[Format-P02]

論　文　内　容　の　要　旨
Synopsis of Thesis

氏　名 Name	DAVILA CRESPO Ana Cecilia
論文題名 Title	AbAdapt: An adaptive approach to predicting antibody-antigen complex structures from sequence (AbAdapt: 抗体-抗原複合体の構造を配列から予測するための適応的方法)

論文内容の要旨

〔目　的(Purpose)〕

Antibodies are a highly diverse class of immune receptors whose binding residues (paratopes) are selected to specifically recognize a given antigen at a given surface patch (epitope). Antibody structures can be predicted from sequences and the binding modes of antibody-antigen complexes can be sampled by existing protein docking methods. However, the scoring of antibody-antigen docked poses starting from unbound homology models has not been systematically optimized for a large and diverse set of input sequences.

〔方法ならびに成績(Methods/Results)〕

To address this need, we have developed AbAdapt, a web server that accepts antibody and antigen sequences, models their 3D structures, predicts epitope and paratope, and then docks the modeled structures using two established docking engines (Piper and Hex). Each of the key steps has been optimized by developing and training new machine-learning models. The sequences from a diverse set of 622 antibody-antigen pairs with known structure were used as inputs for leave-one-out cross validation. The final set of cluster representatives included at least one "Adequate" pose for 550/622 (88.4%) of the queries. The median (IQR) ranks of these "Adequate" poses were 22 (5 to 77). Similar results were obtained on a holdout set of 100 unrelated antibody-antigen pairs.

〔総　括(Conclusion)〕

We have attempted to build a working pipeline out of available tools, assess what worked and what did not, and suggest directions for future improvement. Our efforts were generally encouraging, as indicated by the query coverage and improvement in True ranks across a large and diverse test set, along with the general agreement between LOOCV and holdout benchmarks. Although questions remain about the best balance between Piper and Hex poses, both the only-Piper and Piper-Hex AbAdapt pipelines produced better results than could be obtained by Piper alone. Also, the improvement in epitope prediction performance upon addition of antibody-specific features suggests a way of addressing this long-standing and important problem.

論文審査の結果の要旨及び担当者

（申請者氏名）DAVILA CRESPO ANA CECILIA

論文審査担当者	（職）	氏　名
主　査	大阪大学教授	
副　査	大阪大学~~教授~~准教授	
副　査	大阪大学教授	

論文審査の結果の要旨

The goal of this project was to develop a tool, AbAdapt, that can predict the interaction between an antibody and an antigen. The input can consist of amino acid sequences or structures for either antibody or antigen. The output consists of top-scoring docked models and a list of residue-based scores for the antibody and antigen that represent the probabilities of the residue appearing at the interface (paratope for the antibody and epitope for the antigen). The approach that was used to construct AbAdapt was to break the big problem into many smaller problems. These sub-problems consisted of: antibody and antigen structural modeling (if necessary), initial epitope and paratope prediction, docking using two docking 'engines'(Hex and Piper), scoring the two docking results and selecting the top poses, clustering the top-scoring poses from both Hex and Piper together, scoring the clusters, and, finally re-predicting the epitope probabilities using the docking scores. Each step required construction of a machine-learning model to solve each sub-problem.

The most difficult sub-problem was the last step: incorporating docking information to improve epitope prediction accuracy. This was the area where Ms. Davila's hard work paid off. She was able to improve the "ROC AUC" (a measure of prediction accuracy) from 0.694 to 0.730 in a set of 100 antibody-antigen pairs that had not been "seen" by the machine learning model.

I am not aware of any similar tool, where antibody and antigen sequences can be input and an antibody-specific epitope is predicted. I would thus characterize this as a "big" problem. AbAdapt does not represent a perfect solution. There is still much room for improvement in the prediction accuracy. However, it represents a reasonable approach that, hopefully, will be challenged by others. In this way, the problem will eventually be solved, I believe.

The construction of AbAdapt required several years of work. I believe that reaching the current level of performance shows a high degree of tenacity and diligence by Ms. Davila. The skills acquired during this PhD project will be useful in other areas of science. Ms.Davila deserves her PhD degree.