

グラフニューラルネットワークを用いた B 細胞エピトープの予測		
黒田裕 研究室	学籍番号:21261037	氏名 橘明日香

[背景]

脊椎動物の免疫系では、B 細胞と T 細胞の 2 種類のリンパ球がそれぞれ別の経路で抗原に対する応答を行う。このうち、B 細胞受容体および抗体が認識する抗原分子の特異的領域を B 細胞エピトープという。B 細胞エピトープは、アミノ酸配列が連続している残基からなる線形エピトープと、ポリペプチド鎖の折り畳みによって空間的に近くなつた残基からなる構造依存性エピトープにわけられる。これまでの研究で BCE の 90%以上は構造依存性エピトープだと示されている。

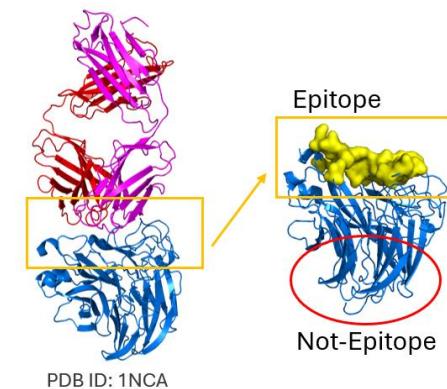
BCE の予測はバイオテクノロジーと臨床応用において重要であるが、実験的同定には多くの時間とコストがかかる。そこで、計算的手法を用いて同定作業を軽減する研究が行われている。本研究では、グラフニューラルネットワークを用いて、構造情報から B 細胞エピトープの特徴抽出と予測を行った。

[材料、手法]

Protein Data Bank より、1 つの完全な抗体と 1 つの抗体以外の構造を含む免疫複合体構造を、262 個取得した。抗体以外の構造を抗原とみなし、pdb2sql を用いて、抗体との接触残基数の平均が 18 残基程度になるように、抗体との距離が 5 Å 以内にある抗原の残基を抽出し、これらをエピトープと定義した。また、エピトープでないエリアの構造は、平均残基数が 18 残基程度になるような空間的に近い残基の集まりを抽出した。エピトープ構造をポジティブデータ、エピトープでない構造をネガティブデータとして、両者をデータセットとした。学習には、機械学習ライブラリである TorchProtein および、グラフニューラルネットワークモデルである GearNet-Edge を用いた。TorchProtein は PyTorch を基盤とした TorchDrug を、タンパク質科学に特化させたプラットフォームで、タンパク質特性予測や、結合親和性予測などのタスクを提供する。GearNet-Edge は TorchDrug で提供されるモデルの一つで、タンパク質の幾何学的構造の事前学習を採用し、タンパク質に特有な特徴量を学習している。また、エッジの関係グラフを用いることでエッジ間の関係も考慮したノードの特徴量の更新を行い、タンパク質の構造表現を効率的に学習することができる。

[結果、展望]

入力データを 5 分割し、クロスバリデーションを行い、モデルを評価した。各フォールドの AUROC の平均は 0.9105 となった。今後は、ネガティブデータ増強およびエピトープの特徴抽出を行う。また、分割した抗原の表面残基に対する本モデルの予測と実験的手法との比較を行い、本モデルの信頼性を検証する。



fold No	k-Fold Cross Validation			AUROC
	Training	Validation	Test	
1	314	36	88	0.9040
2	314	36	88	0.9127
3	316	36	86	0.9180
4	315	35	88	0.9252
5	314	36	88	0.9071
Average				0.9105