GuijunLab-QA
GuijunLab-Human

Evaluation of Protein Complex Quality using DeepUMQA-X Server

Presenter: Fang Liang (PHD candidate)

Mentor: Guijun Zhang (Professor)

College of Information Engineering
Zhejiang University of Technology

CASP16, December 2024

Email: *zgj@zjut.edu.cn*

**http://www.zhanglab-bioinf.com/*

CONTENTS

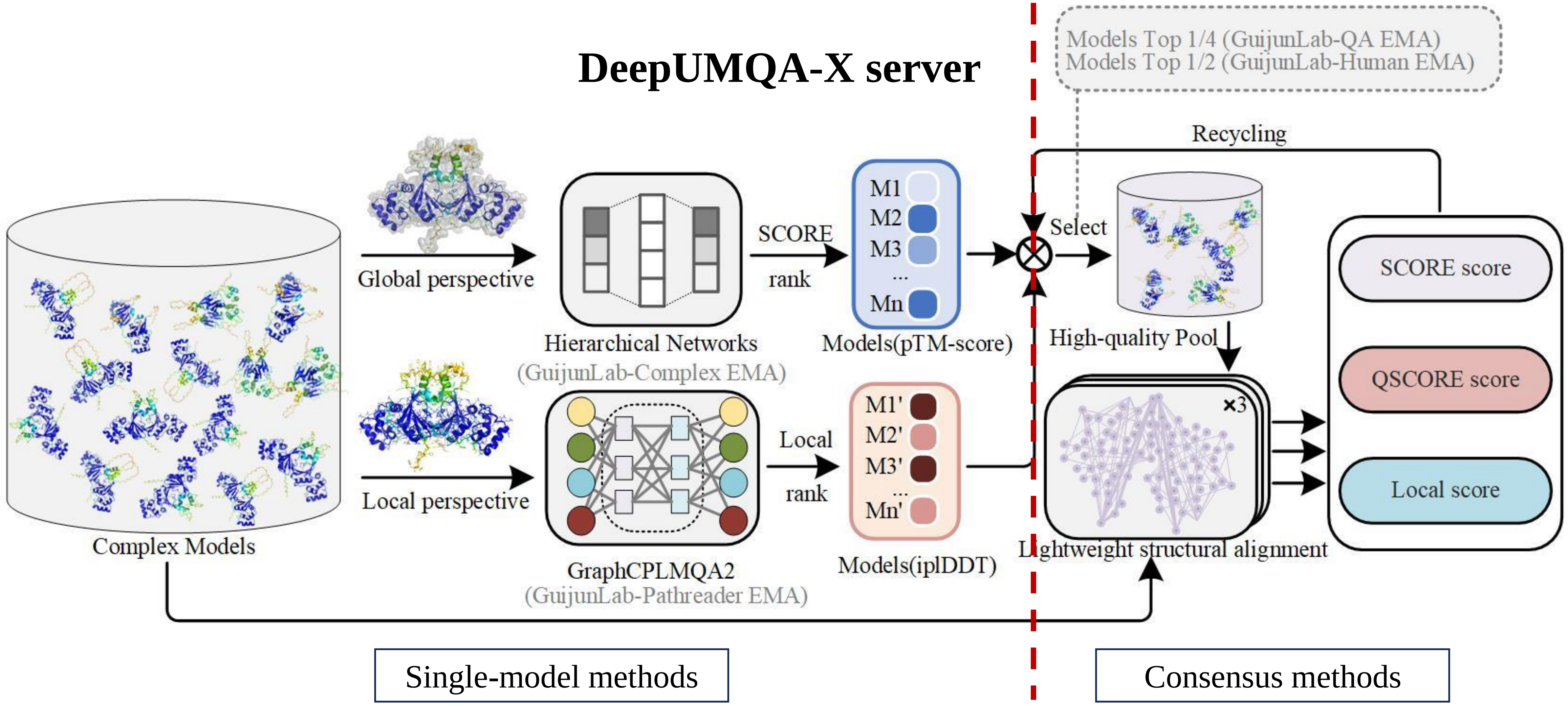
CONTENTS

Motivation

- Consensus EMA methods can show good performance when the quality of the model pool is good. Single-model EMA methods can show better performance than the consensus method when the quality of the model pool is poor.

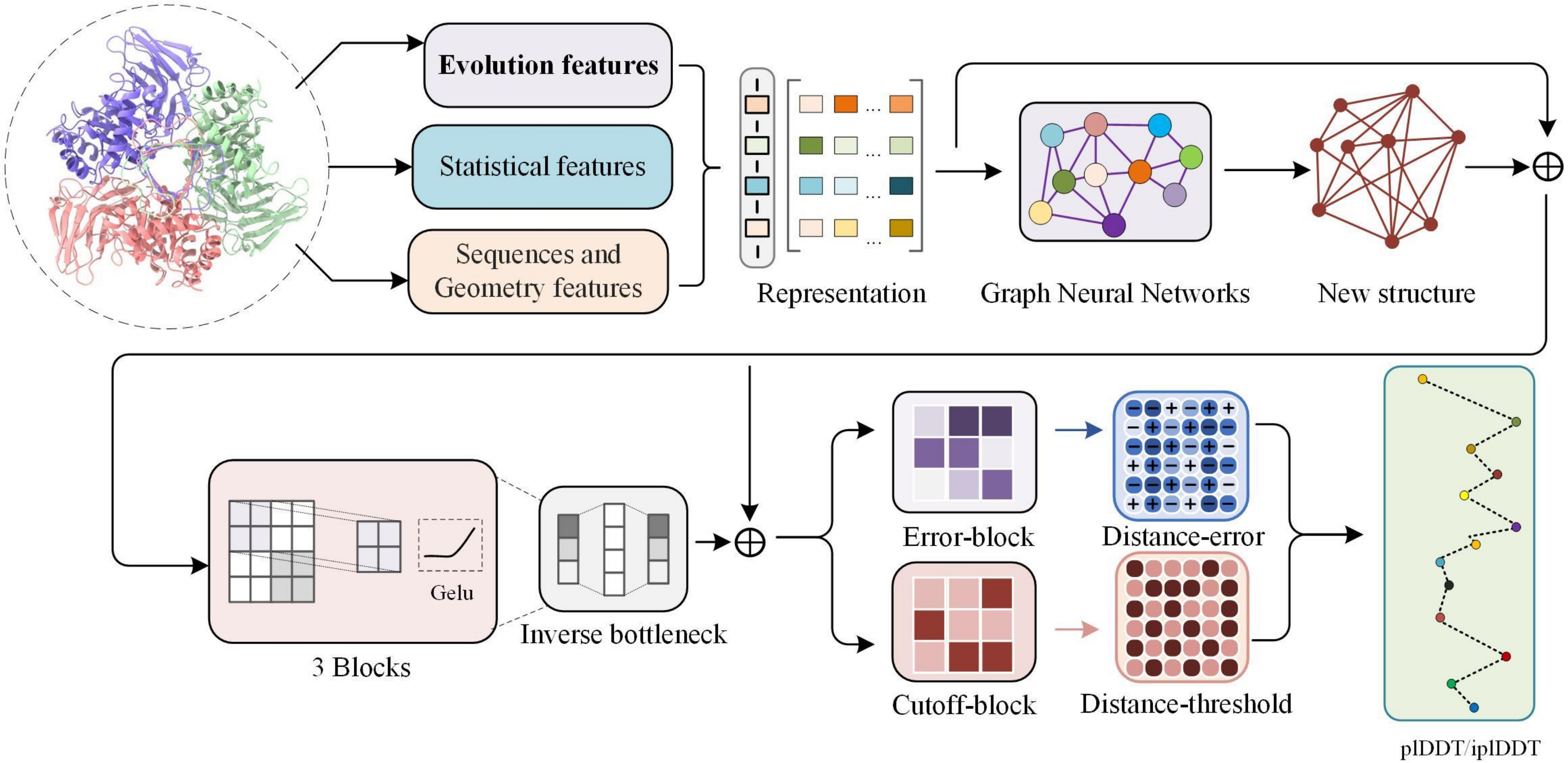
- How to leverage the advantages of single-model methods and consensus methods to improve the performance of protein model quality assessment?

