Proteins: Structure, Function, and Bioinformatics

WILEY

PROTEINS

RESEARCH ARTICLE

Accurate Biomolecular Structure Prediction in CASP16 With Optimized Inputs to State-Of-The-Art Predictors

Wenkai Wang | Yuxian Luo | Zhenling Peng | Jianyi Yang

MOE Frontiers Science Center for Nonlinear Expectations, State Key Laboratory of Cryptography and Digital Economy Security, Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao, China

Correspondence: Zhenling Peng (zhenling@email.sdu.edu.cn) | Jianyi Yang (yangjy@sdu.edu.cn)

Received: 3 April 2025 | **Revised:** 8 July 2025 | **Accepted:** 24 July 2025

Funding: This work was supported by the National Natural Science Foundation of China.

Keywords: CASP16 | protein structure prediction | protein–protein complex | RNA structure prediction

ABSTRACT

Biomolecular structure prediction has reached an unprecedented level of accuracy, partly attributed to the use of advanced deep learning algorithms. We participated in the CASP16 experiments across the categories of protein domains, protein multimers, and RNA monomers, achieving official rankings of first, second, and fourth (top for server groups), respectively. We hypothesized that by leveraging state-of-the-art structure predictors such as AlphaFold2, AlphaFold3, trRosettaX2, and trRosettaRNA2, accurate structure predictions could be achieved through careful optimization of input information. For protein structure prediction, we enhanced the input sequences by removing intrinsically disordered regions, a simple yet effective approach that yielded accurate models for protein domains. However, fewer than 25% of the protein multimers were predicted with high quality. In RNA structure prediction, optimizing the secondary structure input for trRosettaRNA2 resulted in more accurate predictions than AlphaFold3. In summary, our prediction results in CASP16 indicate that protein domain structure prediction has achieved high accuracy. However, predicting protein multimers and RNA structures remains challenging, and we anticipate new advancements in these areas in the coming years.

1 | Introduction

Protein structure prediction has been transformed by the Nobel Prize-winning method AlphaFold2 (AF2) [1]. It is widely accepted that the challenge of single-domain protein structure prediction has been resolved.

However, predicting accurate models for certain types of protein monomers remains challenging, particularly in several specific cases. First, proteins with multiple domains present complexities in predicting domain orientations, even when individual domains are accurately modeled [2]. Second, loop regions often exhibit greater flexibility compared to structured α-helices or β-strands, complicating modeling efforts [3]. Third, underrepresented proteins, such as those from viruses, typically lack sufficient homologous sequences in current databases, resulting in shallow multiple sequence alignments (MSAs) and weaker co-evolution signals, which hinder model accuracy. Fourth, interactions with binding partners must be considered for accurate modeling, as demonstrated by the CASP16 target M1271. Lastly, proteins that adopt alternative conformations or exhibit dynamic structures pose additional challenges for accurate predictions [4].

Another unresolved challenge is the prediction of multimeric protein structures. There are two primary approaches for predicting multimeric structures. The first approach is based on classical molecular docking, which aims to assemble subunit

Wenkai Wang and Yuxian Luo co first authors.

structures by considering the physical and chemical complementarity between different subunits. Notable methods in this category include HDOCK [5] and ClusPro [6], which utilize fast Fourier transforms to expedite the search for optimal binding poses. However, this approach struggles to accommodate substantial conformational changes induced by binding. To address this limitation, the second approach focuses on co-folding, which predicts complex structures directly from protein sequences. Representative methods in this category include AlphaFold-Multimer (AFM) [7], AlphaFold3 (AF3) [8], and RoseTTAFold All-Atom (RFAA) [9]. Notably, AF3 and RFAA are applicable to a wide range of complexes, including protein–protein, protein–nucleic acid, and protein–small molecule interactions. Although AF3 can generate complex models with improved accuracy, accurately modeling a significant portion of targets, such as antibody–antigen interactions, remains challenging. Consistently predicting complex structures with high accuracy is still an open problem.

Inspired by the success of AF2, several deep learning-based approaches have emerged for predicting RNA 3D structures. Notable methods include trRosettaRNA [10], DRfold [11], RhoFold+ [12], DeepFoldRNA [13], RFAA [9], NuFold [14], RoseTTAFoldNA [15], and AF3 [8], which outperform traditional physics-based methods for automated structure prediction. However, challenges persist due to the sparsity of experimental structures, the high flexibility and complexity of RNA structures, and difficulties in nucleotide sequence alignment. As demonstrated by the CASP [16] and RNA-Puzzles experiments [17], human intervention plays a crucial role in generating accurate models. Significant progress is still needed for automated and accurate RNA structure prediction.

We participated in the CASP16 experiments for both protein and RNA structure predictions. Our group achieved top rankings: first for protein domains, second for protein multimers, and fourth (top server group) for RNA monomers. This paper outlines the main methodologies used and summarizes the key results from CASP16.

2 | Methods for Protein Structure Prediction

2.1 | Overall Strategy for Protein Structure Prediction

The protein structure prediction pipeline we employed in CASP16 consists of three main modules: sequence optimization, structure prediction, and model selection (see Figure 1a). The computational workflow begins with sequence optimization, where intrinsically disordered regions (IDRs) are predicted and removed from both the original query sequences and their corresponding multiple sequence alignments (MSAs) to improve structural modeling reliability. The structure prediction module then generates structural models using trRosettaX2 (trX2) [4], AlphaFold2 (AF2), and AlphaFold3 (AF3); both trX2 and AF2 are run locally with the optimized sequences and MSAs, while AF3 is executed via its online server with the optimized sequences. Note that no structure refinement was applied. Finally, the model selection module ranks and clusters the predicted models to prepare them for final submission.

2.2 | Sequence Optimization Module

This module focuses on optimizing three key elements at the sequence level: amino acid sequences, multiple sequence alignments (MSAs), and stoichiometry.

2.3 | Amino Acid Sequence Optimization

Predicted IDRs are removed from the query sequences based on three considerations: IDRs are often unresolved in experimental determinations; IDRs are predicted with low accuracy and may hinder high-accuracy modeling of other structured regions; some targets are too large to be submitted to the AF3 server.

Specifically, we utilized DISOPRED3 [18] to predict IDRs. Removing IDRs at the N- or C-terminus does not impact subsequent structure modeling. However, removing intermediate IDRs (e.g., between residues i and j) can introduce errors, as residues $i-1$ and $j+1$ may be incorrectly connected. For trX2 and AF3, only terminal IDRs are removed. For AF2, we addressed this issue by modifying the *residue_index* feature in AF2's feature dictionary, ensuring that each residue in the optimized sequences retained its original numbering. This modification allows AF2 to handle the structure modeling correctly.

Please note that we did not attempt to reintroduce the removed residues into our submitted models, as these residues are frequently unresolved in experimental structures unless their inclusion could result in more accurate models. For example, the experimental structure of the protein-RNA target M1271 includes only 3276 out of a total of 5990 amino acids and nucleotides. However, when experimental structures partially cover the removed residues, our models may incur penalties (e.g., in H0258). This issue could be addressed by designating the incomplete models as templates in full-length modeling, potentially in future CASPs. However, specifying multimeric templates is currently not supported for AFM/AF3 and other methods, for which a solution is definitely invaluable.

2.4 | MSA Optimization

The MSA construction process is largely consistent with our approach from CASP15 [2]. We searched multiple sequence databases (uniclust30_2018, uniref30, BFD, and manually collected sequences) using HHblits [19], MMseqs2 [20], and JackHMMER [21], generating candidate MSAs for the original sequences. Optimal MSAs were selected by trX2 based on the probability of the top residue pairs in the predicted inter-residue distance matrix. Columns in the MSAs corresponding to predicted IDRs were removed. Notably, all MSAs were generated from the original sequences; generating MSAs from IDR-removed sequences did not yield improvements based on our experience.

2.5 | Stoichiometry Optimization

The stoichiometry for multimeric targets in Phase 0 is inferred using two approaches. The first one is a template-based

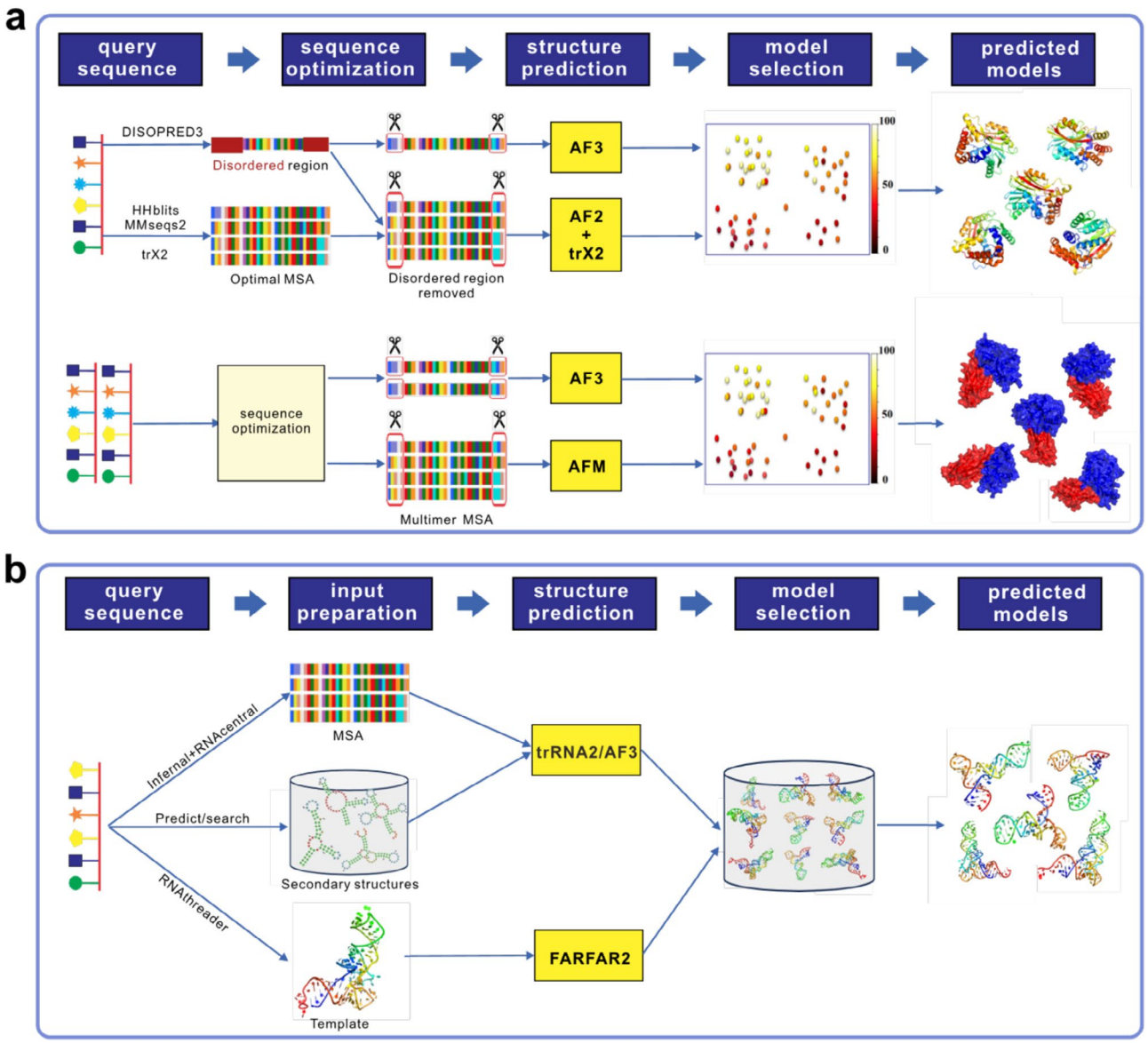

FIGURE 1 | Structure prediction pipelines used in CASP16. (a) Protein structure prediction consists of three modules: Sequence optimization, structure prediction, and model selection. The top and bottom flowcharts in (a) are for monomeric and multimeric protein structure predictions, respectively. (b) RNA structure prediction includes input preparation, structure prediction, and model selection.

approach, where the stoichiometry of a homologous template is transferred to the query target [22]. If no homologous templates are available, we employ an AF3-based trial-and-error method, submitting various subunit combinations to the AF3 server and assigning the stoichiometry that yields the highest-confidence structure model.

2.6 | Structure Prediction Module

2.6.1 | Monomer Structure Prediction

Monomer structures are predicted using trX2, AF2, and AF3. Both trX2 and AF2 are run locally with the optimized sequences and MSAs from the previous module. Since no open-source code for AF3 was available during the CASP16 experiment, we submitted the optimized sequences to the AF3 server to obtain predicted models. If a monomeric target originates from a multimeric target, monomeric models are also derived from those multimeric models.

2.6.2 | Multimer Structure Prediction

Multimer structures are predicted using the AF3 server and AFM. The optimized sequences were submitted to the AF3 server with the default random seed to predict multimeric models. If the confidence score is low (<0.7), we attempt up to five additional random seeds and select the model with the highest confidence score. AFM is run locally using the optimized sequences and MSAs from the previous module. For hetero multimers, the MSA pairing in AFM is toggled based on whether the subunits come from the same or different species, in accordance with our recent study.

2.7 | Model Selection Module

Predicted structure models are clustered based on pairwise similarities: TM-score for monomers and DockQ for multimers. Representative models are selected from the clusters for submission based on their confidence scores: pLDDT for monomers and pTM + ipTM for multimers. The primary distinctions among the groups, Yang-Server, Yang-Multimer, and Yang, lie in this module: models with the highest confidence scores are submitted by the server groups Yang-Server and Yang-Multimer for monomers and multimers, respectively; model rankings are manually reviewed and adjusted by the human group Yang. If the stoichiometry prediction in Phase 0 was incorrect, none of the models generated during this phase were utilized in Phase 1. Conversely, if the prediction was correct, the Phase 0 models were incorporated into the pool of models generated in Phase 1 prior to executing the model selection module.

3 | Methods for RNA Structure Prediction

3.1 | Overview About trRosettaRNA2 Algorithm

We developed trRosettaRNA2 (trRNA2) [23], an enhanced version of trRosettaRNA [10], to directly predict RNA 3D structures from multiple sequence alignments (MSAs) in an end-to-end manner. Considering that the limited RNA 3D structure data may hinder the performance of deep learning-based methods, trRNA2 incorporates a dedicated secondary structure prediction (SS) module. This module is pre-trained on a larger SS dataset (bpRNA [24]) to provide prior base-pairing information. The output of the SS module, along with the MSA embedding, is then fed into 12 RNAformer blocks [10] to update RNA representations and predict 2D geometry. Finally, the 3D structure is generated either by an SS-aware structure module utilizing the updated representations or through energy minimization based on the predicted 2D geometry.

trRNA2 effectively improves RNA 3D structure prediction accuracy compared to trRosettaRNA and achieves competitive performance with AF3, though requiring significantly fewer parameters and computational resources. A comprehensive description of the methodology and the complete benchmark results will be available in a forthcoming publication [23].

3.2 | RNA Structure Prediction by Yang-Server in CASP16

As shown in Figure 1b, the RNA structure modeling in Yang-Server was primarily powered by trRNA2 and RNAthreader [25], a template-based method developed in our lab. For each RNA target, we first predicted the structure using trRNA2. The MSA was generated by Infernal [26] searching against the RNAcentral database [27]. To enhance the diversity of our predictions, we also utilized secondary structures generated by several third-party tools, including SPOT-RNA [28], EternaFold [29], and the template-based method R2DT [30], as well as RNAthreader from our lab, as alternative inputs to trRNA2, in addition to its default prediction. Furthermore, for targets with available 3D templates, we performed template-based modeling using RNAthreader and FARFAR2 [31]. The resulting 3D structures from trRNA2 and RNAthreader (when applicable) were then assessed and ranked using an in-house deep learning-based RNA quality assessment (QA) method (Liu et al. [22]). The top 5 ranked models were submitted as the Yang-Server predictions. For larger RNAs exceeding 400 nucleotides and without templates, additional models from AF3 were included (R1241, R1248, R1250-R1254, R1283, R1290). Similar to proteins, nucleotide sequences for these large targets were manually submitted to the AF3 web server to generate predictions.

4 | Results for Protein Structure Prediction

4.1 | Overall Results of Protein Domains

According to the official assessment, among the 110 groups involved in protein domain structure prediction, Yang-Server ranks first (Figure 2a), while AF3-server is ranked 15th. A direct GDT-TS comparison between Yang-Server and AF3-server is illustrated in Figure 2b (data available in Table S1). Yang-Server outperforms AF3-server for more than half of the targets, confirming the effectiveness of our sequence optimization. The average GDT-TS scores for Yang-Server and AF3-server are 87.61 and 85.28, respectively, indicating that protein domain structure prediction has achieved a high level of accuracy.

However, two hard targets (T1226-D1 and T1207-D1) were modeled with GDT-TS scores below 60 for both Yang-Server and AF3-server. The target T1226-D1 is a small protein (123 residues) from *Turkey Avisivirus*, which was solved by NMR but with poor quality [32]. The accuracy of the Yang-Server model1 is low (GDT-TS 32.79, Figure S1a), while the lower-ranked model, model2, demonstrates significantly higher accuracy (GDT-TS 67.01, Figure S1b), indicating that our model selection scheme still requires improvement.

The target T1207-D1 is a truncated domain consisting of 144 residues (positions 776–919) from a viral polyprotein (Q9YLS4, 2134 residues). The primary modeling challenge lies in the 20 C-terminal residues, which were inaccurately predicted, resulting in a model of medium accuracy (Yang-Server model1 GDT-TS 58.20, Figure S1c). This truncation may hinder the correct modeling of this target. To address this, we extended the model by including an additional 136 residues at the C-terminus (positions 776–1056), enabling the generation of more accurate models. Screening the experimental structure against the pool of structural models revealed that accurate models (with GDT-TS scores over 90) were generated. Unfortunately, the model selection process did not rank these models at the top. Model 5 submitted by Yang-Multimer is one such model, achieving a GDT-TS score of 85.66 (Figure S1d).

4.2 | Overall Results of Protein Multimers

In the official assessment (Figure 2c), our group, Yang-Multimer, is ranked second among 82 groups, with a *Z*-score slightly lower than that of the top group, KiharaLab (14.5 vs. 15.4). The AF3-server is ranked ninth. A direct DockQ comparison of the models predicted by Yang-Multimer and AF3-server is presented

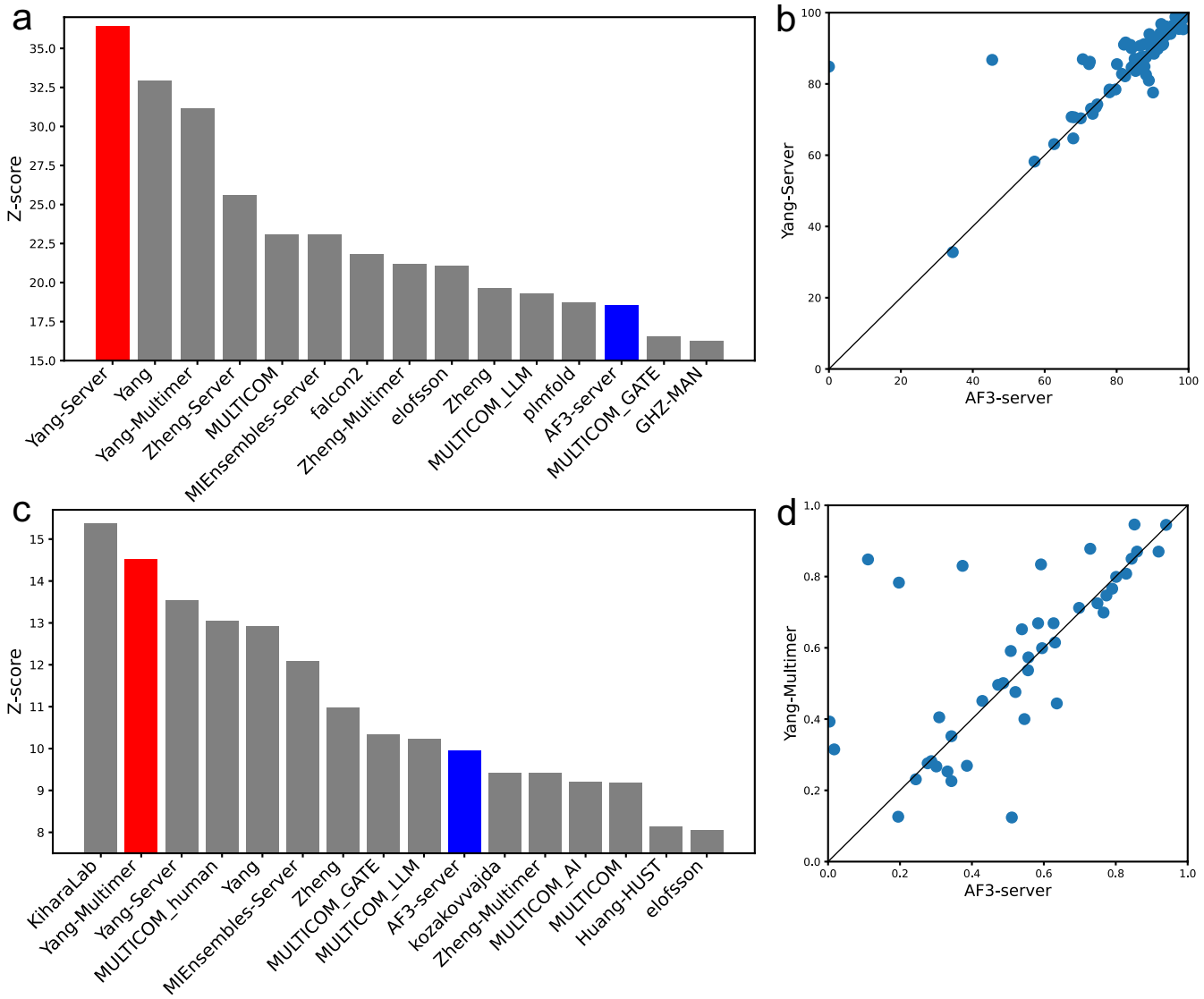

FIGURE 2 | Overall results of the protein structure prediction in the official assessment. (a, c) sum (Z-score > −2.0) of the predicted monomer/multimer models by the top 15 groups. The ranking scores are taken from the official website of CASP16. Yang-Server/Yang-Multimer and AF3-server are highlighted with red and blue bars, respectively. (b, d) Head-to-head GDT-TS/DockQ comparison between Yang-Server/Yang-Multimer and AF3-server for monomer/multimer structure prediction. Note that only Phase 1 models were considered here to reduce the impact of other factors.

in Figure 2d (data available in Table S2). With the exception of two special targets, H1215 (nanobody) and T1249v2o (conformational change), Yang-Multimer predicted other targets with at least acceptable quality (DockQ ≥ 0.23), compared to five by AF3-server. However, only 10 and 7 out of the 44 multimeric targets were predicted with high quality (DockQ ≥ 0.8) by Yang-Multimer and AF3-server, respectively, suggesting that protein multimer structure prediction remains a challenging task. Below, we provide a few examples to illustrate both successes and shortcomings.

