A DATA-DRIVEN APPROACH TO ANTIGEN-ANTIBODY COMPLEX STRUCTURE MODELING USING LABELED VHH ANTIBODIES

Takashi Nagata[1,2], **Hiroyuki Yamazaki**[1,3], **Ryota Maeda**[1,3], **Hirofumi Tsuruta**[1], **Ryotaro Tamura**[1], **Akihiro Imura**[1,3]
[1]COGNANO Inc., [2]Kyoto University, [3]Biorhodes, Inc.
{nagata, akihiroimura}@cognano.co.jp

ABSTRACT

Tumor necrosis factor alpha (TNFα) has been extensively studied using X-ray crystallography, cryo-electron microscopy, and AI-based modeling. Antibodies have also been used as structural probes to investigate TNFα. To enrich antigen-antibody structural data, we immunized alpacas with human TNFα and developed a VHH (single domain antibody) library. VHH antibodies consist of a single chain, which facilitates large-scale data collection. However, accurately modeling antigen-antibody complexes in three-dimensional (3D) remains a challenge. We selected TNFα-binding VHH clones and predicted their 3D structures using ColabFold. By adding our VHH library, we generated multiple sequence alignments (MSAs) with much higher sequence coverage than those using only public databases. The predicted local distance difference test (pLDDT) of structural models for multiple outputs tended to be better, suggesting the usefulness of our antibody library. In this study, we released a labeled antibody dataset for TNFα and attempted structural modeling, which we hope will facilitate advances in antibody engineering and therapeutics.

1 INTRODUCTION

Structural analysis of TNFα has advanced with X-ray crystallography, cryo-electron microscopy, and AI-driven dynamic modeling. Antibodies have also been used as structural probes to study. To improve antigen-antibody structural data, we immunized alpacas with human TNFα and developed a VHH (singledomain antibody) library. Unlike conventional antibodies with heavy and light chains, VHH antibodies are single-chain, enabling easier sequencing and large-scale data collection. Using phage display, we obtained labeled binding data and constructed the worlds largest labeled VHH antibody dataset. Previously, we published similar datasets for IL-6 (Tsuruta et al., 2023) and the SARS-CoV-2 spike protein (Tsuruta et al., 2024).

2 RESULTS

We have expanded and released additional labeled VHH antibody datasets. However, predicting antibody three-dimensional (3D) structures remains a major challenge, particularly in accurately reflecting interactions with antigens. To address this issue, we have been exploring the use of ColabFold (Mirdita et al., 2022) to complement antigen-antibody complex structures. This dataset will be made available at `https://huggingface.co/datasets/COGNANO/AVIDa-hTNFa`.

3 METHODS

We selected clones with strong binding affinity to TNFα (strong binders) from our established antibody library and predicted their 3D structures using LocalColabFold. First, we created clusters of the VHH sequences, including strong binders with a certain homology, and performed multiple sequence alignment (MSA). Using the resulting MSA, we generated structural models and found

that, compared to those based solely on public sequence databases, our models exhibited higher pLDDT between predicted structures. These results suggest the usefulness of our antibody library for predicting antibody structures. We are currently exploring methods to apply this approach to modeling TNFα-antibody complexes.

4 DISCUSSION

In this study, we newly released a labeled antibody dataset specific to human TNFα (577 binder clones in 85 clusters from 2 alpacas) and discussed structural learning strategies incorporating both primary sequence data and 3D structural information. Future research directions include:

- Developing more accurate antigen-antibody complex structure prediction methods
- Utilizing machine learning to analyze antibody-antigen interactions
- Expanding the antibody library to construct even larger datasets

These advancements are expected to greatly improve our understanding of antigen-antibody interactions, facilitating the design and optimization of novel antibody therapeutics.

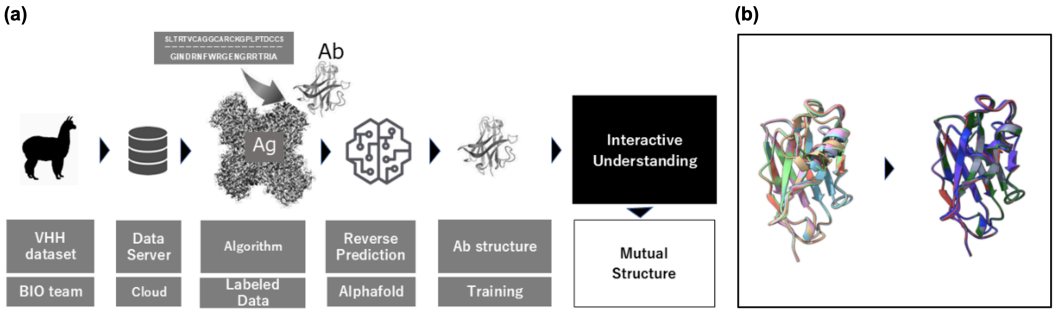

Figure 1: (a) Workflow of Inversely Understanding Ag/Ab Complex (b) Ag/Ab dataset aligns the VHH-Ab shape.

REFERENCES

Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: making protein folding accessible to all. *Nature methods*, 19(6):679–682, 2022.

Hirofumi Tsuruta, Hiroyuki Yamazaki, Ryota Maeda, Ryotaro Tamura, Jennifer N. Wei, Zelda Mariet, Poomarin Phloyphisut, Hidetoshi Shimokawa, Joseph R. Ledsam, Lucy Colwell, and Akihiro Imura. AVIDa-hIL6: A large-scale VHH dataset produced from an immunized alpaca for predicting antigen-antibody interactions. In *Advances in Neural Information Processing Systems 36*, 2023.

Hirofumi Tsuruta, Hiroyuki Yamazaki, Ryota Maeda, Ryotaro Tamura, and Akihiro Imura. A SARS-CoV-2 interaction dataset and VHH sequence corpus for antibody language models. In *Advances in Neural Information Processing Systems 37*, 2024.