CALIFORNIA STATE SCIENCE FAIR
2017 PROJECT SUMMARY

Name(s)

Arjun S. Subramaniam

Project Number

S1524

Project Title

CadML: A New Computational Approach to Optimizing Antibody Affinity for Design of Antibody Therapeutics

Abstract

Objectives/Goals

Antibody therapeutics is a growing field of drugs, involving antibodies that specifically target a pathogen inside the body. Designing such high-affinity antibodies mimics our immune system's response during infection: mutating and selecting antibody variants that bind better to the antigen (target). Yet instead of doing this in vitro, which is inherently random and time-consuming, we could use computational methods to identify mutations on antibodies that increase their affinity to their target. However, predicting the effects of mutations on how well antibodies bind is still an unsolved problem. I propose CadML, a novel, holistic machine learning- based approach.

Methods/Materials

Given an antibody-antigen mutation, CadML incorporates information about the mutation from all levels of protein structure, from the amino acid sequence to structural elements. In addition, I used convolutional neural networks (CNN), a deep learning model, to extract highly informative features from the surrounding environment of the mutation site. All of this information was combined into a final machine learning model, which predicts the change in binding energy caused by the mutation (measuring binding affinity). CadML was written in Python2.7, with protein modeling tools and Python libraries.

Results

I evaluated CadML on the experimentally-verified AB-Bind dataset, consisting of 528 mutations and associated binding affinity changes (favorable and harmful). CadML achieved a Pearson's Correlation of 0.64, identifying favorable, affinity-improving mutations with high accuracy. My method significantly outperformed state-of-the-art prediction methods, including machine learning and simulation-based methods ($P < 0.01$). CadML achieved a correlation of 0.70 on further evaluation of another dataset.

Conclusions/Discussion

CadML combined deep learning with information from all levels of protein structure to find affinity-improving antibody mutations with high accuracy. I extended my analysis by mapping Rituximab, a current antibody therapeutic for leukemia, and finding mutations that could significantly improve its affinity to cancer cells. In addition, I scanned the surface of HIV to provide insight into therapeutic strategies for AIDS. In the future, I would look to integrate CadML into the drug development pipeline, helping bring high-affinity antibody therapies from the benchside to the bedside.

Summary Statement

I built a machine learning model to design antibody therapies that can treat diseases like cancer and HIV by binding effectively to targets inside the body.

Help Received

I worked with Dr. Thomas MacCarthy in the Department of Mathematics and Statistics at Stony Brook University over the summer (Simons Summer Research Program). I proposed the idea and built the algorithm largely independently, with some guidance from Dr. MacCarthy.