A Multi-Layer AI-Assisted in silico Pipeline for Generalizable Enzyme Engineering via Computational Directed Evolution

Bilgin, Gokce Ceyda (School: Bursa Halil Inalcik Bilim ve Sanat Merkezi)

Enzymes are widely used across healthcare, textiles, and environmental technologies, requiring optimization through enzyme engineering. Traditional approaches rely on wet-lab trial-and-error experimentation, which is costly, time-consuming, and dependent on advanced laboratory equipment, limiting accessibility and enzyme-based technology advancement. This project developed a generalizable, computational pipeline and platform to address these limitations by minimizing wet-lab dependency. It consists of four stages. First, enzyme data from the BRENDA database was filtered to construct a balanced dataset. Second, five regression algorithms were trained and benchmarked; the best-performing model was selected to predict whether variants achieve the adaptation. Third, computational directed evolution generated variant sequences through literature-guided mutations applied to the wild-type sequence, with each iteration evaluated against the target property. Achieving target adaptation, however, doesn't guarantee evolutionary plausibility. Therefore, a fourth stage implemented a multi-layer validation system: ESM-2, a protein language model, assessed sequence-level plausibility, while ESMFold evaluated structure-level confidence. Variants scoring high in both layers represent successfully adapted sequences with strong evolutionary plausibility. As a case study, cold adaptation (55°C to 35°C) was performed on Tc-Cut1 (T. cellulosilytica) and heat adaptation (48°C to 68°C) on Pa-LipA (P. aeruginosa), generating new variants with desired adaptation. This low-cost, modular pipeline establishes an accessible framework for computational enzyme engineering, reducing dependence on costly laboratory infrastructure and making advanced protein design accesible to researchers.

Awards Won: