

# Accurate prediction of CDR-H3 loop structures of antibodies with deep learning

Hedi Chen<sup>1</sup>, Xiaoyu Fan<sup>1†</sup>, Shuqian Zhu<sup>1†</sup>, Yuchan Pei<sup>2</sup>, Xiaochun Zhang<sup>1</sup>, Xiaonan Zhang<sup>3</sup>, Lihang Liu<sup>3</sup>, Feng Qian<sup>1\*</sup>, Boxue Tian<sup>1\*</sup>

<sup>1</sup>MOE Key Laboratory of Bioinformatics, State Key Laboratory of Molecular Oncology, School of Pharmaceutical Sciences, Tsinghua University, Beijing, China; <sup>2</sup>Tsinghua Institute of Multidisciplinary Biomedical Research, Tsinghua University, Beijing, China; <sup>3</sup>Department of Natural Language Processing, Baidu International Technology (Shenzhen) Co Ltd, Shenzhen, China

## \*For correspondence:

qianfeng@tsinghua.edu.cn (FQ);  
boxuetian@mail.tsinghua.edu.cn (BT)

<sup>†</sup>These authors contributed equally to this work

**Competing interest:** See page 22

**Funding:** See page 23

## Sent for Review

07 August 2023

## Preprint posted

21 August 2023

## Reviewed preprint posted

18 October 2023

## Reviewed preprint revised

30 January 2024

## Reviewed preprint revised

31 May 2024

## Version of Record published

26 June 2024

**Reviewing Editor:** Volker Dötsch, Goethe University Frankfurt, Germany

© Copyright Chen et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

**Abstract** Accurate prediction of the structurally diverse complementarity determining region heavy chain 3 (CDR-H3) loop structure remains a primary and long-standing challenge for antibody modeling. Here, we present the H3-OPT toolkit for predicting the 3D structures of monoclonal antibodies and nanobodies. H3-OPT combines the strengths of AlphaFold2 with a pre-trained protein language model and provides a 2.24 Å average RMSD<sub>C<sub>α</sub></sub> between predicted and experimentally determined CDR-H3 loops, thus outperforming other current computational methods in our non-redundant high-quality dataset. The model was validated by experimentally solving three structures of anti-VEGF nanobodies predicted by H3-OPT. We examined the potential applications of H3-OPT through analyzing antibody surface properties and antibody–antigen interactions. This structural prediction tool can be used to optimize antibody–antigen binding and engineer therapeutic antibodies with biophysical properties for specialized drug administration route.

## eLife assessment

This article presents H3-OPT, a deep learning method that effectively combines existing techniques for the prediction of antibody structure. This work, supported by **convincing** experiments for validation, is **important** because the method can aid in the design of antibodies, which are key tools in many research and industrial applications.

## Introduction

Antibodies protect the host by recognizing and neutralizing foreign microbes or viruses and provide immunity against future infections. Furthermore, the use of therapeutic antibodies is steadily increasing in clinical treatments for infectious diseases, cancers, and autoimmune disorders (*Meganck and Baric, 2021; Kaplon and Reichert, 2021*). For example, anti-PD-L1 antibodies block PD-L1/PD-1 interactions to inhibit tumor growth in patients (*Brahmer et al., 2012*). A typical monoclonal antibody (mAb) is composed of heavy and light chains, while a nanobody (Nb) has only a single-domain variable heavy chain. The CDR heavy chain 3 (CDR-H3) loop is the most variable region in both length and amino acid sequence, and it plays a central role in antigen binding for both mAbs and Nbs. The development of effective therapeutic antibodies requires the solved structures of candidate antibodies, including the CDR-H3. However, it is both costly and labor intensive to experimentally obtain structures, and thus computationally predicted structures are often used to guide antibody design (*Aguilar Rangel et al.,*

2022). Despite major advances in computational methods, structure prediction for CDR-H3 loops remains challenging (Valdés-Tresanco *et al.*, 2023).

Traditional template-based modeling methods, such as RosettaAntibody (Adolf-Bryfogle *et al.*, 2018), PIGS (Marcatili *et al.*, 2008), ABodyBuilder (Leem *et al.*, 2016), and MODELLER (Eswar *et al.*, 2006), often provide demonstrably inaccurate predictions for CDR-H3 sequences (Almagro *et al.*, 2014; Teplyakov *et al.*, 2014; Almagro *et al.*, 2011), and as a result, alternative artificial intelligence (AI)-based approaches, such as AlphaFold2 (AF2) (Jumper and Hassabis, 2022; Jumper *et al.*, 2021), trRosetta (Du *et al.*, 2021), and RoseTTAFold (Baek *et al.*, 2021), are increasing in popularity. AF2 has shown comparable accuracy to experimentally determined structures by capturing physical and biological information about protein folding, thus providing a versatile deep learning framework for structure prediction. Structure prediction methods using pre-trained protein language models (PLMs), for example, HelixFold-Single (Fang *et al.*, 2022), OmegaFold (Wu *et al.*, 2022a), and ESMFold (Lin *et al.*, 2023), have shown comparable performance to AF2 with accelerated prediction speed. PLMs can be trained with datasets comprising tens of millions of unlabeled protein sequences in a self-supervised manner and can be subsequently applied to a variety of downstream tasks, such as drug-gable protein target prediction (Chen *et al.*, 2023), predicting protein function, and protein design (Hie *et al.*, 2024; Ferruz *et al.*, 2022; Madani *et al.*, 2023). Antibody-specific tools such as IgFold (Ruffolo *et al.*, 2023), tFold-Ab (Wu *et al.*, 2022b), DeepAb (Ruffolo *et al.*, 2022), and NanoNet Cohen *et al.*, 2022 have also been developed to improve accuracy in CDR-H3 prediction. Among them, IgFold leverages sequence representations from PLMs to efficiently predict antibody structures within seconds, and notably, IgFold can provide accuracy comparable to AF2, enabling high-throughput prediction of antibody structures.

In this study, we present H3-OPT, which combines features of AF2 and PLMs to predict antibody structures. We compare H3-OPT with several other antibody structure prediction methods and found that it can provide a lower average RMSD<sub>Cα</sub> for CDR-H3 loops than other algorithms in three subsets of varying difficulty. To further validate our model, we experimentally solved the structures of three anti-VEGF nanobodies predicted by H3-OPT (Zhu *et al.*, 2023). We examined the potential applications of H3-OPT through analyzing antibody surface properties and antibody–antigen interactions. We demonstrate the informative value of high-quality H3 loops for predicting binding affinity and further support the use of H3-OPT as a powerful and versatile tool for studying antigen–antibody interactions. This structural prediction tool can be used to optimize antibody–antigen binding and engineer therapeutic antibodies with biophysical properties.

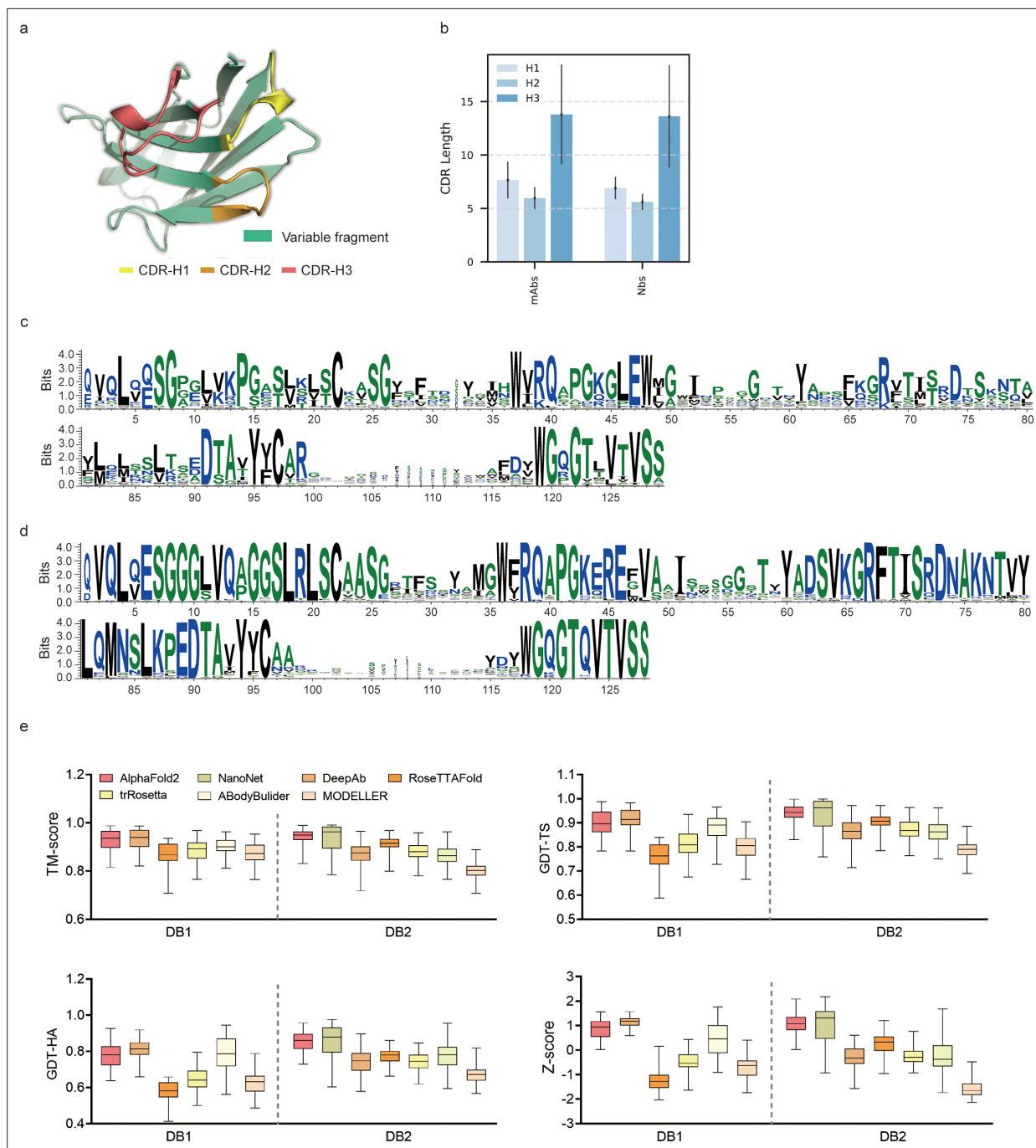
## Results

### High-quality antibody crystal structures in benchmarks

To evaluate the performance of AF2 against currently available methods, we conducted two datasets (DB1 and DB2) with high-resolution (<2.5 Å) X-ray crystal structures from SABdab (Raybould *et al.*, 2020). The CDR length distributions were similar in each dataset (Figure 1b). Additionally, we plotted sequence logos and identified the CDR loops of all heavy chain sequences (Figure 1c and d). These sequence logo plots revealed higher degree of sequence variability of CDR loops, particularly in the CDR-H3, with smaller residue letters than framework regions. Overall, high-quality and diverse antibody datasets allowed us to assess the quality of predicted models and evaluate the strengths and weaknesses of all alternative methods.

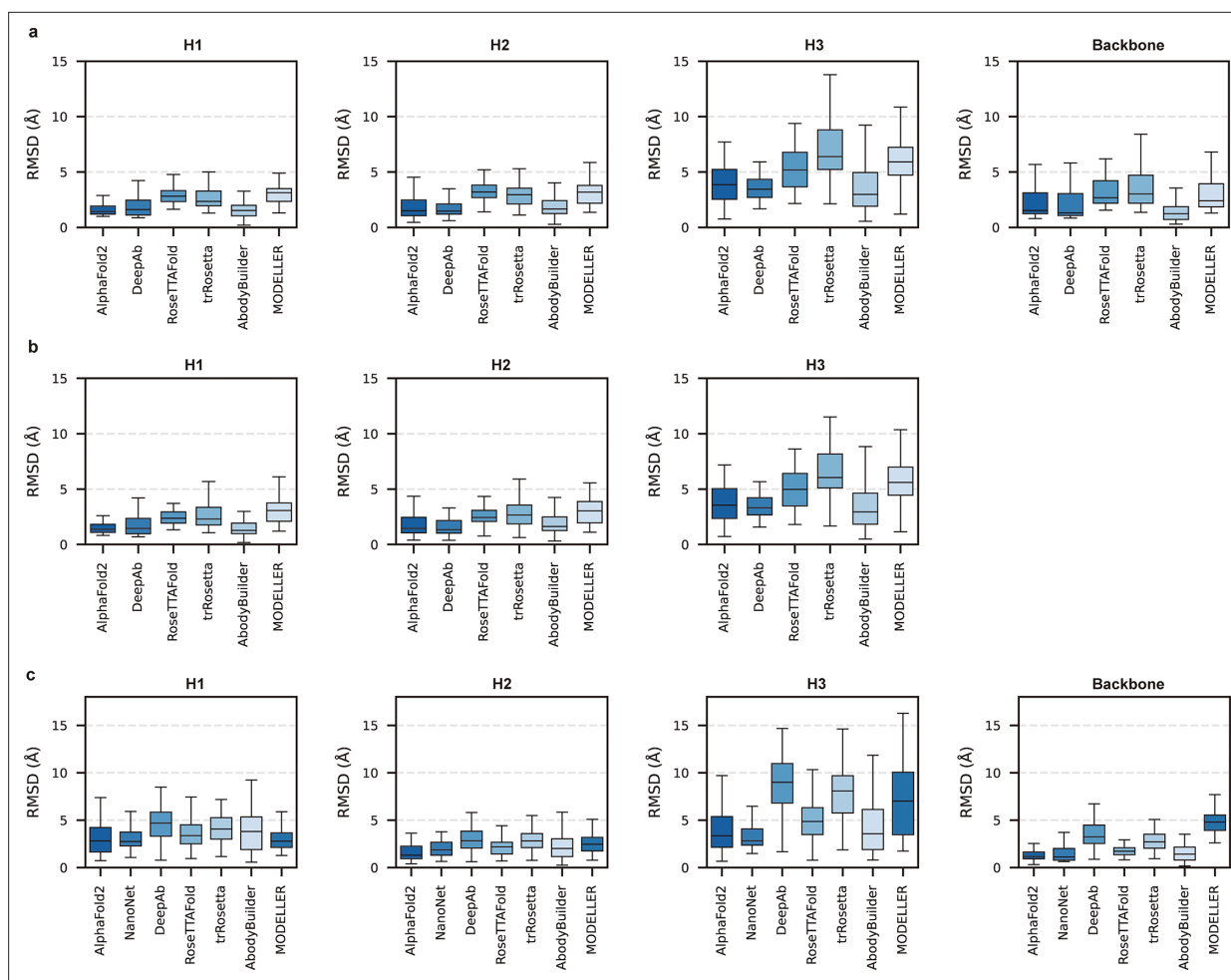
### Accuracy of AF2 for the overall antibody structure predictions was remarkable

To assess the similarity between the predicted models generated by each method and the experimentally determined structure, we first computed the template modeling scores (TM-scores) (Zhang and Skolnick, 2004) for all predicted models of each target. The TM-scores of AF2 predictions in different datasets were as follows: DB1,  $0.93 \pm 0.04$ ; DB2,  $0.94 \pm 0.03$  (Figure 1e). Although the average TM-score of AF2 was slightly lower than those of DeepAb in DB1, it outperformed DeepAb by a large margin in DB2 (0.94 vs 0.87, on average). Then, we used global distance test (GDT) scores (Zemla *et al.*, 1999) to assess the similarity of predicted substructures at different structural cutoffs. The average GDT-TS scores of the models generated by AF2, ABodyBuilder, and DeepAb in DB1 were



**Figure 1.** Accuracy of AF2 on antibody modeling. **(a)** Schematic for CDR heavy chain loops. **(b)** The CDR lengths of monoclonal antibodies (mAbs) ( $n = 47$ ) and nanobodies (Nbs) ( $n = 78$ ). The error bars represent the standard deviation of the data. **(c)** Sequence logo plots of  $V_H$  fragments in DB1. **(d)** Sequence logo plots of  $V_H$  fragments in DB2. The different colors of codes represent the hydrophobicity of amino acids. **(e)** The performance of AF2 on different datasets using various evaluation metrics. In the box plots, the lines at the center of the boxes represent the medians, and the top and bottom lines of the boxes represent the upper and lower quartiles.

0.90, 0.88, and 0.91, respectively. In DB2, the GDT-TS scores of AF2, RoseTTAFold, and NanoNet all exceeded 0.90. Furthermore, we found that the models with high GDT-TS scores also exhibited higher GDT-HA scores compared to the models with low GDT-TS scores. We finally calculated the average Z-scores of GDT-TS, GDT-HA, and TM-score for all datasets to provide comprehensive scores for each method. AF2 outperformed other methods on both DB1 and DB2, with an average Z-score of 0.87 and 1.08, respectively.



**Figure 2.** Accuracy of AF2 on different antibody regions. **(a)** The performance of AlphaFold2 in DB1 relative to other methods after superimposing Fv backbones. **(b)** The performance of H3-OPT in DB1 relative to other methods after superimposing  $V_H$  backbones. **(c)** The performance of H3-OPT in DB2 relative to other methods.

The online version of this article includes the following figure supplement(s) for figure 2:

**Figure supplement 1.** Local accuracy of AlphaFold2 prediction.

## AlphaFold2 accurately predicted the $V_H/V_L$ orientations and CDR-H3 loops of antibodies

Given the vital roles of CDR-H3 loops in antigen recognition, we assessed the local accuracy of CDR-H3 loops among all participants. We first superimposed the backbone atoms of the entire Fv regions to reference structures and calculated the heavy atom RMSDs of CDR-H3 loops ( $\text{RMSD}_{\text{HA}}$ ). The average  $\text{RMSD}_{\text{HA}}$  values of AF2, DeepAb, and ABodyBuilder in DB1 were 3.92, 3.64, and 3.69 Å, respectively, which were lower than those of other methods (**Figure 2a**). Additionally, these  $\text{RMSD}_{\text{HA}}$  values decreased slightly after superimposing the  $V_H$  backbone heavy atoms, indicating that these methods could predict  $V_H/V_L$  orientations more accurately than other alternative methods (**Figure 2b**). Similar to TM-score, DeepAb performed better than AF2 in DB1, but its accuracy decreased substantially in DB2 (**Figure 2c**). AF2, NanoNet, and ABodyBuilder outperformed other methods in DB2, with average  $\text{RMSD}_{\text{HA}}$  values of 3.79, 3.44, and 4.37 Å, respectively.

To better understand the prediction results, we next conducted side-to-side comparisons for the abovementioned methods. The results showed that there were no significant differences in either the backbone or CDR-H3 RMSDs after superimposing the  $V_H$  backbone heavy atoms between DeepAb and AF2 (**Figure 2—figure supplement 1a**). Additionally, AF2 accurately produced the Nb structures with lower backbone RMSDs (all below 4 Å), but provided comparative accuracy in CDR-H3 RMSDs



**Table 1.** The root mean square deviation (RMSD) results of PM6D3 level re-ranking method on 14 same CDR-H3 antibodies.

PDB ID	Ranked 0 RMSD	Lowest energy RMSD	Lowest RMSD	$\Delta_{\text{RMSD}}^*$
4kmt	1.06	1.14	1.05	-0.08
5i19	2.16	1.91	1.77	0.25
5i1l	3.80	3.20	3.19	0.60
5i17	2.86	3.71	2.86	-0.85
5i1d	2.10	2.10	2.02	0.00
5i1c	2.43	1.66	1.45	0.77
5i1a	2.37	0.85	0.59	1.52
5i1i	3.72	3.51	3.51	0.21
5i15	2.16	1.94	1.35	0.22
5i16	3.19	1.70	1.39	1.49
5i18	2.88	2.88	2.88	0.00
5i1e	1.62	1.13	0.92	0.49
5i1g	2.08	2.08	2.00	0.00
5i1h	1.58	1.84	1.32	-0.26

\* $\Delta_{\text{RMSD}}$  was calculated by subtracting the RMSD of predicted model from the RMSD of Ranked\_0 model.

with NanoNet (**Figure 2—figure supplement 1b**). We found that the main reason for this result was that NanoNet predicted wrong main-chain structures at the C-terminus. ABodyBuilder utilized homology modeling-based algorithm for structure prediction, providing similar or better backbone quality than AF2 (**Figure 2—figure supplement 1c and d**). However, it did not significantly improve the accuracy of the CDR3 loops because ABodyBuilder highly relied on the quality of templates (especially when modeling DB2). Taken together, these results indicated that AF2 was a highly effective tool for predicting antibody structures both in mAbs and Nbs, and it produced CDR-H3 loops that were comparable to those generated by AI-based antibody-specific methods.

### Quantum mechanics-based optimization failed to optimize CDR-H3 loops

Due to the lack of consideration of electronic effects in AF2, we hypothesized that the accuracy of CDR-H3 could be further improved by quantum mechanics (QM)-based methods. We applied two QM-based approaches to optimize AF2 models: the first was an energy-based re-ranking method, and the second was loop optimization with QM. For the first approach, we introduced a new dataset

**Table 2.** Accuracy of quantum mechanics (QM)-based re-ranking methods.

Method	Freeze terminal C $\alpha$	CDR*	Phase	Ranked 0 RMSD	Lowest energy RMSD	Lowest RMSD	$\Delta_{\text{RMSD}}$
PM6D3	Y	H3	Gas	2.64	2.76	2.16	-0.12
PM6D3	N	H3	Gas	2.53	2.67	2.03	-0.14
PM6D3	N	H1, H2, H3	Gas	2.50	2.64	2.00	-0.14
B3LYP	N	H3	Gas	2.66	2.87	2.30	-0.21
B3LYP	N	H3	Water	2.66	2.68	2.30	-0.02

RMSD = root mean square deviation.

\*CDR means the energy of which loop is used to re-rank AF2 models.

**Table 3.** Accuracy of quantum mechanics (QM)-based optimization methods.

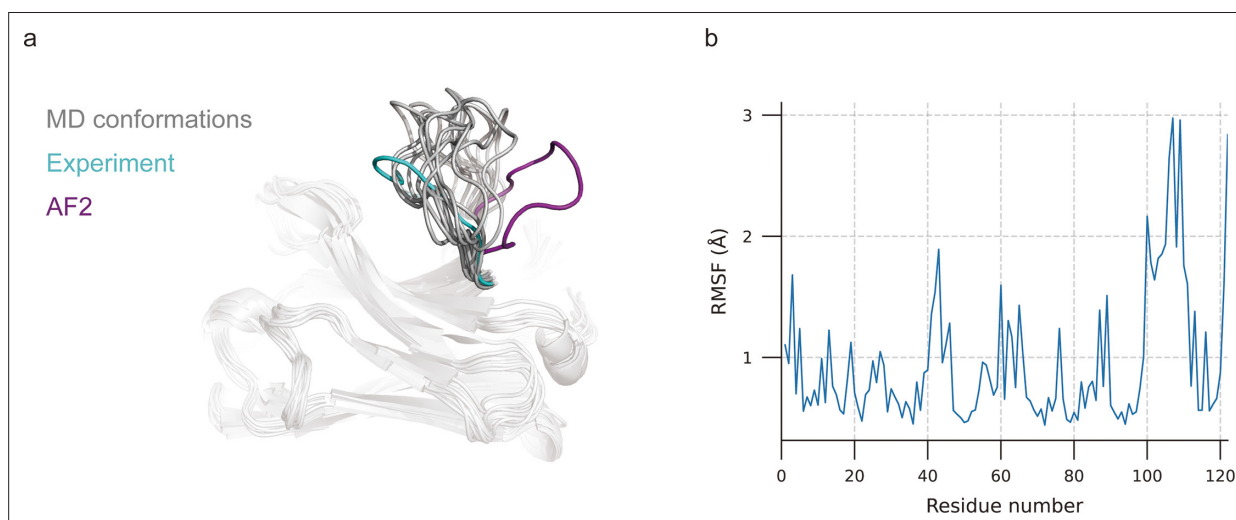
Method	Freeze terminal C $\alpha$	Structure generation method	Phase	Ranked 0 RMSD	Lowest energy RMSD/opted RMSD	Lowest RMSD	$\Delta_{\text{RMSD}}$
PM6D3	Y	/	Gas	1.69	1.74/1.87	1.37	-0.05
B3LYP	N	/	Gas	1.63	1.65/2.55	1.38	-0.02
B3LYP	N	/	Water	1.63	1.58/2.25	1.38	0.05
B3LYP	N	Boltzmann	Gas	1.56	2.05	1.28	-0.49
B3LYP	N	Boltzmann	Water	1.56	1.81	1.28	-0.25
B3LYP	N	Boltzmann, minimized	Gas	1.56	1.96	1.28	-0.40
B3LYP	N	Boltzmann, minimized	Water	1.56	1.84	1.28	-0.28

RMSD = root mean square deviation.

containing 14 antibody sequences with the same CDR-H3 and then run QM optimization for all AF2 models. These models were re-ranked according to the QM energies. As shown in **Table 1**, our results demonstrated that this method outperformed the default ranking criteria for 8 out of 14 targets, achieving an average RMSD improvement of 0.69 Å. However, this re-ranking method did not yield more accurate results when applied to a larger database (DB3) (**Table 2**). We also made additional attempts to improve the re-ranking process by removing constraints on terminal atoms, altering re-ranking criteria, and considering solvent effects. However, these methods did not improve the accuracy of AF2 ( $\Delta_{\text{RMSD}} < 0$ ). Moreover, the QM-based optimization methods also failed to improve the accuracy of H3 loops (**Table 3**). We finally generated structures based on the conformational proportions using Boltzmann distribution of the QM energies, but the accuracy of these structures did not match that of Ranked\_0, with all  $\Delta_{\text{RMSD}}$  below 0. In conclusion, although QM-based methods may not improve the accuracy of CDR-H3 loops in general cases, it still provided opportunities for loop modeling as the development of more accurate physics-based methods.

### Molecular dynamics (MD) simulations could not provide accurate CDR-H3 loop conformations

MD simulations were extensively used to explore stable conformations of proteins in a water environment (**Pan and Aller, 2018**) and were also introduced to loop modeling (**Barozet et al., 2021**). We used MD simulations to search for representative CDR-H3 loop conformations. The simulation



**Figure 3.** Molecular dynamics (MD) generated conformations for benchmark target 7N0R. **(a)** Comparison of CDR-H3 loops of MD (gray), AF2 (pink), and experimentally determined structure (cyan). **(b)** Root mean square fluctuation (RMSF) of antibody residues during simulation. CDR-H3 loop is located in residue number ranging from 98 to 109.

**Table 4.** The accuracy of molecular dynamics (MD)-based CDR-H3 loop optimization in the 10 worst cases of AF2.

PDB ID	C $\alpha$ -RMSD <sub>Ranked_0</sub>	C $\alpha$ -RMSD <sub>MD_opt</sub>	$\Delta$ C $\alpha$ -RMSD
7n0r	10.92	5.62 $\pm$ 0.97	5.30
3juy	6.37	5.71 $\pm$ 0.23	0.66
5y80	6.61	7.59 $\pm$ 0.47	-0.98
7a4t	6.19	7.48 $\pm$ 0.29	-1.29
4nzs	6.57	7.73 $\pm$ 0.26	-1.16
6xzu	7.45	6.34 $\pm$ 0.94	1.11
6x05	6.32	7.48 $\pm$ 0.63	-1.16
3c08	6.68	7.01 $\pm$ 0.11	-0.33
4z9k	9.04	8.01 $\pm$ 0.37	1.03
6oca	7.61	8.01 $\pm$ 0.34	-0.40

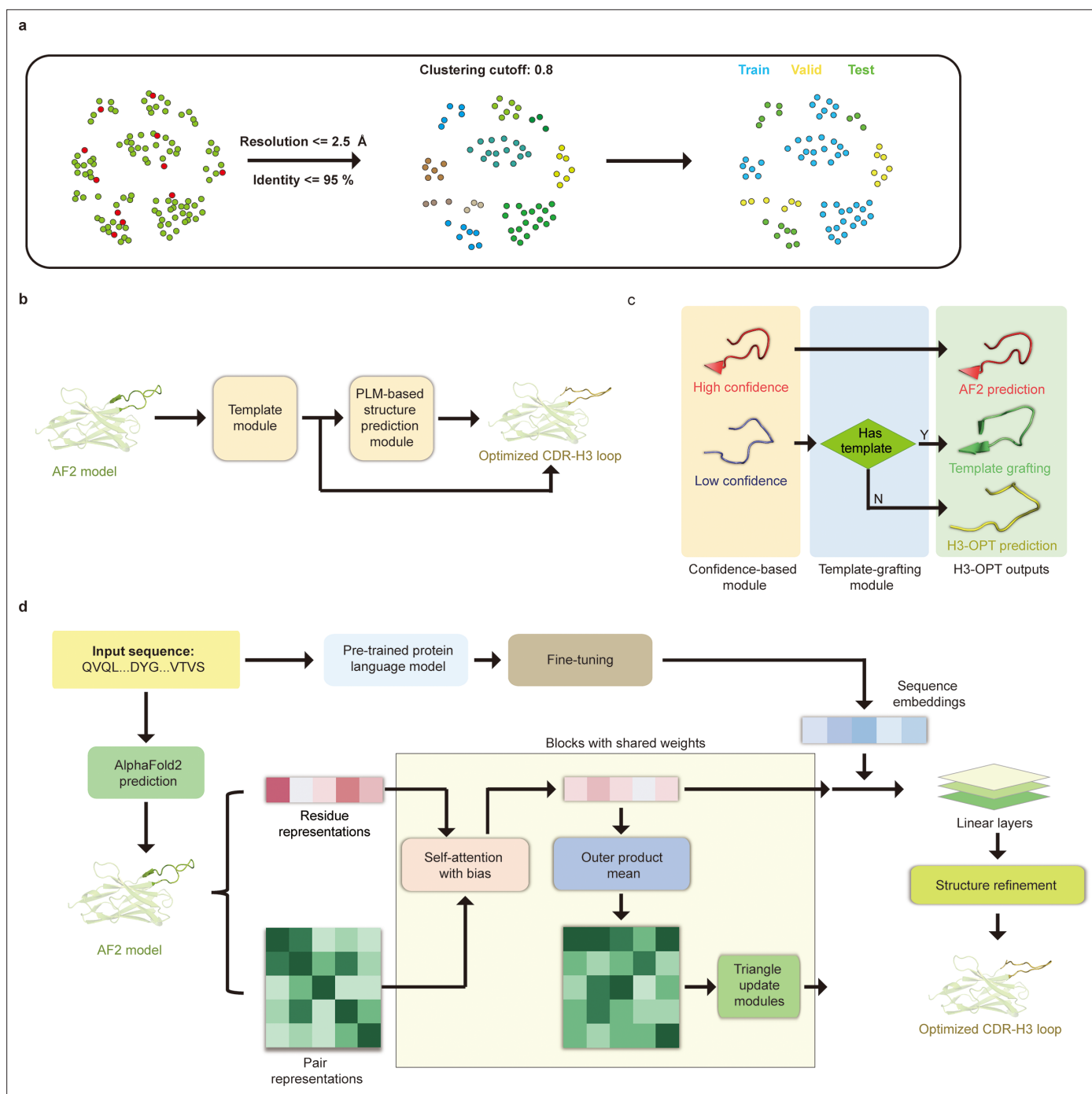
RMSD = root mean square deviation.

systems consist of antibody structures, water, and ions. We selected the top 10 challenging-to-predict targets from the DB1 and DB2 to investigate the dynamic conformations of CDR-H3 loops. The first target was 7N0R, a single-domain antibody that binds the SARS-CoV-2 Nucleocapsid protein with a 12-residue CDR-H3 loop (Ye *et al.*, 2021). We found that MD successfully generated conformations with lower CDR-H3 C $\alpha$ -RMSDs (average value is 5.62 Å, achieving an improvement of 5.30 Å over the AF2 prediction) by correcting the orientation of the CDR-H3 loop (Figure 3a). Additionally, the conformations of 7N0R indicated that the CDR-H3 loop was more flexible than framework regions and other CDR loops, demonstrated by root mean square fluctuation (RMSF) of every residue during simulation (Figure 3b). Despite these promising results for 7N0R, the high CDR-H3 C $\alpha$ -RMSDs of all 10 targets (>5 Å) and poor performance in other cases ( $\Delta$ C $\alpha$ -RMSD < 1.5 Å) (Table 4) both suggested that the conformation search method based on MD simulations failed to generate CDR-H3 loops that closely matched native structures. However, the accuracy of these antibodies with relative long CDR-H3 loops could be improved by H3-OPT through incorporating latent structural information of loop folding from ESM2 or enhanced sampling techniques (Feng *et al.*, 2021).

### The H3-OPT workflow

Prior to training H3-OPT, we first conducted a thorough evaluation of currently available antibody structure prediction tools, including AF2, RoseTTAFold, trRosetta, NanoNet, DeepAb, MODELLER, and ABodyBuilder. AF2 provided high-accuracy predictions for the overall structures of both mAbs and Nbs, with TM-scores and GDT scores >0.9. Based on the relatively higher accuracy of AF2 in structural predictions, its prediction was used to extract structural features of CDR-H3 loops for further optimization by H3-OPT. Using a mAb/Nb sequence and its Fv model structure from AF2 as input, H3-OPT was then used to generate a refined structure. As data quality has large effects on prediction accuracy, we constructed a non-redundant dataset (sequence identity < 0.8) with 1286 high-resolution (<2.5 Å) antibody structures from SAbDab (Dunbar *et al.*, 2014; Figure 4a). The dataset was then divided into training, validation, and test sets based on amino acid sequence identity, which was done by using the UCLUST (Edgar, 2010) software. To assess the prediction results, the test set was split into three subgroups according to the differences in AF2 accuracy (RMSD, a measure of the difference between the predicted structure and an experimental or reference structure) stemming from the length of CDR-H3 sequence: easy-to-predict targets (0–2 Å, Sub1); moderate-difficulty targets (2–4 Å, Sub2); and challenging-to-predict targets (>4 Å, Sub3), with average CDR-H3 loop sequence lengths of 9.12, 11.08, and 16.43, respectively.

The workflow of H3-OPT is depicted in Figure 4b. H3-OPT consists of a template module and a PLM-based structure prediction module (PSPM). The template module determines whether to use PSPM to optimize CDR-H3 and comprises two submodules: a confidence-based module (CBM) and



**Figure 4.** H3-OPT architecture. **(a)** Schematic for dataset preparation. Structures were screened from the SABDab database based on resolution and sequence identity. Clustering of the filtered, high-resolution structures yielded three datasets for training ( $n = 1021$ ), validation ( $n = 134$ ), and testing ( $n = 131$ ). **(b)** The workflow of H3-OPT includes two modules. The template module determines whether to use PLM-based structure prediction module (PSPM), while the PSPM module optimizes the AF2 input structures. **(c)** The template module retains AF2-predicted loops when the confidence score is  $>0.8$  and grafts CDR-H3 loops onto AF2 models for structures with an available template. **(d)** In the PSPM, the network extracts residue-level information and pairwise residue representations from the AF2-predicted models, which are subsequently updated using weight-sharing blocks and concatenated with sequence representations from ESM2. The resulting data is used to predict the 3D coordinates of the H3 loops.

a template-grafting module (TGM). The CBM calculates an AF2 confidence score to evaluate the reliability of CDR-H3 loop sequence inputs through MSA and template searching. The TGM then identifies a template from the H3 template database and grafts the CDR-H3 loop onto the AF2 model if such a template is available (Figure 4c).

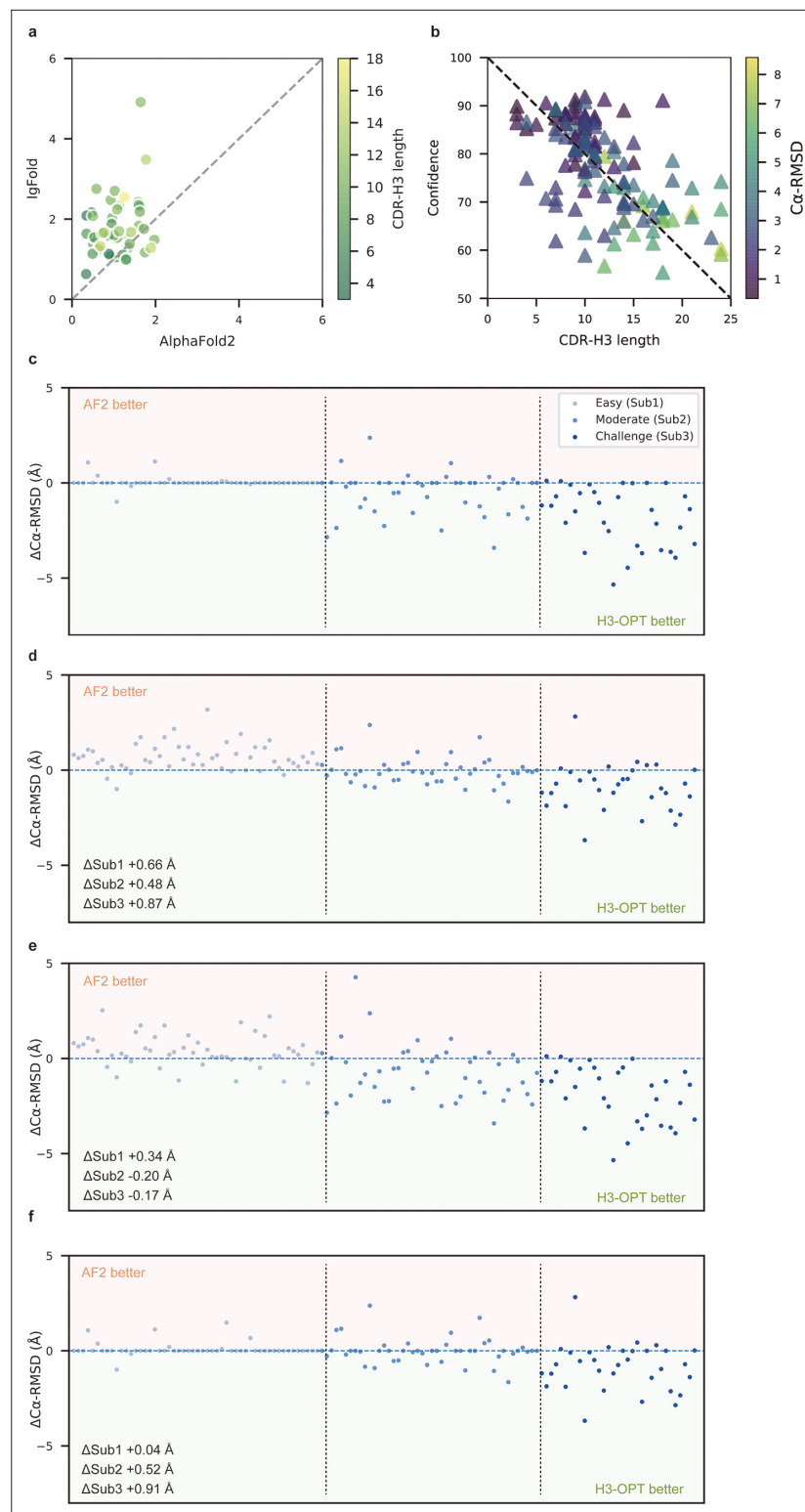
The PSPM contains a series of attention-based components to ensure that the complete structural context is extracted from the input sequences. Specifically, the PSPM employs a row-wise gated attention layer that updates residue-level information and exchanges information within residue pair representations and residue features to infer relationships between the spatial and residue representations. Additionally, by leveraging sequence-level representations from PLMs along with residue-level information, the PSPM can integrate high-information features, such as protein folding patterns, across a vast protein space to predict final three-dimensional (3D) atomic coordinates (**Figure 4d**).

To better understand the contribution of each module to H3-OPT accuracy with varying factors, we next examined how incorporating different approaches through the template and PSPM modules could improve accuracy of H3-OPT. While AF2 showed markedly stronger predictions than IgFold in Sub1 (**Figure 5a**), IgFold provided higher accuracy in Sub3 (**Figure 6a**). In light of previous studies that showed PLM-based models are computationally more efficient but have unsatisfactory accuracy when high-resolution templates and MSA are available (**Lin et al., 2023**), we therefore sought to combine the strengths of AF2 with PLM-based modeling. Similar to IgFold, an initial version of H3-OPT lacking the CBM could not replicate the accuracy of AF2 in Sub1. Interestingly, we observed that the AF2 confidence score of CDR-H3 shared a strong negative correlation with  $C\alpha$ -RMSDs (Pearson correlation coefficient =  $-0.67$ ; **Figure 5b**), which led to us to hypothesize that AF2 models with high-confidence scores might be sufficiently accurate, and therefore do not require further optimization. Following this notion, we incorporated a CBM submodule that could directly retain high-confidence structures from AF2. TGM submodule was added as a means of grafting identical H3 loop templates from public PDBs onto AF2-predicted models to further improve accuracy. Ablation studies in which the CBM or TGM were excluded to determine their respective contributions to the final predictive accuracy (**Figure 5c–f**) revealed that H3-OPT without the template module could generate structures with average  $C\alpha$ -RMSDs of 1.76, 2.76, and 4.86 Å for the Sub1, Sub2, and Sub3 datasets, respectively. Incorporating only the CBM into our model significantly improved  $RMSD_{C\alpha}$  score by 0.62 Å for Sub1, but exerted negligible effects on Sub2 and Sub3. By contrast, including the TGM alone resulted in substantially improved  $RMSD_{C\alpha}$  for Sub2 and Sub3 by 0.68 and 1.04 Å, respectively, suggesting that templates were more effective for relatively long CDR-H3 antibodies. These results suggested that the combination of TGM and CBM modules could leverage available templates to improve prediction accuracy.

Since the large majority of sequences in Sub2 and Sub3 have long CDR-H3 loops with few sequence homologs, attaining high accuracy in structural predictions becomes increasingly challenging for AF2 (**Jumper et al., 2021**). Inspired by IgFold and other PLM-based methods (**Figure 3a**), we thus developed a PSPM module to capture structural information from the sequence embeddings of PLMs (**Rives et al., 2021**). The key innovations of the PSPM for our workflow were the integration of sequence-level representations from PLMs and the simplified architecture of AF2. ProtTrans-T5, AntiBERTy, and ESM2 were initially used without fine-tuning, resulting in overall average  $C\alpha$ -RMSDs of 2.40, 2.49, and 2.32 Å, respectively (**Table 5**). To improve accuracy, we employed a fine-tuning approach for the downstream CDR-H3 structure prediction task. As ESM2 outperformed other PLMs in our test set, we fine-tuned parameters of all ESM2 hidden layers, which resulted in an overall  $RMSD_{C\alpha}$  of 2.24 Å for H3-OPT. It should be noted that most computational models, such as IgFold, froze PLM weights during model training (**Ruffolo et al., 2023**). Analysis of test subsets showed that H3-OPT could provide high accuracy in side chain predictions, high TM-scores, and high GDT scores in Sub2 and Sub3 (**Figure 6b and c**).

To expand its generalizability, we simplified the Evoformer architecture and introduced parameter-sharing to directly predict Cartesian coordinates of CDR-H3 loops through both residue-level features and pair representations. In residue-level representations, rows represented amino acid types and structural features derived from AF2, while columns represented individual residues from the input sequence. The pair representations contained information about the residue pairs, such as  $C\alpha$ - $C\alpha$  distance. These residue-level representations were updated by the attention-based layers and continuously communicated with pair representations; the pair representations were then updated via triangular multiplicative layers (**Jumper et al., 2021**). This simplified architecture improved model efficiency and avoided overfitting. We also applied a structural alignment strategy for feature extraction and model prediction. Given the high conservation of Fv structures, the alignment strategy was designed to effectively capture residue contribution to loop folding, resulting in improved training speed and



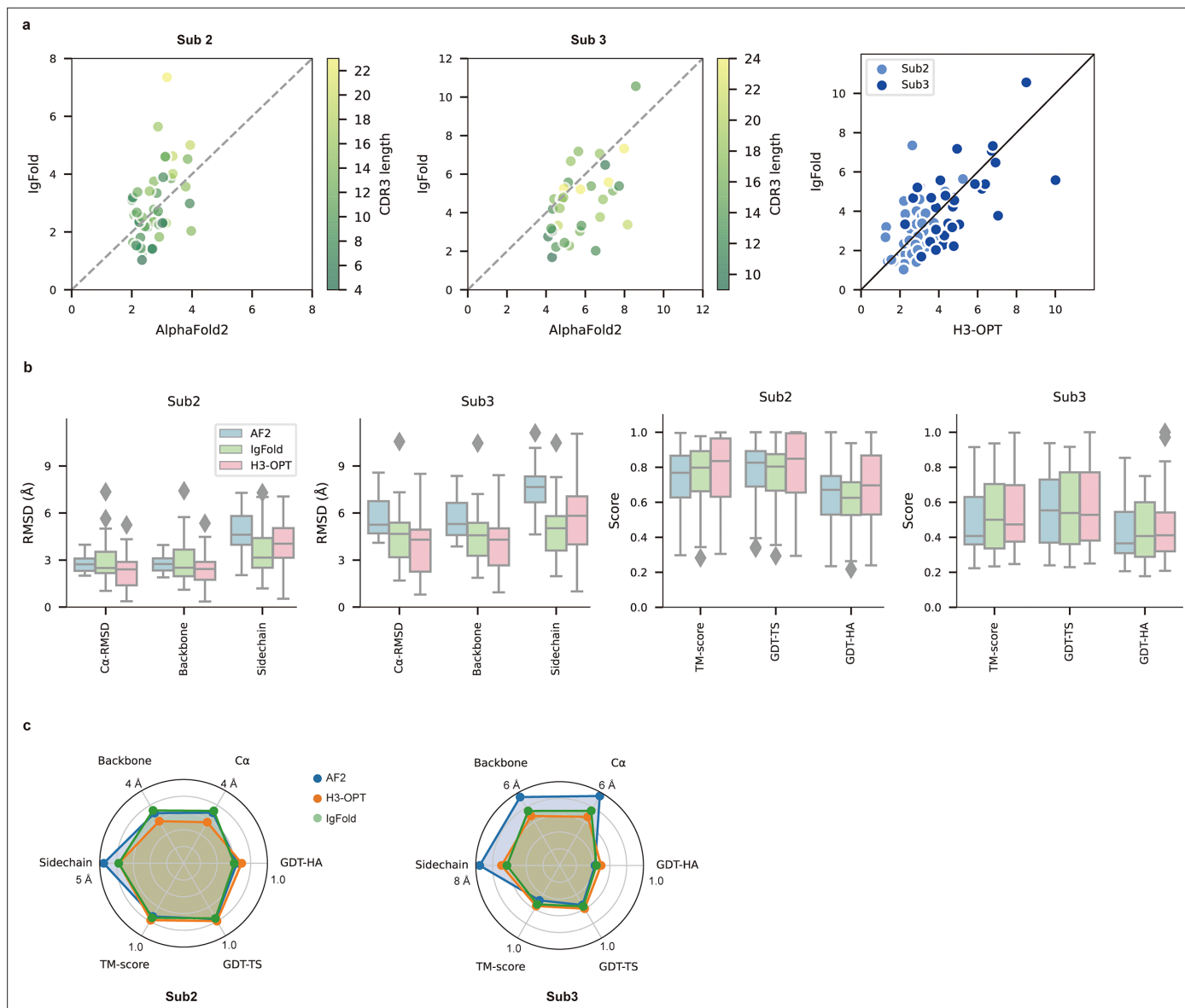


**Figure 5.** Template module and ablation studies. **(a)** Side-by-side comparison of Cα-RMSDs of AF2 and IgFold for Sub1 (n = 52); color scale for data points reflects CDR3 length. AF2 outperformed IgFold for targets left of the dashed diagonal; IgFold outperformed AF2 for targets right of the dashed diagonal. **(b)** Correlations between AF2 confidence score and amino acid sequence length of CDR-H3 loops. Datapoint color indicates Cα-RMSD value for that target. The correlation coefficient for confidence score and CDR-H3 loop length is  $-0.5921$ . **(c)** The accuracy of H3-OPT in three subgroups of the test set.  $\Delta C\alpha$ -RMSDs were calculated by subtracting the  $RMSD_{C\alpha}$  of AF2

Figure 5 continued on next page

Figure 5 continued

from that of H3-OPT. AF2 had higher accuracy for targets above the dashed line; H3-OPT had better accuracy for structures below the dashed line. **(d)** Differences in H3-OPT accuracy without the template module. This ablation study means only PLM-based structure prediction module (PSPM) is used. **(e)** Differences in H3-OPT accuracy without the confidence-based module (CBM). This ablation study means input loop is optimized by template-grafting module (TGM) and PSPM. There are 30 targets in our database with identical CDR-H3 templates. **(f)** Differences in H3-OPT accuracy without the TGM. This ablation study means input loop is optimized by CBM and PSPM.



**Figure 6.** PLM-based structure prediction module (PSPM) module. **(a)** Side-by-side comparison of C $\alpha$ -RMSDs for AF2 and IgFold, IgFold and H3-OPT in the Sub2 ( $n = 46$ ) and Sub3 ( $n = 33$ ) test sets, respectively. **(b)** Comparison of prediction accuracy between AF2 and H3-OPT for Sub2 and Sub3 targets. Metrics including root mean square deviations (RMSDs), template modeling scores (TM-scores), and global distance test (GDT) scores were used to quantitatively assess similarity between predicted and experimental structures. **(c)** Comparison of prediction accuracy between AF2 and H3-OPT using six metrics (RMSD<sub>C $\alpha$</sub> , RMSD<sub>backbone</sub>, RMSD<sub>sidechain</sub>, TM-score, GDT-TS score and GDT-HA score). Radar plots of the mean values of different methods and metrics in predictions of Sub2 and Sub3 targets.

**Table 5.** Performance of H3-OPT with different protein language models (PLMs).

	RMSD <sub>C<math>\alpha</math></sub> (Å)
H3-OPT	2.24 $\pm$ 1.05
AF2	2.85 $\pm$ 0.69
ESM2	2.31 $\pm$ 1.13
Without PLM	2.41 $\pm$ 1.26
AntiBERTy	2.49 $\pm$ 1.42
ProtTrans-T5	2.40 $\pm$ 1.28

RMSD = root mean square deviation.

accuracy. Taken together, H3-OPT provided high-accuracy predictions for CDR-H3 loops, facilitating the development of antibody therapeutics.

### H3-OPT provides higher accuracy CDR-H3 models than existing structure prediction software

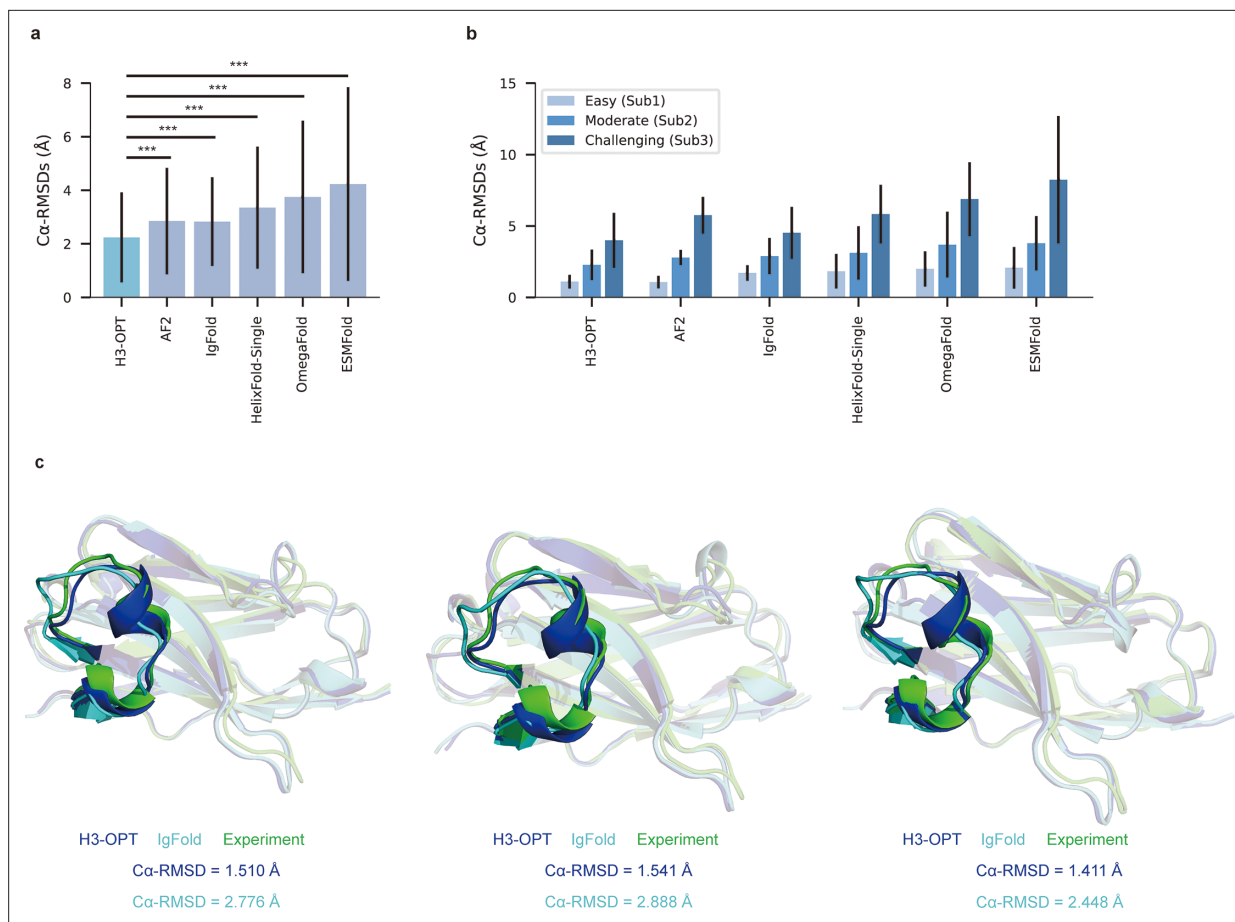
We then evaluated the performance of H3-OPT against currently available methods, including AF2, IgFold, HelixFold-Single, ESMFold, and OmegaFold. H3-OPT achieved an average C $\alpha$  root mean square deviation (RMSD<sub>C $\alpha$</sub> ) of 2.24 Å for CDR-H3, while AF2 and IgFold had CDR-H3 RMSDs of 2.85 and 2.87 Å, respectively (**Figure 7a**). HelixFold-Single, OmegaFold, and ESM-Fold generated comparatively poor predictions, with RMSDs for CDR-H3 of 3.39, 3.75, and 4.23 Å, respectively. As shown in **Figure 7b**, H3-OPT provided higher accuracy predictions than other methods in all three subgroups, with average RMSDs of 1.10, 2.28, and 3.99 Å, respectively. While the predictions of H3-OPT were comparable to AF2 in Sub1, its accuracy was higher than AF2 in Sub2 and Sub3.

In light of these results, we then sought to validate H3-OPT using three experimentally determined structures of anti-VEGF nanobodies, including a wild-type (WT) and two mutant (Mut1 and Mut2) structures, that were recently deposited in Protein Data Bank (**Figure 7c**). Although Mut1 (E45A) and Mut2 (Q14N) shared the same CDR-H3 sequences as WT (Length<sub>CDR-H3</sub> = 17), only minor variations were observed in the CDR-H3. H3-OPT generated accurate predictions with C $\alpha$ -RMSDs of 1.510, 1.541, and 1.411 Å for the WT, Mut1, and Mut2, respectively (the confidence scores of these AlphaFold2-predicted loops were all higher than 0.8, and these loops were accepted as the outputs of H3-OPT by CBM). Subsequent comparison with IgFold showed that AlphaFold2 outperformed IgFold on these targets and IgFold could not accurately predict the short helix in CDR-H3, resulting in C $\alpha$ -RMSDs of 2.776 Å (WT), 2.888 Å (Mut1), and 2.448 Å (Mut2) and more diverse conformations of Mut1 and Mut2. These results indicated that IgFold was capable of learning long-range correlations in protein sequences, but that these long-range correlations could introduce larger errors when MSA and templates were available. These results thus demonstrated that MSA could strongly influence the accuracy for which similar CDR-H3 template structures were available.

### H3-OPT can predict antibody surface properties

To demonstrate how improved CDR-H3 structural accuracy can assist antibody engineering, we applied H3-OPT to predict surface properties. A comparison of H3-OPT with AF2 for identifying surface amino acids (SAAs) by the relative accessible surface areas (rASAs) of CDR-H3 loops showed that H3-OPT could predict SAAs, on average, close to that of the native structures (9.46 vs 9.40, **Figure 8a**), whereas AF2 predicted an average of 9.81 SAAs. Furthermore, H3-OPT predicted closer values to native structures than AF2 in predicting diverse surface properties, such as the distribution of hydrophilic or charged SAAs (**Figure 8b**).

To examine insights H3-OPT could provide into the biophysical properties of antibodies (**Cong et al., 2021**), we next estimated the solvent-accessible surface areas (SASAs). The average SASAs of H3-OPT loops closely resembled that of the native structures, with larger differences in AF2 predictions (**Figure 8c**). Plots of SASA distributions at each alignment position revealed smaller errors in H3-OPT than AF2, with  $\Delta$ SASA ranging from  $-12.17$  to  $9.74$  Å<sup>2</sup> (**Figure 8—figure supplement 1**). In



**Figure 7.** Accuracy of CDR-H3 loop prediction by H3-OPT. (a) The performance of H3-OPT in the test set ( $n_{mAbs} = 119$ ,  $n_{Nbs} = 12$ ) relative to other methods. The  $RMSD_{Ca}$  of H3-OPT was significantly lower than other existing methods ( $p < 0.001$ ). (b) The performance of H3-OPT in structural predictions of three subgroups of the test set ( $n = 52, 46$ , and  $33$ ). (c) H3-OPT structural predictions for three anti-VEGF nanobodies (PDB ID: 8IIU, 8IJZ, 8IJS). The sequence identities of the VH domain and H3 loop are 0.816 and 0.647, respectively, compared with the best template. \*\*\* $p < 0.001$ . RMSD, root mean square deviation.

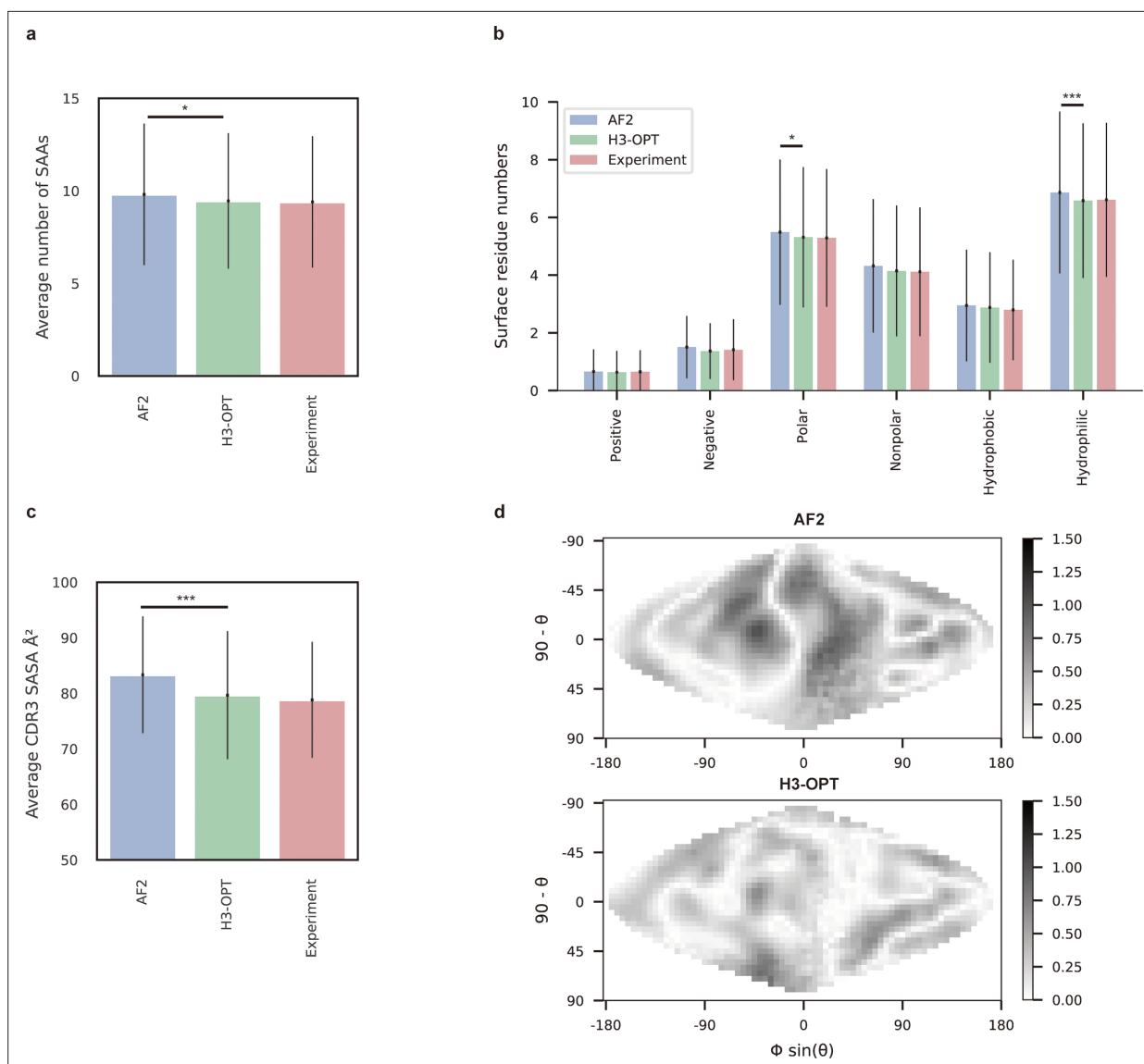
The online version of this article includes the following figure supplement(s) for figure 7:

**Figure supplement 1.** Comparison of accuracy between AF2, H3-OPT, and tFold-Ab methods using the [CAMEO 2022 benchmark dataset \(Leemann et al., 2023\)](#).

addition, comparison of surface charge distributions on the predicted CDR-H3 loops generated by H3-OPT and AF2 in an electrostatic map of a representative structure (PDB 5U3P, **Figure 8d**) showed that H3-OPT predictions were consistent with the experimental structure. These results collectively showed that accurate surface properties predicted by H3-OPT could provide insights into antibody folding and stability, as well as their interactions with antigens.

### H3-OPT can facilitate investigation of antibody–antigen interactions

To explore the potential applications of H3-OPT in studying antibody–antigen interactions, we analyzed contact patterns at predicted binding sites. Given that predicting the structure of an antibody–antigen complex remains challenging, we superimposed  $V_H$  fragments from H3-OPT and AF2 onto experimentally determined native complex structures available in our test set. After identifying the contact residues of antigens by H3-OPT, we found that H3-OPT could substantially outperform AF2 (**Figure 9a**), with a median precision of 0.82 and accuracy of 0.98 compared to 0.71 precision and 0.97 accuracy of AF2. Next, we estimated the distances among interface residues, which are related to binding affinity at the interface. We found that H3-OPT had less error in distance metrics than predictions by AF2 across different distance thresholds, with average mean squared errors of 2.42 and 4.85 Å, respectively (**Figure 9b**). Furthermore, calculation of H3 contact propensities to assess



**Figure 8.** Analysis of surface patches. **(a)** Analysis of surface amino acids for predicted H3 loops. Y-axis represents average number of surface residues for H3 loops ( $n = 131$ ). The surface residues of AF2 models are significantly higher than those of H3-OPT models ( $p < 0.05$ ). **(b)** Histogram of surface patches with different properties predicted by H3-OPT, AF2, or experimentally solved H3 loops. Error bars show standard deviations. H3-OPT models predicted lower values than AF2 models in terms of various surface properties, including polarity ( $p < 0.05$ ) and hydrophilicity ( $p < 0.001$ ). **(c)** Solvent-accessible surface area (SASA) analysis of predicted H3 loops. Values represent the difference in SASA between predicted and experimentally determined H3 structures using AF2 or H3-OPT. The SASA of AF2 models are significantly higher than those of H3-OPT models ( $p < 0.001$ ). **(d)** Comparison of the charged surface patches between H3-OPT and AF2 for target PDB ID: 5U3P. The surface maps compare the surface electrostatic potential of the CDR-H3 loop predicted by H3-OPT or AF2 with the native structure. Darker shading indicates greater difference in electrostatic potential. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

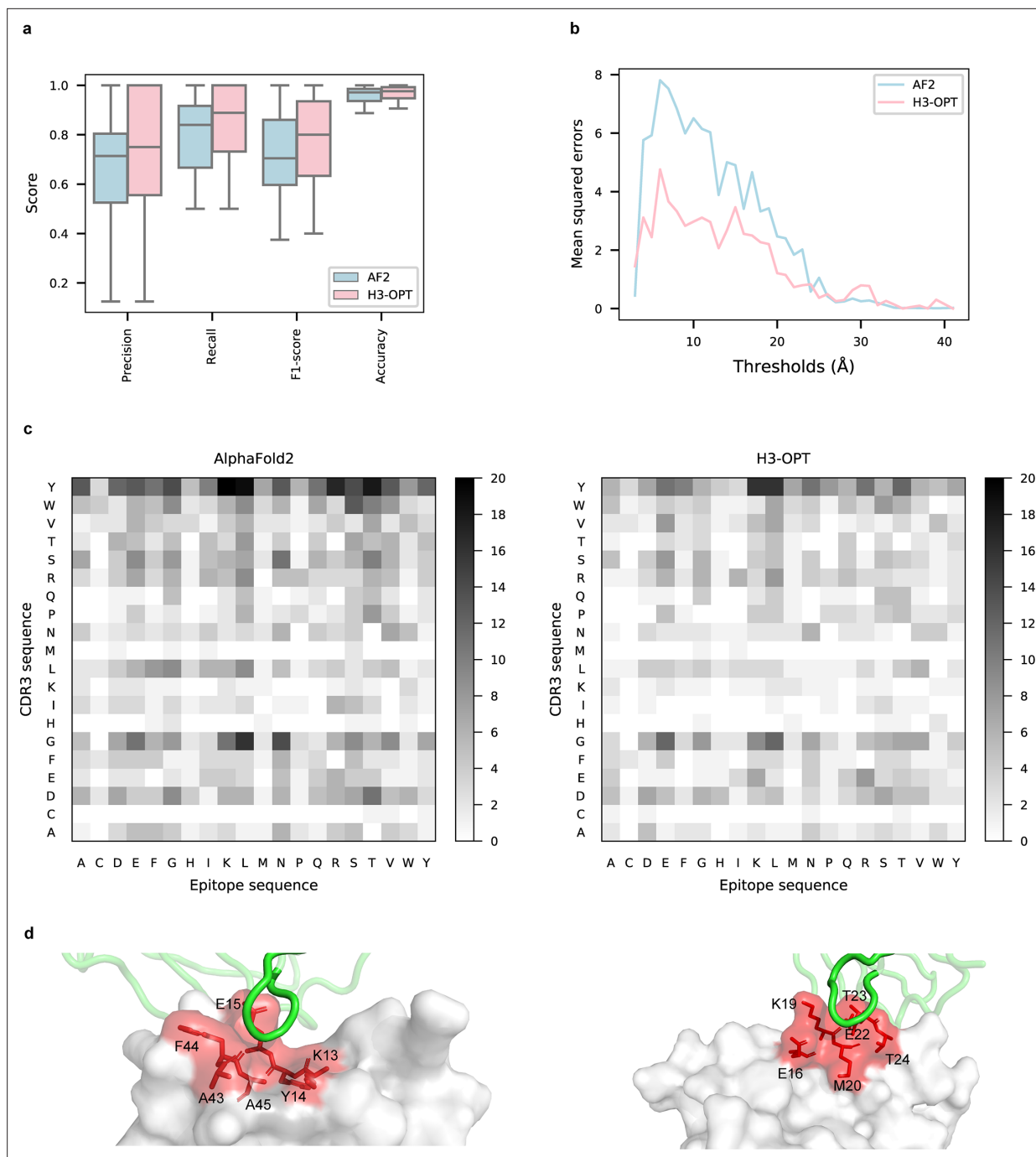
The online version of this article includes the following figure supplement(s) for figure 8:

**Figure supplement 1.** Solvent-accessible surface area (SASA) analysis of predicted H3 loops.

potential differences in binding patterns between the predicted and native structures using high-quality H3 conformations indicated that H3-OPT also displayed higher accuracy in predicting contact propensities (Figure 9c).

Finally, to test whether H3-OPT was reliable for investigation of the detailed mechanisms underlying antigen–antibody binding, we generated contact maps and calculated binding affinities for complex structures obtained by H3-OPT, AF2, or through experiments (see ‘Methods’ for details). We found that contact maps predicted by H3-OPT were consistent with those observed in the





**Figure 9.** Accuracy of H3-OPT predictions of antibody–antigen interactions. **(a)** Performance of H3-OPT in binding site prediction. comparison of prediction accuracy between H3-OPT and AF2 for antibody–antigen binding sites ( $n = 27$ ). Box represents interquartile range (IQR); horizontal line in the center of the box shows median. **(b)** Comparison of the mean squared errors of residue pairs between H3-OPT and AF2 under different distance thresholds. The x-axis represents the experimentally determined distance between pairs of contacting residues at the binding site in the native structure. Y-axis shows mean squared errors of H3-OPT and AF2. **(c)** Heatmaps of the frequency of pairwise residue–residue contacts across antibody–antigen interfaces. This analysis compares contact frequency of H3 loops predicted by AF2 or H3-OPT with the native structure. Darker shading indicates greater difference in contact frequency. **(d)** The predicted H3 loops of two targets interacting with antigens (PDB: 2YC1, 6O9H). The epitopes are highlighted in red and antibody chains are green. H3-OPT could identify the epitopes of different antigens that form the complementary binding interface(s) for the CDR-H3 of antibodies.

The online version of this article includes the following figure supplement(s) for figure 9:

**Figure supplement 1.** RMSD<sub>backbone</sub> during production runs.

**Table 6.** Comparison of binding affinities obtained from molecular dynamics (MD) simulations using AF2 and H3-OPT.

PDB ID	AF2 (kcal/mol)			H3-OPT (kcal/mol)			AF2 (kcal/mol)		H3-OPT (kcal/mol)	
	MM/GBSA	MM/PBSA	AF2 RMSD <sub>C<math>\alpha</math></sub> (Å)	MM/GBSA	MM/PBSA	H3-OPT RMSD <sub>C<math>\alpha</math></sub> (Å)	$ \Delta_{MM/GBSA}^* $	$ \Delta_{MM/PBSA} $	$ \Delta_{MM/GBSA} $	$ \Delta_{MM/PBSA} $
2ghw	-29.20	-33.36	2.7	-14.70	-21.36	3.0	8.63	2.42	23.13	14.42
2yc1	-38.85	-37.73	2.3	-43.80	-48.72	1.5	6.80	18.67	1.85	7.68
3l95	-29.59	-53.35	2.5	-47.22	-68.86	2.5	23.60	11.44	5.97	4.07
3u30	-37.31	-42.41	2.6	-44.94	-50.07	2.5	9.64	2.18	2.01	5.48
4cni	-36.96	-42.93	1.0	-31.92	-40.39	1.3	8.54	7.44	3.50	4.89
4nbz	-36.59	-43.79	1.9	-59.61	-54.23	0.6	10.17	3.30	12.85	13.74
4xnq	-13.55	-17.47	2.7	-31.51	-30.94	0.5	15.40	12.32	2.57	1.15
4ydl	-52.51	-74.25	4.8	-49.17	-73.57	3.6	6.82	6.53	10.15	7.21
5e5m	-59.72	-71.15	3.0	-41.29	-53.70	7.3	0.50	5.79	18.93	11.66
5f7y	-61.76	-69.46	2.7	-60.33	-69.43	1.4	3.95	6.38	5.38	6.41
6kyz	-12.66	-20.32	4.0	-9.36	-17.13	3.7	17.63	17.21	20.93	20.40
6o9h	-39.53	-43.45	2.8	-52.27	-57.51	0.6	10.45	13.31	2.29	0.74
6pyd	-45.87	-58.71	1.0	-35.75	-45.28	1.1	6.29	13.50	3.83	0.06
6u9s	-36.54	-48.66	1.0	-39.79	-44.80	1.3	14.35	10.42	11.11	14.28
Average	/	/	2.6	/	/	2.4	10.20	9.35	8.89	8.01

\* $\Delta_{MM/GBSA}$  (or  $\Delta_{MM/PBSA}$ ) was calculated by subtracting the MM/GBSA (or MM/PBSA) of predicted model from the MM/GBSA (or MM/PBSA) of experimental structure.

experimental structures (**Figure 9d**). Since affinity prediction plays a crucial role in antibody therapeutics engineering, we performed MD simulations to compare the differences in binding affinities between AF2-predicted complexes and H3-OPT-predicted complexes (**Figure 9—figure supplement 1**). Calculation of binding affinities through MD simulations showed that the average affinities of structures obtained by H3-OPT prediction were closer to those of experimentally determined structures than values obtained through AF2 (**Table 6**). These cumulative findings illustrate the informative value of high-quality H3 loops for predicting binding affinity and facilitating antibody engineering, and further support the use of H3-OPT as a powerful and versatile tool for investigating antigen–antibody interactions.

## Discussion

Several platforms for in silico antibody engineering have been developed to improve binding affinity (**Lippow et al., 2007; Riahi et al., 2021; Clark et al., 2006; Shan et al., 2022**), humanization (**Choi et al., 2015; Kurella and Gali, 2014**), and stability (**Tomar et al., 2018; Froning et al., 2020**). At present, engineering these biophysical properties of therapeutic antibodies heavily relies on the availability of the structures of thousands of antibodies. The accurate structure of CDR-H3 is crucial for understanding the mechanism(s) of antigen–antibody interactions. Although AF2 performs well when MSA and templates are available, its predictive capability decreases for targets with long CDR-H3 loops. By contrast, PLMs leverage tens of millions of protein sequences to learn relevant structural patterns, consequently enhancing their capacity to predict orphan antibodies (**Chowdhury et al., 2022**). The results of IgFold indicate that PLMs, which incorporate transformer-based attention mechanisms, are capable of learning both local and global sequence information (**Ruffolo et al., 2023**). While global information can be useful for challenging targets, it may lead to large errors when templates were available, as demonstrated by our case study of anti-VEGF nanobodies (**Figure 7**).

H3-OPT combines the strengths of AF2 and PLMs to predict the structures of mAbs and Nbs, and our results show that combining these features can improve the average RMSD<sub>C $\alpha$</sub>  by 20% over that of AF2 or IgFold. However, H3-OPT is less efficient than PLM-based methods like IgFold because it relies on AF2 structures. To improve efficiency, one possible solution is to cluster the targets based on sequence identity and compute AF2 structures for each cluster, followed by generation of mutant structures. Despite the lower efficiency, H3-OPT offers several notable advantages over IgFold. First, H3-OPT incorporates a simplified version of Evoformer, which improves training efficiency. In addition, H3-OPT employs an alignment-based strategy to extract structural features that are then used to directly predict the C $\alpha$  coordinates of H3 loops, effectively capturing the residue features relevant to loop folding. Moreover, we found that IgFold uses an augmented dataset containing AF2-predicted structures, which could potentially propagate errors during model training. In contrast, H3-OPT was trained by using high-quality and non-redundant antibody structures. Finally, H3-OPT also shows lower C $\alpha$ -RMSDs compared to AF2 or tFold-Ab for the majority (six of seven) of targets in an expanded benchmark dataset, including all antibody structures from CAMEO 2022 (Leemann *et al.*, 2023 ; Figure 7—figure supplement 1).

Although H3-OPT can provide high accuracy with a small training dataset, it is likely that a larger dataset containing high-resolution structures will further improve our model. In the current study, we observed that the accuracy of AF2 CDR-H3 predictions was correlated with the length of the H3 loop. Thus, AF2 can provide accurate predictions for short loops with fewer than 10 amino acids, with PLM-based models offering little or no improvement in such cases. Conversely, antibodies with H3 loop lengths exceeding 25 residues (common in Sub3) pose a long-standing challenge for structural prediction algorithms, and none of the existing methods (including H3-OPT) tested here could provide accurate predictions due to a lack of homologous sequences and a high degree of freedom in this subgroup. Notably, H3-OPT outperformed AF2 in Sub3 because the context of protein sequences learned from ESM2 was fine-tuned to fit antibody structures. Furthermore, we envision that antibodies with H3 loop lengths ranging from 10 to 25 amino acids could be further optimized using PLMs with more model parameters. In addition, we attempted to optimize the H3 loop structure through MD simulations and QM-based methods. However, these approaches generally yielded less accurate predictions than AF2 due to inaccurately described solvent effects and other environmental factors. The development of a more accurate, physics-based method for modeling long loops may be another potentially effective strategy for improving H3-OPT.

Deep generative models are playing a vital role in the field of protein design, leading to several successful applications (Watson *et al.*, 2023; Dauparas *et al.*, 2022; Shin *et al.*, 2021; Luo *et al.*, 2022; Watson *et al.*, 2022; Ingraham *et al.*, 2019). These models capture the intricate relationship between sequences and structures, enabling the generation of novel protein scaffolds not existing in nature. However, the potential immunogenicity concerns associated with these newly designed proteins require comprehensive preclinical and clinical investigations. Consequently, conventional mAbs and Nbs remain the prevailing choices in medicine. Recognizing that the properties of antibodies are heavily influenced by their structures, accurate prediction of the structures of mAbs and Nbs is of great significance in optimizing their therapeutic effectiveness and clinical relevance. To this end, H3-OPT has strong potential to accelerate optimization of antibody–antigen binding as well as engineering of therapeutic antibodies with specialized biophysical properties.

## Methods

### Datasets

High-quality structures are necessary for the accurate assessment of antibody modeling methods. All antibody information was obtained from the SAbDab website. We first downloaded antibody structures with X-ray diffraction resolution of 2.5 Å or less from the RCSB website (<https://www.rcsb.org/>) and deleted redundant sequences with 95% or more identity. Then, we clustered the remaining sequences into 93 clusters using UCLUST (Edgar, 2010) by 65% sequence similarity cutoff and selected centroid sequences to ensure the representativeness and sufficient quantity of the final datasets. We additionally omitted the PDB IDs 5DHV, 3SE9, 4YDJ, 5VF6, 6CT7, 7FAB, 6NMT, 5ILT, and 5UXQ, which were unable to predict using ABodyBuilder. The Fv sequence of each target was identified from initial sequences using the Chothia and Lesk, 1987 definitions. For the Nb dataset, we collected Nb

sequences from a non-redundant dataset (*Zavrtnik and Hadži, 2019*) and deleted all structures with low resolution ( $\geq 2.5$  Å) to construct a high-resolution nanobody structure database. Seven native structures (PDB IDs 6NEX, 6FFJ, 4BUH, 6BHZ, 6UUM, 5O03, and 5C1M) were further excluded from the final dataset owing to the missing residues in CDR-H3 loops. In total, 125 targets were included in the final datasets: 47 mAbs (DB1) and 78 Nbs (DB2). After the model assessment, additional 98 antibodies and 42 nanobodies, which were collected from the test sets of DeepAb and NanoNet, were combined to verify the improvement of the CDR-H3 optimization methods (DB3).

### Sequence logos generation

Sequences of DB1, DB2 V<sub>H</sub> domains were determined by Chothia definitions and aligned using ANARCI (*Dunbar and Deane, 2016*) with the Aho scheme (*Honegger and Plückthun, 2001*). Then, we removed the alignment positions with 85% gaps to obtain a good visualization of alignment statistics. Finally, the alignment outputs of different datasets were submitted to the WebLogo3 server (*Crooks et al., 2004*) to obtain sequence logos of each database. These sequence logos quantitatively showed the conservation of antibody sequences. The height of each letter within the logo corresponds to its base frequency at that particular position. The letters are arranged in descending order of size, with the tallest (or most frequent) letter positioned at the top.

### Benchmarking alternative methods

To evaluate the prediction accuracy of all alternative methods, we benchmarked five AI-based methods (AF2, RoseTTAFold, trRosetta, NanoNet, and DeepAb) and two TBM methods (MODELLER and ABodyBuilder). Identical templates were excluded from TBM templates database and AF2 PDB database. To model full-length V<sub>H</sub> and V<sub>L</sub> antibody, 50 glycines were added as a linker between two chains. The output multiple sequence alignment files of AF2 were then used to predict MODELLER models. ABodyBuilder models were generated using the webserver and modeled by Sphinx. For all methods, recommended parameters were used for all antibody predictions. NanoNet was removed from the DB1 as it was only applied for Nb modeling. We only compared and summarized the target structures that were successfully produced by all participants.

### Model assessment

The residue indexes of different models were aligned and renumbered using MAFFT (*Katoh et al., 2002*) to obtain the same sequence length. We assessed the accuracy of AF2 through varying metrics, including TM-score, GDT-TS score, and GDT-HA score. The structural cutoffs for GDT-TS score were 1, 2, 4, and 8 Å, respectively, while that of the GDT-HA score were 0.5, 1, 2, and 4 Å, respectively. We calculated the average Z-scores of these methods to assess the overall similarities. The average Z-score of each target is given by

$$Z_{ave} = \frac{1}{3}Z_{GDT-TS} + \frac{1}{3}Z_{GDT-HA} + \frac{1}{3}Z_{TM-score}$$

To better evaluate the accuracy of CDR loops, RMSDs were computed after superimposing the Fv or V<sub>H</sub> backbone heavy atoms to their reference structures. All RMSDs were computed using Schrödinger v. 2017-2 (Schrödinger, New York, NY).

### Refinement of the CDR-H3 loop of AF2 through QM methods

We extracted all H3 loop atoms from different AF2 models and re-rank these models based on QM energies. The nitrogen atoms and carbonyl atoms were initially removed from N-terminals and C-terminals, respectively. Next, we added hydrogens to these loops and produced the input files for different QM optimization methods using Schrödinger v. 2017-2. The geometry optimizations were carried out with the Gaussian16 (*Stewart, 2007*) software. We first optimized the CDR-H3 loop by using PM6-D3 (*Grimme et al., 2010; Stephens et al., 1994*) and B3LYP/6-31G\* (*Chothia and Lesk, 1987; Hehre et al., 1972*). The optimization method was divided into three steps: (1) optimizing hydrogens; (2) freezing the terminal C-alpha atoms and optimizing other atoms; and (3) frequency calculations to obtain Gibbs free energies. For an alternative CDR-H3 loop optimization method, we calculated the single-point energies with thermal corrections. The solvation effect can be estimated using the polarizable continuum model (with a dielectric constant  $\epsilon = 78.36$ ). After optimization, the

proportion of each conformation was calculated based on the energies of five AF2 models according to Boltzmann distribution, which was used to generate a Boltzmann-averaged structure. The structures that failed in the optimization process were excluded from the final comparison.

### H3-OPT dataset

Crystal structures of antibodies were obtained from SAbDab, a non-redundant nanobody database (Zavrtanik and Hadži, 2019) and a subset of DeepAb. The structures with resolution  $>2.5$  Å or redundant sequences with 95% or greater identity were removed (Figure 4a). We removed the structures with missing loops in CDR-H3 or lengths  $>45$ . The structures with missing residues at CDR-H3 loops were also dropped. To enhance the generalizability of H3-OPT, the sequences were clustered based on a 90% similarity cutoff by using UCLUST (Edgar, 2010). The resulting clusters were then randomly divided into training, validating, and testing sets to avoid overestimation of performance. The number of sequences in training, validating, and testing sets was 1021 (925 mAbs, 96 Nbs), 134 (122 mAbs, 12 Nbs), and 131 (119 mAbs, 12 Nbs), respectively, with average CDR-H3 loop length of 11.8, 11.9, and 11.7, respectively. Schrödinger v. 2017-2 (Schrödinger), Numpy 1.18.5, and Pandas 1.0.5 were used for data preparation.

### Structure preparation

To address the transformational invariance for predicting 3D coordinates, we applied a structural alignment strategy. Prior to feature extraction, we removed all atoms of light chains in mAbs Fv structures and any non-standard residues in native structures. All antibody  $V_H$  native structures were aligned using StructAlign module in Schrödinger with a randomly selected reference structure (PDB ID: 1GIG). Next, we superimposed AF2-predicted models to their corresponding native structures. This alignment allowed the model to learn the interactions between residues located in H3 loops while disregarding the rotation and translation of the entire structure. The heavy chain fragments were determined using Chothia definitions (Ahdritz et al., 2022a).

### Feature generation

As shown in Table 7, structural features were extracted and aggregated into the following inputs to PSPM of H3-OPT: a residue feature vector of size  $[N_{res}, 34]$  was constructed by concatenating ‘amino acid type’, ‘AF2-predicted coordinates’, ‘AF2-predicted backbone torsion angles’, ‘torsion angles mask’, and ‘predicted residue mask’. The pairwise feature vector of size  $[N_{res}, N_{res}, 60]$  included the ‘pairwise distances’ and ‘pairwise amino acid type’.

**Table 7.** Features of the model.

$N_{res}$  is the number of residues (Jumper et al., 2021).

Feature and shape	Description
Amino acid type $[N_{res}, 21]$	One-hot representation of the input amino acid sequence (including 20 amino acids and unknown).
3D coordinates $[N_{res}, 3]$	$C\alpha$ coordinates of all AlphaFold2-predicted residues
Backbone torsion angles $[N_{res}, 6]$	Sine and cosine encoding of all predicted three backbone torsion angles.
Torsion angles mask $[N_{res}, 3]$	A mask indicating if the angle was presented in the predicted structure.
H3 residue mask $[N_{res}, 1]$	A mask indicating if the residue was located in H3 loop.
Pairwise distances $[N_{res}, N_{res}, 39]$	One hot representation of residue alpha carbon atoms distance. The pairwise distances ranging from 3.25 Å to 50.75 Å were put into 38 bins equally and the last bin contained any larger distances.
Pairwise amino acid type $[N_{res}, N_{res}, 21]$	One-hot representation of the input amino acid sequence.



**Table 8.** Hyperparameters for H3-OPT models.

Model	2	5	1	3	Best
Initial learning rate	1 <sup>-4</sup>	5 <sup>-4</sup>	1 <sup>-3</sup>	5 <sup>-4</sup>	1 <sup>-4</sup>
Hidden layers	64	64	64	64	64
Iterations numbers of Evoformer-like layer	6	6	4	4	4
Average RMSD <sub>Cα</sub> (Å)	2.42	2.36	2.35	2.33	2.24

RMSD = root mean square deviation.

## Network architecture

H3-OPT took the predicted H3 loops of AF2 as input and generate the antibody structure with the optimized CDR-H3 loops as output. H3-OPT was composed of a template module and a PSPM. The template module contains two submodules, that is, CBM and TGM. The CBM calculated the average confidence score of H3 residues and retained AF2 structures with confidence score >0.8. The TGM grafted the template loop structure onto the input AF2 model, when their CDR-H3 sequences were identical. We replaced the predicted CDR-H3 loop generated by AF2 with the template loop by aligning corresponding atoms, which included Cα atoms and carbonyl carbon atoms in the first two residues and Cα atoms and backbone nitrogen atoms in the last two residues.

The input residue-level features are extracted from AF2-predicted structures as mentioned above, resulting in a dimension of  $L \times 34$ , where  $L$  is the padding length of antibody heavy chain. Similarly, the pairwise features has a dimension of  $L \times L \times 60$ . The PSPM network began with a row-wise gated multi-head self-attention layer that updated residue-level features from pair representations. Then, the residue-level information updated the pair representations via outer product mean module. Pair representations were updated by two multiplicative update modules. During training, the weights of the attention layer and the triangle multiplicative update modules were updated four times per epoch. Then, the final hidden states from ESM2 (esm2\_t33\_650M\_UR50D) were passed to a linear layer and then concatenated with the residue-level features to enable the learning of latent information from vast protein sequences. To predict all Cα coordinates for the CDR-H3 loop, the output vector was passed through three linear layers with a hidden size of 64. In summary, PSPM utilized an attention-based mechanism to learn the contribution of individual residues and incorporate sequence representations from PLMs into final predictions. We used OpenFold (Banks et al., 2005; based on GitHub from June 2022: commit 3f57b4a041f063406059f42080ede6d495479617; Ahdriz et al., 2022b) to implement partial networks of AF2.

We employed a fine-tuning strategy to update the sequence representations in ESM2. Initially, we froze all weights of the 33 representation layers in ESM2 and updated only the weights of the attention layers and pair representation update modules. Subsequently, we fixed all weights of the remaining components in H3-OPT and performed fine-tuning exclusively on all hidden layers of ESM2 for the H3 loop prediction task. Finally, we fine-tuned all parameters of H3-OPT to predict the atomic coordinates of all CDR-H3 Cα atoms.

**Table 9.** Average Cα-RMSDs of our test set under different confidence cutoffs.

Cutoff	Cα-RMSD (Å)
0.70	2.46
0.75	2.30
0.80	2.24
0.85	2.17
0.90	2.29
0.95	2.28

RMSD = root mean square deviation.

During training, we trained our model on the training set and validated it on the validation set. The dropout rates of H3-OPT were set to 0.25. Mean squared error (MSE) loss was utilized to train the model. We used Adam optimizer with a learning rate of  $1 \times 10^{-4}$ , weight decay of  $5 \times 10^{-4}$  for training. The model was trained on an NVIDIA V100 super GPU and took 1 hr per epoch over the entire training set with a batch size of 64. We randomly selected 10% of all structures for model validation. The H3-OPT was implemented using PyTorch 1.12.1 in Python 3.7.2. The hyperparameters used during training process for H3-OPT are presented in **Table 8**. The average C $\alpha$ -RMSDs of our test set under different confidence cutoffs in CBM are presented in **Table 9**.

## Structure generation and refinement

We utilized a structure refinement strategy to generate structures of CDR-H3 loops and rectify structural errors. Initially, we used PSPM predictions to modify the C $\alpha$  coordinates of predicted CDR-H3 loops generated by AF2. Then, we minimized the remaining atoms of the CDR-H3 loops, while keeping their C $\alpha$  coordinates fixed, by using the OPLS 2005 force field (*Mirdita et al., 2022*). After the initial energy minimization step, we applied a force constant of 10 kcal/mol to these fixed atoms to guarantee complete relaxation of the entire loops.

## Validation settings

H3-OPT, AF2, HelixFold-Single, IgFold, and ESMFold were compared on the test subset. Given that tFold-Ab is only available on webserver, we compared H3-OPT with tFold-Ab on the **CAMEO 2022 dataset** (*Leemann et al., 2023*). We removed the target antibody structures from the AF2 database before prediction to evaluate the performance of AF2. This ensured that the predictions depended solely on the information derived from other template structures, thus avoiding biased results. Since all other methods did not require MSA searching, we did not exclude these targets from their database. To predict the entire Fv structures, we included 50 glycine residues as a linker between the heavy chain and the light chain for antibodies (*Levy, 2010*). The C $\alpha$ -RMSDs of all predicted H3 loops were calculated after superimposing the V<sub>H</sub> backbone heavy atoms (without CDR-H3) to the reference structures. All structures were generated using publicly available code repositories: AF2 v2.1.1 (based on GitHub from the November 2021 version at GitHub: commit [91b43223422420d1783ed802c8b3a8382a9309fd](https://github.com/af2/af2/commit/91b43223422420d1783ed802c8b3a8382a9309fd); *Zidek et al., 2021*), IgFold (based on GitHub from February 2023: commit [6a09298d165ed1deb438c0b6eefc9cb03ed0eca5](https://github.com/IgFold/IgFold/commit/6a09298d165ed1deb438c0b6eefc9cb03ed0eca5); *Graylab, 2023*), HelixFold-Single (based on GitHub from January 2023: commit [5f39b2c2a4ecc00b89ba05b95dc56212bdd5d886](https://github.com/HelixFold-Single/HelixFold-Single/commit/5f39b2c2a4ecc00b89ba05b95dc56212bdd5d886); *Xiang and Xiaoyao4573, 2023*), and ESMFold v1 (based on GitHub from October 2022: commit [dc823b89c6acb9f67caea53704c2a97524fbd456](https://github.com/sercu/ESMFold/commit/dc823b89c6acb9f67caea53704c2a97524fbd456); *Sercu, 2022*).

