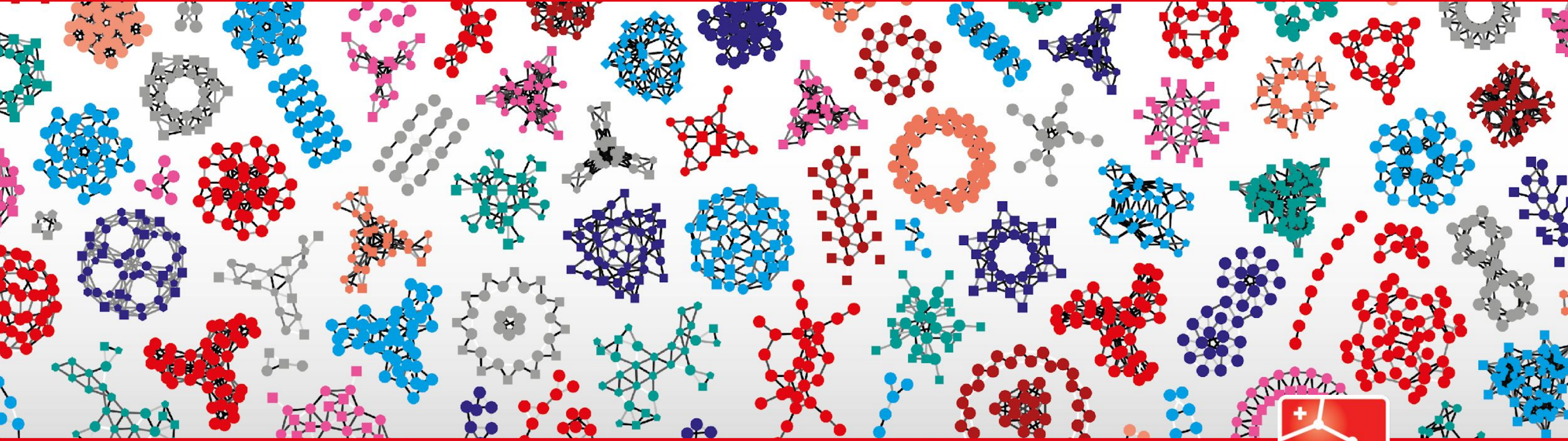


CASP15 EMA

All about assemblies

Gabriel Studer
Gerardo Tauriello
Torsten Schwede

SIB
Swiss Institute of Bioinformatics

BIOZENTRUM
Universität Basel
The Center for
Molecular Life Sciences

CASP7 - A Retrospective

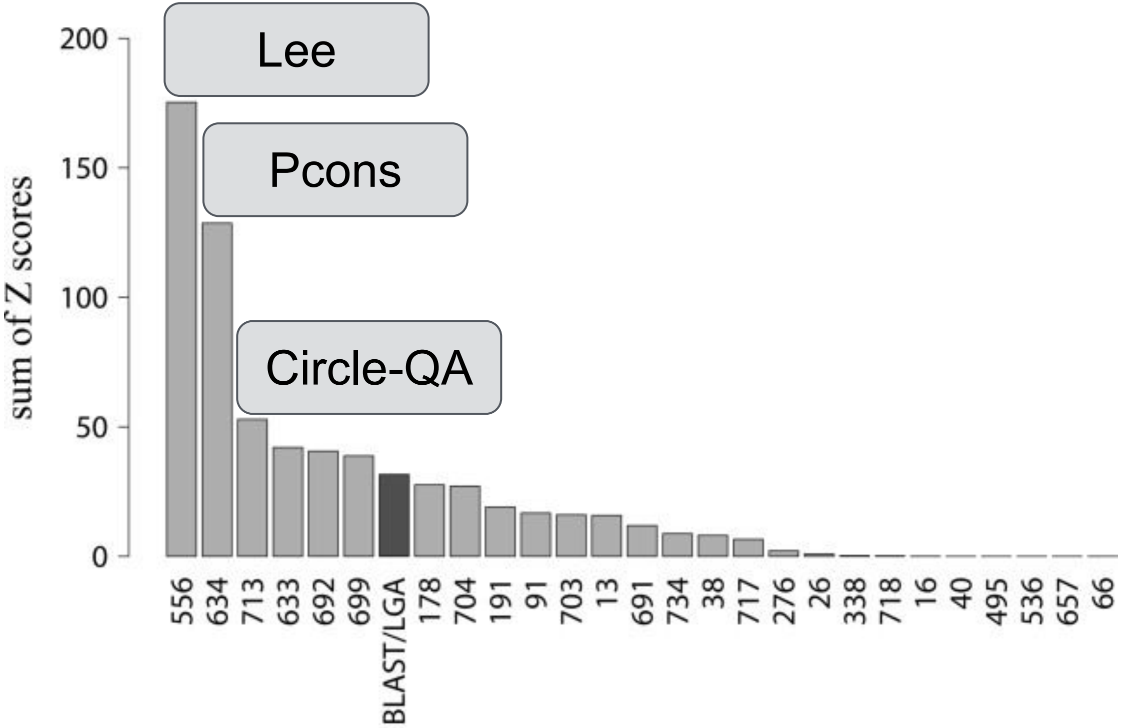


Figure 9 in Cozzetto et al., 2007

Comparison of the performance of the naïve method BLAST/LGA with all other methods submitting QM1 predictions for TBM targets.

CASP9 - A Retrospective

Consensus Baseline

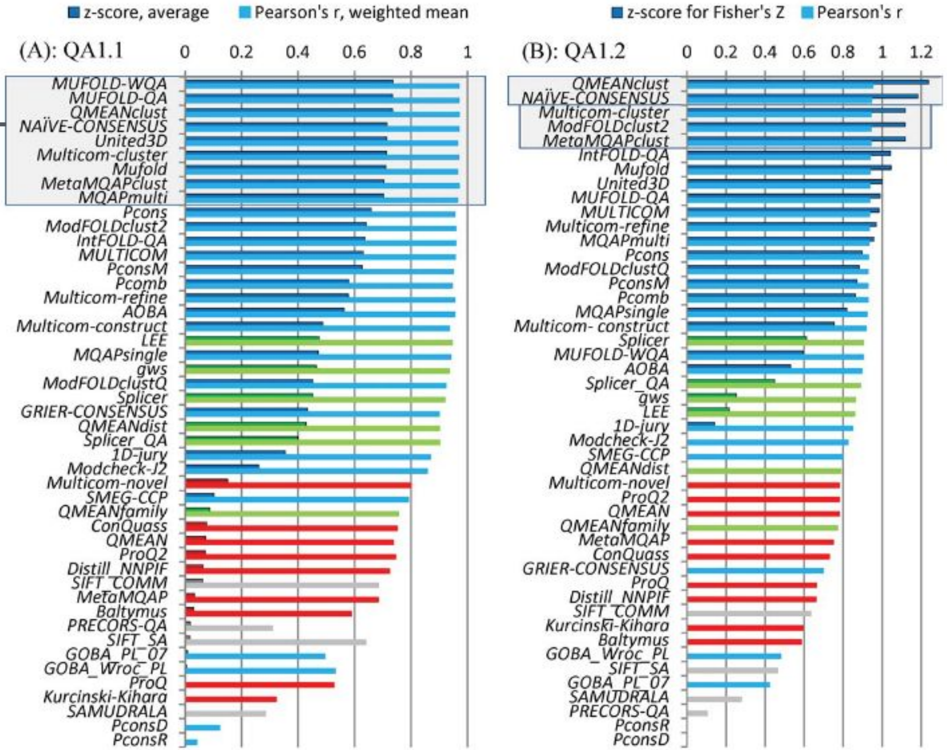
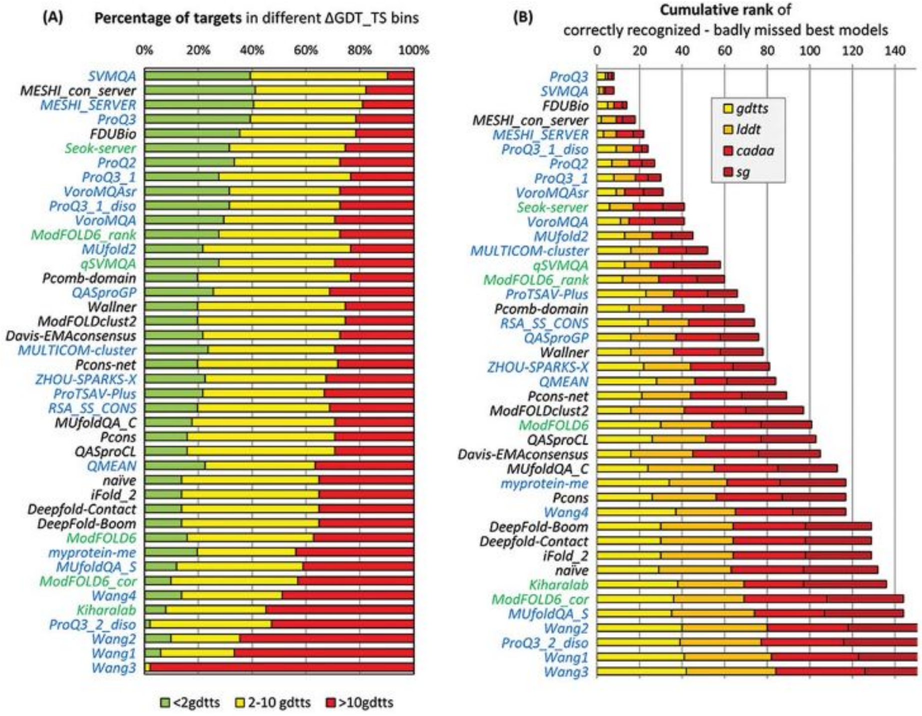


Figure 1 in Kryshtafovych et al., 2011
Consensus methods in blue, Single model methods in red and quasi single model methods in green

CASP12 - A Retrospective



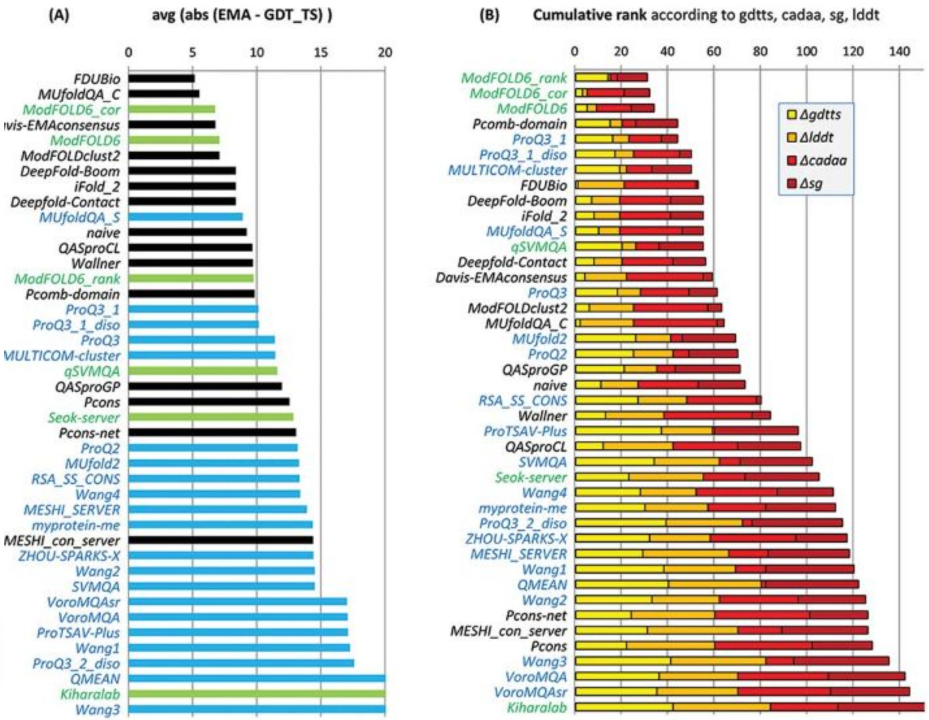


Figure 3 in Kryshtafovych et al., 2017 Figure 4 in Kryshtafovych et al., 2017
Consensus methods in black, single model methods in blue and quasi single model methods in green

From CASP14 to CASP15

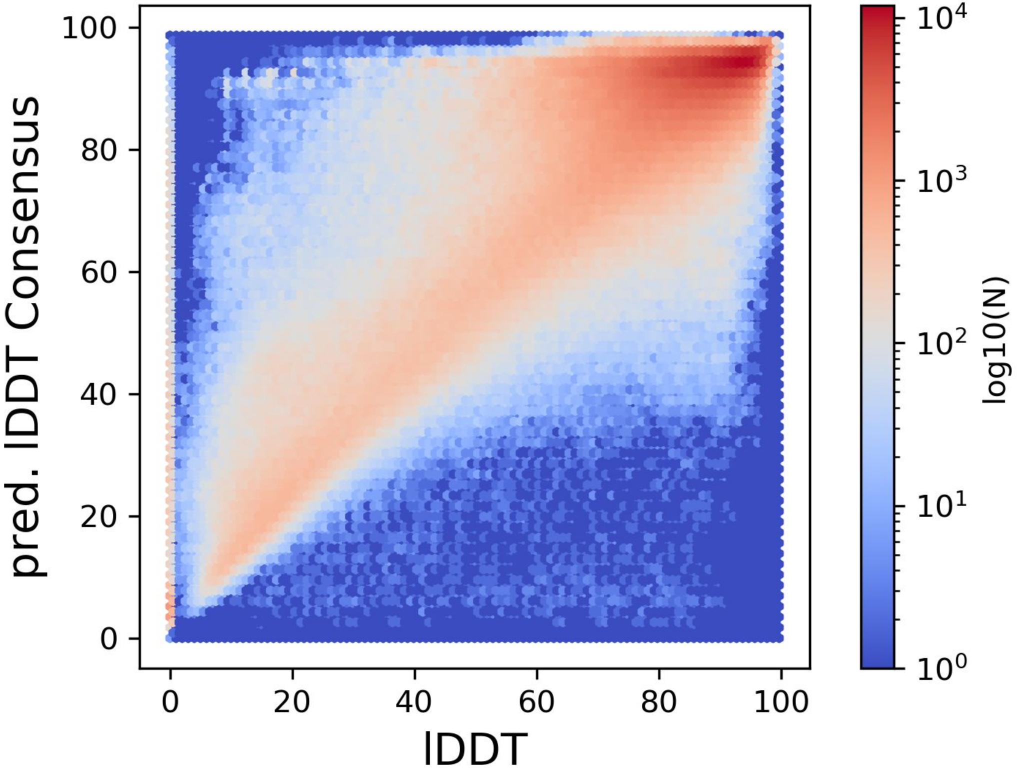
- Further improvement in single model methods
- **!!!Fully integrated self-assessment!!!**
- Consensus still going strong for per-residue assessment - Naive baseline among top performers

Off to new frontiers: Assemblies

From CASP14 to CASP15

But first: A self-assessment story

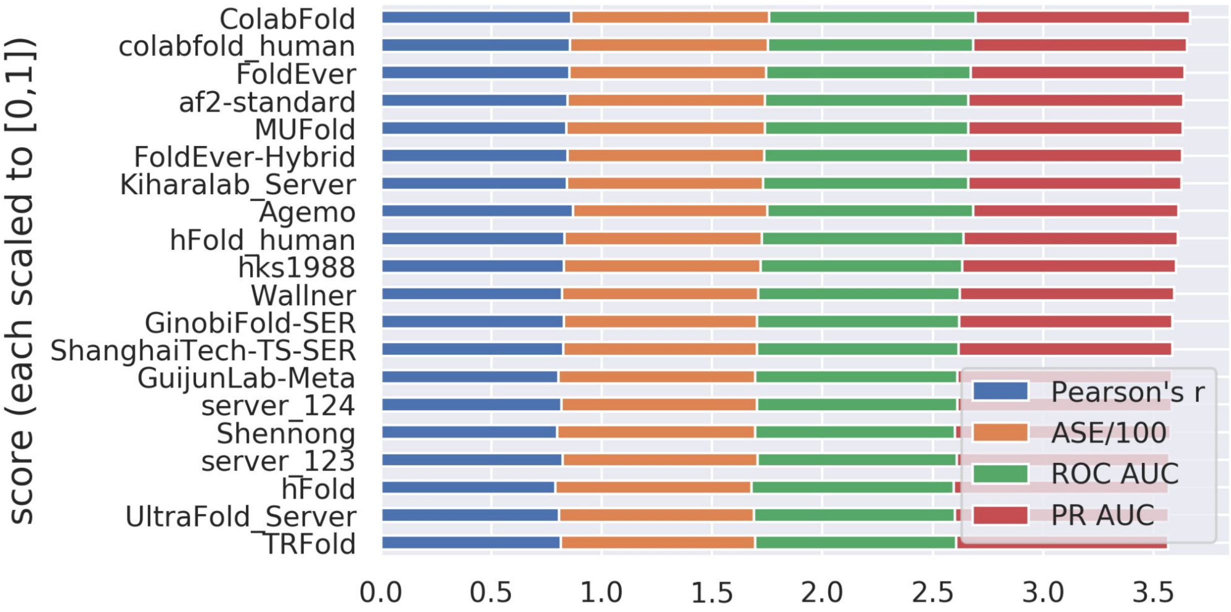
- Consensus is going stronger than ever in per-residue assessment
- Pearson R: 0.83

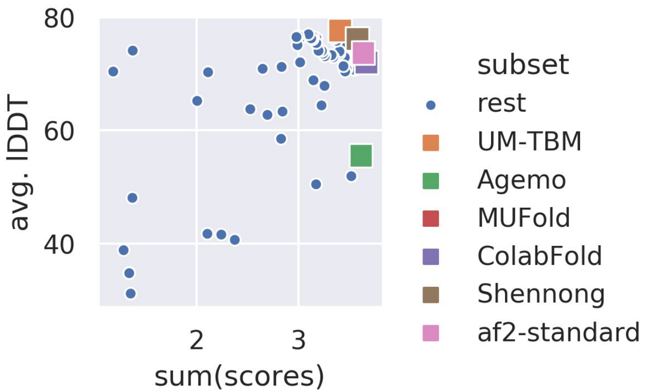


Self-Assessment - Single Chains

Top groups per score:
UM-TBM (avg. IDDT), Agemo (Pearson's r), MUFold (ASE), ColabFold (ROC AUC), Shennong (PR AUC)

Top 20 groups:



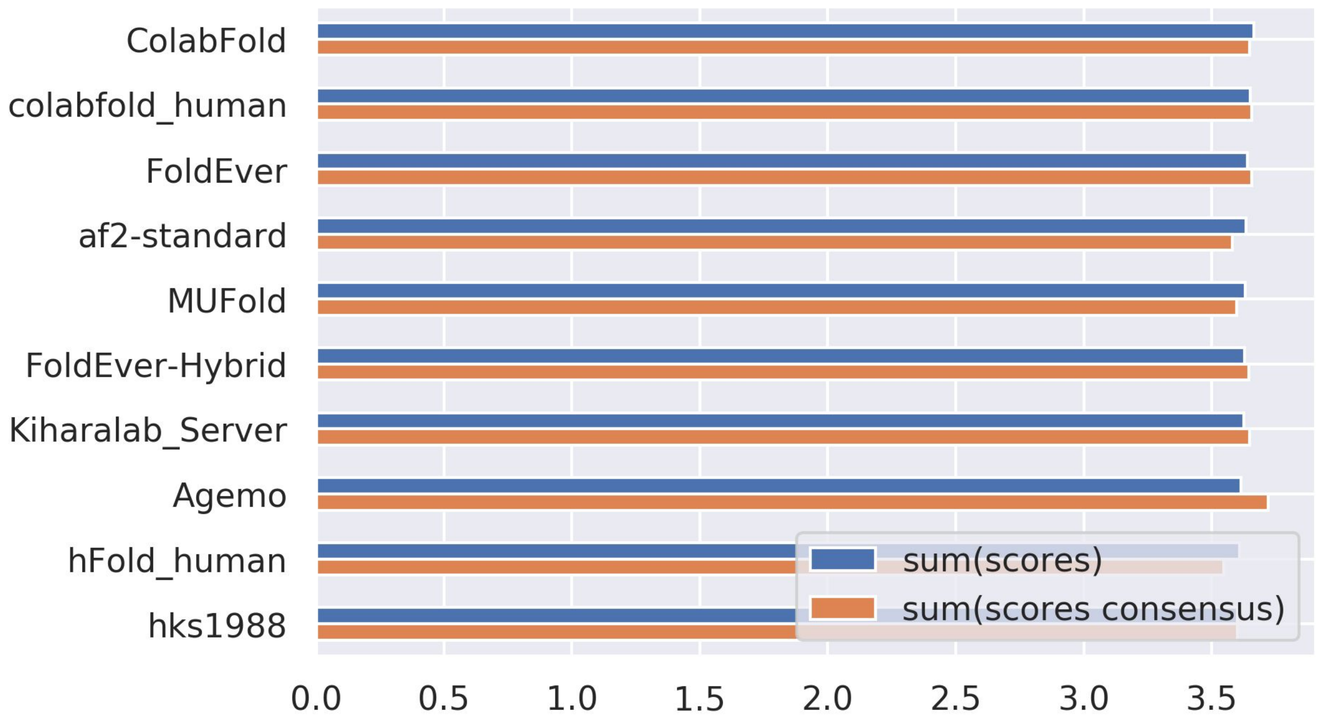


- ➢ **many groups with top scores**
- ➢ **top-IDDT groups not top at self-assess.**

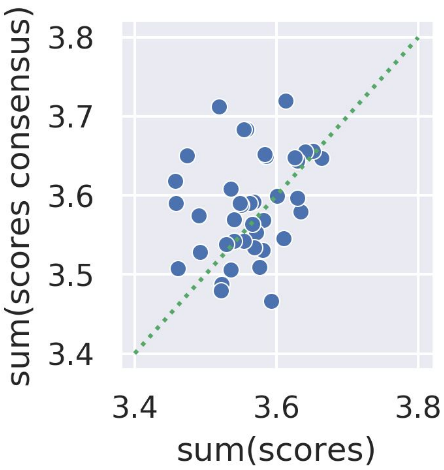
Self-Assessment - Single Chains

How good would a consensus method do on each group's models?

Top 10 groups



Top 40 groups



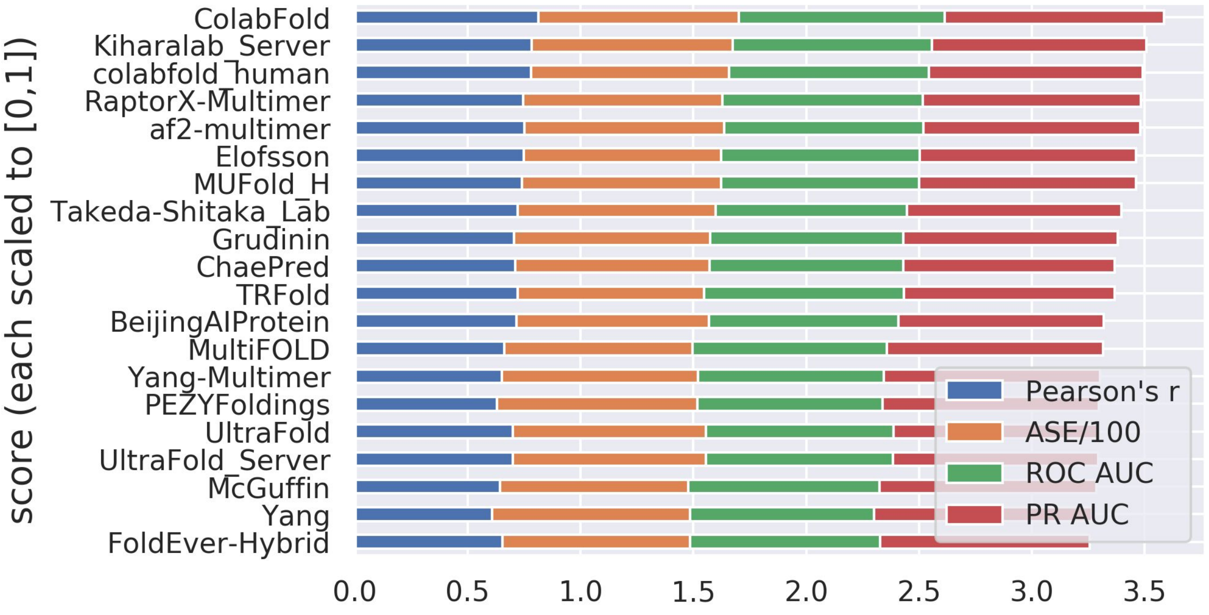
➢ **Top groups would not benefit from consensus**

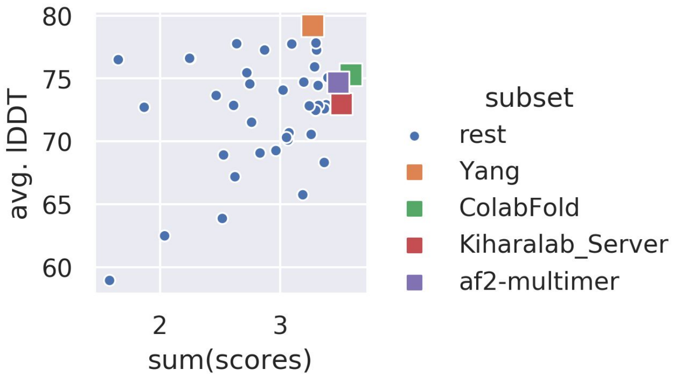
Self-Assessment - Assemblies

Top groups per score:

Yang (avg. lDDT), ColabFold (Pearson's r, ROC AUC, PR AUC), Kiharalab_Server (ASE)

Top 20 groups



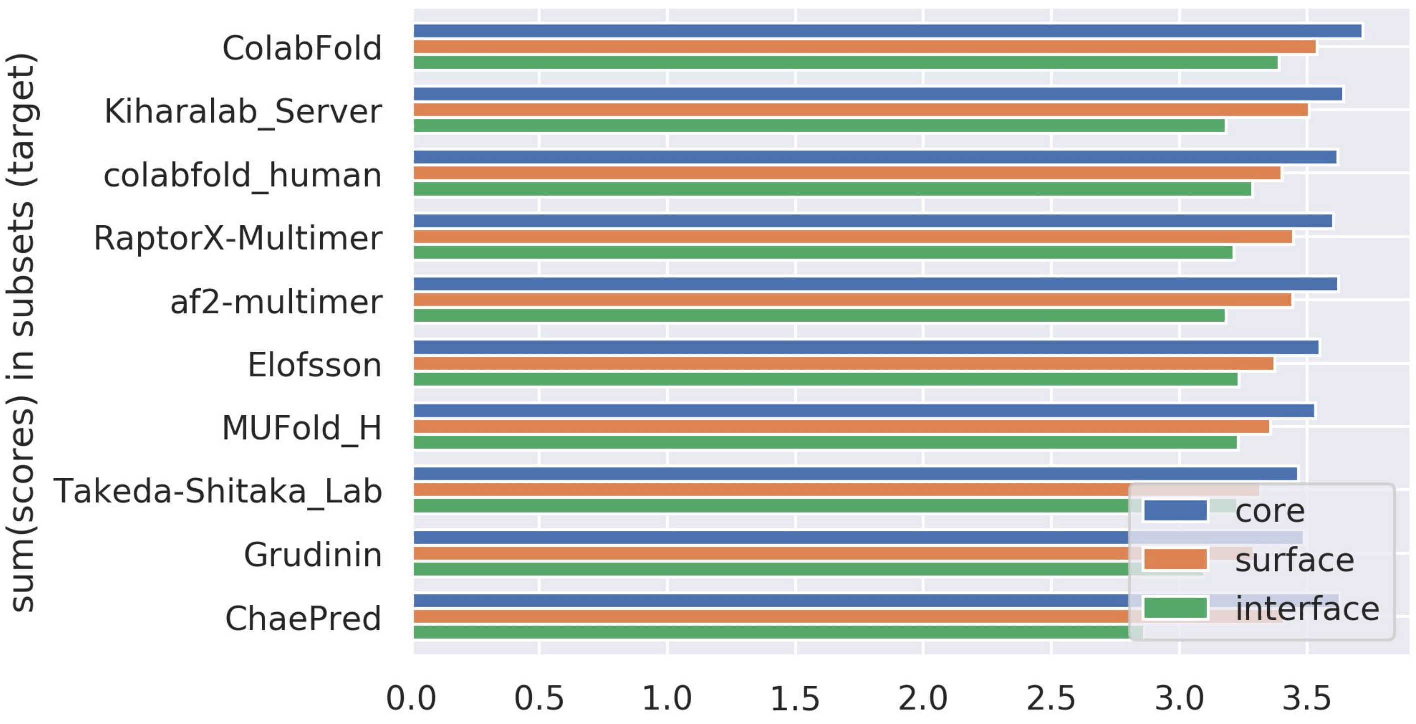


- ➢ **many groups with top scores**
- ➢ **top-lDDT groups not top at self-assess.**

Self-Assessment - Assemblies

Different accuracy in core vs surface vs interface (as seen in target)**?**

Top 10 groups

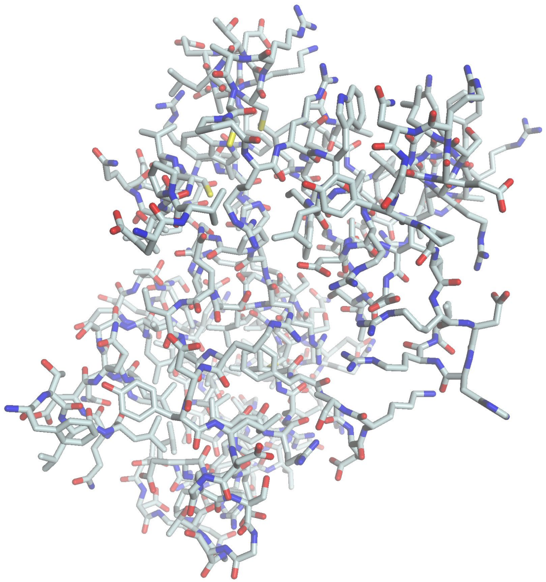


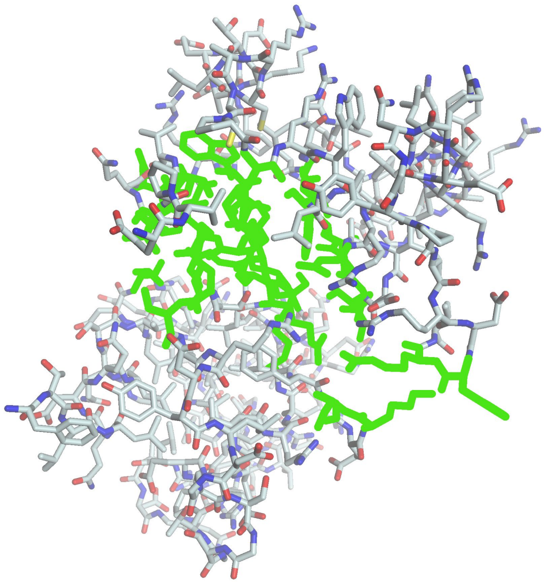
- ➢ **Easiest to predict core**
- ➢ **Hardest to predict interface**

Global - Prediction Task

SCORE: Reflects similarity of the full model/complex to the target upon global superposition (GDT/TM style)

QSCORE: Interface accuracy. Evaluated against contact-based scores (e.g. DockQ, QS-score)





Global - Target Values (QSCORE)

Goal: Fully automated interface evaluation with explicit one-to-one mapping

QS-score (Bertoni et al., 2017):

$$QS{-}score = \frac{\sum\limits_{shared(A,B)} w(min(d_A, d_B))(1 - \frac{|d_A - d_B|}{12})}{\sum\limits_{shared(A,B)} w(min(d_A, d_B)) + \sum\limits_{non-shared(A)} w(d_A) + \sum\limits_{non-shared(B)} w(d_B)}$$

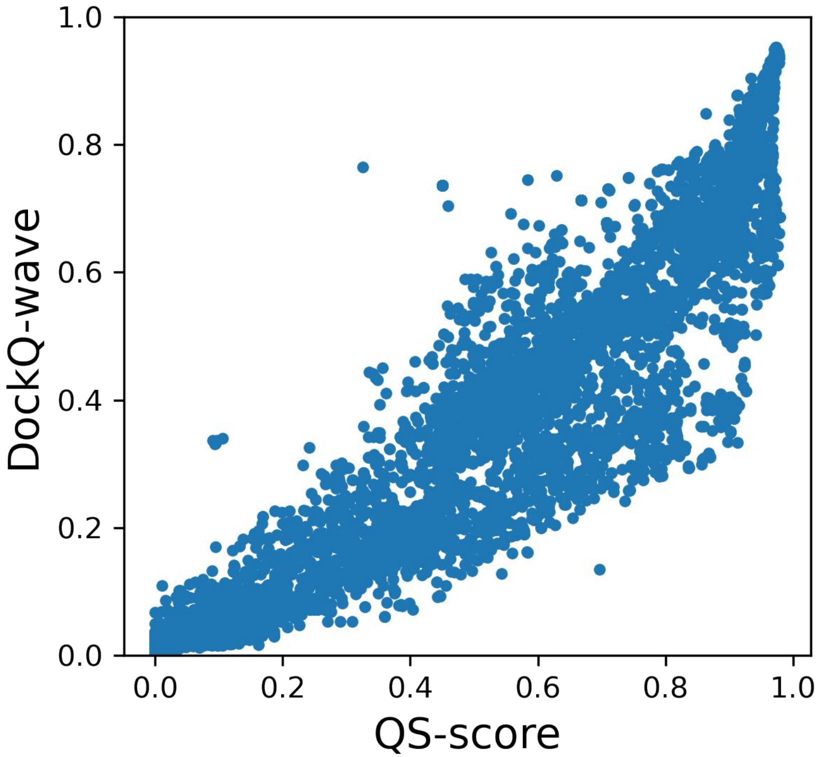
DockQ (Basu et al., 2016):

$$DockQ(F_{nat}, LRMS, iRMS, d_1, d_2) = (F_{nat} + RMS_{scaled}(LRMS, d_1) + RMS_{scaled}(iRMS, d_2))/3$$

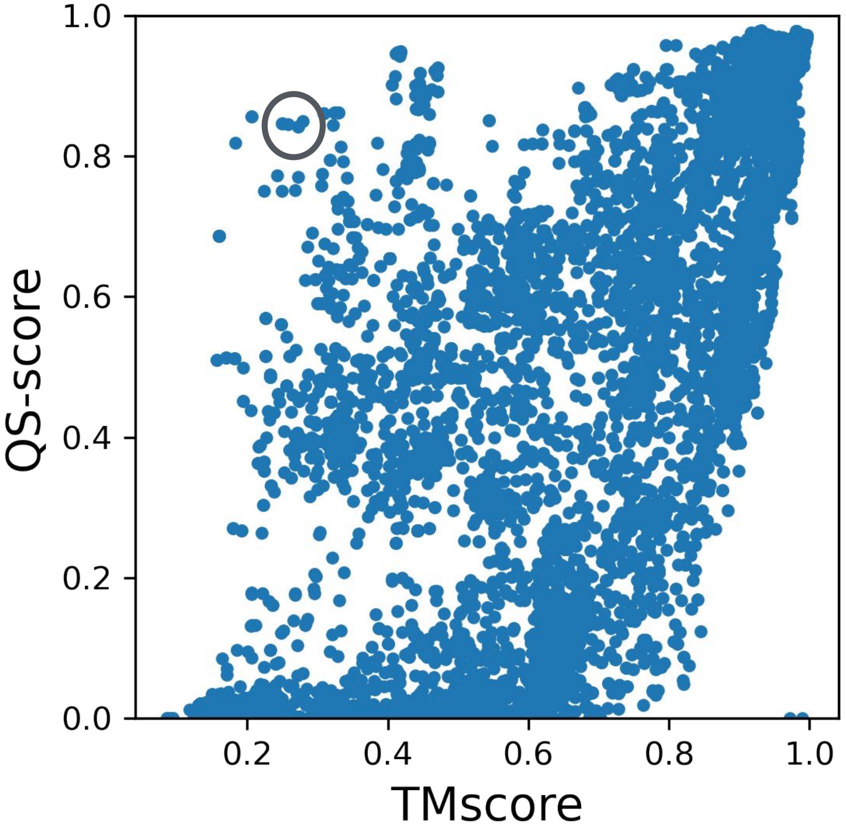
Global - Target Values (QSCORE)

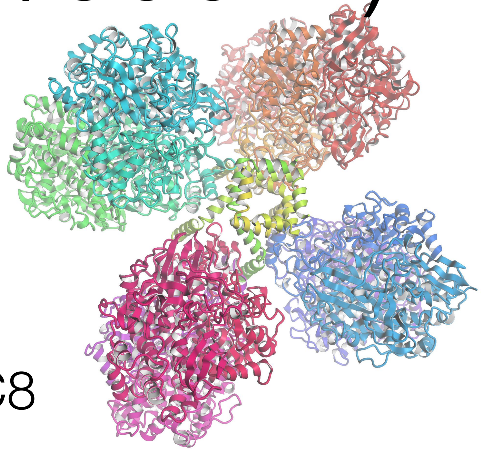
DockQ evaluates single interfaces!
=> weighted average for higher order complexes

DockQ-wave

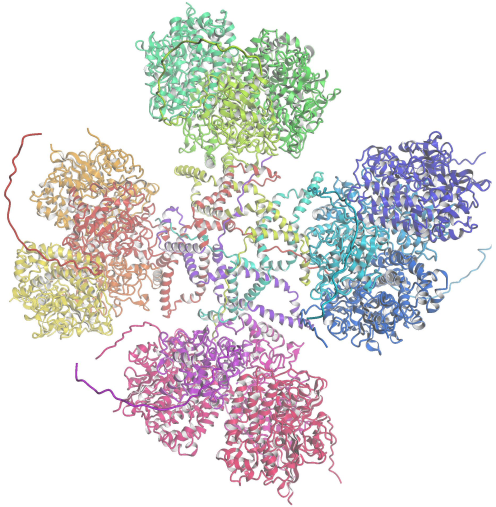


Global - Target Values (QSCORE/SCORE)



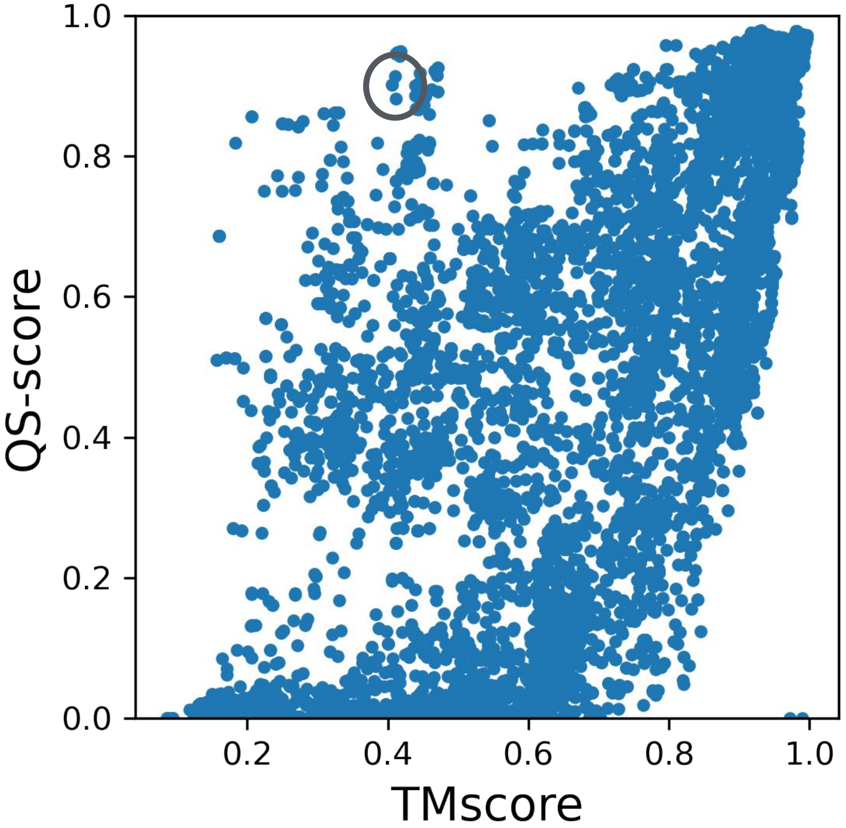


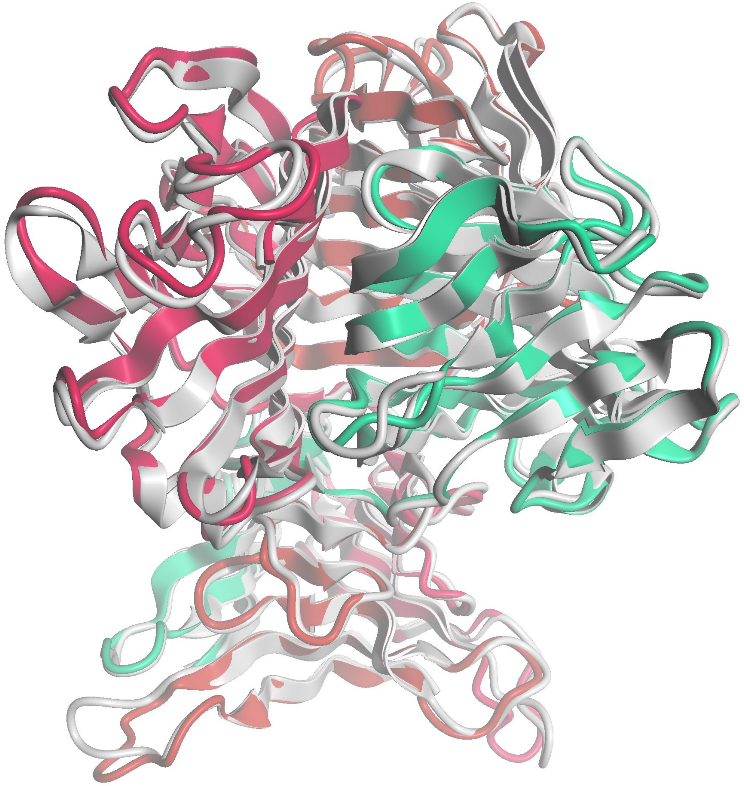
H1114
Stoich.: A4B8C8



H1114TS360_1

Global - Target Values (QSCORE/SCORE)





T1173o (colored)
T1173TS439_4 (white)

Global - Target Values (SCORE)

Goal: Fully automated evaluation with explicit one-to-one mapping

oligo-GDTTS: Not what you think…

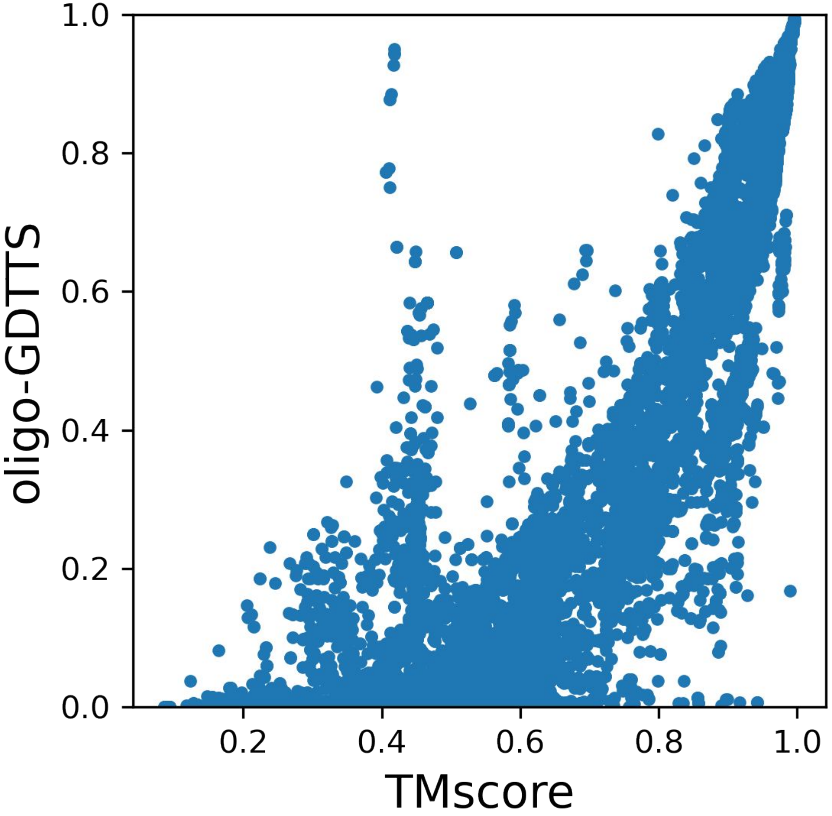
- Perform one-to-one chain mapping
- Use simple Kabsch algorithm to superpose aligned residues (Cα)
- Compute GDTTS with common thresholds (8.0, 4.0, 2.0, 1.0)

TMscore: As computed with USalign by Andriy (Zhang et al., 2022)

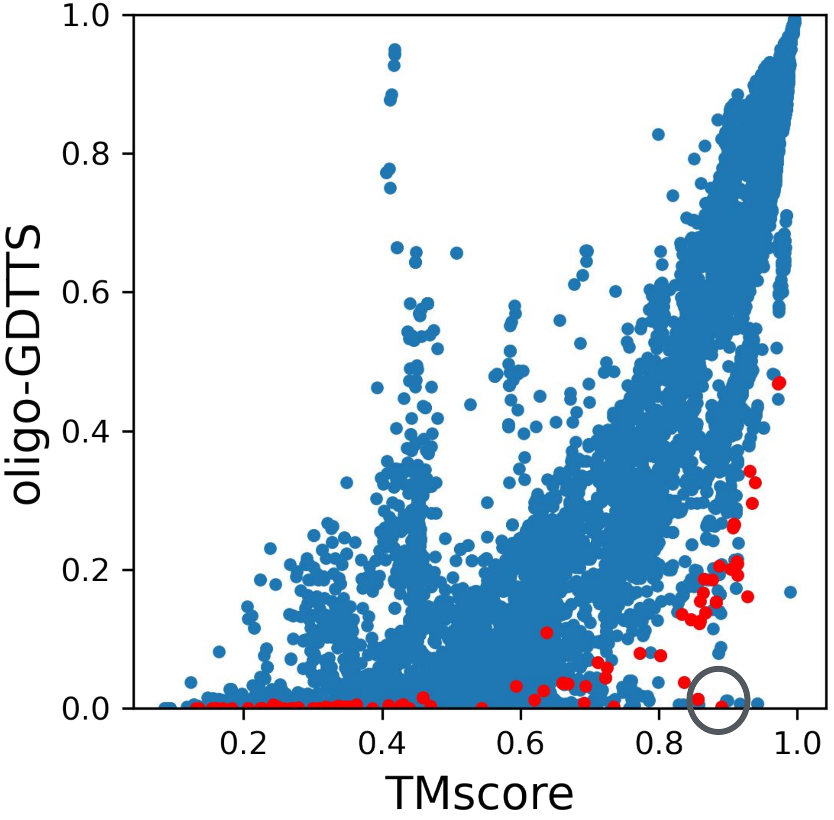
$$TMscore = max\left(\frac{1}{L_{target}}\sum_{i}^{L_{common}}\frac{1}{1+\left(\frac{d_i}{d_0(L_{target})}\right)^2}\right)$$

Remember this!

Global - Target Values (SCORE)



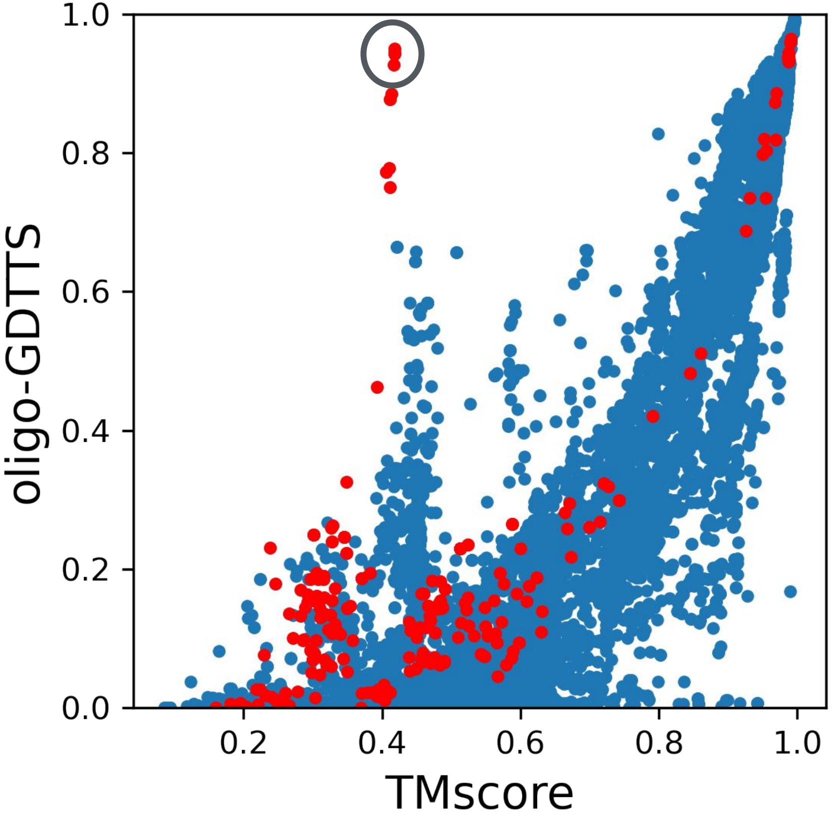
Global - Target Values (SCORE)

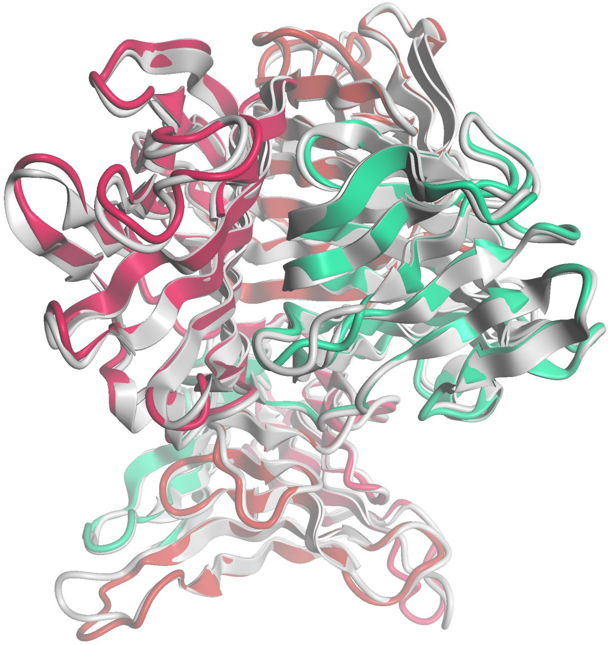


Large complexes (eg. H1114 in red) tend to be “outliers”

H1114TS348_2 - TMscore: 0.89 (d_0>20Å)

Global - Target Values (SCORE)





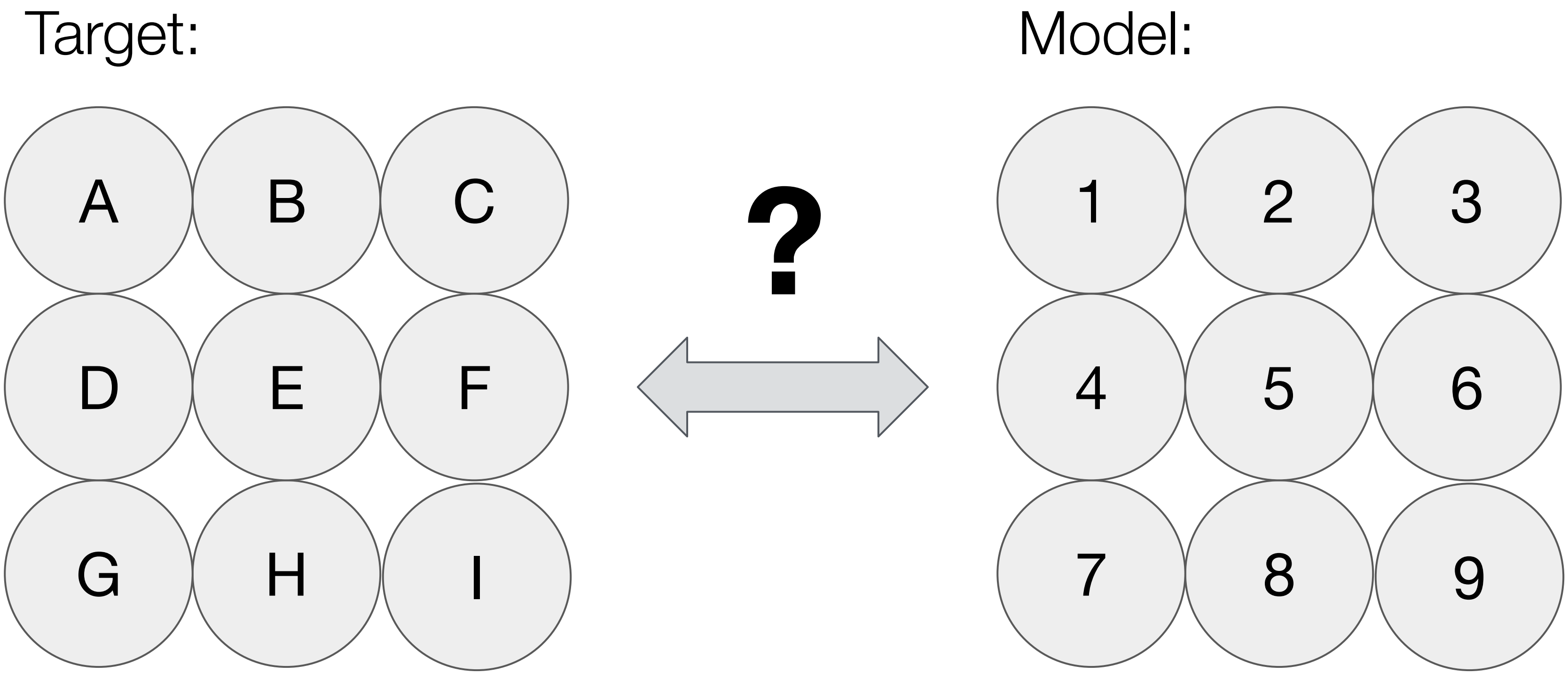
T1173o (colored)
T1173TS439_4 (white)

Global - A Chain Mapping Story

The Problem:

- Original QS-score chain mapping algorithm fails in many cases of CASP15 due to complexity
- Alternatives like USalign (Zhang et al., 2022) and QSalign (Dey et al., 2018) are based on rigid superpositions. QSCORE evaluation would be impossible.

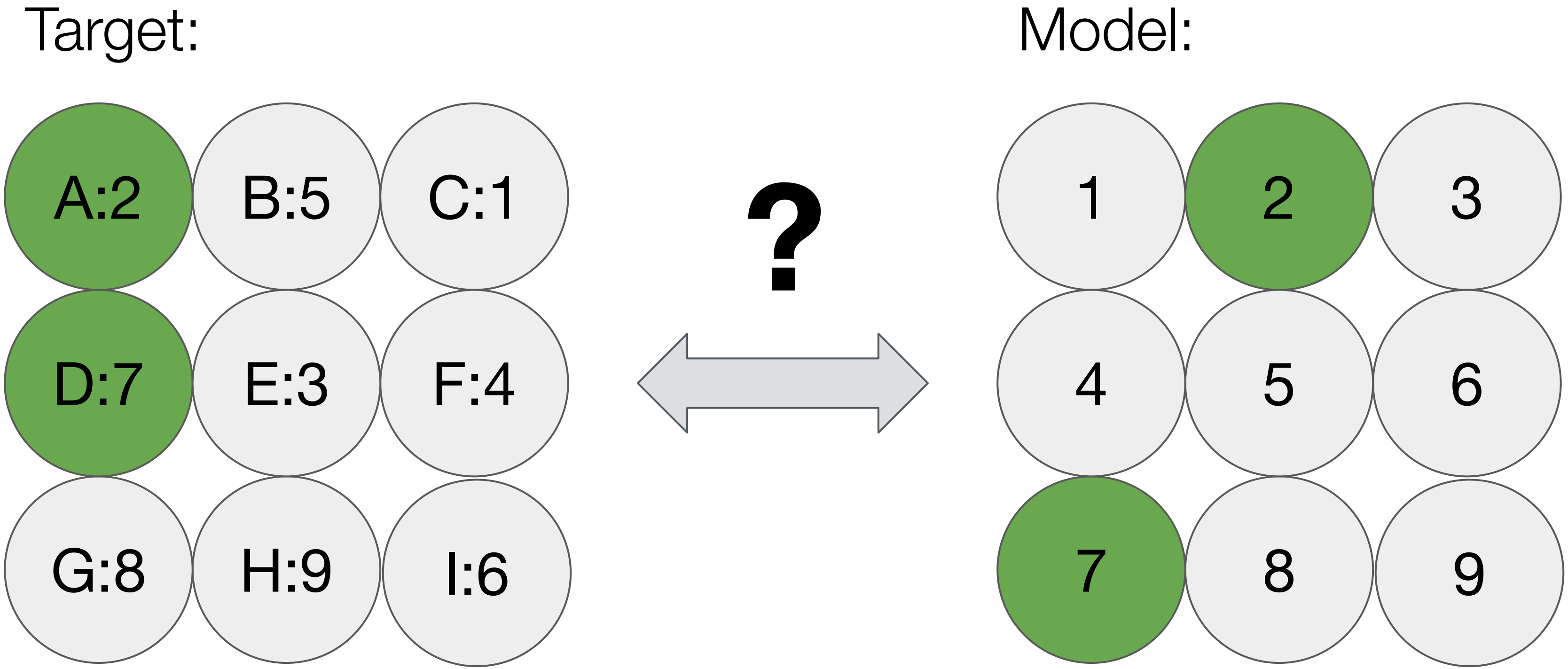
Global - A Chain Mapping Story



Goal: A:1, B:2, C:3, D:4, E:5, F:6, G:7, H:8, I:9

Possible mappings: 9!=362880

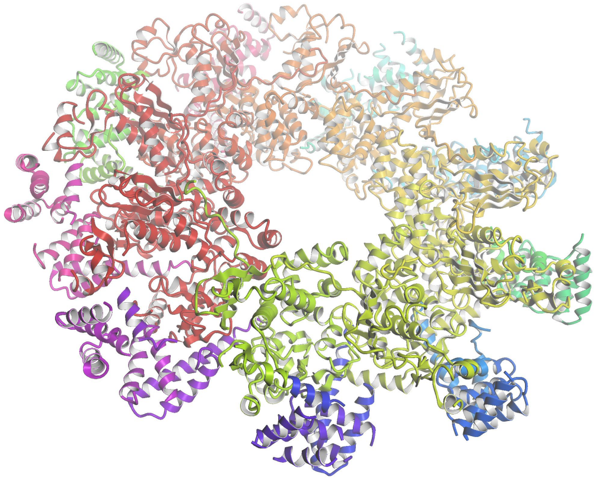
Global - A Chain Mapping Story



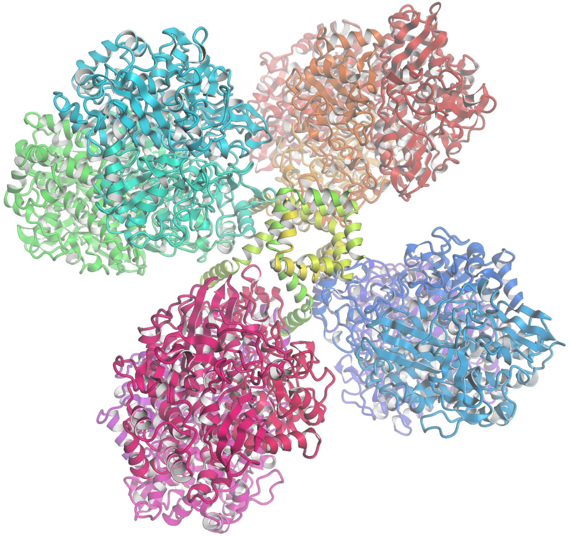
Contact based scores (lDDT, QS-score) are pairwise decomposable

Global - A Chain Mapping Story

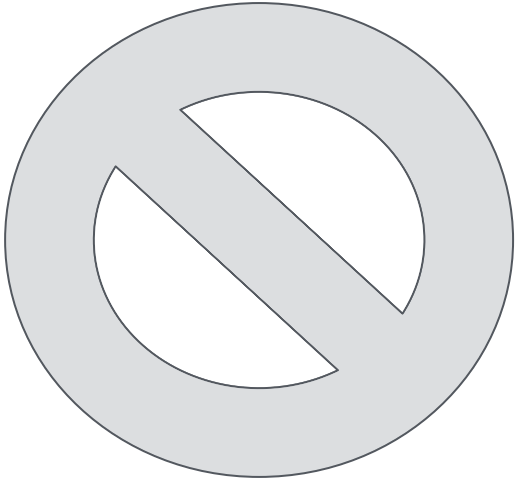
- All CASP targets with <= 12 chains are naively enumerated
- This leaves:



H1111
A9B9C9



H1114
A4B8C8



T1115o
A16

Global - A Chain Mapping Story - Conclusion

- Efficient solution to the chain mapping problem
- QSCORE evaluation would be impossible without
- New efficient implementations of QS-score and lDDT
- New lDDT does oligos and arbitrary compounds (Ligands, RNA/DNA)
- Chain Mapping/lDDT is the basis for further work (Ligand-PLI from Ligand session)

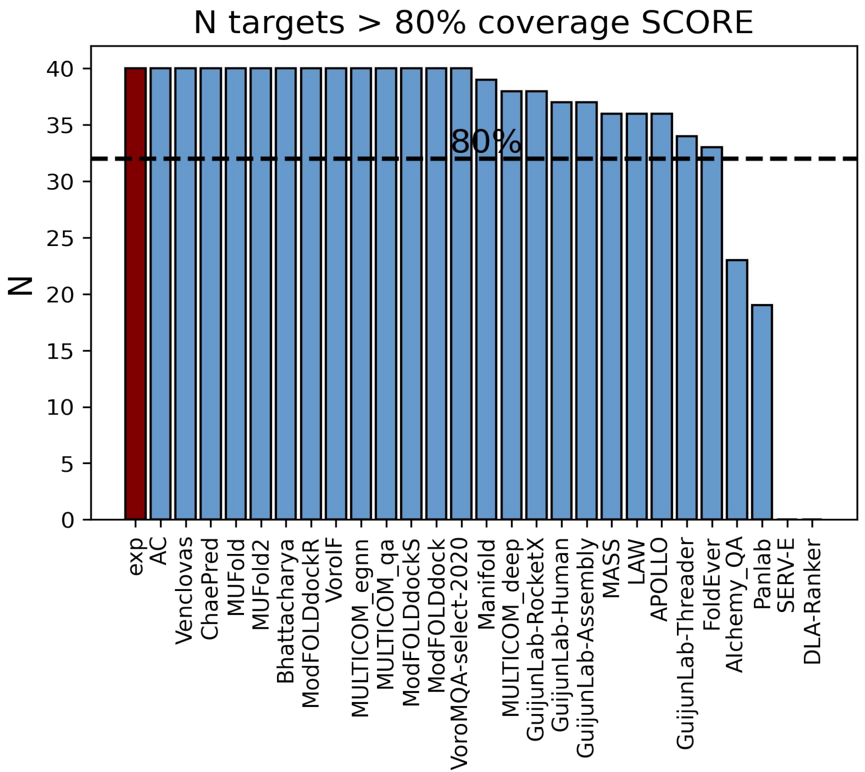
Global - Data Situation

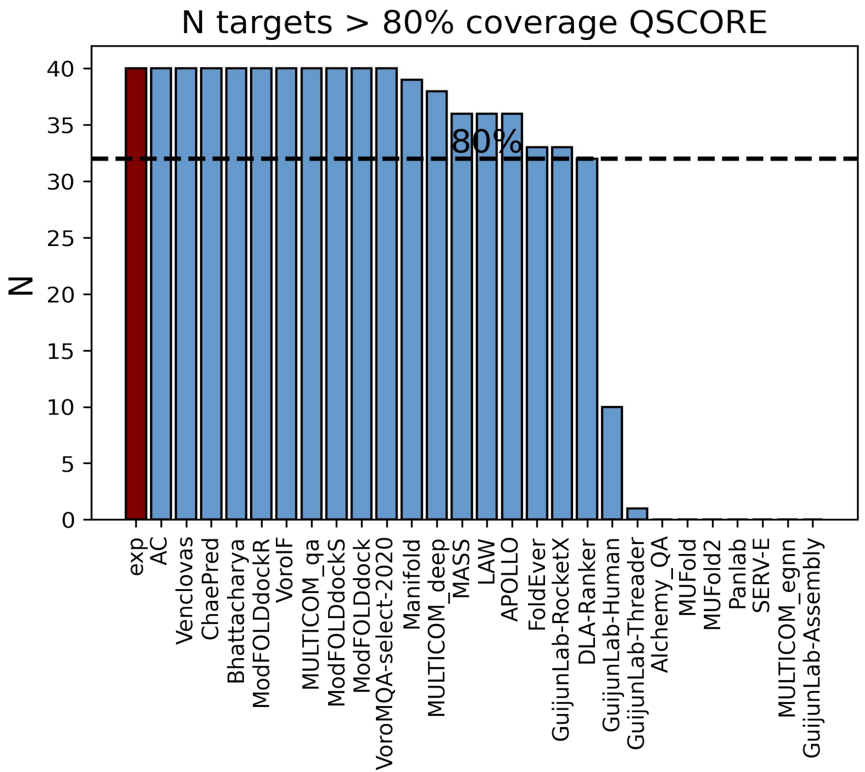
- Largely relied on Ezgi/Burcu/Andriy/Marc
- Total number of models: 11129
- Errors in analysis pipeline: 7
- Wrong Stoichiometry: 650

=> 10472 Models from 40 targets for analysis

Global - Data Situation (per target)

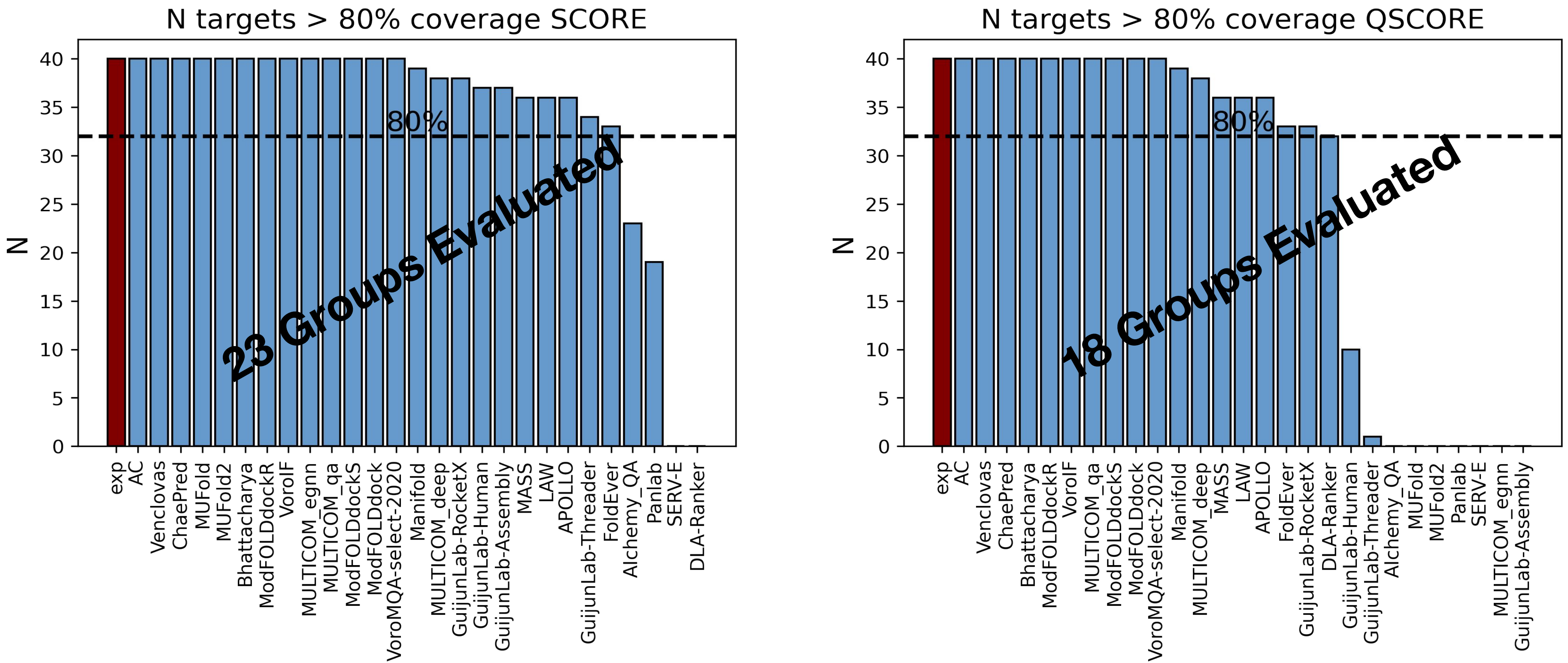
AC Assembly **C**onsensus baseline



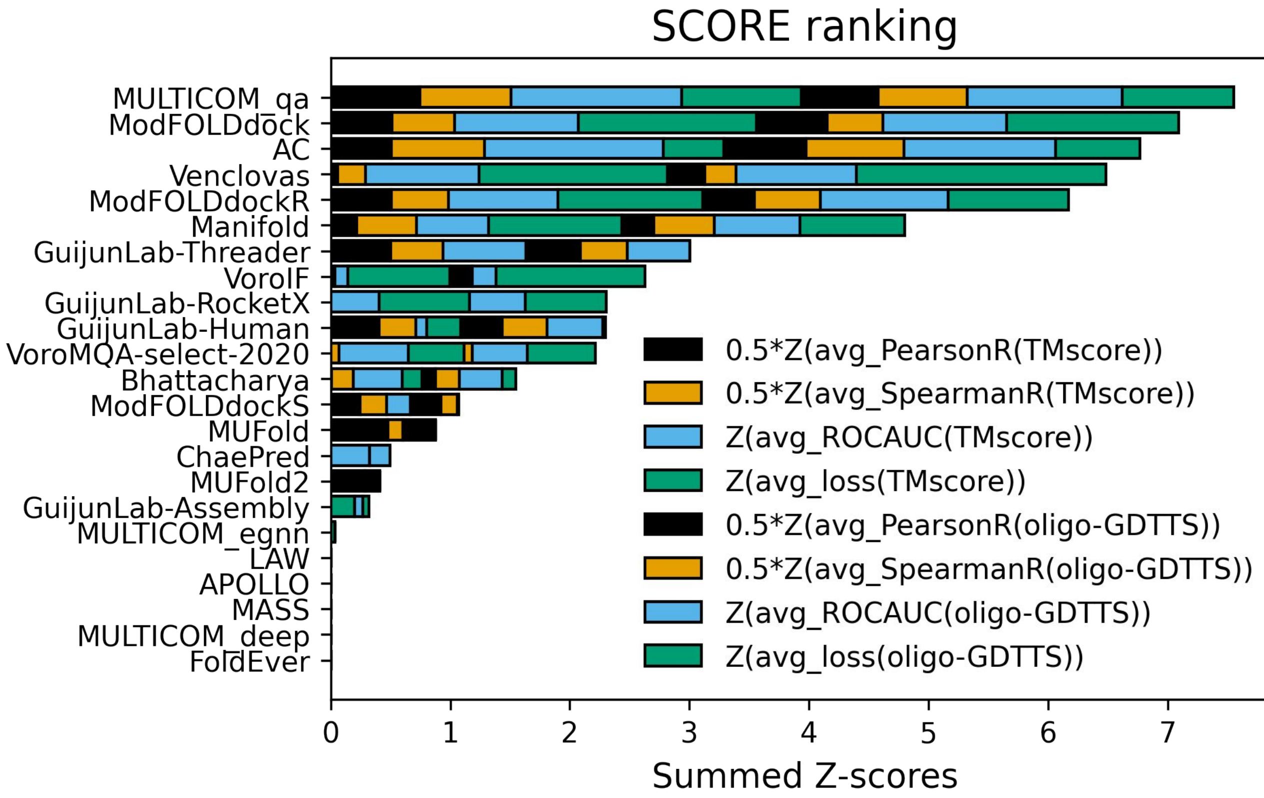


Global - Data Situation (per target)

AC Assembly **C**onsensus baseline

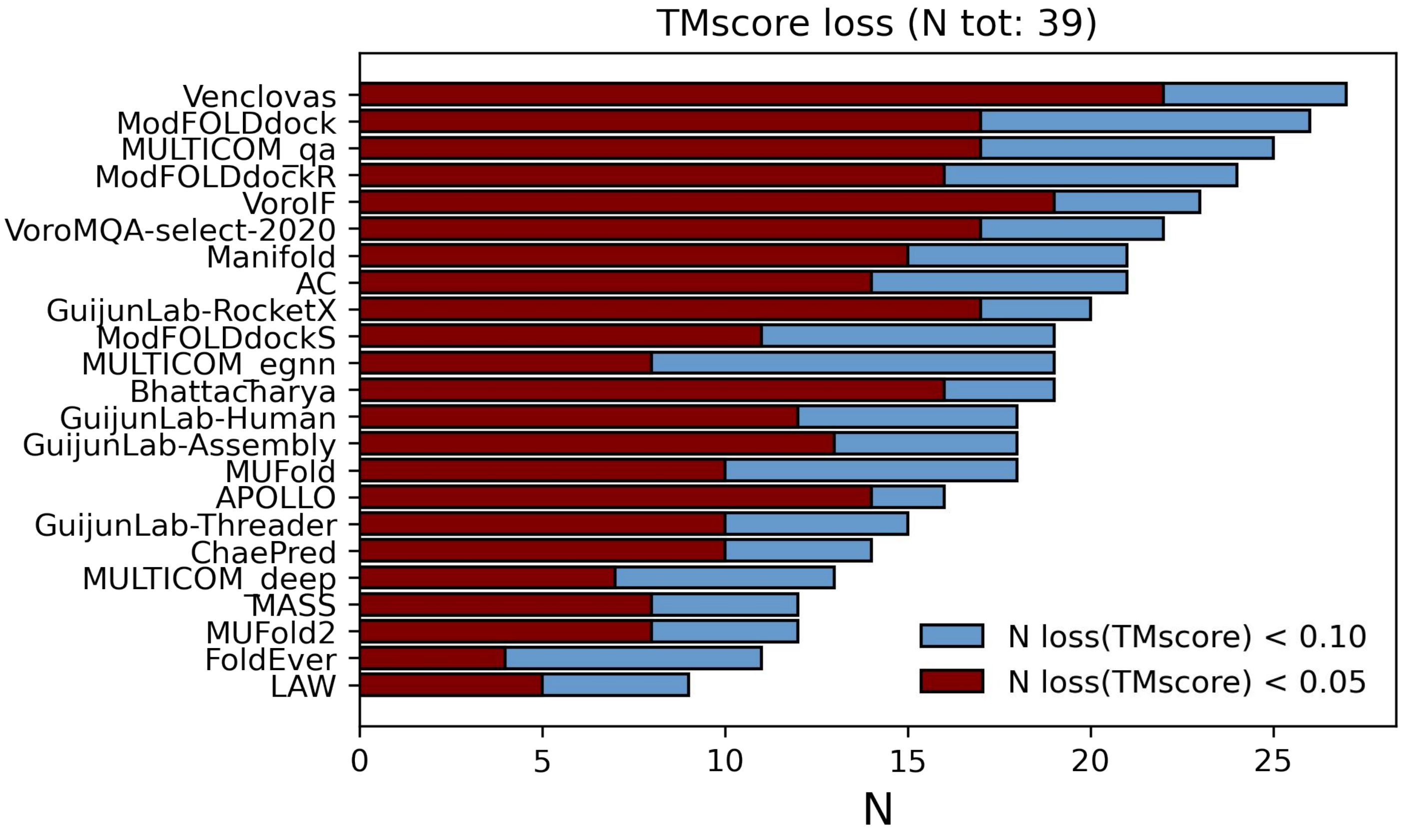


Global - Evaluation (SCORE)

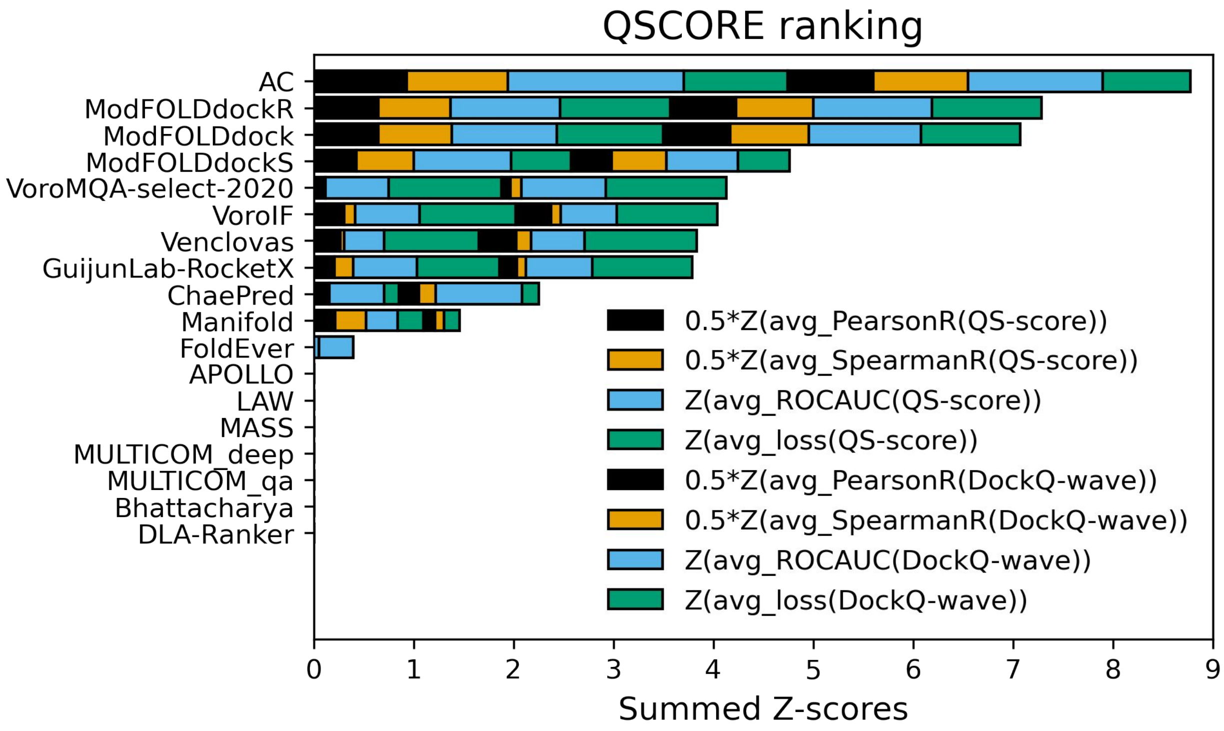


- Considered targets with at least one model with TMscore > 0.6 => N=39
- Z = max(Z, 0.0)
- Adaptive ROCAUC => class threshold at top quartile

Global - Evaluation (SCORE)

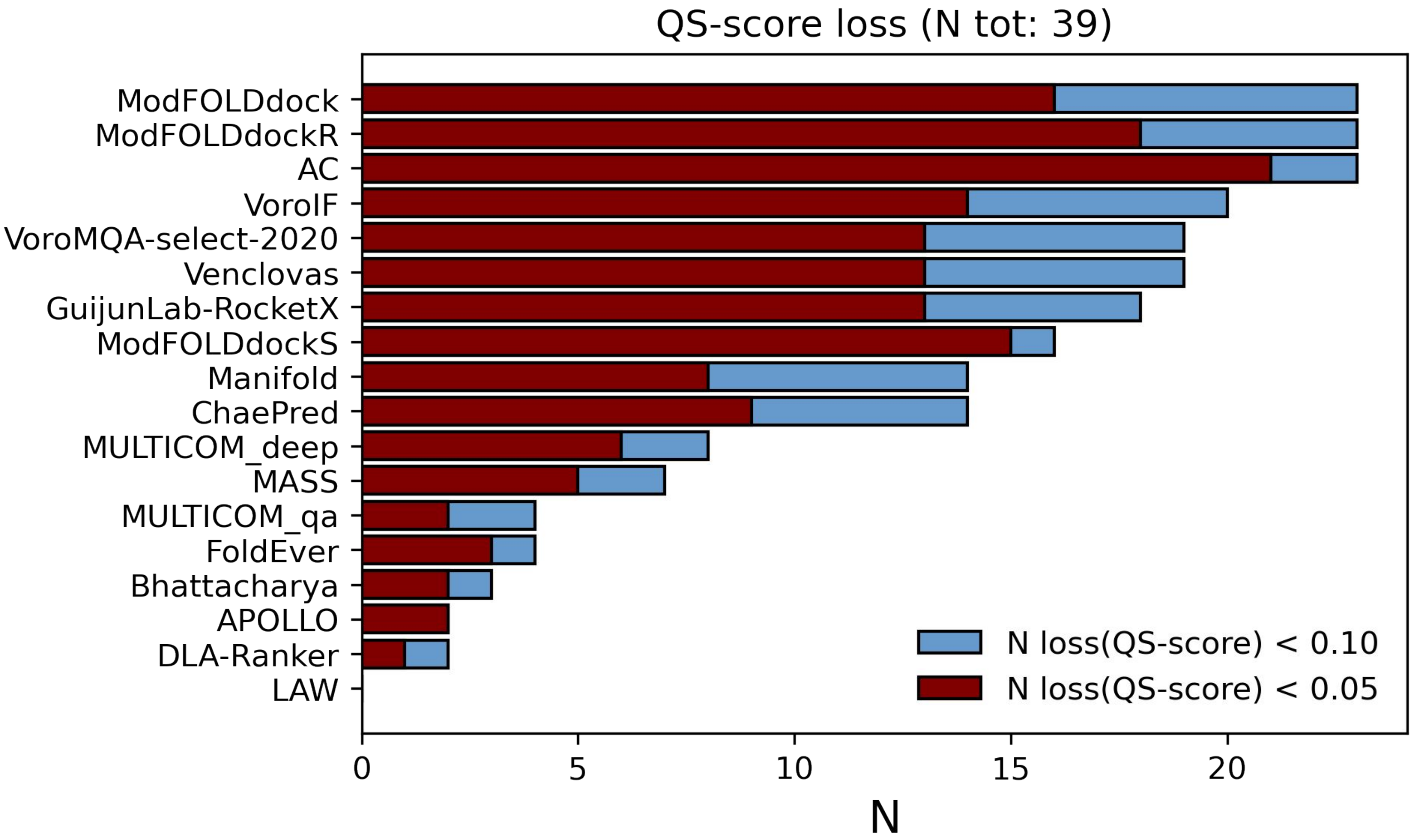


Global - Evaluation (QSCORE)



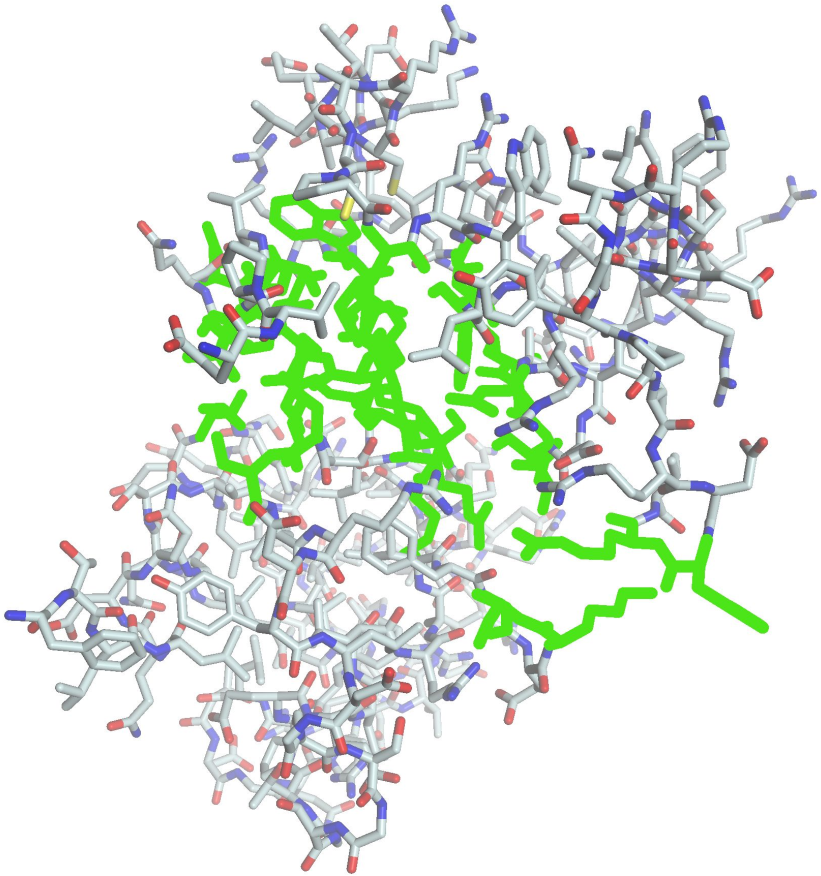
- Considered targets with at least one model with QS-score > 0.6 => N=39
- Z = max(Z, 0.0)
- Adaptive ROCAUC => class threshold at top quartile

Global - Evaluation (SCORE)



Local - Prediction Task

Predict local interface accuracy for interface residues



Local - Target Values

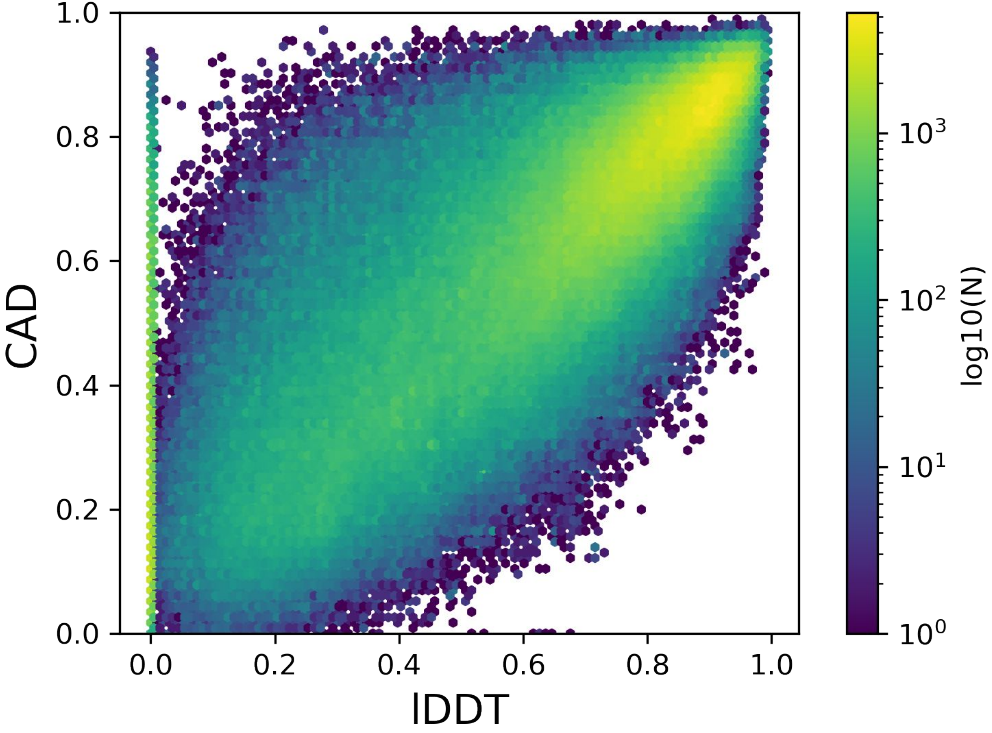
lDDT (Mariani et al., 2013)

- No words needed

CAD (Olechnovič et al., 2013)

- No words needed

Pearson R: 0.83



Local - Target Values

PatchQS/PatchDockQ

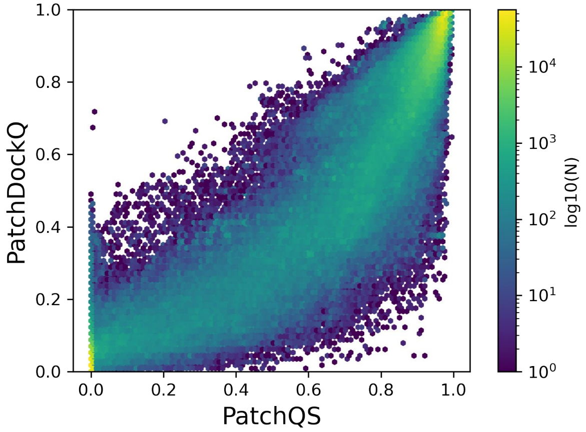
Given residue r in chain A:

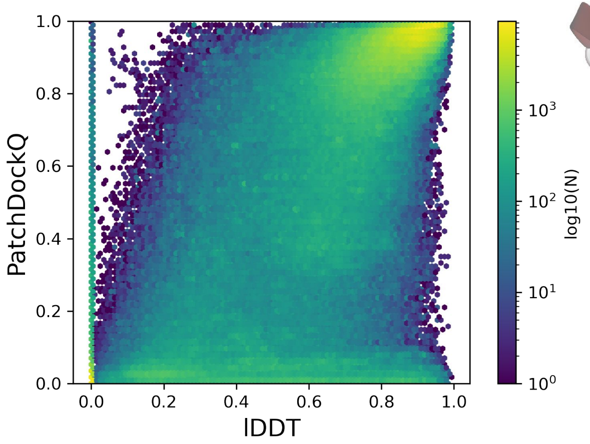
Patch one: (cname=A and 8 <> r) and (12 <> cname!=A)

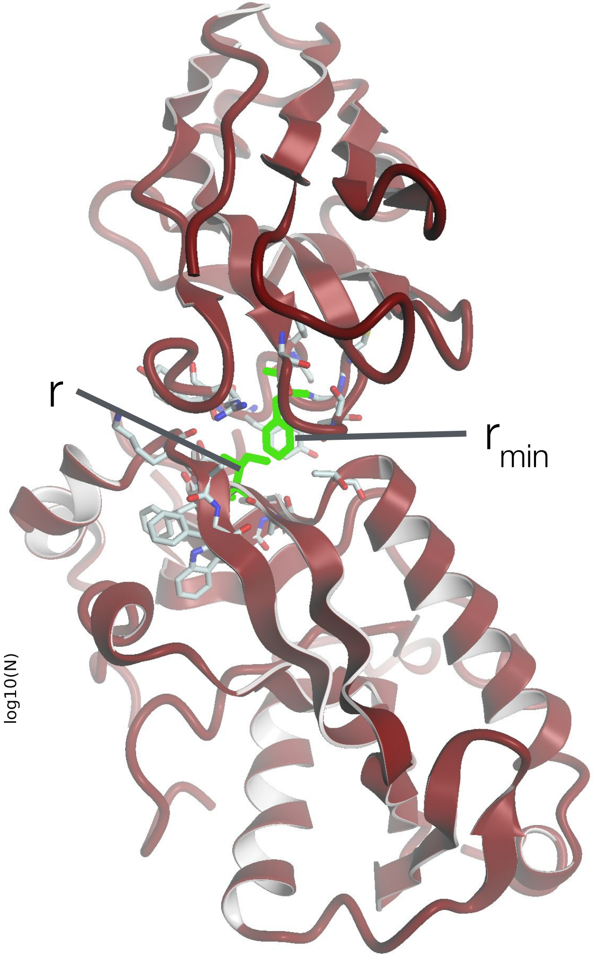
Patch two: (cname!=A and 8 <> r_{min}) and (12 <> A)

With:

- r_{min} : closest residue to r in any chain != A
- <>: within



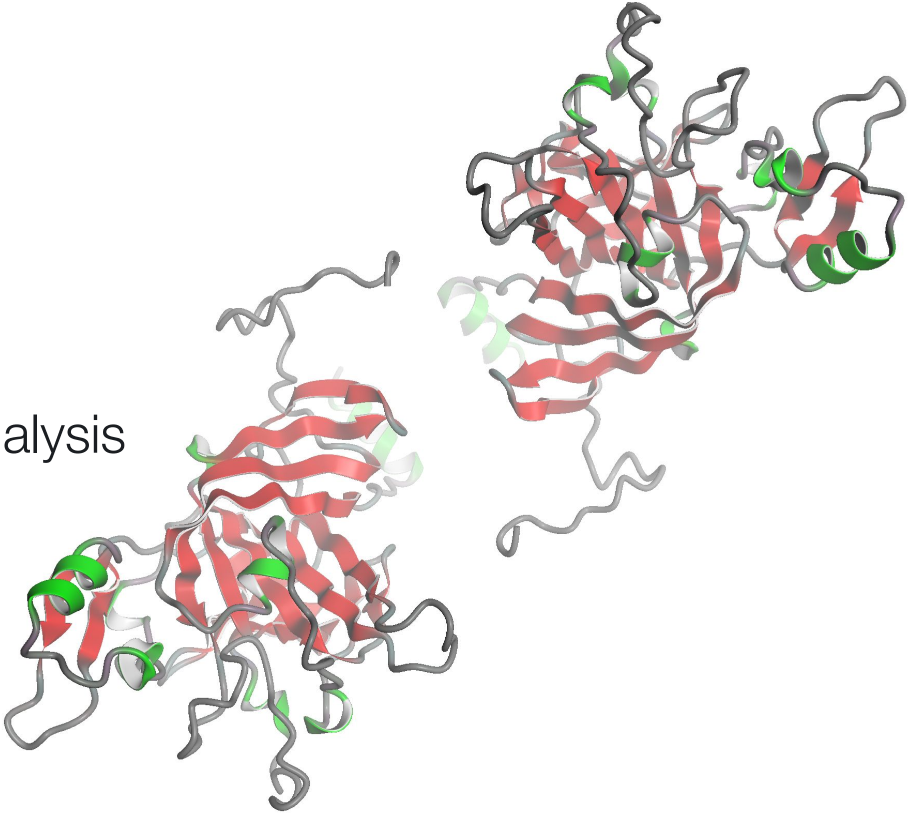




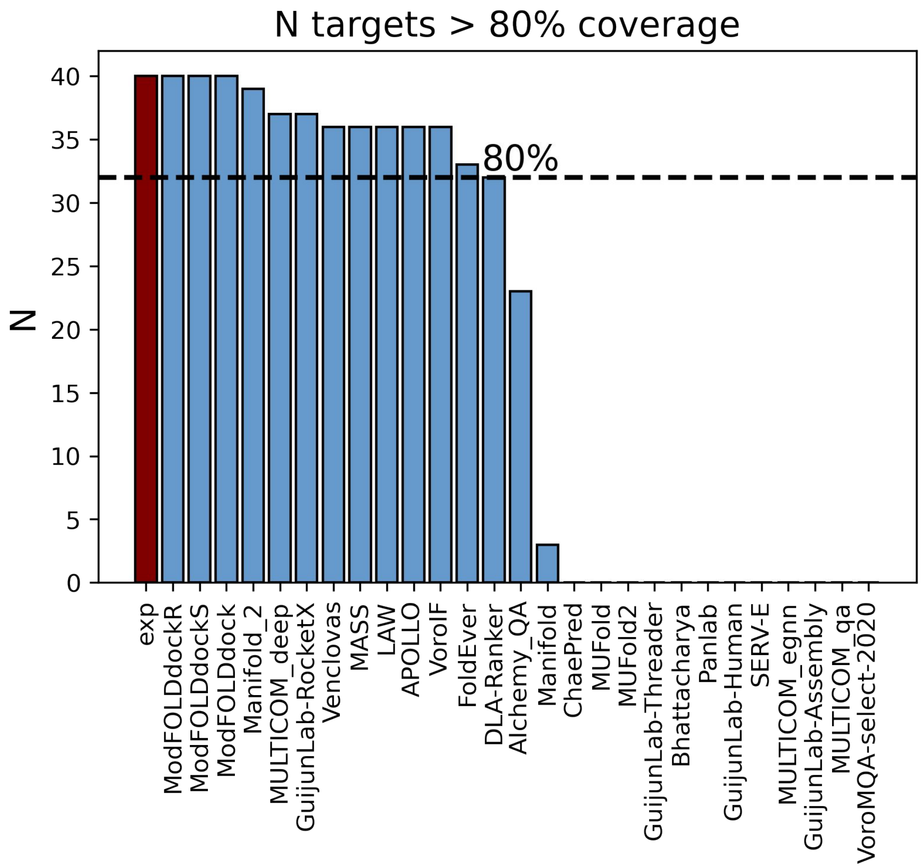
Local - Data Situation

- Total number of models: 11129
- Missing from analysis: 183
    - Errors in analysis pipeline: 7
    - No interface contacts: 176
- Wrong Stoichiometry: 617

