

Ensemble cell-wide kinetic modeling of anaerobic organisms to support fuels and chemicals production

Satyakam Dash^{1*} (sud25@psu.edu), Ali Khodayari^{1*} (auk241@psu.edu), Evert K. Holwerda², Daniel G. Olson², M. Ahsanul Islam³, Matthew Theisen⁴, James C. Liao⁴, Gregory Stephanopoulos³, Lee R. Lynd², and **Costas D. Maranas¹**

¹The Pennsylvania State University; ²Thayer School of Engineering at Dartmouth College; ³Massachusetts Institute of Technology; ⁴University of California, Los Angeles.

<http://maranas.che.psu.edu/>

<https://engineering.dartmouth.edu/lyndlab/>

<http://bamel.scripts.mit.edu/gns/>

<http://www.seas.ucla.edu/~liao/>

Project Goals: The goal of the project is to systematically construct dynamic models of *Clostridium thermocellum* and *Clostridium ljungdahlii* by making use of Ensemble Modeling (EM) paradigm through integration of multiple omic information (metabolomic & fluxomic). These models will be instrumental in exploring genetic interventions for overproduction of biofuel products.

Clostridia have garnered the interest of the bioprocess industry due to their unique metabolism. In particular, *C. thermocellum* can metabolize cellulose into biofuels such as ethanol and *C. ljungdahlii* can metabolize syn gas using the unique Wood-Ljungdahl pathway. Despite their increasing role as bio-production platforms, they remain poorly characterized with significant uncertainty in their metabolic repertoire. To this end, we report progress toward developing cell-wide dynamic metabolic model for both organisms using the Ensemble Modeling (EM) paradigm [1] built on a curated genome-scale metabolic (GSM) model foundation. The existing GSM model of *C. thermocellum* was curated to update cofactors based on experimental observations [2] and subsequently used to construct a second-generation GSM model (*iCth446*). The model contains 446 genes and includes 598 metabolites and 637 reactions, along with gene-protein-reaction associations. The GSM model predicted higher yield of ethanol production for malate knockout ($\Delta mdh \Delta me$) mutant [3] and showed that lactate dehydrogenase knockout (Δldh) mutant did not have an effect on growth rate as observed experimentally [4]. The GSM model was next used to construct a core kinetic metabolic model of the *C. thermocellum*'s central metabolism containing 90 reactions and 76 metabolites with cellobiose as the carbon source under anaerobic growth condition. This core kinetic model accounts for all major biomass precursors. It encompasses the cellobiose degradation pathway, glycolysis/gluconeogenesis, the pentose phosphate (PP) pathway, the TCA cycle, pyruvate metabolism, anaplerotic reactions, alternative carbon metabolism, nucleotide salvage pathway, and 22 substrate level regulatory interactions extracted from BRENDA. Model parameterization was carried out using the EM approach by simultaneously imposing the mutant library data recently measured and provided by the Lynd group. This dataset includes *C. thermocellum* variants with mutations in lactate, malate, acetate, and hydrogen production pathways and combinations thereof defining 22 specific mutants with measured concentrations of various fermentation products such as acetate, lactate, formate, ethanol, hydrogen, carbon dioxide, amino acids and cellobiose (32 measured concentrations per mutant) (see Figure).

The preliminary analysis of the constructed kinetic model showed accurate predictions of cytosolic concentrations in multiple mutant strains, particularly for the pathways with measured data and substrate level regulations during parameterization. For example, the increase in metabolite concentrations (*i.e.*, 1.5 fold) of cellobiose and sugar phosphates by up-regulating amino-acid synthesis pathways under ethanol stress conditions [3] was accurately captured by the model. The up-regulation of amino acid synthesis led to an increase in ammonia uptake, which inhibits the phosphofructokinase reaction and results in fructose phosphate accumulation. Similarly, the fold changes in cytosolic concentrations for thirteen metabolites in the Δldh mutant were predicted accurately by the kinetic model [4]. Nevertheless, a leave-one-out cross-validation test to evaluate the robustness of the estimated parameters revealed that the fidelity of the model predictions remained limited to the mutants located in the proximity of mutations used for training

the model. Thus, integration of additional measured datasets under a variety of perturbation scenarios is required to address this challenge, while the effect of regulatory events must also be properly accounted for to make meaningful predictions. To this end, we are currently expanding the range of the training data by including thirteen additional fermentation datasets under different growth (i.e., media) conditions from the mutant library. At the same time, we use a data-driven approach to account for the action of all known regulators upon change in growth condition by adjusting the level of the enzymes under each condition. In addition, relative mRNA concentrations will be integrated into model to narrow down the range of enzyme levels that will vary during parameter estimation. RNA-seq data will be generated by the Liao group for the wild-type organism, as well as for mutant strains that will be constructed by modulating the activity of selected transcriptional regulators (i.e., increasing and/or decreasing promoter strength). The constructed model will also be analyzed by the Liao group using the ensemble modeling robustness analysis (EMRA) to identify the unrobust enzymes and the metabolic conditions which renders them unstable along with potential metabolic targets for iso-butanol overproduction. A similar effort is also underway for *C. ljungdahlii*. The existing GSM was already used as the scaffold for constructing the core metabolic model, composed of 77 reactions and 63 metabolites (see Figure). The network spans glycolysis, the Wood-Ljungdahl pathway and an incomplete TCA cycle with 41 substrate level regulatory interactions from BRENDA based on other Clostridia species. The Stephanopoulos group is applying ¹³C isotope-assisted labeling experiments and metabolic flux analysis on *C. ljungdahlii* to study its central carbon metabolism and the Wood-Ljungdahl pathway (WLP), degeneracy in its TCA cycle, and the C-N assimilation and influences on the WLP. The generated data will subsequently be used to estimate the core model parameters using the same proposed framework.

The work was supported by the genomic science grant from Department of Energy, USA (grant# DE-SC0012377).

1. Khodayari, A., et al., *A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data*. Metabolic engineering, 2014. **25**: p. 50-62.
2. Zhou, J., et al., *Atypical glycolysis in Clostridium thermocellum*. Applied and environmental microbiology, 2013. **79**(9): p. 3000-8.
3. Yang, S.H., et al., *Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress*. BMC Genomics, 2012. **13**.
4. van der Veen, D., et al., *Characterization of Clostridium thermocellum strains with disrupted fermentation end-product pathways*. Journal of industrial microbiology & biotechnology, 2013. **40**(7): p. 725-734.
5. Nagarajan, H., et al., *Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of Clostridium ljungdahlii*. Microbial cell factories, 2013. **12**: p. 118.

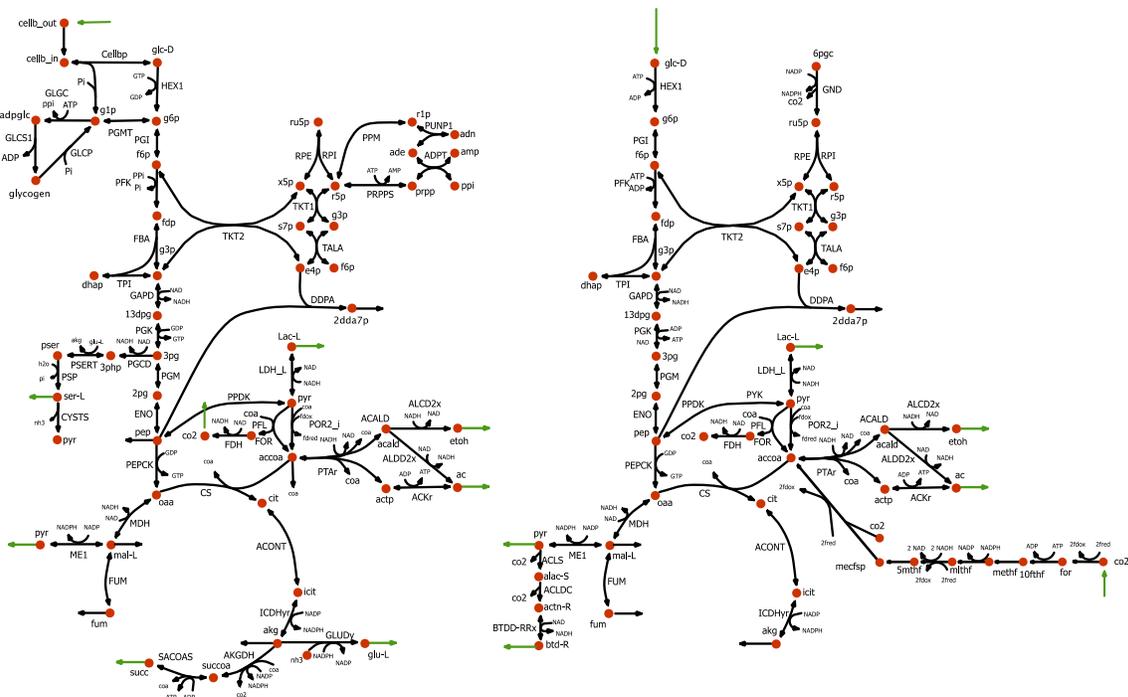


Figure 1: Core metabolic map of *Clostridium thermocellum* and *Clostridium ljungdahlii*. The green arrows represent the metabolite concentrations which are experimentally measured.