## Analysis of SASAs and pairwise residue contacts

We computed the SASAs and rASAs of all CDR-H3 loops by using the ANARCI webserver (*Dunbar and Deane, 2016*) with the default settings. Residue with an rASA >25% was considered surface residues (*Schweke et al., 2022*). Subsequently, we applied AHO alignment software (*Honegger and Plückthun, 2001*) to report the averaged SASA per alignment position. The samples and alignment positions with more than 10 gaps were removed to avoid randomness and bias, reducing the final samples to 123. Finally, the difference in SASAs between H3-OPT and AF2 was calculated by subtracting predicted H3 loops SASAs from the ground truth.

To examine the contribution of H3-OPT in discovering antibody–antigen interactions, we identified contact residue pairs that were within this distance as binding sites by setting a threshold of 5 Å. We next calculated the pairwise residue distance matrices for each individual predicted complex and native structure, where each element represented the closest distance between the heavy atoms of two residues. From these pairwise distances, we derived contact propensity matrices that specifically indicated the presence or absence of interactions between residues.

## Analysis of surface electrostatic potential

We generated 2D projections of CDR-H3 loop's surface electrostatic potential using SURFMAP v2.0.0 (*Salomon-Ferrer et al., 2013*; based on GitHub from February 2023: commit: [e0d51a10debc96775468912ccd8de01e239d1900](https://github.com/SURFMAP/SURFMAP/commit/e0d51a10debc96775468912ccd8de01e239d1900); *Chevrollier et al., 2023*) with default parameters. The 2D surface

maps were calculated by subtracting the surface projection of H3-OPT or AF2-predicted 3 loops to their native structures.

### Binding affinity calculation

We analyzed the binding affinities of antibody–antigen complexes predicted by AF2 and H3-OPT. Their relative binding affinities were calculated through MD simulations. Initially, the missing side chains and loops in the antibody structures were filled using the Protein Preparation Wizard module within Schrödinger software. Subsequently, the tLEaP module of AMBER (Tian *et al.*, 2020) was employed to construct the simulation system with the ff19SB force field (Izadi *et al.*, 2014) and OPC solvent model (Chayen and Saridakis, 2008). Additionally, the simulation system was solvated with a 0.15 M NaCl solution. Energy minimization was performed through a 5000-step steepest descent algorithm, followed by a 5000-step conjugate gradient algorithm. Then, a 400-ps NVT simulation with a time step of 2 fs was performed to gradually heat the system from 0 K to 298 K (0–100 K: 100 ps; 100–298 K: 200 ps; hold 298 K: 100 ps), and a 100-ps NPT simulation with a time step of 2 fs was performed to equilibrate the density of the system. During heating and density equilibration, we constrained the antigen–antibody structure with a restraint value of 10 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. In the production run, 100-ns MD simulations were performed with a time step of 2 fs. The first 50 ns restrains the non-hydrogen atoms of the antigen–antibody complex, and the last 50 ns restrains the non-hydrogen atoms of the antigen, with a constraint value of 10 kcal·mol<sup>-1</sup>·Å<sup>-2</sup>. The distance cutoff for nonbonded interactions was set to 10 Å, and the Berendsen algorithm was utilized to maintain isotropic pressure coupling at 1 bar. The Langevin algorithm was employed to maintain the simulation temperature at 298 K. The relative binding affinities of the antigen–antibody complexes were evaluated using the MMPBSA module of AMBER software, which computed the MM/GBSA energies for the trajectory frames of last 10 ns.

### Crystallization and data collection

The protein expression, purification, and crystallization experiments were described previously (Zhu *et al.*, 2023; Yu *et al.*, 2019). The proteins used in the crystallization experiments were unlabeled. Upon thawing the frozen protein on ice, we performed a centrifugation step to eliminate any potential crystal nucleus and precipitants. Subsequently, we mixed the protein at a 1:1 ratio with commercial crystal condition kits using the sitting-drop vapor diffusion method facilitated by the Protein Crystallization Screening System (TTP LabTech, mosquito). After several days of optimization, single crystals were successfully cultivated at 21°C and promptly flash-frozen in liquid nitrogen. The diffraction data from various crystals were collected at the Shanghai Synchrotron Research Facility and subsequently processed using the aquarium pipeline (Yu *et al.*, 2019).

### Statistics analysis

We conducted two-sided *t*-test analyses to assess the statistical significance of differences between the various groups. Statistical significance was considered when the *p*-values were <0.05. These statistical analyses were carried out using Python 3.10 with the Scipy library (version 1.10.1).

### Acknowledgements

This work was supported by the Tsinghua University Initiative Scientific Research Program (no. 20231080030), Vanke Special Fund for Public Health and Health Discipline Development (2022Z82WKJ009), the Tsinghua-Peking University Center for Life Sciences (no. 20111770319), Tsinghua University – Peking Union Medical College and Hospital Collaboration Foundation (no. 20191080837).

---

### Additional information

#### Competing interests

Xiaonan Zhang, Lihang Liu: Employee of Baidu International Technology (Shenzhen) Co., Ltd. The other authors declare that no competing interests exist.

## Funding

Funder	Grant reference number	Author
Tsinghua University	Initiative Scientific Research Program 20231080030	Boxue Tian
Vanke Special Fund for Public Health and Health Discipline Development	2022Z82WKJ009	Boxue Tian
Tsinghua-Peking University Center for Life Sciences	20111770319	Boxue Tian
Tsinghua University-Peking Union Medical College and Hospital Collaboration Foundation	20191080837	Feng Qian

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## Author contributions

Hedi Chen, Conceptualization, Resources, Data curation, Software, Formal analysis, Supervision, Validation, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing; Xiaoyu Fan, Conceptualization, Software, Methodology; Shuqian Zhu, Yuchan Pei, Data curation, Validation; Xiaochun Zhang, Data curation, Validation, Methodology; Xiaonan Zhang, Lihang Liu, Software, Validation; Feng Qian, Conceptualization, Supervision, Funding acquisition, Project administration, Writing - review and editing; Boxue Tian, Conceptualization, Supervision, Funding acquisition, Methodology, Writing - original draft, Project administration, Writing - review and editing

## Author ORCIDs

Hedi Chen  <http://orcid.org/0000-0003-3922-2156>

Feng Qian  <http://orcid.org/0000-0001-7415-6997>

Boxue Tian  <http://orcid.org/0000-0002-5830-0669>

## Peer review material

Reviewer #2 (Public Review): <https://doi.org/10.7554/eLife.91512.4.sa1>

Reviewer #3 (Public Review): <https://doi.org/10.7554/eLife.91512.4.sa2>

Author response <https://doi.org/10.7554/eLife.91512.4.sa3>

## Additional files

## Supplementary files

- MDAR checklist

## Data availability

The datasets of our study and the codes of H3-OPT are freely available at <https://github.com/chdcg/H3-OPT>, (copy archived at [chdcg, 2024](https://doi.org/10.7554/eLife.91512.4.sa3)).

## References

- Adolf-Bryfogle J**, Kalyuzhnyi O, Kubitz M, Weitzner BD, Hu X, Adachi Y, Schief WR, Dunbrack RL. 2018. RosettaAntibodyDesign (RABD): a general framework for computational antibody design. *PLoS Computational Biology* **14**:e1006112. DOI: <https://doi.org/10.1371/journal.pcbi.1006112>, PMID: 29702641
- Aguiar Rangel M**, Bedwell A, Costanzi E, Taylor RJ, Russo R, Bernardes GJL, Ricagno S, Frydman J, Vendruscolo M, Sormanni P. 2022. Fragment-based computational design of antibodies targeting structured epitopes. *Science Advances* **8**:eabp9540. DOI: <https://doi.org/10.1126/sciadv.abp9540>, PMID: 36367941
- Ahdritz G**, Bouatta N, Floristean C, Kadyan S, Xia Q, Gerecke W, O'Donnell TJ, Berenberg D, Fisk I, Zanichelli N, Zhang B, Nowaczynski A, Wang B, Stepniewska-Dziubinska MM, Zhang S, Ojewole A, Guney ME, Biderman S, Watkins AM, Ra S, et al. 2022a. OpenFold: retraining alphafold2 yields new insights into its learning mechanisms and capacity for generalization. *bioRxiv*. DOI: <https://doi.org/10.1101/2022.11.20.517210>
- Ahdritz G**, Yu D, Wei J, Kadyan S. 2022b. Openfold. 3f57b4a. GitHub. <https://github.com/aqlaboratory/openfold>

- Almagro JC**, Beavers MP, Hernandez-Guzman F, Maier J, Shaulsky J, Butenhof K, Labute P, Thorsteinson N, Kelly K, Teplyakov A, Luo J, Sweet R, Gilliland GL. 2011. Antibody modeling assessment. *Proteins* **79**:3050–3066. DOI: <https://doi.org/10.1002/prot.23130>
- Almagro JC**, Teplyakov A, Luo J, Sweet RW, Kodangattil S, Hernandez-Guzman F, Gilliland GL. 2014. Second antibody modeling assessment (AMA-II). *Proteins* **82**:1553–1562. DOI: <https://doi.org/10.1002/prot.24567>, PMID: 24668560
- Baek M**, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, Millán C, Park H, Adams C, Glassman CR, DeGiovanni A, Pereira JH, Rodrigues AV, van Dijk AA, Ebrecht AC, Opperman DJ, et al. 2021. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**:871–876. DOI: <https://doi.org/10.1126/science.abj8754>, PMID: 34282049
- Banks JL**, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM. 2005. Integrated modeling program, applied chemical theory (IMPACT). *Journal of Computational Chemistry* **26**:1752–1780. DOI: <https://doi.org/10.1002/jcc.20292>, PMID: 16211539
- Barozet A**, Chacón P, Cortés J. 2021. Current approaches to flexible loop modeling. *Current Research in Structural Biology* **3**:187–191. DOI: <https://doi.org/10.1016/j.crstbi.2021.07.002>, PMID: 34409304
- Brahmer JR**, Tykodi SS, Chow LQM, Hwu WJ, Topalian SL, Hwu P, Drake CG, Camacho LH, Kauh J, Odunsi K, Pitot HC, Hamid O, Bhatia S, Martins R, Eaton K, Chen S, Salay TM, Alaparthi S, Grosso JF, Korman AJ, et al. 2012. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *New England Journal of Medicine* **366**:2455–2465. DOI: <https://doi.org/10.1056/NEJMoa1200694>
- Chayen NE**, Saridakis E. 2008. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods* **5**:147–153. DOI: <https://doi.org/10.1038/nmeth.f.203>, PMID: 18235435
- chdcg**. 2024. H3-opt. swl:1:rev:1cb9ec1046644e592232db913923a9a7e9761ba4. Software Heritage. <https://archive.softwareheritage.org/swl:1:dir:c129f69f2cf5b519f6880bae9df45c8dc7c9e40c;origin=https://github.com/chdcg/H3-OPT;visit=swl:1:snp:378dc373ee16a69ec82e34ae188e98a8f482d1dd;anchor=swl:1:rev:1cb9ec1046644e592232db913923a9a7e9761ba4>
- Chen J**, Gu Z, Xu Y, Deng M, Lai L, Pei J. 2023. QuoteTarget: A sequence-based transformer protein language model to identify potentially druggable protein targets. *Protein Science* **32**:e4555. DOI: <https://doi.org/10.1002/pro.4555>, PMID: 36564866
- Chevrollier N**, Schweke H, Lopes A. 2023. Surfmap. e0d51a1. GitHub. <https://github.com/i2bc/SURFMAP>
- Choi Y**, Hua C, Sentman CL, Ackerman ME, Bailey-Kellogg C. 2015. Antibody humanization by structure-based computational protein design. *mAbs* **7**:1045–1057. DOI: <https://doi.org/10.1080/19420862.2015.1076600>, PMID: 26252731
- Chothia C**, Lesk AM. 1987. Canonical structures for the hypervariable regions of immunoglobulins. *Journal of Molecular Biology* **196**:901–917. DOI: [https://doi.org/10.1016/0022-2836\(87\)90412-8](https://doi.org/10.1016/0022-2836(87)90412-8), PMID: 3681981
- Chowdhury R**, Bouatta N, Biswas S, Floristean C, Kharkar A, Roy K, Rochereau C, Ahdriz G, Zhang J, Church GM, Sorger PK, AlQuraishi M. 2022. Single-sequence protein structure prediction using a language model and deep learning. *Nature Biotechnology* **40**:1617–1623. DOI: <https://doi.org/10.1038/s41587-022-01432-w>, PMID: 36192636
- Clark LA**, Boriack-Sjodin PA, Eldredge J, Fitch C, Friedman B, Hanf KJM, Jarpe M, Liparoto SF, Li Y, Lugovskoy A, Miller S, Rushe M, Sherman W, Simon K, Van Vlijmen H. 2006. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. *Protein Science* **15**:949–960. DOI: <https://doi.org/10.1110/ps.052030506>, PMID: 16597831
- Cohen T**, Halfon M, Schneidman-Duhovny D. 2022. NanoNet: Rapid and accurate end-to-end nanobody modeling by deep learning. *Frontiers in Immunology* **13**:958584. DOI: <https://doi.org/10.3389/fimmu.2022.958584>, PMID: 36032123
- Cong Y**, Feng Y, Ni H, Zhi F, Miao Y, Fang B, Zhang L, Zhang JZH. 2021. Anchor-locker binding mechanism of the coronavirus spike protein to human ACE2: insights from computational analysis. *Journal of Chemical Information and Modeling* **61**:3529–3542. DOI: <https://doi.org/10.1021/acs.jcim.1c00241>, PMID: 34156227
- Crooks GE**, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Research* **14**:1188–1190. DOI: <https://doi.org/10.1101/gr.849004>, PMID: 15173120
- Dauparas J**, Anishchenko I, Bennett N, Bai H, Ragotte RJ, Milles LF, Wicky BIM, Courbet A, de Haas RJ, Bethel N, Leung PJY, Huddy TF, Pellock S, Tischer D, Chan F, Koepnick B, Nguyen H, Kang A, Sankaran B, Bera AK, et al. 2022. Robust deep learning-based protein sequence design using ProteinMPNN. *Science* **378**:49–56. DOI: <https://doi.org/10.1126/science.add2187>, PMID: 36108050
- Du Z**, Su H, Wang W, Ye L, Wei H, Peng Z, Anishchenko I, Baker D, Yang J. 2021. The trRosetta server for fast and accurate protein structure prediction. *Nature Protocols* **16**:5634–5651. DOI: <https://doi.org/10.1038/s41596-021-00628-9>, PMID: 34759384
- Dunbar J**, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, Shi J, Deane CM. 2014. SAbDab: the structural antibody database. *Nucleic Acids Research* **42**:D1140–D1146. DOI: <https://doi.org/10.1093/nar/gkt1043>, PMID: 24214988
- Dunbar J**, Deane CM. 2016. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics* **32**:298–300. DOI: <https://doi.org/10.1093/bioinformatics/btv552>, PMID: 26424857
- Edgar RC**. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**:2460–2461. DOI: <https://doi.org/10.1093/bioinformatics/btq461>, PMID: 20709691



