An Integrated Experimental/Informatic Approach to Characterize Proteins of Unknown Function (PUFs) in Bioenergy-Relevant Microbes

U.S. DEPARTMENT OF ENERGY | Office of Science

Background

- Mass spectrometry-based proteomics can identify and quantify thousands of proteins from individual microbial species, but a significant percentage of these are proteins remain unannotated and therefore become classified as **proteins of unknown function (PUFs).**
- Due to the difficulty in extracting meaningful metabolic information, PUFs are often overlooked or discarded during data analysis, even though they might be critically important in functional activities, in particular for metabolic engineering research.

Approach

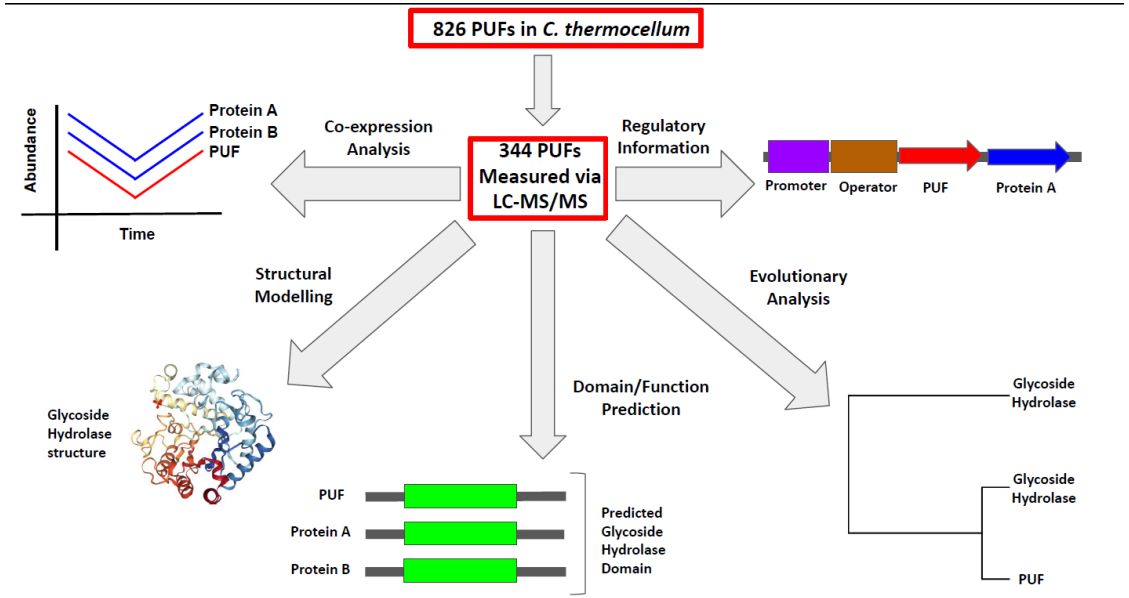
- This study identified putative functions of PUFs via "guilt-by-association" (GBA) by integrating co-expression analysis generated from high-throughput mass spectrometry data with functional annotation approaches, structural modeling, phylogenetic analyses, and gene regulatory information.

Outcome

- Our GBA approach was able to identify putative functions for several PUFs in *C. thermocellum*, with a specific focus on proteins related to cellulose solubilization and biofuel production.
- Overall results revealed 33 PUFs with new tentative functional annotations. This information will be of significant interest to the bioenergy research community who may now validate and characterize targeted PUFs in *C. thermocellum*.

Significance

- This work demonstrates the value of leveraging sequence homology-based annotations with empirical data based on the concept of GBA to predict putative functions for PUFs, opening avenues to further interrogation via targeted engineering experiments.



Overall flowchart of integrated experimental/bioinformatic approach for characterizing proteins of unknown function (PUFs) with a guilt-by-association approach. Proteins that are measured by differential expression in mass spectrometric measurements become candidates for examination with phylogenetic alignments, operon localization, and structural homology methods.

Poudel, S; A Cope; K O'Dell, A Guss; H Seo; C Trinh; R. Hettich. 2021. "Identification and characterization of proteins of unknown function (PUFs) *in Clostridium thermocellum DSM 1313* strains as potential genetic engineering targets," S *Biotech. For Biofuels* 14 Article 116.

THE CENTER FOR BIOENERGY INNOVATION