DeepUMQA-X server Pipeline



Single-model EMA Methods

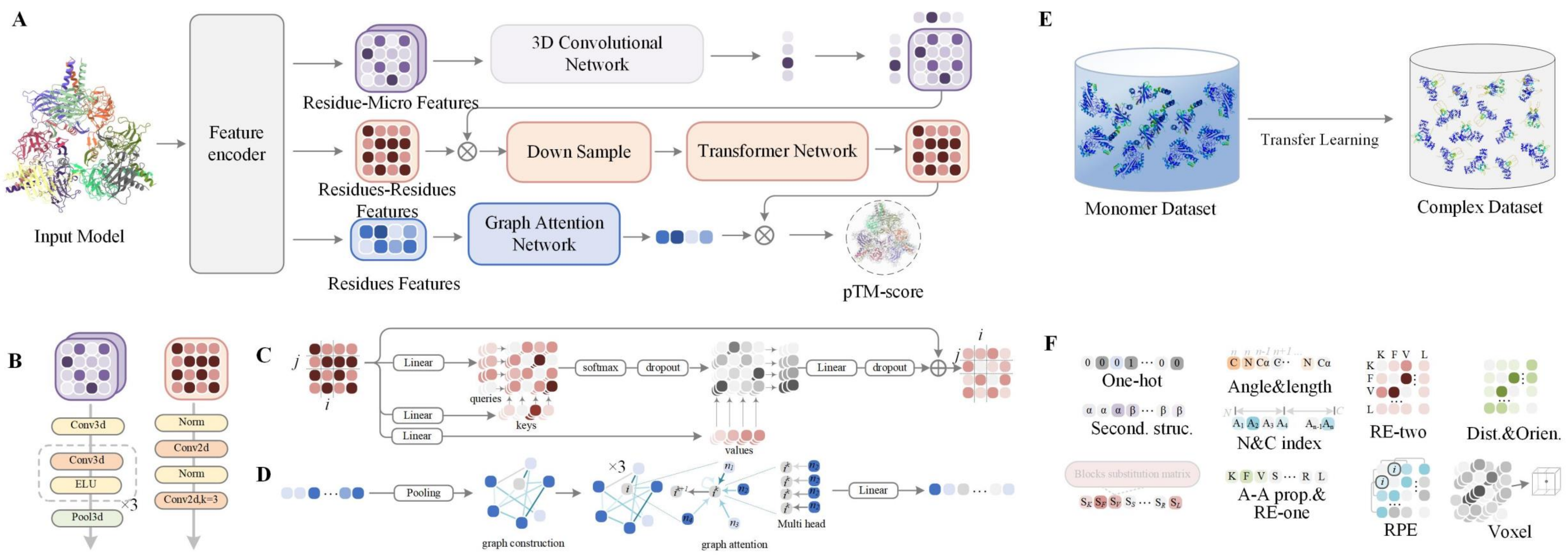
Single-Model Method 1 for lDDT (GraphCPLMQA2)



Local assessment of complex model quality using enhanced network and Evoformer representation based on GraphCPLMQA (GuijunLab-Pathreader EMA)

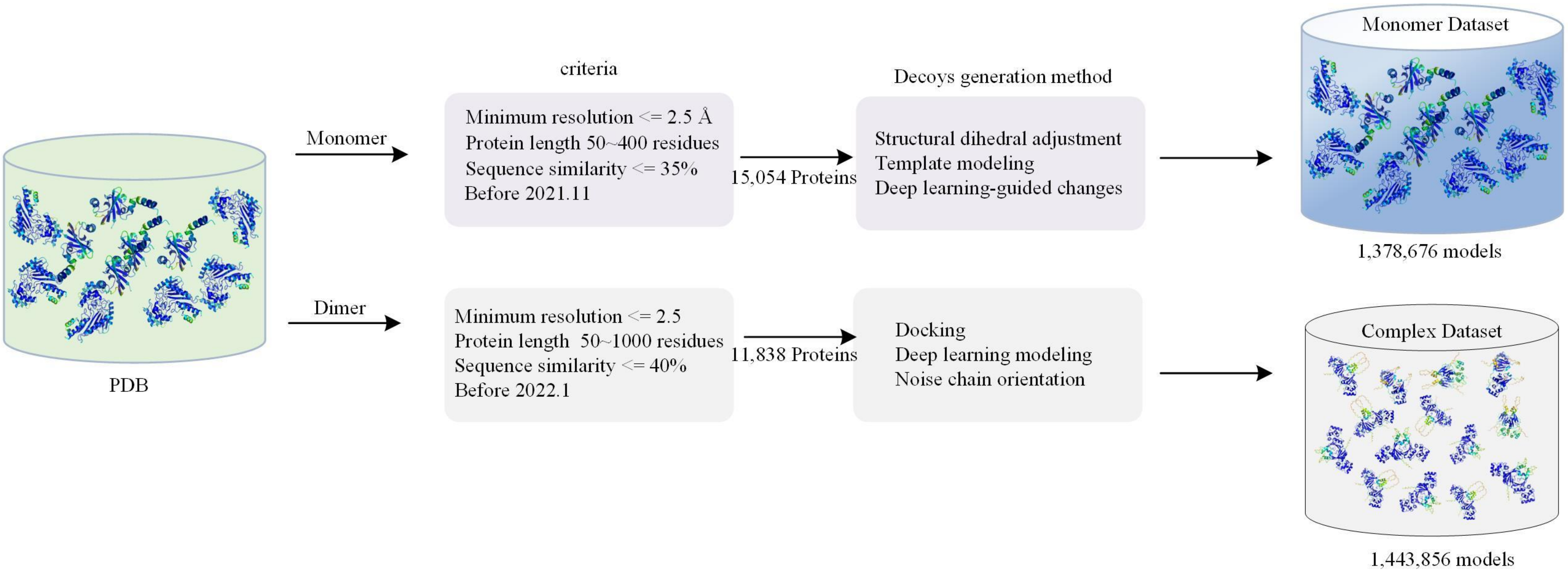
GraphCPLMQA Liu et al., BIB 2024

Single-Model Method 2 for TM-score



Evaluation of protein complex quality based on multivariate representation using hierarchical networks (GuijunLab-Complex EMA)

Single-Model Methods of Dataset



The construction process of monomer database and complex database

Protein Features of Single-Model Methods

Geometry features

- **Triangle Position (USR improved version)**
- Voxelization
- Distance and orientation
- Backbone angle and length
- Second structure

Sequence features

- Amino Acid Properties
- Relative position encoding

Evolution features

- Embedding of protein language models
- MSA Embedding of AlphaFold

Statistical features

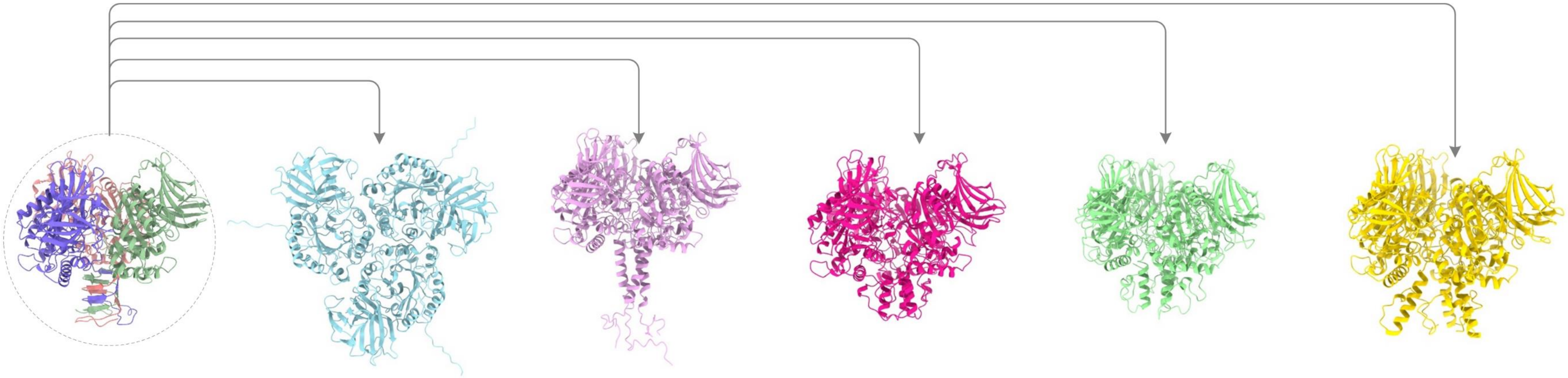
- Rosetta energy
- Blosum62 score matrix