4.3 | What Went Right?

The first example is the protein-RNA complex M1271, which comprises 16 protein subunits (10 unique subunits, denoted as s1–s10) and 1 RNA subunit (Figure 3a). The modeling of the entire complex target exceeds the capabilities of AF2 and trRosettaX2. Therefore, we utilized AF3 to predict the structures of its protein-RNA complex by optimizing the protein sequences. The optimized input sequences used to reproduce our model for this target can be found in Table S3.

The total number of tokens in this target is 5990, exceeding the AF3 server's limit of 5000. Since inter-subunit interactions are crucial for maintaining the conformation of certain subunits in this target, predicting the structure of each subunit separately is unlikely to yield accurate models. Additionally, assembling subunit models into complex models based on docking remains challenging, particularly in cases of induced fit.

The sequence optimization module allowed for the prediction of a reasonable complex structure model for M1271 using AF3 (Figure 3b). Sequence-based disorder prediction indicated

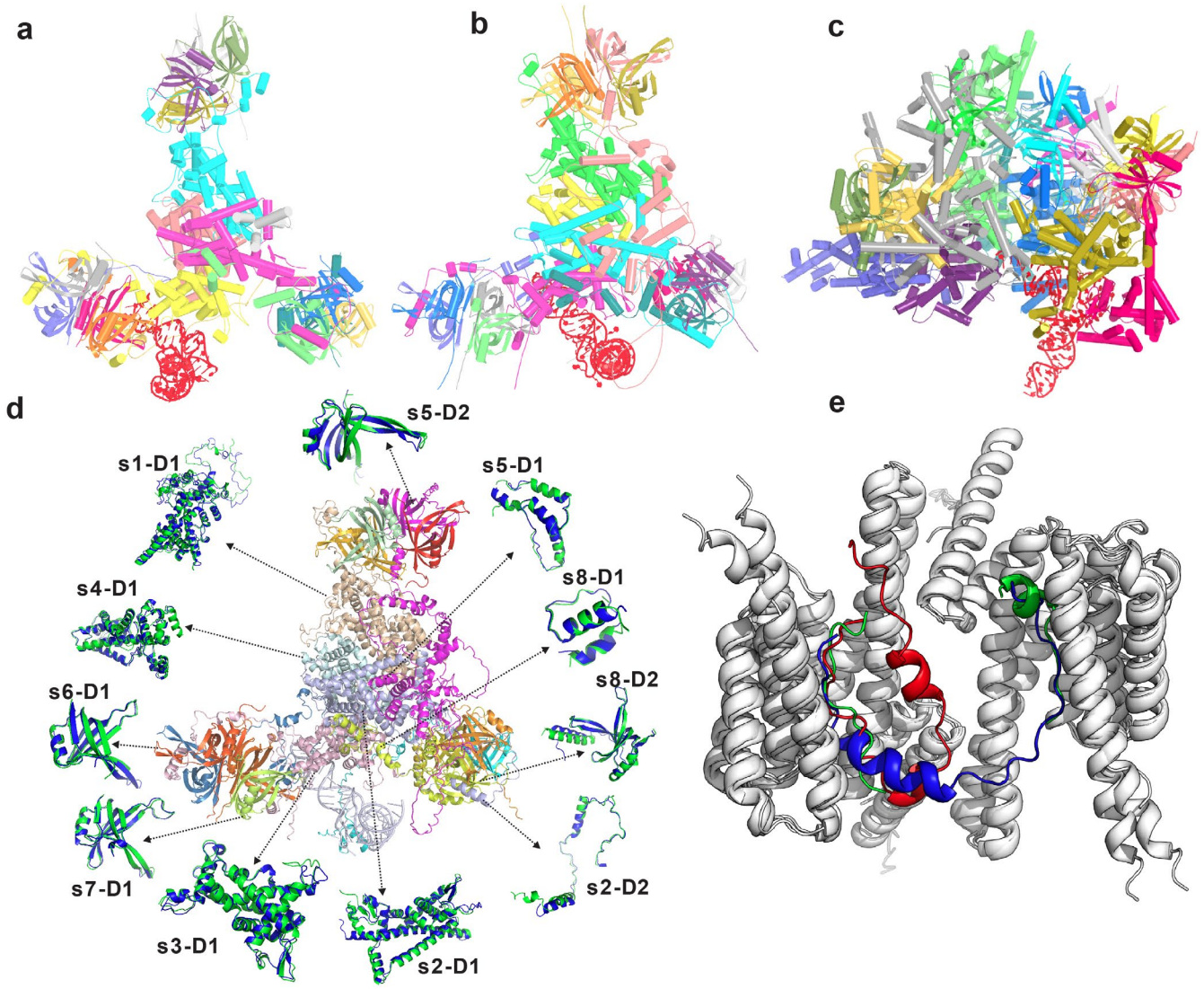

FIGURE 3 | Structure modeling for M1271 and H1258. (a) the experimental structure of M1271. (b) the predicted structural model of M1271 by Yang using optimized sequences via the AF3 server. (c) the predicted structural model of M1271 by AF3, using its original sequences (ran locally after CASP16 as the AF3 server has a limit of 5000 tokens). (d) a detailed analysis of the model in (b). The central structure is colored by chain; for the subunits indicated by arrows, the blue and green cartoons represent the predicted and experimental structures, respectively. (e) the predicted structural models for H1258, by Yang-Multimer and the group AF3-server, focusing on the region (942–977) of the *LRRK2* subunit that interacts with the 14-3-3 subunits (in gray cartoon). The experimental structure is shown in green, while the predicted models are represented in blue (Yang-Multimer) and red (AF3 server).

that many regions in subunits s1, s2, and s4–s7 are disordered. After conducting trial-and-error tests with the AF3 server, a total of 2283 disordered residues were removed from the original target, resulting in an input FASTA sequence with 3707 tokens, which aligns well with the number of tokens (3276) in the experimental structure. The confidence score for the predicted model is approximately 0.6, indicating medium accuracy. According to the official assessment, the global TM-score of our model is 0.732, significantly higher than that of other groups, such as the AF3-server model, which scored 0.389. Due to the token limit of the AF3 web server (5000 tokens), the AF3-server model submitted during CASP16 only included a portion of the subunits for this target. After CASP16, the release of the AF3 source codes allowed us to re-run AF3 locally using the complete sequences of M1271. The predicted model, presented in Figure 3c, exhibits a significantly different overall shape compared to the native structure. This further highlights the effectiveness of our sequence optimization strategy.

A total of 10 protein domains were officially assessed (Figure 3d). The average GDT-TS for our model is 84.64, compared to 73.51 for the AF3-server model.

