Updated genome-scale metabolic model of *Clostridium thermocellum* with standard-conforming organization and improved prediction accuracy

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

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

ctrinh@utk.edu | https://github.com/trinhlab

Abstract

Motivation Our exponentially growing world population demands a sustainable bioeconomy from renewable and carbon neutral production of energy and materials using lignocellulosic biomass and organic wastes. Consolidated bio-processing (CBP) is a promising technology based on a microorganism capable of 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 (Fig. 1). However, its complex and poorly understood metabolism hinders metabolic engineering to achieve high rates, titers, and yields of industrially relevant chemicals, e.g., alcohols and esters.[1,2]

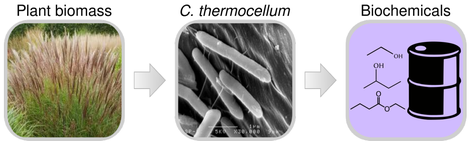

Fig. 1: CBP consists in the direct fermentation of lignocellulosic biomass, removing pretreatment costs that remain a roadblock in biocatalysis technologies.

Approach To unravel the complexity of *C. thermocellum*'s metabolism and enable comprehensive and systematic analysis to drive discovery and strain design, we developed a genome-scale metabolic model using the most recent standards[3] and modeling tools (Fig. 2).[4]

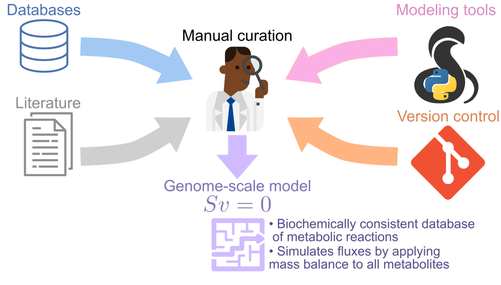

Fig. 2: Genome-scale metabolic modeling involves extensive literature curation in combination with efficient data retrieval and automation tools.

Results 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 published datasets by simulating intracellular fluxes consistent with measured metabolite secretion fluxes and integration of proteomics data.

Methods

Model training

After extensive manual curation of the model (Fig. 2), we gathered a comprehensive flux data set to train model ATP maintenance parameters (Fig. 3a), which demonstrated better growth prediction accuracy under diverse conditions, with respect to the previous model iAT601 (Fig. 3b,c).

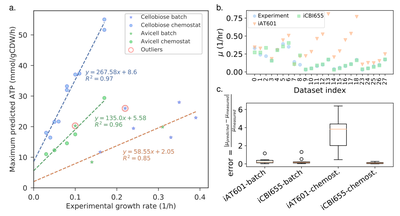

Fig. 3: (**a.**) Training of GAM and NGAM parameters. (**b.**) Comparison of growth prediction error between iCBI655 and iAT601. (**c.**) Error in growth predictions, from **b.**, under batch and chemostat conditions.

Standard-conformance validation with `memote`

We corrected modeling inconsistencies and included extensive metadata in a standard-conforming manner, obtaining a high `memote`[3] score (Fig. 4).

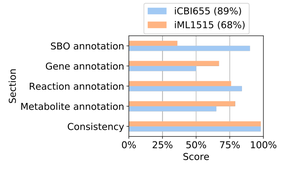

Fig. 4: `memote`[3] scores comparing iCBI655 and iML1515. Overall score noted in legend.

'omics integration

We developed a novel method to make use of proteomics data in combination with a genome-scale model (Fig. 5).

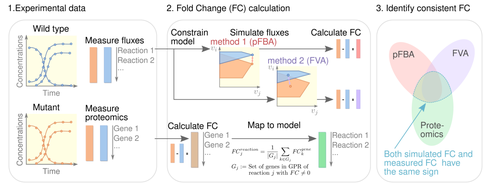

Fig. 5: Procedure to integrate multi-scale into the model. Fold change (FC) is used as an anchor for comparison. FC for all reactions is computed between two conditions using measured fluxes as constraints. Then consistent cases between computed FC and measured FC are identified for further study.

Results

Updated genome-scale model and better phenotype prediction

The iCBI655 model was built starting from the most recently published and comprehensive genome-scale model of *C. thermocellum*, iAT601[5] (Table 1).

	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	[6]	[7]	[5]	This study	[8]

Table 1: Comparison of all genome-scale models of *C. thermocellum* and the latest *E. coli* genome-scale model.

Although genetic manipulation in *C. thermocellum* remains challenging and most studies focus in engineering for overproduction of target compounds, the model was also successfully validated against the few known[1] lethal gene deletions (Table 2.

