Genome-scale metabolic and regulatory network reconstruction of *Pyroccous furiosus*

Dmitry A. Rodionov [1]*(rodionov@sbpdiscovery.org), Ying Zhang [2]*(yingzhang@uri.edu), Ke Zhang, [2] Jason Vailionis, [2] James R. Crosby,[3] Ryan G. Bing,[3] Hailey O'Quinn, [4] Gina L. Lipscomb, [4] Tania N. N. Tanwee,[4] Robert M. Kelly,[3] and **Michael W. W. Adams[4]**

[1]Sanford-Burnham-Prebys Med. Discovery Institute, La Jolla, CA; [2]University of Rhode Island, Kingston, RI; [3]North Carolina State University, Raleigh, NC; [4]University of Georgia, Athens, GA;

Project Goals: The goal of this project is to establish the two non-model microorganisms *Caldicellusiruptor bescii* and *Pyroccous furiosus* as platforms for sustainable production of industrial chemicals using renewable plant biomass. We aim to engineer *C. bescii* and P*. furiosus* to reincorporate CO_2 and H_2 generated during fermentation of lignocellulose as an additional source of carbon and energy, potentially reaching net zero CO_2 emission. Furthermore, *P. furiosus* will be engineered with enzymes from *C. bescii* to allow it to degrade non-pretreated plant biomass. System-wide metabolic and regulatory modeling of both organisms will be used to optimize biomass degradation and desired product yield and selectivity.

Pyrococcus furiosus is an extremely thermophilic, strictly anaerobic, sugar-utilizing archaeon microorganism that grows up to 103°C on starch, laminarin, maltose, trehalose, cellobiose and beta-glucan oligosaccharides, but not crystalline cellulose, xylan or monosaccharides In this study, we applied a subsystems-based approach combining comparative genomics, transcriptional regulon prediction, and genome-scale modeling to reconstruct an integrated view of the metabolic and regulatory network of *P. furiosus*. The complete genomes of over 30 *Pyrococcus and Thermococcus* species were used for ortholog mapping and comparative analysis. Functional gene assignments, genome context analysis, comparative analysis of orthologous genes and DNA upstream regions, gene co-occurrence analysis and protein similarity searches were performed in the SEED environment [1]. We also used genome annotations from Swiss-Prot, KEGG, TCDB, and RegPrecise databases and published experimental data. The previously generated RNASeq datasets for whole-genome gene expression and transcriptional start sites obtained for *P. furiosus* grown on different carbon sources (glucose, maltose, cellobiose) were used for validation of reconstructed transcriptional regulons and for refinement of transporter specificities.

The reconstruction and analyses of genome-scale models (GEMs) combines the stoichiometry of metabolic processes with the definition of condition-specific metabolic constraints into predicting microbial growth and biochemical production. The GEM curation was done with the support of PSAMM software [2] to incorporate the known and predicted metabolic functions of enzymes and transporters. We have obtained a first draft of the *P. furiosus* GEM that contains 409 metabolic genes, 860 metabolites (non-unique) and 538 metabolic reactions, thereby covering 81% of genes with EC number assignments. These include a native RuBisCo enzyme that allows *P. furiosus* to incorporate CO_2 into its central carbon metabolism, a diverse range of catabolic and anabolic pathways, and biomass production equations that incorporate the experimentally calibrated proportion of major cell components. Besides the genes encoding metabolic enzymes, the *P. furiosus* GEM contains 198 genes encoding components of metabolic transporters (89% of total number of putative transporter genes). Growth predictions made by the *P. furiosus* GEM were validated by matching metabolic simulations with growth measurements in batch and chemostat culture using defined media. This

model, alongside our preexisting model of *C. besci* [3], serves as a stepping stone for the engineering of future strains to enable and enhance the yields of bio-based fuels and chemicals.

Five industrial compounds have been selected as engineering targets with the potential to achieve net zero carbon emission. We have developed strategies for the production of these compounds by incorporating enzymes from other thermophiles, as well as engineering the *P. furiosus* RuBisCo into *C. bescii*. Simulations using the two GEMs demonstrated that both *P. furiosus* and *C. bescii* can theoretically achieve negative CO_2 production, while generating products when CO_2 and H_2 are supplied. Additionally, enhanced product formation has been predicted by simulation of multi-gene knockouts and introduction of hydrogenases from *P. furiosus* to *C. bescii*.

Previously, we used a comparative genomics approach to reconstruct the carbohydrate utilization regulatory network in *C. besci* and related bacteria [4]. Here, we used the same approach to identify DNA-binding motifs and reconstruct regulatory networks for 15 out of 65 TFs encoded in *P. furiosus* genome. Most of these TFs are local regulators, while two reconstructed regulons (SurR and TrmBL1) include genes from multiple metabolic pathways. TrmBL1 is a global regulator of the carbohydrate metabolism that co-regulates large sets of genes involved in the starch/maltose utilization, glycolysis and gluconeogenesis. The reconstructed metabolic and regulatory networks are used to generate metabolic models for *C. besci* and *P. furiosus*, and to guide engineering strategies to improve bioproduct formation. The goal is to increase carbon flux from substrates to engineered products, which we hypothesize is key to producing industrially relevant titers and yields.

IMPORTANCE In this study, we developed a predictive model for simulating the metabolism of the non-model organism, *P. furiosus*. The simulation predictions can provide potential directions for the more efficient metabolic engineering design for bio-based chemicals. We will combine modeling and regulatory mechanisms and use extensive experimental validation to enable model parameterization. The systems-wide integration of models at the metabolic and regulatory levels will enable discovery of “non-intuitive” designs that may not be apparent from knowledge-based optimization of targeted pathways.

References

1. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. (2014) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 42:D206-14. doi: 10.1093/nar/gkt1226.

2. Steffensen JL, Dufault-Thompson K, Zhang Y (2016) PSAMM: A Portable System for the Analysis of Metabolic Models. *PLoS Comput Biol* 12. doi:10.1371/journal.pcbi.1004732.

3. Zhang K, Zhao W, Rodionov DA, Rubinstein GM, Nguyen DN, Tanwee TNN, Crosby J, Bing RG, Kelly RM, Adams MWW, Zhang Y. (2021) Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for bio-based chemical production. *mSystems*. 6(3):e0135120. doi: 10.1128/mSystems.01351-20.

4. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM, Tanwee TNN, Bing RG, Crosby JR, Nookaew I, Basen M, Brown SD, Wilson CM, Klingeman DM, Poole FL 2nd, Zhang Y, Kelly RM, Adams MWW. (2021) Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. *mSystems.* 6(3):e0134520. doi: 10.1128/mSystems.01345-20.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program under Award Number DE-SC0019391 and DE-SC0022192