Accurate Immune Protein Structure Prediction by Large Language Model and Transfer Learning

Haicang Zhang

zhanghaicang@sjtu.edu.cn

Shanghai Jiao Tong University School of Medicine https://orcid.org/0000-0001-6268-4258

Tian Zhu

Institute of Computing Technology, Chinese of Academy of Sciences

Milong Ren

Institute of Computing Technology, Chinese of Academy of Sciences

Zaikai He

Institute of Computing Technology, Chinese of Academy of Sciences

Siyuan Tao

Institute of Computing Technology, Chinese of Academy of Sciences

Ming Li

University of Waterloo

Jian Zhang

Shanghai Jiao Tong University School of Medicine https://orcid.org/0000-0002-6558-791X

Dongbo Bu

Institute of Computing Technology, Chinese of Academy of Sciences

Article

Keywords:

Posted Date: July 31st, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7153530/v1

Additional Declarations: There is **NO** Competing Interest.

Accurate Immune Protein Structure Prediction by Large Language Model and Transfer Learning

Tian Zhu[2,3†], Milong Ren[2,3†], Zaikai He[2,3], Siyuan Tao[2,3], Ming Li[4,5], Jian Zhang[1*], Dongbo Bu[2,3,4*] and Haicang Zhang[1,2*]

[1]Department of Pharmaceutical and Artificial-Intelligence Sciences, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China.
[2]SKLP, Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China.
[3]University of Chinese Academy of Science, Beijing, China.
[4]Central China Institute of Artificial Intelligence, Zhengzhou, Henan, China.
[5]David R. Cheriton School of Computer Science, University of Waterloo, Ontario, Canada.

*Corresponding authors: E-mail(s): jian.zhang@sjtu.edu.cn; dbu@ict.ac.cn; zhanghaicang@ict.ac.cn;
†These authors contributed equally to this work.

Abstract

Accurate prediction of immune protein structures is critical for advancing immunotherapy. However, deep learning-based methods like AlphaFold and RosettaFold struggle with immune proteins due to the limited number of solved immune protein structures and the absence of homologous sequences in hypervariable regions. To address these challenges, we introduce ImmuneFold, a transfer-learning framework that leverages a large language model and low-rank adaptation (LoRA) for memory-efficient fine-tuning. ImmuneFold outperforms existing methods, including MSA-based AlphaFold3, in predicting the structures of T-cell receptors,

antibodies, and nanobodies. Additionally, we pair ImmuneFold's predictions with Rosetta energy scoring to develop a zero-shot protocol for TCR–epitope binding prediction, effectively mitigating overfitting issues common in supervised approaches. Experimental evaluations also confirm ImmuneFold's robustness and accuracy in binding prediction. Beyond immune proteins, ImmuneFold provides a scalable framework for adapting advanced models, such as ESMFold and AlphaFold, to other protein families, thereby democratizing access to cutting-edge structural tools for researchers, even those with limited computational resources.

1 Main

T-cell receptors (TCRs) and antibodies are critical proteins in the adaptive immune system, responsible for recognizing and neutralizing specific antigens. TCRs specialize in targeting neoantigens and foreign peptides presented by major histocompatibility complexes (MHCs), while antibodies can bind to the surface of a wide range of antigens. Both have been extensively utilized in the development of personalized immunotherapies and vaccines [1–3]. For example, over a hundred antibody drugs have been approved, and the first TCR-based cancer therapy was recently authorized [4]. Additionally, nanobodies—single-domain antibodies found in camelids —have smaller molecular sizes and offer unique advantages in drug development [5].

All these three types of immune proteins belong to the immunoglobulin (Ig) superfamily. Structurally, their framework regions are conserved and maintain the overall architecture, while the complementarity-determining regions (CDRs) are highly variable in both structure and sequence, primarily determining antigen-specific binding. Upon binding to specific antigens, antibodies, and TCRs form antibody-antigen complexes and TCR-pMHC complexes, respectively. Understanding the structural basis and antigen-specific binding mechanisms of these complexes can enhance structure-based drug design with precise binding specificity and affinity. However, experimental methods like X-ray crystallography and cryo-electron microscopy (cryo-EM) are costly and require pure, stable protein samples. Therefore, there is increasing interest in computational modeling of immune protein structures, especially focusing on the CDRs.

While recent deep learning methods, such as AlphaFold [6] and RoseTTAFold [7] have significantly advanced protein structure prediction, accurate prediction of immune protein structures remains challenging. These methods rely on modeling the evolutionary constraints in the multiple sequence alignments (MSAs) of homologous sequences, thus their predictive accuracy is contingent upon the quality of MSAs. However, the CDRs, particularly CDR3 in the antibody heavy chain or TCR β chain, exhibit high variability and lack sufficient homologous sequences.

To address the limitations of MSA-based methods, language model-based approaches specifically designed for antibody structure prediction have been proposed [8, 9]. Rather than relying on homologous sequences, these methods extract the evolutionary constraints from large-scale language models trained on extensive antibody sequences datasets [10, 11]. They have demonstrated comparable or superior performance while significantly improving efficiency, with structures predicted in seconds. However, the prediction accuracy of these methods is constrained by the scarcity of structure data as only thousands of antibody structures are currently available in the Protein Data Bank (PDB). The challenges are even more pronounced for TCRs. Although millions of CDR sequences are available, the number of full-length TCR variable domain sequences is limited, making it difficult to train effective language models for structure prediction. Furthermore, only a few hundred TCR structures are available for training, which is orders of magnitude fewer than for antibodies.

We introduce ImmuneFold, a specialized structure prediction model for immune proteins. This method leverages knowledge learned from general protein structures, thereby mitigating the issue of limited training data. We utilized a general structure prediction model as our base, trained on non-redundant natural protein datasets, which can learn more general evolutionary and structural rules than a model trained specifically on antibody datasets. We employ parameter-efficient transfer learning (PEFT) for immune proteins, which updates only a subset of parameters rather than the entire model. Compared with other transfer learning methods, PEFT requires significantly less memory and training time while maintaining performance [12–14], making it more accessible for academic research groups with limited computational resources. Originally proposed for large-scale language models like Llama2 [12], PEFT has since been extended to image generation tasks [15] and more recently applied to protein language models for various protein-related tasks [14, 16–18]. In this work, we extend PEFT to the Evoformer-like architecture, a crucial yet computationally expensive component of recent deep learning models for protein structure prediction.

To demonstrate the practical utility of ImmuneFold, we apply it to develop a protocol for predicting TCR-epitope binding. Accurate prediction of TCR-epitope interactions is essential for selecting peptides that trigger immune responses and for engineering TCRs capable of efficiently recognizing specific neoantigens in immunotherapy [19]. However, existing supervised methods [20–22] suffer from severe overfitting due to the scarcity and biased distribution of training binding data. They often fail for TCRs absent from training data, limiting their utility in the practical immunotherapy application. To address this challenge, we integrate the ImmuneFold predictions with Rosseta energy [23] to directly evaluate TCR-epitope binding affinity in a zero-short manner, thereby avoiding the need for training binding data.

2 Results

2.1 Low-rank adaptation of ESMFold for immune protein structure prediction

Within the transfer learning framework, ImmuneFold employs parameter-efficient fine-tuning, specifically the Low-Rank Adaptation (LoRA) technique, to enhance a general protein structure prediction model for immune proteins. Unlike full-parameter training, LoRA updates only two low-rank matrices instead of the entire weight matrix, significantly reducing the number of trainable parameters. Previous methods for antibody structure prediction have exclusively utilized antibody structures as the training set, whereas ImmuneFold emphasizes learning geometric and evolutionary constraints from general protein structures and applies them to immune proteins.

ImmuneFold offers two main advantages. First, our approach does not rely on multiple sequence alignments (MSA) but instead leverages a protein language model, enabling more precise and rapid structural predictions for all types of immune proteins. Second, by leveraging accurate structural predictions, our method can be integrated with the Rosetta Software to support zero-shot prediction of immune protein binding capabilities. Unlike methods that require a specific training set with functional annotations, our approach exhibits superior generalization performance in the functional prediction of immune proteins, particularly for those not present in the training set.

To accommodate various real-world application scenarios, ImmuneFold supports structure prediction for both unbound immune proteins and their bound states when the antigen context is provided. We trained two separate models for TCRs and antibodies, respectively.

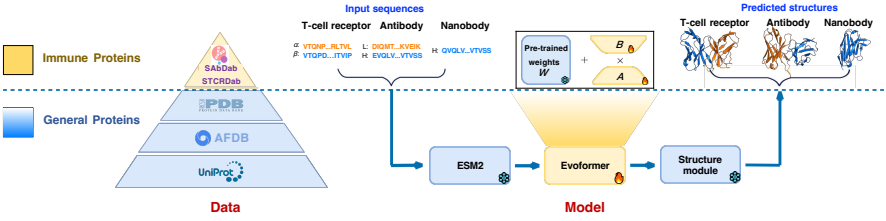

Fig. 1: **Overview of ImmuneFold.** ImmuneFold employs Low-Rank Adaptation (LoRA), a parameter-efficient fine-tuning technique, to fine-tune only the Evoformer module of ESMFold on immune protein structures. Two separate models are fine-tuned: one for TCR structure prediction and the other for antibody and nanobody structure prediction.

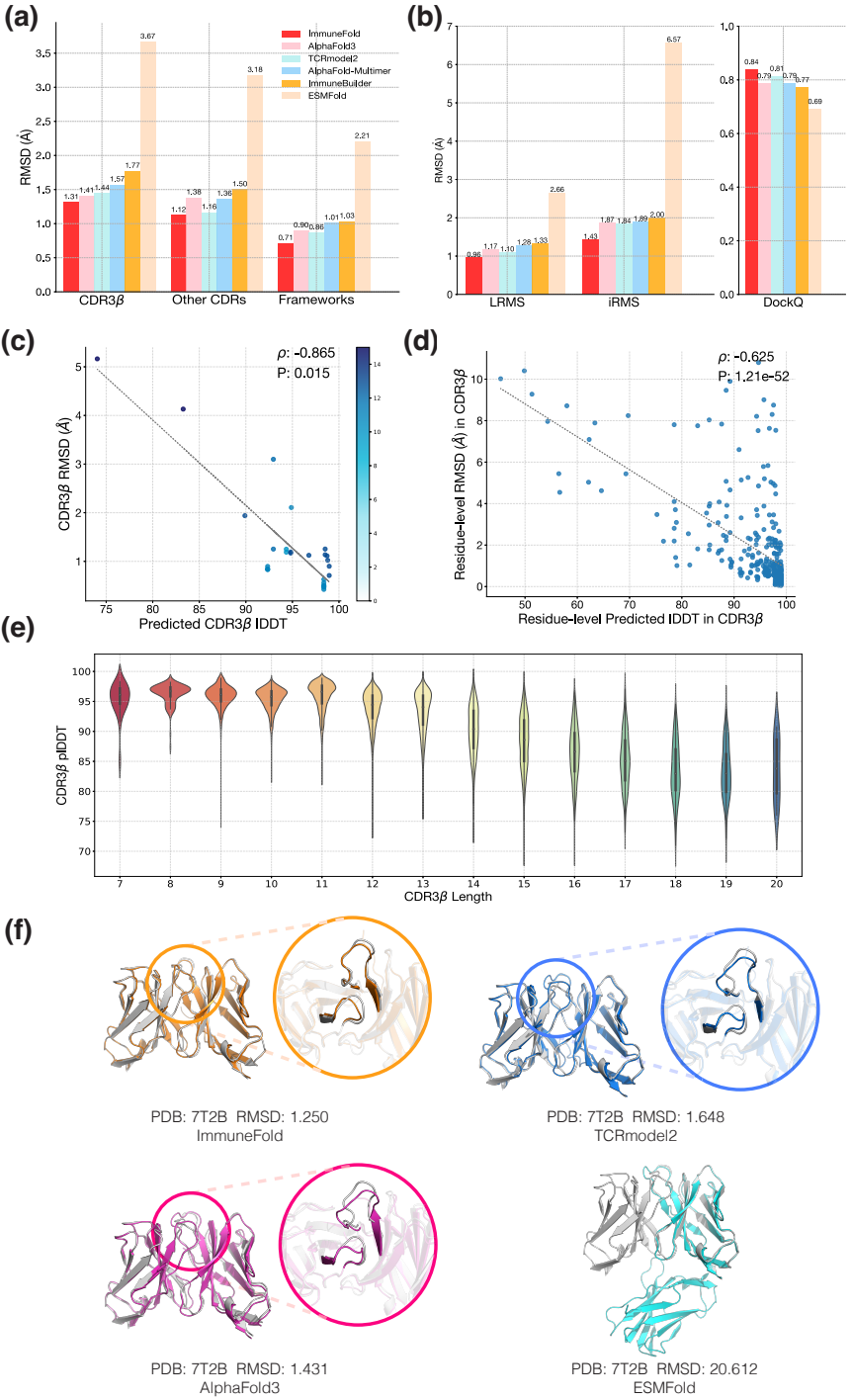
(a)
RMSD (Å)
ImmuneFold
AlphaFold3
TCRmodel2
AlphaFold-Multimer
ImmuneBuilder
ESMFold
CDR3β
Other CDRs
Frameworks
(b)
RMSD (Å)
LRMS
iRMS
DockQ
(c)
ρ: -0.865
P: 0.015
CDR3β RMSD (Å)
Predicted CDR3β lDDT
(d)
ρ: -0.625
P: 1.21e-52
Residue-level RMSD (Å) in CDR3β
Residue-level Predicted lDDT in CDR3β
(e)
CDR3β pLDDT
CDR3β Length
(f)
PDB: 7T2B RMSD: 1.250
ImmuneFold
PDB: 7T2B RMSD: 1.648
TCRmodel2
PDB: 7T2B RMSD: 1.431
AlphaFold3
PDB: 7T2B RMSD: 20.612
ESMFold

Fig. 2: **Evaluation of ImmuneFold for TCR structure prediction. a**, RMSD values for CDR3β, other CDRs, and frameworks of predicted TCR structures. **b**, LRMS, iRMS, and DockQ scores of predicted TCR structures. **c,d**, Pearson correlation between true RMSD values and predicted lDDT scores of CDR3β at the per-target and per-residue level, respectively. **e**, Distribution of predicted lDDT scores for CDR3β across various lengths, based on approximately 32,000 predicted TCR structures from the VDJdb dataset. **f**, Illustrative examples of predicted TCR structures from ImmuneFold (orange), TCRmodel2 (blue), ESMFold (cyan), AlphaFold3 (magenta), and the crystal structure (grey).

2.2 Accurate TCR structure prediction with ImmuneFold

We derive the training data for TCR structure prediction from the STCRDab database [24] and evaluate performance using independent testing sets with no overlap with the training data. We compare our approach with TCR-specific prediction methods, including TCRmodel2 and ImmuneBuilder [25], as well as general protein structure prediction methods such as ESMFold, AlphaFold-Multimer [26] and AlphaFold3 [27].

On the most challenging CDR3β region, ImmuneFold achieves an RMSD of 1.31 Å (Figure 2**a**, Table S1, Figure S1), outperforming all other methods: AlphaFold3 (1.41 Å), TCRmodel2 (1.44 Å), AlphaFold-Multimer (1.57 Å), ImmuneBuilder (1.77 Å), and ESMFold (3.67 Å). Additionally, ImmuneFold attains an RMSD of 1.12 Å on the other CDRs and 0.71 Å on the conserved framework regions, again surpassing the performance of all other approaches.

We further evaluate the relative poses of the α and β chains using docking metrics [28] (Figure 2**b**). ImmuneFold achieves an iRMS of 1.43 Å an LRMS of 0.96 Å, and a DockQ score of 0.84, outperforming TCRmodel2, AlphaFold-Multimer, and AlphaFold3 across all these metrics.

Following AlphaFold2 [6, 29], ImmuneFold also predicts the local distance difference test (pLDDT) between the predicted and ground truth structure as a confidence score. The pLDDT demonstrates high Pearson correlation coefficients of 0.87 and 0.63 with RMSD at the protein and residue levels, respectively (Figure 2**c,d**), indicating that it is a reliable metric for assessing the quality of predicted structures.

We then use `7T2B` as an example to demonstrate the success of our approach (Figure 2**f**). In this case, ImmuneFold achieves an RMSD of just 1.25 Å for the entire structure, with no noticeable deviations in the CDRs between the predicted and native structures. In comparison, TCRmodel2 and AlphaFold3 result in higher RMSDs of 1.65 Å and 1.43 Å, respectively, and show clear distortions in the CDRs. ESMFold even fails to determine the correct relative orientation between the α and β chains, resulting in a significantly larger RMSD of 20.61 Å.

Furthermore, we applied ImmuneFold to predict structures for all the 32,703 TCRs gathered in VDJdb [30] and archived the predicted TCR structures into a database. As of June 2024, only 937 unique TCR-peptide structures have been experimentally determined. Our predicted database represents a 30-fold expansion of the TCR-peptide structure space (Figure 2**e**).

2.3 Zero-shot TCR-epitope binding prediction with ImmuneFold

Accurate prediction of TCR-epitope binding is critical for developing immunotherapies, such as neoepitope identification and TCR sequence engineering [1]. We integrate ImmuneFold with the Rosetta software [23] for TCR-epitope binding prediction (Figure 3**a**). It's important to note that our protocol is a zero-shot learning approach that avoids explicit training on TCR-epitope binding data. In contrast, most existing methods rely on supervised training with limited TCR-epitope binding datasets, leading to poor performance on epitopes not present in the training set [1, 31].

To evaluate ImmuneFold, we compiled a TCR-epitope binding dataset comprising eight subsets, each associated with a distinct wild-type peptide and multiple TCR sequences. Wild-type peptides were utilized as positive samples, whereas decoy peptides filtered using NetMHCpan [32] served as negative samples. We then compared the performance of ImmuneFold with that of recent representative TCR-epitope binding prediction methods, specifically PanPep [20] and TEIM [21].

Across these subsets, ImmuneFold achieved the highest AUROC of 0.69 (Figure 3**b-e**), whereas PanPep and TEIM scored 0.55 and 0.57, respectively. Notably, ImmuneFold achieved an AUROC greater than 0.6 with statistical significance (p-value < 0.001, DeLong's test) in 5 out of the 8 testing subsets, while both PanPep and TEIM reached this level of performance in only 2 subsets each.

We then use the EBV BMLF1 subset as an example to investigate the binding energy of TCRs with wild-type and decoy peptides (Figure 3**f**). The interface between the TCR and the wild-type peptide is buried with more interacting residues, whereas the decoy peptide has a mutation to `Glycine`, resulting in the interface being more exposed to solvent. Consistently, the wild-type exhibits significantly lower Rosetta binding energy than the decoy (Figure 3**g**).

2.4 Accurate prediction of antibody structures using ImmuneFold

To assess the performance of ImmuneFold in predicting antibody and nanobody structures, we compiled a test set consisting of 325 antibody and 112 nanobody structures, with no overlap with the training data. We compared ImmuneFold with general protein structure prediction models, ESMFold,

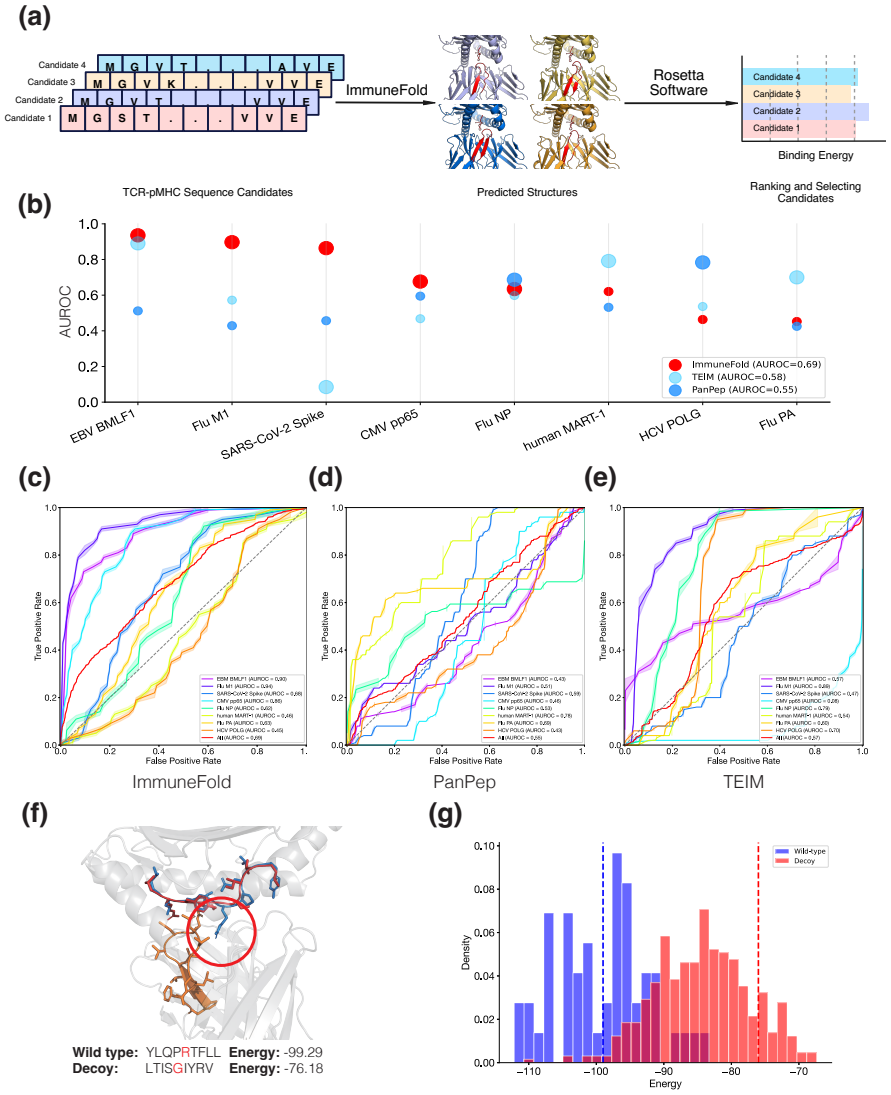

Fig. 3: **Evaluation of ImmuneFold for TCR-epitope binding prediction. a**, Framework of TCR-epitope binding prediction with ImmuneFold. **b**, Evaluation results across eight testing subsets, each associated with a wild-type peptide and multiple TCR sequences. Large dots indicate AUROC values that are statistically significant (p-value < 0.001, DeLong's test), whereas small dots represent non-significant results. **c-e**, Receiver operating characteristic (ROC) curves for ImmuneFold, PanPep, and TEIM across the eight testing subsets. **f**, Two examples of TCR-epitope complexes from EBV BMLF1 dataset with significantly different binding energies. The wild-type structure is shown in blue, and the decoy structure in red. **g**, Distribution of Rosetta binding energies for wild-type and decoy structures within the EBV BMLF1 dataset. The blue and red dashed lines represent the energies of the wild-type and decoy complexes in panel **f**, respectively.

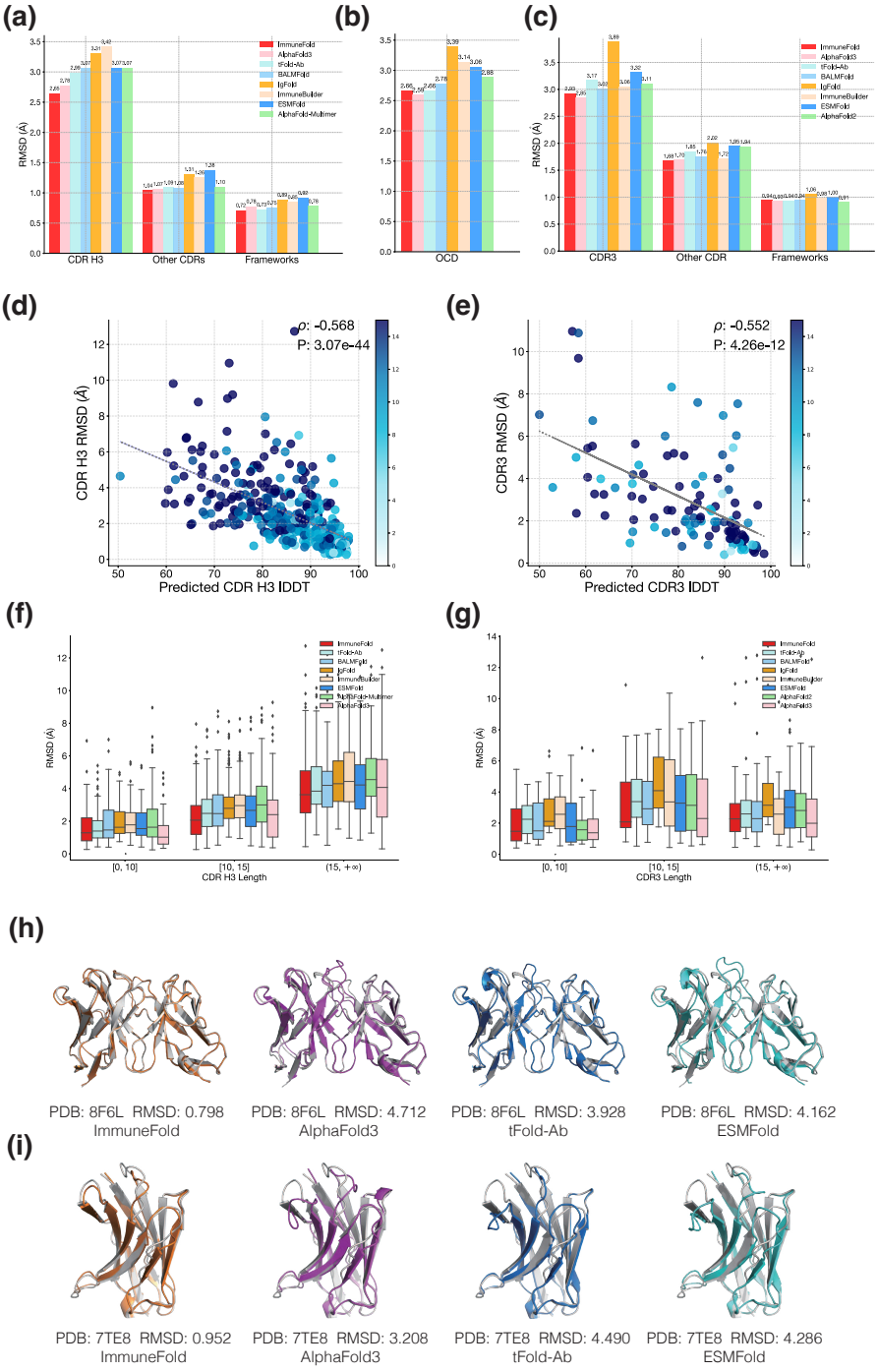
(a)
(b)
(c)
RMSD (Å)
CDR H3
Other CDRs
Frameworks
OCD
CDR3
Other CDR
ImmuneFold
AlphaFold3
tFold-Ab
BALMFold
IgFold
ImmuneBuilder
ESMFold
AlphaFold-Multimer
AlphaFold2
(d)
(e)
ρ: -0.568
P: 3.07e-44
ρ: -0.552
P: 4.26e-12
CDR H3 RMSD (Å)
CDR3 RMSD (Å)
Predicted CDR H3 IDDT
Predicted CDR3 IDDT
(f)
(g)
[0, 10]
[10, 15]
(15, +∞)
CDR H3 Length
CDR3 Length
(h)
PDB: 8F6L RMSD: 0.798
ImmuneFold
PDB: 8F6L RMSD: 4.712
AlphaFold3
PDB: 8F6L RMSD: 3.928
tFold-Ab
PDB: 8F6L RMSD: 4.162
ESMFold
(i)
PDB: 7TE8 RMSD: 0.952
ImmuneFold
PDB: 7TE8 RMSD: 3.208
AlphaFold3
PDB: 7TE8 RMSD: 4.490
tFold-Ab
PDB: 7TE8 RMSD: 4.286
ESMFold

Fig. 4: **Evaluation of ImmuneFold for antibody and nanobody structure prediction. a-b**, RMSD and OCD values of predicted antibody structures. **c**, RMSD values of predicted nanobody structures. **d-e**, Pearson correlations between true RMSD values and predicted lDDT of CDR H3 for predicted antibody and nanobody structures, respectively. **f-g**, RMSD values across various lengths of CDR H3 of predicted antibody and nanobody structures. **h-i**, Examples of predicted antibody structures (**h**) and nanobody structures (**i**) from ImmuneFold (orange), tFold (blue), ESMFold (cyan), AlphaFold3 (magenta), and the crystal structure (grey), where the RMSD values on CDR H3 of antibody and CDR3 of nanobody are labeled.

AlphaFold-Multimer, AlphaFold2, AlphaFold3, as well as other antibody-specific methods, including IgFold, BALMFold [33], ImmuneBuilder, and tFold-Ab [9].

ImmuneFold is the most accurate in predicting the highly variable CDR H3 region of antibodies, achieving an RMSD of 2.65 Å (Figure 4**a,c-d**). The second-best method, AlphaFold3, has an RMSD of 2.78 Å, which is approximately 5% less accurate than ImmuneFold. ImmuneFold also delivers the best performance across the remaining CDRs, the framework regions of antibodies. It achieves comparable performance with AlphaFold3 for nanobodies and outperforms other methods (Figure 4**a-d**, Figure S2). It is noteworthy that tFold-Ab, IgFold, and BALMFold all rely on language models specifically trained on large antibody sequence datasets, whereas ImmuneFold employs ESM2, a general language model trained on diverse protein families. This highlights the ability of general models to efficiently capture evolutionary constraints in antibodies, consistent with previous studies where antibodies are more successfully designed and validated in wet-lab experiments using general models rather than antibody-specific ones [34, 35].

We further evaluate the predicted orientation between heavy and light chains, an important determinant of the overall binding surface. We utilize the orientational coordinates distance (OCD) metric, which summarizes deviations in inter-chain packing angle, inter-domain angle, and heavy-opening and light-opening angles [36]. Among these methods, AlphaFold3 achieved the most accurate orientation predictions, with an OCD of 2.59, ImmuneFold and tFold-Ab closely followed with an OCD of 2.66 (Figure 4**a**). Notably, while AlphaFold3 relies on multiple sequence alignments (MSAs) and template searches, ImmuneFold and tFold-Ab operate on single sequences, enabling significantly more efficient antibody structure prediction.

Consistent with the results for TCR structure prediction, ImmuneFold's predicted lDDT scores exhibit strong Pearson correlations with the true RMSD values at the target level, with coefficients of 0.57 for antibodies and 0.55 for nanobodies (Figure 4**d**–**e**). Furthermore, ImmuneFold maintained consistently comparable accuracy across all CDR H3 and CDR3 lengths compared to AlphaFold3, outperforming other methods (Figure 4**f**–**g**).

We then present two successful examples from ImmuneFold (Figure 4**h,i**). In these cases, the CDR H3 region in the antibody and CDR3 region in the nanobody predicted by ESMFold, tFold-Ab and AlphaFold3 deviate significantly from the native structures and even exhibit steric clashes. In contrast, ImmuneFold provides much more accurate conformations, achieving RMSD values of 0.80 Å for the antibody CDR H3 and 0.95 Å for the nanobody CDR3.

2.5 Evaluation of antibody and nanobody structure prediction in the presence of antigens

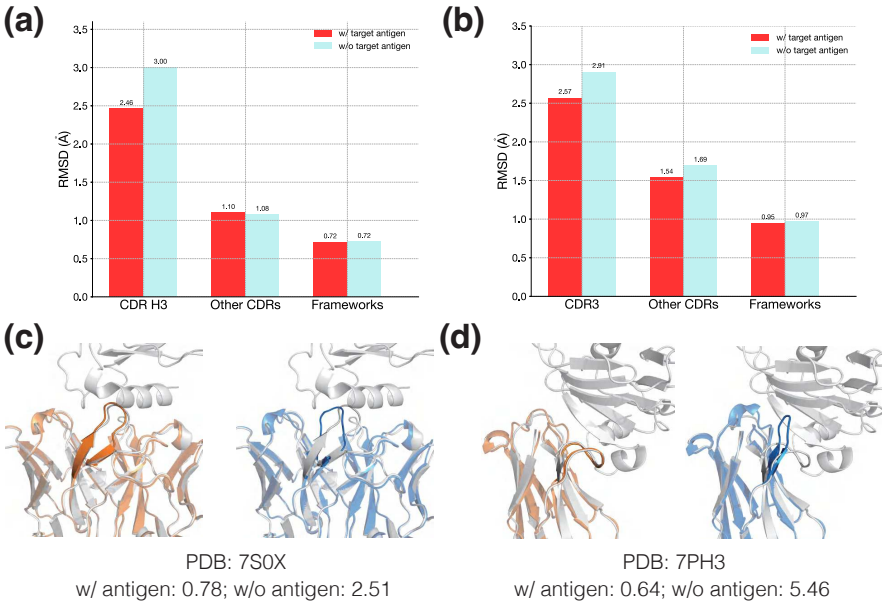

Fig. 5: **Evaluation of antibody and nanobody structure prediction in the presence of target antigens. a,b**, RMSD values of predicted antibody and nanobody structures when modeled with their target antigens. **c,d**, Examples of predicted antibody and nanobody structures, with RMSD values of CDR H3 and CDR3 labeled, respectively. Predictions with antigens are shown in orange, without antigens in blue, and crystal structures are in grey.

We further investigated how structure prediction accuracy is affected by antigen context. To enable ImmuneFold to consider antigen context during prediction, we provided it with the distogram of the antigen structure as initial recycling features (Methods). We then compared the prediction accuracy of bound antibody and nanobody structures against unbound ones.

When the antigen structure was provided, ImmuneFold reduced the RMSD of predicted antibody CDR H3 from 2.91 Å to 2.57 Å, and that of nanobody

CDR3, from 3.00 Å to 2.46 Å (Figure 5**a**,**b**). This indicates its potential application in both antibody virtual screening in the presence of antigen [37, 38] and antibody thermostability optimization in the absence of antigen [39].

We next present two notable cases (Figure 5**c**,**d**) where ImmuneFold significantly improved prediction accuracy with the antigen structure provided, achieving RMSD values of less than 1 Å for the CDR H3 of the antibody and CDR3 of the nanobody.

2.6 Training and inference efficiency of ImmuneFold

Previous studies have shown that the parameter-efficient fine-tuning technique LoRA can significantly improve the training efficiency of large-scale Transformer-based models while maintaining prediction accuracy [12, 14]. Here, we investigate the training efficiency of LoRA for fine-tuning Evoformer-like model architectures compared to full-parameter fine-tuning.

We found that LoRA reduces memory usage by nearly half while maintaining almost the same prediction accuracy (Figure 6**a,b**). With the same computational resources and data throughput, LoRA significantly improves training efficiency. This makes it possible for academic labs and other research groups lacking substantial computational resources to fine-tune recently developed, much larger protein structure prediction models [27].

During inference, ImmuneFold requires approximately 3 seconds on average to predict the structure of an immune protein on an Nvidia A100 GPU. Due to the absence of MSA searching, it achieves up to a 1000-fold speedup compared to MSA-based methods such as AlphaFold3 [27], AlphaFold-Multimer [26], and TCRmodel2 [40] (Figure 6**c**). Notably, LoRA does not introduce additional computational overhead during inference compared to ESMFold, and ImmuneFold demonstrates greater efficiency than other MSA-free approaches. Furthermore, we show that our model exhibits high computational efficiency on antibody and nanobody test sets (Figure S3).

Discussion

We propose ImmuneFold, a novel method for immune protein structure prediction that efficiently fine-tunes ESMFold specifically on immune protein structures. Our results demonstrate that ImmuneFold outperforms existing methods in predicting the structures of various immune proteins, including T-cell receptors, antibodies, and nanobodies. Notably, when integrated with Rosetta software, ImmuneFold can accurately predict TCR-epitope binding, indicating its potential for real-world applications in immunotherapy.

ImmuneFold and other recent deep learning–based methods [8, 25, 33] only predict static structures of immune proteins. However, T-cell receptors and antibodies can exhibit conformational changes upon binding to specific antigens [41–47]. ImmuneFold can be extended in two ways to address this challenge. First, incorporating molecular dynamic simulation data for immune

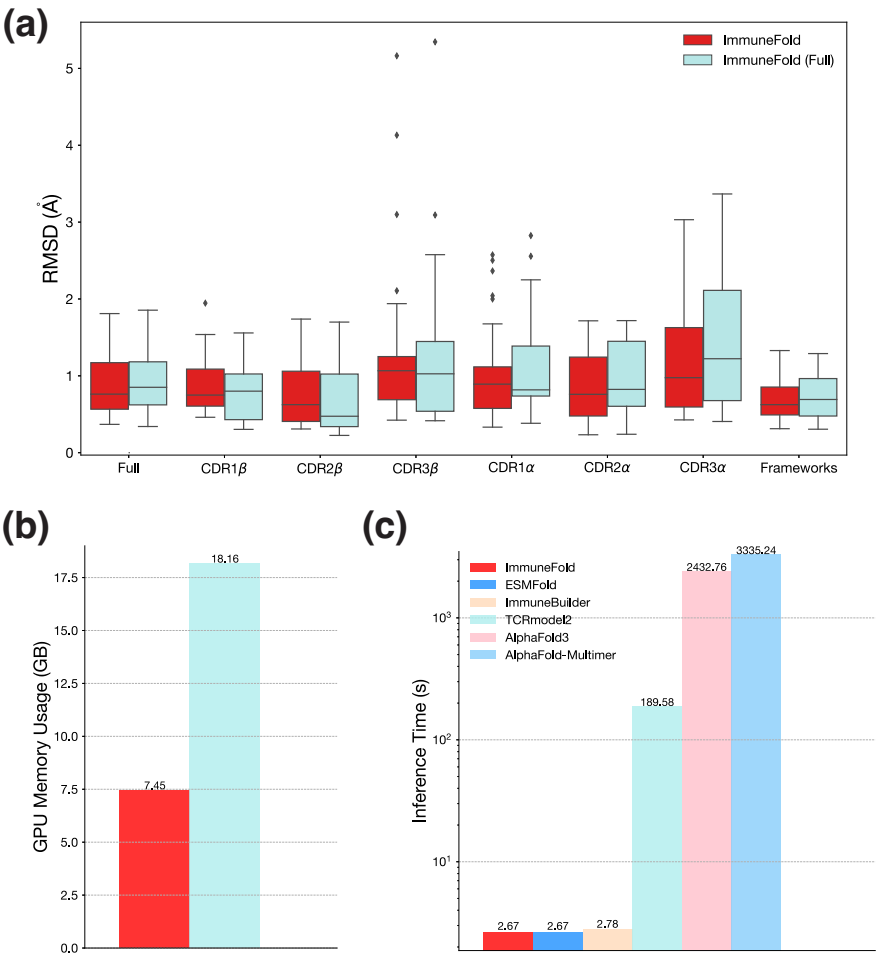

Fig. 6: **Evaluation of ablation study. a**, Evaluation on RMSD for CDR3β, other CDRs, and frameworks for the TCR test sets. **b**, Evaluation of GPU memory usage under different configurations. **c**, Evaluation of inference time across different methods.

proteins during training could enable the model to better capture the conformational diversity of these proteins. Second, our strategy can be extended to fine-tune generative models for protein structure prediction, allowing protein conformations sampling rather than predicting a single static structure [27, 48].

While we have demonstrated that ImmuneFold improves TCR–epitope binding prediction, further efforts are needed to enhance antibody-antigen docking performance, especially when the antigen structure is large and the epitope on the antigen is unknown. Simply fine-tuning ESMFold is insufficient for this task. Antibodies can target a wide range of foreign proteins and

may form considerably large antibody-antigen complexes, whereas TCRs typically target peptides of only 10–20 amino acids. However, ESMFold and its underlying language model, ESM-2, are trained only on monomeric proteins, making it difficult to capture the inter-chain evolutionary constraints critical for modeling large protein complexes [49, 50]. Our future work will primarily focus on improving antibody-antigen docking performance by combining the advantages of ImmuneFold with other cutting-edge methods. For instance, we will explore integrating ImmuneFold with physical energy-based docking methods, as adopted in previous studies [9, 51], or enrich our training dataset with docked protein complex structures [38, 52].

3 Methods

3.1 Datasets

TCR structure datasets. We derived the training and test sets for TCR structure prediction from the STCRDab database [24]. We excluded MHC Class II complex structures, antigen chains whose antigen type was not a peptide, and TCRs lacking the β chain. The dataset was partitioned into training and test sets based on the release date of October 1, 2022. Following previous work [40], we remove the overlap between the training and testing set using the threshold of 99% β sequence similarity for a fair comparison. As a result, the final training and test sets comprised 730 and 27 complexes, respectively.

Antibody and nanobody structure datasets. For antibody and nanobody structure prediction, we construct training and test sets from the SAbDab database [11]. The dataset is split into training and testing sets based on the release date of 1 January 2022. We remove the overlap between the training and testing set using a threshold of 90% sequence similarity. Additionally, we applied a 90% sequence identity threshold within the test set to remove redundant sequences. The final dataset includes a total of 9829 antibody and nanobody structures for training, and 325 antibodies and 112 nanobodies for testing, respectively.

TCR-epitope binding dataset. Following the strategy of previous methods [20, 53], we compiled a testing set for TCR-epitope binding prediction, consisting of 382 positive and 3,438 negative samples. Specifically, the testing set includes eight subsets, each associated with a distinct wild-type peptide and multiple TCR sequences. These wild-type TCR-peptide complexes were utilized as positive samples. For each positive sample, we filtered and selected nine decoy peptides of equivalent lengths using NetMHCpan [32].

Large-scale TCR sequence data for structure prediction. We collected a dataset of 32,703 TCR-peptide sequences from the VDJdb database [30], including only sequences associated with MHC Class I complexes.

3.2 Parameter-efficient transfer learning on ESMFold

ESMFold [50] leverages a large-scale protein language model for protein structure prediction and consists of three main components: a large protein language model, the `Folding Trunk module`, and the `Structure module`. The Folding Trunk is an Evoformer-like architecture adapted from AlphaFold, designed to learn single and pairwise representations of evolutionary and geometric constraints.

Transfer learning can be categorized into four main approaches: instance-based, feature-based, model-based, and relation-based methods [54]. Given the source and target domains, we adopt model-based approaches, which involve fine-tuning the base model ESMFold on immune protein datasets. In our approach, we fine-tune ESMFold on immune protein structures using Low-Rank Adaptation (LoRA) [12]. Specifically, we apply LoRA to the weight matrices of the `Linear` layers within the Folding Trunk, including the `AttentionWithPairBias`, `OutProductMean`, `Transition`, and `TriangularUpdate` modules [6, 50]. All parameters of the language model and the structure module remain frozen during this process.

LoRA restricts parameter updates to a low-rank matrix ΔW, which is added to the original weight matrix $W_0 \in \mathbb{R}^{m\times n}$. This low-rank matrix is factorized into two smaller matrices $B \in \mathbb{R}^{m\times r}$ and $A \in \mathbb{R}^{r\times n}$, where $r \ll m$ and $r \ll n$. The forward computation is expressed as:

$$h(x) = W_0 x + \Delta W x = W_0 x + BAx. \tag{1}$$

Here, B is initialized as zeros, and A is initialized from a `Kaiming Uniform` distribution [55]. After training, the learned A and B are merged into W_0, resulting in no additional computational overhead for inference.

