

## Characterizing Thiamine Diphosphate Dependent Enzymes for Promiscuous C-C Bond Formation Catalysis

Tracey Dinh<sup>1\*</sup>([tracey.dinh@u.northwestern.edu](mailto:tracey.dinh@u.northwestern.edu)), Bradley W. Biggs<sup>1</sup>, Lindsay Caesar<sup>2</sup>, Neil L. Kelleher<sup>2,3,4</sup>, Paul M. Thomas<sup>4</sup>, Linda J. Broadbelt<sup>1</sup>, Keith E.J. Tyo<sup>1</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL; <sup>2</sup>Department of Molecular Biosciences, Northwestern University, Evanston, IL; <sup>3</sup>Department of Chemistry, Northwestern University, Evanston, IL; and <sup>4</sup>Proteomics Center of Excellence, Northwestern University, Evanston, IL

<https://pamspublic.science.energy.gov/CCBond>

**Project Goals:** The goal of this project is to prospect novel carbon-carbon bond ligation reactions in thiamine diphosphate-dependent enzymes using predictive activity models. Carboligase enzymes will be screened for  $\alpha$ -keto acid condensation and machine learning models will be developed to elucidate their catalytic landscape. Activity models will be also be used to predict potential effects of selected enzymes on the *E. coli* metabolome.

**Abstract:** Increasing recognition of non-canonical enzyme activity has revealed potential problems for heterologous expression; however, understanding the potential cell burden due to promiscuous enzyme activity remains a challenge. Toward this end, our team seeks to develop cheminformatics tools that predict enzyme substrate promiscuity and the resulting metabolomic consequences. Merging both experimental and computational screening approaches, this work aims to comprehensively characterize the catalytic landscape of thiamine diphosphate (ThDP)-dependent enzymes and prospect novel promiscuous transformations. We developed a high throughput activity assay to screen carboligase enzymes against diverse sets of  $\alpha$ -keto acid substrates. Activity data from this screen will be used to build support vector machine classifiers for in silico reaction screening. Comprehensive activity landscapes developed from predictive models will provide insight into substrate selectivity and will facilitate enzyme active site engineering. Robust activity models for each carboligase enzyme will allow us to predict promiscuous activity on genome-scale models of host organisms such as *E. coli*.

This work is supported by DOE grant DE-SC0019339.