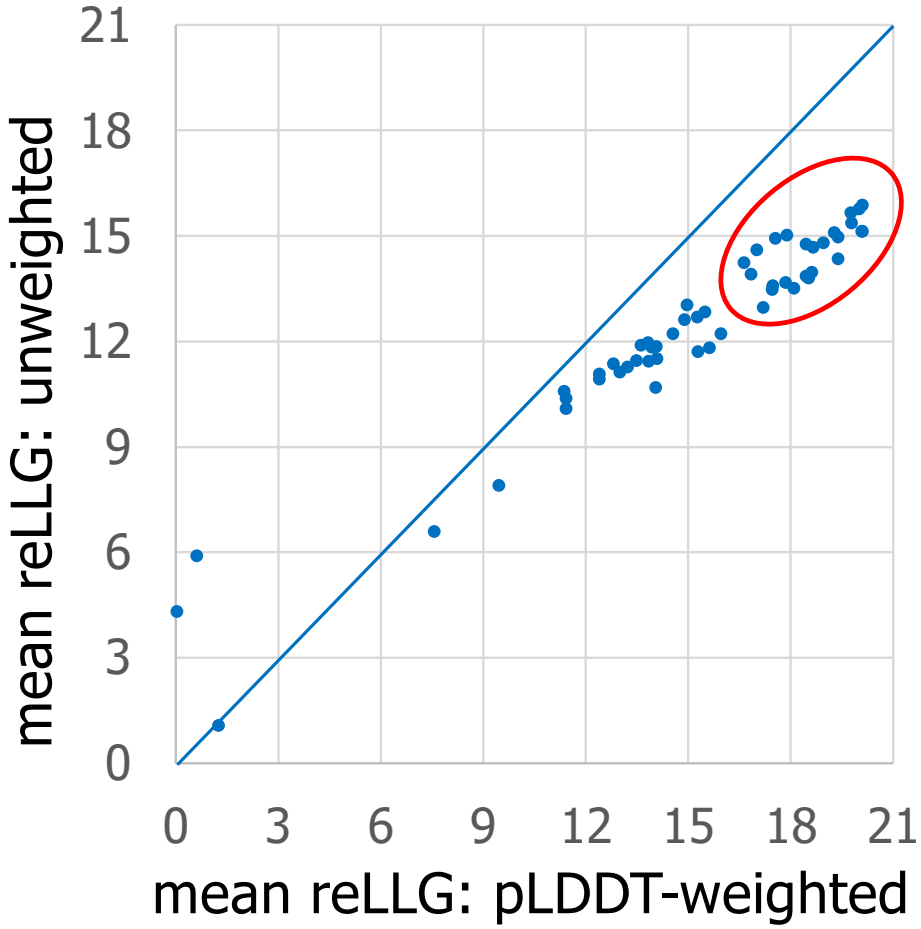
Conclusions from Accuracy Self-Assessment

- AF3 estimates of pLDDT better than AF2
- Most of the best results from unadulterated AF3
- PLMFold could be doing something interesting!

Utility for crystallographic molecular replacement (MR) and cryo-EM docking

- MR and cryo-EM docking both evaluated by log-likelihood-gain (LLG) for how well model explains experimental data
  - LLG scores for MR and docking are closely related
- Compute relative expected LLG scores → reLLG
  - fraction of gold-standard LLG score compared to experimental model
- Evaluate value added by pLDDT
  - ignore pLDDT, set B-factors to constant value
  - convert pLDDT into RMSD then to equivalent B-factor to down-weight less confident parts of model
    - per-atom: use individual atomic pLDDT values (default)
    - per-residue: average pLDDT values per residue before calculation

Confidence weighting improves MR utility



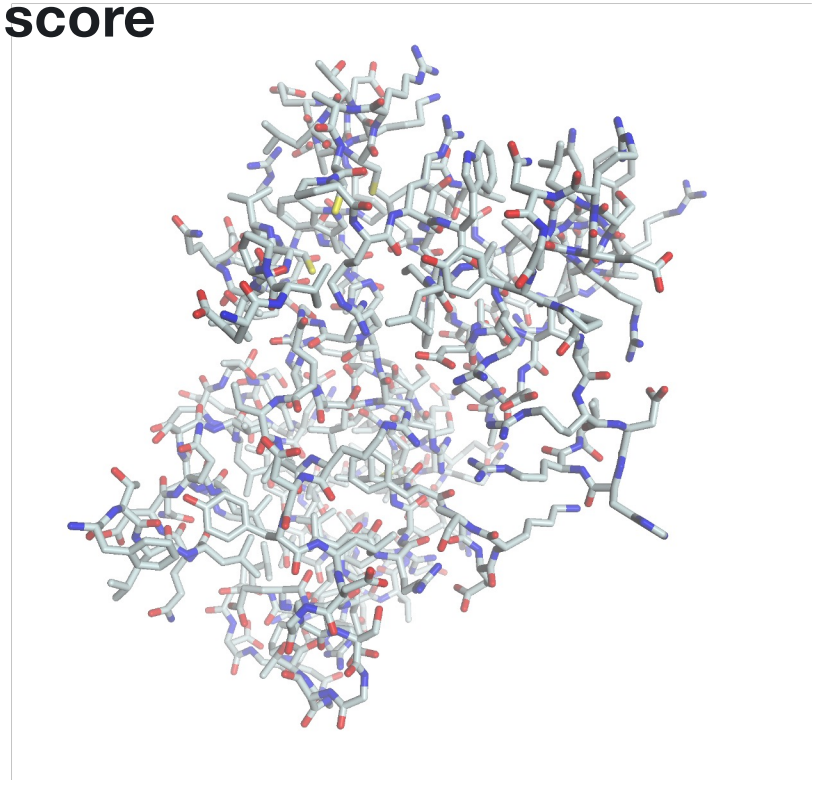
almost all AF3-derived with atomic pLDDT

QMODEs 1 & 2 (Gabriel Studer)

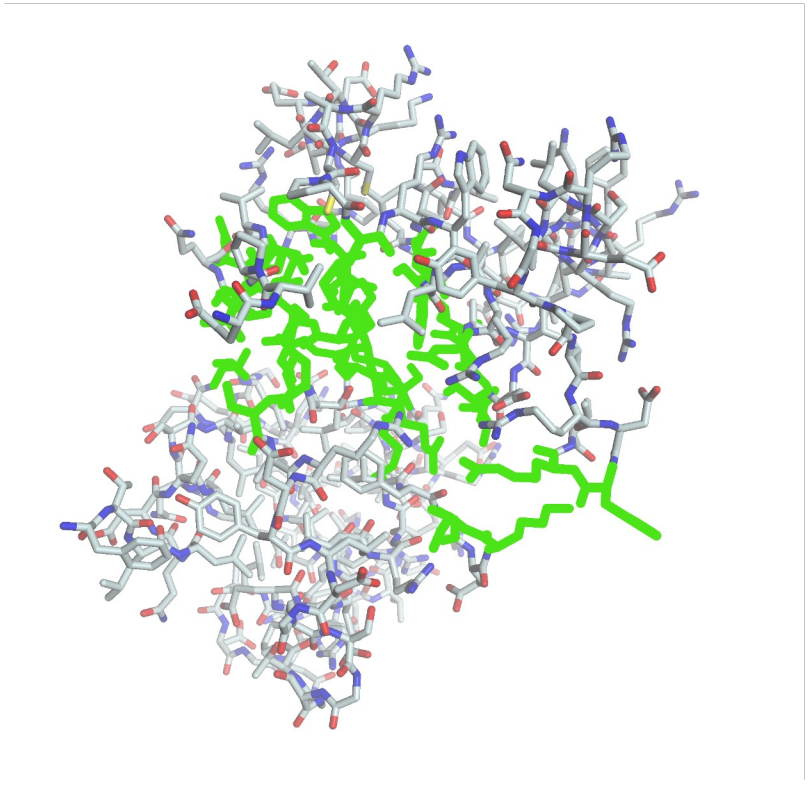
- Quality assessment for multimer models
- Evaluations based on the methods developed in OpenStructure for CASP15
- Chain mapping (a highly combinatorial problem) was improved
- Groups ranked by sum of Z-score (with zero for targets not attempted) rather than mean Z-score for targets attempted

QMODE 1 (Global)

SCORE (0-1): Reflects similarity of the full complex to the target upon global superposition - compare to: **oligo-GDTTS, TM-score**

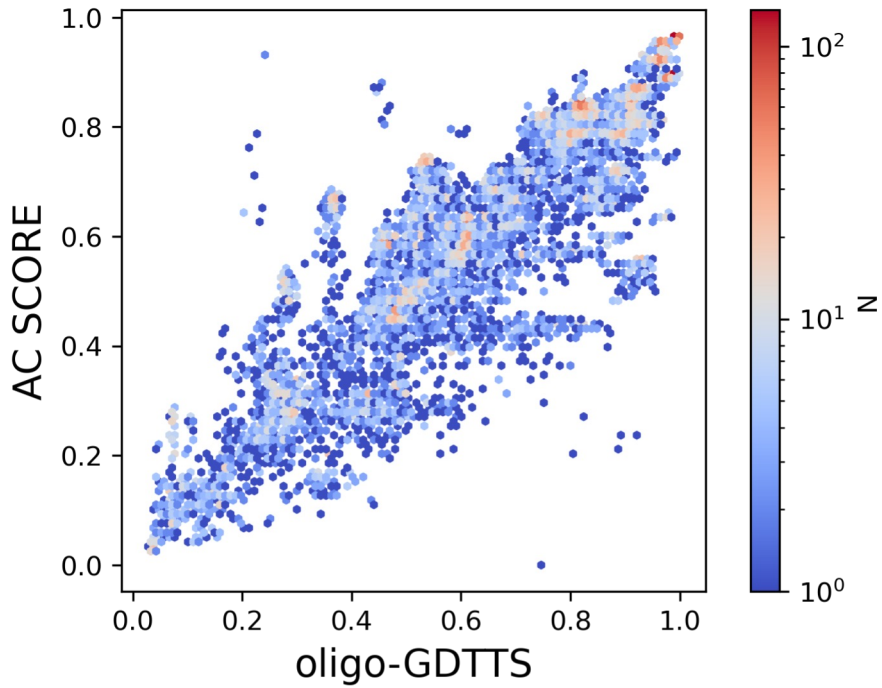


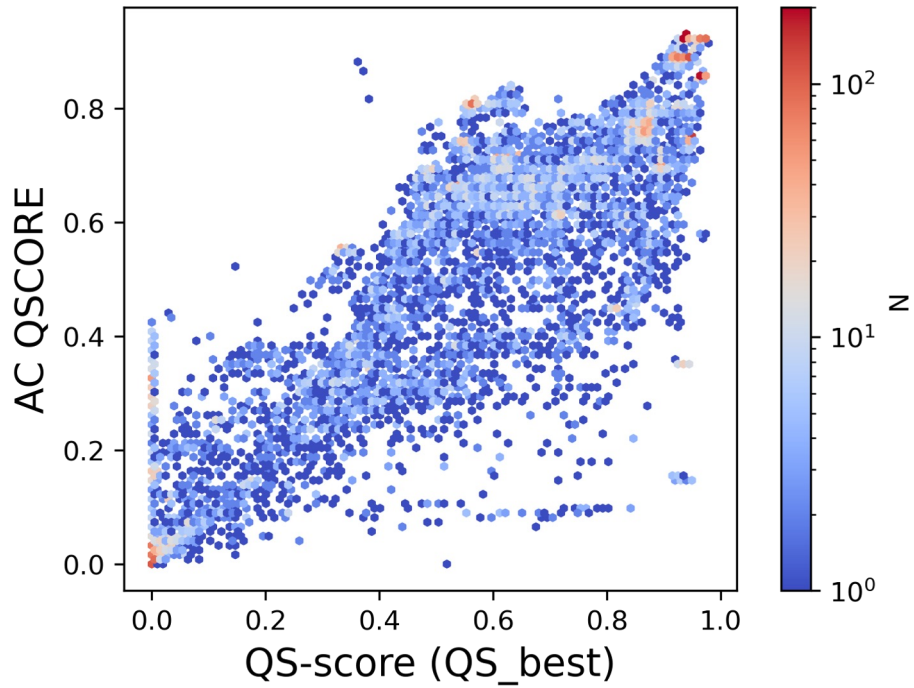
QSCORE (0-1): Interface accuracy – compare to: **QS-score** (QS-best variant), **DockQ-wave**



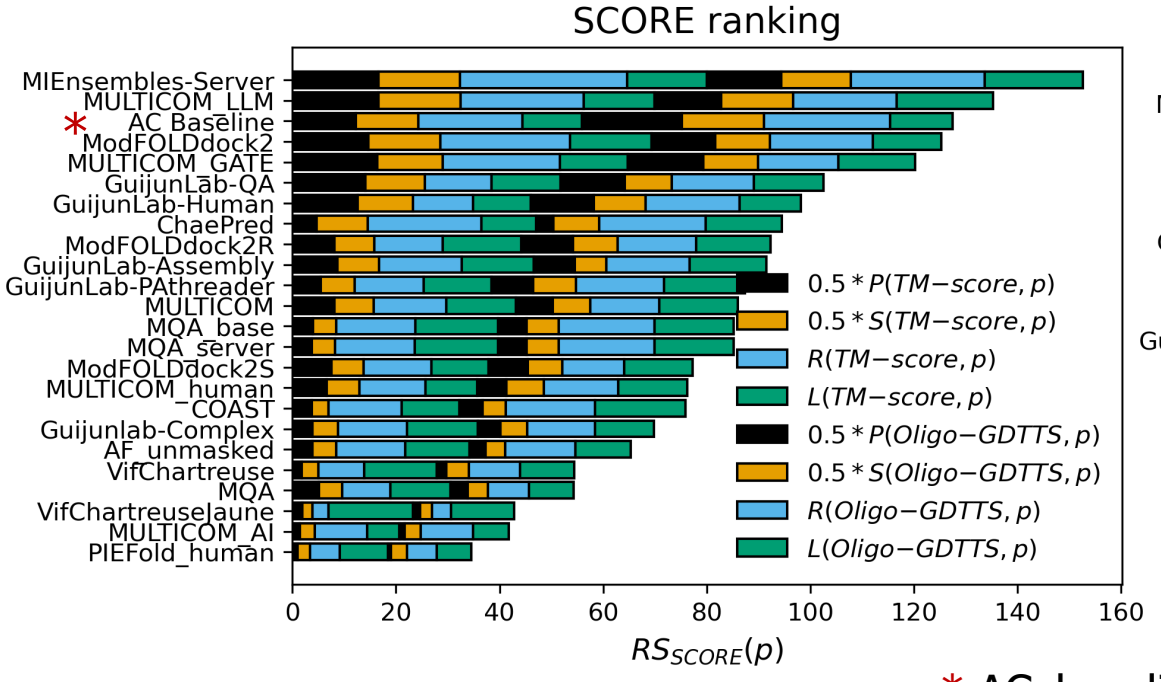
QMODE 1 (Global) - Assembly Consensus

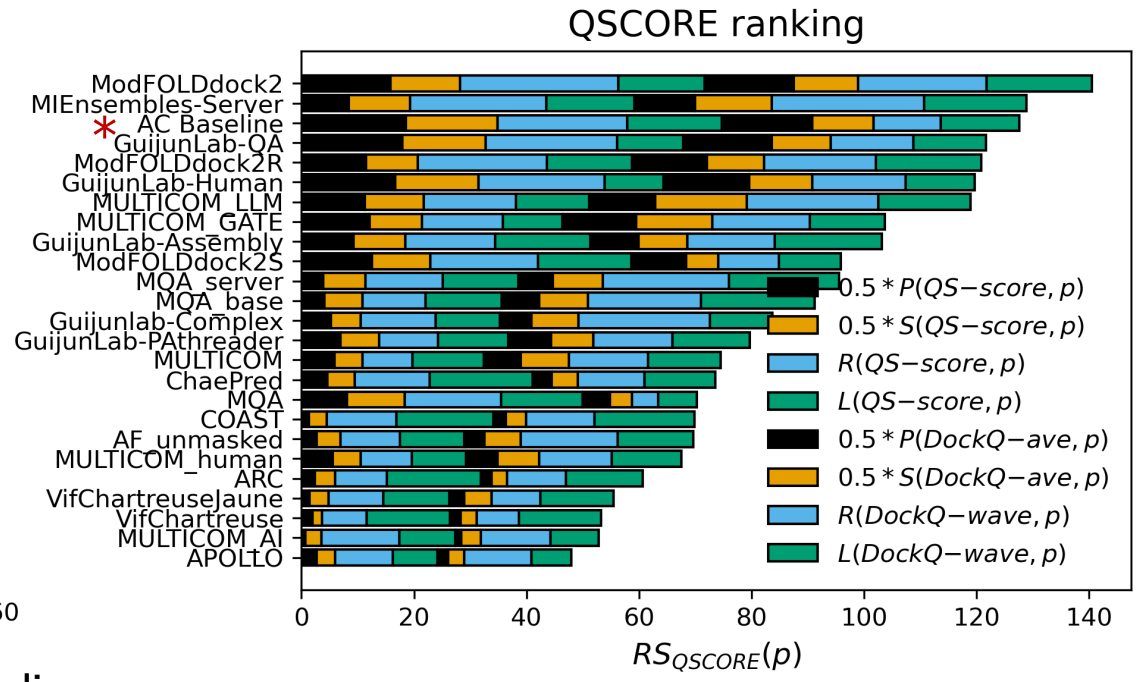
$$S(x) = \frac{1}{N-1} \sum_{y!=x} f(x,y)$$





QMODE 1 (global)



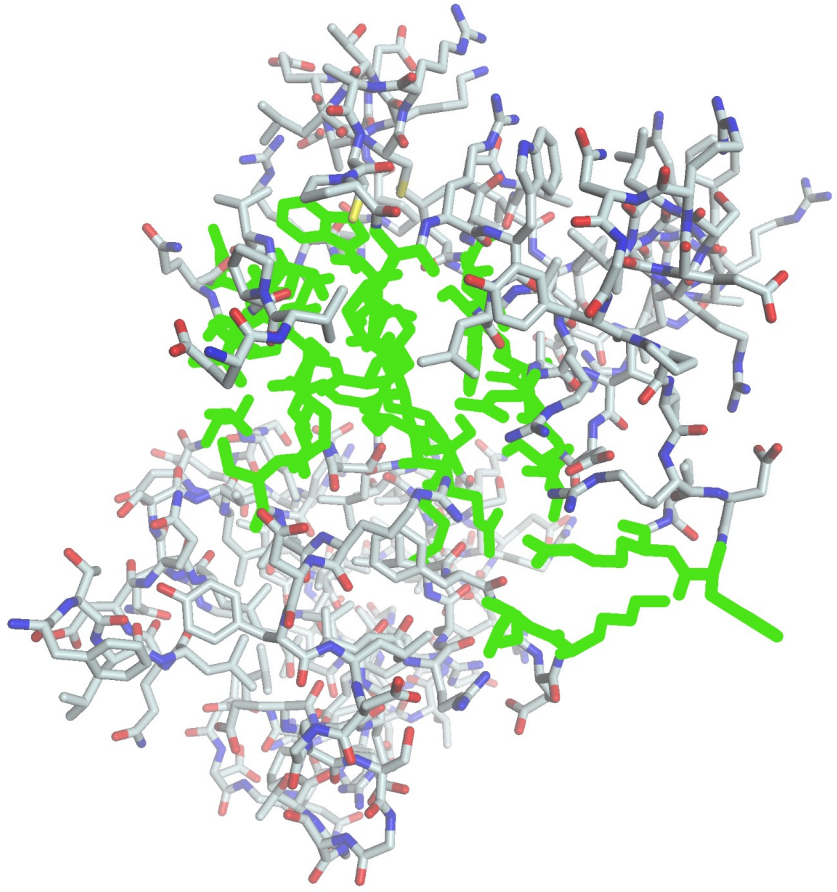


P: Pearson R
S: Spearman R
R: ROC AUC
L: Loss

