

Improved Antibody-Specific Epitope Prediction Using AlphaFold and AbAdapt**

Zichang Xu,^[a] Ana Davila,^[a] Jan Wilamowski,^[a] Shunsuke Teraguchi,^[a, b] and Daron M. Standley^{*,[a]}

Antibodies recognize their cognate antigens with high affinity and specificity, but the prediction of binding sites on the antigen (epitope) corresponding to a specific antibody remains a challenging problem. To address this problem, we developed AbAdapt, a pipeline that integrates antibody and antigen structural modeling with rigid docking in order to derive antibody-antigen specific features for epitope prediction. In this study, we systematically assessed the impact of integrating the state-of-the-art protein modeling method AlphaFold with the AbAdapt pipeline. By incorporating more accurate antibody

models, we observed improvement in docking, paratope prediction, and prediction of antibody-specific epitopes. We further applied AbAdapt-AF in an anti-receptor binding domain (RBD) antibody complex benchmark and found AbAdapt-AF outperformed three alternative docking methods. Also, AbAdapt-AF demonstrated higher epitope prediction accuracy than other tested epitope prediction tools in the anti-RBD antibody complex benchmark. We anticipate that AbAdapt-AF will facilitate prediction of antigen-antibody interactions in a wide range of applications.

Introduction

Highly specific antibody-antigen interactions are a defining feature of adaptive immune responses to pathogens or other sources of non-self molecules.^[1] This adaptive molecular recognition has been exploited to engineer antibodies for various purposes, including laboratory assays and highly specific protein therapeutics.^[2] Despite their widespread use, experimental identification of antibody-antigen complex structures, or their interacting residues, is still a laborious process. Several computational methods for predicting complex models^[3–6] or interface residues on antibody (paratope) or antigen (epitope) have been developed,^[7–10] but the problem of integrating these methods to archive a robust and coherent solution remains challenging. With the recent breakthroughs in protein structural modeling by Deep Learning,^[11–13] we revisit this important problem and assess the impact of state-of-the-art protein

modeling on antibody-antigen docking and binding site prediction.

AbAdapt is a pipeline that combines antibody and antigen modeling with rigid docking and re-scoring in order to derive antibody-antigen specific features for epitope prediction.^[6] As has been reported by others, the rigid docking and scoring steps are sensitive to the quality of the input models.^[14,15] By default, AbAdapt accepts sequences as input and uses Repertoire Builder,^[16] a high-throughput template-based method, for antibody modeling. However, AbAdapt can also accept structures as input for antibodies, antigens, or both. Here, we assessed the effect of using AlphaFold2 antibody models in the AbAdapt pipeline in a large-scale benchmark using leave-one-out cross validation (LOOCV) and also a large and diverse Holdout set. In addition, the improved AbAdapt-AlphaFold2 (AbAdapt-AF) pipeline was tested using a set of recently determined antibodies that target various epitopes on a common antigen: the SARS-Cov-2 spike receptor binding domain (RBD). We found that the use of AlphaFold2 significantly improved the performance of AbAdapt, both at the level of protein structure and predicted binding sites.

Results and Discussion

Improvement in antibody modeling using AlphaFold2

The CDRs constitute the greatest source of sequence and structural variability in antibodies and also largely overlap with their paratope residues. Here, we systematically evaluated the performance of antibody variable region structural models by Repertoire Builder and AlphaFold2 using the LOOCV and Holdout datasets. The accuracy of antibody modeling improved significantly in both the LOOCV (Figure S1A) and Holdout (Figure S1B) sets. The improvement of AlphaFold2 over Reper-

[a] Z. Xu, Dr. A. Davila, J. Wilamowski, Dr. S. Teraguchi, Prof. D. M. Standley
Department of Genome Informatics
Research Institute for Microbial Diseases
Osaka University,
Suita 565-0871 (Japan)
E-mail: standley@biken.osaka-u.ac.jp

[b] Dr. S. Teraguchi
Faculty of Data Science
Shiga University,
Hikone 522-8522 (Japan)

[**] A previous version of this manuscript has been deposited on a preprint server (<https://doi.org/10.1101/2022.05.21.492907>).

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202200303>

© 2022 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

toire Builder was particularly apparent in the modeling of the most challenging CDR loop, CDR-H3: the average RMSD by AlphaFold2 for the LOOCV set dropped from 4.38 Å to 3.44 Å, a 21.50% improvement over Repertoire Builder (Table S1). Similar results were obtained for the Holdout set (4.44 Å to 3.62 Å, a 18.43% improvement). We note that 58% of the 720 queries were released to the PDB before 30 April 2018, which means that these PDB entries could have been used for training AlphaFold2.

As expected, the improvement of CDR loop modeling by AlphaFold2 resulted in improved paratope modeling: the paratope RMSD dropped from 2.69 Å to 2.08 Å (a 22.73% improvement) in the LOOCV set and from 2.83 Å to 2.12 Å (a 25.26% improvement) in the Holdout set (Figure 1A and Table S1). Using a threshold of paratope RMSD > 4 Å to define low-quality models, the ratio of low-quality models by AlphaFold2 dropped from 14.35% to 7.74% in the LOOCV set and 21.0% to 11.0% in the Holdout set, approximately a 2-fold decrease. These improvements are of interest in antibody-antigen docking because we previously observed that AbAdapt docking per-

formance was sensitive to paratope structural modeling errors.^[6] We also observed that the rank 1 AlphaFold2 models had the lowest CDR-H3 and paratope RMSDs while rank 5 models had the highest CDR-H3 and paratope RMSDs in both the LOOCV and Holdout sets (Table S4). This result justifies to use of rank 1 models in the following analysis.

Improvement of AbAdapt pipeline using AlphaFold2 antibody models

After including the more accurate AlphaFold2 antibody models in the AbAdapt pipeline, we evaluated the influence on each of the main steps: paratope prediction, initial epitope prediction, docking, scoring Piper-Hex clusters and antibody-specific epitope prediction.

Paratope prediction

The Area Under the Curve (AUC) of Receiver Operating Characteristic curve (ROC) and Precision-Recall (PR) curves for paratope prediction for the LOOCV dataset were calculated for AbAdapt and AbAdapt-AF (Figure 1B). We found close median ROC AUCs for both training (0.927) and testing (0.913) by AbAdapt-AF which showed slight improvement over AbAdapt (0.923 for training and 0.911 for testing). The median PR AUC for testing improved from 0.748 (AbAdapt) to 0.753 (AbAdapt-AF). In the Holdout set, the median ROC AUC for paratope prediction by AbAdapt-AF was 0.911, which was close to that of the LOOCV set (0.913) (Figure 1B and 1C). A dramatic improvement of median PR AUC was also observed: from 0.706 (AbAdapt) to 0.743 (AbAdapt-AF) (Figure 1C). Thus, introducing more accurate antibody models from AlphaFold2 clearly improved paratope prediction.

Initial epitope prediction

In the case of initial epitope prediction, AbAdapt-AF exhibited a median test ROC AUC of 0.694 in the LOOCV set and 0.695 in the Holdout set which was identical to AbAdapt, as expected, since both pipelines used the same antigen model from Spanner^[17] for training. Meanwhile, these values were far lower than the median training ROC AUC: 0.863 (Table S2). The difference between training and testing suggests that the initial epitope predictor does not generalize well to unseen data and highlights the inherent limitation in epitope prediction without reference to a specific antibody. This issue is addressed in section "Antibody-specific epitope predictions" by performing epitope prediction with specific antibody-derived features.

Hex and Piper docking

We first analyzed the effect of AlphaFold2 models on the frequency of "true" poses produced by Hex^[18] or Piper^[19] in the

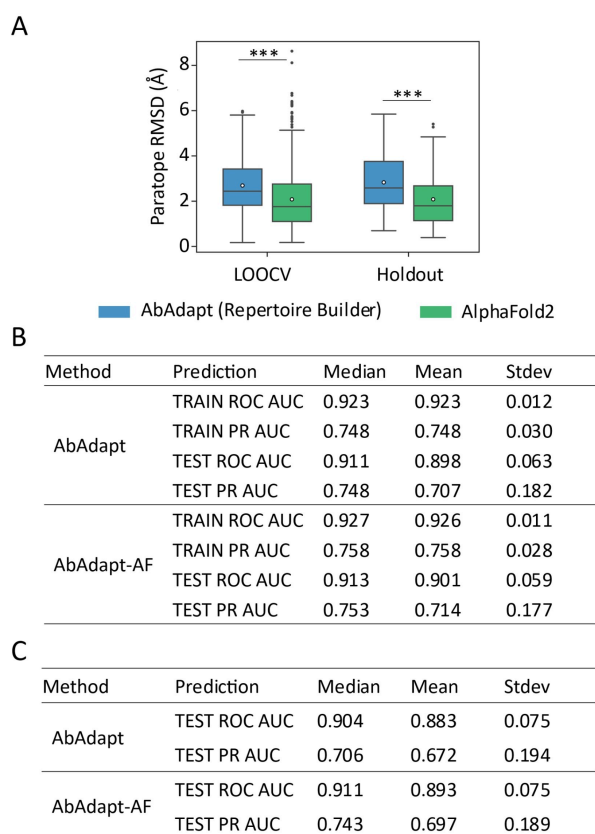


Figure 1. Comparison of the performance of paratope prediction between AbAdapt and AbAdapt-AF. (A) The paratope RMSD of antibody model in LOOCV training set and Holdout set by AbAdapt or AlphaFold2. Wilcoxon matched-pairs signed rank test was performed to compare the corresponding performance between AbAdapt and AbAdapt-AF ($***P \leq 0.001$). The empty circle in each box indicated the average value. Comparison of the paratope prediction performance of antibodies by AbAdapt and AlphaFold2 in LOOCV set (B) and Holdout set (C).

LOOCV set. Here, as in our previous work,^[6] a relatively loose cutoff value of 15 Å for the RMSD of the interface residues (IRMSD), along with epitope and paratope accuracies of 50%, was used to define a “true” pose. To systematically compare the docking performance between AbAdapt and AbAdapt-AF, two more strict IRMSD cutoffs (10 Å and 7 Å) were also utilized. The success ratio was defined as the fraction of queries with at least one “true” pose among all poses (on average 562 poses were generated by AbAdapt-AF). The median true pose ratio of the Hex engine increased modestly from 1.50% to 1.53% (Figure S2C) using an IRMSD cutoff 15 Å, which was not statistically significant (Figure S2A). This is somewhat expected, since Hex was used for local docking, guided by the initial epitope predictions, which themselves were not significantly affected by the use of AlphaFold2 antibody models. In contrast to Hex, the median Piper true pose ratio, which increased from 1.43% to 1.57%, significantly improved upon use of AlphaFold2 antibody models (Figure S2A and S2C). For the Holdout set, the median true pose ratio for both Hex and Piper were not significantly improved by use of AlphaFold2 (Figure S3A and S3C). These results indicate that simply supplying better models does not guarantee improved docking. A similar observation was reported in a recent study using AlphaFold2 models in the ClusPro web server.^[20]

Combined Piper-Hex clustering and scoring

The top Piper and Hex poses from the docking steps were co-clustered and re-scored. Here, we observed an improvement in the AbAdapt-AF pipeline compared with AbAdapt. In the LOOCV set, the median clustered true pose ratio of Hex improved from 1.16% to 1.31% while that for Piper improved 2.32% to 2.78%, resulting in a significantly improved median combined Hex-Piper true pose ratio (2.42% versus 2.83%) for AbAdapt versus AbAdapt-AF, respectively (Figure S2B and S2C). The median rank of the best true pose (with 1 being perfect) dropped significantly from 17 to 13.5, a 20.59% improvement (Figure S2D). We next imposed a stricter (7 Å) IRMSD cutoff; even under this stricter criterion, the median true rank dropped from 54.5 to 43, a 21.10% improvement (Figure S2D). This improvement was also reflected in the ratio of queries with true poses under the stricter 7 Å IRMSD cutoff (success ratio) in the LOOCV set: The AbAdapt-AF success ratio was 55.81%, compared with 46.67% for AbAdapt (Figure 2A).

For the unseen Holdout set, the median clustered true pose ratio for Hex improved but was not statistically significant: 0.56 \pm 8.15% (AbAdapt) versus 1.25 \pm 6.69% (AbAdapt-AF) (Figure S3B and S3C). However, the median clustered true pose ratio of Piper improved significantly: 2.18 \pm 3.97% (AbAdapt) versus 2.66 \pm 3.72% (AbAdapt-AF), resulting in a significantly improved median combined Hex-Piper true pose ratio (2.18 \pm 4.12% versus 2.7 \pm 3.9%) for AbAdapt versus AbAdapt-AF. AbAdapt-AF outperformed AbAdapt in all three IRMSD cutoff values and the success ratio was very close to that of the LOOCV set (Figure 2A and 2B). Taken together, the performance

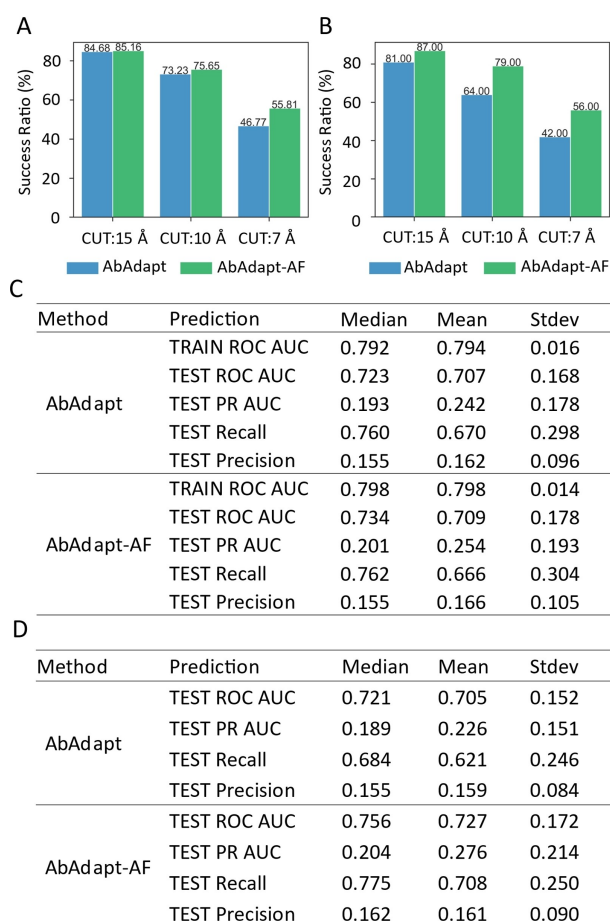


Figure 2. Comparison of the performance of pose clustering and epitope prediction between AbAdapt and AbAdapt-AF. The success ratio of pose clustering in LOOCV set (A) and Holdout set (B) after clustering and combining the pose from Piper and Hex. When setting the true cutoff of interface RMSD (RCUT) 15, 10 and 7 Å, the corresponding success ratio (queries with at least one correctly predicted pose among all poses) was shown above each bar. The final epitope prediction performance in LOOCV set (C) and Holdout set (D) after introducing the post-docking features of specific antibodies.

of scoring and clustering of Hex and Piper poses significantly improved by using AbAdapt-AF.

Antibody-specific epitope predictions

After retraining the epitope predictor with docking features, we observed significant improvement in antibody-specific epitope prediction. In the LOOCV set, the median test ROC AUC increased from 0.694 (initial epitope prediction) to 0.734 (antibody-specific epitope prediction) by AbAdapt-AF, compared with 0.694 (initial epitope prediction) to 0.723 (antibody-specific epitope prediction) by AbAdapt (Figure 2C and Table S2). An even greater improvement was observed in the Holdout set: AbAdapt-AF, the median test ROC AUC improved from 0.695 (initial epitope prediction) to 0.756 (antibody-specific epitope prediction); for AbAdapt, the corresponding values were 0.695 (initial epitope prediction) and 0.721 (antibody-specific epitope

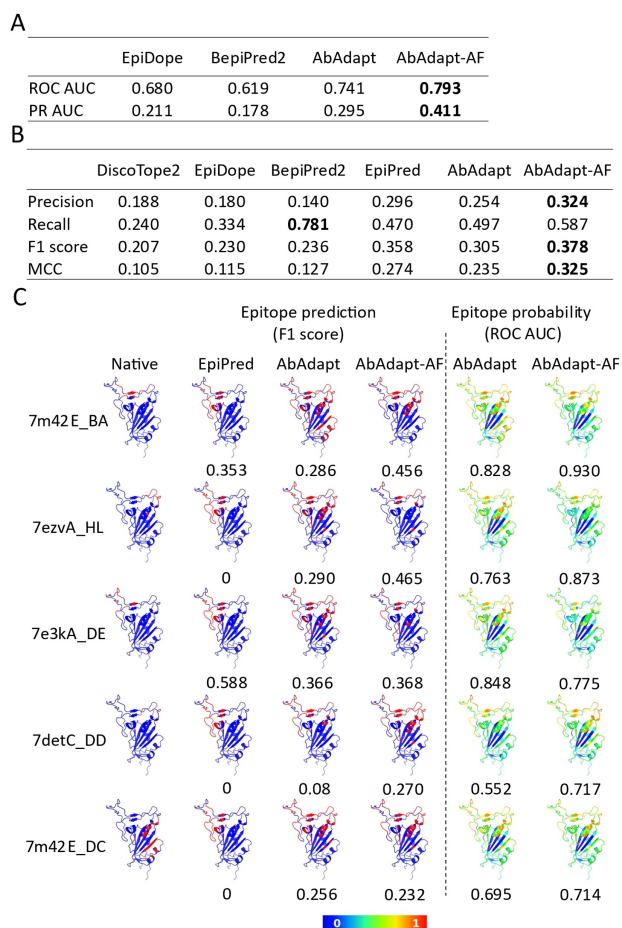


Figure 4. Epitope prediction of 25 anti-RBD antibody complexes. The comparison of epitope prediction performance using probability (A) and a threshold for epitope classification (B). The performance indices are calculated in evaluation metrics and averaged. Bold character indicated the highest value of each item. (C) Epitope map visualization of representative queries. The native epitope (column 1) and predicted epitope by EpiPred (column 2), AbAdapt (column 3), and AbAdapt-AF (column 4) are colored red on the RBD surface. The probability of prediction by AbAdapt and AbAdapt-AF are shown in columns 5 and 6. The value below each prediction is indicated as F1 score (left) and ROC AUC (right).

tested methods when setting a fixed epitope probability threshold (Figure 4B) as described in section “RBD epitope prediction”.

The prediction performances of DiscoTope2, EpiDope, and BepiPred2, which do not take antibody into consideration, were systematically lower than those of the other methods. We noticed that BepiPred2 achieved the highest recall (0.781), but this value was obtained at a precision of only 0.140. Although EpiPred considers antibody information, the results were not very sensitive to the antibodies in this set, as all epitope predictions were in the same region of the RBD (Figure 4C).

Conclusion

In this study, we demonstrated that introducing a more accurate antibody model in the AbAdapt antibody-specific

epitope prediction pipeline had a significantly positive effect at various levels. Nevertheless, there is still much room for improvement. Considering the RBD benchmark, we evaluated six methods for epitope prediction on the SARS-CoV-2 spike RBD, an antigen that can be targeted by one or more antibodies on almost its entire molecular surface.^[22] We found AbAdapt-AF showed higher epitope prediction accuracy than tested methods in this benchmark. Nevertheless, the ROC AUC (0.793) and precision (0.324) values imply that many false-positive epitope predictions are expected. To compare the epitope prediction performance of AbAdapt-AF and other tools fairly and systematically, a broader benchmark containing various antigens will be needed. At present, the use of AlphaFold2 in the context of AbAdapt suggests that this combination represents a robust but incremental way of improving antibody-specific epitope predictions. Newly released antibody modeling tools such as Ablooper^[12] and DeepAb^[13] showed better performance than AlphaFold2 in the Rosetta Antibody Benchmark,^[12] and these methods should also be considered when running AbAdapt with structural models as input.

Experimental Section

Datasets: The preparation of data for LOOCV and Holdout tests was described previously.^[6] In brief, crystal structures of antibody-antigen complexes were gathered from the Protein Data Bank (downloaded May. 24, 2021).^[23] We filtered the complexes using the following criteria: Complete heavy (H) and light (L) variable chain, antigen length equal to or greater than 60 amino acids and resolution less than 4 Å. This filtering process resulted in 1862 Ab–Ag complexes. In the antigen and antibody modeling process by the AbAdapt pipeline, 290 Ab–Ag complexes were discarded due to the poor modeling quality: Root Mean Square Deviation (RMSD) of either the paratope or epitope were higher than 6 Å. Also, antigens that were themselves antibodies were removed (5 queries). Antibody Complementarity-determining regions (CDR) sequences were annotated with the AHo numbering scheme^[24] by the ANARCI program.^[25] Next, the sequence redundancy was removed as follows: A pseudo-sequence of each antibody composed of heavy and light variable domain was constructed and clustered using CD-HIT^[26] at an 85% identity threshold. Any sequences that resulted in failure by third-party software in subsequent steps were also removed (3 queries). This filtering resulted in 720 non-redundant antibody-antigen model pairs. For training, LOOCV was performed using 620 randomly chosen antibody-antigen queries. The remaining 100 queries functioned as an independent Holdout set for testing. We also collected anti-SARS-Cov-2 RBD antibodies on Sept. 5, 2021 from the PDB.^[23] Antibody pseudo-sequences were constructed as described above and clustered with the LOOCV training dataset using CD-HIT at an 85% sequence identity threshold. After superimposing all common epitopes for all pairs of antibodies, the C-alpha root-mean-square deviation (RMSD) was used to cluster the antibodies by single-linkage hierarchical clustering with a 10.0 Å cutoff. We previously used average-linkage hierarchical clustering to classify antibodies that target overlapping epitopes based on structural alignments.^[27] Here, the antigens in the anti-SARS-Cov-2 RBD antibody dataset were aligned well using sequence information, and appeared to form discrete clusters. Thus, we utilized single-linkage hierarchical clustering in this case. Within each cluster, the PDB entry with the most recent release date was chosen as the representative. This

procedure resulted in the identification of 25 novel anti-RBD antibody-antigen complexes.

Epitope and Paratope definition: The epitope (paratope) residues were defined as any residue with at least one heavy atom within 5 Å of the antigen (antibody), as measured by Prody 2.0.^[28]

Antibody modeling by AlphaFold2: The antibodies from the LOOCV and Holdout sets were modeled independently from antigens using the full AlphaFold2 pipeline with default parameters.^[11] We concatenated the sequences of H and L chains in the Fab region via a poly-glycine linker (32G). After cleaving the linker, the rank-one model was used in all subsequent calculations. Antibody CDRs were annotated as described in section "Datasets". To evaluate the modeling accuracy of the CDRs, we first superimposed the +/−4 flanking amino acid residues of each CDR in predicted models onto native antibody structures. Next, we calculated the RMSD of the CDRs by Prody 2.0.^[28] The paratope RMSD was computed similarly after superimposing the paratope of the model onto the native structure.

AbAdapt and AbAdapt-AF pipelines: AbAdapt was run using default parameters, as described previously.^[6] In brief, antibodies were modeled using Repertoire Builder^[16] and antigens were modeled using Spanner.^[17] Here, any templates with bound antibodies overlapping with the true epitope were excluded. Two docking engines (Hex^[18] and Piper^[19]) were used to sample rigid docking degrees of freedom. Machine learning models were used to predict initial epitope and paratopes, score Hex poses, score Piper poses, score clusters of Hex and Piper poses, and predict antibody-specific epitope residues. The description of above machine learning models are summarized in Supplementary Table. 5. For LOOCV calculations, each time an antibody-antigen pair was used as a query, each ML model was re-trained. For the Holdout datasets, training was performed on the entire LOOCV dataset. For the AbAdapt-AF pipeline, the only procedural difference was that AlphaFold2 was used for antibody prediction instead of the default Repertoire Builder method. Naturally, all ML models were re-trained specifically for this use case.

RBD-antibody complex prediction: A model of SARS-Cov-2 RBD (329-532 aa) was built using Spanner^[17] and used as the antigen in all subsequent docking steps. The modeling of 25 anti-RBD antibodies was performed using AlphaFold2 on the ColabFold platform (v1.2) using the AlphaFold2-ptm model type with templates and Amber relaxation.^[29] The rank-one antibody model was used in the downstream workflow of docking and complex modeling. We used three third-party docking methods to benchmark AbAdapt and AbAdapt-AF performance on RBD-antibody docking. The global protein-protein docking program ZDOCK (<https://zdock.umassmed.edu/>, v3.0.2)^[3] was utilized with default parameters for antibody-RBD binding prediction. The top 10 models were retained for analysis. We used the HawkDock server with default parameters (<http://cadd.zju.edu.cn/hawkdock>) for antibody-RBD binding prediction. HawkDock integrates the ATTRACT docking algorithm, the HawkRank scoring function, and the Molecular Mechanics/Generalized Born Surface Area free energy decomposition for protein-protein interface analysis.^[4] The top 100 models were retained for analysis. HDock (<http://hdock.phys.hust.edu.cn/>)^[5] was utilized with the template-free. The top 100 models were retained for analysis. To evaluate the accuracy of all RBD-antibody poses, DockQ (<https://github.com/bjornwallner/DockQ>)^[30] was utilized with CAPRI criteria.^[21]

RBD epitope prediction: We used the following third-party methods for epitope prediction: DiscoTope2, which predicts discontinuous B-cell epitopes (<https://services.healthtech.dtu.dk/service.php?DiscoTope-2.0>),^[7] BepiPred2, which uses a random

forest regression algorithm (<https://services.healthtech.dtu.dk/service.php?BepiPred-2.0>),^[8] EpiDope (<https://github.com/flomock/EpiDope>), which uses a deep neural network to predict linear B-cell epitopes based on antigen amino acid sequence.^[9] For binary labeling of the predicted epitopes, the default thresholds of 0.5 for BepiPred2^[8] and 0.818 for EpiDope^[9] were used. To predict the structural epitopes specific to a given antibody, EpiPred (<http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/epipred>)^[10] was used with 25 anti-RBD antibody models from AlphaFold2 and an RBD model from Spanner. The rank-one epitope prediction from EpiPred was chosen for all subsequent analysis. For AbAdapt-AF and AbAdapt, a threshold of 0.5 for final epitope prediction was used.^[6]

Statistical analysis and Visualization: Wilcoxon matched-pairs signed rank test and descriptive analysis were performed using GraphPad Prism 8 (GraphPad Software, San Diego, CA). For graphing, the Python package Seaborn (v 0.11.0) and Matplotlib (v 3.3.1) were used. We visualized the RBD-antibody complexes and RBD structure with its corresponding epitope map using PyMOL (The PyMOL Molecular Graphics System, v 2.3.3 Schrödinger, LLC.)

Performance measures: To evaluate the performance of epitope and paratope prediction by a threshold-dependent measurement, four performance indices (Precision [Eq. 1], Recall (Sensitivity) [Eq. 2], F1 score [Eq. 3], and Matthews correlation coefficient (MCC) [Eq. 4]) were calculated by the following equations.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \quad (3)$$

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4)$$

Where TP refers to true positive; FN refers to false negative; TN refers to true negative; FP refers to false positive. For Threshold-independent measurement, the Area Under the Curve (AUC) of Receiver Operating Characteristic curve (ROC) and Precision-Recall (PR) curves for paratope and epitope prediction were also calculated, based on the predicted probabilities. The performance indices of epitope and paratope prediction were calculated using the Python package Scikit-learn (v 0.23.2).^[31]

Acknowledgements

We would like to thank all members of the Systems Immunology Lab for very helpful discussion. The computation was performed using Research Center for Computational Science, Okazaki, Japan (Project: 21-IMS-C116 and 22-IMS-C136). This work was supported by Japan Agency for Medical Research and Development (AMED) Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research) under Grant Numbers 22ama121025j0001 and by a Grand-in-Aid for Scientific Research by the Japan Society for the Promotion of Science (JSPS) under Grant Number JP20K06610.

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords: AlphaFold · antibody-antigen docking · antibody-specific epitope prediction · receptor binding domain · SARS-CoV-2

- [1] K. Murphy, C. Weaver, *Janeway's Immunobiology*, 12th ed., New York, Garland Science, 2017.
- [2] J. Maynard, G. Georgiou, *Annu. Rev. Biomed. Eng.* **2000**, *2*, 339–376.
- [3] B. G. Pierce, K. Wiehe, H. Hwang, B. H. Kim, T. Vreven, Z. Weng, *Bioinformatics* **2014**, *30*, 1771–1773.
- [4] G. Weng, E. Wang, Z. Wang, H. Liu, F. Zhu, D. Li, T. Hou, *Nucleic Acids Res.* **2019**, *47*, W322–W330.
- [5] Y. Yan, H. Tao, J. He, S. Y. Huang, *Nat. Protoc.* **2020**, *15*, 1829–1852.
- [6] A. Davila, Z. Xu, S. Li, J. Rozewicki, J. Wilamowski, S. Kotelnikov, D. Kozakov, S. Teraguchi, D. M. Standley, *Bioinformatics* **2022**, vbac015.
- [7] J. V. Kringelum, C. Lundegaard, O. Lund, M. Nielsen, *PLoS Comput. Biol.* **2012**, *8*, e1002829.
- [8] M. C. Jespersen, B. Peters, M. Nielsen, P. Marcatili, *Nucleic Acids Res.* **2017**, *45*, W24–W29.
- [9] M. Collatz, F. Mock, E. Barth, M. Holzer, K. Sachse, M. Marz, *Bioinformatics* **2021**, *37*, 448–455.
- [10] J. Dunbar, K. Krawczyk, J. Leem, C. Marks, J. Nowak, C. Regep, G. Georges, S. Kelm, B. Popovic, C. M. Deane, *Nucleic Acids Res.* **2016**, *44*, W474–478.
- [11] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohli, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
- [12] B. Abanades, G. Georges, A. Bujotzek, C. M. Deane, *Bioinformatics* **2022**, btac016.
- [13] J. A. Ruffolo, J. Sulam, J. J. Gray, *Patterns* **2022**, *3*, 100406.
- [14] I. Anishchenko, P. J. Kundrotas, I. A. Vakser, *Proteins* **2017**, *85*, 470–478.
- [15] P. J. Kundrotas, I. Anishchenko, V. D. Badal, M. Das, T. Dauzhenka, I. A. Vakser, *Proteins* **2018**, *86 Suppl. 1*, 302–310.
- [16] D. Schritt, S. Li, J. Rozewicki, K. Katoh, K. Yamashita, W. Volkmuth, G. Cavet, D. M. Standley, *Mol. Syst. Des. Eng.* **2019**, *4*, 761–768.
- [17] M. Lis, T. Kim, J. J. Sarmiento, D. Kuroda, H. V. Dinh, A. R. Kinjo, K. Amada, S. Devadas, H. Nakamura, D. M. Standley, *Immunome Res.* **2011**, *7*, 1–8.
- [18] G. Macindoe, L. Mavridis, V. Venkatraman, M. D. Devignes, D. W. Ritchie, *Nucleic Acids Res.* **2010**, *38*, W445–449.
- [19] D. Kozakov, R. Brenke, S. R. Comeau, S. Vajda, *Proteins* **2006**, *65*, 392–406.
- [20] U. Ghani, I. Desta, A. Jindal, O. Khan, G. Jones, S. Kotelnikov, D. Padhorny, S. Vajda, D. Kozakov, *bioRxiv* **2021**, <https://doi.org/10.1101/2021.09.07.459290>.
- [21] M. F. Lensink, S. J. Wodak, *Proteins* **2013**, *81*, 2082–2095.
- [22] M. Yuan, H. Liu, N. C. Wu, I. A. Wilson, *Biochem. Biophys. Res. Commun.* **2021**, *538*, 192–203.
- [23] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acids Res.* **2000**, *28*, 235–242.
- [24] A. Honegger, A. Plückthun, *J. Mol. Biol.* **2001**, *309*, 657–670.
- [25] J. Dunbar, C. M. Deane, *Bioinformatics* **2016**, *32*, 298–300.
- [26] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, *Bioinformatics* **2010**, *26*, 680–682.
- [27] Z. Xu, S. Li, J. Rozewicki, K. Yamashita, S. Teraguchi, T. Inoue, R. Shinnakasu, S. Leach, T. Kurosaki, D. M. Standley, *Mol. Syst. Des. Eng.* **2019**, *4*, 769–778.
- [28] S. Zhang, J. M. Krieger, Y. Zhang, C. Kaya, B. Kaynak, K. Mikulska-Ruminska, P. Doruker, H. Li, I. Bahar, *Bioinformatics* **2021**, *37*, 3657–3659.
- [29] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *Nat. Methods* **2022**, *19*, 679–682.
- [30] S. Basu, B. Wallner, *PLoS One* **2016**, *11*, e0161879.
- [31] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, *J. Machine Learning Res.* **2011**, *12*, 2825–2830.

Manuscript received: May 30, 2022

Revised manuscript received: July 27, 2022

Accepted manuscript online: July 27, 2022

Version of record online: August 11, 2022