FastFold: Streamlining End-to-End Deep Learning Protein Domain Prediction on COVID-19 Mutations and Other Universal Applications

Jing, Tim (School: Lynbrook High School)

Computationally determining protein structure and function is promising, but researchers currently lack a complete end-to-end open-source tool capable of sequence alignment, de novo protein structure prediction, and 3D visualization to identify functional differences in domains. The purpose of this study was to satisfy that need efficiently and practically with FastFold while contributing to our understanding of the COVID-19 pathogen. FastFold reads .FASTA DNA/Amino Acid (AA) files and subsequently implements mutation identification, multi-sequence alignment, and secondary structure prediction. Then, FastFold derives Position-Specific-Scoring-Matrix (PSSM) input features using a Biopython consensus sequence that can be incorporated in a deep convolutional neural network utilizing Keras and ResNets to predict inter-residue distances between all sets of amino acids. With the output, FastFold retrieves a fully viewable 3D protein structure with a third-party tool. FastFold is significant as it allows scientists to efficiently predict any protein structure with just a DNA/AA sequence. For instance, FastFold identified notable structural differences in the S1 Subunit of the COVID-19 surface protein, including the presence of structural differences due to various mutations in the UK/South Africa variants, different termination orientations, and a possible lysine interaction site after a few hours of computing. FastFold allows researchers to determine structural differences from any mutation, accelerating research on proteins ranging from Alzheimer's pathogenesis to COVID-19 vaccine efficacy. Crucial conjectures can be determined more efficiently without protein isolation and X-ray crystallography or NMR spectroscopy, creating universal applications in biology.

Awards Won:
Third Award of $1,000