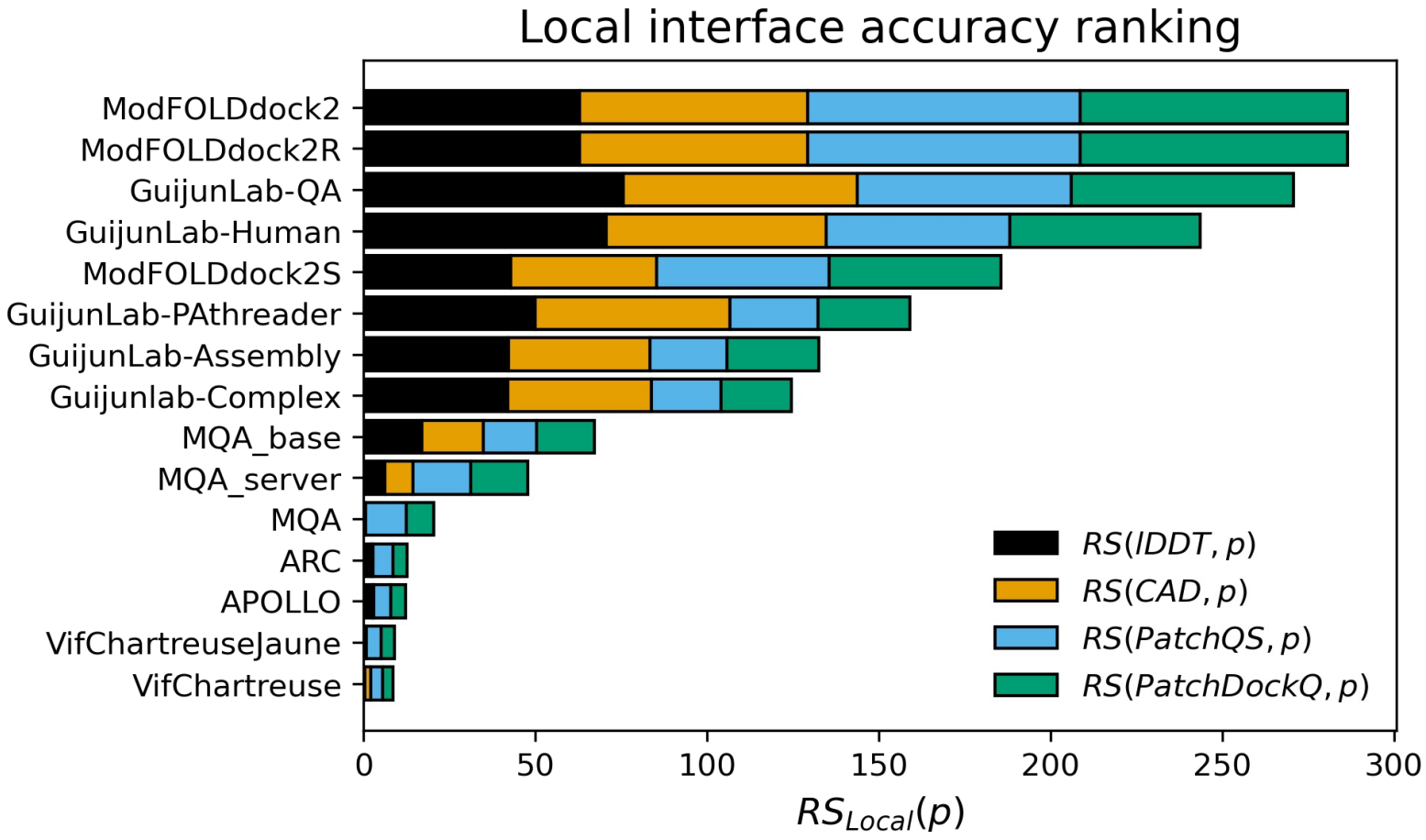
* AC_baseline
MIEnsembles-server
MULTICOM_LLM
GuijunLab_QA
ModFoldDock2

QMODE 2 (Local)

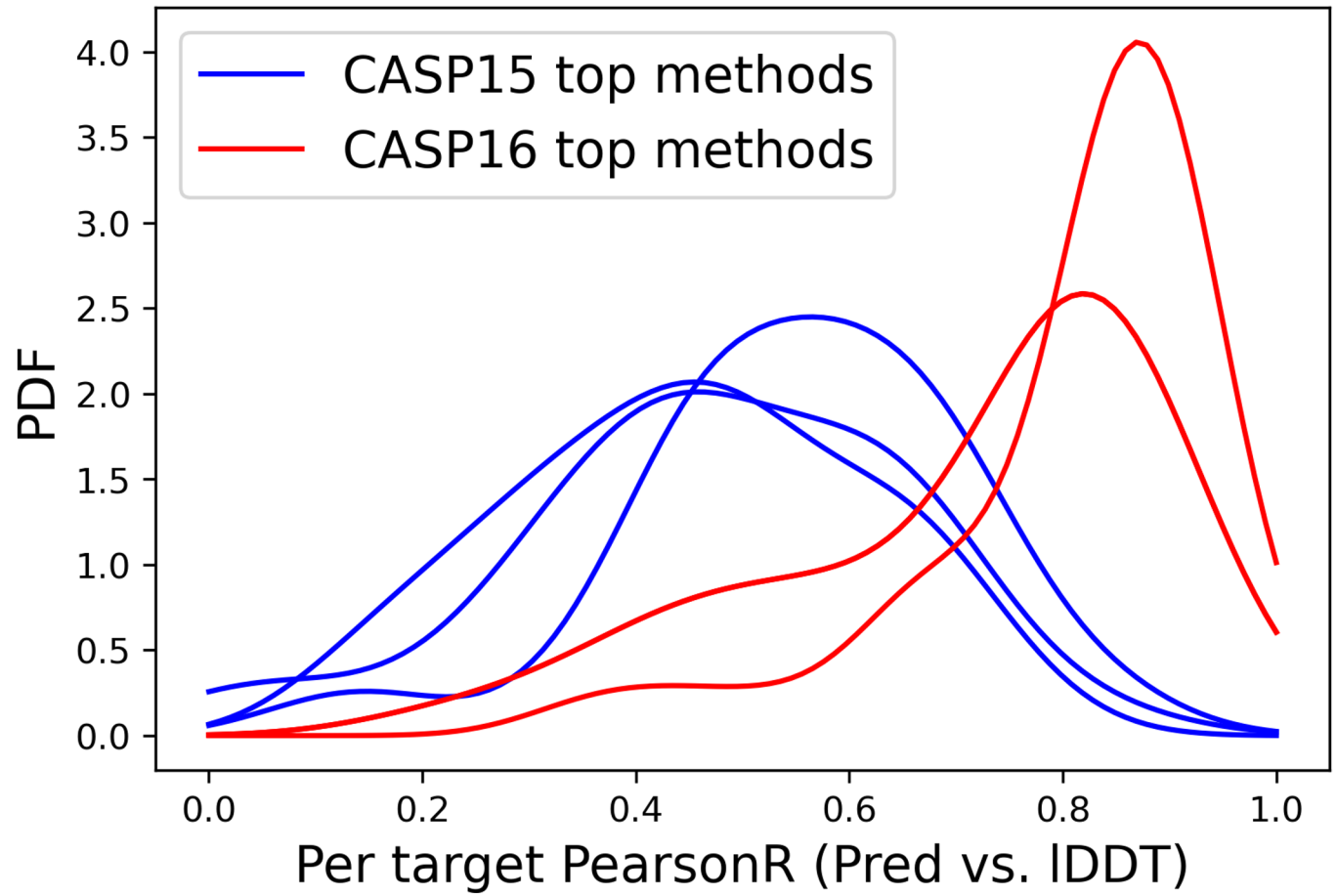
Predict local interface accuracy for interface residues - compared against:
IDDT, CAD, PatchDockQ, PatchQS



