A DEEP LEARNING TOOL FOR PROTEIN ENGINEERING

Huimin Zhao, University of Illinois at Urbana-Champaign, USA
zhao5@illinois.edu

Machine learning has been increasingly used for protein engineering. However, because the general sequence contexts they capture are not specific to the protein being engineered, the accuracy of existing machine learning algorithms is rather limited. In this talk, I will present ECNet (evolutionary context-integrated neural network), a deep-learning algorithm that exploits evolutionary contexts to predict functional fitness for protein engineering. This algorithm integrates local evolutionary context from homologous sequences that explicitly model residue-residue epistasis for the protein of interest with the global evolutionary context that encodes rich semantic and structural features from the enormous protein sequence universe. As such, it enables accurate mapping from sequence to function and provides generalization from low-order mutants to higher-order mutants. We show that ECNet predicts the sequence-function relationship more accurately as compared to existing machine learning algorithms by using ~50 deep mutational scanning and random mutagenesis datasets. Moreover, we used ECNet to guide the engineering of TEM-1 β-lactamase and identified variants with improved ampicillin resistance with high success rates.