=> 10329 Models from 40 targets for analysis

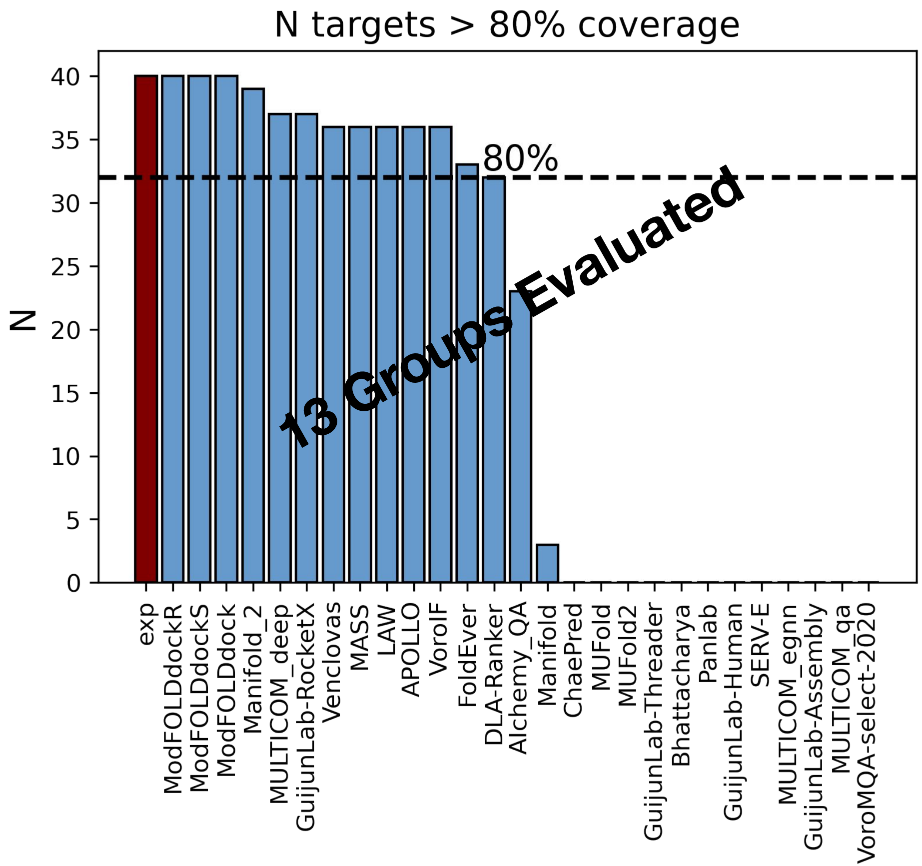


Local - Data Situation



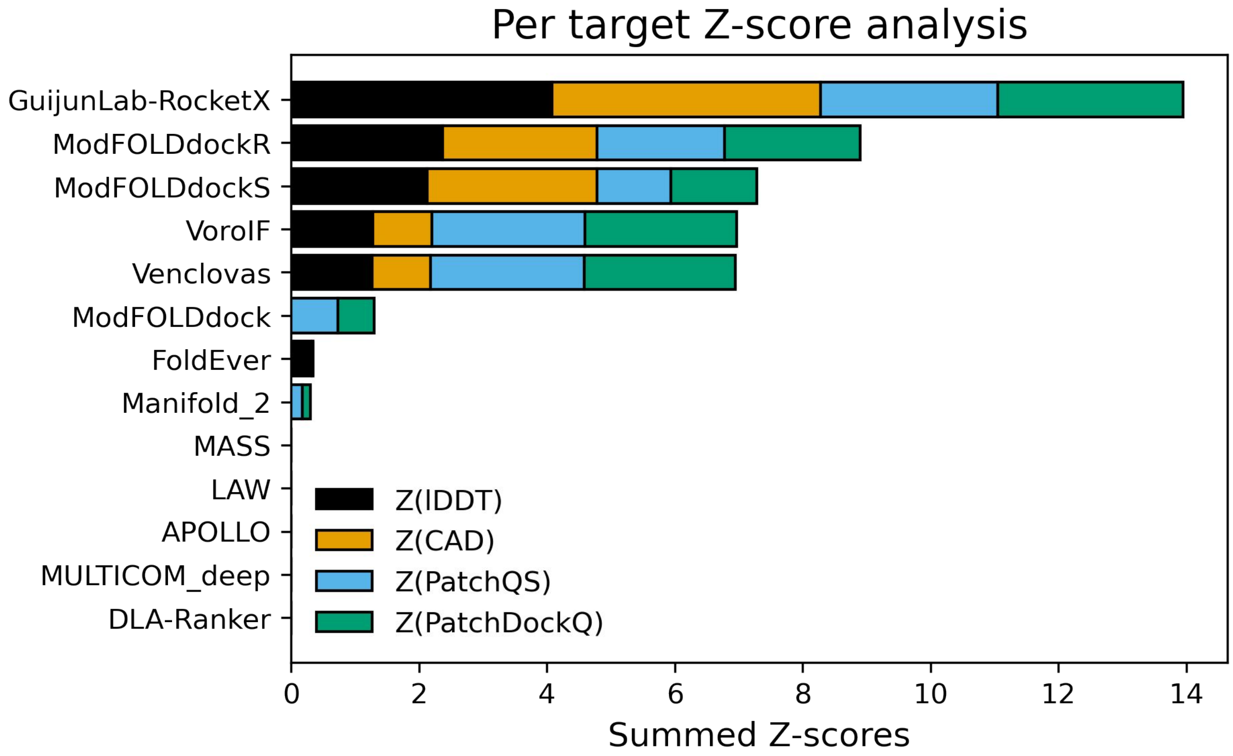
Manifold_2 represents Manifold. Resolved off by one indexing issue in predictions.

Local - Data Situation



Manifold_2 represents Manifold. Resolved off by one indexing issue in predictions.

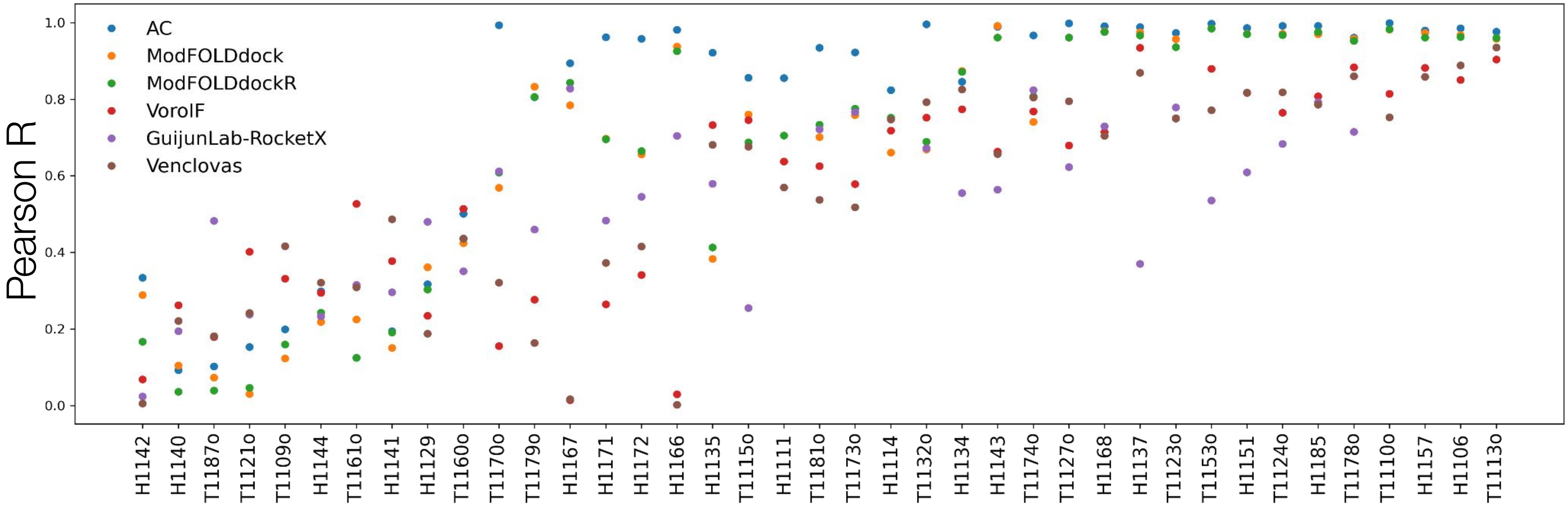
Local - Evaluation Z-score analysis



Z(x) = 0.5*Z(avg_PearsonR(X)) + 0.5*Z(avg_SpearmanR(X)) + Z(avg_ROCAUC(X))

