CABBI
CENTER FOR ADVANCED BIOENERGY AND BIOPRODUCTS INNOVATION

A Generalized Platform for Artificial Intelligence-Powered Autonomous Enzyme Engineering

Background/Objective

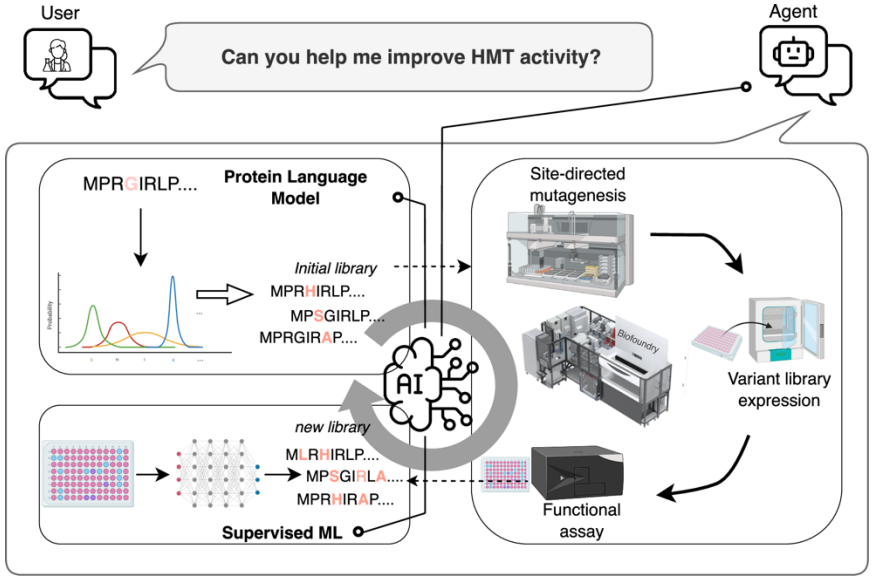
Engineering proteins with desired functions is slow, expensive, and specialist-dependent. We developed a generalized platform for autonomous protein engineering that integrates machine learning (ML) and large language models (LLMs) with biofoundry robotic automation to eliminate the need for human intervention, judgement, and domain expertise, requiring only an input protein sequence and a quantifiable fitness measure.

Approach

Initial protein variants were predicted using unsupervised models, including a protein LLM and an epistasis model. The variant library was constructed and screened using automated modular workflows over a biofoundry. Assay data from each cycle trained an ML model to predict variants for subsequent iterations, followed by next round of biofoundry-driven automated build and test. To validate the system, we engineered two enzymes, *Arabidopsis thaliana* halide methyltransferase (AtHMT) and *Yersinia mollaretii* phytase (YmPhytase), for increased activity compared to the wild-type enzymes.

Results

Both enzymes were engineered for four rounds over four weeks and required fewer than 500 variants per enzyme to construct and characterize. The engineered enzymes showed a 90-fold increase in substrate preference, a 16-fold increase activity for AtHMT, and a 26-fold increase in activity for YmPhytase.



Overview of the generalized platform for autonomous protein engineering.

Significance/Impacts

This autonomous experimentation platform makes engineering proteins faster, cheaper, and specialist-independent and paves the way for rapid advancements across various industries that employ synthetic biology.

Singh et al. 2025 "A Generalized Platform for Artificial Intelligence-powered Autonomous Enzyme Engineering." *Nature Communications*. DOI: 10.1038/s41467-025-61209-y.