digitalSELEX:
in silico Platform for Designing Oligonucleotides

Stephen Hummel[1,3], Jose Bento[2], & Tim van Opijnen[1]

1. Biology Department, Boston College, Chestnut Hill, MA
2. Computer Science Department, Boston College, Chestnut Hill, MA
3. Department of Chemistry and Life Science, United States Military Academy, West Point, NY

Rapid identification of biological and chemical threats as well as rapid development of therapeutics is critical for the warfighter and maintaining combat readiness particularly in remote and austere locations. Antibodies have long been used for identifying pathogens on biosensor devices and as therapeutics to target both viral and bacterial pathogens. Aptamers are single-stranded oligonucleotides, RNA or DNA, and analogous to antibodies for their target recognition and range of applications. These oligonucleotides, however, are typically one-tenth the molecular weight of antibodies and can be chemically synthesized making them immune to batch variations like antibodies.

The gold standard method for identifying aptamers with both high affinity and specificity towards their target protein is achieved through a process known as a **S**ystematic **E**volution of **L**igands by **EX**ponential (SELEX) enrichment. There are multiple SELEX variations but typically consist of exposing an initial oligonucleotide library (~10^{15} molecules) to both target and non-target molecules over success rounds (10-15) prior to sequencing and then further *in vitro* validation.

While the SELEX method can produce high affinity aptamers (dissociation constants in the low nanomolar range), the repetitive nature and stringent counter-selection steps make the process both time consuming and expensive. The in vitro selection process limits applications aspects to only those replicated in the SELEX process (*e.g.,* temperature, pH, and ion concentration). There has been recent work to incorporate ML/AI algorithms into the SELEX process to improve the limited success rate. These combination works have been limited by the massive dataset required for aptamer development still requires *in vitro* selection, meaning environment dependence, and a lack of fidelity on actual nucleotide – amino acid interaction sites. To overcome the low success rate, improve cost, and reduce time, we have developed an *in silico* oligonucleotide design platform, digitalSELEX.

Our platform breaks down a target protein to identify and cluster accessible / biologically relevant atoms. Nucleotides are assigned to the corresponding amino acids and an initial sequence is optimized using a constrained genetic algorithm towards the desired application (*e.g.*, therapeutic or biosensor probe). The sequence then undergoes an *in silico* counter-selection step using molecular docking to optimize the nucleotide sequence with simultaneous random perturbations to ensure high affinity for the target and low affinity for non-target proteins. Once the oligonucleotide is designed, it is chemically synthesized, and the affinity and specificity are determined using flow cytometry. Our digitalSELEX platform reduces the aptamer identification process from several months to a few days while simultaneously reducing cost.