Consensus EMA methods

Structural Alignment of Consensus Methods

A naive way to obtain consensus information between multiple models:

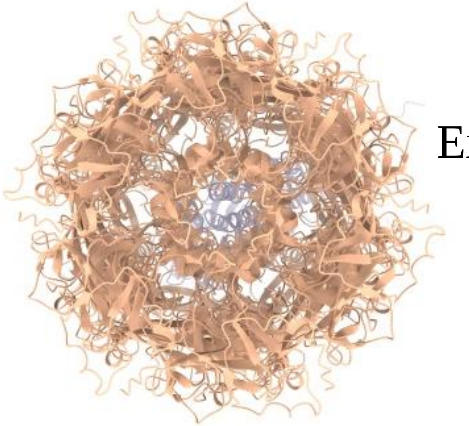
Structural alignment



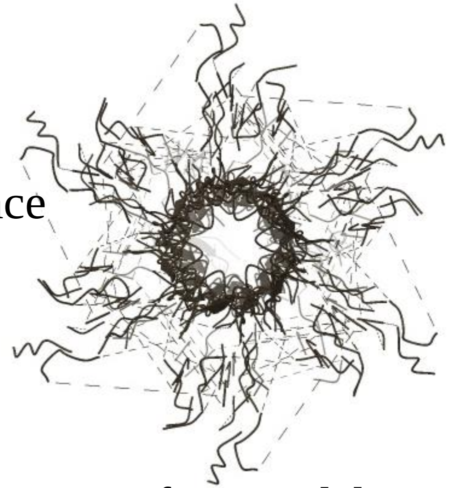
Lightweight Structure Align

How to quickly align large assemblies or massive complex model structures?

- **Interface Sequence Align (At least 10 times faster)**



Extract interface



-AJGL-AJ-PI-FJHGL..

-AJGL-AJ-PI-FJHGL...
-AJ-L-AJ-PIG--HGL…
D-SJLKFG--KL-SGA…
J-L-AJ-PIH--HGL…

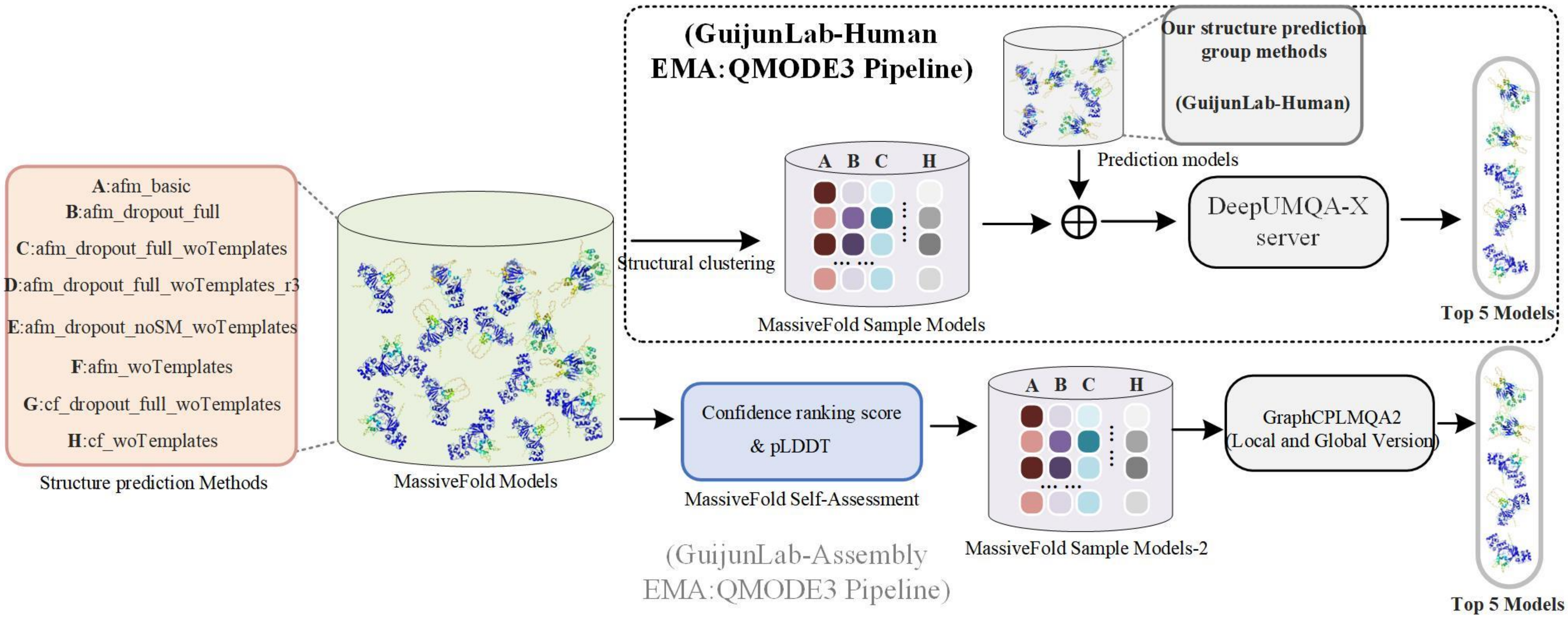
H1227 model-5689 AA

H1227 interface model-1560 AA

Interface sequence

Interface Residues Align

QMODE3: Model Selection Pipeline



- MassiveFold Sample Models (structural clustering)
- MassiveFold Sample Models-2 (MassiveFold Self-Assessment Selection)

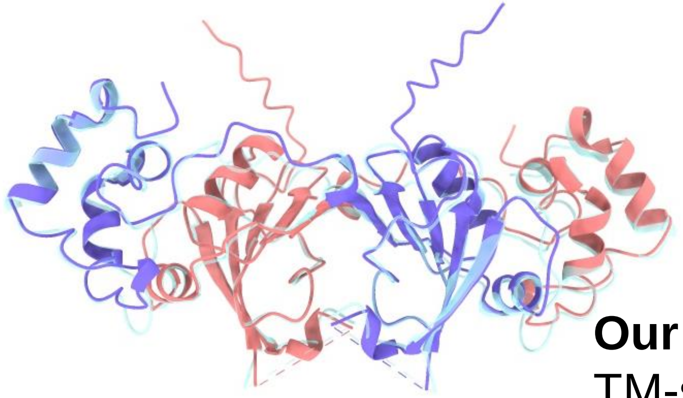
Results

Case Study: T1201o - Homomer

Summary of submitted prediction models

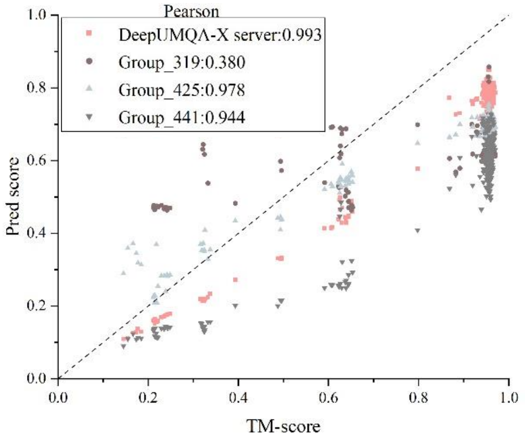
Type	Pearson	AUC	Top1Loss
SCORE	0.97	0.68	0.03
QSCORE	0.96	0.63	0.02
Local	0.81	0.67	-

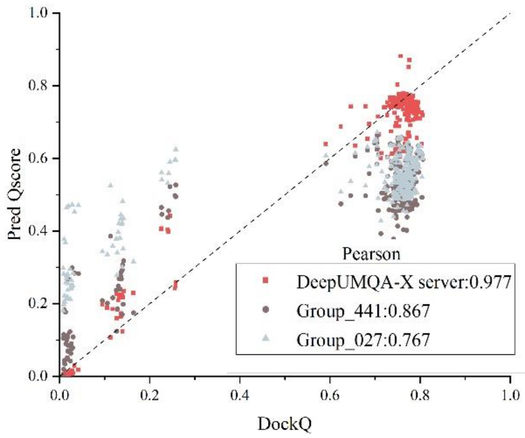
Stoichiometry:A2 (420 AA)

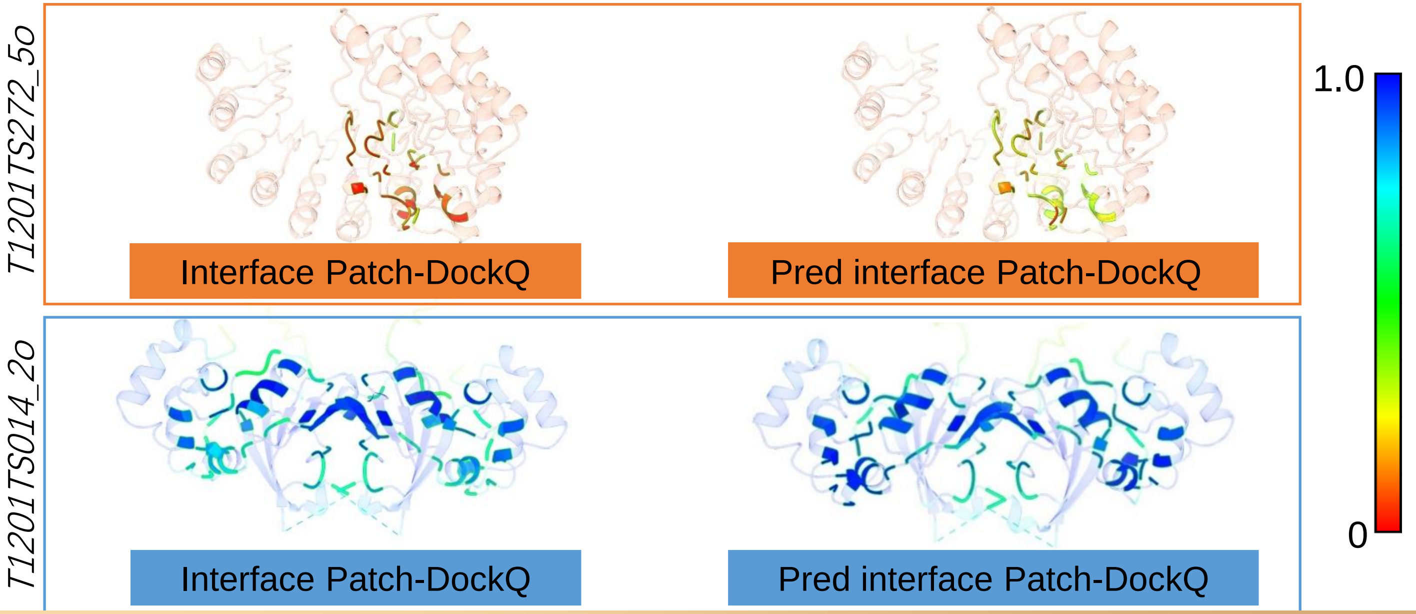


T1201TS293_2o

Our Top1 Model
TM-score: 0.959
DockQ: 0.771







Case Study: T1201o - Homomer

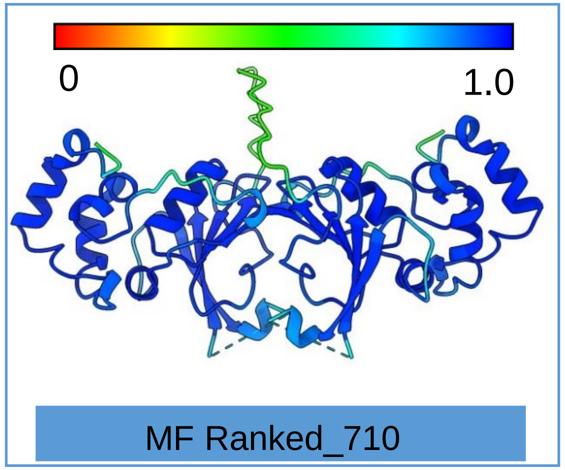
Summary of MassiveFold (MF) sample models

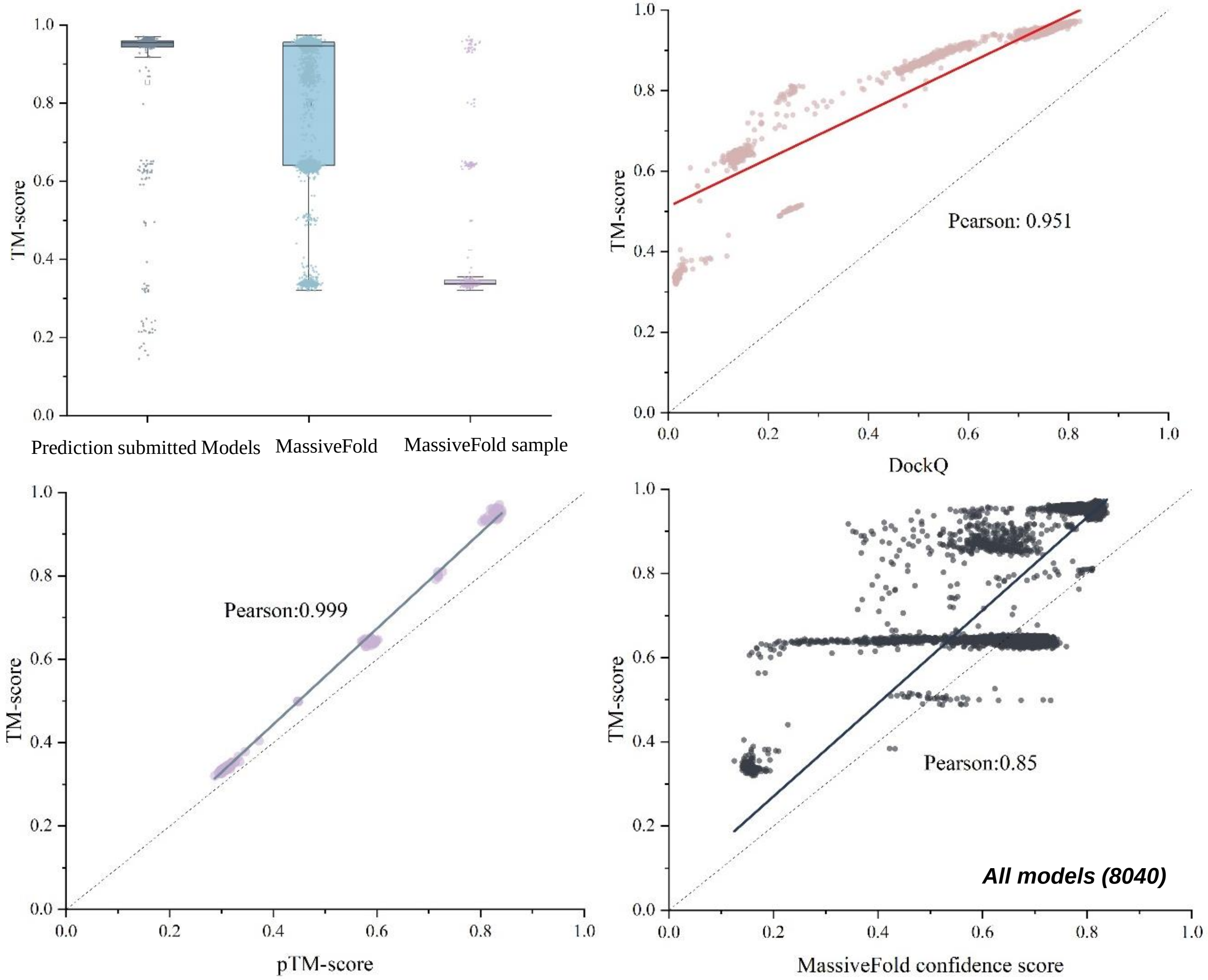
TM-score	Pearson	Spearman	Top1Loss
Ours	0.999	0.926	0.016
MF	0.961	0.370	-

Top 5 Models

Models	TM-score
Our_Top1	0.957
Our_Best	0.970
MF_Top1	0.943
MF_Best	0.961
Best	**0.973**

Our *Top 1*



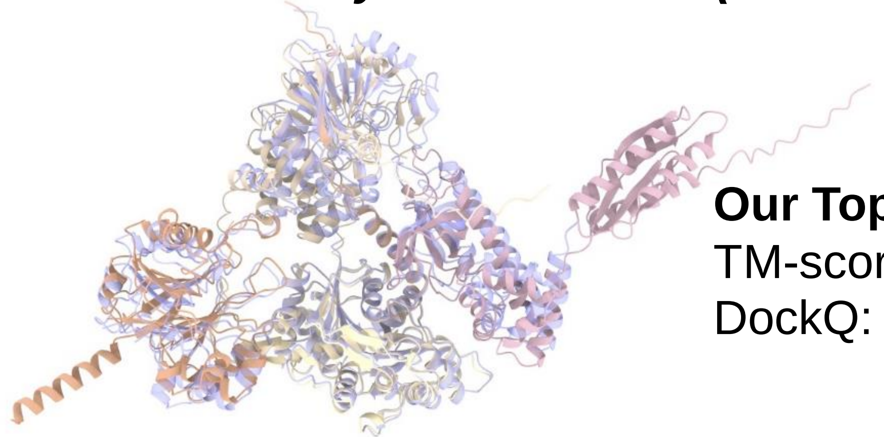


Case Study: H1213 - Heteromer

Summary of submitted prediction models

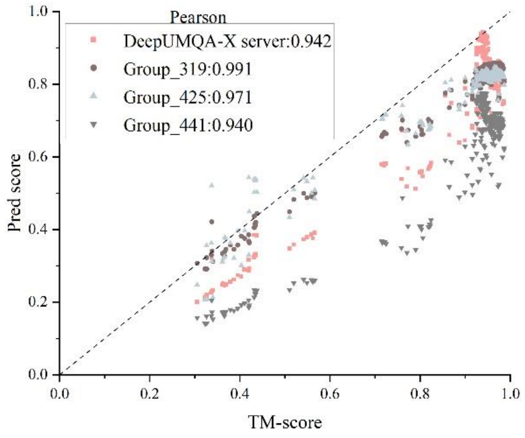
Type	Pearson	AUC	Top1Loss
SCORE	0.944	0.585	0.164
QSCORE	0.960	0.642	0.07
Local	0.789	0.812	-

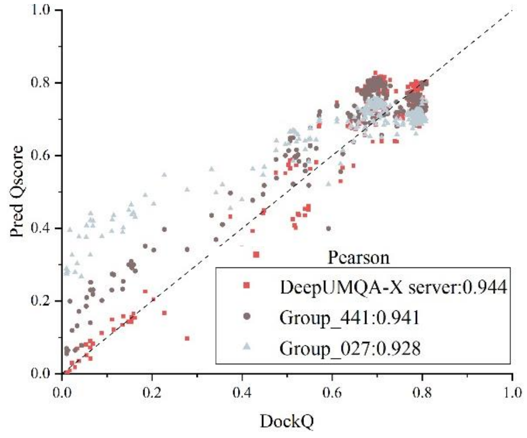
Stoichiometry:A1B1C1D1E1 (1373AA)



H1213TS163_3

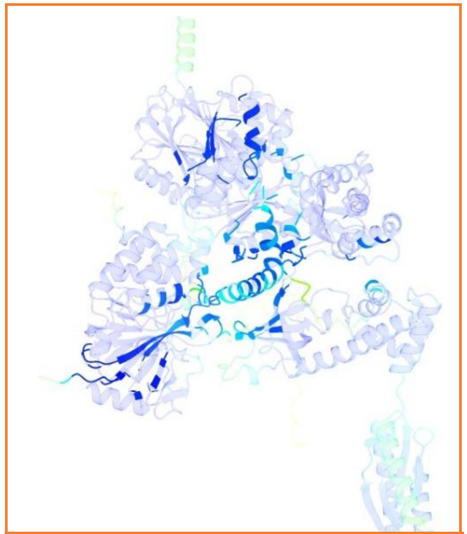
Our Top1 Model
TM-score: 0.946
DockQ: 0.697



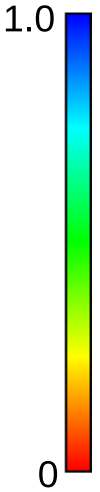


H1213TS198_2

Interface Patch-DockQ



Pred interface Patch-DockQ



Case Study: H1213 - Heteromer

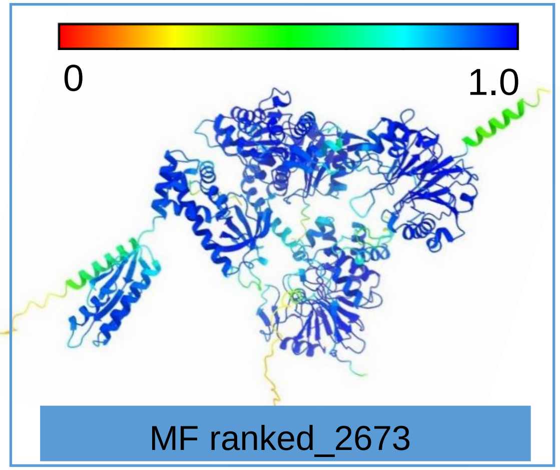
Summary of MassiveFold (MF) sample models

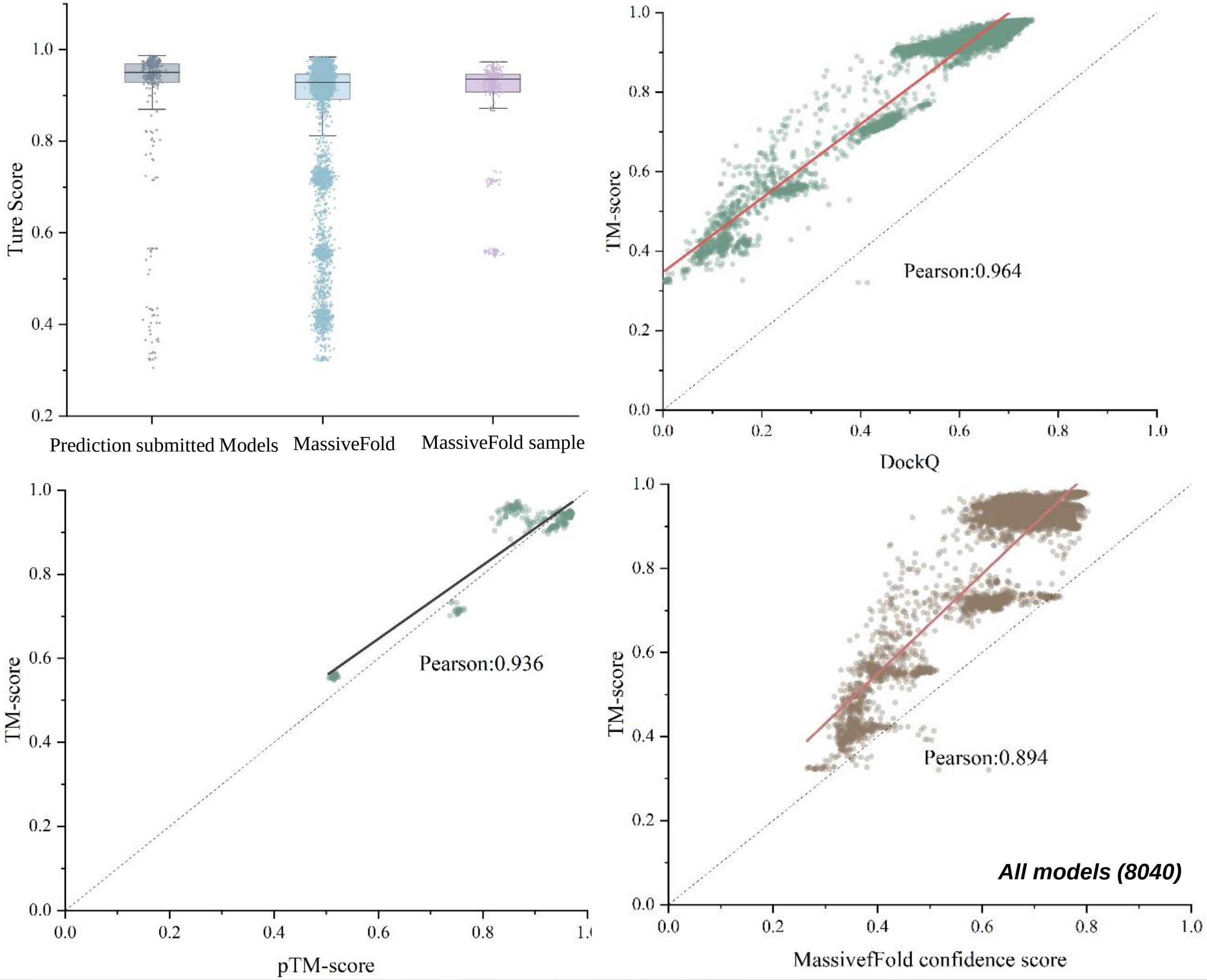
TM-score	Pearson	Spearman	Top1Loss
Ours	0.936	0.450	0.052
MF	0.910	0.509	-

Top 5 Models

Models	TM-score
Our_Top1	0.931
Our_Best	0.979
MF_Top1	0.936
MF_Best	0.978
Best	**0.984**

Our *Top 1*



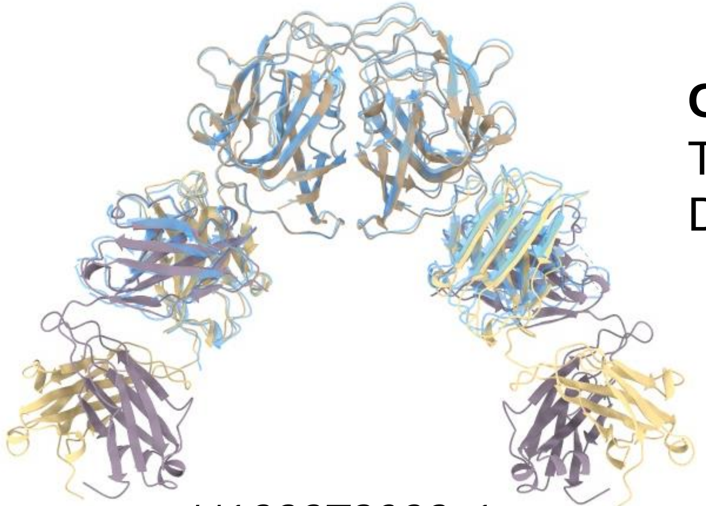


Case Study: H1233 - Antibody antigen

Summary of submitted prediction models

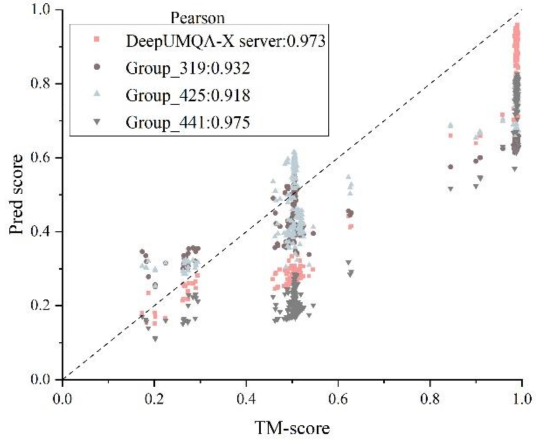
Type	Pearson	AUC	Top1Loss
SCORE	0.972	0.919	0.01
QSCORE	0.952	0.877	0.02
Local	0.803	0.762	-

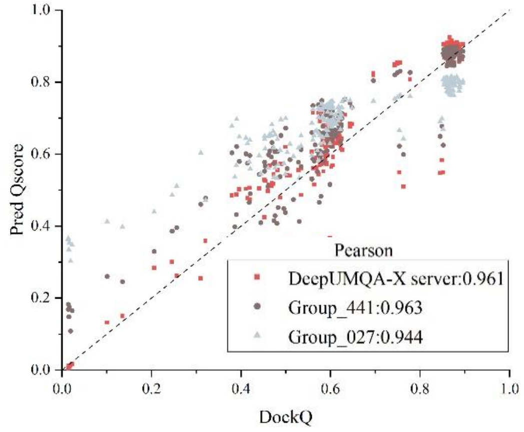
Stoichiometry:A2B2C2 (1316AA)



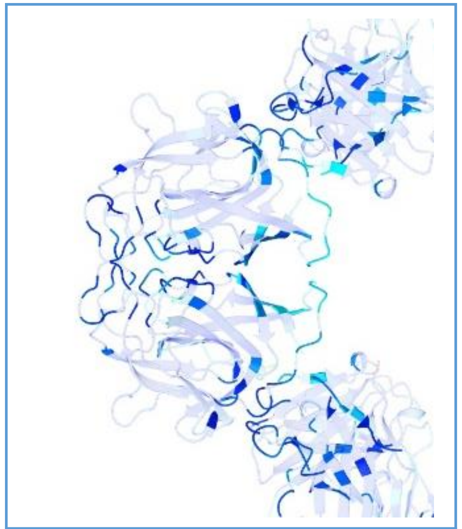
H1233TS028_1

Our Top1 Model
TM-score: 0.990
DockQ: 0.866

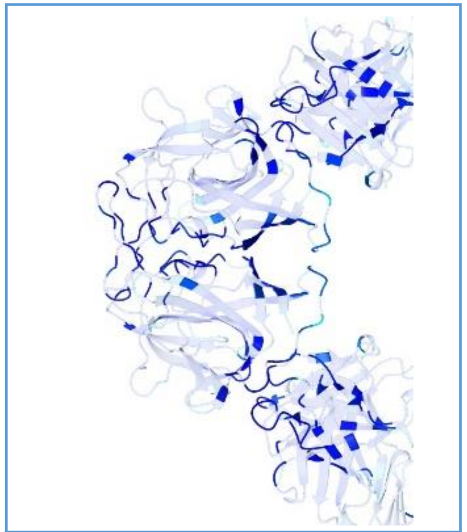




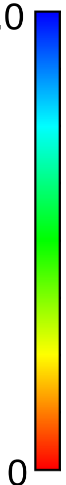
H1233TS272_5



Interface Patch-DockQ



Pred interface Patch-DockQ



Case Study: H1233 - Antibody antigen

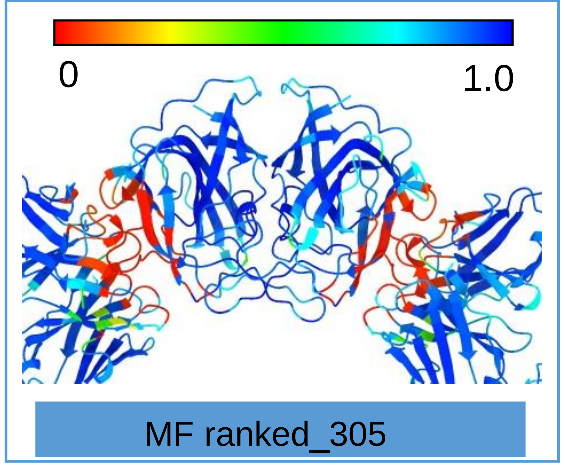
Summary of MassiveFold (MF) sample models

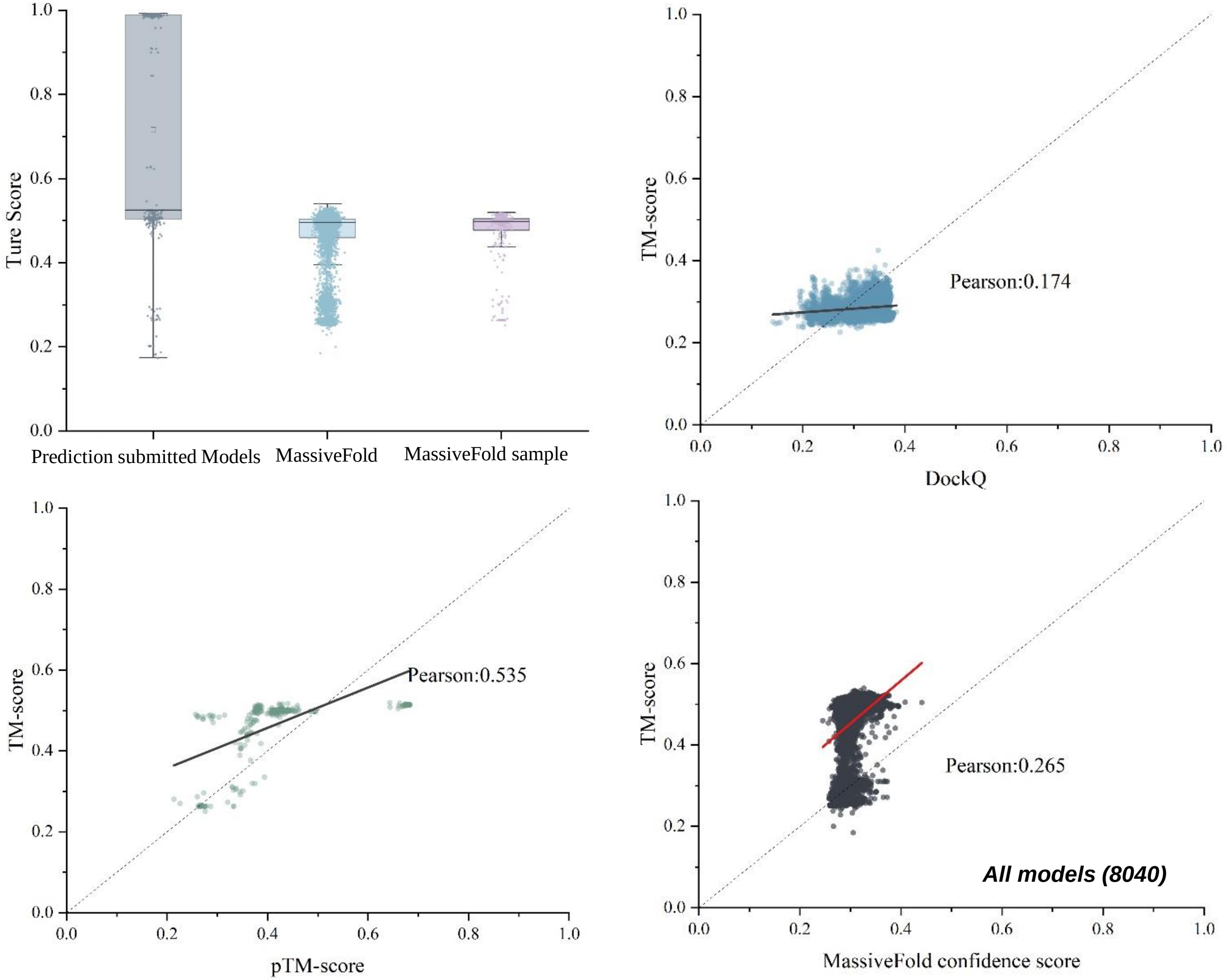
TM-score	Pearson	Spearman	Top1Loss
Ours	0.535	0.759	0.023
MF	0.449	0.539	-

Top 5 Models

Models	TM-score
Our_Top1	0.516
Our_Best	0.516
MF_Top1	0.504
MF_Best	0.505
Best	**0.539**

Our *Top 1*



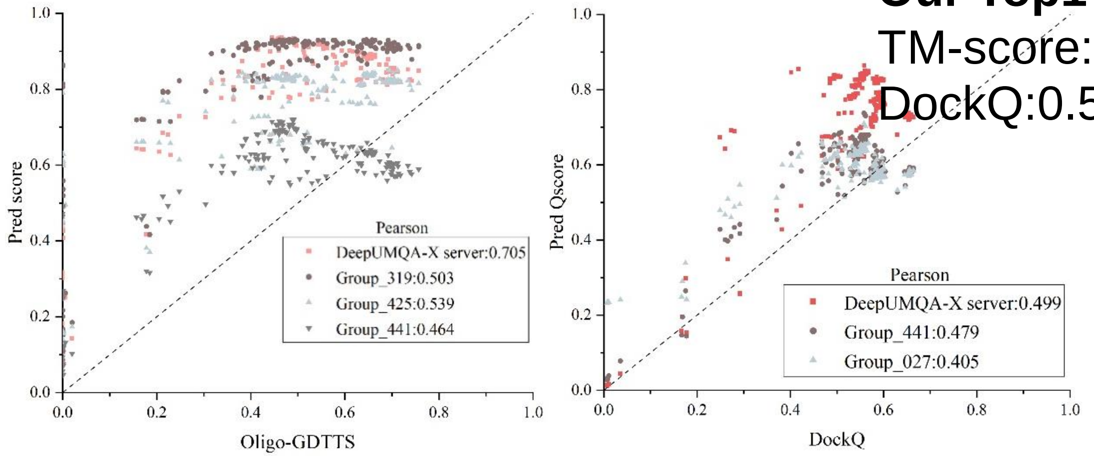


Case Study: H1227 - Large assembly

Summary of submitted prediction models

Type	Pearson	AUC	Top1Loss
SCORE	0.886	0.5	0.15
QSCORE	0.807	0.5	0.08
Local	0.581	0.824	-

Stoichiometry:A1B6 (5689 AA)

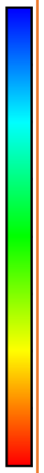


Our Top1 Model
TM-score:0.948
DockQ:0.561

H1227TS022_1

H1227TS051_5

1.0



0

Interface Patch-DockQ

Pred Interface Patch-DockQ

Case Study: H1227 - Large assembly

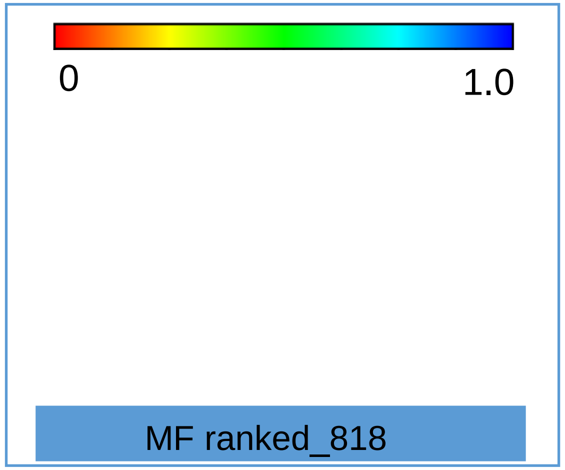
Summary of MassiveFold (MF) sample models-2

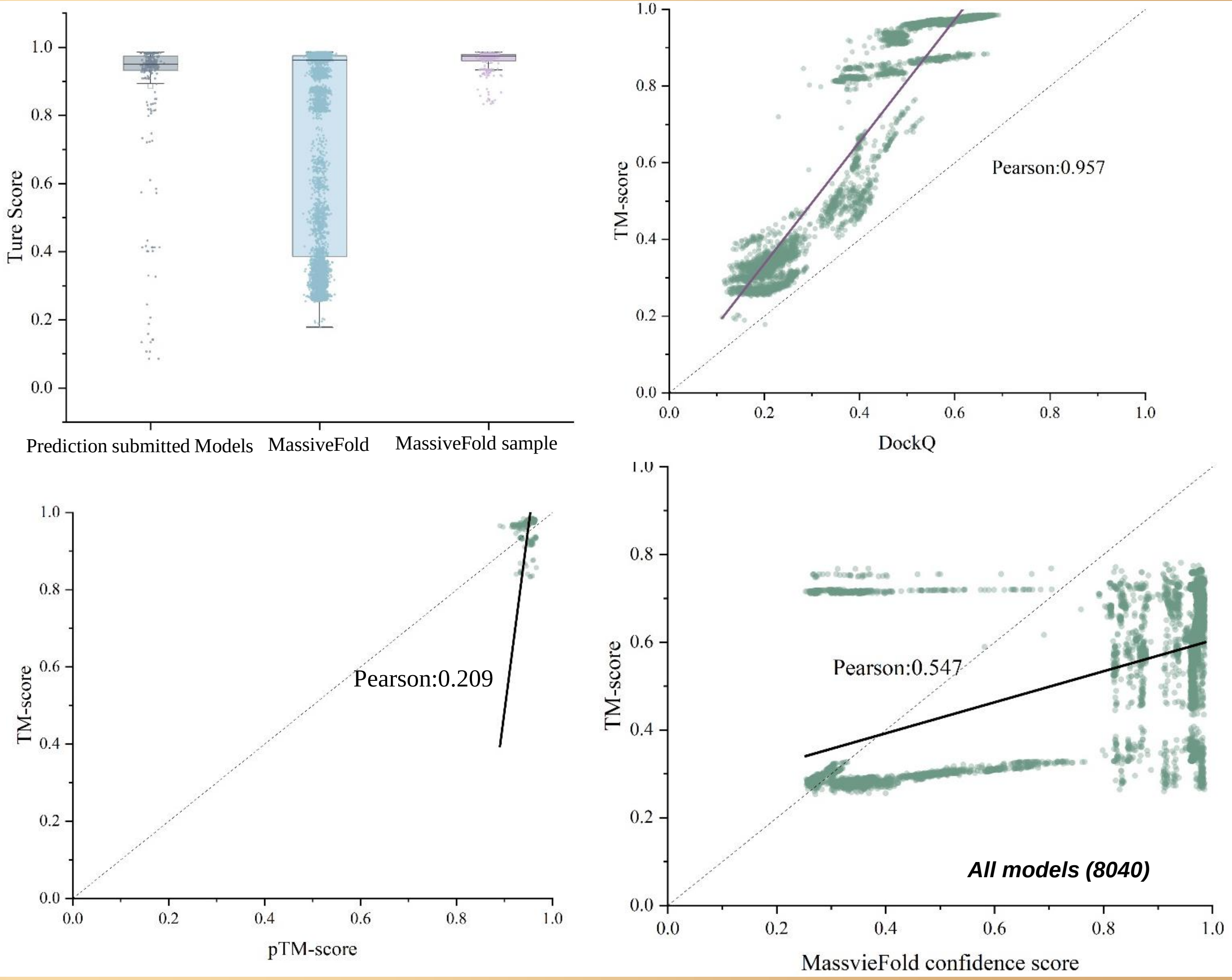
TM-score	Pearson	Spearman	Top1Loss
Ours	0.209	0.488	0.06
MF	0.275	0.391	-

