Midterm II

Andrew Rosen

April 16, 2015

I decided to focus on the paper looking at anti-influenza aptamers [3].

Description

Musafia *et al* examined a new way of fighting the influenza virus. Influenza is a particularly common viral disease that infects millions of people each year. It's particularly dangerous to adults over 65 [1]. The primary mechanism of spreading by coughing or sneezing [4]. Expelled viral particles then bind to the victim's respiratory epithelium, a lining in the respiratory tract.

Musafia *et al* study the use of an aptamer to block the virus from attaching to the respiratory epithelium, inhibiting infection. Aptamers are a type of molecule called oligonucleotides. These molecules are single-stranded DNA or RNA of approximately 15-45 nucleotides in length. Aptamers are highly discriminating and can be created to bind to very specific targets, while not binding to closely related "targets."[1]

The process of generating aptamers is called "Systematic Evolution of Ligands by EXonential enrichment," or SELEX. This process generates a library of 10^{13} to 10^{15} random oligonucleotides, each 20-100 nucleotides long. The library is searched for sequences that can interact with the target; those few are chosen for the enrichment process.

This paper examines a DNA aptamer called BV02 and compares it to generated aptamers of similar length and other properties. BV02 had been previously shown to be effective at preventing influenza infection by blocking the viral hemagglutinin [2]. This aptamer was chosen by Musafia *et al* as the starting point and base comparison to the other aptamers created by the authors.

The method description involves a lengthy description of preparation for biochemical experiments. I won't pretend to understand the specifics, but this process describes how Musafia *et al* created their solutions and how a scientist[2] could recreate it. The experiment to establish the efficacy of aptamers was conducted on mice. The aptamers were applied to the mice premixed with the influenza virus. I don't have the experience to judge if this is an accurate representation of infection "in the wild."

This process was performed for a total of 96 aptamers. Once the relative binding affinities [3] were established, these aptamers were unceremoniously slapped into a database. Every fourth aptamer was noted as a member of the test set, and the remaining aptamers were used to create the training set.

[1] As far as I understand it.

[2] Unfortunate grad student.

[3] compared to BV02's ability to bind to viral hemagglutinin.

Various features were isolated which impacted the aptamers ability to bind to viral hemagglutinin. The effectiveness was impacted primarily by the aptamer's size, secondary structure, and the presence of repeated C nucleotides, among others. QSAR was used to create a mathematical model to predict the effectiveness of the apatamer's binding.

Effectiveness

Musafia *et al*'s results seem extremely promising. They found 14 aptamers that had 15 times the binding affinity of BV02 [3]. They were able to find the 6 most relevant descriptors which defined the effectiveness of the binding. Their algorithm was able to successfully predict the binding affinity of the aptamers.

Relevence To Term Project

Our project focuses on generating potential siRNA candidates for attacking a family of viruses. This paper focuses on using a machine learning algorithm to predict the effectiveness of an aptamer binding to influenza. The relevance to our project is limited; the workflow for creating siRNA requires no machine learning and is fairly straightforward.

Improvements

The first thing that raised my eyebrows was that SELEX generated 10^{13} to 10^{15} random sequences. That seems like an excessively large search space of random material, but this may be standard operating procedure for this field (again, I'm not a biologist). Since the user will be selecting which of these random sequences should be enriched, there's presumably some kind of scoring mechanism that can be used as a fitness function.

Perhaps it would be possible to craft a genetic algorithm that does the same thing, but with a much smaller population than 10^{13} each generation. Select the best 100 sequences and apply crossover and mutation we stop seeing improvement for a number of generations.

References

[1] Centers for Disease Control, Prevention (CDC, et al. Estimates of deaths associated with seasonal influenza—united states, 1976-2007. *MMWR. Morbidity and mortality weekly report*, 59(33):1057, 2010.

[2] Sung Ho Jeon, Basak Kayhan, Tamar Ben-Yedidia, and Ruth Arnon. A dna aptamer prevents influenza infection by blocking the receptor binding region of the viral hemagglutinin. *Journal of Biological Chemistry*, 279(46):48410–48419, 2004.

[3] Boaz Musafia, Rony Oren-Banaroya, and Silvia Noiman. Designing anti-influenza aptamers: Novel quantitative structure activity relationship approach gives insights into aptamer–virus interaction. *PloS one*, 9(5):e97696, 2014.

[4] I Stephenson and M Zambon. The epidemiology of influenza. *Occupational medicine*, 52(5):241–247, 2002.

Additional comments

Under the "Training and Test Data Sets for QSAR Study" section, fourth is misspelled as "forth."