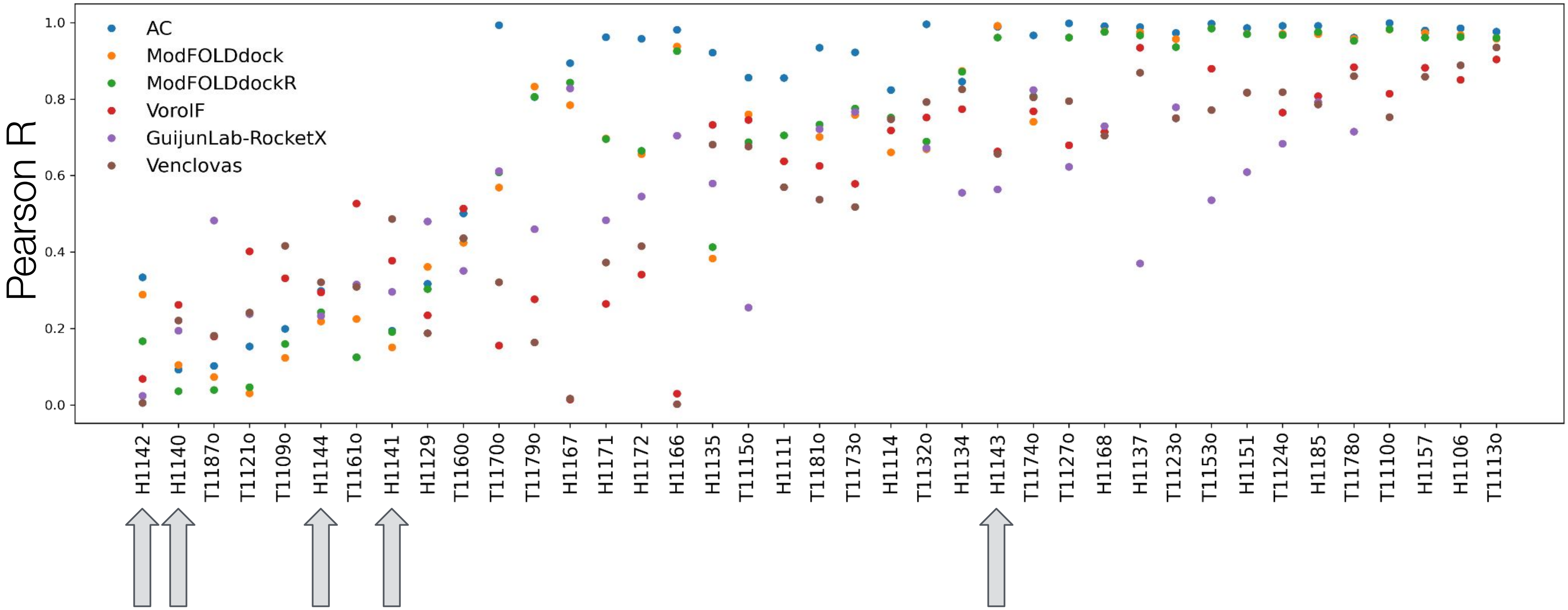
The target matters

QS-score vs QSCORE



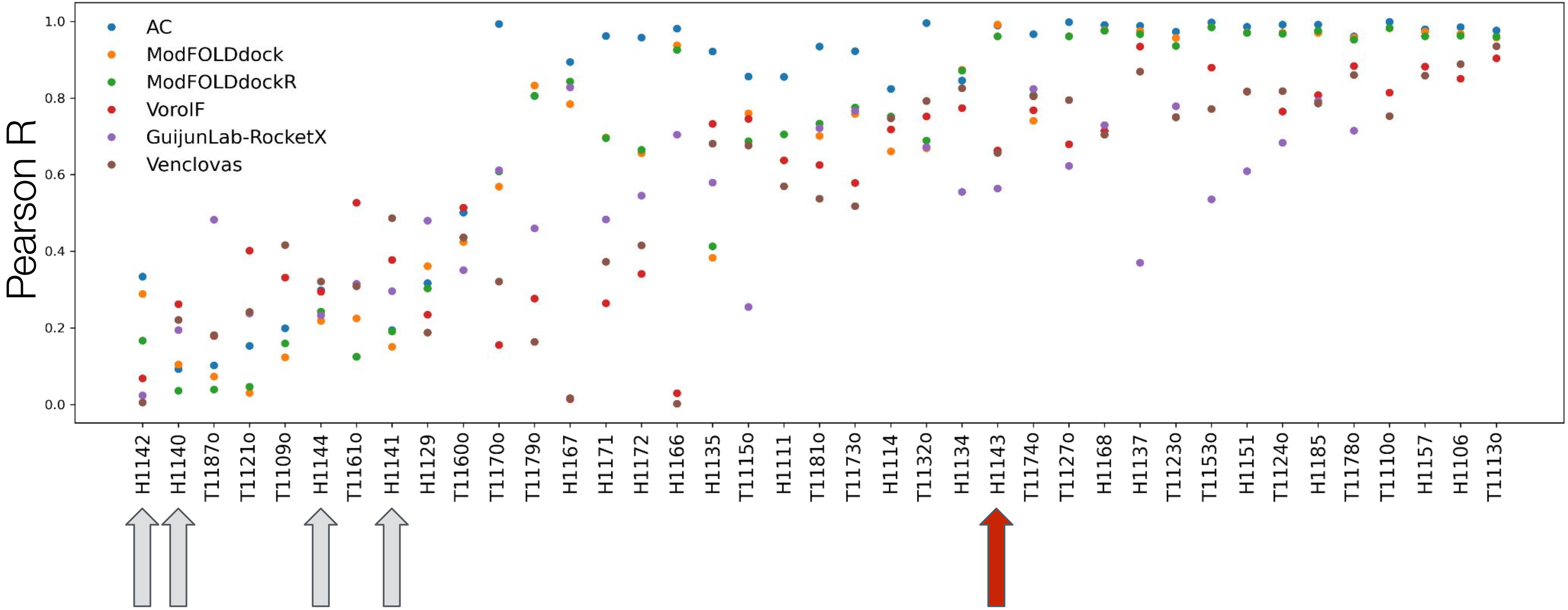
The target matters - Nanobodies

QS-score vs QSCORE

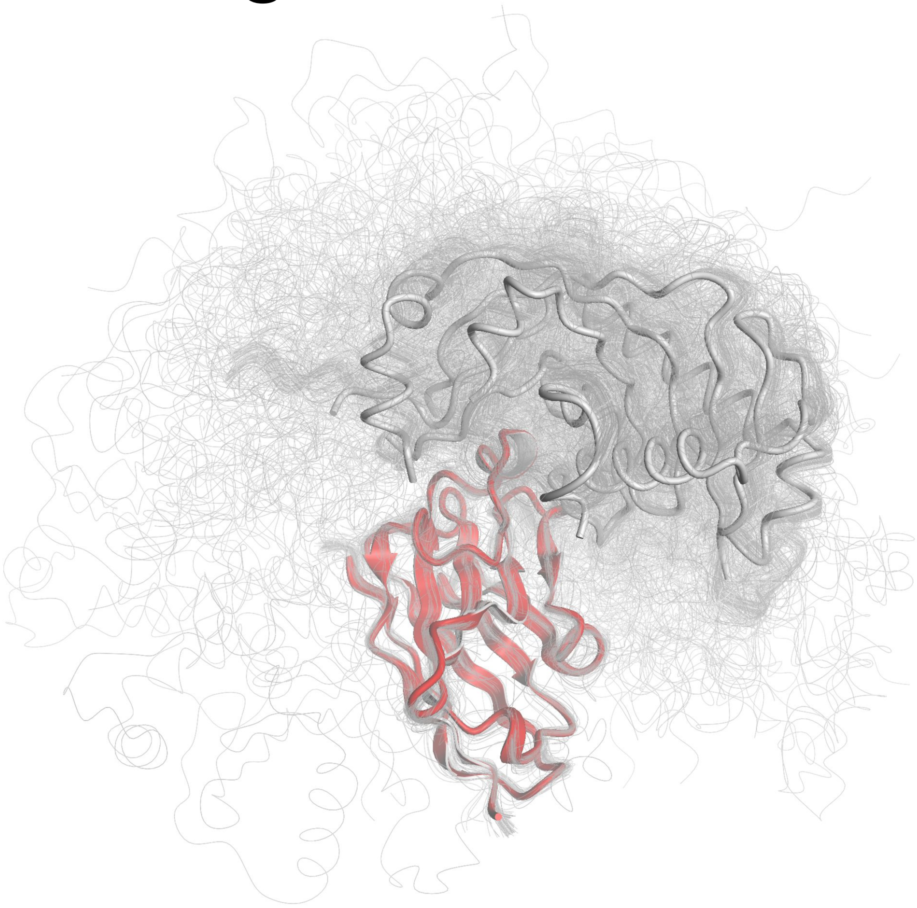


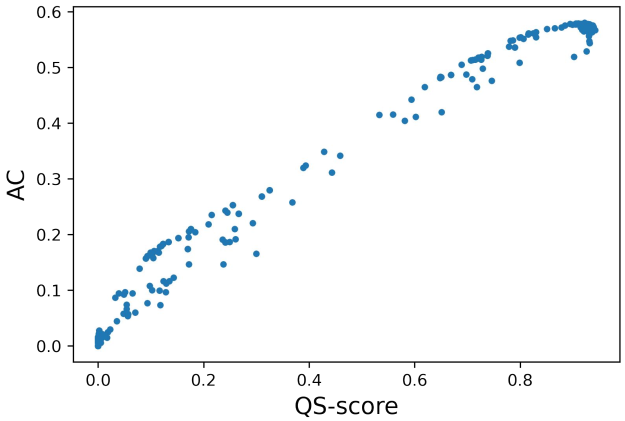
The target matters - Nanobodies

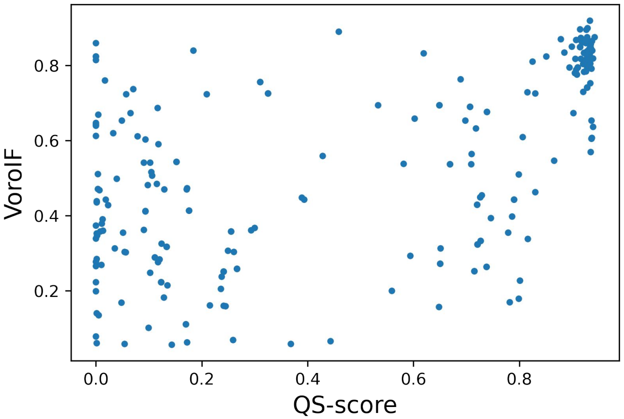
QS-score vs QSCORE



The target matters - Nanobodies - H1143

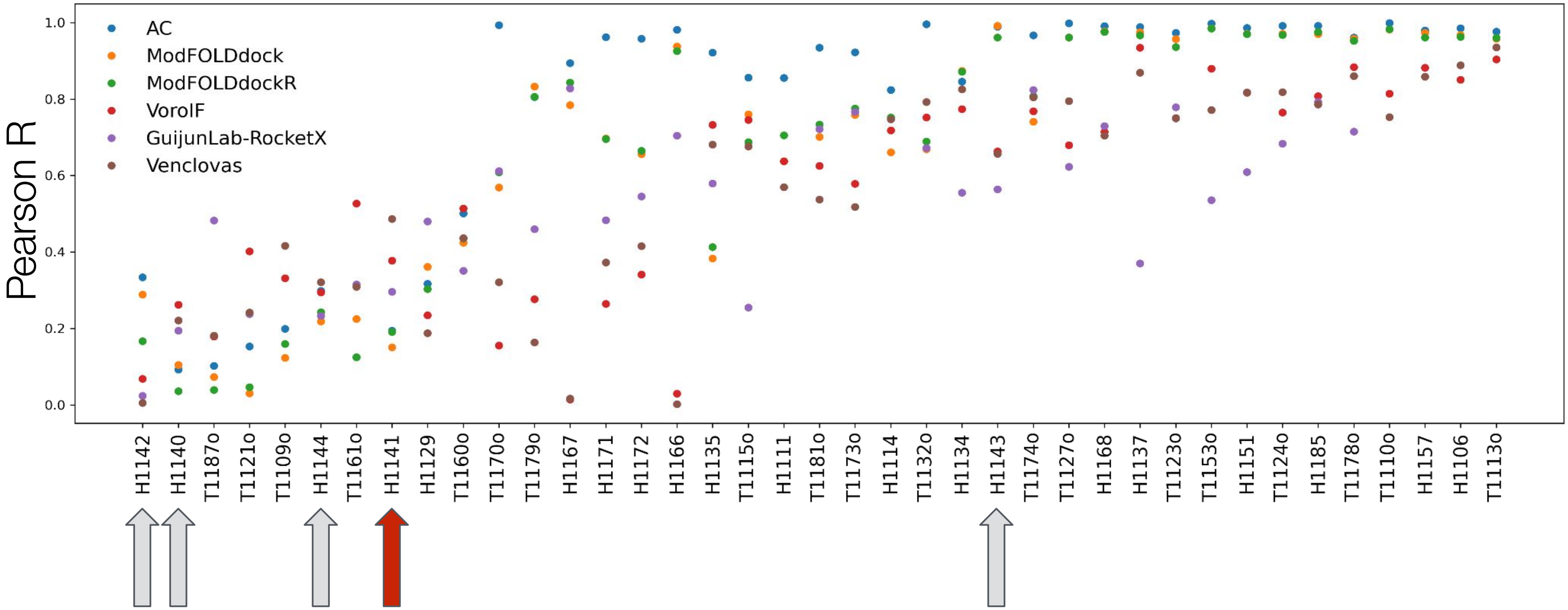




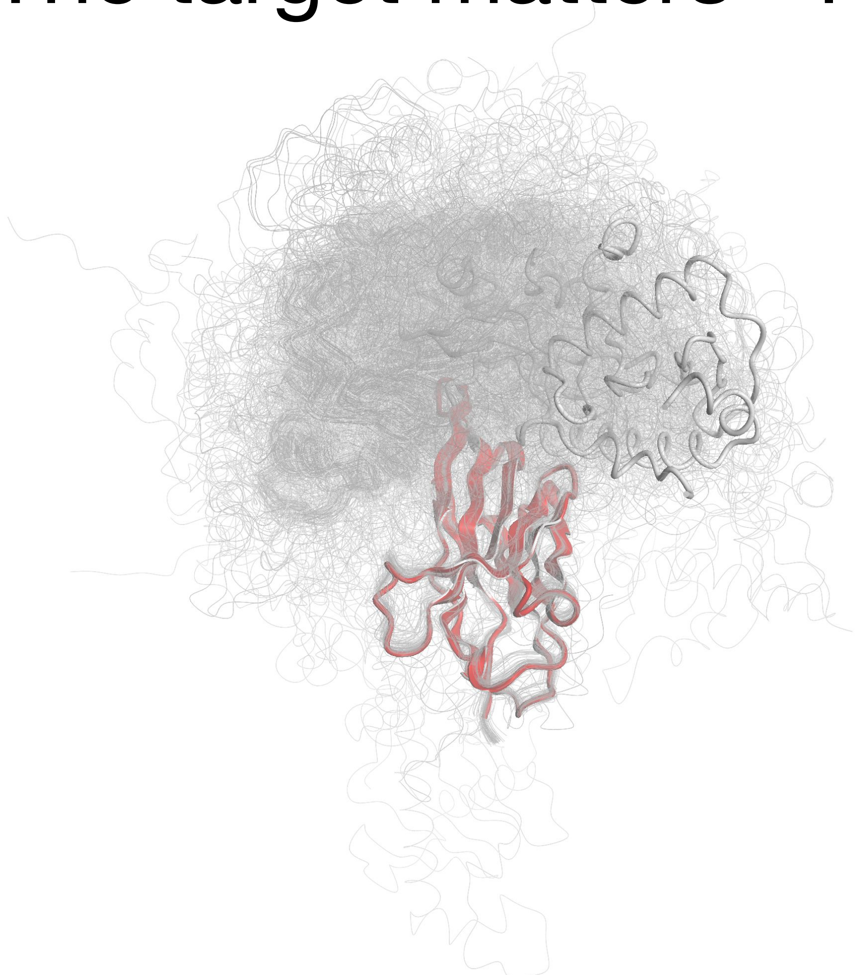


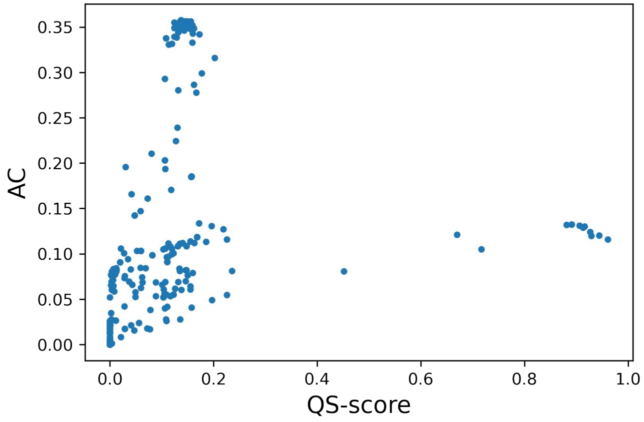
The target matters - Nanobodies

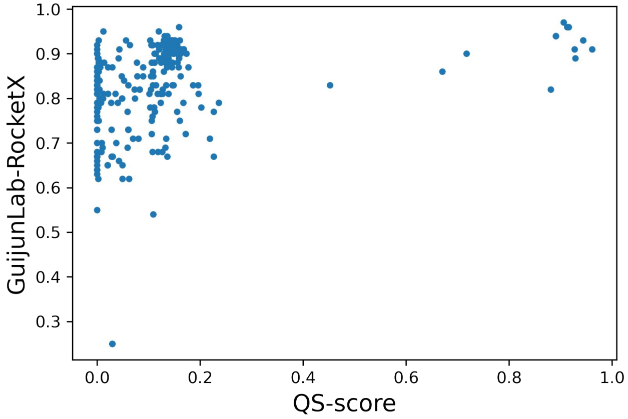
QS-score vs QSCORE



The target matters - Nanobodies - H1141

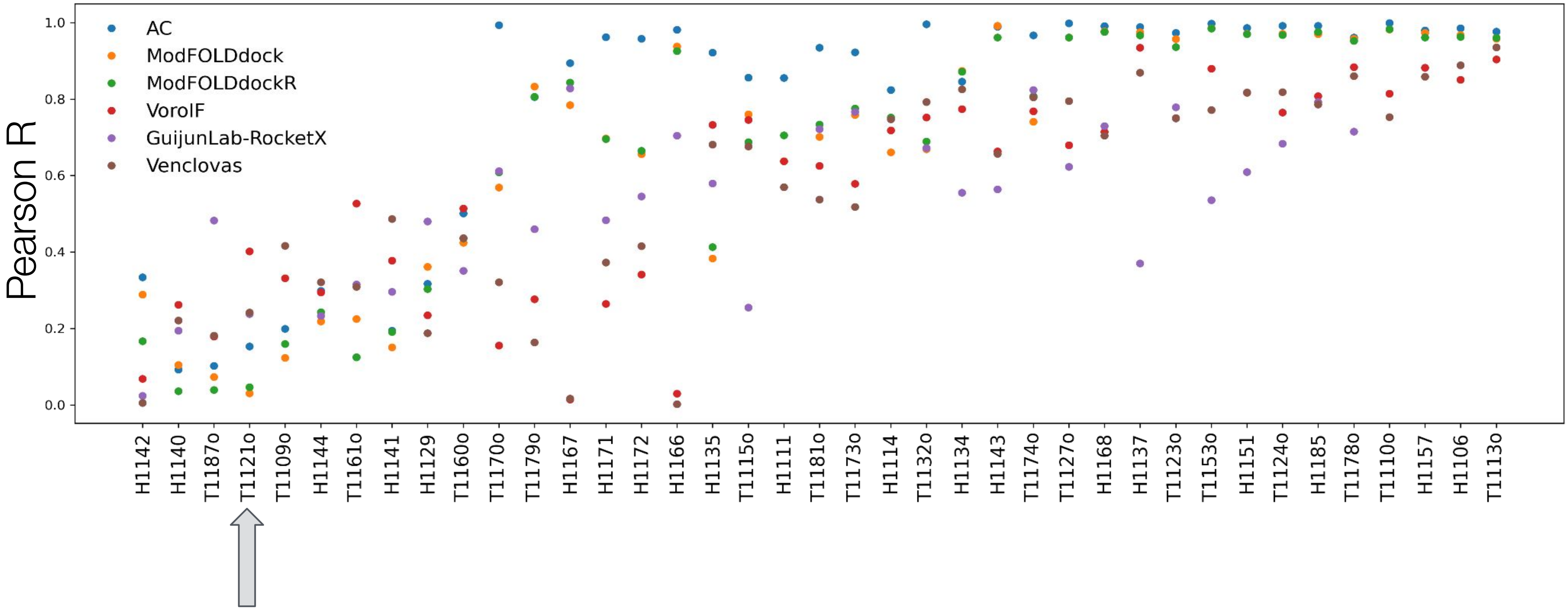




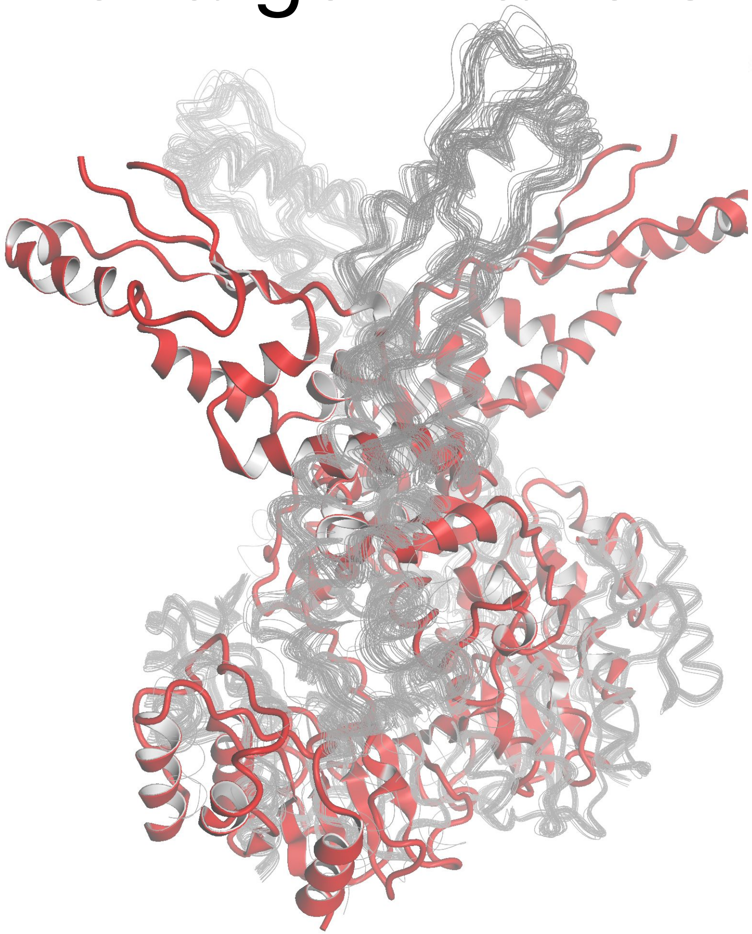


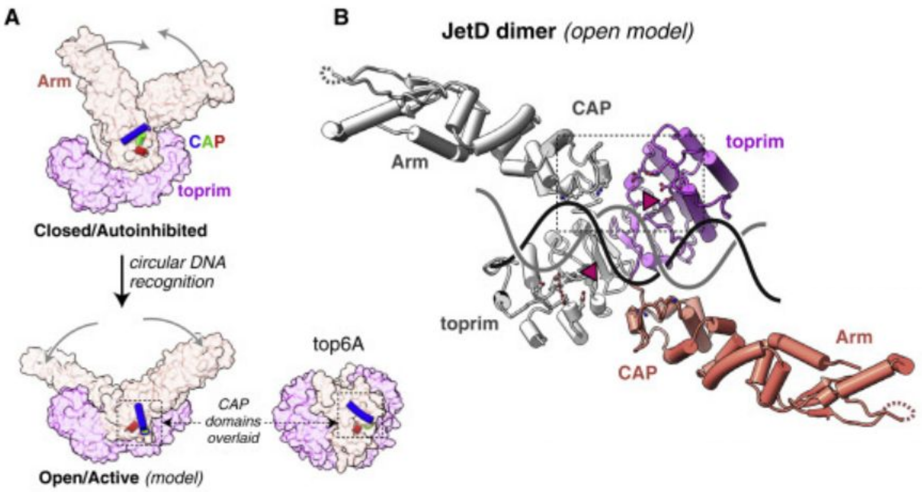
The target matters - Flexibility? T1121o

QS-score vs QSCORE



The target matters - Flexibility? T1121o





"Our structural data on JetABC and JetD reveal a single conformation of these complexes, which is insufficient to fully model or explain the observed biochemical activities of Wadjet. In particular, the structure of JetD reveals an inactive closed state, and we are so far unable to directly measure DNA binding by JetD or visualize its putative open state that binds and cleaves DNA"
Deep et al., 2022

Acknowledgements

Schwede People!


Gerardo Tauriello


Ezgi Karaca


CASP Wizard


Torsten Schwede


Burcu Özden

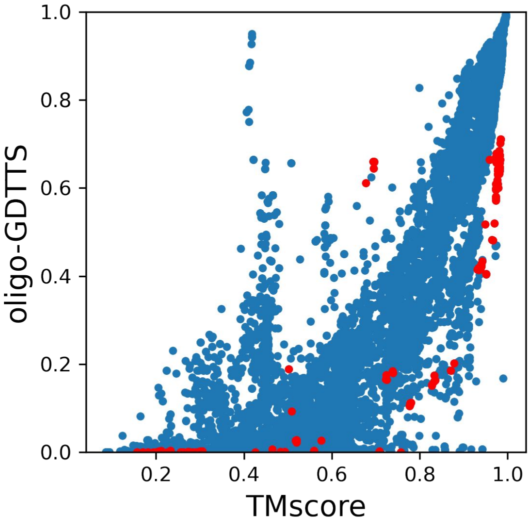

Marc Lensink

Assessment Code: https://git.scicore.unibas.ch/schwede/casp15_ema

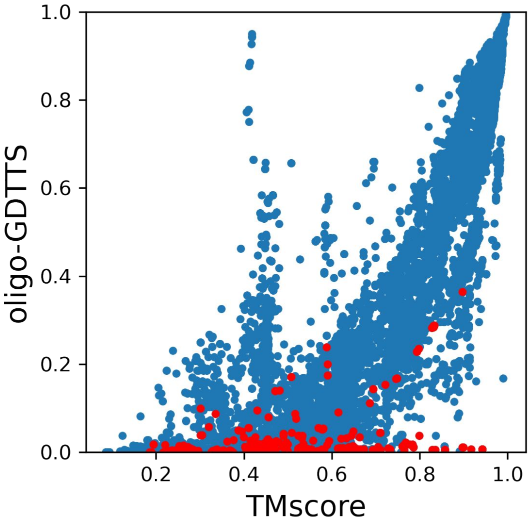
Backup slides

Large Complexes - TMscore vs oligo-GDTTS

H1111



H1135



T1115