Top 5 Models

Models	TM-score
Our_Top1	0.980
Our_Best	0.981
MF_Top1	0.942
MF_Best	0.942
Best	**0.986**

Our *Top 1*



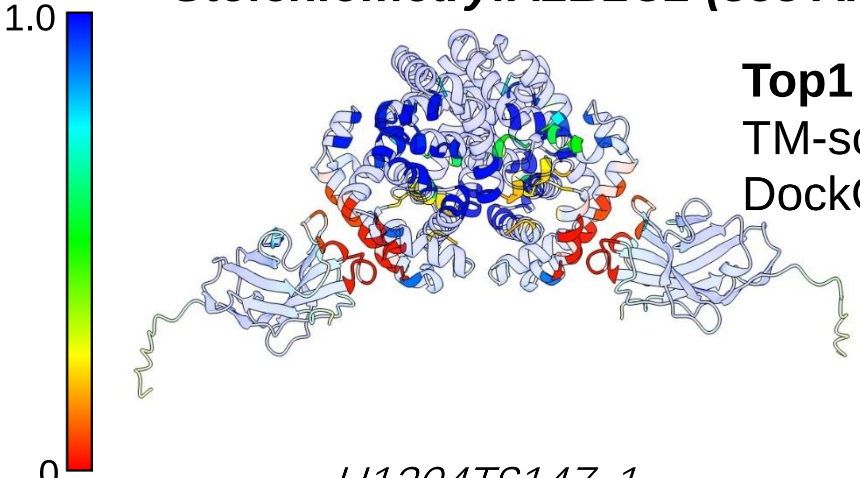


Case Study: H1204 - Nanobody Complex

Summary of submitted prediction models

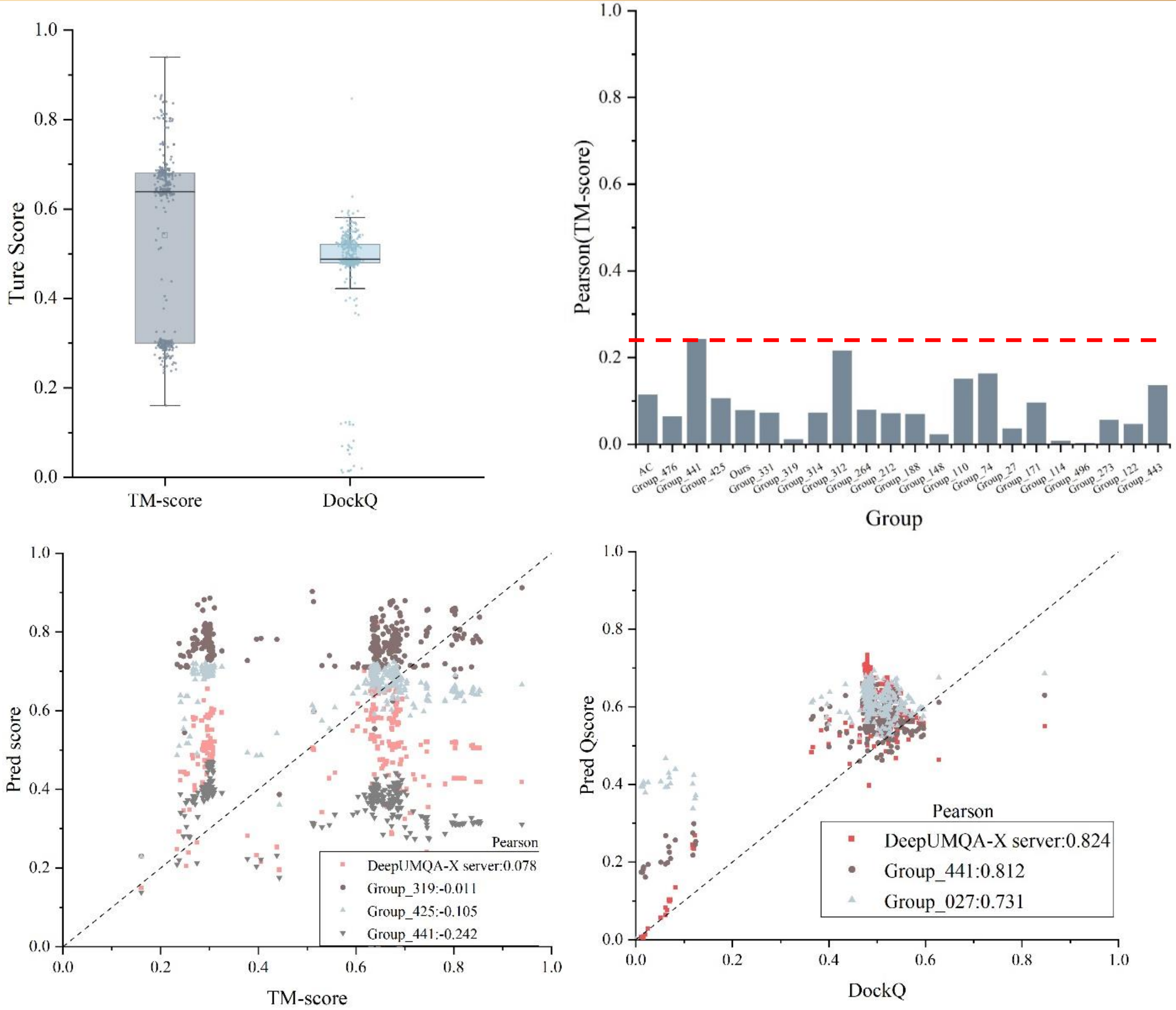
Type	Pearson	AUC	Top1Loss
SCORE	0.131	0.500	0.300
QSCORE	0.803	0.530	0.416
Local	0.771	0.889	-

Stoichiometry:A2B2C2 (858 AA)



Top1 Model
TM-score:0.640
DockQ:0.480

H1204TS147_1



Conclusions

Work

- **DeepUMQA-X Server**
- GraphCPLMQA2 (Improved version)
- Global scoring Method
- Interface Sequence Align （Lightweight Structure Align ）

Assessment Challenges

- Large Assemblies
- Antibody-antigen complex
- **Ensemble structures**

Group member and acknowledgements

Professor
Guijun Zhang

Dong Liu

Xuanfeng Zhao

Haodong Wang

Fang Liang

Meng Sun

Tianyou Zhang

Lei Xie

Minghua Hou

Xinyue Cui

Suhui Wang

GuijunLab-EMA Group
& PhD students

Guijun Zhang Lab members

CASP organizers

中华人民共和国科学技术部
Ministry of Science and Technology (MoST)

Ministry of Science and Technology

College of information engineering,
Zhejiang University of Technology

BioMap