The second example, H1258, is a complex structure formed by *LRRK2* and 14-3-3 (A1B2). With complete sequences, the model predicted by AF3 is inaccurate, reflected in a confidence score of less than 0.4. In this AF3 model, the DockQ score for the interface between *LRRK2* and 14-3-3 is below 0.2 (see Figure 3e, red cartoon). To address this, we extracted residues 861–1014 from *LRRK2* and re-modeled their interaction with 14–3-3 using AF3. This approach yielded a model with a significantly improved

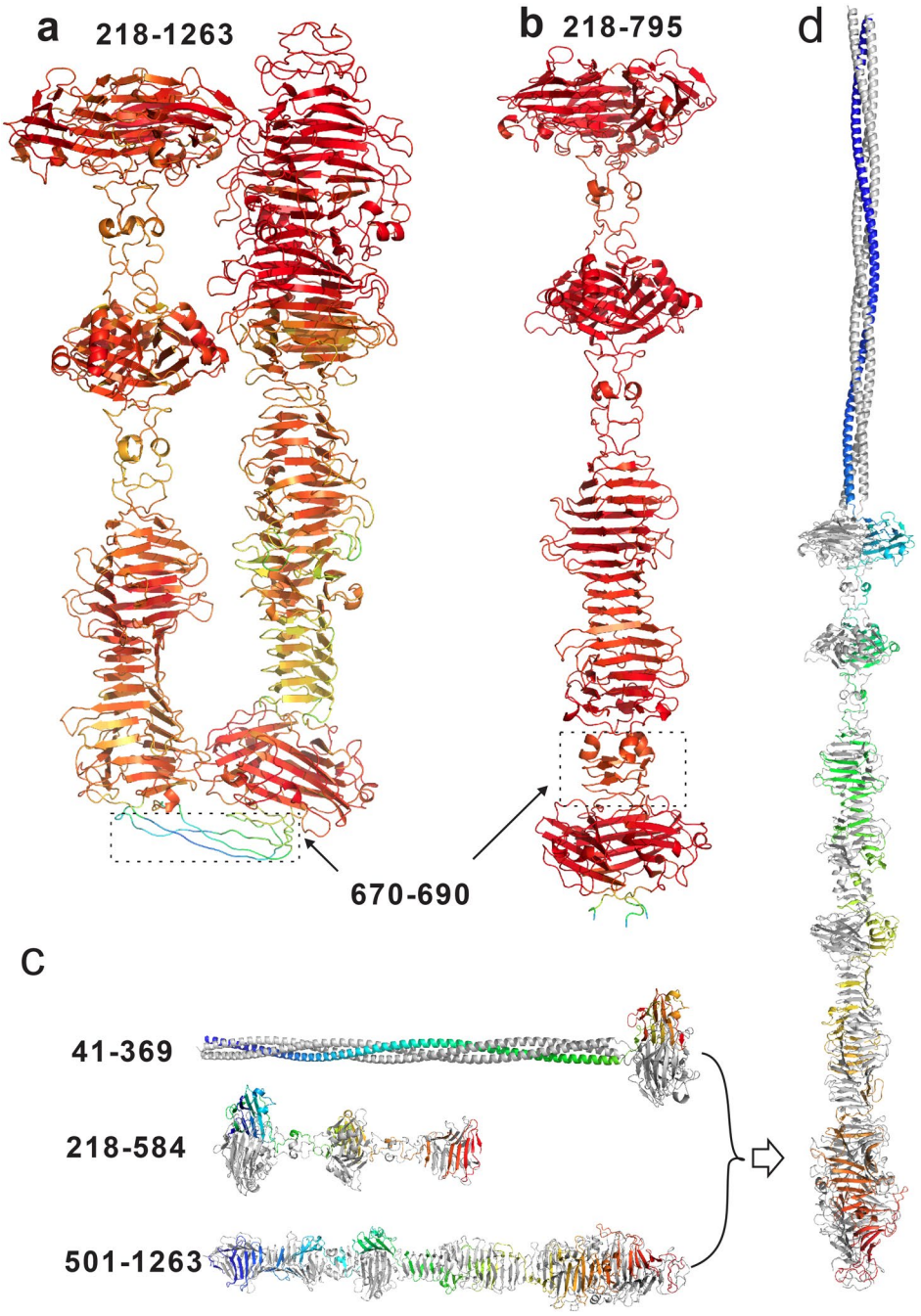

FIGURE 4 | Predicted structure models for the target T0257o. (a, b) predicted complex models for T0257o with sequences (a) 218–1263, and (b) 218–795. The complex models are colored by confidence (i.e., pLDDT), in which red/blue indicates high/low confidence. (c) complex models predicted using overlapping fragments. (d) predicted complex model of T0257o by assembling the models of the overlapping fragments. The rainbow color in (c, d) indicates the direction of each fragment.

confidence score greater than 0.9. The DockQ score for the interface in our model exceeds 0.7 (see Figure 3e, blue cartoon), which is substantially higher than that of the default AF3 model. This result suggests the importance of our sequence optimization. However, as previously mentioned, our model may still exhibit lower scores when the full-length structure is assessed. One potential solution is to construct a full-length model using the partial model as a template. Unfortunately, specifying multimeric templates is currently unsupported in AFM/AF3 and other methods. We are actively exploring ways to overcome this limitation.

The third example is T0257o. The sequence optimization module successfully predicted this target as a homo trimer. However, when predicting the trimer model with the optimized sequence (positions 41–1236), the model exhibited two shoulders, with an overall confidence score of about 0.5. The region connecting these shoulders was modeled with low accuracy. Even after removing the 217 N-terminal residues that formed a long alpha helix, a similar model was obtained. The residues 670–690 were modeled with low confidence, where the structure bends (highlighted by the dashed box in Figure 4a). We hypothesize that this may result from incorrect co-evolution signals between the

