Metabolic Network Modeling of *Clostridium thermocellum* for Systems Biology and Metabolic Engineering

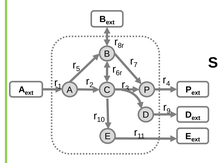
R. ADAM THOMPSON [2,3], SERGIO GARCIA[1,2], SANJEEV DAHAL[2,4], INTAWAT NOOKAEW[2,4], DONOVAN S. LAYTON[1,2], ADAM M. GUSS[2,3,5], DAN G. OLSON[2,6], LEE R. LYND[2,6], and CONG T. TRINH[1,2,3]

[1] Department of Chemical and Biomolecular Engineering, University of Tennessee; Knoxville, Tennessee; [2]BioEnergy Science Center, Oak Ridge, Tennessee; [3]Bredesen Center for Interdisciplinary Research and Education; [4]Comparative Genomics Group, Biosciences Division, Oak Ridge National Laboratory ; [5]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA, [6]Thayer School of Engineering, Dartmouth College, Hanover, NH, USA

Abstract: Clostridium *thermocellum* is a gram-positive thermophile that can directly convert lignocellulosic material into commercially relevant chemicals such as biofuels. Its metabolism contains many branches and redundancies, which limit the production of biofuels at industrially relevant yields and titers. In order to guide the experimental efforts required to overcome these barriers, we built two models of *C. thermocellum* metabolism. Through an extensive literature review, we first constructed a model of the core metabolism of C. *thermocellum*. This model was experimentally validated and served to investigate the range of phenotypes of *C. thermocellum* in response to significant perturbation of energy and redox pathways. The results revealed a complex, robust redox metabolism of *C. thermocellum*. By incorporating experimental data into this core model, we identified redox bottlenecks hindering high-yield ethanol production in *C. thermocellum*. With the recently published sequence of a genetically-tractable strain C. thermocellum DSM 1313, the KEGG database as a scaffold, and further literature review, we expanded the core model into a genome scale model (iAT601). This model constitutes a knowledge base for the organism, including detailed metabolic information, as well as gene protein reaction association. These features allow us to conduct studies on the impact of secondary metabolisms, isozymes, media composition, and provide a more solid basis for computational strain design. We used several sets of experimental data to train the model, e.g., estimation of the ATP requirement for growth-associated maintenance (13.5 mmol ATP/g DCW/hr) and cellulosome synthesis (57 mmol ATP/g cellulosome/hr). Using our tuned model, we predicted the experimentally observed differences in cell biomass yield based on which cellodextrin species is assimilated. We further employed our tuned model to analyze the experimentally quantified differences in fermentation profiles (i.e., the ethanol to acetate ratio) between cellobiose- and cellulose-grown cultures, for which we inferred potential regulatory mechanisms to explain the phenotypic differences. Finally, we used the model to design over 250 genetic modification strategies with the potential to optimize ethanol production, 6,155 for hydrogen production, and 28 for isobutanol production. Our developed genome-scale model iAT601 is capable of accurately predicting complex cellular phenotypes under a variety of conditions by integration of low- and high-throughput data, and serves as a high-quality platform for model-guided strain design to produce industrial biofuels and chemicals of interest.

Theory

Stoichiometric Metabolic Modeling



$S =$

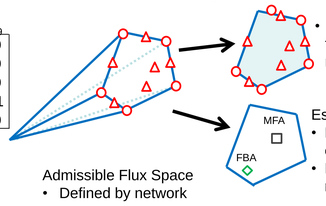
	r_1	r_2	r_3	r_4	r_5	r_{6f}	r_7	r_{8f}	r_9
A	1	-1	0	0	-1	0	0	0	0
B	0	0	0	0	1	-1	-1	-1	0
C	0	1	-1	0	0	1	0	0	0
D	0	0	1	0	0	0	0	0	-1
P	0	0	1	-1	0	0	2	0	0

$\boldsymbol{r} = [r_1\ r_2\ r_3\ r_4\ r_5\ r_{6f}\ r_7\ r_{8f}\ r_9\ r_{10}\ r_{11}]^T$

$\mathbf{S} \cdot \boldsymbol{r} = \mathbf{0}$

$\boldsymbol{r}_{irrev} \geq \mathbf{0}$

Alternative strategies



- Elementary Mode Analysis finds basis pathways to study network properties.

Estimate flux distributions:
- Metabolic Flux Analysis (Best experimental data fit)
- Flux Balance Analysis(Growth rate maximization)

Admissible Flux Space
- Defined by network

Modeling Methods

Flux Balance Analysis

$$\max\ r_{BIO}$$
$$\text{s.t.}$$
$$\sum_j S_{i,j} r_j = 0\ \forall\ \text{metabolites}\ i$$
$$r_j^{min} \leq r_j \leq r_j^{max}\ \forall\ \text{reactions}\ j$$

- Where reactions are bound by r^{min} and r^{max}

Metabolic Flux Analysis

$$\mathbf{r}_{n\times 1} = \mathbf{EM}_{n\times k} \cdot \boldsymbol{w}_{k\times 1}$$

- **r** is the flux distribution vector, **EM** is the elementary mode matrix, and **w** ($\mathbb{R}^+$) is the vector of weighting factors

$$\mathbf{r}_m = \mathbf{EM}_m \cdot \boldsymbol{w}$$
$$\boldsymbol{w} = pinv(\mathbf{EM}_m) \cdot \mathbf{r}_m$$

- Weighting factors can be calculated from the measured fluxes ($\mathbf{r}_m$) and the Moore-Penrose pseudoinverse of $\mathbf{EM}_m$

Metabolic Flux Ratio Analysis

$$\mathbf{MFR}_{i\leftarrow} = \frac{r_{j,i\leftarrow}}{\sum_k r_{k,i\leftarrow}}$$

- **r** is the incoming flux(es) of node *i* through reaction *j*. can be defined similarly where outgoing flux(es) of node *i* is represented with the opposite arrow.
- Allows for quantitative differentiation of flux distributions at key metabolic nodes.

Constrained Minimal Cut Sets

- $\mathbf{T} \cdot \mathbf{r} \leq \mathbf{b}$ Target flux polyhedron. Fluxes below specified product formation rate.
- $\mathbf{D} \cdot \mathbf{r} \leq \mathbf{d}$ Desired flux polyhedron. Fluxes with growth rate above desired minimum.
- Enumerate all reaction knockouts satisfying constraints.

Questions We Can Address:

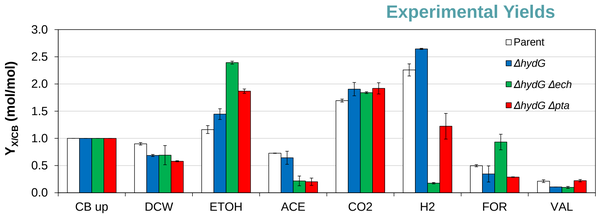
- What media/culturing conditions are optimal for production of target metabolites?
- What genotypes can constrain the phenotypic space of *C. thermocellum* for high-yield biofuel production?
- What understanding can we gain of metabolic and regulatory phenotypes from OMICs data?
- How can we balance thermodynamics, kinetics, and enzyme levels to optimize biofuel production?

Primary References

- Thompson et. al, Biot. Biofuels, under review.
- Thompson et. al, Metab. Eng., 2015.

Core Metabolic Model Capabilities

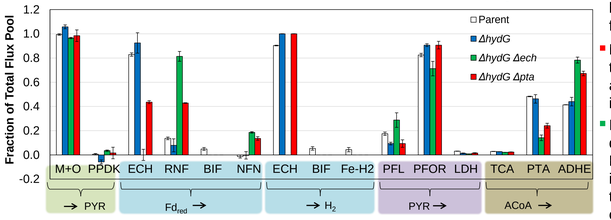
Experimental Yields



- Experimental yields between parent and Δ*hydG* are not significantly altered.
- In Δ*hydG* Δ*pta,* reducing acetate yield forces carbon from acetyl-CoA to ethanol, pulling electrons away from hydrogen production. The lack of ATP generation via acetyl-kinase manifests in a lower DCW yield.
- The significant drop in hydrogen yield in Δ*hydG* Δ*ech* is correlated with a dramatic increase in ethanol yield, as well as an unexpected increase in formate yield.

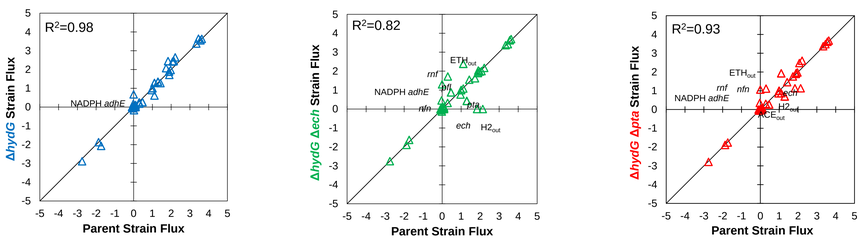
Increase in ethanol yield is more dependent on constrained electron flow than constrained carbon flow.

