Received: 3 May 2023 | Revised: 17 July 2023 | Accepted: 31 July 2023

DOI: 10.1002/prot.26570

RESEARCH ARTICLE

PROTEINS WILEY

Improved protein structure prediction with trRosettaX2, AlphaFold2, and optimized MSAs in CASP15

Zhenling Peng[1] | Wenkai Wang[2] | Hong Wei[2] | Xiaoge Li[1] | Jianyi Yang[1]

[1]MOE Frontiers Science Center for Nonlinear Expectations, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China

[2]School of Mathematical Sciences, Nankai University, Tianjin, China

Correspondence
Jianyi Yang, MOE Frontiers Science Center for Nonlinear Expectations, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao 266237, China.
Email: yangjy@sdu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Numbers: T2225007, T2222012

Abstract

We present the monomer and multimer structure prediction results of our methods in CASP15. We first designed an elaborate pipeline that leverages complementary sequence databases and advanced database searching algorithms to generate high-quality multiple sequence alignments (MSAs). Top MSAs were then selected for the subsequent step of structure prediction. We utilized trRosettaX2 and AlphaFold2 for monomer structure prediction (group name Yang-Server), and AlphaFold-Multimer for multimer structure prediction (group name Yang-Multimer). Yang-Server and Yang-Multimer are ranked at the top and the fourth, respectively, for monomer and multimer structure prediction. For 94 monomers, the average TM-score of the predicted structure models by Yang-Server is 0.876, compared to 0.798 by the default AlphaFold2 (i.e., the group NBIS-AF2-standard). For 42 multimers, the average DockQ score of the predicted structure models by Yang-Multimer is 0.464, compared to 0.389 by the default AlphaFold-Multimer (i.e., the group NBIS-AF2-multimer). Detailed analysis of the results shows that several factors contribute to the improvement, including improved MSAs, iterated modeling for large targets, interplay between monomer and multimer structure prediction for intertwined structures, etc. However, the structure predictions for orphan proteins and multimers remain challenging, and breakthroughs in this area are anticipated in the future.

KEYWORDS
AlphaFold2, multiple sequence alignment, multimer structure prediction, protein structure prediction, trRosetta

1 | INTRODUCTION

The accuracy of protein structure prediction has reached an unprecedented level due to the development of artificial intelligence (AI) algorithms, the accumulation of big protein sequence/structure data, and the advance of hardware.[1] The major methods for protein structure prediction are shifting from template-based modeling (e.g., I-TASSER[2]) and fragment assembly (e.g., Rosetta[3]) to AI-based prediction (e.g., AlphaFold2,[4] RoseTTAFold,[5] and trRosetta[6]).

However, multimer structure prediction remains notoriously difficult. Traditionally, protein–protein docking is used to generate the multimer structure with the input of monomer structures. The docking-based approach relies on highly accurate monomer structure models, and conformational flexibility is hardly captured. Inspired by the success in monomer structure prediction, AI is introduced to improve multimer structure prediction. One example is the 'fold-and-dock' approach, in which folding and docking are used alternatively to predict the multimer structure.[7] The 'simultaneous folding' approach represents a significant advancement as it eliminates the need for the docking step, as exemplified by AlphaFold-Multimer[8] and RoseTTAFold.[5] Like monomer structure prediction, these methods also depend

Zhenling Peng, Wenkai Wang, and Hong Wei are co-first authors.

on the availability of high-quality multiple sequence alignments, which significantly restricts their potential applications.

Our group participated in the CASP15 experiment for both monomer and multimer structure prediction. During CASP15, we submitted predictions as both server (Yang-Server, Yang-Multimer) and human (Yang) groups. The methods used by the human group are similar to the server groups. Our human interventions are mainly in terms of more model generation and model order adjustment, which work well for a few targets. Nevertheless, on average, the submitted models from our human group are slightly worse than the models from our server groups. The analyses in this article are for the server groups only.

2 | METHODS

2.1 | trRosettaX2

We develop trRosettaX2 (denoted by trRX2 below), an improved version of trRosettaX,[9] inspired by the success of AlphaFold2 (AF2). trRX2 uses a transformer-based neural network, called trFormer, to predict inter-residue geometry from the raw multiple sequence alignment (MSA) to reduce information loss caused by handcrafted features (e.g., the MSA cannot be recovered from the MSA-derived co-evolution features in trRosetta). The network trFormer is similar to Evoformer in AF2 but has only 12 blocks due to limited hardware conditions. A few modifications were made in trFormer compared with Evoformer. The first one involves utilizing tied row-wise attention for updating the MSA representation. The second one is the parallel organization of row- and column-wise operations in the MSA updating, rather than the sequential manner in Evoformer. The third one incorporates convolution operations as a complement to the triangle updates.

The predicted inter-residue geometry is used to generate tertiary structure models based on energy minimization, as done in trRosetta.[6] The major reason for using the energy minimization step is to reduce the complexity of the network training, so that we could implement the algorithm under limited computer hardware condition. Tests on the CASP14 targets suggest that trRX2 outperforms RoseTTAFold and has comparable accuracy to AF2.[1] More details about the trRX2 methodology and benchmark results will be published elsewhere (Wang et al., in preparation).

2.2 | MSA generation

Three different methods are used to generate MSAs for regular targets (Figure 1). (i) The first one is based on HHblits search against three HMM profile databases (uniclust30_2018_08, uniref30_2202_02, and bfd).[10] A maximum of 25 MSAs are obtained from this search with various cutoffs for *e*-value (10^{-40}, 10^{-10}, 10^{-3}, 1), coverage (50%, 75%) and sequence identity (90% and 100%).[6,11] The top MSA is selected based on the average probability of the top 15*L* residue pairs in the predicted distance map, that is, with the metrics *mP*20 defined in ref. [12]. (ii) The second one is based on MMseqs2 search against two expandable profile databases (uniref30_2202 and colabfold_envdb_202108),[13] which generates three MSAs (one from each database and a combined version). In general, the above MSAs are enough to build accurate models for easy targets. (iii) The last one is mainly designed for hard targets. A FASTA

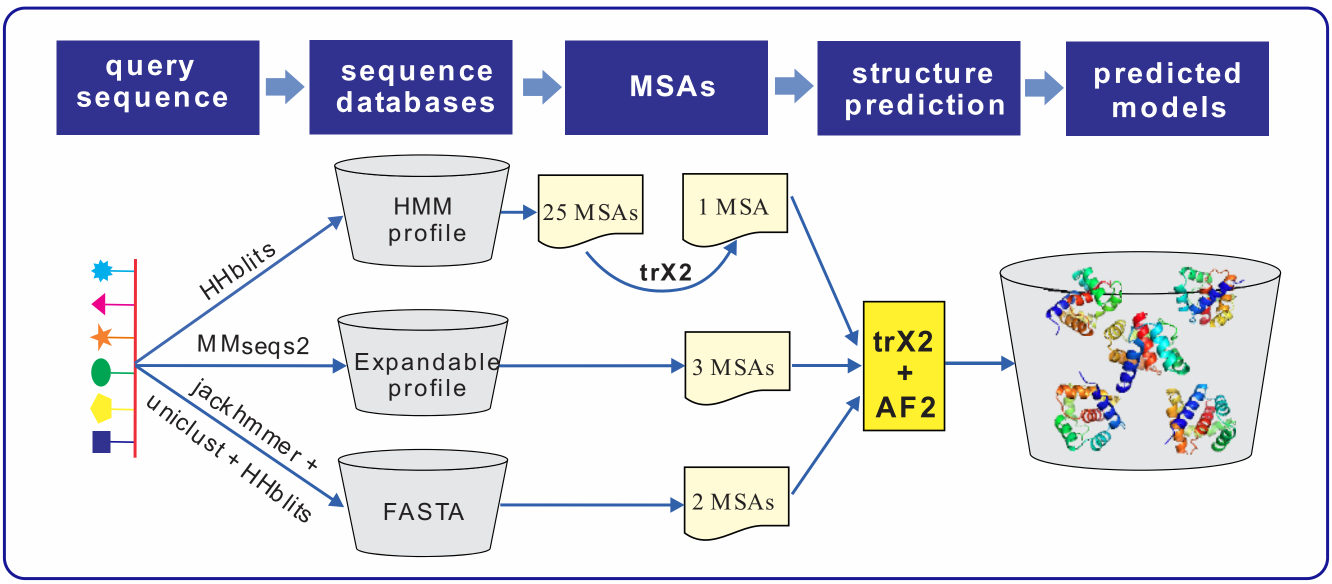


FIGURE 1 The MSA curation and structure prediction protocol by Yang-Server in CASP15. A maximum of 30 MSAs are generated from three different sequence databases. To speed up, the top MSA from the HMM profile database is selected by trRosettaX2 (trRX2) before running the structure prediction pipeline trRX2/AF2. In our benchmark tests, inclusion of homologous templates did not bring significant improvement (for both trRX2 and AF2, unpublished data), which may suggest that the template information is already implied in the MSAs. As a result, structure templates are not used in our monomer structure prediction pipeline.

sequence database was collected manually from various resources, which is searched with the procedure described below.

To generate MSAs from the FASTA database, we adopted a three-step approach. (a) We use the program jackhmmer[14] to search against the FASTA database to detect sequence relatives. Full-length hits are then extracted, forming a smaller database of candidate homologues. (b) The candidate homologues are converted into HMM profile database using the software uniclust.[15] (c) Two MSAs are then generated by the HHblits search against this custom-built HMM profile database with two *e*-value cutoffs (10^{-3}, 1).

2.3 | MSA curation for large targets

We developed a composite modeling approach for large targets, which involves running DISOPRED3[16] to predict disordered residues. We then remove disordered regions and generate an initial MSA (MSA_i) for the new target sequence. An initial model is built with MSA_i, and a two-step approach is introduced to improve the model if it is predicted with low confidence, inspired by the fact that accurate domain-based models could be built after splitting multi-domain targets into domains.

First, we cut the target into domains based on UniDoc,[17] and each domain is submitted to the default pipeline for both MSA generation and structure prediction. The MSAs of the domains are concatenated to construct a full-length MSA (MSA_d), which is combined with the original MSA (MSA_i) to construct the final MSA (MSA_f) in case MSA_i is shallow (i.e., with a limited number of homologous sequences). Second, final models are generated using the MSA_f. High-accuracy domain models are also used as custom templates in this step. This protocol was used to build improved models for a few targets.

2.4 | Monomer structure prediction by Yang-Server

In total, the six MSAs (see Figure 1 for the source) from the MSA search engine are submitted to trX2 and AF2 to predict structure models. A total of 10 models (5 from trX2 and 5 from AF2) are obtained from each MSA. Templates are not used in both trX2 and AF2 as they did not seem to improve the predictions significantly in our benchmark tests (unpublished observation, and the discussions in Section 3.3). Default values are used for other AF2 options (e.g., 'num_recycles'). The 60 models from all MSAs are then ranked based on the quality assessment (QA) scores from two QA methods, DeepUMQA[18] and QDistance.[19] To increase the diversity of the submitted models, the models are clustered into at least five clusters based on their pairwise structural similarity. In each cluster, the model with the highest QA score is selected for submission. Note that the best models are not necessarily selected based on the above protocol, probably because the QA methods are not perfect yet.

2.5 | Multimer structure prediction by Yang-Multimer

In Yang-Multimer, we utilize AlphaFold-Multimer (AFM) to predict the structure of multimers. A few modifications were made on the original version of AlphaFold (v2.0), including the use of custom MSAs and skipping of the MSA pairing step. The MSA pairing was skipped based on our benchmark with AFM. We did not observe significant difference in AFM's modeling results between using and not using MSA pairing, which as supported in a recent study.[20] The MSAs of the monomers are obtained from Yang-Server. For some large targets (e.g. H1111) that surpass the modeling capacity of AFM, the multimer structures were predicted based on superimposition over multimer templates, with predicted monomer structures from Yang-Server.

3 | RESULTS AND DISCUSSION

We mainly compare our predictions with those by the default AF2, which was implemented by the server group NBIS-AF2-standard/NBIS-AF2-multimer. Note that the NBIS-AF2-standard/NBIS-AF2-multimer submissions were run in a quite modest/simple configuration and so do not necessarily represent the best performance that can be achieved with AlpahFold2/AlphaFold-Multimer. In addition, to compare with NBIS-AF2-standard conveniently, the accuracies of the NBIS-AF2-standard models are used to define the target difficulties, which are different with CASP's definitions.

3.1 | Overall results of monomer structure prediction

A total of 132 groups participated in the TS prediction category of the CASP15 experiment. Our group, named Yang-Server, is ranked at the top (the top 10 groups are shown in Figure 2A), according to the default ranking scheme (https://predictioncenter.org/casp15/zscores_final.cgi). As stated in the CASP15 Abstract book, the top 20 groups used either the original AF2 or a retrained version, except for the BAKER group, which is based on an improved version of RoseTTAFold[5] and ranked 13th.

A total of 109 targets were included in the official assessment; but the experimental structures of 15 targets are not available to the predictors. For the 94 targets (shown in Figure 3A), the average TM-score[21] of the submitted models by Yang-Server is 0.876, compared to 0.798 by NBIS-AF2-standard (ranked 44th). There is no significant difference between the Yang-Server and the NBIS-AF2-standard models when the NBIS-AF2-standard models are accurate (TM-score > 0.8), with the exception of the target T1137s4-D3 (0.17 vs. 0.89, the green point in Figure 3A). The low TM-score of the Yang-Server model for this target was due to the removal of the wrongly predicted disordered residues at the C-terminus, which is trivial to fix. However, when the NBIS-AF2-standard models are less accurate

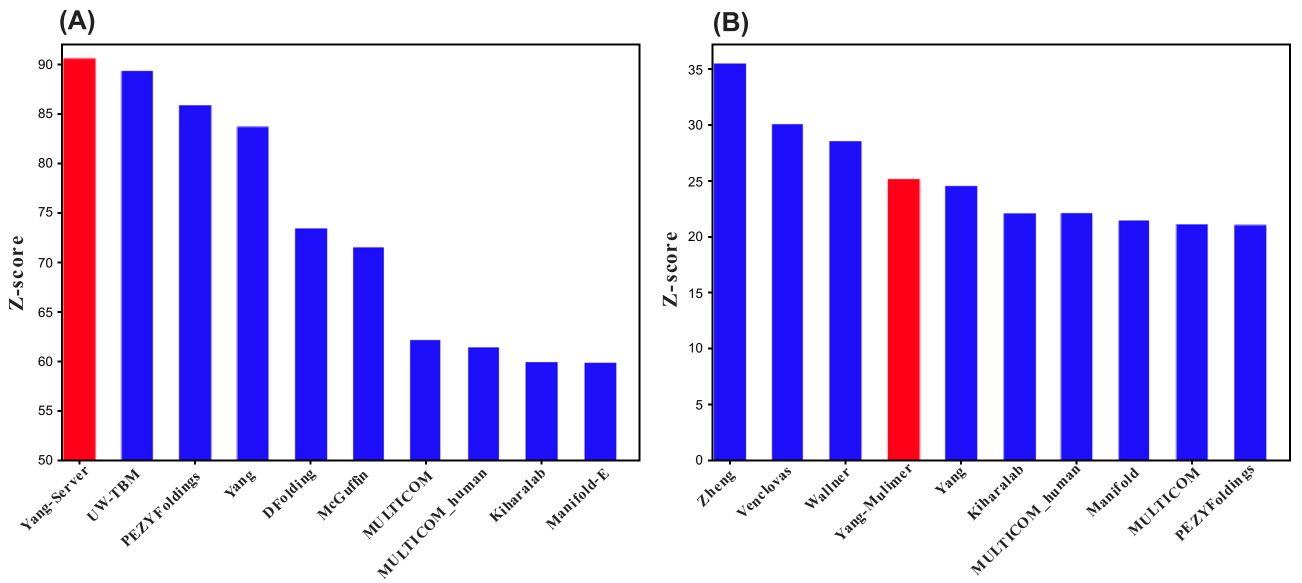


FIGURE 2 Top 10 structure prediction groups for monomers (A) and multimers (B). The ranking scores are taken from the official website of CASP15. Yang-Server and Yang-Multimer are highlighted in red bars.

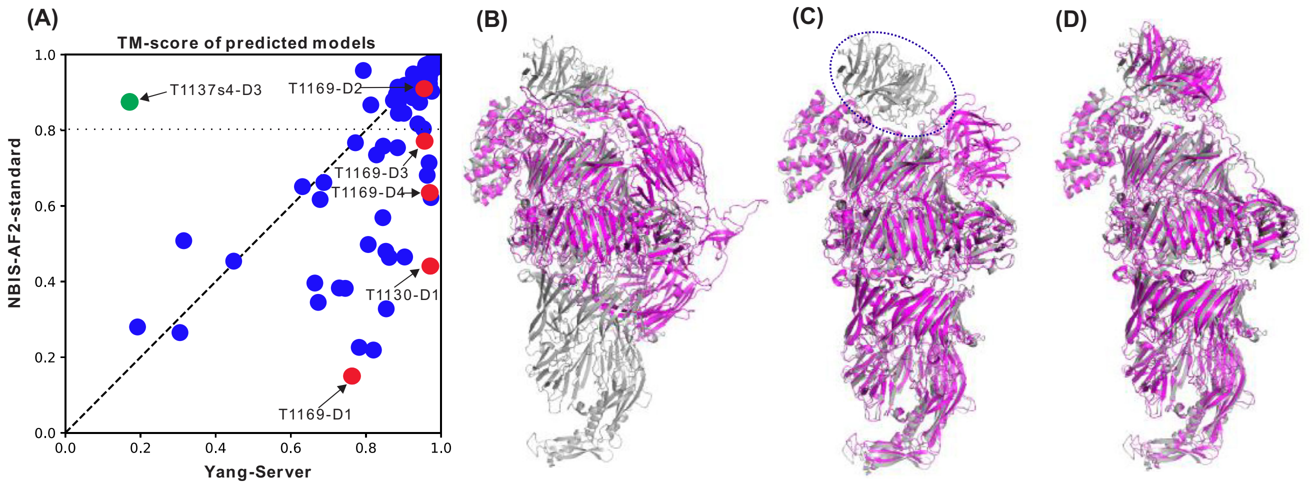


FIGURE 3 Results of monomer structure prediction by Yang-Server and NBIS-AF2-standard. (A) TM-scores of the predicted models for 94 targets. Special targets are highlighted in red/green points. (B–D) The predicted structure models for a large target (T1169) with different methods. (B) NBIS-AF2-standard model. (C) AF2 model predicted locally with the Yang-Server_MSA. The highlighted region is the N-terminal domain of the target (T1169-D1). (D) Yang-Server model. The predicted models and the native structure are shown in magenta and gray cartoons, respectively. Note that only the assessed residues (1–2735) are shown in (B–D).

(TM-score ≤ 0.8), the Yang-Server models are much more accurate than the NBIS-AF2-standard models for most targets. A detailed analysis on the modeling results for a few example targets is given below.

3.2 | What went right?

There are 31 targets which the TM-scores of the NBIS-AF2-standard models are lower than or equal to 0.8. The mean TM-score of the Yang-Server models for these targets is 0.76, which is compared to 0.48 of the NBIS-AF2-standard models. For 27 out of the 31 targets, the Yang-Server models are more accurate than the NBIS-AF2-standard models. A few factors contribute to the improvement, which are explained with a few examples below.

3.2.1 | Improved MSA

For target T1130 (highlighted in red point in Figure 3A), no homologous sequences could be detected from the first two profile databases and the predicted model by AF2/trRX2 has a low accuracy (TM-score of ∼0.4). Other protein language model-based single-sequence folding

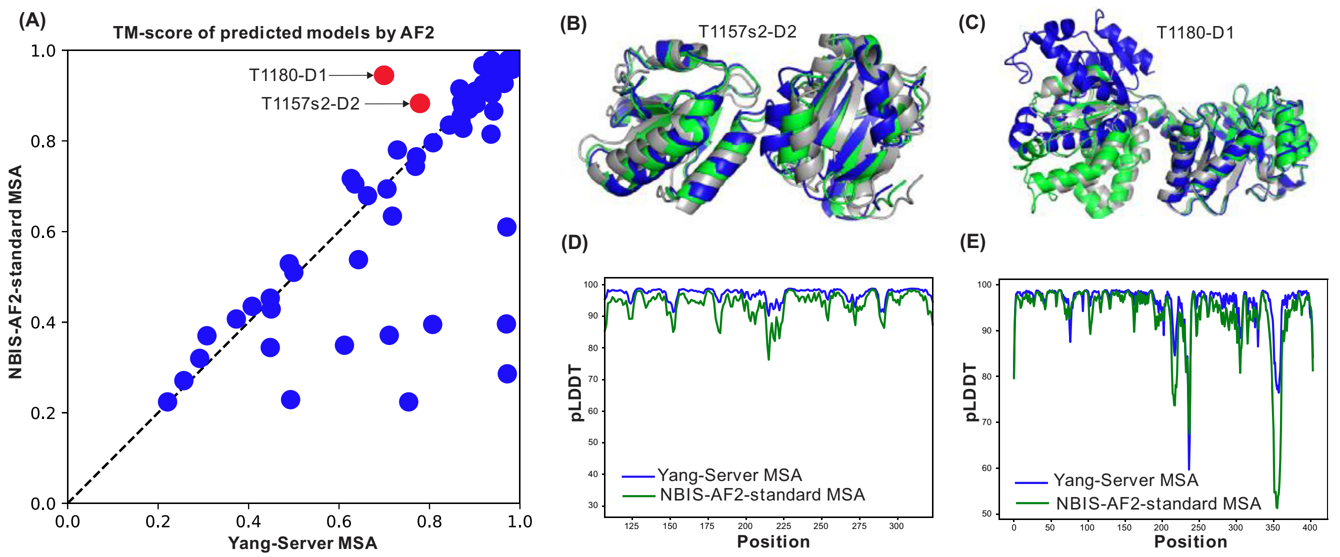


FIGURE 4 AF2 models built locally with two different sets of MSAs. (A) TM-score of the AF2 models by Yang-Server_MSAs and NBIS-AF2-standard_MSAs. (B, C) AF2 models with Yang-Server_MSA and NBIS-AF2-standard_MSA are in blue and green cartoons, respectively. The experimental structure is shown in gray cartoon. (D, E) The local confidence scores (pLDDT) for the predicted models shown in (B, C).

protocols, such as trRosettaX-Single,[22] could not produce reliable models as well. However, 397 homologous sequences were detected from the FASTA sequence database, which resulted in a highly accurate model with a TM-score greater than 0.9 by both AF2 and trRX2.

3.2.2 | A composite modeling approach for large targets

We developed a composite modeling approach for large targets, which involves running DISOPRED3[16] to predict disordered residues. We then remove disordered regions and generate MSAs with the approach described in Section 2.3. This approach works well for a few large targets. For example, T1169 is the largest monomer target in CASP15 with 3364 amino acids, which was proved challenging for the default AF2. This target was cut into four domains in the official assessment (the red points in Figure 3A). The predicted full-length model by NBIS-AF2-standard has a low TM-score of ~0.5 (Figure 3B). After deleting disordered residues at both termini, the size of the target was reduced to 2900, and other regions were predicted correctly resulting in an improved model with a TM-score of ~0.8 (see Figure 3C). With our composite modeling approach, the orientation of the N-terminal domain (i.e., T1169-D1) was fixed, and other regions were also improved. The final model matched the native structure very well (Figure 3D), with a TM-score of 0.95 for the assessed regions (1–2753).

3.3 | What went wrong?

To concentrate on the impact of MSA, we re-ran AF2 on different MSAs with identical configurations. The MSAs for NBIS-AF2-standard were downloaded from the website http://duffman.it.liu.se/casp15/. For each target, the three MSAs (bfd_uniclust_hits, mgnify_hits, and uniref90_hits) are combined and filtered to remove identical sequences by HHblits (with options '-id 99 -cov 0.1'). The processed MSAs (denoted by NBIS-AF2-standard_MSA) are available together with our MSAs (denoted by Yang-Server_MSA) at the trRosetta server: https://yanglab.nankai.edu.cn/trRosetta/benchmark/.

It turns out that Yang-Server_MSAs yield more accurate models than NBIS-AF2-standard_MSAs for most targets. The average TM-score of the predicted models by AF2 is 0.832 and 0.796 for Yang-Server_MSAs and NBIS-AF2-standard_MSAs, respectively (see Figure 4A). However, there are a few targets (e.g., T1157s2-D2 and T1180-D1, highlighted in red points in Figure 4A) that the Yang-Server_MSAs result in less accurate models than the NBIS-AF2-standard_MSAs (Figure 4B,C). Though the pLDDTs (Figure 4D,E) of the models from the Yang-Server_MSAs are slightly higher than those from the NBIS-AF2-standard_MSAs, the models by the former are less accurate than the latter (TM-score: 0.779 vs. 0.883 for T1157s2-D2; 0.7 vs. 0.954 for T1180-D1). Note that both targets contain two domains, and the domain-level TM-scores for individual domain models are similar for both MSAs. This reflects the issue of our strategy in MSA curation for multi-domain targets, which will be fixed in the future.

Based on the above observation, we would like to make a concluding remark on MSA. There is no single way to generate optimal MSAs for all targets. A researcher should be encouraged to try various MSA generation methods (as done in this work). Then an unsolved issue is how to effectively select the optimal MSA from several candidate MSAs, especially for multi-domain targets, which deserves more attention in future.

In addition, the models generated from the NBIS-AF2-standard_MSAs locally (without templates) are similar to the submitted

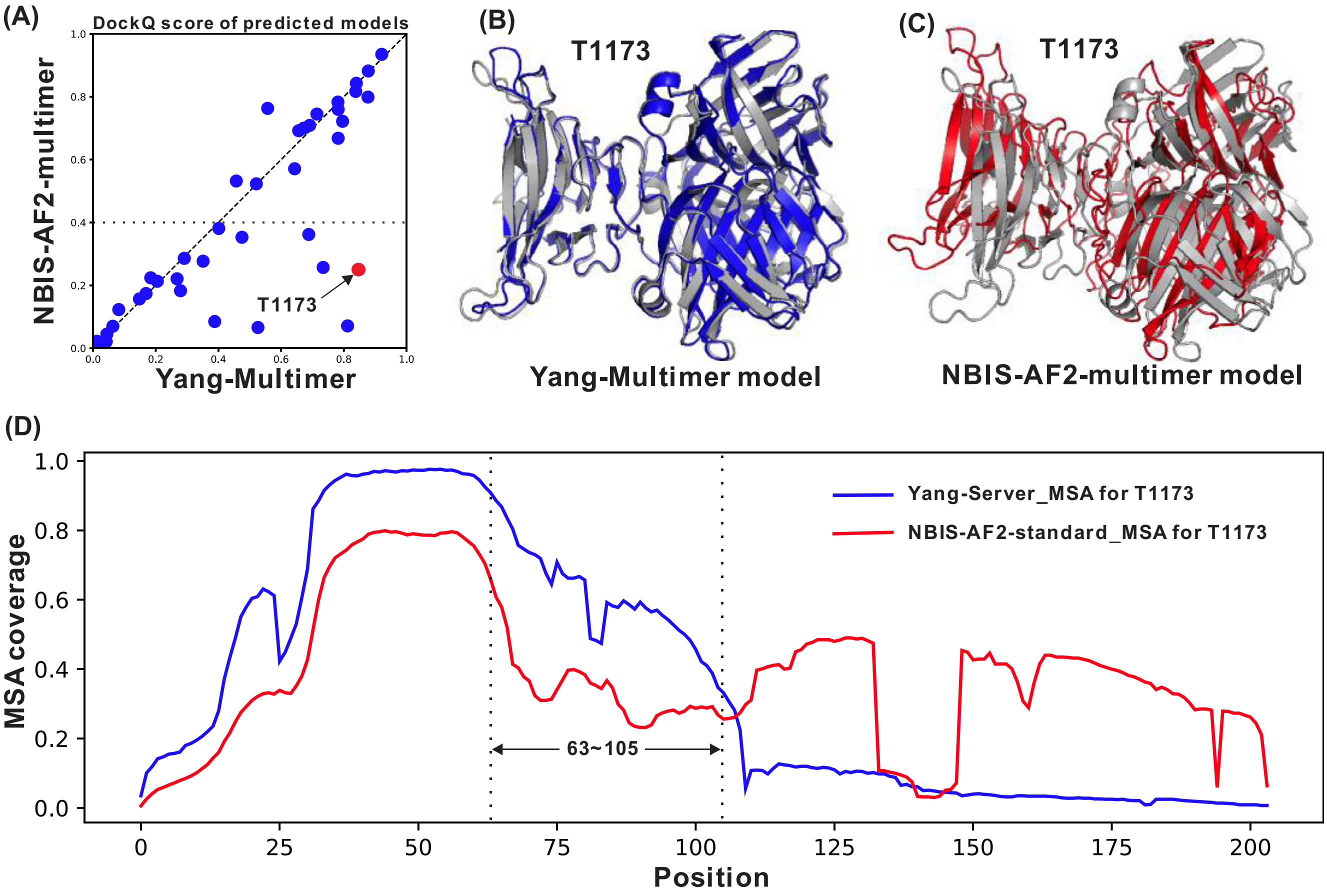


FIGURE 5 Results of the multimer structure prediction by Yang-Multimer and NBIS-AF2-multimer. (A) DockQ scores of the predicted models for 42 multimeric targets. (B)/(C) The predicted models superimposed onto native structure (native structure is in gray cartoon, while predicted models are in blue/red cartoon) for the homo-trimmer target T1173. (D) MSA coverage for the target T1173.

NBIS-AF2-standard models (with templates), as indicated by the average TM-score of 0.796 versus 0.798. This suggests that the inclusion of templates does not make a big difference overall, though it may be helpful for certain targets. For example, for the above target T1180-D1, the domain orientation could be fixed by including templates in the modeling (when using the Yang-Server_MSA). However, the use of templates might hurt as well (e.g., limit the sampling of conformational space), especially for targets with multiple conformations. As a result, we prefer not using templates during the CASP season.

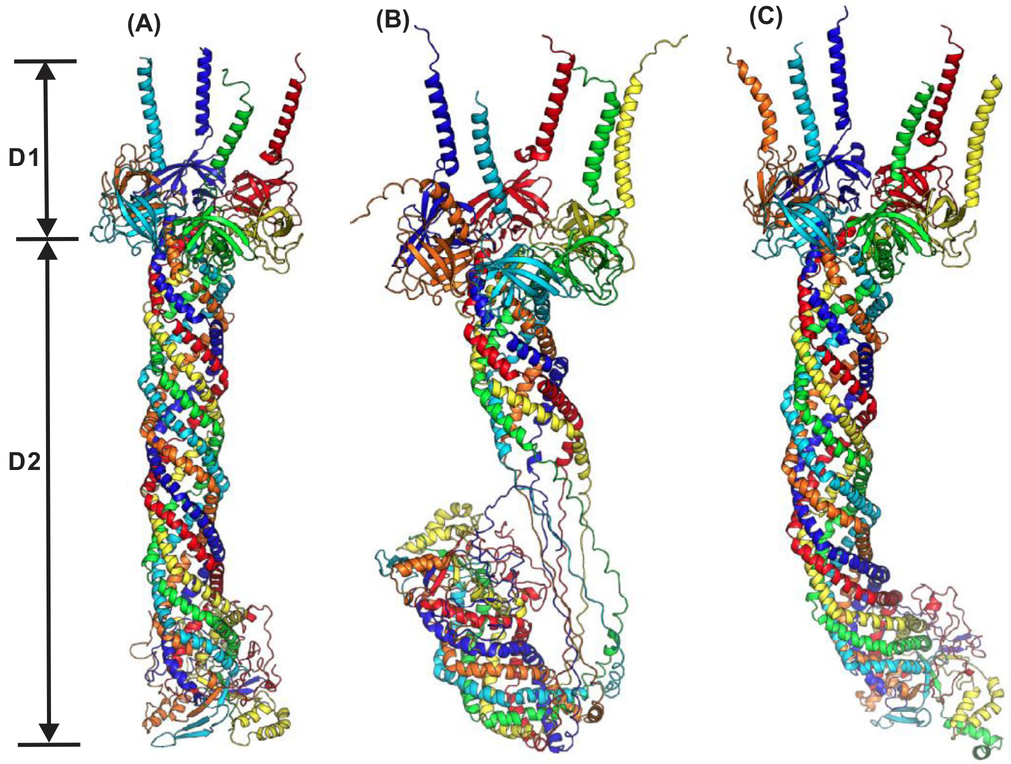
3.4 | Results of multimer structure prediction

Our multimer structure prediction method, Yang-Multimer, predicts the multimer structure based on AFM with the input of MSAs generated by Yang-Server. A total of 43 targets are included in the official assessment (Yang-Multimer submitted models for 42 targets). Overall, Yang-Multimer is ranked fourth among 87 participating groups (the top 10 groups are shown in Figure 2B), outperforming NBIS-AF2-multimer (ranked 31st) significantly.

When the DockQ scores of the NBIS-AF2-multimer models are higher than 0.4, there is no significant difference between our models and NBIS-AF2-multimer models (Figure 5A). For 11 targets, the DockQ scores[23] for the predicted models by both Yang-Multimer and NBIS-AF2-multimer are less than 0.2. This indicates that multimer structure prediction is in general more difficult than monomers.[24] Progress is expected to be made in this area in the future. Nevertheless, Yang-Multimer improves the modeling for a few challenging targets that NBIS-AF2-multimer does not work well, which is explained with a few examples below.

3.4.1 | Improved MSA

The target T1173 is a homo-trimer and each chain consists of two domains: T1173-D1 (1–62) and T1173-D2 (63–204). The NBIS-AF2-standard MSA (2702 sequences) contains more homologous sequences than the Yang-Server MSA (1743 sequences, from the FASTA database). For the first domain, the predicted structure models are all accurate with both MSAs. However, for residues 63–105 in the



(D)

Interface	A-B	B-C	C-D	D-E	E-F	F-A	Average
NBIS-AF2-multimer	0.289	0.25	0.189	0.267	0.197	0.197	0.243
Yang-Multimer	0.37	0.418	0.369	0.455	0.409	0.434	0.409

FIGURE 6 Comparison between the Yang-Multimer and NBIS-AF2-multimer models for the first six chains of the target H1137. (A–C) The experimental structure, the Yang-Multimer model and the NBIS-AF2-multimer model, respectively. Chains (A–F) are indicated by different colors. Table (D) lists the DockQ scores of the predicted interfaces between different chains.

second domain, the coverage of the NBIS-AF2-standard_MSA is lower than the Yang-Server_MSA (Figure 5D). These residues are the linker between both domains (inside each chain) and are located at the interface between chains. A high-quality alignment for these residues is necessary to generate models with high accuracy. It turns out that the Yang-Multimer model (Figure 5B) are significantly more accurate than the NBIS-AF2-multimer model (Figure 5C) for this target (DockQ score 0.847 vs. 0.25).

3.4.2 | Iterative modeling

The target H1137 is an *ATP-binding cassette transporter complex* consisting of 10 chains, with the first six chains (s1–s6, Figure 6A) forming a hetero hexamer with intertwined interactions in the second domain (D2). The NBIS-AF2-multimer model has low accuracy, especially for the residues in D2 (Figure 6B), probably due to the large size of this target (2731 amino acids). On the other hand, when predicting the structures separately for individual chains with AF2, only the first domains (D1) can be predicted correctly, and the orientations of the second domains are wrong.

To address the above issue, we predicted the structure for H1137 based on an iterative modeling. In the first iteration, a hexamer model for the domains D2 was predicted with AFM (after removing predicted disordered residues in chains C, D, and F). In the second iteration, the hexamer structure for both D1 and D2 was predicted by AFM with the D2 models as custom templates. With this procedure, the models for chains in D2 were improved dramatically (TM-score increases by 0.4–0.5). For the whole hexamer structure, the average DockQ score of the model predicted by Yang-Multimer is >0.4, compared to 0.243 by NBIS-AF2-multimer.