- Eswar N**, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. 2006. Comparative protein structure modeling using Modeller. *Current Protocols in Bioinformatics* **Chapter 5**:5.6.1–5.6.30. DOI: <https://doi.org/10.1002/0471250953.bi0506s15>, PMID: 18428767
- Fang X**, Wang F, Liu L, He J, Lin D, Xiang Y, Zhang X, Wu H, Li H, Song L. 2022. HelixFold-single: msa-free protein structure prediction by using protein language model as an alternative. *arXiv*. DOI: <https://doi.org/10.21203/rs.3.rs-1969991/v1>
- Feng JJ**, Chen JN, Kang W, Wu YD. 2021. Accurate structure prediction for protein loops based on molecular dynamics simulations with RSFF2C. *Journal of Chemical Theory and Computation* **17**:4614–4628. DOI: <https://doi.org/10.1021/acs.jctc.1c00341>, PMID: 34170125
- Ferruz N**, Schmidt S, Höcker B. 2022. ProtGPT2 is a deep unsupervised language model for protein design. *Nature Communications* **13**:4348. DOI: <https://doi.org/10.1038/s41467-022-32007-7>, PMID: 35896542
- Froning K**, Maguire J, Sereno A, Huang F, Chang S, Weichert K, Frommelt AJ, Dong J, Wu X, Austin H, Conner EM, Fitchett JR, Heng AR, Balasubramaniam D, Hilgers MT, Kuhlman B, Demarest SJ. 2020. Computational stabilization of T cell receptors allows pairing with antibodies to form bispecifics. *Nature Communications* **11**:2330. DOI: <https://doi.org/10.1038/s41467-020-16231-7>, PMID: 32393818
- Graylab**. 2023. Igfold. 6a09298. GitHub. <https://github.com/Graylab/IgFold>
- Grimme S**, Antony J, Ehrlich S, Krieg H. 2010. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *The Journal of Chemical Physics* **132**:154104. DOI: <https://doi.org/10.1063/1.3382344>, PMID: 20423165
- Hehre WJ**, Ditchfield R, Pople JA. 1972. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **56**:2257–2261. DOI: <https://doi.org/10.1063/1.1677527>
- Hie BL**, Shanker VR, Xu D, Bruun TUJ, Weidenbacher PA, Tang S, Wu W, Pak JE, Kim PS. 2024. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology* **42**:275–283. DOI: <https://doi.org/10.1038/s41587-023-01763-2>, PMID: 37095349
- Honegger A**, Plückthun A. 2001. Yet another numbering scheme for immunoglobulin variable domains: an automatic modeling and analysis tool. *Journal of Molecular Biology* **309**:657–670. DOI: <https://doi.org/10.1006/jmbi.2001.4662>
- Ingraham J**, Garg V, Barzilay R, Jaakkola T. 2019. Generative models for graph-based protein design. *Advances in Neural Information Processing Systems*. .
- Izadi S**, Anandakrishnan R, Onufriev AV. 2014. Building water models: a different approach. *The Journal of Physical Chemistry Letters* **5**:3863–3871. DOI: <https://doi.org/10.1021/jz501780a>, PMID: 25400877
- Jumper J**, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583–589. DOI: <https://doi.org/10.1038/s41586-021-03819-2>, PMID: 34265844
- Jumper J**, Hassabis D. 2022. Protein structure predictions to atomic accuracy with AlphaFold. *Nature Methods* **19**:11–12. DOI: <https://doi.org/10.1038/s41592-021-01362-6>, PMID: 35017726
- Kaplon H**, Reichert JM. 2021. Antibodies to watch in 2021. *mAbs* **13**:1860476. DOI: <https://doi.org/10.1080/19420862.2020.1860476>, PMID: 33459118
- Katoh K**, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30**:3059–3066. DOI: <https://doi.org/10.1093/nar/gkf436>, PMID: 12136088
- Kurella VB**, Gali R. 2014. Structure guided homology model based design and engineering of mouse antibodies for humanization. *Bioinformatics* **10**:180–186. DOI: <https://doi.org/10.6026/97320630010180>, PMID: 24966517
- Leem J**, Dunbar J, Georges G, Shi J, Deane CM. 2016. ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. *mAbs* **8**:1259–1268. DOI: <https://doi.org/10.1080/19420862.2016.1205773>, PMID: 27392298
- Leemann M**, Sagasta A, Eberhardt J, Schwede T, Robin X, Durairaj J. 2023. Automated benchmarking of combined protein structure and ligand conformation prediction. *Proteins* **91**:1912–1924. DOI: <https://doi.org/10.1002/prot.26605>, PMID: 37885318
- Levy ED**. 2010. A simple definition of structural regions in proteins and its use in analyzing interface evolution. *Journal of Molecular Biology* **403**:660–670. DOI: <https://doi.org/10.1016/j.jmb.2010.09.028>, PMID: 20868694
- Lin Z**, Akin H, Rao R, Hie B, Zhu Z, Lu W, Smetanin N, Verkuil R, Kabeli O, Shmueli Y, dos Santos Costa A, Fazel-Zarandi M, Sercu T, Candido S, Rives A. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379**:1123–1130. DOI: <https://doi.org/10.1126/science.ade2574>
- Lippow SM**, Wittrup KD, Tidor B. 2007. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nature Biotechnology* **25**:1171–1176. DOI: <https://doi.org/10.1038/nbt1336>, PMID: 17891135
- Luo S**, Su Y, Peng X, Wang S, Peng J, Ma J. 2022. Antigen-Specific Antibody Design and Optimization with Diffusion-Based Generative Models for Protein Structures. *Advances in Neural Information Processing Systems*. 9754–9767. DOI: <https://doi.org/10.1101/2022.07.10.499510>
- Madani A**, Krause B, Greene ER, Subramanian S, Mohr BP, Holton JM, Olmos JL, Xiong C, Sun ZZ, Socher R, Fraser JS, Naik N. 2023. Large language models generate functional protein sequences across diverse families. *Nature Biotechnology* **41**:1099–1106. DOI: <https://doi.org/10.1038/s41587-022-01618-2>, PMID: 36702895
- Marcatili P**, Rosi A, Tramontano A. 2008. PIGS: automatic prediction of antibody structures. *Bioinformatics* **24**:1953–1954. DOI: <https://doi.org/10.1093/bioinformatics/btn341>

- Meganck RM**, Baric RS. 2021. Developing therapeutic approaches for twenty-first-century emerging infectious viral diseases. *Nature Medicine* **27**:401–410. DOI: <https://doi.org/10.1038/s41591-021-01282-0>, PMID: [33723456](https://pubmed.ncbi.nlm.nih.gov/33723456/)
- Mirdita M**, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. 2022. ColabFold: making protein folding accessible to all. *Nature Methods* **19**:679–682. DOI: <https://doi.org/10.1038/s41592-022-01488-1>, PMID: [35637307](https://pubmed.ncbi.nlm.nih.gov/35637307/)
- Pan L**, Aller SG. 2018. Allosteric role of substrate occupancy toward the alignment of P-glycoprotein nucleotide binding domains. *Scientific Reports* **8**:14643. DOI: <https://doi.org/10.1038/s41598-018-32815-2>, PMID: [30279588](https://pubmed.ncbi.nlm.nih.gov/30279588/)
- Raybould MIJ**, Marks C, Lewis AP, Shi J, Bujotzek A, Taddese B, Deane CM. 2020. Thera-SAbDab: the therapeutic structural antibody database. *Nucleic Acids Research* **48**:D383–D388. DOI: <https://doi.org/10.1093/nar/gkz827>, PMID: [31555805](https://pubmed.ncbi.nlm.nih.gov/31555805/)
- Riahi S**, Lee JH, Wei S, Cost R, Masiero A, Prades C, Olfati-Saber R, Wendt M, Park A, Qiu Y, Zhou Y. 2021. Application of an integrated computational antibody engineering platform to design SARS-CoV-2 neutralizers. *Antibody Therapeutics* **4**:109–122. DOI: <https://doi.org/10.1093/abt/tbab011>, PMID: [34396040](https://pubmed.ncbi.nlm.nih.gov/34396040/)
- Rives A**, Meier J, Sercu T, Goyal S, Lin Z, Liu J, Guo D, Ott M, Zitnick CL, Ma J, Fergus R. 2021. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *PNAS* **118**:e2016239118. DOI: <https://doi.org/10.1073/pnas.2016239118>, PMID: [33876751](https://pubmed.ncbi.nlm.nih.gov/33876751/)
- Ruffolo JA**, Sulam J, Gray JJ. 2022. Antibody structure prediction using interpretable deep learning. *Patterns* **3**:100406. DOI: <https://doi.org/10.1016/j.patter.2021.100406>, PMID: [35199061](https://pubmed.ncbi.nlm.nih.gov/35199061/)
- Ruffolo JA**, Chu LS, Mahajan SP, Gray JJ. 2023. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications* **14**:2389. DOI: <https://doi.org/10.1038/s41467-023-38063-x>, PMID: [37185622](https://pubmed.ncbi.nlm.nih.gov/37185622/)
- Salomon-Ferrer R**, Case DA, Walker RC. 2013. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **3**:198–210. DOI: <https://doi.org/10.1002/wcms.1121>
- Schweke H**, Mucchielli MH, Chevrollier N, Gosset S, Lopes A. 2022. SURFMAP: a software for mapping in two dimensions protein surface features. *Journal of Chemical Information and Modeling* **62**:1595–1601. DOI: <https://doi.org/10.1021/acs.jcim.1c01269>, PMID: [35349266](https://pubmed.ncbi.nlm.nih.gov/35349266/)
- Sercu T**. 2022. Esm. dc823b8. GitHub. <https://github.com/facebookresearch/esm>
- Shan S**, Luo S, Yang Z, Hong J, Su Y, Ding F, Fu L, Li C, Chen P, Ma J, Shi X, Zhang Q, Berger B, Zhang L, Peng J. 2022. Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization. *PNAS* **119**:e2122954119. DOI: <https://doi.org/10.1073/pnas.2122954119>
- Shin JE**, Riesselman AJ, Kollasch AW, McMahon C, Simon E, Sander C, Manglik A, Kruse AC, Marks DS. 2021. Protein design and variant prediction using autoregressive generative models. *Nature Communications* **12**:2403. DOI: <https://doi.org/10.1038/s41467-021-22732-w>, PMID: [33893299](https://pubmed.ncbi.nlm.nih.gov/33893299/)
- Stephens PJ**, Devlin FJ, Chabalowski CF, Frisch MJ. 1994. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry* **98**:11623–11627. DOI: <https://doi.org/10.1021/j100096a001>
- Stewart JJP**. 2007. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **13**:1173–1213. DOI: <https://doi.org/10.1007/s00894-007-0233-4>, PMID: [17828561](https://pubmed.ncbi.nlm.nih.gov/17828561/)
- Tepljakov A**, Luo J, Obmolova G, Malia TJ, Sweet R, Stanfield RL, Kodangattil S, Almagro JC, Gilliland GL. 2014. Antibody modeling assessment II. Structures and models. *Proteins* **82**:1563–1582. DOI: <https://doi.org/10.1002/prot.24554>
- Tian C**, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Miguez AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C. 2020. ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation* **16**:528–552. DOI: <https://doi.org/10.1021/acs.jctc.9b00591>, PMID: [31714766](https://pubmed.ncbi.nlm.nih.gov/31714766/)
- Tomar DS**, Singh SK, Li L, Broulidakis MP, Kumar S. 2018. In silico prediction of diffusion interaction parameter ( $k_D$ ), a key indicator of antibody solution behaviors. *Pharmaceutical Research* **35**:193. DOI: <https://doi.org/10.1007/s11095-018-2466-6>, PMID: [30128780](https://pubmed.ncbi.nlm.nih.gov/30128780/)
- Valdés-Tresanco MS**, Valdés-Tresanco ME, Jiménez-Gutiérrez DE, Moreno E. 2023. Structural modeling of nanobodies: a benchmark of state-of-the-art artificial intelligence programs. *Molecules* **28**:3991. DOI: <https://doi.org/10.3390/molecules28103991>
- Watson JL**, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, Wicky BIM, Hanikel N, Pellock SJ, Courbet A, Sheffler W, Wang J, Venkatesh P, Sappington I, Torres SV, Lauko A, et al. 2022. Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models. *bioRxiv*. DOI: <https://doi.org/10.1101/2022.12.09.519842>, PMID: [35132409](https://pubmed.ncbi.nlm.nih.gov/35132409/)
- Watson JL**, Juergens D, Bennett NR, Trippe BL, Yim J, Eisenach HE, Ahern W, Borst AJ, Ragotte RJ, Milles LF, Wicky BIM, Hanikel N, Pellock SJ, Courbet A, Sheffler W, Wang J, Venkatesh P, Sappington I, Torres SV, Lauko A, et al. 2023. De novo design of protein structure and function with RFdiffusion. *Nature* **620**:1089–1100. DOI: <https://doi.org/10.1038/s41586-023-06415-8>, PMID: [37433327](https://pubmed.ncbi.nlm.nih.gov/37433327/)
- Wu R**, Ding F, Wang R, Shen R, Zhang X, Luo S, Su C, Wu Z, Xie Q, Berger B, Ma J, Peng J. 2022a. High-Resolution de Novo Structure Prediction from Primary Sequence. *bioRxiv*. DOI: <https://doi.org/10.1101/2022.07.21.500999>

- Wu J**, Wu F, Jiang B, Liu W, Zhao P. 2022b. tFold-Ab: fast and accurate antibody structure prediction without sequence homologs. *bioRxiv*. DOI: <https://doi.org/10.1101/2022.11.10.515918>
- Xiang Y**, xiaoyao4573. 2023. Paddlehelix. 5f39b2c. GitHub. [https://github.com/PaddlePaddle/PaddleHelix/tree/5f39b2c2a4ecc00b89ba05b95dc56212bdd5d886/apps/protein\\_folding/helixfold-single](https://github.com/PaddlePaddle/PaddleHelix/tree/5f39b2c2a4ecc00b89ba05b95dc56212bdd5d886/apps/protein_folding/helixfold-single)
- Ye Q**, Lu S, Corbett KD. 2021. Structural Basis for SARS-CoV-2 Nucleocapsid Protein Recognition by Single-Domain Antibodies. *bioRxiv*. DOI: <https://doi.org/10.1101/2021.06.01.446591>, PMID: 34100017
- Yu F**, Wang Q, Li M, Zhou H, Liu K, Zhang K, Wang Z, Xu Q, Xu C, Pan Q, He J. 2019. Aquarium: an automatic data-processing and experiment information management system for biological macromolecular crystallography beamlines. *Journal of Applied Crystallography* **52**:472–477. DOI: <https://doi.org/10.1107/S1600576719001183>
- Zavrtanik U**, Hadži S. 2019. A non-redundant data set of nanobody-antigen crystal structures. *Data in Brief* **24**:103754. DOI: <https://doi.org/10.1016/j.dib.2019.103754>, PMID: 30976633
- Zemla A**, Venclovas C, Moutl J, Fidelis K. 1999. Processing and analysis of CASP3 protein structure predictions. *Proteins* **37**:22–29. DOI: [https://doi.org/10.1002/\(SICI\)1097-0134\(1999\)37:3+<22::AID-PROT5>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-0134(1999)37:3+<22::AID-PROT5>3.0.CO;2-W)
- Zhang Y**, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. *Proteins* **57**:702–710. DOI: <https://doi.org/10.1002/prot.20264>
- Zhu S**, Fan S, Tang T, Huang J, Zhou H, Huang C, Chen Y, Qian F. 2023. Polymorphic nanobody crystals as long-acting intravitreal therapy for wet age-related macular degeneration. *Bioengineering & Translational Medicine* **8**:e10523. DOI: <https://doi.org/10.1002/btm2.10523>
- Zidek A**, Tomlinson H, Ward T. 2021. Alphafold. 91b4322. GitHub. <https://github.com/google-deepmind/alphafold>