We set the rank r to 16 for single representations and 8 for pair representations. This configuration reduces the number of trainable parameters to 2.7% of that required for full-parameter fine-tuning, enabling efficient adaptation of ESMFold to immune protein structures with minimal computational cost.

Following ESMFold, we employ an offset of 512 for relative positional encoding between different chains, supplying these as features to the `Evoformer` module. However, unlike ESMFold, which adds a linker of 25 glycines before extracting ESM2 language embeddings, we adapt the code of the original ESM2 model to handle position offsets directly.

In conclusion, our approach offers two main advantages. First, we initialize the model weights from ESMFold, enabling the use of learned presentations from massive protein structures beyond immune proteins. Second, LoRA updates only two low-rank matrices instead of the entire weight matrix, significantly reducing the number of trainable parameters [12, 56]. Compared to full-parameter fine-tuning, which requires extensive computational resources, LoRA is more accessible for academic research groups without sacrificing prediction performance [15, 57].

3.3 Immune protein structure prediction with antigen context

ImmuneFold supports structure prediction of bound antibodies or TCRs in the presence of specific antigens. We adapt the original ESMFold architecture to incorporate antigen structure during the network recycling stage.

AlphaFold2 [6] and subsequent methods like ESMFold [50] utilize network recycling to iteratively refine predicted structures, significantly improving prediction accuracy. In this process, the distogram of residue-residue distances and other features from the previous predictions are fed back into the model, providing feedback for further refinement.

For ImmuneFold, in the first recycling cycle, we initialize the antibody or TCR distogram with zeros and compute the distogram for the antigen structure directly. In subsequent cycles, the antibody or TCR distogram is updated with the latest predictions, while the antigen's distogram remains unchanged.

3.4 Training losses and other hyper-settings

Following AlphaFold-Multimer [26], we incorporate the interface Frame Aligned Point Error (FAPE) into the original loss functions used by ESMFold. The training loss is expressed as:

$$\mathcal{L} = 0.3\mathcal{L}_{\text{distogram}} + 1.0\mathcal{L}_{\text{FAPE}} + 1.0\mathcal{L}_{\text{interface}} + 0.01\mathcal{L}_{\text{pLDDT}} + 0.01\mathcal{L}_{\text{pTM}} \quad (2)$$

We train a single model for antibody and nanobody structure predictions. To adapt to various application scenarios, we randomly remove either the light chain or the antigen chain in antibody-antigen complex structures during training. To accommodate cases where the antigen structure is available in the application, we replace the predicted antigen structure with the ground truth during the calculation of distogram features in the recycling stage [6].

For training in the TCR dataset, we augment the data by randomly removing the peptide chain, the MHC chain, or both.

We utilize the Adam optimizer [58] with an initial learning rate of 5×10^{-4}. Each model is trained for 2×10^4 steps with a batch size of 32 structures. LoRA fine-tuning requires ∼10 days per model using eight Nvidia A800 GPUs, whereas the full-parameter fine-tuning under the same training settings requires ∼20 days per model.

3.5 TCR-epitope binding prediction

For TCR-epitope binding prediction, we first predict the complex structure using ImmuneFold and then evaluate the binding energies using Rosetta software. We directly compute the binding free energy (ΔG) of the predicted complex structures, following a method similar to that used in ATLAS [59]:

$$\Delta G = \Delta G_{\text{TCR-pMHC}} - (\Delta G_{\text{TCR-MHC}} + \Delta G_{\text{peptide}}) \quad (3)$$

where $\Delta G_{\mathrm{TCR-pMHC}}$ denotes the free energy of the entire TCR-pMHC complex, computed using the `ref2015` scoring function [59] in PyRosetta [60]. The terms $\Delta G_{\mathrm{TCR-MHC}}$ and $\Delta G_{\mathrm{peptide}}$ represent the energies for isolated structures of peptide and TCR-MHC, respectively.

3.6 Evaluation metrics

To evaluate the performance of immune structure predictions, we compute the root mean squared deviation (RMSD) between predicted and experimentally determined structures, focusing solely on backbone heavy atoms (N, C_α, C, O). Specifically, we align the entire variable (Fv) regions of the predicted and crystal structures using the `Kabsch algorithm` [61], then compute the RMSD separately for each CDR and framework region. We define the CDR and framework regions using the IMGT numbering scheme [62] for TCR structures and the Chothia numbering scheme [63] for antibody and nanobody structures.

To assess the relative positioning between the α and β chains in TCRs, we employ the interface RMSD (iRMS), ligand RMSD (LRMS), and DockQ metrics [28, 64]. For antibodies, we utilize the Orientational Coordinate Distance (OCD) to evaluate the relative position between heavy and light chains [36].

For the evaluation of TCR-epitope binding prediction, we use AUROC to assess sorting accuracy. AUROC measures the area under the curve with the false positive rate (FPR) on the x-axis and the true positive rate on the y-axis.

Data Availability

The training and test datasets utilized in this study are derived from the STCRDab (https://opig.stats.ox.ac.uk/webapps/stcrdab-stcrpred/) and SAbDab (https://opig.stats.ox.ac.uk/webapps/sabdab-sabpred/sabdab) databases. Data of TCR-epitope binding prediction is obtained from the link (https://github.com/phbradley/TCRdock/tree/main/datasets_from_the_paper). Additionally, the VDJdb database is accessed via its website (https://vdjdb.cdr3.net/).

Code Availability

The ImmuneFold software is available on GitHub at https://github.com/CarbonMatrixLab/immunefold.

Acknowledgments

We acknowledge the financial support from the National Key R&D program of China (grant no. 2023YFF1205103) and the National Natural Science Foundation of China (grant no. 32370657). We also acknowledge the financial support from the Development Program of China (grant no. 2020YFA0907000) and the National Natural Science Foundation of China (grant nos. 32271297

and 62072435) to D.B. We thank Beijing Paratera Co., Ltd, and the ICT Computing-X Center for providing computational resources.

We also thank Dr. Yufeng Shen and Chungong Yu for the valuable discussions on our work.

Author Contributions

H.Z. conceived the project and reimplemented ESMFold with LoRA integration. T.Z. and M.R. adapted the training procedures for structure prediction of antibody-antigen complexes and TCR–pMHC complexes, and developed the zero-shot protocol for TCR–epitope binding prediction. All authors contributed to the experimental design, while T.Z., M.R., and Z.H. conducted the experiments and wrote the manuscript. H.Z., D.B., J.Z., M.L., and S.T. reviewed and revised the manuscript.

Competing interests

The authors declare no competing interests.

References

[1] Weber, A., Pélissier, A. & Martínez, M. R. T-cell receptor binding prediction: A machine learning revolution. *ImmunoInformatics* 100040 (2024) .

[2] Kaplon, H., Chenoweth, A., Crescioli, S. & Reichert, J. M. *Antibodies to watch in 2022*, Vol. 14, 2014296 (2022).

[3] Lu, R.-M. *et al.* Development of therapeutic antibodies for the treatment of diseases. *Journal of biomedical science* **27**, 1–30 (2020) .

[4] Singh, N. Approval of the first TCR-based cell therapy. *Molecular Therapy* (2024) .

[5] Jovčevska, I. & Muyldermans, S. The therapeutic potential of nanobodies. *BioDrugs* **34** (1), 11–26 (2020) .

[6] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596** (7873), 583–589 (2021) .

[7] Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373** (6557), 871–876 (2021) .

[8] Ruffolo, J. A., Chu, L.-S., Mahajan, S. P. & Gray, J. J. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications* **14** (1), 2389 (2023) .

[9] Wu, F. *et al.* Fast and accurate modeling and design of antibody-antigen complex using tFold. *bioRxiv* 2024–02 (2024) .

[10] Olsen, T. H., Boyles, F. & Deane, C. M. Observed Antibody Space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Science* **31** (1), 141–146 (2022) .

[11] Dunbar, J. *et al.* SAbDab: the structural antibody database. *Nucleic acids research* **42** (D1), D1140–D1146 (2014) .

[12] Hu, E. J. *et al.* *LoRA: Low-rank adaptation of large language models* (2022).

[13] Zeng, S., Wang, D., Jiang, L. & Xu, D. Parameter-efficient fine-tuning on large protein language models improves signal peptide prediction. *Genome Research* gr–279132 (2024) .

[14] Sledzieski, S. *et al.* Democratizing protein language models with parameter-efficient fine-tuning. *Proceedings of the National Academy of Sciences* **121** (26), e2405840121 (2024) .

[15] Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. *High-resolution image synthesis with latent diffusion models*, 10684–10695 (2022).

[16] Zeng, S., Wang, D., Jiang, L. & Xu, D. *Prompt-based learning on large protein language models improves signal peptide prediction*, 400–405 (Springer, 2024).

[17] Chen, B. *et al.* xTrimoPGLM: unified 100B-scale pre-trained transformer for deciphering the language of protein. *arXiv preprint arXiv:2401.06199* (2024) .

[18] Wang, Z. & Shen, Y. SABRE: Self-Attention Based model for predicting T-cell Receptor Epitope Specificity. *bioRxiv* 2023–10 (2023) .

[19] La Gruta, N. L., Gras, S., Daley, S. R., Thomas, P. G. & Rossjohn, J. Understanding the drivers of MHC restriction of T cell receptors. *Nature Reviews Immunology* **18** (7), 467–478 (2018) .

[20] Gao, Y. *et al.* Pan-peptide meta learning for T-cell receptor–antigen binding recognition. *Nature Machine Intelligence* **5** (3), 236–249 (2023) .

[21] Peng, X. *et al.* Characterizing the interaction conformation between T-cell receptors and epitopes with deep learning. *Nature Machine Intelligence* **5**, 1–13 (2023) .

[22] Gielis, S. *et al.* Detection of enriched T cell epitope specificity in full T cell receptor sequence repertoires. *Frontiers in immunology* **10**, 2820 (2019) .

[23] Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation* **13** (6), 3031–3048 (2017) .

[24] Leem, J., de Oliveira, S. H. P., Krawczyk, K. & Deane, C. M. STCRDab: the structural T-cell receptor database. *Nucleic acids research* **46** (D1), D406–D412 (2018) .

[25] Abanades, B. *et al.* ImmuneBuilder: Deep-Learning models for predicting the structures of immune proteins. *Communications Biology* **6** (1), 575 (2023) .

[26] Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer. *biorxiv* 2021–10 (2021) .

[27] Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 1–3 (2024) .

[28] Basu, S. & Wallner, B. DockQ: a quality measure for protein-protein docking models. *PlOS ONE* **11** (8), e0161879 (2016) .

[29] Mariani, V., Biasini, M., Barbato, A. & Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29** (21), 2722–2728 (2013) .

[30] Shugay, M. *et al.* VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic acids research* **46** (D1), D419–D427 (2018) .

[31] Dens, C., Laukens, K., Bittremieux, W. & Meysman, P. The pitfalls of negative data bias for the T-cell epitope specificity challenge. *bioRxiv* (2023) .

[32] Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic acids research* **48** (W1), W449–W454 (2020) .

[33] Jing, H. *et al.* Accurate prediction of antibody function and structure using bio-inspired antibody language model. *Briefings in Bioinformatics* **25** (4), bbae245 (2024) .

[34] Hie, B. L. *et al.* Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology* (2023) .

[35] Shanker, V. R., Bruun, T. U., Hie, B. L. & Kim, P. S. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science* **385** (6704), 46–53 (2024) .

[36] Marze, N. A., Lyskov, S. & Gray, J. J. Improved prediction of antibody VL–VH orientation. *Protein Engineering, Design and Selection* **29** (10), 409–418 (2016) .

[37] Honorato, R. V. *et al.* The HADDOCK2. 4 web server for integrative modeling of biomolecular complexes. *Nature Protocols* 1–23 (2024) .

[38] Harmalkar, A., Lyskov, S. & Gray, J. J. Reliable protein-protein docking with AlphaFold, Rosetta, and replica-exchange. *eLife* **13** (2024) .

[39] Hutchinson, M. *et al.* Enhancement of antibody thermostability and affinity by computational design in the absence of antigen. *bioRxiv* 2023–12 (2023) .

[40] Yin, R. *et al.* TCRmodel2: high-resolution modeling of T cell receptor recognition using deep learning. *Nucleic Acids Research* **51** (W1), W569–W576 (2023) .

[41] Adhikary, R., Yu, W., Oda, M., Zimmermann, J. & Romesberg, F. E. Protein dynamics and the diversity of an antibody response. *Journal of Biological Chemistry* **287** (32), 27139–27147 (2012) .

[42] Jimenez, R., Salazar, G., Baldridge, K. K. & Romesberg, F. E. Flexibility and molecular recognition in the immune system. *Proceedings of the National Academy of Sciences* **100** (1), 92–97 (2003) .

[43] Thielges, M. C., Zimmermann, J., Yu, W., Oda, M. & Romesberg, F. E. Exploring the energy landscape of antibody- antigen complexes: protein dynamics, flexibility, and molecular recognition. *Biochemistry* **47** (27), 7237–7247 (2008) .

[44] Tang, Y. & Cao, Y. Modeling the dynamics of antibody–target binding in living tumors. *Scientific Reports* **10** (1), 16764 (2020) .

[45] Jay, J. W. *et al.* IgG antibody 3D structures and dynamics. *Antibodies* **7** (2), 18 (2018) .

[46] Kass, I., Buckle, A. M. & Borg, N. A. Understanding the structural dynamics of TCR-pMHC interactions. *Trends in immunology* **35** (12), 604–612 (2014) .

[47] Adhikary, R. *et al.* Structure and dynamics of stacking interactions in an antibody binding site. *Biochemistry* **58** (27), 2987–2995 (2019) .

[48] Jing, B., Berger, B. & Jaakkola, T. *AlphaFold Meets Flow Matching for Generating Protein Ensembles* (2024).

[49] Zhu, J., He, Z., Li, Z., Ke, G. & Zhang, L. Uni-Fold MuSSe: De Novo Protein Complex Prediction with Protein Language Models. *bioRxiv* 2023–02 (2023) .

[50] Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **379** (6637), 1123–1130 (2023) .

[51] Harmalkar, A. *et al.* Docking with Rosetta and deep learning approaches in CAPRI rounds 47-55 (2024) .

[52] Hitawala, F. N. & Gray, J. J. What has AlphaFold3 learned about antibody and nanobody docking, and what remains unsolved? *bioRxiv* 2024–09 (2024) .

[53] Bradley, P. Structure-based prediction of T cell receptor: peptide-MHC interactions. *eLife* **12**, e82813 (2023) .

[54] Iman, M., Arabnia, H. R. & Rasheed, K. A review of deep transfer learning and recent advancements. *Technologies* **11** (2), 40 (2023) .

[55] He, K., Zhang, X., Ren, S. & Sun, J. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*, 1026–1034 (2015).

[56] Lin, Y. *et al.* LoRA Dropout as a Sparsity Regularizer for Overfitting Reduction. *arXiv preprint arXiv:2404.09610* (2024) .

[57] Touvron, H. *et al.* Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288* (2023) .

[58] Kingma, D. P. & Ba, J. *Adam: A method for stochastic optimization*, International Conference on Learning Representations (2015). URL https://doi.org/10.48550/arXiv.1412.6980.

[59] O'Meara, M. J. *et al.* Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with rosetta. *Journal of chemical theory and computation* **11** (2), 609–622 (2015) .

[60] Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26** (5), 689–691 (2010) .

[61] Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography* **32** (5), 922–923 (1976) .

[62] Lefranc, M.-P. *et al.* IMGT, the international ImMunoGeneTics information system. *Nucleic acids research* **37** (suppl_1), D1006–D1012 (2009) .

[63] Dondelinger, M. *et al.* Understanding the significance and implications of antibody numbering and antigen-binding surface/residue definition. *Frontiers in immunology* **9**, 2278 (2018) .

[64] Méndez, R., Leplae, R., De Maria, L. & Wodak, S. J. Assessment of blind predictions of protein–protein interactions: current status of docking methods. *Proteins: Structure, Function, and Bioinformatics* **52** (1), 51–67 (2003) .

4 Supplementary Materials

4.1 Details on running the compared methods

ImmuneBuilder: We utilized the test script provided in the ImmuneBuilder GitHub repository (https://github.com/oxpig/ImmuneBuilder) and conducted TCRBuilder2, AbBuilder2, and NanoBuilder2 for TCR, antibody, and nanobody structure prediction, respectively, using the default settings.

TCRmodel2: We used the test script provided in the TCRmodel2 GitHub repository (https://github.com/piercelab/tcrmodel2). For a fair comparison, we employed AlphaFold-Multimer in TCRmodel2 with only model 1, rather than an ensemble of 5 models, and set the template time cutoff to 1 October 2022.

AlphaFold-Multimer: We utilized the test script provided in the AlphaFold repository (https://github.com/google-deepmind/alphafold). For a fair comparison, we used AlphaFold-Multimer with only model 1 and set the template time cutoff to 1 January 2022 for antibodies and 1 October 2022 for T-cell receptors.

AlphaFold2: We used the test script from the AlphaFold repository (https://github.com/google-deepmind/alphafold). For a fair comparison, we utilized only model 1 and set the template time cutoff to January 1, 2022.

AlphaFold3: We employed the test script provided in the AlphaFold3 repository (https://github.com/google-deepmind/alphafold3) and utilized their pre-trained checkpoint directly. To ensure robustness, we executed five independent runs using different random seeds and selected the best result based on the `ranking_score`.

ESMFold: We utilized the test script provided in the ESM GitHub repository (https://github.com/facebookresearch/esm) with the default settings.

IgFold: We utilized the test script provided in the IgFold GitHub repository (https://github.com/Graylab/IgFold) with the default settings.

BALMFold: We utilized all predicted structures provided by the BALMFold server (https://beamlab-sh.com/models/BALMFold).

tFold-Ab: We utilized the test script provided in the tFold GitHub repository (https://github.com/TencentAI4S/tfold) with all default parameters.

PanPep: We utilized the test script provided in the Panpep GitHub repository (https://github.com/bm2-lab/PanPep) in the zero-shot setting.

TEIM: We utilized the test script provided in the TEIM-seq GitHub repository (https://github.com/pengxingang/TEIM) with the sequence-level interaction.

Supplementary Figures

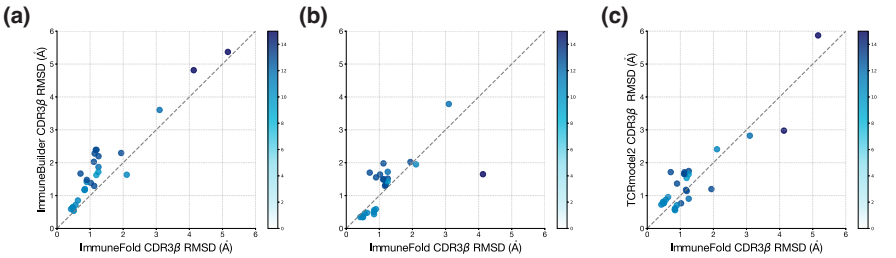

Fig. S1: Head-to-head comparison with ImmuneBuilder, AlphaFold3 and TCRmodel2 on T-cell receptor dataset.

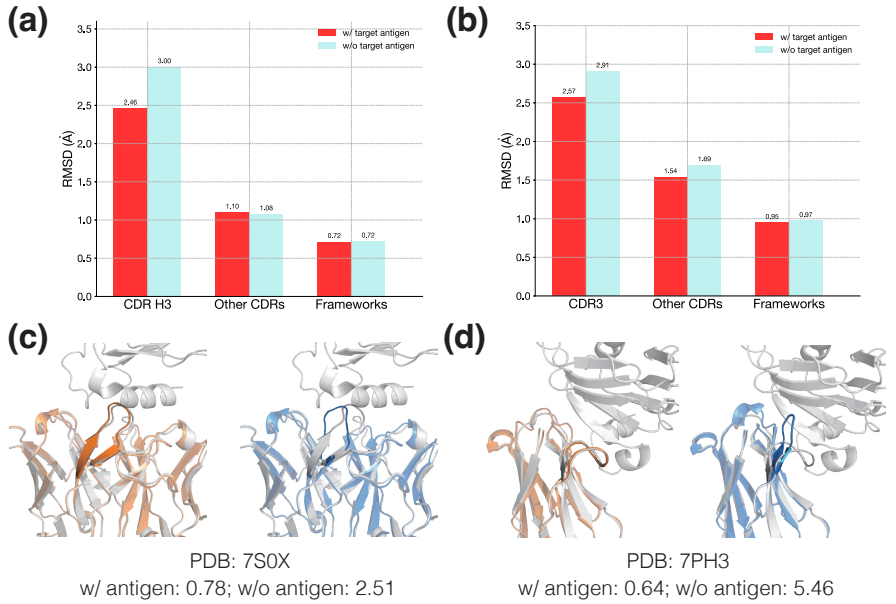

Fig. S2: Head-to-head comparison with ImmuneBuilder, AlphaFold3 and tFold-Ab. **a-b**, Results on antibody dataset. **c-d**, Results on nanobody dataset.

Supplementary Tables

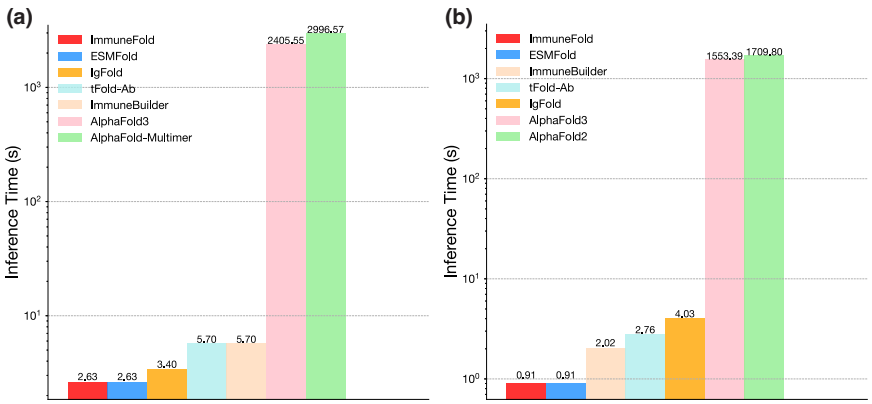

Fig. S3: Evaluation of inference time across different methods. **a**, inference time on antibody test dataset. **b**, inference time on nanobody test dataset.

Table S1: TCR structure prediction accuracy.

Methods	**CDR1β**	**CDR2β**	**CDR3β**	**CDR1α**	**CDR2α**	**CDR3α**	**Framework**
ESMFold	2.69	3.10	3.67	2.66	2.97	3.44	2.21
ImmuneBuilder	**0.85**	0.90	1.77	1.35	1.04	1.92	1.03
TCRmodel2	0.93	**0.75**	1.44	**0.94**	1.00	1.42	0.86
AlphaFold-Multimer	1.05	0.90	1.57	1.13	0.98	1.70	1.01
AlphaFold3	0.98	1.00	1.41	1.01	1.02	1.78	0.90
ImmuneFold	0.87	0.76	**1.31**	1.06	**0.88**	**1.27**	**0.71**

Table S2: Neoepitope benchmark.

Organism	**MHC**	**Epitope length**	**Eptiope sequence**	**Antigen**
human	HLA-A*02:01	9	GILGFVFTL	Flu M1
human	HLA-A*02:01	9	GLCTLVAML	EBV BMLF1
human	HLA-A*02:01	9	NLVPMVATV	CMV pp65
human	HLA-A*02:01	9	YLQPRTFL	SARS-CoV-2 Spike
human	HLA-A*02:01	10	ELAGIGILTV	human MART-1
human	HLA-A*02:01	10	KLVALGIGAV	HCV POLG
mouse	H2-Db	9	ASNENMETM	Flu NP
mouse	H2-Db	10	SSLENFRAYV	Flu PA

Table S3: Antibody structure prediction accuracy.

Methods	CDR H1	CDR H2	CDR H3	CDR L1	CDR L2	CDR L3	Framework
ESMFold	1.37	1.33	3.07	1.24	0.88	1.36	0.92
AlphaFold-Multimer	1.16	1.05	3.07	0.91	0.77	1.08	0.78
ImmuneBuilder	1.38	1.23	3.42	1.03	0.93	1.20	0.85
IgFold	1.38	1.22	3.31	1.10	0.91	1.39	0.89
tFold-Ab	1.20	1.04	2.99	0.88	0.77	1.06	0.73
BALMFold	1.18	1.03	3.07	0.86	0.76	1.08	0.75
AlphaFold3	**1.14**	1.02	2.78	0.88	**0.75**	1.02	0.78
ImmuneFold	**1.14**	**1.00**	**2.65**	**0.65**	**0.75**	**1.00**	**0.72**

Table S4: Nanobody structure prediction accuracy.

Methods	CDR1	CDR2	CDR3	Framework
ESMFold	2.06	1.54	3.32	1.00
AlphaFold2	2.14	1.42	3.11	**0.91**
ImmuneBuilder	1.82	1.44	3.06	0.98
IgFold	2.14	1.65	3.89	1.06
tFold-Ab	1.99	1.47	3.17	0.94
BALMFold	1.85	1.44	3.02	0.94
AlphaFold3	1.83	**1.31**	**2.85**	0.93
ImmuneFold	**1.79**	1.34	2.93	0.94

Table S5: Antibody structure prediction accuracy given antigen contexts.

Methods	CDR H1	CDR H2	CDR H3	CDR L1	CDR L2	CDR L3	Framework
w/ antigen	1.11	0.98	2.46	0.95	0.71	1.03	0.72
w/o antigen	1.18	1.08	3.00	0.86	0.76	1.05	0.72

Table S6: Nanobody structure prediction accuracy given antigen contexts.

Methods	CDR1	CDR2	CDR3	Framework
w/ antigen	1.65	1.25	2.57	0.95
w/o antigen	1.81	1.34	2.91	0.97