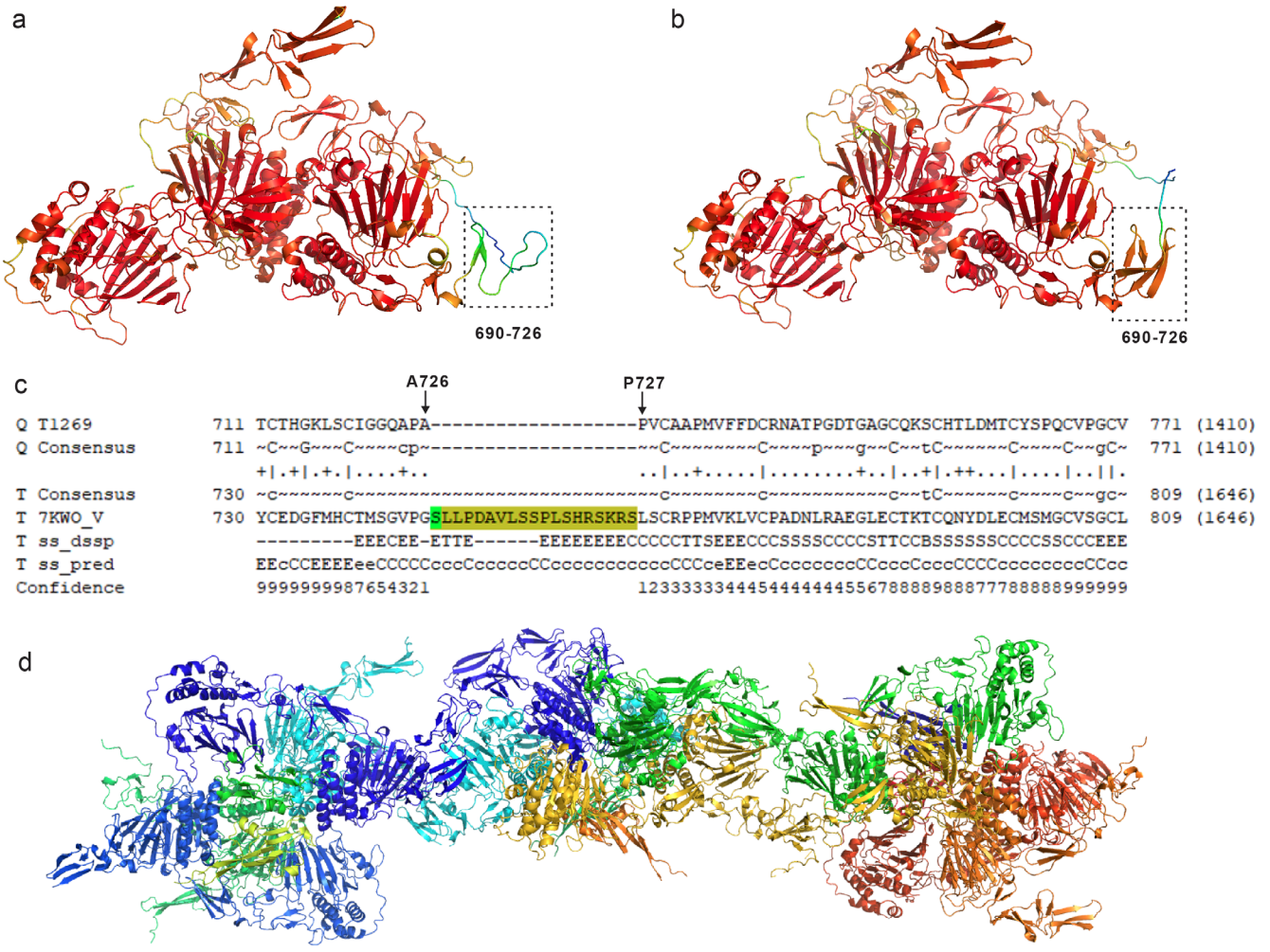

FIGURE 5 | Predicted structure models for the target T1269. (a) and (b) are the monomeric models before and after segment insertion, respectively. The structure is colored by confidence (i.e., pLDDT), in which red/blue indicates high/low confidence. (c) is the alignment between T1269 and the template 7KWO_V, indicating a segment insertion between A726 and P727 is needed for the target. (d) is the filament structure model built with the improved model in (b) and the template 7A5O.

two shoulders. Consequently, we submitted a shorter sequence (positions 218–795) to the AF3 server, which produced a model with a significantly higher confidence score (0.84, Figure 4b). This model no longer bends at residues 670–690, confirming our hypothesis.

Based on the above observation, we divided the sequence into three overlapping segments (Figure 4c) and submitted them to the AF3 server to generate structural models. All segments were modeled with high confidence, with the overlapping regions acting as "glue" to connect the individual segments. Specifically, the full-length model (Figure 4d) was predicted by iteratively superimposing the segment models using the TM-score program based on the overlapping regions. The resulting model may exhibit steric clashes or breaks at the segment boundaries, which could be improved through refinement. The DockQ score for this model is 0.783, compared to 0.197 for the default AF3-server model.

The fourth example is T1269, which is a filament. The sequence optimization indicated that residues after position 1250 are largely disordered, which were thus removed during the modeling. When predicting it as a monomer, the model achieved a relatively high confidence score (0.81). However, the residues 690–726 in the model were estimated with very low accuracy (pLDDT < 30, Figure 5a). We attempted to improve the quality by modeling the target as a dimer, trimer, and tetramer, but without success. Given that multiple templates exist for this target, we examined the alignments between this target and the templates. The alignment with template 7KWO_V (Figure 5c) indicated a segment deletion between A726 and P727. Based on this alignment, we manually inserted a segment of 19 amino acids into the target sequence and modeled the structure with the new sequence. The inserted segment was simply removed from the model for submission. Thus, A726 and P727 will be distant in the predicted structure without forming a peptide bond. In fact, the segment G721-P727 was not resolved in the experimental structure of this target, reflecting that similar difficulty might be faced during the structure determination by Cryo-EM. This modification significantly improved the pLDDT scores for residues 690–726 in the new model to over 80 (Figure 5b). With this enhanced model and the filament template 7A5O, we successfully constructed its filament structure (Figure 5d).

4.4 | What Went Wrong?

We encountered challenges in predicting accurate models for a few targets. We provide two such examples: H0258 and H0227, which were caused by inappropriate sequence optimization and incorrect assignment of stoichiometry, respectively.

For target H0258, we correctly predicted its stoichiometry as A1B2. However, the sequence optimization revealed that residues 893–1014 in subunit A were disordered, a finding further confirmed by low pLDDT scores in the monomer model (Figure S2a). These residues were removed before constructing a complex model for the target. Unfortunately, they are crucial for interactions with subunits B and C, making it impossible to build an accurate complex model; the maximum confidence score achieved was around 0.3. Recognizing this issue during Phase 1, we reintroduced the removed residues, resulting in a significantly more confident model (Figure S2a). As noted in the official report [33], our model for H1258 significantly outperformed other groups in terms of interface quality, which was also analyzed in the previous section.

For target H0227, the stoichiometry was incorrectly predicted as A6B6, which led to an inaccurate model. An experimental structure exists for subunit B, which is a hexamer. This prior information suggests that the stoichiometry should be "A?B6." Due to symmetry considerations, we mistakenly predicted the stoichiometry as A6B6, resulting in a model with a confidence score of approximately 0.45 (Figure S2b), without thoroughly exploring other combinations during Phase 0. With the release of the correct stoichiometry information in Phase 1, we re-predicted the complex structure, generating a new model with a higher confidence score (~0.65, Figure S2c).

The examples discussed above highlight the necessity of exploring alternative sequences and stoichiometries, unless highly confident models have been achieved through default optimization. In addition to these examples, several other targets, such as antibody/nanobody-involved complexes (e.g., H1204 and H1232), present significant challenges. Combining physics-based sampling with deep learning methods, as demonstrated by Dr. Dima Kozakov's group, may enhance the modeling accuracy.

5 | Results for RNA Structure Prediction

The Yang-Server predictions for RNA were primarily based on trRNA2 and RNAthreader. Yang-Server submitted models for 31 of the 36 officially assessed RNA monomer targets. Overall, Yang-Server ranked 4th among the 64 participating groups and was the best server group, surpassing AF3-server (ranked 9th; Figure 6a). Note that the Yang-Server results for large RNAs (>400 nucleotides) without templates included the AF3 predictions. For the remaining targets, Yang-Server maintained a consistent ranking, still outperforming AF3-server (Figure 6b), illustrating the robustness of our prediction strategy. The following examples illustrate the effectiveness of Yang-Server in RNA structure prediction.

5.1 | Secondary Structure Prior Knowledge

The target R1256 is an *SL5 RNA* from the 5′ proximal region of coronaviruses, which exhibits 30 alternative conformations. As shown in Figure 6c, even without SS information, trRNA2 achieved competitive RMSDs with AF3 (~20 Å) for this target. The SS generated by the SS module in trRNA2 (denoted by trRNA-SS) closely resembled that extracted from the experimental 3D structure (Figure S3a). Using this prior information as input, the final Yang-Server model is significantly more accurate than the best AF3-server model, achieving an RMSD of 7.5 Å. Furthermore, the Yang-Server model exhibited a much lower clashscore (2.0) than the AF3-server model (12.1). According to the Z-score ranking, the Yang-Server model for this target was the most accurate prediction across all groups (Figure S3b). This example highlights the ability of trRNA2 to generate accurate predictions by exploring SS prior information, a capability not readily available in AF3.

5.2 | Template-Based Modeling

The target R1281 is a *2F RNA 6-helix bundle dimer*, a mutated version of a solved RNA structure (PDB ID: 7PTK), which was a synthetic RNA target in CASP15. Despite the availability of template information, the prediction of this large RNA remained challenging. No group achieved an RMSD below 17 Å. For this challenging target, we identified 3D templates and performed template-based modeling using the automated RNAthreader method (Figure 6d). This prediction was the best model in terms of diverse metrics, including RMSD, TM-score, and GDT-TS. In contrast, the AF3-server performed below average, with a *Z*-score of zero (Figure S3c).

In addition, we observed that similar results could be obtained by feeding the SS of the template into the trRNA2 model (Figure 6d). This example demonstrates the potential of combining template information with deep learning-based methods like trRNA2 to improve RNA 3D structure prediction, which can be further investigated in depth in the future.

6 | Conclusions

We have entered an exciting era of biomolecular structure prediction. By utilizing state-of-the-art structure predictors, including AlphaFold2, AlphaFold3, trRosettaX2, and trRosettaRNA2, we explored the potential for generating accurate structural models through careful optimization of input information in CASP16.

For protein structure prediction, we enhanced the input sequences by removing intrinsically disordered regions—a simple yet effective approach that yielded accurate models for protein domains across most targets. However, only a small fraction (<25%) of the protein multimers were predicted with high quality, indicating that predicting protein multimers remains a significant challenge. In the future, it will be beneficial to explore various strategies to enhance the modeling process. This includes comprehensive model sampling using AFsample [34] and alternative MSA generation through structure-based

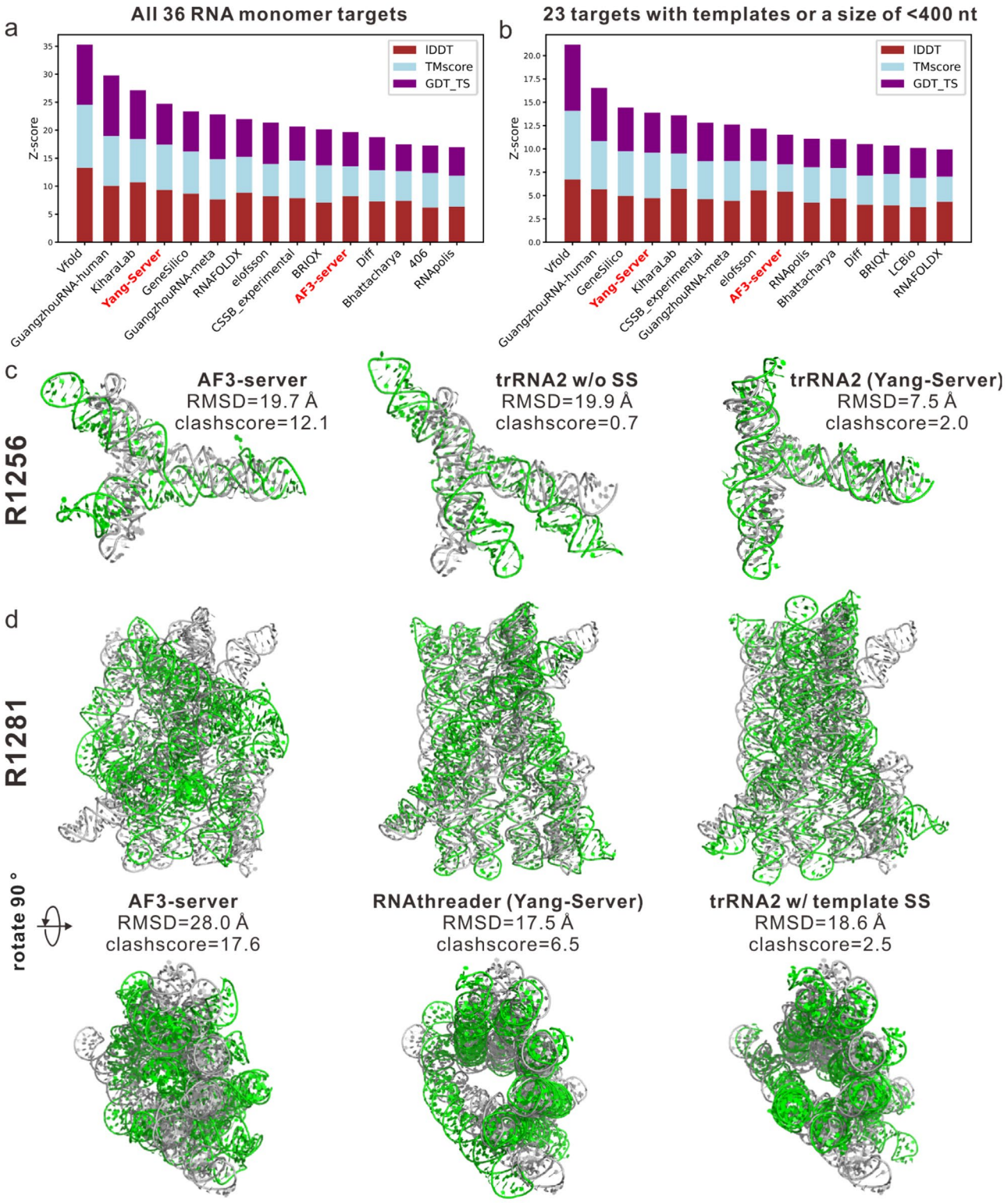

FIGURE 6 | Results for the RNA structure prediction by Yang-Server. (a, b) Top 15 RNA structure prediction groups ranked by summed *Z*-score (> 0.0) on all 36 officially assessed targets and 23 targets with < 400 nucleotides or with available templates. Server groups are highlighted in red bold in the *x*-axis labels. (c, d) Two examples, R1256 and R1281, to illustrate the Yang-Server performance. The predicted and experimental structures are shown in green and gray cartoons, respectively.

alignment in MULTICOM [35]. Another way is to develop an AlphaFold-independent system, using novel algorithms and physics-enhanced neural networks [32, 33].

In RNA structure prediction, trRosettaRNA2, with improved secondary structure inputs, successfully generated the most accurate models for several RNA targets in CASP16. While

automated predictions are not yet as accurate as those of leading human experts Vfold [36, 37], deep learning holds great promise for advancing RNA structure prediction in the coming years.

Author Contributions

Wenkai Wang: formal analysis, investigation, methodology, writing – original draft. **Yuxian Luo:** methodology, writing – original draft, software. **Zhenling Peng:** supervision, methodology, writing – review and editing, formal analysis. **Jianyi Yang:** conceptualization, methodology, project administration, supervision, formal analysis, writing – review and editing.

Acknowledgments

We would like to express our gratitude to the organizers, assessors, and target providers of the CASP16 experiment. This work is supported by the following funding sources: National Key Research and Development Program of China (2024YFA0916901), National Natural Science Foundation of China (NSFC T2225007, T2222012, 32430063), Postdoctoral Fellowship Program and China Postdoctoral Science Foundation (BX20240212), and Fundamental Research Funds for the Central Universities.

Data Availability Statement

The data that support the findings of this study are openly available in CASP16 at https://predictioncenter.org/casp16/.

References

1. J. Jumper, R. Evans, A. Pritzel, et al., "Highly Accurate Protein Structure Prediction With AlphaFold," *Nature* 596, no. 7873 (2021): 583–589.

2. Z. Peng, W. Wang, H. Wei, X. Li, and J. Yang, "Improved Protein Structure Prediction With trRosettaX2, AlphaFold2, and Optimized MSAs in CASP15," *Proteins: Structure, Function, and Bioinformatics* 91, no. 12 (2023): 1704–1711.

3. H. Woo, Y. Kim, and C. Seok, "Protein Loop Structure Prediction by Community-Based Deep Learning and Its Application to Antibody CDR H3 Loop Modeling," *PLoS Computational Biology* 20, no. 6 (2024): e1012239.

4. C. Xiang, W. Wang, Z. Peng, and J. Yang, "Generating Dynamic Structures Through Physics-Based Sampling of trRosettaX2-Predicted Inter-Residue Geometries," bioRxiv 2025.05.28.656531, 2025.

5. Y. Yan, D. Zhang, P. Zhou, B. Li, and S. Y. Huang, "HDOCK: A Web Server for Protein-Protein and Protein-DNA/RNA Docking Based on a Hybrid Strategy," *Nucleic Acids Research* 45, no. W1 (2017): W365–W373.

6. D. Kozakov, D. R. Hall, B. Xia, et al., "The ClusPro Web Server for Protein-Protein Docking," *Nature Protocols* 12, no. 2 (2017): 255–278.

7. R. Evans, M. O'Neill, A. Pritzel, et al., "Protein Complex Prediction With AlphaFold-Multimer," bioRxiv 2021.10.04.463034, 2021.

8. J. Abramson, J. Adler, J. Dunger, et al., "Accurate Structure Prediction of Biomolecular Interactions With AlphaFold 3," *Nature* 630, no. 8016 (2024): 493–500.

9. R. Krishna, J. Wang, W. Ahern, et al., "Generalized Biomolecular Modeling and Design With RoseTTAFold All-Atom," *Science* 384, no. 6693 (2024): eadl2528.

10. W. Wang, C. Feng, R. Han, et al., "trRosettaRNA: Automated Prediction of RNA 3D Structure With Transformer Network," *Nature Communications* 14, no. 1 (2023): 7266.

11. Y. Li, C. Zhang, C. Feng, R. Pearce, P. Lydia Freddolino, and Y. Zhang, "Integrating End-to-End Learning With Deep Geometrical Potentials for Ab Initio RNA Structure Prediction," *Nature Communications* 14, no. 1 (2023): 5745.

12. T. Shen, Z. Hu, S. Sun, et al., "Accurate RNA 3D Structure Prediction Using a Language Model-Based Deep Learning Approach," *Nature Methods* 21, no. 12 (2024): 2287–2298.

13. R. Pearce, G. S. Omenn, and Y. Zhang, "De Novo RNA Tertiary Structure Prediction at Atomic Resolution Using Geometric Potentials From Deep Learning," bioRxiv 2022.05.15.491755, 2022.

14. Y. Kagaya, Z. Zhang, N. Ibtehaz, et al., "NuFold: End-To-End Approach for RNA Tertiary Structure Prediction With Flexible Nucleobase Center Representation," *Nature Communications* 16, no. 1 (2025): 881.

15. M. Baek, R. McHugh, I. Anishchenko, H. Jiang, D. Baker, and F. DiMaio, "Accurate Prediction of Protein-Nucleic Acid Complexes Using RoseTTAFoldNA," *Nature Methods* 21, no. 1 (2024): 117–121.

16. R. Das, R. C. Kretsch, A. J. Simpkin, et al., "Assessment of Three-Dimensional RNA Structure Prediction in CASP15," *Proteins* 91, no. 12 (2023): 1747–1770.

17. F. Bu, Y. Adam, R. W. Adamiak, et al., "RNA-Puzzles Round V: Blind Predictions of 23 RNA Structures," *Nature Methods* 22, no. 2 (2025): 399–411.

18. D. T. Jones and D. Cozzetto, "DISOPRED3: Precise Disordered Region Predictions With Annotated Protein-Binding Activity," *Bioinformatics* 31, no. 6 (2015): 857–863.

19. M. Remmert, A. Biegert, A. Hauser, and J. Söding, "HHblits: Lightning-Fast Iterative Protein Sequence Searching by HMM-HMM Alignment," *Nature Methods* 9, no. 2 (2012): 173–175.

20. M. Steinegger and J. Söding, "MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets," *Nature Biotechnology* 35, no. 11 (2017): 1026–1028.

21. L. S. Johnson, S. R. Eddy, and E. Portugaly, "Hidden Markov Model Speed Heuristic and Iterative HMM Search Procedure," *BMC Bioinformatics* 11, no. 1 (2010): 431.

22. Y. Luo, H. Wu, H. Wei, Z. Peng, and J. Yang, "Predicting the Oligomeric State of Proteins Using Multiple Templates Detected by Complementary Alignment Methods," *Proteins: Structure, Function, and Bioinformatics* 93, no. 12 (2025): 2138–2149.

23. W. Wang, Z. Peng, and J. Yang, "Predicting RNA 3D Structure and Conformers Using a Pre-Trained Secondary Structure Model and Structure-Aware Attention," bioRxiv 2025.04.09.647915, 2025.

24. P. Danaee, M. Rouches, M. Wiley, D. Deng, L. Huang, and D. Hendrix, "bpRNA: Large-Scale Automated Annotation and Analysis of RNA Secondary Structure," *Nucleic Acids Research* 46, no. 11 (2018): 5381–5394.

25. Z. Du, Z. Peng, and J. Yang, "RNA Threading With Secondary Structure and Sequence Profile," *Bioinformatics* 40, no. 2 (2024): btae080.

26. E. P. Nawrocki and S. R. Eddy, "Infernal 1.1: 100-Fold Faster RNA Homology Searches," *Bioinformatics* 29, no. 22 (2013): 2933–2935.

27. The Rnacentral Consortium, "RNAcentral: A Hub of Information for Non-Coding RNA Sequences," *Nucleic Acids Research* 47, no. D1 (2019): D221–D229.

28. J. Singh, J. Hanson, K. Paliwal, and Y. Zhou, "RNA Secondary Structure Prediction Using an Ensemble of Two-Dimensional Deep Neural Networks and Transfer Learning," *Nature Communications* 10, no. 1 (2019): 5407.

29. H. K. Wayment-Steele, W. Kladwang, A. I. Strom, et al., "RNA Secondary Structure Packages Evaluated and Improved by High-Throughput Experiments," *Nature Methods* 19, no. 10 (2022): 1234–1242.

30. B. A. Sweeney, D. Hoksza, E. P. Nawrocki, et al., "R2DT Is a Framework for Predicting and Visualising RNA Secondary Structure Using Templates," *Nature Communications* 12, no. 1 (2021): 3494.

31. A. M. Watkins, R. Rangan, and R. Das, "FARFAR2: Improved De Novo Rosetta Prediction of Complex Global RNA Folds," *Structure* 28, no. 8 (2020): 963–976.e6.

32. R. Yuan, J. Zhang, A. Kryshtafovych, et al., "CASP16 Protein Monomer Structure Prediction Assessment," bioRxiv 2025.05.29.656942, 2025.

33. J. Zhang, R. Yuan, A. Kryshtafovych, et al., "Assessment of Protein Complex Predictions in CASP16: Are we Making Progress?," bioRxiv 2025.05.29.656875, 2025.

34. B. Wallner, "AFsample: Improving Multimer Prediction With AlphaFold Using Massive Sampling," *Bioinformatics* 39, no. 9 (2023): btad573.

35. J. Liu, P. Neupane, and J. Cheng, "Improving AlphaFold2- and AlphaFold3-Based Protein Complex Structure Prediction With MULTICOM4 in CASP16," *Proteins: Structure, Function, and Bioinformatics* 94, no. 1 (2026): 131–141.

36. S. Zhang, J. Li, Y. Zhou, and S. J. Chen, "Enhancing RNA 3D Structure Prediction in CASP16: Integrating Physics-Based Modeling With Machine Learning for Improved Predictions," *Proteins: Structure, Function, and Bioinformatics* 94, no. 1 (2026): 239–248.

37. R. C. Kretsch, A. M. Hummer, S. He, et al., "Assessment of Nucleic Acid Structure Prediction in CASP16," bioRxiv 2025.05.06.652459, 2025.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information.