



# Genome-Scale Metabolic Model of *Caldicellulosiruptor bescii* Reveals Optimal Metabolic Engineering Strategies for Bio-based Chemical Production

Ke Zhang,<sup>a</sup> Weishu Zhao,<sup>a</sup> Dmitry A. Rodionov,<sup>b,c</sup> Gabriel M. Rubinstein,<sup>d</sup> Diep N. Nguyen,<sup>d</sup> Tania N. N. Tanwee,<sup>d</sup> James Crosby,<sup>e</sup> Ryan G. Bing,<sup>e</sup> Robert M. Kelly,<sup>e</sup>  Michael W. W. Adams,<sup>d</sup>  Ying Zhang<sup>a</sup>

<sup>a</sup>Department of Cell and Molecular Biology, College of the Environment and Life Sciences, University of Rhode Island, Kingston, Rhode Island, USA

<sup>b</sup>Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, California, USA

<sup>c</sup>A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

<sup>d</sup>Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia, USA

<sup>e</sup>Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina, USA

Ke Zhang and Weishu Zhao contributed equally to this article. The order of the co-first authors was determined alphabetically.

**ABSTRACT** Metabolic modeling was used to examine potential bottlenecks that could be encountered for metabolic engineering of the cellulolytic extreme thermophile *Caldicellulosiruptor bescii* to produce bio-based chemicals from plant biomass. The model utilizes subsystems-based genome annotation, targeted reconstruction of carbohydrate utilization pathways, and biochemical and physiological experimental validations. Specifically, carbohydrate transport and utilization pathways involving 160 genes and their corresponding functions were incorporated, representing the utilization of C5/C6 monosaccharides, disaccharides, and polysaccharides such as cellulose and xylan. To illustrate its utility, the model predicted that optimal production from biomass-based sugars of the model product, ethanol, was driven by ATP production, redox balancing, and proton translocation, mediated through the interplay of an ATP synthase, a membrane-bound hydrogenase, a bifurcating hydrogenase, and a bifurcating NAD- and NADP-dependent oxidoreductase. These mechanistic insights guided the design and optimization of new engineering strategies for product optimization, which were subsequently tested in the *C. bescii* model, showing a nearly 2-fold increase in ethanol yields. The *C. bescii* model provides a useful platform for investigating the potential redox controls that mediate the carbon and energy flows in metabolism and sets the stage for future design of engineering strategies aiming at optimizing the production of ethanol and other bio-based chemicals.

**IMPORTANCE** The extremely thermophilic cellulolytic bacterium, *Caldicellulosiruptor bescii*, degrades plant biomass at high temperatures without any pretreatments and can serve as a strategic platform for industrial applications. The metabolic engineering of *C. bescii*, however, faces potential bottlenecks in bio-based chemical productions. By simulating the optimal ethanol production, a complex interplay between redox balancing and the carbon and energy flow was revealed using a *C. bescii* genome-scale metabolic model. New engineering strategies were designed based on an improved mechanistic understanding of the *C. bescii* metabolism, and the new designs were modeled under different genetic backgrounds to identify optimal strategies. The *C. bescii* model provided useful insights into the metabolic controls of this organism thereby opening up prospects for optimizing production of a wide range of bio-based chemicals.

**KEYWORDS** bio-based chemical production, *Caldicellulosiruptor*, central carbon metabolism, metabolic engineering, metabolic modeling, redox balance

**Citation** Zhang K, Zhao W, Rodionov DA, Rubinstein GM, Nguyen DN, Tanwee TNN, Crosby J, Bing RG, Kelly RM, Adams MWW, Zhang Y. 2021. Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for bio-based chemical production. *mSystems* 6:e01351-20. <https://doi.org/10.1128/mSystems.01351-20>.

**Editor** Zarath M. Summers, ExxonMobil Research and Engineering

**Copyright** © 2021 Zhang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ying Zhang, [yingzhang@uri.edu](mailto:yingzhang@uri.edu).

For a companion article on this topic, see <https://doi.org/10.1128/mSystems.01345-20>.

**Received** 21 December 2020

**Accepted** 4 May 2021

**Published** 1 June 2021

Global population growth and increasing industrial development have generated a greater demand for energy and resources (1). However, the supply of conventional fossil fuels is finite and their use is burdened with environmental issues. Therefore, their replacement with a sustainable energy supply becomes more and more pressing. Plant biomass can be used to generate fuels and other valuable chemicals (e.g., ethanol, biodiesel, plastics, and industrial chemicals) that can serve as renewable alternatives to many of the products derived from petroleum or natural gas and do so in environmentally benign processes. Inedible plants (such as switchgrass and poplar) are the ideal biomass sources for industrial biotechnology, since they are abundant, renewable, economical, and do not compete with food supply. However, plant biomass is extremely recalcitrant to deconstruction and thus requires chemical or enzymatic pretreatment to make the carbohydrate content accessible to prototypical industrial microbes. This creates an economic barrier for industrial development, as well as generating negative environmental ramifications (2). Clearly, alternative microbes that can degrade nonpretreated plant biomass, although rare in nature, are needed to meet this challenge through metabolic engineering to generate the needed biofuels and other bioproducts.

*Caldicellulosiruptor bescii* is a promising platform microorganism for converting lignocellulose to bio-based chemicals, based on its ability to degrade nonedible plant biomass, both cellulose and hemicellulose, without any enzymatic or chemical pretreatments (3, 4). *C. bescii* is an extremely thermophilic, strictly anaerobic, Gram-positive bacterium (5). It is the most thermophilic cellulolytic bacterium known to date with an optimal growth temperature ( $T_{\text{opt}}$ ) of 78 to 80°C and a maximum growth temperature ( $T_{\text{max}}$ ) of 90°C. *C. bescii* utilizes a wide range of simple and complex carbohydrates, both C5- and C6-based (5). Its high growth temperature creates several advantages for industrial applications, including (i) reduced risks of phage infection and contamination (since no known phages survive at such high temperatures), (ii) higher solubility of plant biomass polysaccharides, and (iii) the potential to distill volatile products directly from fermentation broths, which also minimizes product toxicity (6).

Engineered *C. bescii* strains are capable of producing ethanol (7–11) or acetone and  $\text{H}_2$  (12) by fermenting simple sugars (i.e., cellobiose and maltose), crystalline cellulose, or plant biomass (i.e., switchgrass and poplar) as the sole carbon and energy source. Wild-type *C. bescii* does not contain genes required for ethanol production, so foreign genes have been introduced through a uracil auxotrophy system enabled by development of *pyrFA* or *pyrE* mutants for the construction of engineered strains (13, 14). Recently, the *pyrE* deletion background, strain MACB1018, is shown to have a more stable genetic background than the *pyrFA* deletion background, strain JWC005 (15). A lactate dehydrogenase (*ldh*) knockout strain, MACB1034, is constructed with the *pyrE* mutant, providing a stable genetic background, and is used as the parent strain for the ethanol production (14). A bifunctional acetaldehyde/alcohol dehydrogenase (AdhE) from *Clostridium thermocellum* ( $T_{\text{opt}} = 55$  to 60°C) was expressed heterogeneously in *C. bescii*, enabling the engineered *C. bescii* strain to produce ethanol from crystalline cellulose, albeit at a suboptimal growth temperature of 60°C (10). In recent work, heterogeneous expression of genes encoding the enzymes, aldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus* ( $T_{\text{opt}} = 100^\circ\text{C}$ ) and alcohol dehydrogenase (AdhA) from *Thermoanaerobacter* sp. strain X514 ( $T_{\text{opt}} = 72^\circ\text{C}$ ), enabled *C. bescii* to produce ethanol from crystalline cellulose optimally at 75°C (16). However, further development of *C. bescii* strains to produce bio-based chemicals at industrially relevant titers requires a better understanding of fermentation patterns to inform metabolic engineering.

Genome-scale models (GEMs) provide a systems-level view of the metabolism and are frequently applied to optimize the production of bio-based chemicals in engineered microorganisms. This approach has been used to model thermophilic bacteria that are capable of degrading cellulose or hemicellulose, such as *Clostridium thermocellum* (cellulose;  $T_{\text{opt}} = 55$  to 60°C [17, 18]) and *Thermoanaerobacterium saccharolyticum* (hemicellulose;  $T_{\text{opt}} = 70^\circ\text{C}$  [19]), as well as the mesophile *Clostridium cellulolyticum* (20),

which degrades both cellulose and hemicellulose. In contrast to these organisms, *C. bescii* is much more thermophilic, growing up to 90°C (3) and, more importantly, utilizes C5-based hemicellulose, the second major carbohydrate in plant biomass, in addition to C6-based cellulose (21). This is a major advantage over thermophilic species that have only a cellulose-degrading or a hemicellulose-degrading metabolism (22–24). The adaptation of *C. bescii* to high temperature also represents an advantage over mesophilic species (i.e., *Clostridium cellulolyticum*) due to the better compatibility of thermophiles with industrial processing conditions.

