Development of a Genome-Scale Metabolic Model of *Clostridium thermocellum* and Its Applications for Integration of Multi-Omics Datasets and Computational Strain Design

Sergio Garcia[1,2], R. Adam Thompson[2,3,4†], Richard J. Giannone[2,5], Satyakam Dash[2,6], Costas D. Maranas[2,6] and Cong T. Trinh[1,2,3,4*]

[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] Bredesen Center for Interdisciplinary Research and Graduate Education, The University of Tennessee, Knoxville, TN, United States, [4] Oak Ridge National Laboratory, Oak Ridge, TN, United States, [5] Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, United States, [6] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA, United States

Solving environmental and social challenges such as climate change requires a shift from our current non-renewable manufacturing model to a sustainable bioeconomy. To lower carbon emissions in the production of fuels and chemicals, plant biomass feedstocks can replace petroleum using microorganisms as biocatalysts. The anaerobic thermophile *Clostridium thermocellum* is a promising bacterium for bioconversion due to its capability to efficiently degrade lignocellulosic biomass. However, the complex metabolism of *C. thermocellum* is not fully understood, hindering metabolic engineering to achieve high titers, rates, and yields of targeted molecules. In this study, we developed an updated genome-scale metabolic model of *C. thermocellum* that accounts for recent metabolic findings, has improved prediction accuracy, and is standard-conformant to ensure easy reproducibility. We illustrated two applications of the developed model. We first formulated a multi-omics integration protocol and used it to understand redox metabolism and potential bottlenecks in biofuel (e.g., ethanol) production in *C. thermocellum*. Second, we used the metabolic model to design modular cells for efficient production of alcohols and esters with broad applications as flavors, fragrances, solvents, and fuels. The proposed designs not only feature intuitive push-and-pull metabolic engineering strategies, but also present novel manipulations around important central metabolic branch-points. We anticipate the developed genome-scale metabolic model will provide a useful tool for system analysis of *C. thermocellum* metabolism to fundamentally understand its physiology and guide metabolic engineering strategies to rapidly generate modular production strains for effective biosynthesis of biofuels and biochemicals from lignocellulosic biomass.

Keywords: *Clostridium thermocellum*, biofuels, genome-scale model, metabolic model, omics integration, modular cell design, ModCell

1. INTRODUCTION

Global oil reserves will be soon depleted (Shafiee and Topal, 2009), and climate change could become a major driver of civil conflict (Hsiang et al., 2011). These challenges to security and the environment need to be addressed by replacing our current non-renewable production of energy and materials for a renewable and carbon neutral approach (Ragauskas et al., 2006). The gram-positive, thermophilic, cellulolytic, strict anaerobe *C. thermocellum* is capable of efficient degradation of lignocellulosic biomass to produce biofuels and biomaterial precursors, making this organism an ideal candidate for consolidated bioprocessing (CBP), where production of lignocellulosic enzymes, saccharification, and fermentation take place in a single step (Olson et al., 2012). However, its complex and poorly understood metabolism remains the main roadblock to achieve industrially competitive titers, rates, and yields of biofuels such as ethanol (Tian et al., 2016) and isobutanol (Lin et al., 2015).

For the past decade, significant efforts have been dedicated to characterize and manipulate the central metabolism of *C. thermocellum*, due to increasing interest in developing this organism as a CBP manufacturing platform for biofuels production (Akinosho et al., 2014). *C. thermocellum* possesses atypical central metabolism, characterized by the important roles of pyrophosphate and ferredoxin (Zhou et al., 2013), which makes redirection of both carbon and electron flows for biofuel production challenging to achieve. Specifically, the metabolic network of *C. thermocellum* contains various reactions to regulate intracellular concentration levels of NADH, NADPH, and reduced ferredoxin. These cofactors are used as electron donors with high specificity throughout metabolism. To maintain redox balance, *C. thermocellum* also possesses several hydrogenases to oxidize these reduced cofactors to molecular hydrogen that is secreted by the cell. Removal of these hydrogenases through deletion of *ech* (encoding the ferredoxin-dependent hydrogenase, ECH) and *hydG* (associated with the bifurcating hydrogenase, BIF, and bidirectional hydrogenase, H2ASE_syn) was successfully applied to increase ethanol yield by electron rerouting (Biswas et al., 2015). Thompson et al. (2015) characterized the $\Delta hydG\Delta ech$ strain in depth by flux analysis of its core metabolism, concluding that the major driver for ethanol production was redox rather than carbon balancing. In particular, the conversion of reduced ferredoxin to NAD(P)H is likely the most rate limiting step. In a subsequent study, Lo et al. (2017) over-expressed *rnf* (encoding the ferredoxin-NAD oxidoreductase, RNF) in the $\Delta hydG\Delta ech$ strain that is expected to enhance NADH supply, but did not achieve improved ethanol yield.

In an attempt to redirect carbon and electron flows for enhanced ethanol production, Deng et al. (2013) manipulated the pyruvate node and malate shunt of *C. thermocellum*. By converting phosphoenolpyruvate (*pep*) to oxaloacetate (*oaa*) and then to pyruvate (*pyr*), this shunt can interchange one mole of NADPH with one mol of NADH generated from glycolysis. Interestingly, the authors noted that replacement of the malate shunt by alternative pathways not linked to NADPH increased

ethanol production and carbon recovery but reduced amino acid formation, confirming the role of the malate shunt as an NADPH source in *C. thermocellum*.

Sulfur metabolism also plays a key role in redox metabolism of *C. thermocellum* and has been investigated for its role in ethanol production. Sulfate, a component of *C. thermocellum* media, serves as an electron acceptor, which is capable of oxidizing sulfate to sulfite and then sulfide. Thompson et al. (2015) demonstrated that the strain $\Delta hydG\Delta ech\Delta pfl$ could not grow in a conventional defined medium due to its inability to secrete hydrogen or formate, but was able to rescue growth by sulfate supplementation to the culture medium. More recently, Biswas et al. (2017) reported an increase in final sulfide concentration and over-expression of the associated sulfate uptake and reduction pathway in the $\Delta hydG$ strain, but did not observe a significant difference in final sulfide concentration in $\Delta hydG\Delta ech$. Remarkably, neither of the strains consumed cysteine from the medium, unlike the wild-type. Sulfide can be converted to cysteine by CYSS (cysteine synthase) or homocysteine and then methionine by SHSL2 (succinyl-homoserine succinate lyase) and METS (methionine synthase), but the connection between the cessation of cysteine uptake and sulfate metabolism remains unclear.

Overall, the complex interactions of *C. thermocellum* metabolic pathways remain challenging to understand and engineer with conventional methods, and hence require a quantitative systems biology approach to decipher. To this end, several genome-scale metabolic models (GSMs) of *C. thermocellum* have been developed. The first GSM, named iSR432, was constructed for the strain ATCC27405 and applied to identify gene deletion strategies for high ethanol yield (Roberts et al., 2010). This model was then further curated into iCth446 (Dash et al., 2017). More recently, Thompson et al. developed the iAT601 genome-scale model (Thompson et al., 2016) for the strain DSM1313, which is genetically tractable (Argyros et al., 2011). The iAT601 model was used to identify genetic manipulations for high ethanol, isobutanol, and hydrogen production (Thompson et al., 2016), and to understand growth cessation prior to substrate depletion observed under high-substrate loading fermentations that simulate industrial conditions (Thompson and Trinh, 2017). In addition to these core and genome-scale steady-state metabolic models, a kinetic model of central metabolism, k-ctherm118, was recently developed and used to elucidate the mechanisms of nitrogen limitation and ethanol stress (Dash et al., 2017). Due to the biotechnological relevance of the *Clostridium* genus, GSMs have also been developed for other species (Dash et al., 2016), including *C. acetobutylicum* (Senger and Papoutsakis, 2008; Salimi et al., 2010; McAnulty et al., 2012; Wallenius et al., 2013; Dash et al., 2014; Yoo et al., 2015; Lee and Trinh, 2019), *C. beijerinckii* (Milne et al., 2011), *C. butyricum* (Serrano-Bermúdez et al., 2017), *C. cellulolyticum* (Salimi et al., 2010), and *C. ljungdahlii* (Nagarajan et al., 2013).

In this study, we developed an updated genome-scale metabolic model of *C. thermocellum*, named iCBI655, with more comprehensive and precise metabolic coverage, enhanced prediction accuracy, and extensive documentation. This model

is a human-curated database that coherently represents all the available genetic, genomic, and metabolic knowledge of *C. thermocellum* from both experimental literature and bioinformatic predictions. Furthermore, the model can be applied not only to enable metabolic flux simulation but also to provide a framework to contextualize disparate datasets at the system level. As a demonstration for the model application, we first developed a quantitative multi-omics integration protocol and used it to fundamentally study redox metabolism and potential redox bottlenecks critical for production of biofuels (e.g., ethanol) in *C. thermocellum*. Furthermore, we used the model, in combination with the previously developed ModCell tool (Garcia and Trinh, 2019c), to design modular (chassis) cells (Garcia and Trinh, 2019b) for alcohol and ester production.

2. RESULTS

2.1. Development of an Upgraded *C. thermocellum* Genome-Scale Model Named iCBI655

The iCBI655 model was developed using the published iAT601 model (Thompson et al., 2016) as a starting point. The model improvements include updated metabolic pathways, new annotation, and new extensive documentation. A detailed account of these changes can be found in the **Supplementary Datasheet 1**. Here, we highlight the most relevant modifications.

2.1.1. Modeling Updates

To facilitate model usage and reduce human error, the identifiers of reactions and metabolites were converted from KEGG into BiGG human-readable form (King et al., 2015). Additionally, reaction and metabolite identifiers were linked to the modelSEED database (Henry et al., 2010) that enables analysis through the KBase web interface (Arkin et al., 2018). The gene identifiers and functional descriptions were updated to the most current annotation (NCBI Reference Sequence: NC_017304.1). Metabolite formulas and charges from the modelSEED database (Henry et al., 2010) were included in the model and reactions were systematically corrected for charge and mass balance by the addition of protons and water.

2.1.2. Metabolic Updates

The automated construction process used in the previous model introduced several inconsistencies that were corrected in the current model. We removed reactions that were blocked and non-gene-associated, apparently introduced during automated gap-filling. Two notable examples are (i) the blocked selenate pathway which lacks experimental evidence (e.g., selenoproteins have not been found in *C. thermocellum*), and (ii) blocked reactions involving molecular oxygen (e.g., oxidation of Fe^{2+} to Fe^{3+}) that are not possible in strict anaerobes like *C. thermocellum*. Furthermore, tRNA cycling reactions were unblocked by including tRNA in the biomass reaction (Reimers et al., 2017). Metabolite isomers were examined and consolidated under the same metabolite identifier when possible, leading to the removal of duplicated reactions and the elimination of gaps.

Transport and exchange reactions were updated to reflect the export of amino acids and uptake of pyruvate as observed during fermentation experiments (Holwerda et al., 2014).

In terms of specific reactions, oxaloacetate decarboxylase was eliminated from the model in accordance with recent findings (Olson et al., 2017). The stoichiometries of pentose-phospate reactions, including sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase (FBA3) and sedoheptulose 1,7-bisphosphate ppi-dependent phosphofructokinase (PFK3_ppi), were corrected (according to experimental evidence, Rydzak et al., 2012) from the previous model by ensuring mass balance and avoiding lumping multiple steps into one reaction. Transaldolase (TALA) was removed from the model due to lack of annotation for this gene in *C. thermocellum*.

Several modifications were also performed in key bioenergetic reactions. The reactions catalyzed by membrane-bound enzymes, including inorganic diphosphatase (PPA) (Zhou et al., 2013) and membrane-bound ferredoxin-dependent hydrogenase (ECH) (Calusinska et al., 2010), were corrected to capture proton translocation. Furthermore, hydrogenase reactions were updated to ensure ferredoxin association for all cases and remove those reactions that do not involve ferredoxin and only use NAD(P)H as a cofactor, based on our recent understanding of *C. thermocellum* metabolism (Biswas et al., 2017). Gene-protein-reaction associations were updated to represent experimental knowledge. For instance, the hydrogenases BIF (CLO11313_RS09060-09070) and H2ASE (CLO1313_RS12830, CLO1313_RS02840) require the maturase Hyd (CLO1313_RS07925, CLO1313_RS11095, CLO1313_RS12830) to be functional, and the maturase itself requires all of its subunits to operate, which enables accurate representations of *hydG* deletion genotypes (Biswas et al., 2015).

Two hypothetical reaction modifications were introduced to ensure consistency with reported phenotypes. First, to enable growth without the need for succinate secretion, as observed in experimental data (**Supplementary Datasheet 2**), the reaction homoserine-O-trans-acetylase (HSERTA) was added to enable methionine biosynthesis (essential for growth). Although this reaction is not currently known to be associated with any gene in *C. thermocellum*, it is present as a gene-associated reaction in other *Clostridium* GSMs (Nagarajan et al., 2013). Next, the reaction deoxyribose-phosphate aldolase (DRPA) was removed based on a systematic analysis (section 4.4) to ensure correct lethality prediction of the $\Delta hydG\Delta ech\Delta pfl$ mutant strain as well as the correct prediction of growth recovery in this mutant by addition of external electron sinks such as sulfate or ketoisovalerate (**Table 1**). The correct prediction of $\Delta hydG\Delta ech\Delta pfl$-associated phenotypes is critical to successfully use the model for computational strain design (Long et al., 2015; Ng et al., 2015; Maranas and Zomorrodi, 2016; Wang and Maranas, 2018; Garcia and Trinh, 2019a,b,c).

2.2. Comparison of iCBI655 Against Other Genome-Scale Models

We compared iCBI655 with the previous GSMs of *C. thermocellum* and the highly-curated GSM iML1515 of

TABLE 1 | Comparison of mutant growth rates predicted by iAT601 and iCBI655.

Gene deletions	Medium	Percent of W.T. growth rate (%)		
		iAT601	*iCBI655*	Experiment
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	+

Experimental values are taken from Thompson et al. (2015); for some mutants whose growth recovery, not growth rate, was reported, they are presented with "+". W.T., wildtype; MTC, Medium for Thermophilic Clostridia.

TABLE 2 | Comparison of all genome-scale metabolic models of *C. thermocellum* and the latest *E. coli* model.

	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	Roberts et al., 2010	Dash et al., 2017	Thompson et al., 2016	This study	Monk et al., 2017

the extensively studied bacterium *Escherichia coli* (**Table 2**). The increased number of genes in iCBI655 with respect to iAT601 cover a variety of functions, including hydrogenase chaperones, cellulosome and cellulase, ATP synthase, and transporters. Remarkably, iCBI655 has a smaller percentage of blocked reactions than iAT601, indicating higher biochemical consistency. The number of metabolites in iCBI655 is smaller than those in iAT601 mainly due to the removal of metabolites that did not appear in any reaction, duplicated metabolites (e.g., certain isomers), and blocked pathways added automatically during gap-filling without any gene association. *C. thermocellum* DSM1313 has 2911 protein coding genes, 22% of which is captured by iCB655, while *E. coli* MG1655 has 4240 genes, 35% of which is included in iML1515. Overall, iCBI655 has the increased coverage of the metabolic functionality of *C. thermocellum* but remains far from the well-studied *E. coli*.

2.3. Training of Model Parameters Under Diverse Conditions

Growth and non-growth associated maintenance (GAM and NGAM) are parameters that capture the consumption of ATP toward cell division and homeostasis, respectively. These are known to be condition-specific; however, genome-scale models do not include a mechanistic description that allows to determine these ATP consumption rates as part of the simulation. Instead, GAM is incorporated into the biomass pseudo-reaction and

NGAM has its own pseudo-reaction that hydrolyzes ATP at a rate tuned by the constraint parameters.
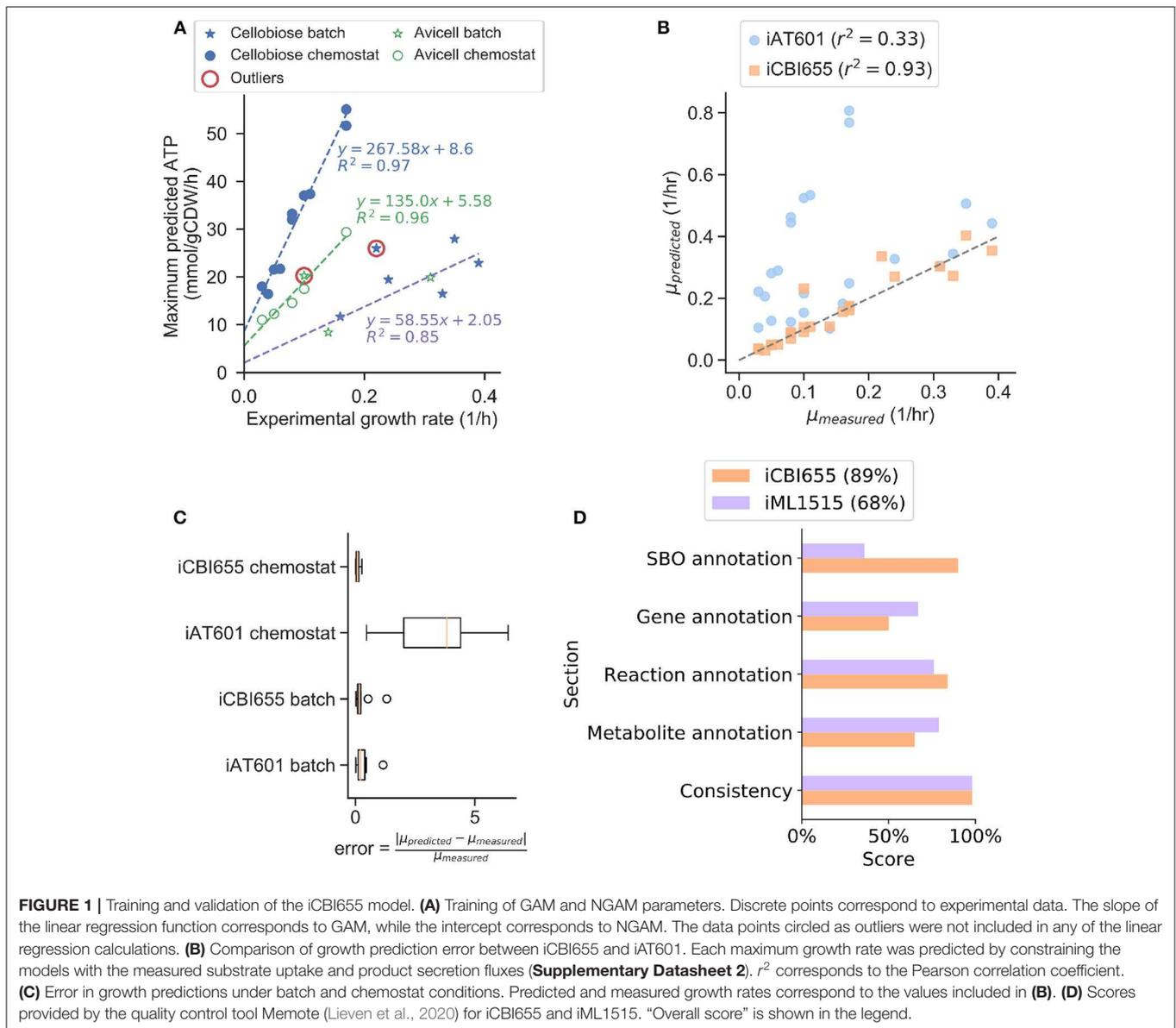
To increase model prediction accuracy for various conditions, we trained GAM and NGAM parameters of iCBI655 using an extensive dataset of 28 extracellular fluxes (**Supplementary Datasheet 2**) measured during the growth phase under different reactor configurations, carbon sources, and gene deletion mutants. This approach is based on the method used to train the iML1655 *E. coli* model (Monk et al., 2017). Remarkably, we observed highly linear trends under three different conditions, including chemostat reactor with cellobiose as a carbon source, chemostat reactor with cellulose as a carbon source, and batch reactor with either cellobiose or cellulose as a carbon source (**Figure 1A**). This model training has led to improved growth rate prediction of iCBI655 as compared to iAT601 that has previously been trained with only a smaller dataset (**Figure 1B**). Specifically, the iAT601 training dataset was limited to batch conditions; hence, the inaccurate predictions of iAT601 were observed for chemostat conditions (**Figure 1C**).

2.4. Assessment of Model Quality and Standard Compliance With Memote

The field of metabolic network modeling suffers from a lack of standard enforcement and quality control metrics that limit model reproducibility and applicability. To address this issue, Lieven et al. (2020) recently developed the Memote framework that systematically tests for standards and best practices in GSMs. We applied Memote to both the iCBI655 and *E. coli* iML1515 models for comparison (**Figure 1D**). This analysis produced five independent scores that assess model quality. The *consistency score* measures basic biochemical requirements, such as mass and charge balance of metabolic reactions, and it was near 100% for both models. Additionally, the different annotation scores quantify how many elements in the model contain relevant metadata. More specifically, the *systems biology ontology (SBO) annotation* indicates if an object in the model refers to a metabolite, reaction, or gene, while the respective *annotation scores* of these elements correspond to properties (e.g., name, chemical formula, etc.) and identifiers linking them to relevant databases (e.g., KEGG Kanehisa and Goto, 2000 or modelSEED Henry et al., 2010). The *overall score* is computed as a weighted average of all the individual scores with additional emphasis on the *consistency score*. In summary, the high scores obtained by iCBI655 indicate the quality of the model and ensure its applicability for future studies.

2.5. Model-Guided Analysis of Proteomics and Flux Datasets Sheds Light on Redox Metabolism Critical for Biofuel Production in *C. thermocellum*

For the first application of the genome-scale metabolic model, we aimed to understand the complex redox metabolism and potential redox bottlenecks critical for enhanced biofuel production in *C. thermocellum*. We used the model as a scaffold to analyze proteomics and metabolic flux data collected for the *C. thermocellum* wild-type and Δ*hydG*Δ*ech* strains. The

FIGURE 1 | Training and validation of the iCBI655 model. **(A)** Training of GAM and NGAM parameters. Discrete points correspond to experimental data. The slope of the linear regression function corresponds to GAM, while the intercept corresponds to NGAM. The data points circled as outliers were not included in any of the linear regression calculations. **(B)** Comparison of growth prediction error between iCBI655 and iAT601. Each maximum growth rate was predicted by constraining the models with the measured substrate uptake and product secretion fluxes (**Supplementary Datasheet 2**). r^2 corresponds to the Pearson correlation coefficient. **(C)** Error in growth predictions under batch and chemostat conditions. Predicted and measured growth rates correspond to the values included in **(B)**. **(D)** Scores provided by the quality control tool Memote (Lieven et al., 2020) for iCBI655 and iML1515. "Overall score" is shown in the legend.

$\Delta hydG\Delta ech$ mutant was engineered to redirect electron flow from hydrogen to ethanol by removal of primary hydrogenases (Biswas et al., 2015; Thompson et al., 2015). Previous studies of $\Delta hydG\Delta ech$ based on analysis of secretion fluxes (Thompson et al., 2015) or omics (Biswas et al., 2017) suggested the presence of redox bottlenecks in this mutant but did not identify which specific pathways and cofactors (i.e., NADH vs. NADPH) are responsible. We aim to solve this problem through integrated and quantitative analysis of omics and fluxes at the genome scale.
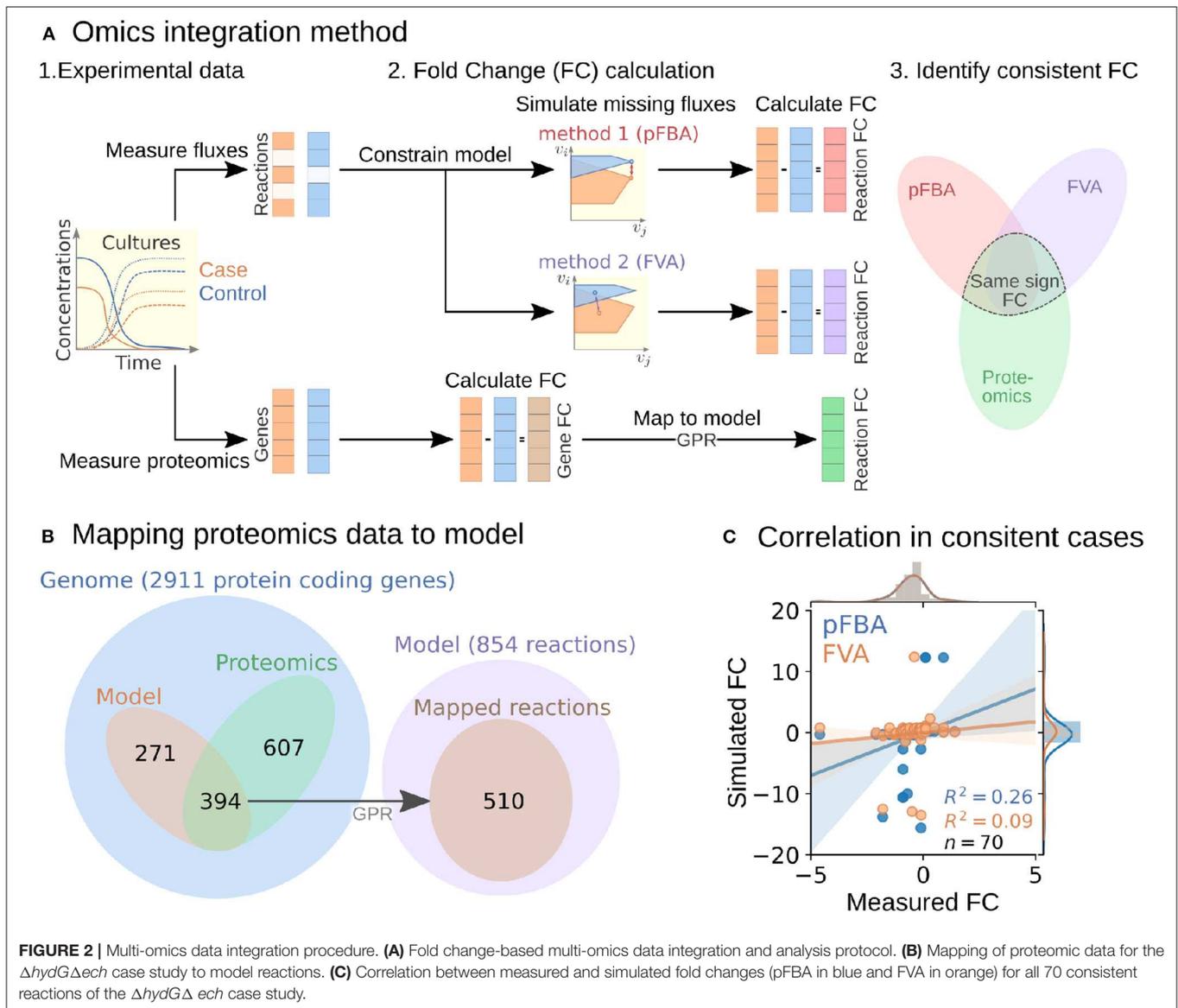
2.5.1. Development of Fold Change-Based Omics Integration Protocol

To perform the analysis, we formulated an omics integration protocol anchored at the quantification of fold change (FC) between case and control samples (**Figure 2A**). In this approach, we first compared FCs between simulated intracellular fluxes and measured omics data. Next, we identified *consistent reactions* with FCs of the same sign and different from zero in both measured proteomics and simulated fluxes for further analysis (section 4.6).

To start the FC-based omics integration, we obtained measured FCs by mapping the measured proteomics data to 510 out of the 856 reactions in the model through the gene-protein-reaction (GPR) associations (**Figure 2B**). Then, we identified 70 consistent reactions by comparing measured FCs with two types of simulated FCs: (i) parsimonious flux balance analysis (pFBA) that determines the most efficient flux distribution (assuming all enzymes are roughly as efficient) and (ii) flux variability analysis (FVA) that identifies the feasible flux range of each reaction.
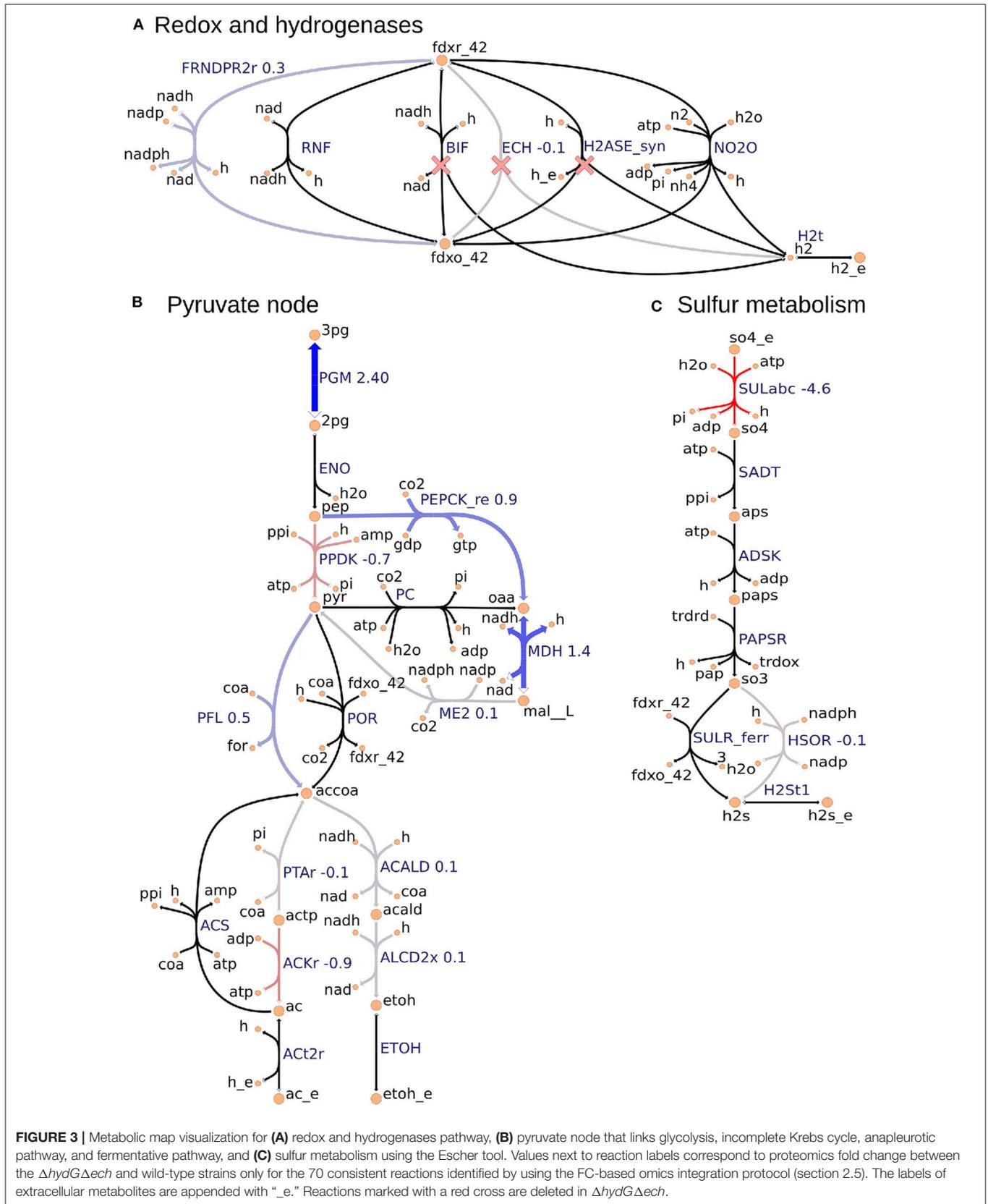
The Pearson correlation coefficients between simulated and measured FCs for the consistent reactions were 0.26 and 0.09 for pFBA and FVA, respectively (**Figure 2C**). In general,

FIGURE 2 | Multi-omics data integration procedure. **(A)** Fold change-based multi-omics data integration and analysis protocol. **(B)** Mapping of proteomic data for the Δ*hydG*Δ*ech* case study to model reactions. **(C)** Correlation between measured and simulated fold changes (pFBA in blue and FVA in orange) for all 70 consistent reactions of the Δ*hydG*Δ*ech* case study.

the FVA reaction flux ranges remained mostly unchanged, suggesting that pFBA is a better representation of actual metabolic fluxes as previously observed (Machado and Herrgård, 2014). The top consistent reactions with the highest proteomics FCs (**Table S1**) belong primarily to the central metabolism of *C. thermocellum* (**Figure 3**). Interestingly, discrepancies in magnitude between flux and protein FCs for consistent reactions could be used to identify bottlenecks. For example, for a given enzyme, a small increase in flux combined with a large increase in translation could be an indicator of low catalytic efficiency; alternatively, such discrepancy could also point at an upstream thermodynamic bottleneck. Similar comparisons between simulated flux and omics has previously been used to identify regulatory mechanisms (Bordel et al., 2010). Overall, the identification and analysis of consistent reactions is an effective approach to gain certainty on the activity changes of metabolic pathways between conditions.

2.5.2. FC-Based Omics Integration Reveals Redirection of Electron Flow for NADPH Supply in Δ*hydG*Δ*ech* Strain

Our identification of the consistent reactions by using the FC-based omics integration protocol revealed coherent indications of increased NADPH biosynthesis in the Δ*hydG*Δ*ech* mutant with respect to the wild-type across three major metabolic areas: (i) an increased protein level of FRNDPR2r (also known as NFN) that converts one mol of reduced ferredoxin (*fdxr_42*) and one mole of NADH into two moles of NADPH (**Figure 3A**), (ii) an increased protein level of all three malate shunt enzymes and a decreased protein level of the alternative route PPDK (**Figure 3B**), and (iii) a decreased protein level of sulfur transporter and of HSOR that oxidizes sulfite into sulfide consuming NADPH (**Figure 3C**). These observations are consistent with the failure of *rnf* over-expression to enhance ethanol production (Lo et al., 2017), since RNF

FIGURE 3 | Metabolic map visualization for **(A)** redox and hydrogenases pathway, **(B)** pyruvate node that links glycolysis, incomplete Krebs cycle, anapleurotic pathway, and fermentative pathway, and **(C)** sulfur metabolism using the Escher tool. Values next to reaction labels correspond to proteomics fold change between the Δ*hydG*Δ*ech* and wild-type strains only for the 70 consistent reactions identified by using the FC-based omics integration protocol (section 2.5). The labels of extracellular metabolites are appended with "_e." Reactions marked with a red cross are deleted in Δ*hydG*Δ*ech*.

produces NADH but the key cofactor bottleneck seems to be NADPH. Furthermore, a direct look at the proteomics data revealed that RNF subunits (Clo1313_0061-Clo1313_0066) had a statistically significant decrease in protein levels of the mutant (**Supplementary Datasheet 2**). The preference of $\Delta hydG\Delta ech$ toward NADPH could be due to the cofactor specificity of the remaining redox balancing pathways (e.g., isobutanol), thermodynamics and protein cost constraints, or a combination of both. While the contribution of thermodynamic constraints is beyond the scope of this study, a recent analysis (Dash et al., 2019) of the ethanol production pathway in *C. thermocellum* highlighted the importance of engineering strategies to increase NADPH, for instance, introduction of NADPH-linked GAPDH that converts glyceraldehyde-3-phosphate to 3-phospho-D-glyceroyl phosphate in glycolysis or overexpression of NADPH-FNOR that transfers electrons from reduced ferredoxin to NADPH. The contribution of alternative redox balancing pathway toward the increased NADPH biosynthesis will be examined next.

2.5.3. Analysis of Simulated Fluxes Reveals the Role of NADPH in Redox Balancing

The analysis based on consistent reactions strongly indicates that NADPH production is important in the $\Delta hydG\Delta ech$ mutant to achieve redox balance. However, the pathways oxidizing NADPH remain unknown since not all reactions in the model could be mapped to proteomics measurements and carbon recovery was lower in the mutant strain (Thompson et al., 2015). To identify these pathways, we examined the simulated fluxes of all reactions (instead of only consistent reactions) that differed in value between wild-type and mutant, and limited this search to exchange reactions and reactions that involve NADPH (**Table S2**). These simulated fluxes predicted an increase in the isobutanol pathway, including keto-acid reductoisomerase (KARA1) that consumes NADPH and isobutanol secretion (EX_ibutoh_e). The isobutanol pathway can consume NADPH through several enzymes (Lin et al., 2015) and has increased flux during overflow metabolism at high-substrate loading (Holwerda et al., 2014; Thompson and Trinh, 2017). The model also predicted a decrease in valine secretion (EX_val__L_e), since the isobutanol pathway competes with the valine pathway after KARA1. Remarkably, this prediction is consistent with the lower valine secretion measured in $\Delta hydG\Delta ech$ (Biswas et al., 2017). A certain amount of NADPH is likely oxidized by the mutated alcohol-dehydrogenase enzyme observed after short adaptation in $\Delta hydG$ that is compatible with both NADH and NADPH (Biswas et al., 2015). However, this feature is not captured by the model since in general gene knock-outs are simulated by blocking the associated reactions. Overall this analysis indicates that $\Delta hydG\Delta ech$ likely increases isobutanol secretion to alleviate redox imbalance.
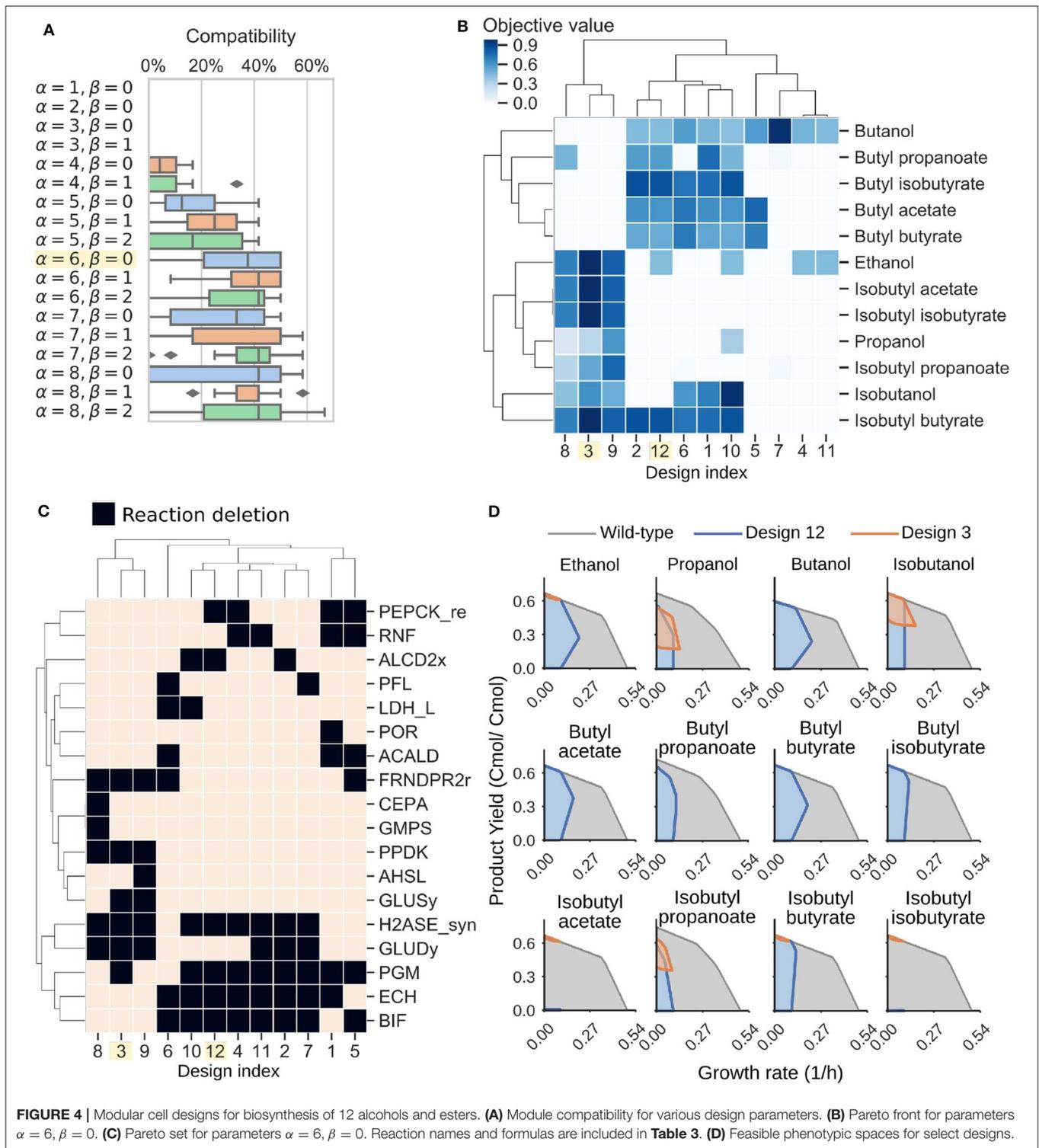
Taken altogether, model-guided data analysis illustrates the power of the model as contextualization tool and provides new insights into the redox bottlenecks present in *C. thermocellum* that are critical in the production of reduced molecules. The integration of omics and fluxes led to the resolution of NADPH as the key cofactor in redox bottleneck of $\Delta hydG\Delta ech$. It

helped identify specific pathways that undergo major changes in protein levels, providing interesting target reactions for further engineering. Generally, the developed FC-based omics integration protocol can be applied to different omics data types due to its simplicity. The method does not require one to formulate or assume a quantitative relationship between omics measurements and simulated fluxes. Furthermore, fold change in biomolecule concentrations implemented in the method is currently much easier to measure in a quantitatively reliable manner for many molecules than case-specific absolute concentrations.

2.6. Model-Guided Design of Modular Production Strains for Biofuel Synthesis

Another common application of genome-scale models is strain design (Long et al., 2015; Ng et al., 2015; Maranas and Zomorrodi, 2016; Wang and Maranas, 2018; Garcia and Trinh, 2019a,b,c). We used the iCBI655 model combined with the ModCell tool (section 4.9) to design *C. thermocellum* modular production strains for efficient biosynthesis of alcohols and esters. Briefly, with ModCell we aim to design a modular (chassis) cell that can be rapidly combined with exchangeable pathway modules in a plug-and-play fashion to obtain modular production strains exhibiting target phenotypes with minimal strain optimization cycles (Trinh, 2012; Trinh et al., 2015; Garcia and Trinh, 2019b,c). In this study, the target phenotype for modular production strains is growth-coupled to product synthesis (*wGCP*), that corresponds to the minimum product synthesis rate at the maximum growth rate. The ModCell mathematical formulation, computational algorithm, and implementation were described in details previously (Garcia and Trinh, 2019a,c, 2020). The design variables to attain the target phenotypes involve genetic manipulations of two types: (i) reaction deletions, constrained by the parameter α, that corresponds to gene knock-outs; and (ii) module reactions, constrained by the parameter β, that corresponds to reactions deleted in the modular cell but added back to specific modules to enhance the compatibility of the modular cell. Once these two parameters are specified, the solution to the problem is a set of Pareto optimal designs named Pareto front. In a Pareto optimal design, the performance (i.e., objective value) of a given module can only be increased at the expense of lowering the performance of another module. To characterize the practicality of each design, we say a modular cell is compatible with certain modules if the design objective is above a specific threshold (e.g., 0.5 in this study). Hence, the *compatibility* of a design corresponds to the number of compatible modules.

