A Multi-Stage Computational Pipeline for Designing Therapeutic Peptides Against Alpha-Synuclein Aggregation, Utilizing a Geometry-Aware Machine Learning Model Trained on ProLIF Data

Ezhilarasan, Samanyu (School: Nikola Tesla STEM High School)

Parkinson's disease is the world's most rapidly growing neurodegenerative disease, and the cause is largely attributed to the misfolding of a protein in the brain known as alpha-synuclein (a-syn). Creating a targeted peptide therapy to prevent the misfolding of a-syn into a toxic molecule would be the first step in decreasing the appearance of Parkinson's cases worldwide. However, current drug-discovery efforts are hindered by slow, costly experimental screening pipelines and a lack of fully integrated computational methods that are capable of rapidly generating and validating novel therapeutic candidates from scratch. As such, the main objective of this study was to utilize a thorough computational workflow to screen for promising de novo peptide candidates and create a geometry-aware machine learning model that provides a quick and accurate validation system for potential candidates. ProtGPT2 was used to generate a batch of 170 biologically-sound peptides, which were then evolved in an evolutionary algorithm over 1000 generations, with the top 200 candidates from the final population being taken to AutoDock Vina to dock. The peptides were docked against the NAC region of a-syn and ranked based on their binding affinities to find the top 10 candidates, with PEP185 achieving a ?G of -10.17 kcal/mol. 776 small molecules acting as active and failed inhibitors of a-syn aggregation were docked against the NAC region and the protein-ligand interaction fingerprints (PLIFs) were used to train a Random Forest model that achieved an AUC of 0.751 ± 0.031. These results suggest the possibility of an improved accuracy rate with more training, providing a single computational method capable of generating accurate validation scores for peptide inhibitors.

Awards Won: