2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI) | 979-8-3503-1129-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/CCCAI59026.2023.00025

A benchmark study of protein folding algorithms on nanobodies

Shibo Liang
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China.
201905100101@stumail.sztu.edu.cn

Ziquan Liang
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China.
201902010219@stumail.sztu.edu.cn

Zecheng Wu
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China.
201902010314@stumail.sztu.edu.cn

Feijuan Huang
Shenzhen Institute of Translational Medicine,Shenzhen 2nd People's Hospital, Shenzhen, China. fjhuang@email.szu.edu.cn

Xu Wang
College of Communication Engineering, Jilin University, Changchun, Jilin, China.
wangxu2020@mails.jlu.edu.cn

Yuanzhe Cai
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China.
caiyuanzhe@sztu.edu.cn

Bingding Huang
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China.
huangbingding@sztu.edu.cn

Xin Wang
College of Big Data and Internet
Shenzhen Technology University
Shenzhen, China.
wangxin@sztu.edu.cn
ORCID: 0000-0001-8866-5223

***Abstract*—Nanobodies, also known as single-domain or VHH antibodies, are recombinant variable domains derived from heavy-chain-only antibodies. They exhibit desirable characteristics, including small size, high solubility, exceptional stability, rapid blood clearance, and deep tissue penetration, rendering them valuable tools for disease diagnosis and treatment. In recent years, several deep-learning-based methods for protein structure prediction have been developed, requiring only protein sequences as input. Notable examples include AlphaFold2, RoseTTAFold, DeepAb, NanoNet, and tFold, which have demonstrated remarkable performance in protein or antibody/nanobody prediction. In this study, we analyzed 60 nanobody samples with known experimental 3D structures from the Protein Data Bank (PDB). The accuracy of these algorithms was assessed using two metrics: RMSD and TM-score. Our findings revealed that NanoNet and tFold, particularly NanoNet, exhibit outstanding performance.**

Keywords—Nanobody; benchmark; protein folding; algorithms.

I. INTRODUCTION

In 1993, Belgian scientists discovered a new type of antibody in camelids, which was later found in dromedaries and alpacas as well[1]. This antibody structure lacked the L chain C region, the H chain C region (CH1), and two L chain V regions, and was only composed of two H chain V regions and H chain C regions (CH2 and CH3). This distinct structure of the antibody led to it being referred to as a heavy-chain antibody[1,2]. The V domain of H chain antibodies (VHH)[3], also known as nanobodies (Nbs), has a molecular mass of approximately 15 kDa based on in vitro recombinant protein expression, which is one-tenth that of conventional antibodies, while still retaining the total antigen-binding capacity[4]. Nbs have several advantages over conventional antibody Fab and single chain antibody fragments, including weaker immunogenicity, lower production cost, greater water solubility, superior tissue permeability, and stability[4,5]. These properties enable Nbs to reach sites that are inaccessible to conventional antibodies, such as inside tumor cells[6] and the blood-brain barrier[7], thereby expanding their potential applications.

Currently, the experimental determination of protein structures is carried out using costly and time-consuming techniques such as X-ray crystallography,[8] cryo-electron microscopy (cryo-EM),[9] and nuclear magnetic resonance.[10] Over the last six decades, these labor-intensive techniques have succeeded in determining the structures of only ~170,000 out of the 200 million known proteins across all organisms.[11] Therefore, developing a method capable of accurately predicting protein structure from amino acid sequences alone would be of immense benefit to many biomedical research fields, including antibody drug discovery. In recent years, deep-learning-based protein structure prediction algorithms have emerged that can predict 3D protein structures from 2D protein sequences. Among them, AlphaFold2 (AF2) was the first to achieve relatively high accuracy.[12]

AF2[12] is a deep-learning-based program developed by the Google DeepMind Team, which emerged victorious in the protein structures competition at Critical Assessment of Structure Prediction (CASP) 14 in November 2020. Differing from experimentally obtained structures by only an average of one atom's width, AF2 reached a comparable level of protein structure prediction to humans using sophisticated instruments such as cryo-EM.

AF2's algorithm can be divided into three stages: feature extraction, encoding, and decoding. Upon being given a protein's amino acid sequence, AF2 searches the database for homologous sequences to obtain a feature representation of the sequence and amino acid pairs. Next, the encoder constructs multiple sequence alignment (MSA)[13] and pair matrices, and the information in the two matrices is updated. Finally, the

979-8-3503-1129-7/23/$31.00 ©2023 IEEE
DOI 10.1109/CCCAI59026.2023.00025

encoder encodes the protein structure's 3D coordinates using relative positions.

RoseTTAFold[14], developed by David Baker's team at the University of Washington, was first published online in July 2021. While its network architecture is generally similar to AF2, it achieves comparable performance through a different approach. RoseTTAFold uses a three-track network that sequentially transforms and integrates information at the 1D sequence level, 2D distance map level, and 3D coordinate level.

DeepAb[15], developed by Jeffrey A. Ruffolo's team at Johns Hopkins University, was first published online in June 2021. Its algorithm comprises two phases. The first phase is a deep residual convolutional network (ResNet)[16] and a long and short-term memory network (LSTM)[17] that predict the relative distance and direction (2D structure) between residual pairs in the V region (Fv). The second phase is a distance and angle-constrained approach based on the Rosetta minimization protocol for predicting 3D structures. The network only requires H and L chain sequences as input and is designed with interpretable attention components to provide insight into the model's prediction regions.

NanoNet[18], developed by Tomer Cohen's team at the Hebrew University of Jerusalem, is the first modeling approach optimized for VHH. The model was constructed by training the neural network using a large amount of data from the Ab's VH domain, the Vβ domain of the T cell receptor (TCR), and VHH[19].

TFold is a protein folding component of Tencent's iDrug, which is an artificial intelligence (AI) Lab. The tFold algorithm incorporates three innovative techniques to enhance modeling accuracy. Firstly, it uses multi-source fusion technology to mine the co-evolutionary multiple sequence alignment (MSA) information. Secondly, a deep cross-attention residual network is employed to significantly improve the prediction of crucial 2D structural details, such as residue-residue distances and orientation matrices. Finally, it utilizes a novel template-based free modeling method to combine the structural information in the 3D models obtained through free and template-based modeling, thereby considerably enhancing the accuracy of the final 3D model[20].

To benchmark the five protein prediction algorithms for Nb 3D structures accurately and visually, we used two popular methods to calculate and analyze their strengths and weaknesses: root mean squared difference (RMSD) and template modeling I scores(TM-score)[22]. The comparison between the experimental (gold standard) and prediction results is shown in Figure 3.

Root mean squared deviation (RMSD) is a widely used parameter for evaluating the similarity between protein structures by measuring the difference in atomic positions.[21] It is important to note that the RMSD calculation requires both protein structure files to have the same atomic order and number. However, NanoNet produces a protein structure file that only contains information for five atoms (N, Cα, C, O, Cβ), resulting in an unequal number of atoms compared to the experimental file that contains all atomic coordinates. In order to include NanoNet in the comparison of RMSD differences among the five methods and the experimental file, we generated protein structure file duplicates that only contain Cα atomic coordinates from the complete atomic protein structure prediction files of the other four methods. Subsequently, we computed their backbone+Cβ RMSD by employing these backbone+Cβ files (Fig. 4B).

The TM-score is a measure that provides an estimate of the similarity between two protein structures,[22] and was developed as an alternative to the commonly used RMSD measure. Unlike RMSD, the TM-score is intended to be a more reliable measure of the overall similarity between protein structures, especially when comparing full-length structures.[23]

II. Results

A. All-atom

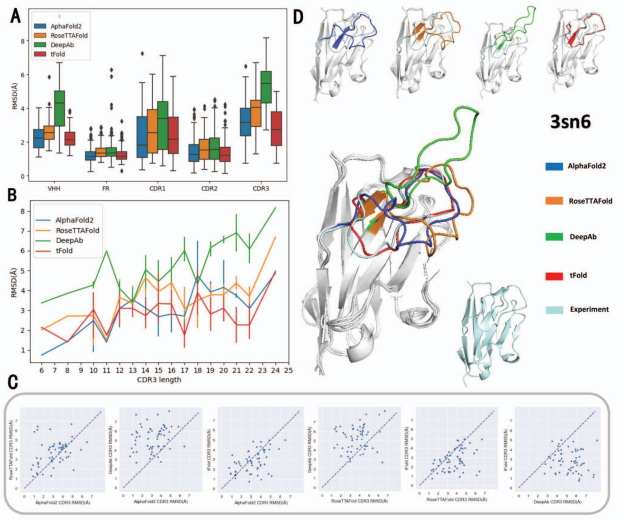

Fig. 1. RMSD and structure comparison of the four all-atom prediction methods with the experimental structure. (A) Boxplot of RMSD for the full VHH structure, FR, CDR1, CDR2, and CDR3. (B) Length comparison of CDR3 loop RMSD in the line graph. (C) Scatterplots of pairwise RMSD comparisons for CDR3 loops among the four different methods. (D) Visual comparison of the predicted 3D Nb structures in PyMOL.

The DeepAb method had the poorest RMSD for predicting the VHH with a mean of 4.19±1.39 Å (Table 1). However, the mean RMSD of all four methods did not differ significantly in the FR, which is more structurally stable compared to the CDRs (Table 1). Predictions of the FR spatial structure were more accurate for all four methods.

In terms of the CDR1 loop region, AF2 and tFold had the best results with mean RMSD of 2.35±1.54 Å and 2.44±1.35 Å, respectively (Table 1). In contrast, DeepAb and RoseTTAFold performed the worst, with mean RMSD of 3.14±1.60 Å and 2.80±1.52 Å, respectively (Table 1).

TABLE I. Mean RMSD between experimental and all-atom predicted whole VHH and FR, CDR1, CDR2, and CDR3 structures for each method.

Method	VHH	FR	CDR1	CDR2	CDR3
AlphaFold2	2.25(±0.69)	1.19(±0.42)	2.35(±1.54)	1.52(±0.99)	3.17(±1.32)
RoseTTAFold	2.59(±0.77)	1.42(±0.40)	2.80(±1.52)	1.69(±0.95)	3.74(±1.27)
DeepAb	4.19(±1.39)	1.48(±0.64)	3.14(±1.60)	1.78(±1.05)	5.28(±1.28)
tFold	2.20(±0.59)	1.25(±0.39)	2.44(±1.35)	1.45(±0.96)	2.79(±1.14)

The RMSD values for CDR2 were relatively low among all four methods, and did not differ significantly (Fig. 1A and Table 1). Furthermore, CDR2 had the smallest average length compared to CDR1 and CDR3 in the test dataset.

CDR3 loop region, which has longer residue lengths and a highly variable nature, was more challenging to predict. Here, tFold had the best performance with a mean RMSD of 2.79±1.14 Å, followed by AF2 with a mean RMSD of 3.17±1.32 Å. In contrast, DeepAb had the poorest performance with a mean RMSD of 5.28±1.28 Å, followed by RoseTTAFold with a mean RMSD of 3.74±1.27 Å (Table 1).

We also investigated the relationship between CDR length and RMSD by analyzing the CDR3 loop length and its corresponding RMSD (Fig. 1B). We observed a positive correlation between CDR3 length and RMSD, where longer CDR3 loops resulted in larger RMSD, and this trend was observed for all methods.

In addition, we conducted one-to-one comparisons of RMSD for CDR3 loops among the four methods (Fig. 1C). Comparing AF2 with RoseTTAFold, we found that the results were mostly clustered around the diagonal, with AF2 performing slightly better overall. When comparing AF2 with DeepAb, we found that AF2 outperformed DeepAb. Similarly, tFold performed better than AF2, whereas RoseTTAFold performed better than DeepAb. Finally, comparing tFold with DeepAb, we found that tFold significantly outperformed DeepAb.

We also employed PyMOL to visualize the predicted structure of each method. Figure 1D highlights the CDR3 loop region where DeepAb's prediction differed significantly from the experimental structure, while tFold's predicted structure for the same region was largely consistent with the experimental structure.

B. Backbone+Cβ

The boxplots comparing the Backbone+Cβ RMSD for all methods are presented in Figure 2A. The results showed that DeepAb performed the worst, while AF2, tFold, and NanoNet demonstrated superior performance. Although NanoNet performed best on the full VHH structure, it also exhibited more outliers.

In the CDR1 loop region, NanoNet had the best performance (mean RMSD of 1.32±0.92 Å), followed by AF2 with a mean RMSD of 1.37±1.16 Å (Table 2). Conversely, DeepAb remained the worst performing method, with a mean RMSD of 1.96±1.16 Å, followed by RoseTTAFold with a mean RMSD of 1.72±1.03 Å. tFold demonstrated a middle-performing method with a mean RMSD of 1.58±0.97 Å.

In the CDR2 loop region (Fig. 2A), tFold was the best performing method with a mean RMSD of 0.72±0.71 Å, followed by NanoNet with a mean RMSD of 0.79±0.67 Å. DeepAb was the worst performing method with a mean RMSD of 1.06±0.86 Å, followed by RoseTTAFold with a mean RMSD of 0.90±0.71 Å. AF2 was in the middle with a mean RMSD of 0.77±0.77 Å.

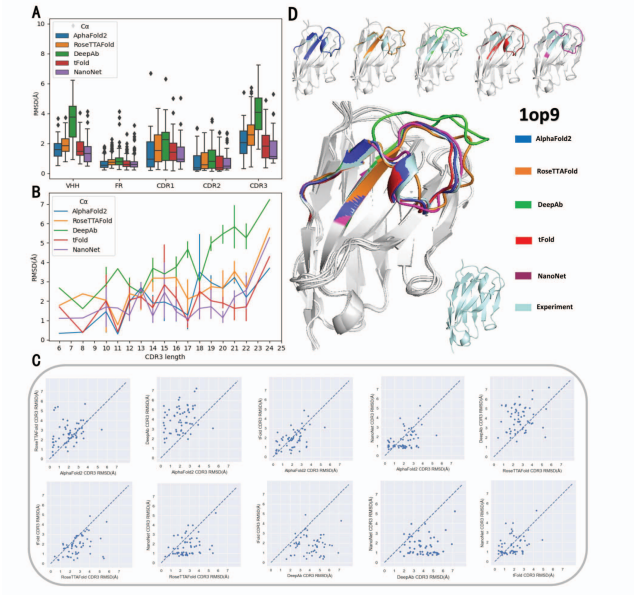

Fig. 2. RMSD and Backbone+Cβ structure comparison of the five methods with the experimental structure. (A) Boxplot of RMSDs for the full VHH structure and FR, CDR1, CDR2, and CDR3. (B) Length comparison of CDR3 loops RMSD in the line graph. (C) Scatterplots of pairwise RMSD comparisons for CDR3 loops among the five different methods. (D) Visual comparison of the predicted Nb 3D structures in PyMOL.

The CDR3 loop region (Fig. 2A) exhibited the best performance by NanoNet, with a mean RMSD of 1.67±0.97 Å. However, a large RMSD outlier was also observed in the CDR3 loop (Fig. 2A). tFold was the second-best performing method, with a mean RMSD of 1.86±0.99 Å. Conversely, DeepAb remained the worst performing method, with a mean RMSD of 4.04±1.28 Å, followed by RoseTTAFold with a mean RMSD of 2.68±1.13 Å. AF2, with a mean RMSD of 2.07±1.12 Å, was the method that performed at the median level (Table 2).

TABLE II. MEAN RMSD BETWEEN THE EXPERIMENTAL AND BACKBONE+Cβ PREDICTED WHOLE VHH AND FR, CDR1, CDR2, AND CDR3 STRUCTURES FOR EACH METHOD. TABLE TYPE STYLES

Method	VHH	FR	CDR1	CDR2	CDR3
AlphaFold2	1.60(±0.65)	0.65(±0.33)	1.37(±1.16)	0.77(±0.77)	2.07(±1.12)
RoseTTAFold	1.90(±0.66)	0.80(±0.34)	1.72(±1.03)	0.90(±0.71)	2.68(±1.13)
DeepAb	3.62(±1.52)	0.88(±0.51)	1.96(±1.16)	1.06(±0.86)	4.04(±1.28)
tFold	1.63(±0.66)	0.68(±0.34)	1.58(±0.97)	0.72(±0.71)	1.86(±0.99)
NanoNet	1.41(±0.82)	0.67(±0.42)	1.32(±0.92)	0.79(±0.67)	1.67(±0.97)

The comparison of RMSD with CDR3 loop length (Fig. 2B) indicated that NanoNet outperformed the other methods. Likewise, a comparison of CDR3 loop RMSD among the different methods (Fig. 2C) demonstrated that NanoNet achieved the best performance. However, the differences

between NanoNet, tFold, and AF2 are minimal, as most points fall along the diagonal in these comparisons. DeepAb consistently showed the poorest performance among all methods, while AF2 outperformed RoseTTAFold albeit by a small margin. NanoNet exhibited better performance than AF2, and tFold's performance was similar to AF2's. Additionally, tFold showed superior performance compared to RoseTTAFold, and NanoNet exhibited superior performance to RoseTTAFold. Furthermore, NanoNet performed slightly better than tFold.

C. TM-score

In the comparison of all-atom structures, tFold had the highest TM-scores, followed closely by AF2, while DeepAb performed the worst and RoseTTAFold was in the middle (Fig. 3A), consistent with the RMSD comparisons. On the other hand, NanoNet outperformed the other methods in the comparison of Backbone+Cβ structures (Fig. 3B).

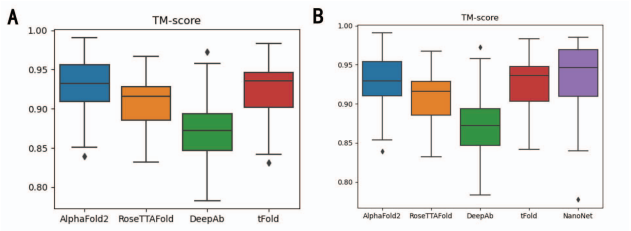

Fig. 3. Boxplot comparison of TM-scores for all-atom and Backbone+Cβ structures for each method. (A) Boxplots of TM-scores for all-atom structures predicted by each method. (B) Boxplots of TM-scores for the Backbone+Cβ structures predicted by each method.

III. Materials and Methods

A. Obtaining the PDB experimental file

We retrieved the amino acid sequences and experimental PDB files of Nbs from the official PDB website,[25] eliminating any duplicates and retaining the entries with the highest resolution. Subsequently, we utilized the five structure prediction algorithms to generate prediction PDB files based on the Nb amino acid sequences. Finally, we compared the predicted PDB files with the experimental ones to calculate the RMSD and TM-scores, using the latter as the reference standard (Fig. 4).

B. RMSB calculation

The experimental and prediction PDB files were analyzed using PyMOL (https://github.com/schrodinger/pymol-open-source) to compute the RMSD for each prediction algorithm. PyMOL aligns the atoms between the predicted and experimental PDB files automatically. By default, PyMOL removes atoms with large distance differences before computing the RMSD using the outlier rejection cutoff of 2.0 Å in PyMOL's align parameter. This filtering threshold can remove all abnormal values, resulting in indistinguishable RMSD outcomes. To ensure a fair and reasonable comparison while retaining the maximum number of atoms in the experimental PDB file for the RMSD calculation, we set the cutoff to 10. This threshold prevented the removal of atoms with differences up to 10 and ensured that the RMSD results were accurate and reliable.

C. The main flow of the method

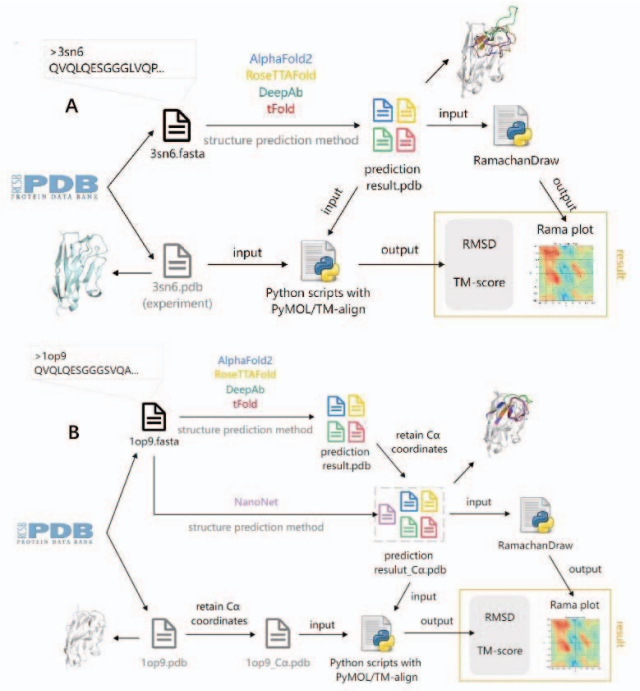

Fig. 4. A schematic diagram of the two comparisons. (A) Schematic of how scores were calculated for all-atom protein models. (B) Schematic of how scores were calculated for Backbone+Cβ protein models.

D. CDR loops segmentation by the international immunogenetics information system (IMGT) scheme

The Nb protein sequences were divided into seven regions (FR1, CDR1, FR2, CDR2, FR3, CDR3, and FR4) according to the IMGT annotation method.[26]

E. PDB file partitioning based on CDR loops

The seven regions were used to segment the experimental PDB files obtained from the official PDB website. The processed experimental structure PDB files containing Backbone+Cβ atoms were also segmented into CDR regions. The segmentation process generated seven separate PDB files for each FR or CDR loop region, resulting in a total of seven PDB files for Nb.

F. TM-score calculation

The TM-score was calculated using TM-align.[22] The experimental and predicted PDB files were passed to the TM-align method to calculate the TM score.

IV. Discussion

A. Performance on predictive accuracy

Our study evaluated the performance of five deep learning Nb-folding algorithms in predicting VHH structures, and the results indicated that NanoNet had the highest accuracy, followed by tFold, AF2, and RoseTTAFold, while DeepAb performed the worst based on RMSD (as shown in Figs. 1A and 2A). The CDR loop regions, especially CDR3 loops, were the main regions where prediction errors occurred

and had a significant impact on the overall prediction results, followed by CDR1 loops.

Notably, NanoNet not only predicted the overall framework but also the local variable regions most accurately (as shown in Fig. 2A-B and Table 2). Furthermore, we obtained the same ranking order of the algorithms when using TM-scores (Fig. 3A-B).

B. Computational time

It is noteworthy that NanoNet's algorithm exhibits the highest speed for model prediction, with a mere 4.65 seconds required for predicting 60 input sequences due to the parallel processing capacity of NanoNet. Conversely, using DeepAb on a local server with four NVIDIA Tesla V100 SXM2 graphics processing units in single-chain mode, the average prediction time for a single Nb sequence was approximately 3 minutes, whereas the average prediction time for the same Nb sequence using AF2's Colab notebook was around 5 minutes (https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AF2.ipynb). Furthermore, using RoseTTAFold's online Robetta platform (https://robetta.bakerlab.org, last accessed: July 14, 2022), the average prediction time for a single Nb sequence ranged from 1 to 55 minutes, with each account limited to 20 predictions per day. In comparison, using the tFold platform (https://drug.ai.tencent.com/console/en/tfold?type=predict, last accessed: July 28, 2022) typically required 20-30 minutes for predicting a single Nb sequence, although occasionally the duration could extend to 9-16 hours. Regrettably, this online platform permits only 10 tasks per account per day, and only one outcome could be downloaded every 24 hours, greatly hindering its efficiency.

C. Robustness

Regarding the algorithm's functional level, NanoNet solely enables the spatial prediction of five atoms, including nitrogen (N), Cα, carbon (C), oxygen (O), and beta-carbon (Cβ), for each amino acid backbone structure of the Nb (VHH) formed by the VH region of the common Y-shaped lgG Ab. Conversely, DeepAb allows for predicting the VL and VH double chains of the AB's Fv, or only the single chain VH or VL. In contrast, the other three folding algorithms (AF2, RoseTTAFold, and tFold) enable the prediction of the entire atomic protein structure of the Ab.

Our folding predictions with Nb 1I3U demonstrated that DeepAb, RoseTTAFold, and tFold could not accommodate wildcard "X" characters in the amino acid sequence. Consequently, the "X" generated a 'ValueError' that halted the program. To address this issue, we removed the invalid 'X' character from the FASTA sequence. Moreover, during our examination of the input amino acid sequences, we discovered that some Nb FASTA files downloaded from the PDB database had polyhistidine tags at their C-terminal end. If these files were used directly for folding structure prediction, NanoNet's RMSD prediction error would be 2.87±1.97 Å, which is considerably higher than when the tag was deleted (RMSD=1.38±0.82 Å).

D. Relevance of the predicted outcomes and algorithm differences

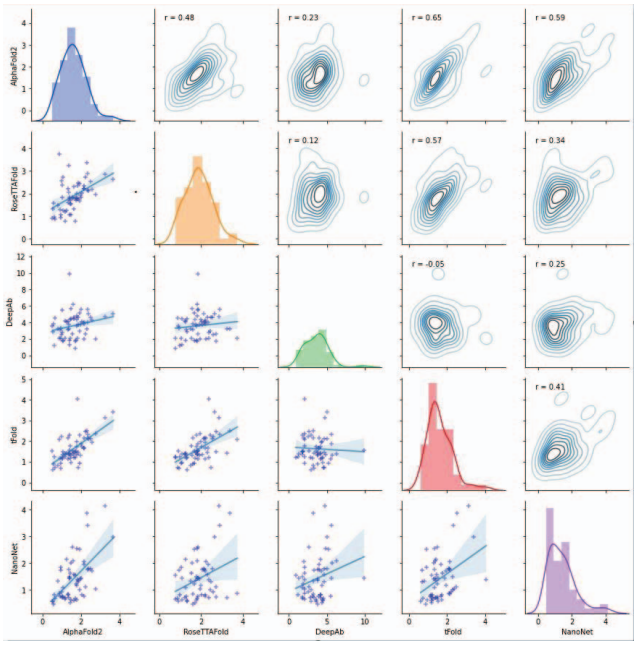

Fig. 5. Pairwise Pearson's correlations between Nb structure predictions by the five protein folding algorithms.

The scatterplots displayed below the diagonal of the matrix indicate the linear correlation of Root Mean Square Deviations (RMSD) between pairs of protein folding algorithms. On the other hand, the scatterplots above the diagonal of the matrix display the kernel density of data between pairs of protein folding algorithms, along with the corresponding Pearson's correlation coefficient (r). The density of the kernel for two algorithms with higher correlations is more concentrated along the diagonal, as demonstrated by the nuclear density map of AF2 and tFold (located in the top row, second from the right; r=0.65). The RMSD density barplot for each algorithm is exhibited along the diagonal.

The correlation of root-mean-square deviations (RMSD) presents a straightforward yet effective approach for determining the proximity of pairs of algorithms. This is particularly relevant for contemporary machine learning techniques, particularly those related to protein folding, which tend to have opaque inner workings. Our findings indicate that AF2 exhibited the strongest correlation with tFold (r=0.65), followed by AF2 with NanoNet (r=0.59), RoseTTAFold with tFold (r=0.57), and AF2 with RoseTTAFold (r=0.48; see Fig. 5).

E. Conclusions

RoseTTAFold and AF2 represent advancements over the AF1 algorithm. Both algorithms share some similarities in their algorithmic frameworks. However, their results were not strongly correlated, possibly due to differences in their approaches, such as RoseTTAFold's inclusion of a 3D track in the three-track block and variations in database selection. By contrast, AF2 demonstrates a high level of accuracy in

predicting the protein structure, particularly in stable regions such as Nb FRs.

The algorithms used did not yield the best all-atom Nb folding results. Surprisingly, the tFold algorithm, developed by Tencent AI Lab and released before RoseTTAFold and AF2, produced superior results when comparing all-atom structures. Similar to AF2, tFold's folding process is divided into three homology modeling steps.[12] Its predictions showed a high degree of correlation with those of AF2 and RoseTTAFold. Moreover, tFold's accuracy in predicting CDR3s was significantly better than that of AF2 and RoseTTAFold. This suggests that tFold may have been optimized specifically for highly variable regions, such as CDR3.

Although NanoNet is a lightweight folding algorithm that relies on ResNet,[16] its predictions were remarkably similar to those of AF2, which employs more sophisticated structures that incorporate an attention algorithm.[12,24] Table 2 demonstrates that NanoNet achieved the highest accuracy, particularly in CDR3 regions, underscoring its efficacy in predicting Nb folding structures and emphasizing the significance of selecting an appropriate training dataset for this particular subfield. Furthermore, the simplified Backbone+Cβ structure upon which NanoNet's calculations are based contributed to its impressive speed.

The study found that the DeepAb folding algorithm had the poorest performance. Both DeepAb and its predecessor, DeepH3, were originally designed for predicting folding structures of traditional antibody Fv regions, and as such, are not well-suited for predicting the folding structures of nanobodies. This is likely due to the fact that the algorithm was mainly developed with double-stranded antibodies in mind, and its training datasets primarily consisted of double-chained antibodies. Although DeepAb does have a single-chain mode, its performance in predicting nanobody folding structures was still unsatisfactory.

F. Perspectives

New algorithms also continue to emerge in this newly emerging field such as IgFold and ABlooper, prediction speed, and RMSD-based accuracy, which will greatly benefit protein and Nb 3D structure predictions, especially for high-variability such as CDR3.

ACKNOWLEDGMENT

We thank Chunhai Xue, Leixin Zhu, Zelong Li, and Huanqing Long for data collection and all the members of the Wang lab for helpful discussions.

REFERENCES

[1] Hamers-Casterman C, Atarhouch T, Muyldermans S al, et al. Naturally occurring antibodies devoid of light chains. Nature. 1993;363(6428):446-448.

[2] Hassanzadeh-Ghassabeh G, Devoogdt N, De Pauw P, Vincke C, Muyldermans S. Nanobodies and their potential applications. Nanomed. 2013;8(6):1013-1026.

[3] Nguyen VK, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline VHH and specific mechanisms enlarge the antigen-binding repertoire. EMBO J. 2000;19(5):921-930.

[4] Muyldermans S. Single domain camel antibodies: current status. Adv Comb Biol. 2001;74(4):277-302. doi:10.1016/S1389-0352(01)00021-6

[5] Nguyen VK, Desmyter A, Muyldermans S. Functional heavy-chain antibodies in Camelidae. Published online 2001.

[6] Jovčevska I, Muyldermans S. The therapeutic potential of nanobodies. BioDrugs. 2020;34(1):11-26.

[7] Li T, Bourgeois J, Celli S, et al. Cell‐penetrating anti‐GFAP VHH and corresponding fluorescent fusion protein VHH‐GFP spontaneously cross the blood‐brain barrier and specifically recognize astrocytes: application to brain imaging. FASEB J. 2012;26(10):3969-3979.

[8] Ilari A, Savino C. Protein structure determination by x-ray crystallography. Bioinformatics. Published online 2008:63-87.

[9] Milne JL, Borgnia MJ, Bartesaghi A, et al. Cryo‐electron microscopy–a primer for the non-microscopist. FEBS J. 2013;280(1):28-45.

[10] Clore GM, Gronenborn AM. Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscop. Crit Rev Biochem Mol Biol. 1989;24(5):479-564.

[11] McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallogr Sect F Struct Biol Commun. 2014;70(1):2-20.

[12] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2

[13] Pei J. Multiple protein sequence alignment. Curr Opin Struct Biol. 2008;18(3):382-386.

[14] Baek M, DiMaio F, Anishchenko I, et al. Accurate prediction of protein structures and interactions using a 3-track neural network. Science. 2021;373(6557):871-876. doi:10.1126/science.abj8754

[15] Ruffolo JA, Sulam J, Gray JJ. Antibody structure prediction using interpretable deep learning. Patterns. 2022;3(2):100406.

[16] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. Published online December 10, 2015. doi:10.48550/arXiv.1512.03385

[17] Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. ArXiv Prepr ArXiv150801991. Published online 2015.

[18] Cohen T, Halfon M, Schneidman-Duhovny D. NanoNet: Rapid and accurate end-to-end nanobody modeling by deep learning. Front Immunol. 2022;13:958584. doi:10.3389/fimmu.2022.958584

[19] Cohen T, Halfon M, Schneidman-Duhovny D. NanoNet: Rapid End-to-End Nanobody Modeling by Deep Learning at Sub Angstrom Resolution. Bioinformatics; 2021. doi:10.1101/2021.08.03.454917

[20] Zheng L, Lan H, Shen T, et al. tFold-TR: Combining Deep Learning Enhanced Hybrid Potential Energy for Template-Based Modeling Structure Refinement. ArXiv Prepr ArXiv210504350. Published online 2021.

[21] Carugo O. How root-mean-square distance (rmsd) values depend on the resolution of protein structures that are compared. J Appl Crystallogr. 2003;36(1):125-128.

[22] Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. Nucleic Acids Res. 2005;33(7):2302-2309.

[23] Hooft RW, Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot. Bioinformatics. 1997;13(4):425-430.

[24] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. Adv Neural Inf Process Syst. 2017;30.

[25] Sussman JL, Lin D, Jiang J, et al. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Crystallogr D Biol Crystallogr. 1998;54(6):1078-1084.

[26] Lefranc MP, Pommié C, Ruiz M, et al. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev Comp Immunol. 2003;27(1):55-77. doi:10.1016/s0145-305x(02)00039-3