To design *C. thermocellum* modular cells, we first evaluated a range of design parameters α and β with an increasing number of genetic manipulations (**Figure 4A**). As expected, increasing the number of deletions leads to more compatible designs, at the expense of more complexity in the implementation. We selected an intermediate point of $\alpha = 6$, $\beta = 0$ for further analysis. This Pareto front is composed of 12 designs that can be clustered into two groups (**Figure 4B**). The first group (e.g., designs 3, 8, and 9) are compatible with all products except butanol and its derived

FIGURE 4 | Modular cell designs for biosynthesis of 12 alcohols and esters. **(A)** Module compatibility for various design parameters. **(B)** Pareto front for parameters $\alpha = 6, \beta = 0$. **(C)** Pareto set for parameters $\alpha = 6, \beta = 0$. Reaction names and formulas are included in **Table 3**. **(D)** Feasible phenotypic spaces for select designs.

esters, whereas the second group (e.g., designs 1, 2, 10, and 12) have high objective values for butanol and its derived esters.

To understand the characteristics of each group, we can inspect the deletions of each design (**Figure 4C**, **Table 3**). Designs 3, 8, and 9 all have in common H2ASE_syn, GLUDy, PPDK, and

FRNDPR2r deletion, while the last two deletions never appear in design 1, 2, 10, or 12. The majority of deletion targets are central metabolic reactions (**Table 3**). The common targets include deletion of hydrogenases that appear in the cluster of designs 2, 4, 7, 10, 11, and 12 with the $\Delta hydG \Delta ech$ genotype discussed

TABLE 3 | Reaction deletions sorted by appearance frequency (counts) in the designs of the Pareto front for $\alpha = 6$, $\beta = 0$.

ID	Name	Formula	Counts (%)
PGM	Phosphoglycerate mutase	2pg_c ↔ 3pg_c	75
H2ASE_syn	Bidirectional [NiFe] Hydrogenase (Fe-H2)	h2_c + nadp_c ↔ h_c + nadph_c	75
ECH	(FeFe)-hydrogenase, ferredoxin dependent, membrane-bound	2.0 fdxr_42_c + 3.0 h_c ↔ 2.0 fdxo_42_c + h2_c + h_e	66.7
BIF	Bifurcating Hydrogenase	2.0 fdxr_42_c + 3.0 h_c + nadh_c ↔ 2.0 fdxo_42_c + 2.0 h2_c + nad_c	66.7
GLUDy	Glutamate dehydrogenase (NADP)	glu__L_c + h2o_c + nadp_c ↔ akg_c + h_c + nadph_c + nh4_c	50
FRNDPR2r	Ferredoxin: nadp reductase (NFN)	2.0 fdxr_42_c + h_c + nadh_c + 2.0 nadp_c ↔ 2.0 fdxo_42_c + nad_c + 2.0 nadph_c	41.7
RNF	Ferredoxin:NAD oxidoreductase (membrane bound)	2.0 fdxr_42_c + 2.0 h_c + nad_c ↔ 2.0 fdxo_42_c + h_e + nadh_c	33.3
PEPCK_re	Phosphoenolpyruvate carboxykinase (GTP)	co2_c + gdp_c + pep_c → gtp_c + oaa_c	33.3
ALCD2x	Alcohol dehydrogenase (ethanol)	acald_c + h_c + nadh_c → etoh_c + nad_c	25
ACALD	Acetaldehyde dehydrogenase (acetylating)	accoa_c + h_c + nadh_c → acald_c + coa_c + nad_c	25
PPDK	Pyruvate, phosphate dikinase	amp_c + 2.0 h_c + pep_c + ppi_c → atp_c + pi_c + pyr_c	25
GLUSy	Glutamate synthase (NADPH)	akg_c + gln__L_c + h_c + nadph_c → 2.0 glu__L_c + nadp_c	16.7
PFL	Pyruvate formate lyase	coa_c + pyr_c → accoa_c + for_c	16.7
LDH_L	L-lactate dehydrogenase	h_c + nadh_c + pyr_c → lac__L_c + nad_c	16.7
POR	Pyruvate-ferredoxin oxidoreductase	coa_c + 2.0 fdxo_42_c + pyr_c → accoa_c + co2_c + 2.0 fdxr_42_c + h_c	8.3
CEPA	Cellobiose phosphorylase	cellb_c + pi_c → g1p_c + glc__D_c	8.3
GMPS	GMP synthase	atp_c + nh4_c + xmp_c → amp_c + gmp_c + 3.0 h_c + ppi_c	8.3
AHSL	O-Acetyl-L-homoserine succinate-lyase	achms_c + cys__L_c ↔ ac_c + cyst_L_c + h_c	8.3

earlier or removal of reactions that form fermentative byproducts such as ALCD2x and ACALD (ethanol), PFL (formate), LDH_L (lactate). Interestingly, ACKr or PTA (acetate) does not appear in this list, likely because acetate production can serve as a regulatory valve for redox metabolism, especially in a modular cell that must be compatible with products of diverse degrees of reduction.

More interestingly, we also found important branch-point deletion reactions (Stephanopoulos and Vallino, 1991) in central metabolism that have not yet been explored for strain design. Most prominently, these reactions include GLUDy, PEPCK_re, and PPDK, which appear with percentage frequencies of 50%, 33.3%, and 25%, respectively (**Table 3**). Both PEPCK_re and PPDK present two alternative routes that influence the ratio of NADPH to NADH, which is relevant to control metabolic fluxes though the specific dependencies of certain enzymes toward each redox cofactor. Since GLUDy consumes NADPH and is a key reaction in amino-acid metabolism, this enzyme and related ones (e.g., GLUSy) are interesting targets for practical implementation. We speculate the two product groups emerge likely because the butanol pathway strictly depends on NADH due to the reactions ACOAD1z (acyl-CoA dehydrogenase) and HACD1 (3-hydroxyacyl-CoA dehydrogenase), while the ethanol, propanol, and isobutanol pathways are more flexible in their use of NADH or NADPH. The designs 3, 8, and 9 perform poorly with butanol, and are also the only ones containing PPDK deletion. This deletion forces pep to pyruvate flux through the malate shunt that converts NADH to NADPH. Engineering of the cofactor specificities of the butanol pathway can be used to build one modular cell compatible with all products under consideration.

Two representative designs from the groups mentioned earlier are 3 and 12. Their feasible growth and production phenotypes reveal a tight coupling between product formation and growth rate (**Figure 4D**). This phenotype enables pathway optimization through adaptive laboratory evolution, as previously done for ethanol (Tian et al., 2016), overcoming one of the main challenges of *C. thermocellum* engineering that is optimization of enzyme expression levels. Hence, the proposed modular cells can also serve as platforms for pathway selection and optimization. In summary, this analysis demonstrates the potential of the model to identify non-intuitive metabolic engineering strategies that can be key to build effective modular platform strains for the production of biofuels and biochemicals in *C. thermocellum*.

3. CONCLUSIONS

In this study, we developed a genome-scale metabolic model of the biotechnologically relevant organism *C. thermocellum*. Model development followed standards and best practices to ensure reproducibility and accessibility. We demonstrated the enhanced predictions of the model for diverse fermentation conditions and gene lethality. Genome-scale models have a broad range of applications in systems biology, including metabolic engineering, physiological discovery, phenotype interpretation, and studies of evolutionary processes (Feist and Palsson, 2008; Palsson, 2015). To illustrate the model applications, we chose to tackle the challenge of disparate data integration and interpretation at the systems level. We developed a fold-change-based omics integration method for this purpose, and used it to identify routes in central metabolism that were selected to increase NADPH generation in the $\Delta hydG\Delta ech$ strain. This analysis revealed the importance of NADPH cofactor over its alternatives and provided new engineering targets for enhanced biosynthesis of

reduced products in *C. thermocellum*. We also illustrated the use of the model to design *C. thermocellum* modular cells, using the ModCell tool (Garcia and Trinh, 2019b). The proposed designs cover C2 through C4 alcohols and their derived esters, which are key target molecules for renewable production with *C. thermocellum* (Peters, 2018). The proposed designs feature a combination of previously-explored and novel strategies to couple target metabolite production with cellular growth. Like the well-developed genome-scale models (Monk et al., 2017; Lu et al., 2019) of the important organisms *Escherichia coli* and *Saccharomyces cerevisiae* broadly used for strain engineering both in academia (Blazeck and Alper, 2010) and industry (Yim et al., 2011), we anticipate the iCBI655 genome-scale model will also provide a versatile tool for systems metabolic engineering of *C. thermocellum*.

4. METHODS

4.1. Model Curation

The genome scale model iCBI655 was constructed from iAT601 (Thompson et al., 2016) by following the standard GSM development protocol (Thiele and Palsson, 2010). Reaction and metabolite identifiers were mapped from KEGG to BiGG using the BiGG API (King et al., 2015). Metabolite charges were obtained from modelSEED when available, and otherwise calculated using the Chemaxon pKa plugin (Szegezdi and Csizmadia, 2007) for a pH of 7.2 (Thiele and Palsson, 2010). The biomass objective function was consolidated into one pseudo-reaction avoiding the use of intermediate pseudo-metabolites present in iAT601. Reactions were assigned with a confidence level based on a standard genome-scale model annotations (Thiele and Palsson, 2010).

4.2. Metabolic Flux Simulations

Constraint-based metabolic network modeling (Palsson, 2015) is based on the feasible flux space, Ω_k, defined by network stoichiometry and flux bounds that represent thermodynamic constraints and measured values:

$$\Omega_k := \{v_{jk} \in \mathbb{R} :$$

$$\sum_{j \in \mathcal{J}} S_{ij} v_{jk} = 0 \ \forall i \in \mathcal{I} \tag{1}$$

$$l_{jk} \leq v_{jk} \leq u_{jk} \ \forall j \in \mathcal{J}\} \tag{2}$$

Here $\mathcal{I}$ and $\mathcal{J}$ are the sets of metabolites and reactions in the model, respectively, and v_{jk} is the metabolic flux (mmol/gCDW/h) through reaction j in the simulation condition k. Constraint (1) enforces mass balance for all metabolites in the network, where S_{ij} represents the stoichiometric coefficient of metabolite i in reaction j. Constraint (2) enforces lower and upper bounds l_{jk} and u_{jk}, respectively, for each reaction j in the network.

In different simulation conditions, k, S_{ij} remains fixed given the structure of the network for all $i, j \in \mathcal{I}, \mathcal{J}$. However, certain bounds u_{jk} and l_{jk} are modified to represent specific metabolic constraints. For example, to apply measured reaction fluxes such as in the case of GAM and NGAM calculation or the

omics integration protocol (section 4.6), l_{jk} and u_{jk} are specified using the experimentally measured average (μ_{jk}) and standard deviation (σ_{jk}), which for normally distributed samples with 3 replicates produces an interval with a confidence level above 90% (3-4). Similarly, to represent a certain gene deletion mutant k, the bounds are set to be $u_{jk} = l_{jk} = 0$ for the associated reaction j.

$$l_{jk} = \mu_{jk} - \sigma_{jk} \ \forall j \in \text{Measured}_k \tag{3}$$

$$u_{jk} = \mu_{jk} + \sigma_{jk} \ \forall j \in \text{Measured}_k \tag{4}$$

The feasible flux space Ω_k can be explored in different ways; (Trinh et al., 2009; Palsson, 2015) for instance, an optimization objective is often defined to identify specific flux distributions $v_{jk}^{\text{sim}} \ \forall j \in \mathcal{J}$:

$$v_{jk}^{\text{sim}} \in \arg\max \left\{ \sum_{j \in \mathcal{J}} c_j v_{jk} : v_{jk} \in \Omega_k \right\} \forall j \in \mathcal{J} \tag{5}$$

Here c_j is the coefficient of reaction j in the linear objective function, which is changed according to the simulation context. For example, to train GAM and NGAM (**Figure 1A**) the objective was set to maximize flux through the ATP hydrolysis reaction, i.e., $c_j = 1$ for j corresponding to ATP hydrolysis reaction, and 0 otherwise. To evaluate growth prediction accuracy (**Figures 1B,C**), the objective was set to maximize growth, i.e., $c_j = 1$ for j corresponding to growth pseudo-reaction and 0 otherwise.

4.3. Simulation of Different Growth Environments

The model is configured to generally represent different medium and reactor conditions by modifying three features. The first feature involves model boundaries specifying which metabolites may enter the intracellular environment (i.e., present in the growth medium) or may exit the intracellular environment (i.e., secreted by *C. thermocellum*). This feature can be adjusted through u_{jk} and l_{jk} for exchange reactions. In our simulations, only essential metabolites required for *in silico* growth may be consumed and only commonly observed metabolites may be produced, unless otherwise noted. The second feature involves biomass objective function. iCBI655 contains 3 possible biomass reactions: (i) BIOMASS_CELLOBIOSE used for growth in cellobiose with cellulosan constituting 2% of cell dry weight (CDW) (Zhang and Lynd, 2005), (ii) BIOMASS_CELLULOSE used for growth on cellulose with cellulosan constituting 20% of CDW (Zhang and Lynd, 2005), and (iii) BIOMASS_NO_CELLULOSOME, a biomass function that does not consider cellulosan production and only used as a control. The combination of cellulosome and protein fractions accounts for 52.85% of the CDW in all cases (Roberts et al., 2010; Thompson et al., 2015). Cellobiose conditions were used in all simulations unless otherwise noted. The third feature involves GAM/NGAM. Three sets of these parameters are considered including batch, chemostat-cellulose, and chemostat-cellobiose, based on fitting the model to experimental data. Batch conditions were used in all simulations unless otherwise noted.

For growth on cellulose, the experimentally measured glucose-equivalent uptake was represented in the model through the following pseudo-reactions: *3 glceq_e → cell3_e; 4 glceq_e → cell4_e; 5 glceq_e → cell5_e;* and *6 glceq_e → cell6_e.* Here, *cell3_e, cell4_e, cell5_e,* and *cell6_e* are cellodextrin polymers with 3, 4, 5, and 6 glucose monomers, respectively. These polymers can be imported inside the cell through the oligo-cellulose transport ABC system. The model is free to use any cellodextrin length, although utilization of longer cellodextrins results in higher ATP yield (Zhang and Lynd, 2005; Thompson et al., 2016).

4.4. Single-Reaction Deletion Analysis to Match Experimentally Observed Phenotype

A core model of *C. thermocellum* (Thompson et al., 2015) correctly predicted the experimentally observed lethality of $\Delta hydG\Delta ech\Delta pfl$; however, the iAT601 genome-scale model built by extension of this core model failed, suggesting that the genome-scale model has alternative active pathways leading to the false growth prediction *in silico*. To resolve this false positive prediction in iCBI655, we calculated the maximum growth rates for all possible additional single reaction deletions in the $\Delta hydG\Delta ech\Delta pfl$ mutant. This analysis resulted in three possible additional reaction deletions that are predicted to be lethal (i.e., maximum growth rate prediction below 20% of the simulated wild-type value Palsson, 2015), including the removal of (i) glycine secretion (EX_gly_e), (ii) 5,10-methylenetetrahydrofolate oxidoreductase (MTHFC), and (iii) deoxyribose-phosphate aldolase (DRPA). For the first removal, addition of sulfate or ketoisovalerate in the growth medium of $\Delta hydG\Delta ech\Delta pfl$ fails to predict growth recovery as observed experimentally (Thompson et al., 2015), making this option invalid. Likewise, the second removal is invalid because it makes PFL an essential reaction in the wild-type strain; however, experimental evidence demonstrates that Δpfl mutant is able to grow (Papanek et al., 2015). The last option was chosen since it correctly predicts growth recovery of $\Delta hydG\Delta ech\Delta pfl$ by sulfate or ketoisovalerate addition in the growth medium, and does not make PFL essential in the wild-type strain.

4.5. Model Comparison

The *C. thermocellum* and *E. coli* models were obtained from their respective publications in SBML format. Blocked reactions were calculated by allowing all exchange reactions to have an unconstrained flux (i.e., $lb_{jk} = -1{,}000$, $ub_{jk} = 1{,}000$ $\forall j \in Exchange$). This procedure enables the most general scenario which produces the smallest number of blocked reactions in each model. Additional details can be found in **Supplementary Datasheet 1**.

4.6. Omics Integration Protocol

The omics integration protocol developed in this study consists of three steps: (i) simulation of fold changes, (ii) mapping of measured gene fold changes to reactions, and (iii) comparison of measured and simulated fold changes.

4.6.1. Calculation of Simulated Fold Changes

To simulate metabolic fluxes, lower and upper bounds (2) are constrained according to experimental data as described in section 4.2. Then, for the pFBA method, a quadratic optimization problem (6) is solved, leading to a unique flux distribution v_{jk}^{pFBA} $\forall j \in \mathcal{J}$.

$$v_{jk}^{\mathrm{pFBA}} \in \arg\min \left\{ \sum_{j \in \mathcal{J}} v_{jk}^2 : v_{jk} \in \Omega_k \right\} \forall j \in \mathcal{J} \qquad (6)$$

For the FVA method, a sequence of linear programming problems is solved where each flux is minimized (7) and maximized (8):

$$v_{jk}^{\min} \in \arg\min \left\{ v_{jk} : v_{jk} \in \Omega_k \right\} \ \forall j \in \mathcal{J} \qquad (7)$$

$$v_{jk}^{\max} \in \arg\max \left\{ v_{jk} : v_{jk} \in \Omega_k \right\} \ \forall j \in \mathcal{J} \qquad (8)$$

Note that for computation we applied the loop-less FVA method (Schellenberger et al., 2011; Chan et al., 2018), as implemented in cobrapy (Ebrahim et al., 2013), that introduces additional constraints in Ω_k to remove thermodynamically infeasible cycles from all feasible flux distributions.

FVA produces a flux range $[v_{jk}^{\min}, v_{jk}^{\max}]$ for each reaction $j \in \mathcal{J}$. To compare between states k (e.g., wild-type and mutant), we define the *FVA center*, a scalar variable that generally indicates a change in this range (9).

$$v_{jk}^{\mathrm{FVA}} = \frac{v_{jk}^{\max} + v_{jk}^{\min}}{2} \qquad (9)$$

The FVA center is a heuristic analysis with the main purpose of determining whether a reaction exhibits an upward shift (center increase) or a downward shift (center decrease) between two conditions k. It should be emphasized that the FVA center, v_{jk}^{FVA}, does not attempt to quantify the fraction of overlap between ranges nor to identify what type of shift might occur from all possible permutations. Unlike v_{jk}^{pFBA}, v_{jk}^{FVA} does not necessarily represent a feasible flux distribution of Ω_k. Furthermore, the FVA center could potentially fail to capture hypothetical permutations of fluxes. Despite these considerations, the FVA center remains a useful heuristic to analyze simulated fold changes.

Finally, to determine the fold change for either pFBA or FVA simulated fluxes, the conventional procedure for fold change calculation in omics data is emulated. First, values are floored to avoid very large (or infinite) fold changes in cases with very small magnitude change. This is accomplished through a flooring piece-wise function (10), where $\epsilon = 0.0001$ is the minimum value and x is an arbitrary scalar variable.

$$floor(x) = \begin{cases} x + \epsilon & \text{if } 0 < x < \epsilon \\ x - \epsilon & \text{if } 0 > x > -\epsilon \\ x & \text{otherwise} \end{cases} \qquad (10)$$

Then, the fluxes are normalized to the substrate uptake rate $v_{\mathrm{uptake},k}$ and fold change is calculated in $\log_2$ space (11).

$$FC_j^{\mathrm{sim}}(v_{j,\mathrm{mut}}, v_{j,\mathrm{wt}}) = \log_2\left[floor\left(\frac{v_{j,\mathrm{mut}}}{|v_{\mathrm{uptake,mut}}|}\right)\right]$$
$$- \log_2\left[floor\left(\frac{v_{j,\mathrm{wt}}}{|v_{\mathrm{uptake,wt}}|}\right)\right] \quad (11)$$

4.6.2. Calculation of Measured Fold Changes
Fold change between case and control samples, FC_l, is calculated in $\log_2$ space for each gene $l \in \mathcal{L}$, where $\mathcal{L}$ is the set of genes in the model. These gene fold changes can be mapped to metabolic reaction fold changes using the gene-protein reaction associations (GPR), given $\mathcal{G}_j$ as the set of genes with $FC_l \neq 0$ in the GPR of reaction j:

$$FC_j^{\mathrm{meas}} = \frac{1}{\mathrm{card}(\mathcal{G}_j)} \sum_{l \in \mathcal{G}_j} FC_l \quad (12)$$

4.6.3. Identification of Consistent Fold Changes
A reaction j is said to have a consistent fold change if the measured fold change has the same sign of at least one of the simulated fold changes, more formally:

$$\mathcal{M} := \left\{ j \in \mathcal{J} : \left(\left[(FC_j^{\mathrm{sim,pFBA}} < 0) \vee (FC_j^{\mathrm{sim,FVA}} < 0)\right] \wedge (FC_j^{\mathrm{meas}} < 0) \right) \right.$$
$$\left. \vee \left(\left[(FC_j^{\mathrm{sim,pFBA}} > 0) \vee (FC_j^{\mathrm{sim,FVA}} > 0)\right] \wedge (FC_j^{\mathrm{meas}} > 0) \right) \right\}$$
$$(13)$$

where $\mathcal{M} \subseteq \mathcal{J}$ is the set of consistent reactions which is considered for further analysis and the simulated fold changes are re-defined for brevity (14-15).

$$FC_j^{\mathrm{sim,pFBA}} := FC_j^{\mathrm{sim}}(v_{j,\mathrm{mut}}^{\mathrm{pFBA}}, v_{j,\mathrm{wt}}^{\mathrm{pFBA}}) \quad (14)$$
$$FC_j^{\mathrm{sim,FVA}} := FC_j^{\mathrm{sim}}(v_{j,\mathrm{mut}}^{\mathrm{FVA}}, v_{j,\mathrm{wt}}^{\mathrm{FVA}}) \quad (15)$$

4.7. Software Implementation
Model development was performed using Python and Jupyter notebooks with open-source Python libraries including cobrapy (Ebrahim et al., 2016). The sequence of upgrades and improvements can be seen in the Git version control records. The repository is available online through Github (https://github.com/trinhlab/ctherm-gem) and in **Supplementary Datasheet 1**.

4.8. Proteomics Data Collection
C. thermocellum wild-type and $\Delta hydG\Delta ech$ strains were cultured in batch reactors and metabolic fluxes were calculated as previously described (Thompson et al., 2015). For proteomics measurements, the wild-type and mutant strains were cultured in MNM and MTC media (Kridelbaugh et al., 2013), respectively. While both wild-type and mutant were originally cultured in MTC (Thompson et al., 2015), the wild-type had to be cultured separately in MNM medium due to insufficient volume for proteomics sampling in the MTC culture. MTC has higher

nitrogen and trace mineral concentrations, but previous studies have shown no effect on growth rates (Kridelbaugh et al., 2013). During the mid-exponential growth phase 10 mL samples were collected, centrifuged, and the resulting pellet was stored at $-20\,°C$. Cell pellets were then prepared for LC-MS/MS-based proteomic analysis. Briefly, proteins extracted via SDS, boiling, and sonic disruption were precipitated with trichloroacetic acid (Giannone et al., 2015b). The precipitated protein was then resolubilized in urea and treated with dithiothreitol and iodoacetamide to reduce and block disulfide bonds prior to digestion with sequencing-grade trypsin (Sigma-Aldrich). Following two-rounds of proteolysis, tryptic peptides were salted, acidified, and filtered through a 10 kDa MWCO spin column (Vivaspin 2; GE Healthcare) and quantified by BCA assay (Pierce).

For each LC-MS/MS run, 25 µg of peptides were loaded via pressure cell onto a biphasic MudPIT column for online 2D HPLC separation and concurrent analysis via nanospray MS/MS using a LTQ-Orbitrap XL mass spectrometer (Thermo Scientific) operating in data-dependent acquisition (one full scan at 15 k resolution followed by 10 MS/MS scans in the LTQ, all one µscan; monoisotopic precursor selection; rejection of analytes with an undecipherable charge; dynamic exclusion = 30 s) (Giannone et al., 2015a).

Eleven salt cuts (25, 30, 35, 40, 45, 50, 65, 80, 100, 175, and 500 mM ammonium acetate) were performed per sample run with each followed by 120 min organic gradient to separate peptides.

Resultant peptide fragmentation spectra (MS/MS) were searched against the *C. thermocellum* DSM1313 proteome database concatenated with common contaminants and reversed sequences to control false-discovery rates using MyriMatch v.2.1. (Tabb et al., 2007). Peptide spectrum matches (PSM) were filtered by IDPicker v.3 (Ma et al., 2009) to achieve a peptide-level FDR of <1 % per sample run and assigned matched-ion intensities (MIT) based on observed peptide fragment peaks. PSM MITs were summed on a per-peptide basis and those uniquely mapping to their respective proteins were imported into InfernoRDN (Taverner et al., 2012). Peptide intensities were log2-transformed, normalized across replicates by LOESS, standardized by median absolute deviation, and median centered across all samples. Peptide abundance data were then assembled to proteins via RRollup and further filtered to maintain at least two values in at least one replicate set. Protein abundances were then used for the modeling efforts describe herein.

All raw and database-searched LC-MS/MS data pertaining to this study have been deposited into the MassIVE proteomic data repository and have been assigned the following accession numbers: MSV000084488 (MassIVE) and PXD015973 (ProteomeXchange). Data files are available upon publication (ftp://massive.ucsd.edu/MSV000084488/).

4.9. Modular Cell Design
The ModCell formulation, computational algorithm, and implementation followed the previous reports (Garcia and Trinh, 2019a,c, 2020). The iCBI655 model with cellobiose as a carbon source in the batch reactors (**Supplementary Datasheet 4**) was used as an input for modular cell design. The alcohol pathways

were curated from recent literature (Holwerda et al., 2014; Lin et al., 2015; Loder et al., 2015), where adapted Adh can use either NADH or NADPH as an electron donor to synthesize the target alcohol (Biswas et al., 2015). The esters-producing pathways require an alcohol acetyltransferase (AAT) reaction to condense an alcohol and acyl-CoA that are already present in the alcohols-producing pathways. Even though a thermostable AAT has not yet been reported in literature to function at high temperature, an engineered chloramphenicol acetyl transferase (CAT) can be repurposed as a thermostable AAT (Seo et al., 2019, 2020). The ModCell software is available online at https://github.com/TrinhLab/ModCell2.

DATA AVAILABILITY STATEMENT

All raw and database-searched LC-MS/MS data pertaining to this study have been deposited into the MassIVE proteomic data repository and have been assigned the following accession numbers: MSV000084488 (MassIVE) and PXD015973 (ProteomeXchange). Data files are available upon publication (ftp://massive.ucsd.edu/MSV000084488/).

AUTHOR CONTRIBUTIONS

CT managed the project. SG and CT conceived the study, designed experiments, and analyzed the data. SG, RT, RG, and SD performed the experiments. SG prepared the draft with co-authors' inputs. All read, wrote, and approved the final draft.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00772/full#supplementary-material

Supplementary Datasheet 1 | Software used to develop, configure, and analyze iCBI655.

Supplementary Datasheet 2 | Flux dataset used to train the iCBI655 model and proteomics dataset for the wild-type and Δ*hydG*Δ*ech* strains.

Supplementary Datasheet 3 | Supplementary tables.

Supplementary Datasheet 4 | iCBI655 model in various formats for cellobiose growth conditions and map of central metabolic pathways in Escher format.

REFERENCES

Akinosho, H., Yee, K., Close, D., and Ragauskas, A. (2014). The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Front. Chem.* 2:66. doi: 10.3389/fchem.2014.00066

Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R., Feinberg, L. F., Olson, D. G., et al. (2011). High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Appl. Environ. Microbiol.* 77, 8288–8294. doi: 10.1128/AEM.00646-11

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: the United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163

Biswas, R., Wilson, C. M., Giannone, R. J., Klingeman, D. M., Rydzak, T., Shah, M. B., et al. (2017). Improved growth rate in *Clostridium thermocellum* hydrogenase mutant via perturbed sulfur metabolism. *Biotechnol. Biofuels* 10:6. doi: 10.1186/s13068-016-0684-x

Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., and Guss, A. M. (2015). Elimination of hydrogenase active site assembly blocks h 2 production and increases ethanol yield in *Clostridium thermocellum*. *Biotechnol. Biofuels* 8:20. doi: 10.1186/s13068-015-0204-4

Blazeck, J., and Alper, H. (2010). Systems metabolic engineering: genome-scale models and beyond. *Biotechnol. J.* 5, 647–659. doi: 10.1002/biot.200900247

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.* 6:e1000859. doi: 10.1371/journal.pcbi.1000859

Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* 156, 1575–1588. doi: 10.1099/mic.0.032771-0

Chan, S. H., Wang, L., Dash, S., and Maranas, C. D. (2018). Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *Bioinformatics* 34, 4248–4255. doi: 10.1093/bioinformatics/bty446

Dash, S., Khodayari, A., Zhou, J., Holwerda, E. K., Olson, D. G., Lynd, L. R., et al. (2017). Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnol. Biofuels* 10:108. doi: 10.1186/s13068-017-0792-2

Dash, S., Mueller, T. J., Venkataramanan, K. P., Papoutsakis, E. T., and Maranas, C. D. (2014). Capturing the response of *Clostridium acetobutylicum* to chemical stressors using a regulated genome-scale metabolic model. *Biotechnol. Biofuels* 7:144. doi: 10.1186/s13068-014-0144-4

Dash, S., Ng, C. Y., and Maranas, C. D. (2016). Metabolic modeling of clostridia: current developments and applications. *FEMS Microbiol. Lett.* 363, 1–10. doi: 10.1093/femsle/fnw004

Dash, S., Olson, D. G., Chan, S. H. J., Amador-Noguez, D., Lynd, L. R., and Maranas, C. D. (2019). Thermodynamic analysis of the pathway for ethanol production from cellobiose in *Clostridium thermocellum*. *Metab. Eng.* 55, 161–169. doi: 10.1016/j.ymben.2019.06.006

Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Shaw, A. J., and Lynd, L. R. (2013). Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. *Metab. Eng.* 15, 151–158. doi: 10.1016/j.ymben.2012.11.006

Ebrahim, A., Brunk, E., Tan, J., O'brien, E. J., Kim, D., Szubin, R., et al. (2016). Multi-omic data integration enables discovery of hidden biological regularities. *Nat. Commun.* 7:13091. doi: 10.1038/ncomms13091

Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). Cobrapy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74

Feist, A. M., and Palsson, B. Ø. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat. Biotechnol.* 26:659. doi: 10.1038/nbt1401

Garcia, S., and Trinh, C. T. (2019a). Comparison of multi-objective evolutionary algorithms to solve the modular cell design problem for novel biocatalysis. *Processes* 7:361. doi: 10.3390/pr7060361

Garcia, S., and Trinh, C. T. (2019b). Modular design: implementing proven engineering principles in biotechnology. *Biotechnol. Adv.* 37:107403. doi: 10.1016/j.biotechadv.2019.06.002

Garcia, S., and Trinh, C. T. (2019c). Multiobjective strain design: a framework for modular cell engineering. *Metab. Eng.* 51, 110–120. doi: 10.1016/j.ymben.2018.09.003

Garcia, S., and Trinh, C. T. (2020). Harnessing natural modularity of cellular metabolism to design a modular chassis cell for a diverse class of products by using goal attainment optimization. *ACS Synth. Biol.* 9, 1665–1681. doi: 10.1021/acssynbio.9b00518

Giannone, R. J., Wurch, L. L., Heimerl, T., Martin, S., Yang, Z., Huber, H., et al. (2015a). Life on the edge: functional genomic response of *Ignicoccus hospitalis* to the presence of *Nanoarchaeum equitans*. *ISME J.* 9:101. doi: 10.1038/ismej.2014.112

Giannone, R. J., Wurch, L. L., Podar, M., and Hettich, R. L. (2015b). Rescuing those left behind: recovering and characterizing underdigested membrane and hydrophobic proteins to enhance proteome measurement depth. *Anal. Chem.* 87, 7720–7728. doi: 10.1021/acs.analchem.5b01187

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28:977. doi: 10.1038/nbt.1672

Holwerda, E. K., Thorne, P. G., Olson, D. G., Amador-Noguez, D., Engle, N. L., Tschaplinski, T. J., et al. (2014). The exometabolome of *Clostridium thermocellum* reveals overflow metabolism at high cellulose loading. *Biotechnol. Biofuels* 7:155. doi: 10.1186/s13068-014-0155-1

Hsiang, S. M., Meng, K. C., and Cane, M. A. (2011). Civil conflicts are associated with the global climate. *Nature* 476:438. doi: 10.1038/nature10311

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2015). Bigg models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44, D515-D522. doi: 10.1093/nar/gkv1049

Kridelbaugh, D. M., Nelson, J., Engle, N. L., Tschaplinski, T. J., and Graham, D. E. (2013). Nitrogen and sulfur requirements for *Clostridium thermocellum* and *Caldicellulosiruptor bescii* on cellulosic substrates in minimal nutrient media. *Bioresour. Technol.* 130, 125–135. doi: 10.1016/j.biortech.2012.12.006

Lee, J.-W., and Trinh, C. T. (2019). Microbial biosynthesis of lactate esters. *Biotechnol. Biofuels* 12:226. doi: 10.1186/s13068-019-1563-z

Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., et al. (2020). Memote for standardized genome-scale metabolic model testing. *Nat. Biotechnol.* 38, 272–276. doi: 10.1038/s41587-020-0446-y

Lin, P. P., Mi, L., Morioka, A. H., Yoshino, K. M., Konishi, S., Xu, S. C., et al. (2015). Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metab. Eng.* 31, 44–52. doi: 10.1016/j.ymben.2015.07.001

Lo, J., Olson, D. G., Murphy, S. J.-L., Tian, L., Hon, S., Lanahan, A., et al. (2017). Engineering electron metabolism to increase ethanol production in *Clostridium thermocellum*. *Metab. Eng.* 39, 71–79. doi: 10.1016/j.ymben.2016.10.018

Loder, A. J., Zeldes, B. M., Garrison, G. D., Lipscomb, G. L., Adams, M. W., and Kelly, R. M. (2015). Alcohol selectivity in a synthetic thermophilic n-butanol pathway is driven by biocatalytic and thermostability characteristics of constituent enzymes. *Appl. Environ. Microbiol.* 81, 7187–7200. doi: 10.1128/AEM.02028-15

Long, M. R., Ong, W. K., and Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* 34, 135–141. doi: 10.1016/j.copbio.2014.12.019

Lu, H., Li, F., Sánchez, B. J., Zhu, Z., Li, G., Domenzain, I., et al. (2019). A consensus *S. cerevisiae* metabolic model yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat. Commun.* 10, 1–13. doi: 10.1038/s41467-019-11581-3

Ma, Z.-Q., Dasari, S., Chambers, M. C., Litton, M. D., Sobecki, S. M., Zimmerman, L. J., et al. (2009). Idpicker 2.0: Improved protein assembly with high discrimination peptide identification filtering. *J. Proteome Res.* 8, 3872–3881. doi: 10.1021/pr900360j

Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* 10:e1003580. doi: 10.1371/journal.pcbi.1003580

Maranas, C. D., and Zomorrodi, A. R. (2016). *Optimization Methods in Metabolic Networks*. Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781119188902

McAnulty, M. J., Yen, J. Y., Freedman, B. G., and Senger, R. S. (2012). Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism *in silico*. *BMC Syst. Biol.* 6:42. doi: 10.1186/1752-0509-6-42

Milne, C. B., Eddy, J. A., Raju, R., Ardekani, S., Kim, P.-J., Senger, R. S., et al. (2011). Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* ncimb 8052. *BMC Syst. Biol.* 5:130. doi: 10.1186/1752-0509-5-130

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., et al. (2017). IML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35:904. doi: 10.1038/nbt.3956

Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D. R., Ebrahim, A., et al. (2013). Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. Microbial cell factories 12:118. doi: 10.1186/1475-2859-12-118

Ng, C. Y., Khodayari, A., Chowdhury, A., and Maranas, C. D. (2015). Advances in *de novo* strain design using integrated systems and synthetic biology tools. *Curr. Opin. Chem. Biol.* 28, 105–114. doi: 10.1016/j.cbpa.2015.06.026

Olson, D. G., Hörl, M., Fuhrer, T., Cui, J., Zhou, J., Maloney, M. I., et al. (2017). Glycolysis without pyruvate kinase in *Clostridium thermocellum*. *Metab. Eng.* 39, 169–180. doi: 10.1016/j.ymben.2016.11.011

Olson, D. G., McBride, J. E., Shaw, A. J., and Lynd, L. R. (2012). Recent progress in consolidated bioprocessing. *Curr. Opin. Biotechnol.* 23, 396–405. doi: 10.1016/j.copbio.2011.11.026

Palsson, B. Ø. (2015). *Systems Biology: Constraint-Based Reconstruction and Analysis*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139854610

Papanek, B., Biswas, R., Rydzak, T., and Guss, A. M. (2015). Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. *Metab. Eng.* 32, 49–54. doi: 10.1016/j.ymben.2015.09.002

Peters, N. K. (2018). *Bioenergy Research Centers*. Technical report, USDOE Office of Science (SC), Washington, DC. doi: 10.2172/1471709

Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., et al. (2006). The path forward for biofuels and biomaterials. *Science* 311, 484–489. doi: 10.1126/science.1114736

Reimers, A.-M., Lindhorst, H., and Waldherr, S. (2017). A protocol for generating and exchanging (genome-scale) metabolic resource allocation models. *Metabolites* 7:47. doi: 10.3390/metabo7030047

Roberts, S. B., Gowen, C. M., Brooks, J. P., and Fong, S. S. (2010). Genome-scale metabolic analysis of clostridium thermocellum for bioethanol production. *BMC Syst. Biol.* 4:31. doi: 10.1186/1752-0509-4-31

Rydzak, T., McQueen, P. D., Krokhin, O. V., Spicer, V., Ezzati, P., Dwivedi, R. C., et al. (2012). Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. *BMC Microbiol.* 12:214. doi: 10.1186/1471-2180-12-214

Salimi, F., Zhuang, K., and Mahadevan, R. (2010). Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. *Biotechnol. J.* 5, 726–738. doi: 10.1002/biot.201000159

Schellenberger, J., Lewis, N. E., and Palsson, B. Ø. (2011). Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys. J.* 100, 544–553. doi: 10.1016/j.bpj.2010.12.3707

Senger, R. S., and Papoutsakis, E. T. (2008). Genome-scale model for *Clostridium acetobutylicum*: Part I. Metabolic network resolution and analysis. *Biotechnol. Bioeng.* 101, 1036–1052. doi: 10.1002/bit.22010

Seo, H., Lee, J.-W., Garcia, S., and Trinh, C. T. (2019). Single mutation at a highly conserved region of chloramphenicol acetyltransferase enables isobutyl acetate production directly from cellulose by *Clostridium thermocellum* at elevated temperatures. *Biotechnol. Biofuels* 12:245. doi: 10.1186/s13068-019-1583-8

Seo, H., Nicely, P. N., and Trinh, C. T. (2020). Endogenous carbohydrate esterases of *Clostridium thermocellum* are identified and disrupted for enhanced

isobutyl acetate production from cellulose. *Biotechnol. Bioeng.* 117, 2223–2236. doi: 10.1002/bit.27360

Serrano-Bermúdez, L. M., Barrios, A. F. G., Maranas, C. D., and Montoya, D. (2017). *Clostridium butyricum* maximizes growth while minimizing enzyme usage and ATP production: metabolic flux distribution of a strain cultured in glycerol. *BMC Syst. Biol.* 11:58. doi: 10.1186/s12918-017-0434-0

Shafiee, S., and Topal, E. (2009). When will fossil fuel reserves be diminished? *Energy Policy* 37, 181–189. doi: 10.1016/j.enpol.2008.08.016

Stephanopoulos, G., and Vallino, J. J. (1991). Network rigidity and metabolic engineering in metabolite overproduction. *Science* 252, 1675–1681. doi: 10.1126/science.1904627

Szegezdi, J., and Csizmadia, F. (2007). "Method for calculating the PKA values of small and large molecules," in *Abstracts of Papers of The American Chemical Society*, Vol. 233 (Washington, DC: Amer Chemical Soc).

Tabb, D. L., Fernando, C. G., and Chambers, M. C. (2007). Myrimatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *J. Proteome Res.* 6, 654–661. doi: 10.1021/pr 0604054

Taverner, T., Karpievitch, Y. V., Polpitiya, A. D., Brown, J. N., Dabney, A. R., Anderson, G. A., et al. (2012). Danter: an extensible r-based tool for quantitative analysis of-omics data. *Bioinformatics* 28, 2404–2406. doi: 10.1093/bioinformatics/bts449

Thiele, I., and Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* 5:93. doi: 10.1038/nprot.2009.203

Thompson, R. A., Dahal, S., Garcia, S., Nookaew, I., and Trinh, C. T. (2016). 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. *Biotechnol. Biofuels* 9:194. doi: 10.1186/s13068-016-0607-x

Thompson, R. A., Layton, D. S., Guss, A. M., Olson, D. G., Lynd, L. R., and Trinh, C. T. (2015). Elucidating central metabolic redox obstacles hindering ethanol production in *Clostridium thermocellum. Metab. Eng.* 32, 207–219. doi: 10.1016/j.ymben.2015.10.004

Thompson, R. A., and Trinh, C. T. (2017). Overflow metabolism and growth cessation in *Clostridium thermocellum* DSM1313 during high cellulose loading fermentations. *Biotechnol. Bioeng.* 114, 2592–2604. doi: 10.1002/bit.26374

Tian, L., Papanek, B., Olson, D. G., Rydzak, T., Holwerda, E. K., Zheng, T., et al. (2016). Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum. Biotechnol. Biofuels* 9:116. doi: 10.1186/s13068-016-0528-8

Trinh, C. T. (2012). Elucidating and reprogramming *Escherichia coli* metabolisms for obligate anaerobic n-butanol and isobutanol production. *Appl. Microbiol. Biotechnol.* 95, 1083–1094. doi: 10.1007/s00253-012-4197-7

Trinh, C. T., Liu, Y., and Conner, D. J. (2015). Rational design of efficient modular cells. *Metab. Eng.* 32, 220–231. doi: 10.1016/j.ymben.2015.10.005

Trinh, C. T., Wlaschin, A., and Srienc, F. (2009). Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl. Microbiol. Biotechnol.* 81, 813–826. doi: 10.1007/s00253-008-1770-1

Wallenius, J., Viikilä, M., Survase, S., Ojamo, H., and Eerikäinen, T. (2013). Constraint-based genome-scale metabolic modeling of *Clostridium acetobutylicum* behavior in an immobilized column. *Bioresour. Technol.* 142, 603–610. doi: 10.1016/j.biortech.2013.05.085

Wang, L., and Maranas, C. D. (2018). Mingenome: an *in silico* top-down approach for the synthesis of minimized genomes. *ACS Synth. Biol.* 7, 462–473. doi: 10.1021/acssynbio.7b00296

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., et al. (2011). Metabolic engineering of *Escherichia coli* for direct production of 1, 4-butanediol. *Nat. Chem. Biol.* 7:445. doi: 10.1038/nchembio.580

Yoo, M., Bestel-Corre, G., Croux, C., Riviere, A., Meynial-Salles, I., and Soucaille, P. (2015). A quantitative system-scale characterization of the metabolism of *Clostridium acetobutylicum. MBio* 6:e01808-15. doi: 10.1128/mBio.01808-15

Zhang, Y.-H. P., and Lynd, L. R. (2005). Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7321–7325. doi: 10.1073/pnas.0408734102

Zhou, J., Olson, D. G., Argyros, D. A., Deng, Y., van Gulik, W. M., van Dijken, J. P., et al. (2013). Atypical glycolysis in *Clostridium thermocellum. Appl. Environ. Microbiol.* 79, 3000–3008. doi: 10.1128/AEM.04037-12