論　文　内　容　の　要　旨

Synopsis of Thesis

氏　名 Name	徐　子暢
論文題名 Title	Improved Antibody-Specific Epitope Prediction Using AlphaFold and AbAdapt (AlphaFoldおよびAbAdaptを使用した抗体特異的エピトープ予測の改善)

論文内容の要旨

〔目　的(Purpose)〕

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. With the recent breakthroughs in protein structural modeling by Deep Learning, we revisit this important problem and assess the impact of state-of-the-art protein modeling on antibody-antigen docking and binding site prediction.

〔方法ならびに成績(Methods/Results)〕

We selected 720 non-redundant antibody-antigen pairs. For training of the machine learning model, leave-one-out cross-validation (LOOCV) was performed using 620 randomly chosen queries. The remaining 100 queries functioned as an independent Holdout set for testing. We also identified 25 novel anti-SARS-Cov-2 RBD antibody-antigen complexes for further analysis. The antibodies from the LOOCV and Holdout sets were modeled independently from antigens using the full AlphaFold2 (AF) pipeline and the rank-one model was used in all subsequent calculations. In the AbAdapt-AF pipeline, antibodies were modeled using AF and antigens were modeled using Spanner. Two docking engines (Hex and Piper) were used to sample rigid docking poses. 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.

We first systematically evaluated the performance of antibody variable region structural models by AF. The improvement of AF over Repertoire Builder was apparent in the six CDR loops and this improvement resulted in improved paratope modeling: the paratope RMSD dropped from 2.69 Å to 2.08 Å in the LOOCV set. After including the more accurate AF antibody models in the AbAdapt pipeline, we observed improvement in paratope prediction: The median PR AUC improved from 0.748 (AbAdapt) to 0.753 (AbAdapt-AF) in the LOOCV set. Next, we analyzed the effect of AF models on the sampling, clustering, and scoring of "true" poses produced by Hex or Piper. In the LOOCV set, the median clustered true pose ratio of Hex improved modestly 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. 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. An even greater improvement was observed in the Holdout set: the median test ROC AUC improved from 0.695 (initial epitope prediction) to 0.756 (antibody-specific epitope prediction) by AbAdapt-AF. We next assessed the performance of AbAdapt-AF in RBD benchmark. We found AbAdapt-AF outperformed other three docking methods and demonstrated Abadapt-AF achieves higher epitope prediction accuracy than other tested tools.

〔総　括(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. We observed improvement in docking, paratope prediction, and prediction of antibody-specific epitopes. We anticipate that AbAdapt-AF will facilitate prediction of antigen-antibody interactions in a wide range of applications.

[Format-H07]

論文審査の結果の要旨及び担当者

（申請者氏名）徐　子暢

論文審査担当者	（職）	氏　名	
	主　査	大阪大学教授　Daron M. Standley	Daron Standley
	副　査	大阪大学教授　山崎 晶	山崎 晶 署名
	副　査	大阪大学教授　茂呂　和世	茂呂和世 名

論文審査の結果の要旨

In this research project, Zichang's goal was to construct a bioinformatics tool, AbAdapt, that can predict the epitope of an antibody-antigen pair. He utilized docking in order to sample antibody-antigen binding modes; then, he a machine learning model for each stage of the docking in order to enhance the scoring and epitope prediction. The results were comparable with other state-of-the-art methods and showed that docking improved the epitope prediction beyond that that of antigen-based predictions (Davila et al Bioinformatics Advances 2022).

Zichang then took advantage of the recent advancement in protein structural modeling using AlphaFold to improve the AbAdapt antibody-antigen modeling pipeline (Xu et al ChemBioChem 2022). This research is timely and significant. For example, we can analyze BCR repertoire sequence data (e.g. comparing COVID-19 and healthy repertoires), identify disease-specific antibodies. In such cases, it would be helpful to have this tool to analyze antibody-antigen binding. Indeed, Abadapt was used to predict binding modes for infection-enhancing antibodies identified by Arase and co-workers (Liu et al. Cell 2021).

Based on the above, I believe that Zichang Xu's research Worthy of a dissertation.