ENGINEERING PROTEINS WITH 3D CONVOLUTIONAL NEURAL NETWORKS 8

Daniel Diaz, The University of Texas at Austin, USA
danny.diaz@utexas.edu

An extremely important task in biotechnology is the ability to engineer proteins by introducing mutations into their sequences, which ultimately alters their folded structure and function. In nature, this process occurs via random mutation and selection, also known as evolution. In the Ellington lab, we have trained a self-supervised, 3D convolutional neural networks (http://www.mutcompute.com) over the entire Protein structure DataBank (PDB) to learn the optimal pockets for a given amino acid within a given protein are or should be. In a manner similar to other 3DCNNs, our MutCompute algorithm can 'see' which of 20 amino acids is best 'fit' for a given pocket on a protein. In most cases (ca. 80%), MutCompute returns the wild-type (natural) amino acid, but in some cases, it uses its broad computer vision to adjudge that a different amino acid might be a better fit for a pocket. In these cases, have successfully introduced mutations to improve function, basically anticipating what evolution might have done on a longer timescale. We have used this insight to engineer a wide variety of proteins, from improving the fluorescence of blue-fluorescent protein by 10-fold, to increasing the stability of a plastic-eating enzyme by over 10 degrees Celsius (to the point where it can completely degrade plastic packaging within 48 hours), to creating stabilized enzymes that have been used in COVID19 diagnostics. Currently, we are developing methodologies to tailor our CNN models to antibody-antigen and protein-ligand interactions with the goal of accelerating substrate specificity engineering for antibody maturation and enzyme catalysis, respectively.