

Structural bioinformatics

DeepFold: enhancing protein structure prediction through optimized loss functions, improved template features, and re-optimized energy function

Jae-Won Lee^{1,2,†}, Jong-Hyun Won^{1,2,†}, Seonggwang Jeon^{1,2,†}, Yujin Choo^{2,3,†}, Yubin Yeon^{1,2,†}, Jin-Seon Oh^{2,3,†}, Minsoo Kim^{4,†}, SeonHwa Kim⁵, InSuk Joung⁶, Cheongjae Jang⁷, Sung Jong Lee⁸, Tae Hyun Kim¹, Kyong Hwan Jin⁵, Giltae Song⁹, Eun-Sol Kim¹, Jejoong Yoo⁴, Eunok Paek¹, Yung-Kyun Noh^{1,10,*}, Keehyoung Joo^{10,*}

¹Department of Computer Science, Hanyang University, Seoul 04763, Korea

²Center for Advanced Computation, Korea Institute for Advanced Study, Seoul 02455, Korea

³Department of Artificial intelligence, Hanyang University, Seoul 04763, Korea

⁴Department of Physics, Sungkyunkwan University, Suwon 16419, Korea

⁵School of Electrical Engineering, Korea University, Seoul 02841, Korea

⁶Standigm Inc., Seoul 06234, Korea

⁷Artificial Intelligence Institute, Hanyang University, Seoul 04763, Korea

⁸Basic Science Research Institute, Changwon National University, Changwon 51140, Korea

⁹School of Computer Science and Engineering, Pusan National University, Busan 46241, Korea

¹⁰School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Korea

*Corresponding authors. Center for Advanced Computation, Korea Institute for Advanced Study, Seoul 02455, Korea. E-mail: newton@kias.re.kr (K.J.); Department of Computer Science, Hanyang University, Seoul 04763, Korea. E-mail: nohyung@hanyang.ac.kr (Y.-K.N.)

[†]Equal contribution

Associate Editor: Arne Elofsson

Abstract

Motivation: Predicting protein structures with high accuracy is a critical challenge for the broad community of life sciences and industry. Despite progress made by deep neural networks like AlphaFold2, there is a need for further improvements in the quality of detailed structures, such as side-chains, along with protein backbone structures.

Results: Building upon the successes of AlphaFold2, the modifications we made include changing the losses of side-chain torsion angles and frame aligned point error, adding loss functions for side chain confidence and secondary structure prediction, and replacing template feature generation with a new alignment method based on conditional random fields. We also performed re-optimization by conformational space annealing using a molecular mechanics energy function which integrates the potential energies obtained from distogram and side-chain prediction. In the CASP15 blind test for single protein and domain modeling (109 domains), DeepFold ranked fourth among 132 groups with improvements in the details of the structure in terms of backbone, side-chain, and Molprobability. In terms of protein backbone accuracy, DeepFold achieved a median GDT-TS score of 88.64 compared with 85.88 of AlphaFold2. For TBM-easy/hard targets, DeepFold ranked at the top based on Z-scores for GDT-TS. This shows its practical value to the structural biology community, which demands highly accurate structures. In addition, a thorough analysis of 55 domains from 39 targets with publicly available structures indicates that DeepFold shows superior side-chain accuracy and Molprobability scores among the top-performing groups.

Availability and implementation: DeepFold tools are open-source software available at <https://github.com/newtonjoo/deepfold>.

1 Introduction

Proteins serve as the essential building blocks of biological organisms, acting as the machinery of life. Protein functionality is highly dependent on their 3D structures; therefore, predicting and understanding these structures is crucial. The prediction of complete protein structures from their amino acid sequences alone has been a long-standing challenge, as demonstrated by the biennial CASP competition since 1994.

Two primary computational approaches have been employed to address this challenge: template-based modeling (TBM) and free modeling (FM). TBM constructs a protein structure by

using structures of homologous sequences, initially introduced in a study by Browne *et al.* (Browne *et al.* 1969). In TBM methods, finding a highly similar sequence from a given sequence is critical to producing accurate structures (Pearce and Zhang 2021). A conventional method for this is to align sequences using a position-specific score matrix (PSSM) (Altschul *et al.* 1997), or sequence profile Hidden Markov Models (HMMs) (Krogh *et al.* 1994) which made huge impact on searches for similar sequences and structures. These template structures serve as structural restraints for target structure modeling through sequence-structure alignment and optimization of empirical molecular

Received: 14 June 2023; Revised: 17 November 2023; Editorial Decision: 20 November 2023; Accepted: 22 November 2023

© The Author(s) 2023. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

mechanics force fields (Šali and Blundell 1993, Case *et al.* 2005, Huang *et al.* 2016). Subsequent refinement, particularly of side-chain structures, is performed using empirical potentials employing molecular dynamics refinement or more efficient global optimization algorithms (Joo *et al.* 2007, Hong *et al.* 2017). Numerous groups have contributed to this field up until CASP12 (Joo *et al.* 2007, Adhikari *et al.* 2017, Hong *et al.* 2017, Ovchinnikov *et al.* 2017, Wang *et al.* 2017b, Zhang *et al.*, 2017).

As an alternative, FM does not require known templates and instead exploits thermo-physical properties of protein structures. Molecular dynamics methods are predominant in this area (Zhang *et al.* 2011, Robustelli *et al.* 2018). Based on the assumption that native structures exhibit the lowest energy conformations (Anfinsen 1973), researchers attempt to identify the lowest energy conformations. However, this approach is practical only for short sequence lengths (Duan and Kollman 1998).

Another route to protein structure modeling without template structures has been prediction of contacts between residues which typically utilizes the correlated mutations of residues from multiple sequence alignments for the target sequence (Morcos *et al.* 2011, Ekeberg *et al.* 2013). An important breakthrough has been an idea to incorporate deep neural networks to harness the contact information hidden in the MSA's (Wang *et al.* 2017a, b) which led to the surprising appearance and success of AlphaFold1 in CASP13 (Senior *et al.* 2020). Recently, deep neural network models called AlphaFold2 (Jumper *et al.* 2021) (AF2) and RoseTTAFold (Baek *et al.* 2021) have made significant breakthroughs in this problem. These neural networks use attention modules (Bahdanau *et al.* 2014, Vaswani *et al.* 2017) to inject long-range contact information from MSAs and templates into vector representations, successfully constructing a complete protein structure based on the representation. In CASP14, AF2 demonstrated striking prediction performance, showing high accuracy in predicted protein structures on most of the targets (Kwon *et al.* 2021, Pereira *et al.* 2021). However, despite AF2's remarkable results on average, there still remain some targets for which the predictions of AF2 are rather poor, especially in the cases when MSAs are insufficient for proper modeling. Therefore there are still rooms for further improvement in deep neural network models, both in terms of efficiency and quality of the predicted structures. In particular, improving the accuracy of side-chain modeling would be important for high accuracy modeling and real applications.

Several attempts have been made to improve AF2 in various ways since its advent. One avenue is to address the efficiency of training and inference. UniFold (Li *et al.* 2022) and OpenFold (Ahdritz *et al.* 2022) are the two initial works in this field. While UniFold is a slightly modified model of AF2 with its own training system, OpenFold is a reconstructed version of AF2 with accelerated training/inference time. ColabFold (Mirdita *et al.* 2022), ParaFold (Zhong *et al.* 2021), and FastFold (Cheng *et al.* 2022) are another model that demonstrated fast MSA searches and inference speed. Some works have attempted to build models that cover certain limitations of AF2. ESMFold (Lin *et al.* 2023), EMBER2 (Weissenow *et al.* 2022), and HelixFold (Fang *et al.* 2023) presented end-to-end models that can predict complete structures without MSAs or templates, which is not possible with AF2. With the recent success of diffusion models (Ho *et al.* 2020, Song *et al.* 2021), various efforts have been made

toward predicting protein structures using generative models (Liu *et al.* 2022, Trippe *et al.* 2023, Watson *et al.* 2023). These models showed some potentials but still have limitations compared to AF2-based methods in terms of accuracy and effectiveness.

Even with such a long history and remarkable advances in protein structure prediction, improvement in the quality of detailed structures such as side-chains has always been in strong demand, along with the quality of protein backbone structures, especially in the larger community of life sciences and industry. In this study, we address this problem by constructing a protein structure prediction model called DeepFold. In order to achieve more accurate backbone and side-chains with enhancement of the overall quality of protein structures, we modified the losses of the side-chain torsion angles and FAPE (frame aligned point error). And we introduced new loss functions for side chain confidence and secondary structure prediction. In addition, we replaced the template feature generation of AF2 by integrating a refined version of the alignment method, CRFalign (Lee *et al.* 2022), which builds upon the established principles of conditional random fields (Peng and Xu 2009). This refined version incorporates new feature vectors and gradient boosted regression trees. Lastly, re-optimization of the predicted structures is performed using a powerful global optimization algorithm called conformational space annealing (CSA) (Lee *et al.* 1997, Joo *et al.* 2007). We used a molecular mechanics energy function that integrates the potential energies obtained from the distogram and side-chain prediction. For validation, we benchmarked the prediction protocol of DeepFold on CASP13/14 targets which showed consistent improvement in side-chain torsions of proteins with a modest increase in backbone accuracy. In CASP15 blind competition, DeepFold achieved quite promising result, ranking fourth out of 132 teams in the assessors' formula metric. Further analysis revealed that DeepFold outperformed top-performing groups in terms of detailed protein structure, including side-chains and Molprobit scores.

2 Materials and methods

Figure 1 shows the flow diagram of DeepFold for protein structure prediction. Given an amino acid sequence, DeepFold starts with generating MSAs and templates, which are processed as input to the DeepFold network. Then, Evoformer (Jumper *et al.* 2021) network generates single and pair representations for the sequence that are fed into Structure modules to produce protein 3D structures. Those structures are re-optimized by CSA with an energy function including molecular mechanics force field and additional potentials from distogram and side-chain predictions. Details of DeepFold, with focus on modifications of AF2 are described below.

2.1 MSA and template feature generation

The input features for DeepFold are generated from MSAs and templates. For the MSAs, DeepFold uses standard MSA generation procedure of AF2 using JackHMMER (Johnson *et al.* 2010) with MGnify (Richardson *et al.* 2023), JackHMMER with UniRef90 (Suzek *et al.* 2015), and HHblits (Remmert *et al.* 2012) with UniClust30 (Mirdita *et al.* 2017)/BFD (Steinegger *et al.* 2019). As for the templates, we collected template candidates from HHPRED. In addition, we ranked all the structures in the PDB40 database (see Dataset

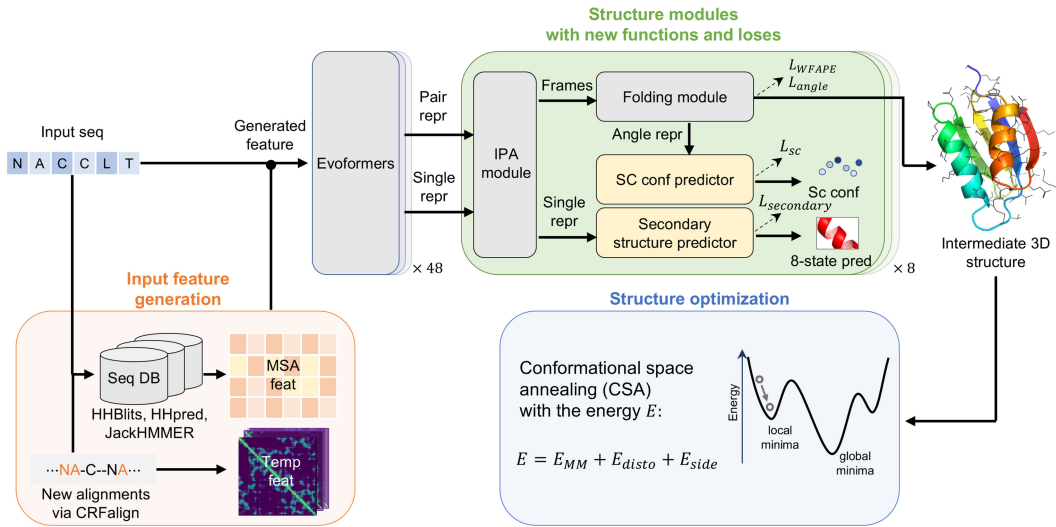


Figure 1. Prediction flow diagram of DeepFold. Given an amino acid sequence, the method first generates input features using MSAs and templates, where the MSAs are obtained from HHblits, JackHMMER, and HHpred, and the templates/alignments are generated by CRFalign. Protein 3D structures are predicted by DeepFold network and then final structures are re-optimized by conformational space annealing (CSA). (See the main text for details.)

section) based on their structural similarity to our query protein using DeepAlign (Nolle *et al.* 2020) with the AF2 model. From these ranked results, we selected the top 20 structures as additional template candidates. These aggregated template candidates were then re-ranked using CRFalign (Lee *et al.* 2022), a sequence-structure alignment method based on pairwise conditional random fields and gradient boosted regression trees. We utilized up to four of these templates and their respective alignments as input for the DeepFold network. In protein structure prediction, templates have been shown to be especially beneficial for targets with similar templates (Wu *et al.* 2022).

2.2 Sequentially conditioned torsion angle loss

Since the side-chain torsion angles χ^1, \dots, χ^4 are sequentially dependent in terms of 3D coordinates, we modified the torsion angle loss considering this sequential relationship in side chain angles by introducing the pre-torsion angle error \tilde{E}_{χ^k} :

$$\tilde{E}_{\chi^k} := \left(\tilde{E}_{\chi^{k-1}} + E_{\chi^k} - \frac{\tilde{E}_{\chi^{k-1}} \cdot E_{\chi^k}}{\sqrt{2}} \right), \text{ for } k = 2, 3, 4, \quad (1)$$

with the initial condition $\tilde{E}_{\chi^1} = E_{\chi^1}$, where the E_{χ^k} is the square-root angle error defined as:

$$E_{\chi^k} := \sqrt{\|\alpha^{\chi^k} - \hat{\alpha}^{\chi^k}\|_2} \text{ for } k = 1, 2, 3, 4. \quad (2)$$

Here, $\alpha^{\chi^k} \in \mathbb{R}^2$ is a cosine/sine vector representation of χ^k in a single residue:

$$\alpha^{\chi^k} = (\cos \chi^k, \sin \chi^k)^\top. \quad (3)$$

The sequential properties of the pre-torsion angle error \tilde{E}_{χ^k} in Equation (1) can be seen as follows. In the case of χ^1 where no preceding angles exist, the error is reduced to just the angle error $E_{\chi^1} = \sqrt{\|\alpha^{\chi^1} - \hat{\alpha}^{\chi^1}\|_2}$. For the case of χ^2, χ^3 , or χ^4 , the error references the previous error $\tilde{E}_{\chi^{k-1}}$ in the calculation of the current error, such that the current error \tilde{E}_{χ^k} increases (i.e.

is nonzero) even if the current angle difference E_{χ^k} is zero. Therefore, even though the model predicts correctly χ^2 angles, if it predicts χ^1 incorrectly, the torsion error of χ^2 will be increased. Lastly, summing the pre-torsion angle error \tilde{E}_{χ^k} for each of the χ angles, we can define the overall angle loss as follows

$$\mathcal{L}_{\text{angle}} = \frac{1}{N} \left(\sum E_{\chi^1} + \sum \tilde{E}_{\chi^2} + \sum \tilde{E}_{\chi^3} + \sum \tilde{E}_{\chi^4} \right), \quad (4)$$

where N is the number of residues and the summations are over the numbers of k th side-chain torsion angles respectively ($k = 1 \dots 4$).

2.3 The weighted FAPE loss

The Frame-aligned point error (FAPE) in AF2 plays an essential role in aligning atom positions in the generated rigid frame of each residue (i.e. the predicted local frame of each residue). The FAPE loss of AF2 is defined as

$$\mathcal{L}_{\text{FAPE}}(T_i, \mathbf{x}_j) = \min \left(\|T_i^{-1} \cdot \mathbf{x}_j - T_i^{-1(\text{true})} \cdot \mathbf{x}_j^{(\text{true})}\|^2, 10\text{\AA} \right), \quad (5)$$

where $T_i = (R_i, \vec{t}_i)$ is the predicted local frame (which represents the rotation R_i and translation \vec{t}_i of i th residue) and atom coordinates $\mathbf{x}_j \in \mathbb{R}^3$, which measures the squared error between local coordinates $T_i^{-1} \cdot \mathbf{x}_j$ and its ground truth $T_i^{-1(\text{true})} \cdot \mathbf{x}_j^{(\text{true})}$. Note that the loss has a clamping value of 10 Å that washes out the loss values larger than 10 Å. In DeepFold, we introduce the weighted FAPE loss as follows:

$$\mathcal{L}_{\text{WFAPE}}(T_i, \mathbf{x}_j) = w_{ij} \mathcal{L}_{\text{FAPE}}(T_i, \mathbf{x}_j), \quad (6)$$

where the weight w_{ij} is determined by following sigmoid-type function

$$w_{ij} = \frac{1}{(1 + \exp(-2(-d_{ij} + v)))} + b, \quad (7)$$

in which d_{ij} is the distance with maximal probability from the distogram, between C_β atoms of residues i and j . The hyperparameters v , b control the inflection point of the sigmoid function and an additional weighting constant, respectively. In this work, we used $v = 12.0$, and $b = 1.5$ which were determined by iterative grid searches. This modified loss gives larger weight on the closer atoms.

2.4 Secondary structure loss

We introduce secondary structure loss in DeepFold network by implementing small multi-layer perceptrons in the structure module as shown in Fig. 1. This network uses single representations as inputs which were obtained by IPA (Invariant Point Attention) in structure module (Jumper *et al.* 2021). And they return as output the prediction results of 8-state secondary structure of each residue (Kabsch and Sander 1983). To train this network, we defined a loss function for the secondary structure by cross-entropy as follows:

$$\mathcal{L}_{\text{secondary}} = -\frac{1}{N} \sum_i \sum_{c=1}^8 y_i^c \log p_i^c, \quad (8)$$

where y_i^c is the true secondary structure of i -th residue with c -state, and p_i^c is the predicted probability that the 8-state secondary structure of the i -th residue is in c -state.

2.5 Side-chain confidence loss

Similarly to the plddt in AF2, we defined another confidence measure focussing on the side chains of protein structure as follows. Side chain confidence s_i for each residue $i = 1, \dots, N$ is defined as:

$$s_i = \frac{1}{1 + (\delta_i/\delta_0)^2} \quad (9)$$

where δ_i is the difference between the true and predicted χ^1 angles $\delta_i = \chi_i^{1(\text{true})} - \chi_i^{1(\text{pred})}$. Here, δ_0 is a reference value of angle difference which was set as 12° in this work. In order to predict s_i , we implemented another multi-layer perceptron network in the structure module which returns as output the probability \hat{p}_i^s for s_i to be in the specific bin among 50 bins in the interval $[0,1]$. This can be compared with the true values of s_i , from which a loss function can be formed. The loss function is defined as the cross-entropy loss between one-hot encoded s_i and predictions:

$$\mathcal{L}_{\text{sc}} := -\frac{1}{N} \sum_{i=1}^N p_i^{s\top} \log \hat{p}_i^s, \quad (10)$$

where the p_i^s and \hat{p}_i^s are one-hot encoded representations of s_i and prediction \hat{s}_i , respectively. In this work, the side chain confidence score for a single sequence is defined by the average score over all residues.

With the modified loss functions and newly added loss terms, we arrive at the following combined loss \mathcal{L} which can be used for training the DeepFold model:

$$\mathcal{L} = c_1 \mathcal{L}_{\text{angle}} + c_2 \mathcal{L}_{\text{WFAPE}} + c_3 \mathcal{L}_{\text{secondary}} + c_4 \mathcal{L}_{\text{sc}} + \mathbf{c}^\top \mathcal{L}_{\text{others}}, \quad (11)$$

where the $\mathcal{L}_{\text{others}} = (\mathcal{L}_{\text{aux}}, \mathcal{L}_{\text{dist}}, \mathcal{L}_{\text{msa}}, \mathcal{L}_{\text{conf}}, \mathcal{L}_{\text{exp resolved}}, \mathcal{L}_{\text{vio}})$ are the other losses in AF2 with the default weights $\mathbf{c} = (0.5, 0.3, 2.0, 0.01, 0.01, 1.0)$ as they used in (Jumper *et al.* 2021). Supplementary Table S1 shows the relative weights of various loss terms in our modifications.

2.6 Dataset, training, and validation

We used the latest PDB database (Feb. 2022) for training. We clustered the sequences of PDB using CD-HIT with 40% sequence identity, which resulted in 31,911 protein chains. We further filtered 23,366 chains of high-resolution (with $< 2.5 \text{ \AA}$) for a fine-tuning dataset. The obtained sequences were cropped to 256 and 384 residue sizes as in the AF2 for training. Five DeepFold models were selected from training with various training schedules. For all the trained models, we employed the Uni-fold (a trainable version of AF2) training system, where the trainings were started from the AF2 parameters and then further optimized in the style of transfer learning. Full details are described in Supplementary Section S1.

2.7 Re-optimization by conformational space annealing

Once 3D structures are inferred from the DeepFold networks, we perform global optimization using conformational space annealing (CSA) with the full atom force field, distance restraints, and side chain torsion restraints generated by the networks. Full details are described in Supplementary Section S2.

3 Results

In order to evaluate the performance of DeepFold, we conducted a blind test in the CASP15 competition. As our Group Names, we used DFolding and DFolding-server, where the latter was a server prediction protocol which was performed without the final CSA re-optimization due to the 3-day deadline. We present our results obtained from the official evaluation data, which was provided by CASP15 assessors of Single Protein and Domain Modeling category [15th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (<https://predictioncenter.org/casp15/>)]. Traditionally, all targets are divided into domains, which are classified as TBM and FM domains. In the CASP15, total of 109 domains, which were officially assessed, consist of 62 TBM domains (including 47 TBM-easy and 14 TBM-hard), and 47 FM domains (8 FM/TBM and 39 FM). We also present our own assessment of 55 domains (39 targets) including 27 TBM, 27 FM domains, and 1 domain that was not classified, for which native structures were publicly available.

The official metric to compare the performance of all Groups is called *sum Z-score* which is a weighted sum of the Z-scores for each of the domains as in the following CASP15 assessor's formula (Read and Chavali 2007, Mariani *et al.* 2011, Olechnovič *et al.* 2013, Croll *et al.* 2019, Millán *et al.* 2021, Pereira *et al.* 2021) [GDT-HA (Global Distance Test for High Accuracy), reLLG (Relative eLLG score), ASE (Accuracy Self Estimate), LDDT (Local Distance Difference Test), AA (CAD-score, all atoms), SG (The average value of Sphere Grinder Score), SC_{error} (Side Chain Dihedral Error Score), Molprobit (Aggregated Molprobit Score), BB_{error}

(Backbone Dihedral Error Score), DipDiff (The average difference between the local DipScores)],

$$\begin{aligned} \text{sum Z-score} &= \frac{1}{6}(\text{GDT-HA} + \text{reLLG} + \text{ASE}) \\ &+ \frac{1}{16}(\text{LDDT} + \text{AA} + \text{SG} + \text{SC}_{\text{error}}) \\ &+ \frac{1}{12}(\text{Molprb} + \text{BB}_{\text{error}} + \text{DipDiff}). \end{aligned} \quad (12)$$

In addition to the above metric, GDT-TS is typically used as a standard measure of modeling accuracy (Global Distance Test for Tertiary Structure) to evaluate the accuracy of global backbone trace of the structure. Figure 2 illustrates the official rankings of DFolding, together with other top 50 among 132 Groups including BAKER (RosettaFold), ColabFold, NBIS-AF2-standard and OpenFold (the latter three are basically standard AF2-based protocols with modification of MSA generation). In Fig. 2a, DFolding was ranked fourth according to the assessor's formula, while in (b), DFolding was at the top in terms of sum Z-score of GDT-TS for 64 domains, classified in the TBM-easy and TBM-hard target. In the figure, DFolding-server was also included, indicating the performance difference in comparison with DFolding. Note that CSA re-optimization is not used in DFolding-server protocol due to the time limit of three-days.

Table 1 shows the results of top 10 performing groups from the Fig. 2a, in addition to three AF2 based methods. All metrics used in assessor's formula are shown in the average values instead of Z-scores. Despite the differences of backbone accuracies, the values of the metrics such as SC_{error} , AA, BB_{error} , and DipDiff are showing very little differences between groups. In particular, the value of SC_{error} (which means side-chain error for protein structure models, where lower is better) is not very distinguishable in spite of a significant difference in backbone accuracy such as GDT-TS. This is because the criterion of correctness in the metric of SC_{error} is too broad. Below, we will discuss our own evaluation for side-chain accuracies for 55 domains, where native structures are available in public. In the Molprobity score showing protein-likeness, DeepFold achieved a score of 0.97 (lower is better) and performed well among the top groups.

Figure 3 shows the effect of CSA re-optimization by comparing DFolding (y -axis) versus DFolding-server (x -axis) in

terms of GDT-TS, LDDT, SC error, and Molprobity, where we used the official metrics provided by CASP15 assessors. DFolding prediction exhibits better performance in average scores for GDT-TS, LDDT, and Molprobity, while difference of SC_{error} is very minor. It is due to the same reason mentioned above that the criterion of side-chain error is too broad. In GDT-TS, DFolding performed well for the relatively hard targets (GDT-TS less than about 70), which can be attributed to the result of CSA re-optimization. It also shows significant improvement in Molprobity score.

We now apply further analysis of our own on the performance of DeepFold with the 55 domains for which the native structures are available. In Table 2, GDT-TS and GDT-HA were calculated by using the TM-score program instead of using the official data. Side-chain accuracies of χ^1 and χ^2 were calculated by using 10° criterion of correctness, instead of using SC_{error} used in the official results. All values presented in the table are cumulative Z-scores for each metric, in line with the CASP15 assessor's methodology (higher values are better). The results from DFolding demonstrate a balanced performance in both backbone and side-chain metrics when compared to leading groups. The table shows that the sum Z-score of DFolding in Molprobity value is significantly higher than all the other top performing groups. Furthermore, the table shows that DFolding achieved higher scores in LDDT, GDT-TS, and GDT-HA than the well-known methods such as AF2, BAKER, ColabFold, and OpenFold.

Figure 4 shows a comparison between the accuracies of DFolding and DFolding-server for 55 domains. On average, DFolding outperforms DFolding-server in terms of GDT-TS and Molprobity value, while the sidechain scores are similar in terms of χ^1_{acc} and χ^2_{acc} . This indicates that CSA re-optimization process of DFolding resulted in a minor improvement for the overall structure of the 55 protein domains.

As mentioned in the Section 2, DFolding used CRFalign for the generation of template features. We can evaluate the quality of template features for 55 domains where the native structures are available. Figure 5 shows the difference between template features of DFolding and those of AF2. In order to show the difference between DFolding and AF2 in terms of the quality of templates, we used the best template among the (up to) four templates for each target. Figure 5a shows the TM-scores of the templates of DFolding in comparison with those of AF2's templates. Average TM-score of DFolding

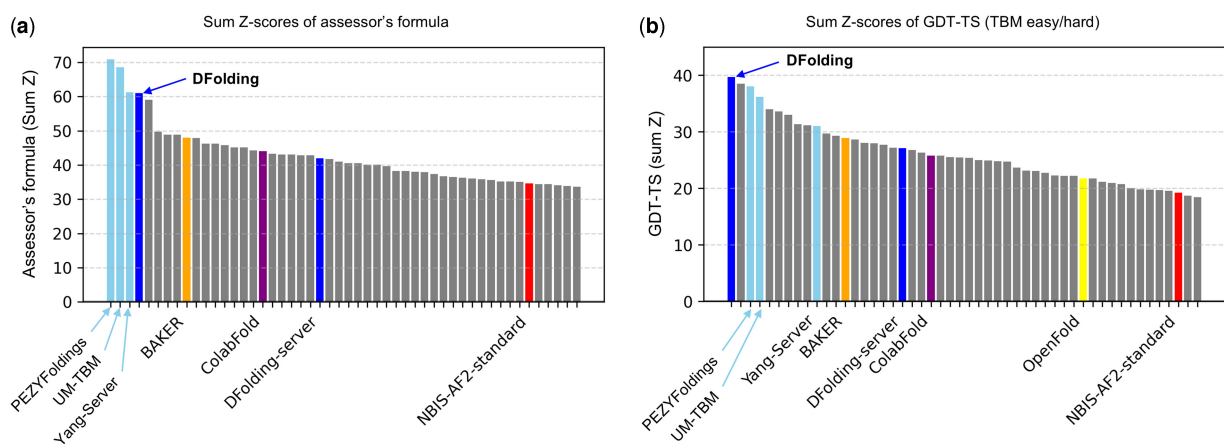


Figure 2. CASP15 rankings in terms of sum Z of (a) the assessor's formula [Equation 12] and (b) GDT-TS for TBM easy/hard targets (62 domains in total). In each figure, only top 50 out of 132 teams are shown.

Table 1. The official CASP15 results of top 10 performing groups, in addition to three AF2 based methods sorted by assessor's formula in Equation (12).^a

Rank	Methods	sum Z	GDT-TS	GDT-HA	reLLG	ASE	LDDT	AA	SG	SC _{error}	Molprobability	BB _{error}	DipDiff
1	PEZYFoldings	70.83	81.89	68.00	12.65	88.39	0.79	0.77	90.10	9.59	1.34	9.38	0.02
2	UM-TBM	68.56	83.75	69.49	12.26	89.04	0.81	0.78	92.06	9.42	1.22	9.21	0.01
3	Yang-Server	61.28	83.92	69.48	12.45	85.89	0.79	0.78	91.76	9.50	1.36	9.29	0.02
4	DFolding (DeepFold)	61.07	80.33	66.39	10.53	87.22	0.78	0.76	88.79	9.44	0.97	9.21	0.03
5	Yang	59.01	83.07	68.92	12.99	86.13	0.79	0.78	91.01	9.51	1.38	9.29	0.02
6	McGuffin	49.73	79.82	66.34	11.28	85.22	0.77	0.76	88.24	9.43	1.59	9.21	0.01
7	MULTICOM	48.83	80.20	66.41	11.99	75.55	0.78	0.77	89.98	9.44	1.38	9.21	0.01
8	MULTICOM_refine	48.83	79.25	65.46	11.37	85.75	0.79	0.77	89.75	9.43	1.16	9.21	0.01
9	BAKER	47.99	79.33	63.81	8.59	85.72	0.77	0.75	89.20	9.46	0.75	9.21	0.03
10	MULTICOM_human	47.90	79.99	66.13	11.08	75.81	0.78	0.76	89.68	9.44	1.38	9.21	0.01
17	ColabFold	44.02	75.36	62.33	10.77	89.75	0.76	0.75	86.57	9.43	1.34	9.21	0.01
45	NBIS-AF2-standard	34.63	75.81	61.74	9.59	89.84	0.76	0.75	87.21	9.44	1.28	9.21	0.00
53	OpenFold	33.25	73.71	60.34	9.13	87.79	0.74	0.74	84.72	9.53	1.72	9.30	0.00

^a All of the 11 metrics in the formula are also shown as the average (over all the target) of the raw scores. References for each of the predictors listed are as follows: PEZYFoldings (Oda 2023), UM-TBM (Peng et al. 2023), Yang-Server (Zheng et al. 2023), McGuffin (McGuffin et al. 2023), MULTICOM (Liu et al. 2023), ColabFold (Mirdita et al. 2022), and OpenFold (Ahdritz et al. 2022). The top scores in each metric are marked in bold.

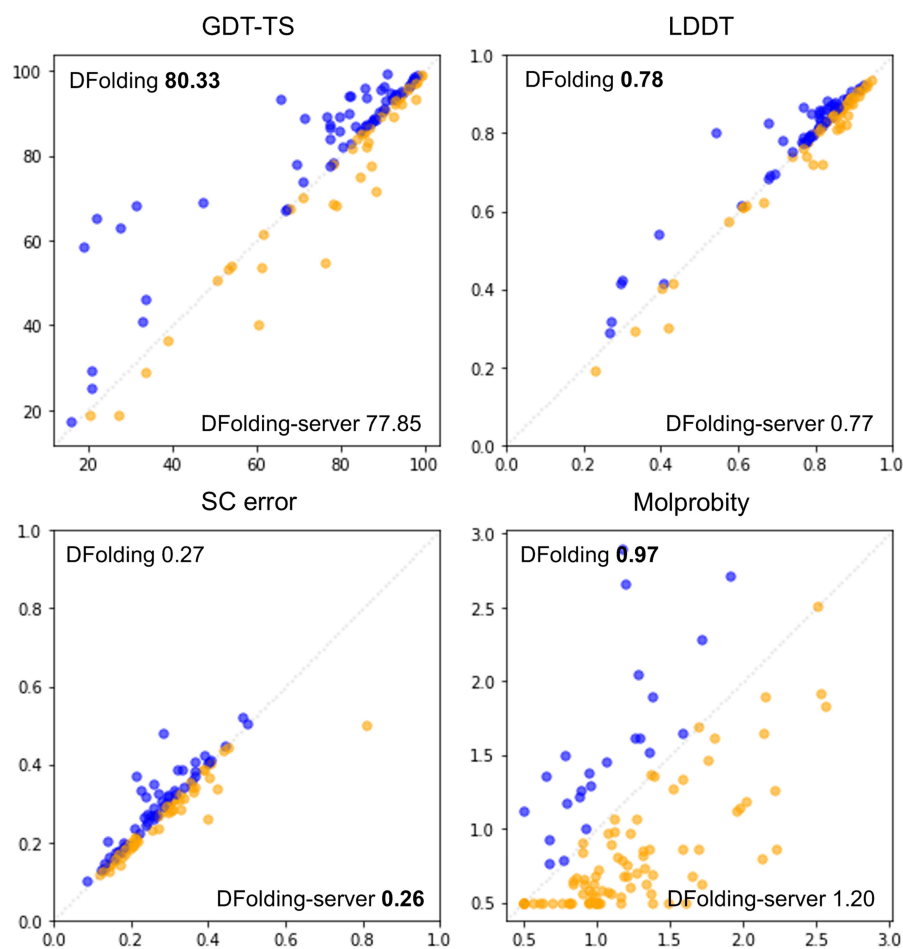


Figure 3. Comparison between the accuracies of DFolding (y-axis) versus DFolding-server (without CSA re-optimization, x-axis). The CASP15 official results of GDT-TS, LDDT, SC error, and Molprobability are used. Higher values are better for GDT-TS, LDDT, while lower values are better for SC_{error} and Molprobability. All numbers within the figures are average values over 109 domains.

templates was 0.6042 while that of AF2 was 0.5521. Figure 5b shows a comparison of the quality of CRFalign alignments and those of AF2's. The alignment quality was measured based on the pairwise distance of residues. First, for the native sequence, collect pairs of residues and measure pairwise distances. The collected pairs are >3 in sequence separation, with mutual distances <8 Å. Second, for the residue

pairs at the same location with the native sequence, pairwise distance is calculated for the aligned template sequence. Then, we calculate the ratio of the template residue pairs that have a similar distance to the native residue pairs' distance. If the difference of a residue pair distance between native and template is less than a threshold, we label that the template residue pair distance is similar to that of native's. In the case of (b), we set

Table 2. CASP15 results of top 10 teams in Table 1, evaluated on 55 domains, for which the native structures are publicly available.^a

Methods	LDDT	GDT-TS	GDT-HA	χ_{acc}^1	χ_{acc}^2	Molprobability	counts
PEZYFoldings	38.19	51.80	54.93	27.73	27.22	26.99	53
UM-TBM	36.41	47.91	53.86	27.49	31.27	27.71	55
Yang-Server	24.69	49.19	52.44	35.14	35.80	17.34	55
DFolding (DeepFold)	31.98	46.76	48.99	29.00	29.42	50.51	55
Yang	22.78	45.03	47.66	33.65	40.38	24.81	55
McGuffin	29.75	44.32	49.21	24.56	29.05	17.03	55
MULTICOM	29.41	38.33	43.88	22.78	27.23	24.20	55
MULTICOM_refine	29.70	28.98	30.73	23.98	23.32	31.21	55
BAKER	15.93	30.73	27.01	35.14	32.44	28.81	55
MULTICOM_human	27.20	37.32	41.77	22.03	22.98	24.34	55
ColabFold	24.93	20.42	21.60	21.40	22.57	21.17	55
NBIS-AF2-standard	21.52	14.87	14.57	22.69	20.29	23.74	54
OpenFold	16.46	14.09	15.16	18.03	21.07	19.67	55

^a All scores in the table are Z-sum scores. GDT-TS and GDT-HA in the table are calculated by TM-score (Zhang and Skolnick 2005). The top scores in each metric are marked in bold.

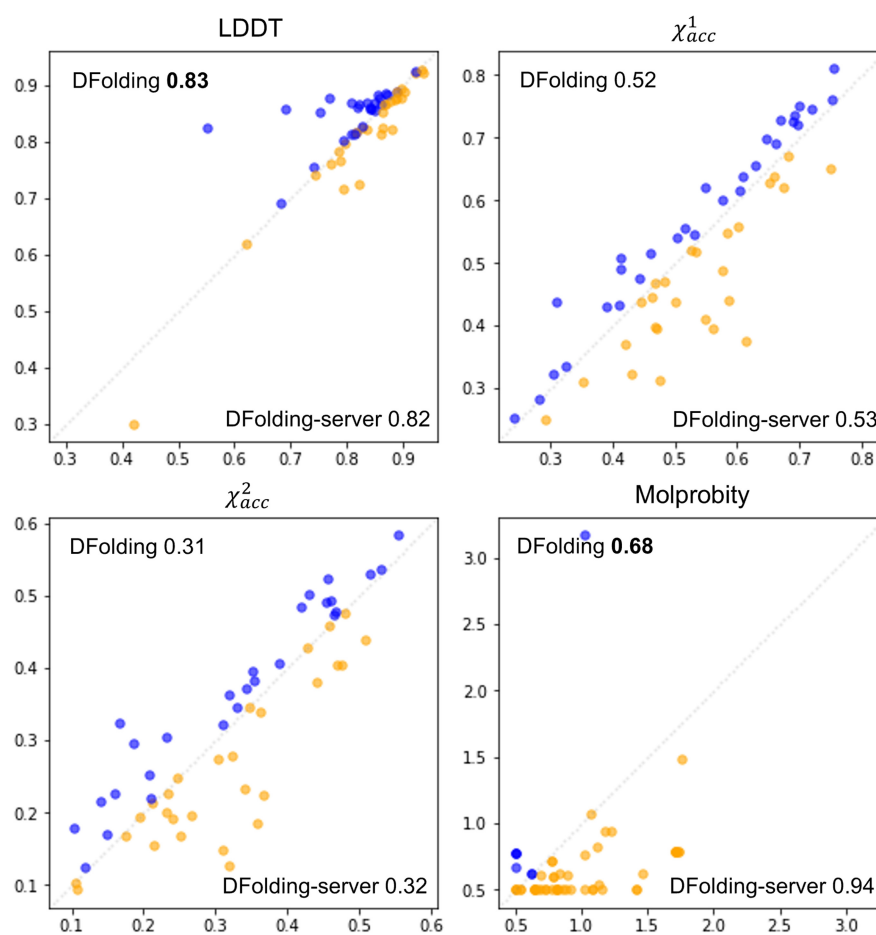


Figure 4. Comparison between the accuracies of DFolding (y-axis) and DFolding-server (x-axis) for publicly available 55 domains. All numbers within the figures are average values over 55 domains. Lower value is better for Molprobability.

the threshold as 3 Å. We can see that the alignment quality of DFolding using CRFalign shows significant improvement over that of AF2, especially in the cases of hard targets (when the ratio of alignment quality is less than about 0.7). However, because the predicted structure is greatly affected by the quality of the MSA, the superiority of template information does not guarantee the improvement of the final structure. To further verify this, we added the final results of the 18 domains in Fig. 5b where the difference in values is >0.15 in Supplementary Section S7. For example, Target T1123,

which belongs to the FM/TBM class, was among the relatively well-predicted cases. As illustrated in Fig. 6a, DFolding ranked first with a GDT-TS score of 86.80, while DFolding-server was placed in the third position with a GDT-TS score of 83.64, outperforming other groups. Figure 6c shows the best template, 5EWO (York *et al.* 2016), which was found using CRFalign. The template exhibited higher structural similarity to the native structure (TM-score=0.69) than AF2 (TM-score=0.31). Figure 6b displays the predictions of DFolding and AF2 in cartoon style with the native structure.

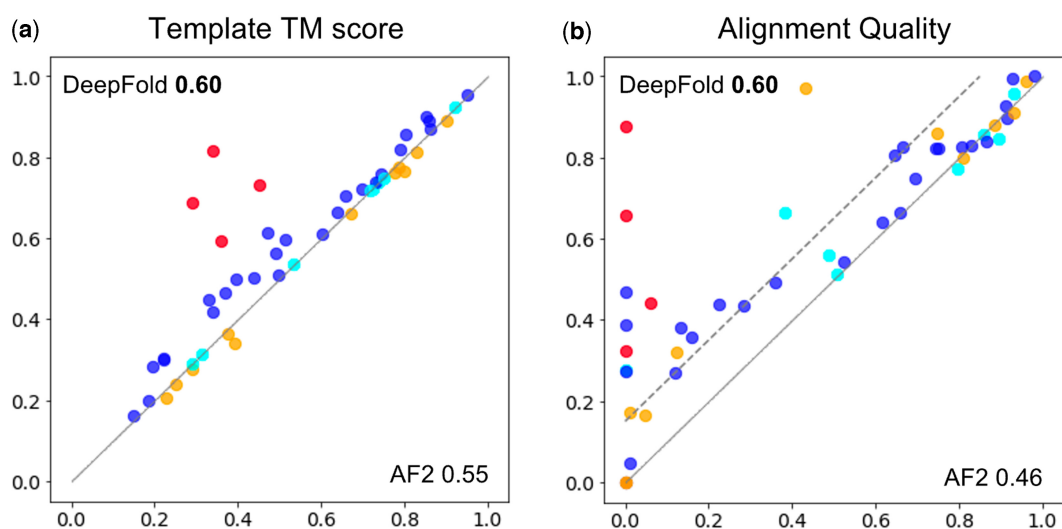


Figure 5. Comparison of the templates of the 55 domains in CASP15. (a) TM-score comparison is shown between the templates found by DeepFold (through CRFalign) versus the ones found by AF2. Each point represents the template with the highest TM-score obtained by the respective method for a given target. Orange colors are the domains with a lower TM-score for CRFalign than the AF2 method, and cyan colors indicate the domains for which the same templates are found by both methods. Red color is the target T1123. (b) A comparison of the alignment qualities of CRFalign and AF2 is shown. All numbers within the figures are average values over 55 domains. The dotted line is indicating a difference > 0.15 , and as a result, we had significantly better template information for 18 domains.

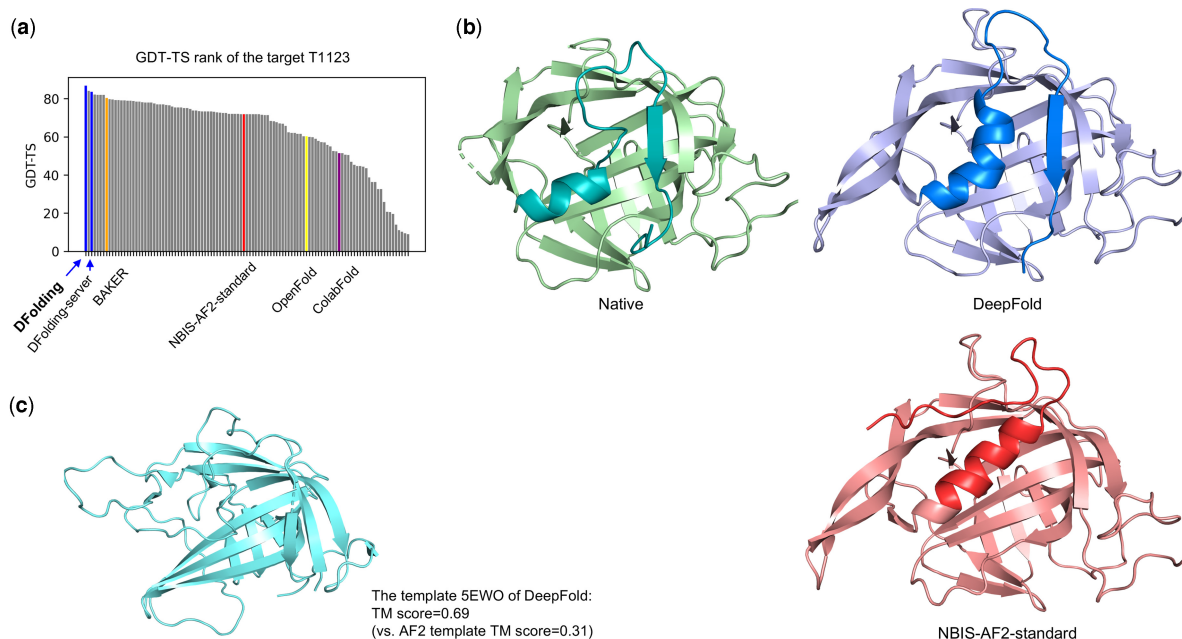


Figure 6. The target T1123 in CASP15 for which DeepFold predictions showed the best GDT-TS scores. (a) DeepFold human and server models ranked at the top three in a row. (b) The predicted 3D structures of T1123 and the native structure (colored as green). Compared to AF2, the DeepFold prediction correctly predicts helix and β -sheets near the C-terminal region, highlighted in blue. (c) The template 5EWO obtained by CRFalign shows a high structural similarity to the native structures (TM = 0.69) compared with the template in AF2 (TM = 0.31).

It can be observed that our models successfully generated β -sheets (shown in blue color) near the C-terminal, while AF2 failed to predict that region.

3.1 What went wrong

DFolding failed to generate correct fold for some targets especially for T1125 and T1130. T1125 was a challenging target, likely due to its large size of 1200 residues, which consists of 6 domains (five FM domains and one TBM-hard). Only a few groups achieved good scores (<https://predictioncenter.org/casp15/results.cgi?view=tb-sel>). For T1125-D1, the average

GDT-TS of all groups was 28.44, but that of top three groups had scores around 80. The Neff score of MSA [i.e. represents a measure of the diversity or the effective number of sequences within the alignment, with higher scores indicating a more diverse set of sequences in the MSA (Meier and Söding 2014).] generated by AF2 standard was 1.39 with only five sequences including the target itself, resulting in low plddt scores ranging from 0.39 to 0.44. For the single domain target T1130 with 198 residues (FM domain), DFDfolding exhibited poor performance with 29.40 of GDT-TS along with most of the groups, while only 3 groups achieved around 95. For this

target, the Neff score of AF2 is 0.0 with no additional homologous sequences. And no template structures were found as can be expected for an FM domain. It is highly probable that the three successful groups identified high-quality MSAs through their extended efforts for the target. We also attempted to generate MSAs using other available web servers, but with limited success. As a result, it appears that MSA quality is crucial for the successful prediction of structures using AF2-based methods.

3.2 What went right

We found that the improvement of DeepFold's prediction accuracy can be attributed to the following three elements: modified loss functions, choice of better templates, and CSA re-optimizations. To begin with, the positive effect of modified loss functions is confirmed by ablation studies, as shown in the [Supplementary Section S3](#). To summarize, modified loss functions improved the side-chain accuracy while maintaining the backbone accuracy. The choice of better templates and alignments (partly through CRFalign) improved the backbone accuracy of the specific targets. Finally, CSA re-optimization improved backbone accuracy and molprobit scores through structure optimization.

4 Conclusion

In this work, we have presented DeepFold, a pipeline for predicting protein structures. Our primary objective was to improve the accuracy of backbone and side-chain prediction, and we achieved this by making several modifications to the AF2 pipeline. These modifications included optimizing the loss functions with focus on torsion angle losses and additional loss functions introduced to improve side chain accuracy and secondary structure prediction with updated template features. Further re-optimization was performed by CSA using a molecular mechanics energy function which integrates the potential energies obtained from distogram and side-chain prediction. In the blind test at the CASP15 competition, DeepFold improves the details of the protein structure in both backbone and side-chains, ranking fourth in the category of Single Protein and Domain Modeling (109 domains). In addition, DeepFold's structure prediction showed the best results in terms of backbone accuracy of GDT-TS for the targets in the TBM class (62 among 109 domains), which are more important in practical applications. Our thorough analysis of 55 domains with publicly available ground-truth structures revealed that DeepFold showed the best side-chain accuracy and Molprobit scores in the top-performing groups. It can be seen that the effect of modification and retraining of the DeepFold network and the final re-optimization have made significant differences. In particular, it can be seen that re-optimization further improves the details of protein structure such as backbone, side-chain and Molprobit.

There remain limitations in predicting certain targets, such as very large target like T1125, and T1130 both with low-quality MSAs. This is due to the well-known fact that the performance of AF2-based methods is heavily dependent on the quality of MSAs. It is essential to evaluate the quality of the obtained MSA and enhance the quality of low-quality MSAs by employing techniques such as protein language models ([Lin et al. 2023](#)) or generative models ([Liu et al. 2022](#)).

Furthermore, in order to predict larger protein structures properly, reliable protein domain parsing must be considered.

In conclusion, our work on DeepFold represents an improvement upon the foundation laid by AF2. Through our modifications to the AF2 pipeline, we have achieved higher accuracy in terms of backbones and side-chains. We believe that our work has advanced the field of protein structure prediction, which we expect to be a valuable tool for the protein modeling community.

Acknowledgement

We thank the KIAS Center for Advanced Computation for providing computing resources.

Supplementary data

[Supplementary data](#) are available at *Bioinformatics* online.

Conflict of interest

None declared.

Funding

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [2021-0-02068, 2020-0-01373]; the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (MSIT) [NRF-2018 R1D1A1B07049312] and the 2023 KT ICT AI2XL Laboratory R&D Fund project funded by KT [KT award B230000177].

Data availability

The data underlying this article is available at <https://github.com/newtonjoo/deepfold>.

References

- Adhikari B, Hou J, Cheng J *et al.* Protein contact prediction by integrating deep multiple sequence alignments, coevolution and machine learning. *Proteins Struct Funct Bioinform* 2017;**86**:84–96.
- Ahdritz G, Bouatta N, Kadyan S *et al.* Openfold: retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *bioRxiv*, <https://doi.org/10.1101/2022.11.20.517210>, 2022, preprint: not peer reviewed.
- Altschul SF, Madden TL, Schäffer AA *et al.* Gapped blast and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389–402.
- Anfinsen CB. Principles that govern the folding of protein chains. *Science* 1973;**181**:223–30.
- Baek M, DiMaio F, Anishchenko I *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021;**373**:871–6.
- Bahdanau D, Cho K, Bengio Y *et al.* Neural machine translation by jointly learning to align and translate. *arXiv*, arXiv:1409.0473, 2014, preprint: not peer reviewed.
- Browne WJ, North ACT, Phillips DC *et al.* A possible three-dimensional structure of bovine α -lactalbumin based on that of hen's egg-white lysozyme. *J Mol Biol* 1969;**42**:65–86.

- Case DA, Cheatham TE, Darden T *et al.* The amber biomolecular simulation programs. *J Comput Chem* 2005;26:1668–88.
- Cheng S, Zhao X, Lu G *et al.* Fastfold: reducing alphafold training time from 11 days to 67 hours. arXiv, arXiv: 2203.00854, 2022, preprint: not peer reviewed.
- Croll TI, Sammito MD, Kryshchukovych A *et al.* Evaluation of template-based modeling in CASP13. *Proteins Struct Funct Bioinform* 2019;87:1113–27.
- Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 1998;282:740–4.
- Ekeberg M, Lökvist C, Lan Y *et al.* Improved contact prediction in proteins: using pseudolikelihoods to infer potts models. *Phys Rev E* 2013;87:012707.
- Fang X, Wang F, Liu L *et al.* A method for multiple-sequence-alignment-free protein structure prediction using a protein language model. *Nat Mach Intell* 2023;5:1087–96.
- Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. *Adv Neural Info Process Syst* 2020;33:6840–51.
- Hong SH, Joung I, Flores-Canales JC *et al.* Protein structure modeling and refinement by global optimization in CASP12. *Proteins Struct Funct Bioinform* 2017;86:122–35.
- Huang J, Rauscher S, Nawrocki G *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* 2016;14:71–3.
- Johnson LS, Eddy SR, Portugaly E *et al.* Hidden markov model speed heuristic and iterative hmm search procedure. *BMC Bioinformatics* 2010;11:431–8.
- Joo K, Lee J, Lee S *et al.* High accuracy template based modeling by global optimization. *Proteins Struct Funct Bioinform* 2007;69:83–9.
- Jumper J, Evans R, Pritzel A *et al.* Highly accurate protein structure prediction with alphafold. *Nature* 2021;596:583–9.
- Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–637.
- Krogh A, Brown M, Mian IS *et al.* Hidden Markov Models in computational biology: applications to protein modeling. *J Mol Biol* 1994;235:1501–31.
- Kwon S, Won J, Kryshchukovych A *et al.* Assessment of protein model structure accuracy estimation in CASP14: old and new challenges. *Proteins Struct Funct Bioinform* 2021;89:1940–8.
- Lee J, Scheraga HA, Rackovsky S *et al.* New optimization method for conformational energy calculations on polypeptides: conformational space annealing. *J Comput Chem* 1997;18:1222–32.
- Lee SJ, Joo K, Sim S *et al.* Crfalign: a sequence-structure alignment of proteins based on a combination of hmm-hmm comparison and conditional random fields. *Molecules* 2022;27:3711.
- Li Z, Liu X, Chen W *et al.* Uni-fold: an open-source platform for developing protein folding models beyond alphafold. bioRxiv, <https://doi.org/10.1101/2022.08.04.502811>, 2022, preprint: not peer reviewed.
- Lin Z, Akin H, Rao R *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* 2023;379:1123–30.
- Liu J, Guo Z, Wu T *et al.* Improving alphafold2-based protein tertiary structure prediction with multicom in CASP15. *Commun Chem* 2023;6:188.
- Liu Y, Chen L, Liu H *et al.* De novo protein backbone generation based on diffusion with structured priors and adversarial training. bioRxiv, <https://doi.org/10.1101/2022.12.17.520847>, 2022, preprint: not peer reviewed.
- Mariani V, Kiefer F, Schmidt T *et al.* Assessment of template based protein structure predictions in CASP9. *Proteins Struct Funct Bioinform* 2011;79:37–58.
- McGuffin LJ, Edmunds NS, Genc AG *et al.* Prediction of protein structures, functions and interactions using the intFOLD7, multifold and modfolddock servers. *Nucleic Acids Res* 2023;51:W274–80.
- Meier A, Söding J. Context similarity scoring improves protein sequence alignments in the midnight zone. *Bioinformatics* 2014;31:674–81.
- Millán C, Keegan RM, Pereira J *et al.* Assessing the utility of CASP14 models for molecular replacement. *Proteins Struct Funct Bioinform* 2021;89:1752–69.
- Mirdita M, von den Driesch L, Galiez C *et al.* Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* 2017;45:D170–6.
- Mirdita M, Schütze K, Moriwaki Y *et al.* Colabfold: making protein folding accessible to all. *Nat Methods* 2022;19:679–82.
- Morcos F, Pagnani A, Lunt B *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci USA* 2011;108:E1293–301.
- Nolle T, Seeliger A, Thoma N *et al.* Deepalign: alignment-based process anomaly correction using recurrent neural networks. In: Dustdar S, Yu E, Salinesi C *et al.* (eds), *Advanced Information Systems Engineering: 32nd International Conference, CAiSE 2020*, June 8–12, 2020. Grenoble, France: Springer, 2020, 319–33.
- Oda T. Improving protein structure prediction with extended sequence similarity searches and deep-learning-based refinement in CASP15. *Proteins Struct Funct Bioinform* 2023;91:1712–23.
- Olechnovič K, Kulberkytė E, Venclovas C *et al.* Cad-score: a new contact area difference-based function for evaluation of protein structural models. *Proteins Struct Funct Bioinform* 2013;81:149–62.
- Ovchinnikov S, Park H, Kim DE *et al.* Protein structure prediction using rosetta in CASP12. *Proteins Struct Funct Bioinform* 2017;86:113–21.
- Pearce R, Zhang Y. Toward the solution of the protein structure prediction problem. *J Biol Chem* 2021;297:100870.
- Peng J, Xu J. Boosting protein threading accuracy. In: Batzoglu S (ed.), *Research in Computational Molecular Biology: 13th Annual International Conference, RECOMB 2009*, May 18–21, 2009. Tucson, AZ, USA: Springer, 2009, 31–45.
- Peng Z, Wang W, Hong W *et al.* Improved protein structure prediction with trRosettaX2, alphafold2, and optimized msas in CASP15. *Proteins Struct Funct Bioinform* 2023;91:1704–11.
- Pereira J, Simpkin AJ, Hartmann MD *et al.* High-accuracy protein structure prediction in CASP14. *Proteins Struct Funct Bioinform* 2021;89:1687–99.
- Read RJ, Chavali G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. *Proteins Struct Funct Bioinform* 2007;69:27–37.
- Remmert M, Biegert A, Hauser A *et al.* Hhblits: lightning-fast iterative protein sequence searching by hmm-hmm alignment. *Nat Methods* 2012;9:173–5.
- Richardson L, Allen B, Baldi G *et al.* MGnify: the microbiome sequence data analysis resource in 2023. *Nucleic Acids Res* 2023;51:D753–9.
- Robustelli P, Piana S, Shaw DE *et al.* Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci USA* 2018;115:E4758–66.
- Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993;234:779–815.
- Senior AW, Evans R, Jumper J *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* 2020;577:706–10.
- Song Y, Sohl-Dickstein J, Kingma DP *et al.* Score-based generative modeling through stochastic differential equations. *International Conference on Learning Representations*, 2021.
- Steinegger M, Mirdita M, Söding J *et al.* Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nat Methods* 2019;16:603–6.
- Suzek BE, Wang Y, Huang H *et al.*; UniProt Consortium. Uniref clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 2015;31:926–32.
- Trippe BL, Yim J, Tischer D *et al.* Diffusion probabilistic modeling of protein backbones in 3d for the motif-scaffolding problem. *International Conference on Learning Representations*, 2023.
- Vaswani A, Shazeer N, Parmar N *et al.* Attention is all you need. *Adv Neural Inf Process Syst* 2017;30. https://papers.nips.cc/paper_files/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

- Wang S, Sun S, Li Z *et al.* Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol* 2017a;**13**: e1005324.
- Wang S, Sun S, Xu J *et al.* Analysis of deep learning methods for blind protein contact prediction in CASP12. *Proteins Struct Funct Bioinform* 2017b;**86**:67–77.
- Watson JL, Juergens D, Bennett NR *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* 2023;**620**:1089–100.
- Weissenow K, Heinzinger M, Rost B *et al.* Protein language-model embeddings for fast, accurate, and alignment-free protein structure prediction. *Structure* 2022;**30**:1169–77.e4.
- Wu F, Jing X, Luo X *et al.* Improving protein structure prediction using templates and sequence embedding. *Bioinformatics* 2022;**39**:btac723.
- York RL, Yousefi PA, Bogdanoff W *et al.* Structural, mechanistic, and antigenic characterization of the human astrovirus capsid. *J Virol* 2016;**90**:2254–63.
- Zhang C, Mortuza SM, He B *et al.* Template-based and free modeling of i-TASSER and QUARK pipelines using predicted contact maps in CASP12. *Proteins Struct Funct Bioinform* 2017;**86**:136–51.
- Zhang J, Liang Y, Zhang Y *et al.* Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure* 2011;**19**:1784–95.
- Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 2005;**33**:2302–9.
- Zheng W, Wuyun Q, Freddolino PL *et al.* Integrating deep learning, threading alignments, and a multi-msa strategy for high-quality protein monomer and complex structure prediction in CASP15. *Proteins Struct Funct Bioinform* 2023;**91**: 1684–703.
- Zhong B, Su X, Wen M *et al.* Parafold: paralleling alphafold for large-scale predictions. *International Conference on High Performance Computing in Asia-Pacific Region Workshops*, 2021.