# An Updated Genome-scale Metabolic Model of *Clostridium thermocellum* that Simulates All Metabolic Fluxes under Diverse Conditions, Enabling New Insights in Data Analysis and Biocatalytic Strain Design

**Sergio Garcia[1,2], Satyakam Dash[1,3], Costas Maranas[1,3], R. Adam Thompson,[4] Richard Giannone,[1,5] and Cong T. Trinh[1,2,4]**

[1]Center for Bioenergy Innovation, Oak Ridge National Laboratory, Oak Ridge; [2]Department of Chemical and Biomolecular Engineering, The University of Tennessee, Knoxville, TN;
[3]Department of Chemical Engineering, Pennsylvania State University, PA; [4]Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville and Oak Ridge National Laboratory, Oak Ridge; [5]Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge;

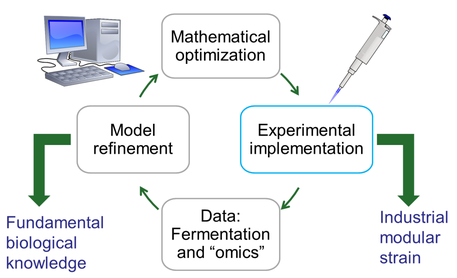
## ABSTRACT

Our exponentially growing world population requires a sustainable bioeconomy from renewable and carbon neutral production of energy and materials using lignocellulosic biomass and organic wastes. Consolidated bioprocessing (CBP) is a promising technology that utilizes a CBP microorganism capable of performing biomass hydrolysis and fermentation in a single step. *Clostridium thermocellum* is a gram positive thermophilic CBP bacterium capable of efficient degradation of untreated lignocellulosic biomass, such as poplar or switchgrass, to produce biofuels and biomaterial precursors. However, *C. thermocellum* has complex and poorly understood metabolism hindering metabolic engineering to achieve high rates, titers, and yields of industrially relevant chemicals, e.g., alcohols and esters. In this study, we developed an updated genome-scale model *of C. thermocellum*, named iCBI655, to account for recent discoveries in the metabolism *of C. thermocellum*, improve the predictability of the model by training it with a broad dataset of experimental fluxes and against known lethality phenotypes, and increase its accessibility and reproducibility through extensive documentation and standard-conforming model organization. Furthermore, we illustrated the use of the model to generate biological insights from multi-scale datasets by simulating intracellular fluxes consistent with measured metabolite secretion fluxes and integration of proteomics data. We anticipate the new model will be useful for metabolic engineering and studying physiology, metabolism, and regulation of *C. thermocellum*.

## Modeling principles

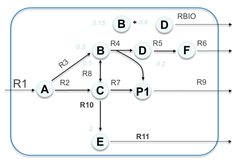
### What is a genome-scale metabolic model and why is it useful?

- A genome-scale metabolic model is a curated, genetically- and biochemically-consistent database of metabolism of an organism.
- The model is part of a design cycle which incorporates our current knowledge and helps us understand experimental data in a holistic and quantitative manner.




### How can the model be used to calculate fluxes?

- Principles of mass balance, thermodynamics, and mathematical optimization enable the use of the database to calculate metabolic fluxes.
- Only reaction stoichiometry information is required.



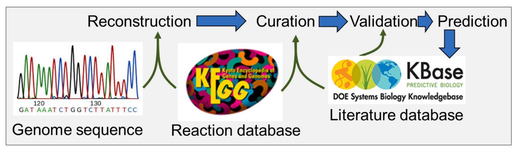

$S_{ij}$ ≔ Stoichiometric coefficient of metabolite $i$ in reaction $j$
$v_j$ ≔ Flux of reaction $j$ (mmol/gCDW/hr)
$C_i$ ≔ Concentration of metabolite $i$
$l_j$ ≔ Lower flux bound of reaction $j$
$u_j$ ≔ Upper flux bound of reaction $j$

*In – out = acc.* Steady state assumption

$$\sum_{j \in J} S_{ij} v_j = \frac{dC_i}{dt} = 0 \text{ for all } i \in Metabolites$$ Mass balance

$$l_j \leq v_j \leq u_j \text{ for all } j \in Reactions$$ Reaction rate bounds

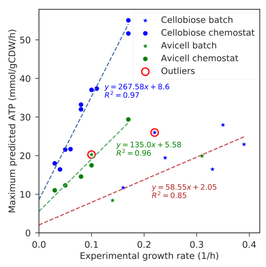
### Software tools




- Multiple tools automate the construction of the model. However a lot of manual inspection has to be performed.
- The previous *C. thermocellum* model, iAT601 (Thompson et al. 2016) was used as a starting point to build the updated model, iCBI655.
- The model construction process is documented in detail using version control software and available at https://github.com/trinhlab.

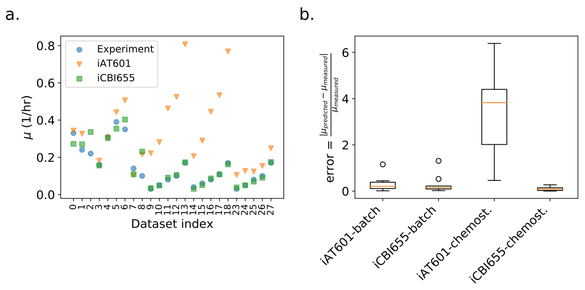
## Model training and validation

### Training model parameters with comprehensive experimental flux dataset




- The model parameters (GAM and NGAM) were trained with an extensive dataset of measured fermentation fluxes from literature.
- We identified a good fit for three separate conditions, cellobiose chemostat, cellulose chemostat, and batch reactor.
- This model training helps accurately predict growth rate and other phenotypes under diverse conditions.

### Comparison of growth-prediction accuracy with respect to previous *C. thermocellum* model




- The updated model can predict growth rate more accurately under diverse conditions than the previous model.

### Prediction accuracy of gene lethality in important biotechnological mutants

| Gene deletions | Medium | Fraction of W.T. growth rate (%) *iAT601* | *iCBI655* | *In vivo* |
|---|---|---|---|---|
| *hydg* | MTC | 100 | 100 | 73 |
| *hydg-ech* | MTC | 85 | 85 | 67 |
| *hydg-pta-ack* | MTC | 100 | 100 | 48 |
| *hydG-ech-pfl* | MTC | 58 | 0 | 0 |
| *hydG-ech-pfl* | MTC + fumarate | 377 | 726 | 0 |
| *hydG-ech-pfl* | MTC + sulfate | 58 | 65 | + |
| *hydG-ech-pfl* | MTC + ketoisovalerate | 97 | 101 | + |

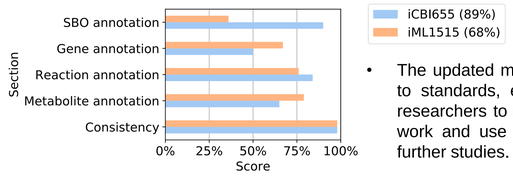
- The updated model can reconcile lethality phenotypes in key redox mutants used for ethanol production.

### Comparison to previous models

| | iSR432 | iCth446 | iAT601 | iCBI665 | iML1515 |
|---|---|---|---|---|---|
| Strain | ATCC27405 | ATCC27405 | DSM1313 | DSM1313 | MG1655 |
| Genes | 432 | 446 | 601 | 665 | 1515 |
| Metabolites | 583 | 599 | 903 | 795 | 1877 |
| Reactions | 632 | 660 | 872 | 854 | 2712 |
| Blocked reactions | 39.2% | 32.1% | 40.8% | 35.1% | 9.8% |
| Reference | [roberts2010] | [dash2017] | [thompson2016] | This study | [monk2017] |

- The increased number of genes and reduced number of blocked reactions indicates better coverage of metabolic function.

### Standard conformance score with *Memote*

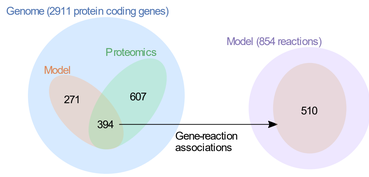



- The updated model conforms to standards, enabling other researchers to reproduce this work and use the model for further studies.

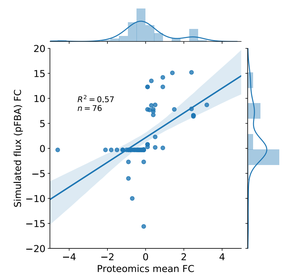
## Application example: New insights form extracellular metabolite and proteomics datasets

### Dataset and analysis approach

- The wildtype strain is compared to the *hydG-ech* deletion strain, a mutant for increased ethanol production.
- This mutant was cultured in batch reactors (Thompson et al. 2015) and extracellular metabolite profiles were measured together with proteomics.
- The model serves as a scaffold to integrate these two disparate data types, together with existing knowledge of *C. thermocellum.*
- The proteomics data maps to 510 (60%) of the reactions in the model.




### Focus on reactions where proteomics fold change is consistent with simulated flux fold change




- Fold change (FC) is computed between the wildtype and mutant form proteomics data in the conventional way.
- The model is constrained to obey measured fluxes and the remaining fluxes are simulated. This is done for the wildtype and mutant separately. Then FC in flux for each reaction is calculated.
- 76 reactions have consistent FC (i.e., both FC agree in sign and are different than 0) in both proteomics and flux simulations.
- Discrepancies in FC magnitude can help discover enzyme catalytic efficiency or regulation by metabolite concentration.

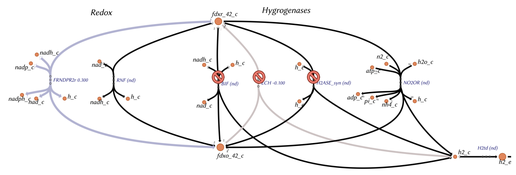
### Fold change contextualization in interactive map

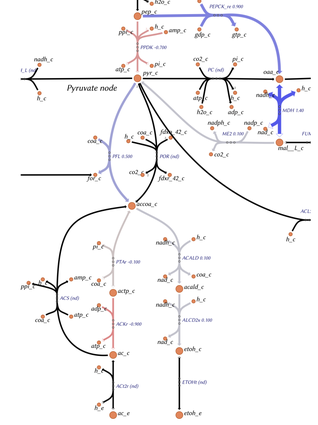
Top downregulated enzymes

| ID | Formula | Fold change *proteomics* | *pFBA* | *FVA center* |
|---|---|---|---|---|
| SULabc | atp_c + h2o_c + so4_e → adp_c + h_c + pi_c + so4_c | -4.6 | -0.3 | 0.8 |
| UMPK | atp_c + h_c + ump_c → adp_c + udp_c | -2.1 | -0.3 | 0.0 |
| CYSS | acser_c + h2s_c → ac_c + cys__L_c | -1.8 | -0.3 | -0.5 |
| UAG4Ei | uacgam_c → udpacgal_c | -1.5 | -0.3 | -0.1 |
| IGPDH | eig3p_c → h2o_c + imacp_c | -1.2 | -0.3 | -0.1 |
| ACGK | acglu_c + atp_c → acg5p_c + adp_c | -1.2 | -0.3 | -0.3 |
| CTPS1 | atp_c + nh4_c + utp_c → adp_c + ctp_c + 2.0 h_c + pi_c | -1.2 | -0.3 | 0.0 |
| FE3abc | atp_c + fe3_e + h2o_c → adp_c + fe3_c + h_c + pi_c | -1.0 | -0.3 | -0.1 |
| NADS2 | atp_c + dnad_c + gln__L_c + h2o_c → amp_c + glu__L_c + 2.0 h_c + nad_c + ppi_c | -0.9 | -0.3 | -0.1 |
| QULNS | dhap_c + iasp_c → 2.0 h2o_c + h_c + pi_c + quln_c | -0.9 | -0.3 | -0.1 |
| ACKr | actp_c + adp_c → ac_c + atp_c | -0.9 | -2.7 | -0.1 |

Top upregulated enzymes

| ID | Formula | Fold change *proteomics* | *pFBA* | *FVA center* |
|---|---|---|---|---|
| RPI | r5p_c ↔ ru5p__D_c | 3.2 | 8.7 | 11.2 |
| PGM | 2pg_c ↔ 3pg_c | 2.4 | 15.2 | 14.5 |
| DHR | dhor__S_c + h2o_c ↔ cbasp_c + h_c | 2.4 | 7.9 | 7.1 |
| GHMT2r | gly_c + h2o_c + mlthf_c ↔ ser__L_c + thf_c | 1.5 | 7.8 | 11.9 |
| MDH | mal__L_c + nad_c ↔ h_c + nadh_c + oaa_c | 1.4 | 15.1 | 14.5 |
| PEPCK_re | co2_c + gdp_c + pep_c → gtp_c + oaa_c | 0.9 | 0.1 | 0.1 |
| TPI | g3p_c ↔ dhap_c | 0.9 | 14.2 | 14.0 |
| VOR2b | 3moh_c + coa_c + 2.0 fdxo_42_c → co2_c + 2.0 fdxr_42_c + h_c + ibcoa_c | 0.9 | 12.3 | 0.8 |
| PFL | coa_c + pyr_c → accoa_c + for_c | 0.5 | 0.2 | 0.8 |
| DCTPD | dctp_c + h2o_c + h_c ↔ dutp_c + nh4_c | 0.5 | 2.3 | 1.5 |
| PSAT | akg_c + pser__L_c ↔ 3php_c + glu__L_c | 0.5 | 8.7 | 15.0 |




- An interactive metabolic map was built from the model using the software *Escher.* This map allows to visualize any from of data that can be mapped to reactions or metabolites. In this case fold change values measured in proteomics data are included.
- Observations regarding the mutant include:
  - PEPCK and the malate shunt are used instead of PPDK to generate pyruvate. This pathway consumes nadh (MDH) and produces napdh (ME2).
  - FRNDPR2r (ferredoxin nadph reductase, NFN) increases, which converts 2 moles of the accumulated reduced ferredoxin into oxidized ferredoxin using 1 mole of nadh and producing 2 moles of nadph.
- Interpretation: The redox imbalance caused by an accumulation of reduced ferredoxin leads to its redox potential being transferred to nadph, likely activating new pathways capable of recycling nadph. This is a previously unobserved mechanism to cope with the redox imbalance caused by the deletion of hydrogenases.

## CONCLUSIONS

- Developed the genome-scale model, iCBI655, that reflects the most current genetic and metabolic knowledge of the consolidated bioprocessing organism *Clostridium thermocellum* DSM1313.
- Collected a comprehensive dataset of extracellular fluxes to train model parameters, leading to increase prediction accuracy across culture conditions with respect to previous models.
- The iCBI655 model correctly predicts lethality of key gene deletion mutants relevant for model-guided design of biocatalytic strains.
- The iCBI655 model emphasizes accessibility and reproducibility through interactive metabolic maps and comprehensive documentation.
- Illustrated the use of the iCBI655 model to gain insights of intracellular fluxes when only external secretion fluxes were measured for wild type and mutant strains. Proteomics data was integrated to gain additional confidence in the simulation.

The Center for Bioenergy Innovation (CBI) is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.