

Machine Learning-based Prediction of Enzyme Substrate Scope: Application to Bacterial Nitrilases

Zhongyu Mou¹, Jason Eakes¹, Connor Cooper², Carmen Foster¹, Robert Standaert¹, Mircea Podar¹, Mitchel Doktycz¹, and Jerry Parks¹

¹Oak Ridge National Laboratory

²The University of Tennessee Knoxville-Oak Ridge National Laboratory Graduate School of Genome Science and Technology

May 7, 2020

Abstract

Predicting the range of substrates accepted by an enzyme from its amino acid sequence is challenging. Although sequence- and structure-based annotation approaches are often accurate for predicting broad categories of substrate specificity, they generally cannot predict which specific molecules will be accepted as substrates for a given enzyme, particularly within a class of closely related molecules. Combining targeted experimental activity data with structural modeling, ligand docking, and physicochemical properties of proteins and ligands with various machine learning models provides complementary information that can lead to accurate predictions of substrate scope for related enzymes. Here we describe such an approach that can predict the substrate scope of bacterial nitrilases, which catalyze the hydrolysis of nitrile compounds to the corresponding carboxylic acids and ammonia. Each of the four machine learning models (linear regression, random forest, gradient-boosted decision trees, and support vector machines) performed similarly (average ROC = 0.9, average accuracy = ~82%) for predicting substrate scope for this dataset. The approach is intended to be highly modular with respect to physicochemical property calculations and software used for docking and modeling.

Hosted file

nitrilases_submit_final.pdf available at <https://authorea.com/users/319169/articles/448933-machine-learning-based-prediction-of-enzyme-substrate-scope-application-to-bacterial-nitrilases>