BRC Science Highlight
January 2022

EnZymClass: Substrate Specificity Prediction Tool of Plant Acyl-ACP Thioesterases Based on Ensemble Learning

Background/objective

Plant acyl-acyl carrier protein (ACP) thioesterases (TEs) are enzymes that produce valuable oleochemical precursors in microbial and plant hosts. However, many acyl-ACP TEs exhibit activity that is not selective toward high-value products of interest. Testing for the desired activity can be an expensive, time-consuming process due to the low throughput of available screening methods. Computationally predicting an enzyme's function could accelerate this process, however standard machine learning (ML) approaches are challenged by inadequate amounts of high-quality information on enzymes and sequences required by these techniques. We designed a ML method, Ensemble method for enZyme Classification (EnZymClass), to characterize acyl-ACP TEs using a small training dataset of disparate sequence-function information.

Approach

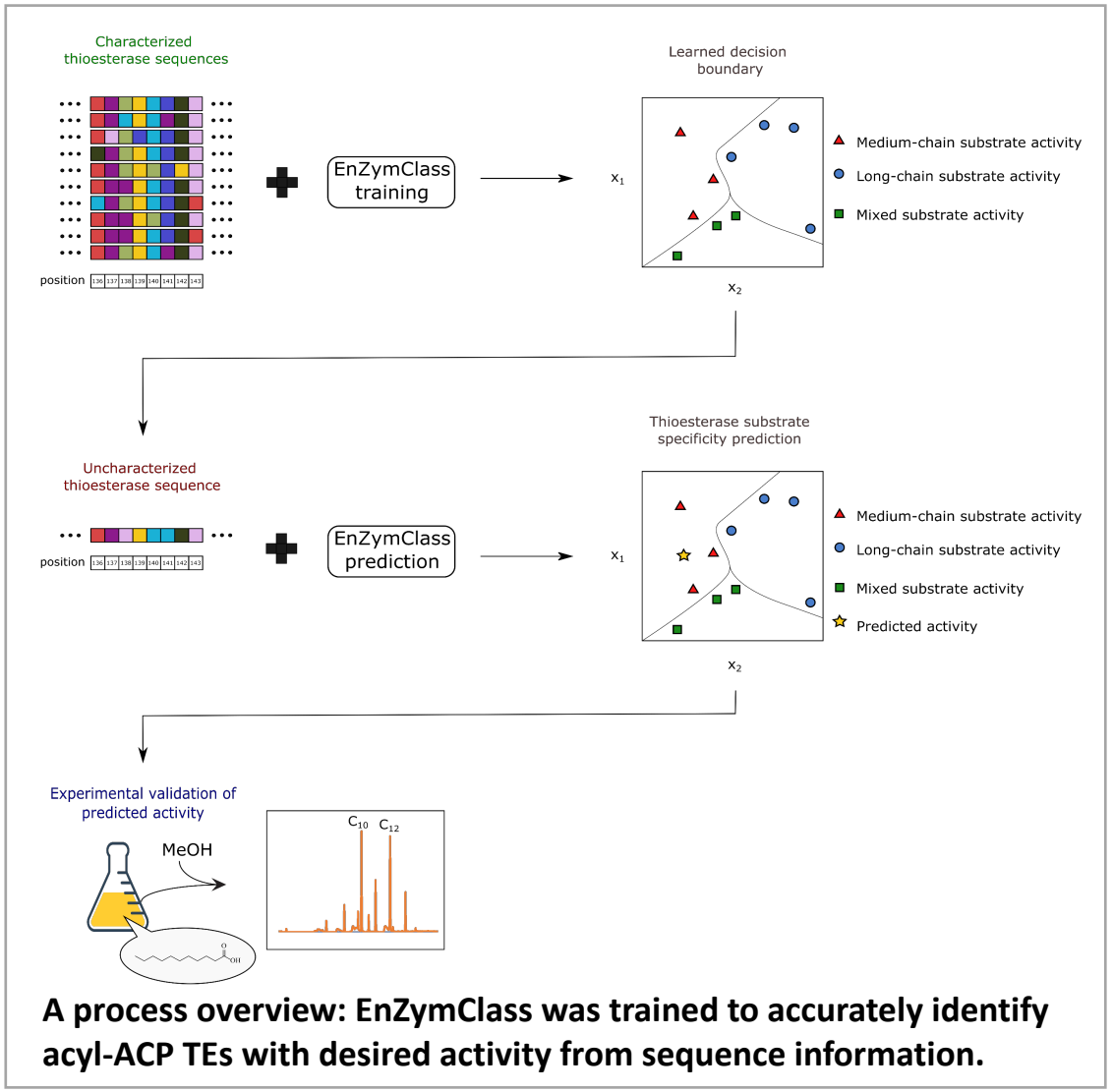
- Trained EnZymClass with sequence-function information on acyl-ACP TEs from literature.
- Used EnZymClass to identify which acyl-ACP TEs had desired substrate activity among uncharacterized sequences in an enzyme database.
- Experimentally verified predictions from model.

Results

- EnZymClass identified two acyl-ACP TEs with desired substrate activity among a database containing 617 TE sequences.
- Engineered one of the TEs to improve titer over wildtype (WT) by 4.2-fold.

Significance

- New ML methods like EnZymClass can use existing limited datasets to accelerate bioprospecting efforts and thus facilitate selection of appropriate enzymes for further engineering.



A process overview: EnZymClass was trained to accurately identify acyl-ACP TEs with desired activity from sequence information.

Banerjee, D., Jindra, M.A., Linot, A.J., Pfleger, B.F., Maranas, C.D. Dec. 16, 2021. "EnZymClass: Substrate Specificity Prediction Tool of Plant Acyl-ACP Thioesterases Based on Ensemble Learning." *Current Research in Biotechnology* 4: 1-9. DOI: 10.1016/j.crbiot.2021.12.002.

CABBI CENTER FOR ADVANCED BIOENERGY AND BIOPRODUCTS INNOVATION
U.S. DEPARTMENT OF ENERGY