METAFoR Analysis



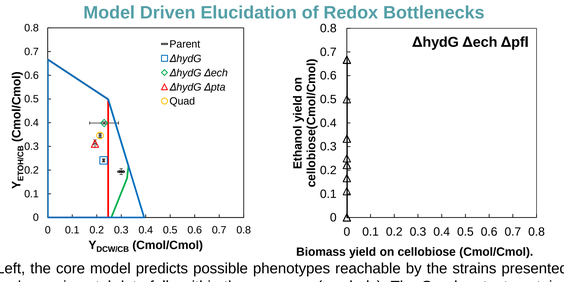
- METAFoR analysis reiterates the robustness of hydrogen production, as ECH can compensate for the lack of hydG related activity.
- METAFoR analysis of Δ*hydG* Δ*pta* shows that the push of carbon to ethanol instead of acetate pulls electrons with it. This is seen in the increase in flux through PFOR, RNF, and NFN.
- In Δ*hydG* Δ*ech,* RNF and NFN fluxes increase dramatically to recycle reduced ferredoxin. However, there is still an apparent redox imbalance that favors diverting pyruvate flux through the redox neutral PFL reaction instead of PFOR.

Alterations of Flux Distribution



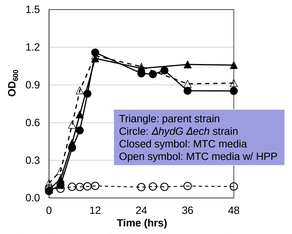
Major outliers are associated with redox metabolism with Δ*hydG* Δ*ech* being the most effective perturbation presented here:

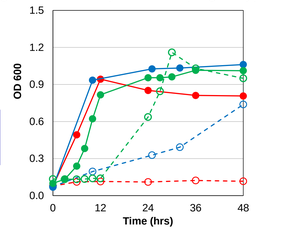
Model Driven Elucidation of Redox Bottlenecks



Biomass yield on cellobiose (Cmol/Cmol).

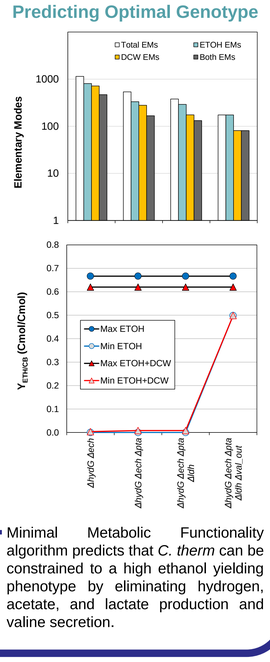
- Left, the core model predicts possible phenotypes reachable by the strains presented (Lines) and experimental data falls within these spaces (symbols). The Quad mutant contains Δ*hydG* Δ*pta* Δ*pfl* Δ*ldh*. Right, the model predicts no growth with a Δ*hydG* Δ*ech* Δ*pfl* genotype.





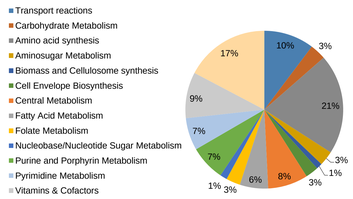
- Left, culturing with the PFL inhibitor hypophosphite (HPP) abolishes growth in the Δ*hydG* Δ*ech* strain as predicted by the model. Right, using exogenous electron sinks restores growth at different efficacies. Fumarate (NADH probe, red) does not restore growth. Isovalerate (NADPH probe, blue) only slightly restores growth. Sulfate (Ferredoxin probe, green) restores growth after a slight lag, indicating Fd to NAD(P)H turnover is the limiting factor. Open circles correspond to HPP addition.

Predicting Optimal Genotype

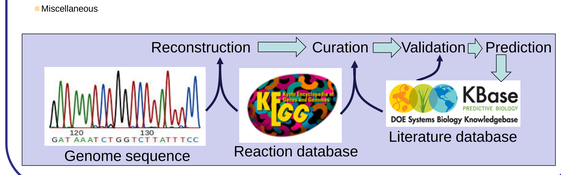


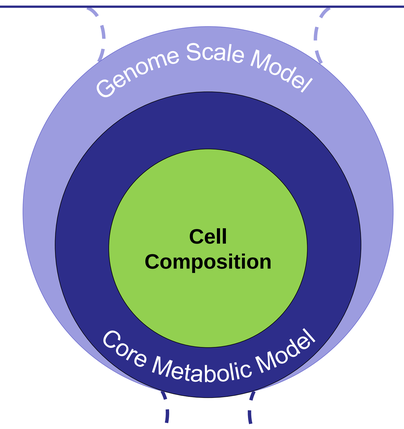
- Minimal Metabolic Functionality algorithm predicts that *C. therm* can be constrained to a high ethanol yielding phenotype by eliminating hydrogen, acetate, and lactate production and valine secretion.

*i*AT601

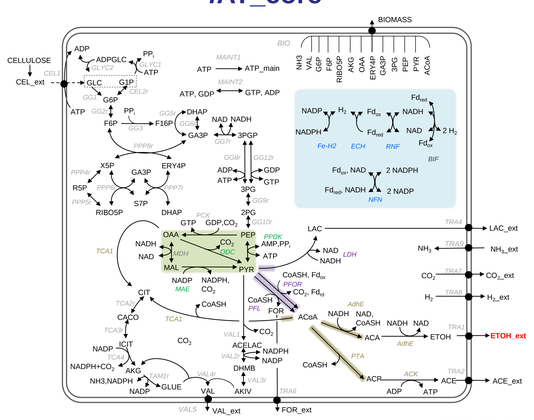


►*The refined genome scale model consists of 871 reactions encoded by 601 genes, spanning multiple KEGG ontology categories* ◄





*i*AT_core



Most Relevant Reaction Abbreviations

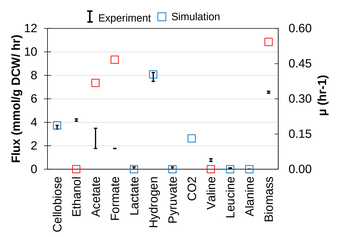
- pyruvate phosphate dikinase (PPDK)
- oxaloacetate decarboxylase (ODC)
- malic enzyme (MAE)
- pyruvate formate lyase (PFL)
- pyruvate ferredoxin oxidoreductase (PFOR)
- lactate dehydrogenase (LDH)
- alcohol dehydrogenase (AdhE)
- phosphotransacetylase (PTA)
- citrate synthase (TCA1)
- Ni-Fe energy conserving hydrogenase (ECH)
- Rnf-type NADH:Fd oxidoreductase (RNF)
- NADH-Fdrd:NADP+ oxidoreductase (NFN)
- NADH-Fdrd bifurcating hydrogenase (BIF)
- Fe-Fe NADPH-depenent hydrogenase (Fe-H2)

Metabolic nodes of interest:
→PYR
PYR→
ACoA→
Fdrd→
→H2

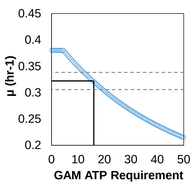
Genome Scale Metabolic Model Capabilities

Finding ATP Requirements: Cellobiose vs Cellulose

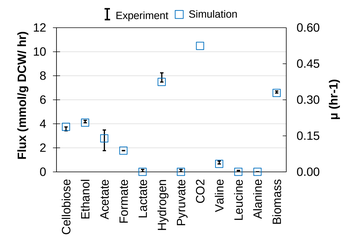
Cellobiose



►*Initially, cellobiose was used as the sole constraint for flux balance analysis, which lead to poor predictions (highlighted in red)* ◄

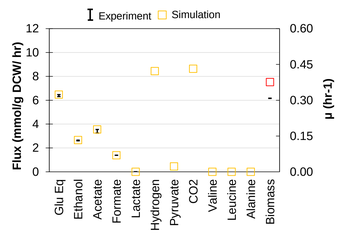


►*Setting fermentation parameters, we sequentially varied the growth associated maintenance term to fit the experimental data* ◄

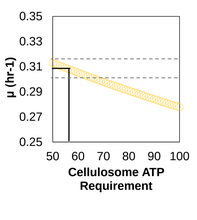


►*Combining experimental constraints with the fitted GAM coefficient gives good agreement to Flux Balance Analysis values* ◄

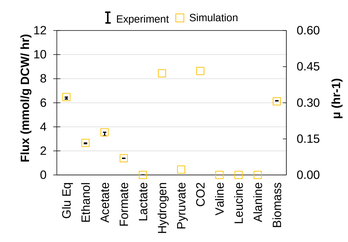
Cellulose



►For simulation of cellulose growth, we set the cellulosome to be 20% of DCW, while setting flux constraints to experimental values. This leads to poor growth prediction (highlighted in red) ◄

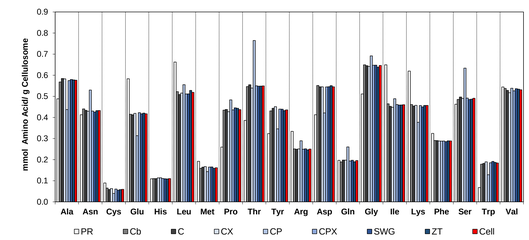


►*In a similar manner to above, we calibrated the ATP requirement for cellulosome production. This value had to be calibrated because secretion and turnover of the cellulosome costs more ATP than normal cellular proteins* ◄



►*Combining experimental fluxes, our tuned GAM coefficient, and the tuned cellulosome coefficient leads to accurate phenotype prediction from the simulations* ◄

Determination of Cellulosome Compositon

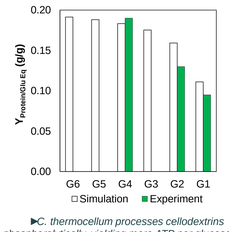


►*To determine cellulosome composition, proteomics data from multiple substrate conditions were compared* ◄

►*Abundance values and protein sequences allowed for the calculation of amino acid requirements, and the median value across conditions was used for our in silico cellulosome component* ◄

►*PR: GEM Dry cell weight protein composition, Cb: Experimental data from growth on cellobiose, C: Experimental growth on cellulose, CX: Cellulose / Xylan, CP: Cellulose / pectin, CPX: Cellulose / Pectin / Xylan, SWG: Switchgrass, ZT: Z-Trim dietary fiber, Cell: median values for application in cellulosome component* ◄

Effect of Cellodextrin Length



►*C. thermocellum processes cellodextrins phosphorolytically, yielding more ATP per glucose unit as the oligomers increase in length* ◄

►Our results simulating protein yield correlate well with literature values ◄

Strain Design Via Constrained Minimal Cut Sets

►*Minimal cut sets are metabolic engineering strategies which seek to maximize product yields while maintaining a minimum growth rate. We set two constraints for all of them, minimum growth rate, and minimum product yield* ◄

Products	Target cut set sizes	# Strain Designs
Ethanol	6	67
	7	185
Hydrogen	4	12
	5	221
	6	1105
	7	4816
Isobutanol	7	28

►*With our genome scale model, we are able to find over 200 strategies for optimizing ethanol production by eliminating 6 or 7 genes* ◄

►*Hydrogen production can be optimized with as little as 4 cuts, while isobutanol needs 7 cuts* ◄

Top 15 knockout set strategies

Increasing number of knockouts

cMCS #	Ethanol: Max Growth rate (1/h)	Ethanol: Product Yield Range at Maximum Growth	cMCS #	Hydrogen: Max Growth rate (1/h)	Hydrogen: Product Yield Range at Maximum Growth	cMCS #	Isobutanol: Max Growth rate (1/h)	Isobutanol: Product Yield Range at Maximum Growth
1	0.0007	1.1533 -1.1535	1	0.0108	2.5361 -3.5824	1	0.0127	0.75912 -0.76201
2	0.0031	1.1483 -1.1494	2	0.0157	2.5248 -2.53	2	0.0005	0.59999 -0.60014
3	0.0007	1.1533 -1.1535	3	0.0138	2.5168 -3.54	3	0.0011	0.80701 -0.80728
4	0.0034	1.277 -1.2779	4	0.0011	2.718 -3.2431	4	0.0011	0.80701 -0.80728
5	0.0062	1.2731 -1.2748	5	0.0136	2.5173 -3.5406	5	0.0011	0.807 -0.80728
6	0.0071	1.2718 -1.2737	6	0.0104	2.3806 -3.5481	6	0.0011	0.807 -0.80728
7	0.0071	1.272 -1.2737	7	0.0104	2.3806 -3.5481	7	0.0038	0.77612 -0.78073
8	0.0077	1.2973 -1.3	8	0.032	3.518 -3.5192	8	0.0037	0.77613 -0.78073
9	0.0077	1.2973 -1.3	9	0.0302	3.5231 -3.5243	9	0.0038	0.77612 -0.78073
10	0.0077	1.2973 -1.3	10	0.0166	3.0709 -3.3436	10	0.0037	0.77613 -0.78073
11	0.0077	1.2973 -1.3	11	0.0034	2.5547 -2.5561	11	0.0005	0.59998 -0.60013
12	0.0056	1.2741 -1.2754	12	0.0089	2.5949 -2.5986	12	0.0078	0.60593 -0.60836
13	0.0035	1.2769 -1.2778	13	0.0225	2.5084 -2.5156	13	0.0008	0.58602 -0.58643
14	0.0269	1.5779 -1.6104	14	0.0166	3.0709 -3.3436	14	0.0127	0.75912 -0.76201
15	0.005	1.215 -1.2158	15	0.0169	2.5219 -2.5274	15	0.0127	0.75912 -0.76201

►*These strategies are excellent starting points for strain design and require more in depth analysis* ◄

Fractional reaction appearance in ethanol cMCSs