Genome-Scale Metabolic and Regulatory Network Reconstruction of *Caldicellulosiruptor bescii* Leads to the Reconstruction of Predictive Models for Bioproduct Generation

Dmitry A. Rodionov,[1]*(rodionov@sbpdiscovery.org), Ying Zhang, [2](yingzhang@uri.edu), Weishu Zhao, [2] Ke Zhang, [2] James R. Crosby,[3] Ryan G. Bing,[3] Diep M. N. Nguyen,[4] Tania N. N. Tanwee,[4] Gabriel M. Rubinstein,[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: We are using systems biology-guided approaches to develop a microbial metabolic engineering platform for the non-model organism, *Caldicellulosiruptor bescii*, the most thermophilic lignocellulose-degrading organism known with an optimal growth temperature near 80°C. This work leverages recent breakthrough advances in the development of molecular genetic tools for this organism, complemented by a deep understanding of its metabolism and physiology gained over the past decade of study in the PIs' laboratories. We are applying the latest metabolic reconstruction and modeling approaches to optimize biomass to product conversion using switchgrass as the model plant and acetone and other fermentation products as targets. The overarching goal is to demonstrate that a non-model microorganism, specifically an extreme thermophile, can be a strategic metabolic engineering platform for industrial biotechnology using a systems biology-based approach.

Caldicellulosiruptor bescii is an extremely thermophilic, strictly anaerobic, gram-positive bacterium. It is the most thermophilic cellulolytic bacterium known to date (T_{opt} 78°C, T_{max} 90°C), and it can use a wide range of simple and complex carbohydrates. Its ability to degrade plant biomass without enzymatic or chemical pretreatments and at a high optimum growth temperature offers several advantages for industrial applications (1). Engineered *C. bescii* strains have been shown to produce desired bioproducts, such as ethanol, from un-pretreated plant biomass through consolidated bioprocessing (CBP). However, efficient metabolic engineering requires in-depth understanding of its metabolic and transcriptional regulatory networks. Previous experimental studies identified a variety of carbohydrate-active enzymes in *C. bescii* and related *C. saccharolyticus* species, while prior transcriptomic experiments identified their putative carbohydrate uptake transporters.

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 *C. bescii*. The reconstruction of carbohydrate utilization regulatory network includes the predicted binding sites for 34 mostly local transcription factors and points to the regulatory mechanisms controlling expression of genes involved in degradation of plant biomass (1). The Rex and CggR regulons control the central glycolytic and primary redox reactions. The identified transcription factor binding sites and regulons were validated with transcriptomic and transcription start site experimental data for *C. bescii* grown on cellulose, cellobiose, glucose, xylan, and xylose. The XylR and XynR regulons control xylan-induced transcriptional response of genes involved in degradation of xylan and xylose utilization. The reconstructed regulons informed the carbohydrate utilization reconstruction analysis and improved

functional annotations of 51 transporters and 11 catabolic enzymes. Using gene deletion, we confirmed that the shared ATPase component MsmK is essential for growth on oligo- and polysaccharides, but not for the utilization of monosaccharides (2).

The reconstruction of regulatory networks of carbohydrate utilization was complemented with a genome-scale model (GEM) of the *C. bescii* metabolism (3). The model was used to examine potential bottlenecks that could be encountered for metabolic engineering of *C. bescii* to produce bio-based chemicals from plant biomass. The model utilizes subsystems-based genome annotation, targeted reconstruction of carbohydrate utilization pathways, and biochemical and physiological based experimental validations. Specifically, carbohydrate transport and utilization pathways involving 160 genes and their corresponding functions were incorporated, representing the utilization of C5/C6 monosaccharides, disaccharides, and polysaccharides such as cellulose and xylan. The model predicted that optimal production from biomass-based sugars of the model product, ethanol, was driven by ATP production, redox balancing, and proton translocation, mediated through the interplay of an ATP synthase, a membrane-bound hydrogenase, a bifurcating hydrogenase and a bifurcating NAD- and NADP-dependent oxidoreductase. These mechanistic insights guided the design and optimization of new engineering strategies for product optimization, which were tested in the *C. bescii* model, showing a two-fold increase in ethanol yields (3).

References

1. Crosby, J. R., Laemthong, T., Lewis, A. M., Straub, C. T., Adams, M. W. W. and Kelly, R. M. (2019) Extreme thermophiles as emerging metabolic engineering platforms. *Curr. Opin. Biotechnol.* **59,** 55-64
2. Rodionov, D. A., Rodionova, I. A., Rodionov, V. A., Zhang, K., Rubinstein, G. M., Tanwee, T. N. N., Crosby, J., Bing, R. G., Nookaew, I., Basen, M., Brown, S. D., Wilson, C., Klingeman, D. M., Poole, F. L., Zhang, Y., Kelly, R. M. and Adams, M. W. W. (2020) "Transcriptional regulation of biomass degradation and carbohydrate utilization in *C. bescii*" *mSystems* (submitted 12/2020)
3. Zhang, K., Zhao, W., Rodionov, D., Rubinstein, G. M., Nguyen, D. N., Tanwee, T. N. N., Crosby, J., Bing, R. G., Kelly, R. M., Adams, M. W. W. and Zhang, Y. (2020) "Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for ethanol production" *mSystems* (submitted 12/2020)

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