TOM: Transformer Optimization of mRNA Initiator Region- Increasing Custom Protein Production for Cheaper Biopharmaceuticals, Helping Climate Change, Better mRNA Vaccine Design + More!

Gupta , Ritvik (School: Johnston High School)

Globally, countless organisms suffer from infections, diseases, and environmental threats. Custom proteins are at the heart of solutions to these challenges. Custom proteins are life-saving treatments like antibodies, anti viral/cancer peptides, essential hormones, vaccines, and even offer hope for sustainability by breaking down plastics and capturing carbon. Recombinant Protein Production enables scientists to insert custom DNA into host cells to produce specific proteins. Current industry-standard tools to optimize recombinant DNA use the Codon Adaptation Index (CAI), yet doing this changes local mRNA secondary structure and Minimum Free Energy (MFE) at the translation initiation region, creating hairpins that block ribosome loading and initiation, thus limiting protein production. mRNA secondary structure around the start codon (disrupting ribosome docking/initiation) is the true rate-limiting barrier to protein synthesis, and it's 22x more correlated to protein production than CAI. This project introduces TOM, a novel Transformer deep learning model that optimizes mRNA secondary structure at the critical translation initiation region. Over 40 million natural E. Coli initiation sequences were downloaded, where through a rigorous data filtration effort, only 10,000 ideal, non redundant, and naturally occurring sequences were used to train the model. TOM is benchmarked on MFE, adenine count, codon usage, and a negative element analysis against industry standard optimizers and significantly outperforms. TOM's improvement of the MFE at the initiation region offers a ~4x increase in protein production by addressing the rate-limiting step of translation initiation. This research opens doors to cheaper pharmaceuticals and a way to mitigate the climate crisis.

Awards Won: