AptaBLE: An Enhanced Deep Learning Platform for Aptamer Protein Interaction Prediction and Design

Sawan Patel[1*] **Keith Fraser**[2,3,4*] **Zhangzhi Peng**[5*]

Owen Yao[1] **Adam Friedman**[1] **Pranam Chatterjee**[5,6,7†] **Sherwood Yao**[1†]

[1] Atom Bioworks
[2] Department of Biological Sciences, Rensselaer Polytechnic Institute
[3] Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute
[4] Future of Computing Institute, Rensselaer Polytechnic Institute
[5] Department of Biomedical Engineering, Duke University
[6] Department of Computer Science, Duke University
[7] Department of Biostatistics and Bioinformatics, Duke University

Abstract

Aptamers are short, single stranded DNA or RNA sequences that bind to specific targets such as proteins, small molecules, as well as cells, with high affinity and specificity. These aptamers can serve as molecular recognition elements, making them valuable in therapeutic applications, diagnostics, and targeted drug delivery as alternatives to antibodies. Aptamers are typically developed through in vitro SELEX. Despite the levels of success that have been realized through the development and enhancements in SELEX, several challenges persist. To overcome these limitations, we have developed AptaBLE (Aptamer Binding LanguagE), a large language model capable of predicting aptamer-protein interactions and generating novel aptamer sequences against diverse protein targets. Here we demonstrate how AptaBLE leverages fused embeddings to score aptamer-protein binding in a structure-agnostic fashion. We report on performance gains that can be realized via AptaBLE when compared to other deep learning methods. Lastly, we highlight how AptaBLE can be used in a generative capacity to produce novel aptamers with favorable binding profiles.

*Contributed equally.
†Corresponding author

Machine Learning for Structural Biology Workshop, NeurIPS 2024.