There are several other factors that may improve the structure prediction for multimers, such as more sampling for antibody structure prediction (H1140–H1144) with the option '-num_multimer_predictions_per_model', specifying monomer models (after removing low-confidence regions) as custom templates (H1129) by adding them into the template library, and the use of homologous templates for larger targets (H1111), etc. Details for these examples are skipped here.

4 | CONCLUSIONS

The recent CASP15 experiments provide a way to track progress in protein structure prediction over the past 2 years. We have conducted an in-depth analysis of the modeling results made by our group (Yang-Server and Yang-Multimer). Overall, the predicted models by Yang-Server is about 10% more accurate than AF2. For easy targets, highly accurate structure models can be built with most of the existing methods. For hard targets, it is beneficial to generate high-quality MSAs with complementary algorithms and sequence databases. However, consistently predicting multimer structures with high accuracy remains a challenge, and breakthroughs in this area are anticipated in the future.

AUTHOR CONTRIBUTIONS

Zhenling Peng: Investigation; methodology; writing – original draft; writing – review and editing; supervision; formal analysis. **Wenkai Wang:** Software; writing – review and editing; data curation. **Hong Wei:** Data curation; writing – review and editing. **Xiaoge Li:** Investigation; writing – review and editing; data curation. **Jianyi Yang:** Conceptualization; methodology; writing – review and editing; writing – original draft; project administration; supervision; formal analysis.

ACKNOWLEDGMENTS

We would like to thank the CASP15 organizers and assessors for the great efforts in this CASP experiments. This work is supported in part by the National Natural Science Foundation of China (NSFC T2225007 and T2222012).

DATA AVAILABILITY STATEMENT

The final MSAs for the CASP15 targets used by our group are available for download at the trRosetta server. The full pipeline we used in CASP15 will be available through the trRosetta server.

REFERENCES

1. Peng Z, Wang W, Han R, Zhang F, Yang J. Protein structure prediction in the deep learning era. *Curr Opin Struct Biol*. 2022;77:102495.
2. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*. 2015;43(W1): W174-W181.
3. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using Rosetta. *Methods Enzymol*. 2004;383:66-93.
4. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589.
5. Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. 2021;373(6557):871-876.
6. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci*. 2020;117(3):1496-1503.
7. Baek M, Anishchenko I, Park H, Humphreys IR, Baker D. Protein oligomer modeling guided by predicted interchain contacts in CASP14. *Proteins*. 2021;89(12):1824-1833.
8. Evans R, O'Neill M, Pritzel A, et al. Protein complex prediction with AlphaFold-Multimer. *BioRxiv*. 2021.
9. Su H, Wang W, du Z, et al. Improved protein structure prediction using a new multi-scale network and homologous templates. *Adv Sci*. 2021;8:e2102592.
10. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 2012;9(2):173-175.
11. Wu Q, Peng Z, Anishchenko I, Cong Q, Baker D, Yang J. Protein contact prediction using metagenome sequence data and residual neural networks. *Bioinformatics*. 2020;36(1):41-48.
12. Du Z, Peng Z, Yang J. Toward the assessment of predicted interresidue distance. *Bioinformatics*. 2022;38(4):962-969.
13. Steinegger M, Soding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol*. 2017; 35(11):1026-1028.
14. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res*. 2013;41(12):e121.
15. Mirdita M, von den Driesch L, Galiez C, Martin MJ, Söding J, Steinegger M. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res*. 2017;45(D1): D170-D176.
16. Jones DT, Cozzetto D. DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics*. 2015; 31(6):857-863.
17. Zhu K, Su H, Peng Z, Yang J. A unified approach to protein domain parsing with inter-residue distance matrix. *Bioinformatics*. 2023;39(2): btad070.
18. Guo SS, Liu J, Zhou XG, Zhang GJ. DeepUMQA: ultrafast shape recognition-based protein model quality assessment using deep learning. *Bioinformatics*. 2022;38(7):1895-1903.
19. Ye LS, Wu P, Peng Z, Gao J, Liu J, Yang J. Improved estimation of model quality using predicted inter-residue distance. *Bioinformatics*. 2021;37(21):3752-3759.
20. Gao M, Nakajima An D, Parks JM, Skolnick J. AF2Complex predicts direct physical interactions in multimeric proteins with deep learning. *Nat Commun*. 2022;13(1):1744.
21. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004;57(4):702-710.
22. Wang WK, Peng ZL, Yang JY. Single-sequence protein structure prediction using supervised transformer protein language models. Nature computational. *Science*. 2022;2(12):804.
23. Basu S, Wallner B. DockQ: a quality measure for protein-protein docking models. *PLoS One*. 2016;11(8):e0161879.
24. Callaway E. After AlphaFold: protein-folding contest seeks next big breakthrough. *Nature*. 2023;613(7942):13-14.

How to cite this article: Peng Z, Wang W, Wei H, Li X, Yang J. Improved protein structure prediction with trRosettaX2, AlphaFold2, and optimized MSAs in CASP15. *Proteins*. 2023; 91(12):1704-1711. doi:10.1002/prot.26570