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	+

Table 2: Comparison of mutant growth rate prediction between iAT601 and iCBI655. To simulate mutant genotypes for growth rate prediction, gene deletions were applied and growth rate was maximized without constraining secretion fluxes to known values, to recreate simulations for strain design were such additional constraints are not available. *In vivo* values are taken form Thompson et al.[1], where growth rate in some cases was not reported, but growth recovery was reported, this is indicated with the "+" symbol.

Integrating extracellular metabolite and proteomics datasets provides novel biological insights

We applied the proteomics integration method (Fig. 5) to compare a wild-type strain with the Δ*hydG*-Δ*ech* deletion mutant, that removes all major hydrogenases (BIF, H2ASE_syn, and ECH) to redirect electrons towards ethanol production. The analysis revealed the following features:

- Redox and hydrogenase metabolism (Fig. 6b): FRNDPR2r (a.k.a. NFN) translation increases to convert one mol of *fdxr_42* and one mol of *nadh* to two moles of *nadph*.
- Pyruvate metabolism (Fig. 6c): PPDK translation decreases while the malate shunt (PEPCK, MDH, ME2) increases. While both pathways convert *pep* to *pyr*, the later converts one mol of *nadh* formed in glycolysis to *nadph*.
- Sulfur metabolism (Fig. 6d): Sulfate reduction diminishes (lower translation of ABC uptake transporter and lower translation of HSOR) to preserve *nadph*.
- Overall, this reveals that the Δ*hydG*-Δ*ech* copes with redox imbalance by increasing *nadph* production which is oxidized in byproduct secreting pathways (e.g., isobutanol).
- Future engineering strategies can be focused in further constraining the undesired pathways that consume *nadph* by directly targeting them or by targeting the sources of *nadph* discovered here.

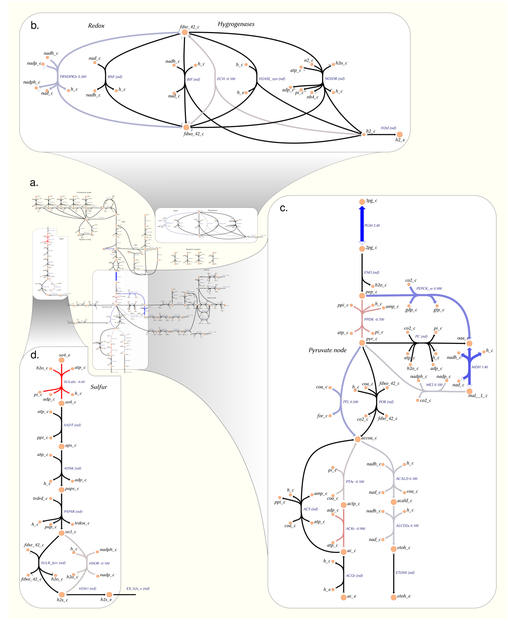

Fig. 6: Consistent reactions (i.e., both simulated flux and proteomics FC have the same sign) are colored according to proteomics FC value, which is also included next to reaction labels. (**a.**) Overall map of central metabolism (**b.**) Redox and hydrogenase metabolism. (**c.**) Pyruvate metabolism. (**d.**) Sulfur metabolism.

References

1. Thompson, R. A. *et al.* Elucidating central metabolic redox obstacles hindering ethanol production in Clostridium thermocellum. *Metabolic engineering* **32,** 207–219 (2015).
2. Thompson, R. A. & Trinh, C. T. Overflow metabolism and growth cessation in Clostridium thermocellum DSM1313 during high cellulose loading fermentations. *Biotechnology and bioengineering* **114,** 2592–2604 (2017).
3. Lieven, C. *et al.* Memote: A community-driven effort towards a standardized genome-scale metabolic model test suite. *BioRxiv*, 350991 (2018).
4. Ebrahim, A. *et al.* COBRApy: constraints-based reconstruction and analysis for python. *BMC systems biology* **7,** 74 (2013).
5. Thompson, R. A. *et al.* Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome. *Biotechnology for biofuels* **9,** 194 (2016).
6. Roberts, S. B. *et al.* Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. *BMC systems biology* **4,** 31 (2010).
7. Dash, S. *et al.* Development of a core Clostridium thermocellum kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnology for biofuels* **10,** 108 (2017).
8. Monk, J. M. *et al.* iML1515, a knowledgebase that computes Escherichia coli traits. *Nature biotechnology* **35,** 904 (2017).

Acknowledgements: This research was funded by the NSF CAREER Award (NSF#1553250) and the Center of Bioenergy Innovation (CBI), U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The funders had no role in the study design, data collection and analysis, or decision to publish.

Poster presented at MPA 2019. Riga, Latvia.