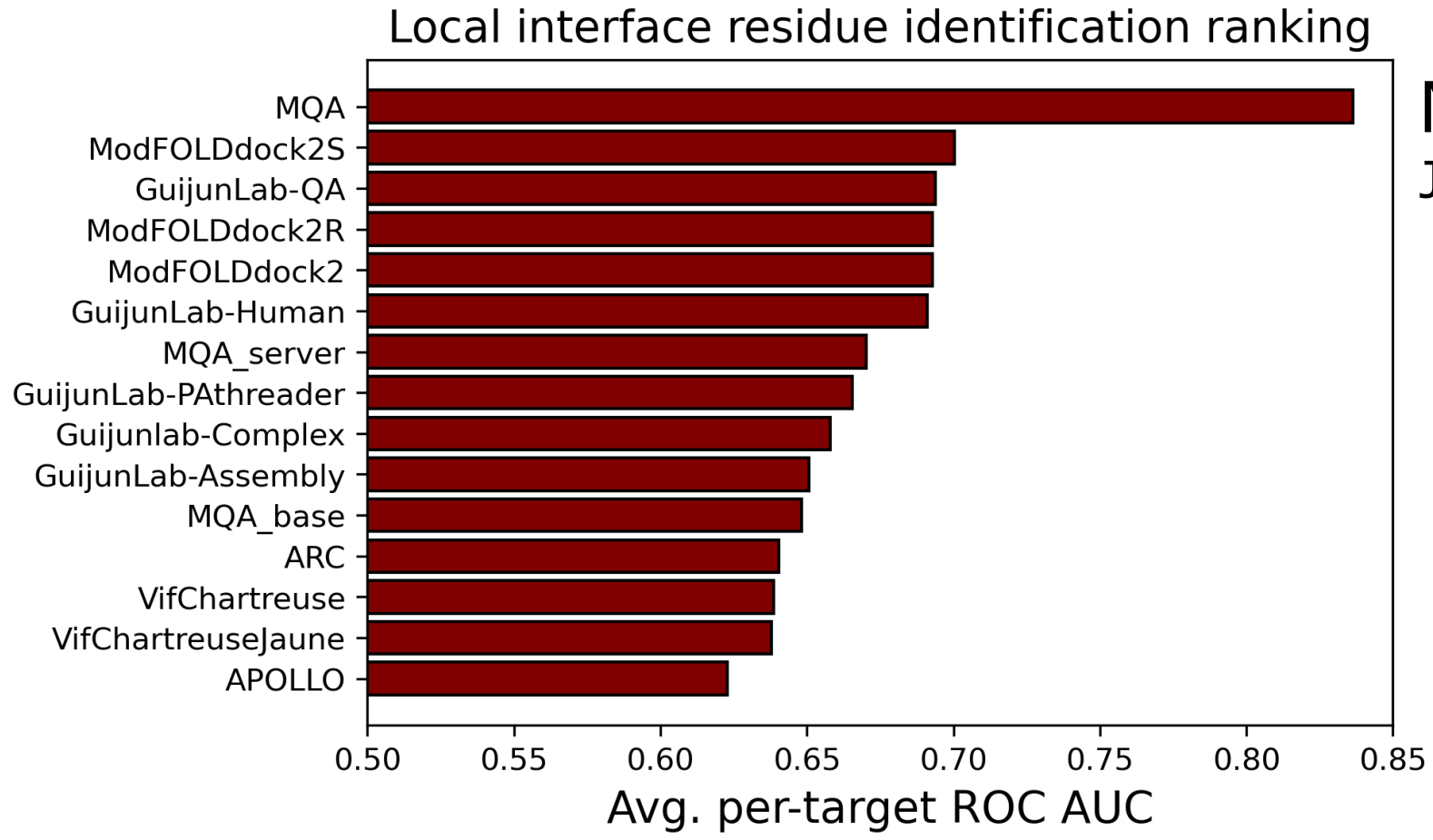
QMODE 2 (local)



QMODE 2 (Local)



Accuracy in identifying interface residues



MQA:

Jun Liu, Yang Zhang

QMODE1/2 summary

- Similar groups at top compared to CASP15
- Some signs of improvement
- Details from presentations!