Extensive biochemical and physiological studies of *C. bescii* have revealed several unique features in the central carbon and redox metabolism of this organism. Its central carbon metabolism is characterized by the presence of a glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GOR), which uses ferredoxin as an electron carrier for oxidizing glyceraldehyde-3-phosphate (GAP) to 3-phosphoglycerate. The GOR pathway functions in parallel to the canonical GAP dehydrogenase (GAPDH), which uses NAD as the electron carrier (25). A second potentially confounding aspect to *C. bescii* metabolism is the fate of the reduced ferredoxin ( $\text{Fd}_{\text{red}}$ ) produced by GOR and pyruvate ferredoxin oxidoreductase (POR), since  $\text{Fd}_{\text{red}}$  can be oxidized by three redox related enzyme complexes (10, 26). One of these is conventional membrane-bound [NiFe]-hydrogenase (MBH), which couples  $\text{Fd}_{\text{red}}$  oxidation to  $\text{H}_2$  production (9, 27). However, the other two are examples of a recently discovered type of energy-conserving enzyme that carries out electron bifurcation wherein an exergonic reaction is used to drive an endergonic reaction (28). The first is a bifurcating [FeFe]-hydrogenase (BF- $\text{H}_2$ ase) that reversibly couples  $\text{H}_2$  production to energy-yielding  $\text{Fd}_{\text{red}}$  oxidation and energy-requiring NADH oxidation (29). The second example is NADH-dependent ferredoxin NADP oxidoreductase (BF-Nfn); this also simultaneously oxidizes  $\text{Fd}_{\text{red}}$  and NADH, while coupling these two reactions to the reduction of NADP for biosynthetic purposes (30). The complex interconversion of the oxidized and reduced forms of these electron carriers, and their respective connections to the central carbon metabolism of *C. bescii* suggests that these processes are tightly regulated (25). To enhance NADH generation in *C. bescii*, required for ethanol production using AdhE, we also expressed heterogeneously a membrane-bound, sodium ion-pumping reduced ferredoxin NAD oxidoreductase (Rnf\_Na) from *Thermoanaerobacter* sp. strain X514 ( $T_{\text{opt}} = 60^\circ\text{C}$ ). It was hypothesized that this would use  $\text{Fd}_{\text{red}}$  produced from carbohydrate oxidation to increase the production of NADH for NADH-dependent ethanol production; however, the introduction of Rnf\_Na had only a marginal effect in enhancing ethanol yields (10).

Further enhancement of ethanol production in *C. bescii* obviously requires a fundamental understanding of the redox processes at a systems level, and that is the goal of the research described here. Overall, the unique metabolic and physiological characteristics of *C. bescii* motivated the application of GEM reconstruction for guiding the metabolic engineering designs of this organism for bioproduct generation. This study represents the first metabolic modeling of *C. bescii* using a subsystems-level approach. Model validations were performed using experimental physiological and metabolic data. The *C. bescii* model was further refined to accurately simulate use of carbon and energy sources, including both simple (glucose, cellobiose) and complex (cellulose, xylan) carbohydrates. The model was then used to identify metabolic bottlenecks and predict novel engineering strategies for the optimization of bio-based chemicals by resolving constraints in redox balancing.

## RESULTS

**Overview of the genome-scale metabolic model GEM-iCbes.** The genome-scale metabolic model of *C. bescii*, GEM-iCbes, contains 610 genes, 718 metabolites, and 714 metabolic reactions (see Data Set S1 in the supplemental material). Compared to existing models of other plant biomass-degrading thermophilic bacteria that can use either cellulose or hemicellulose (but not both), the GEM-iCbes included a higher percentage of protein-coding genes (CDS) over the genome (24%) and across all metabolic genes

predicted by the COG classification (87%) (Table 1). Reactions in the GEM-iCbes included 589 cytosolic reactions, 60 transmembrane reactions, and 12 extracellular reactions, all associated with specific gene annotations. A high consistency score (99%) was achieved with GEM-iCbes when analyzed using the Memote consistency check (31), which is comparable to the consistency score of 98%, for a recently updated model of *C. thermocellum*, iCBI655 (18) (Table 1).

The metabolic reconstruction of *C. bescii* involved extensive manual curation, combining evidence from putative gene orthology and biochemical/physiological experiments in *C. bescii* and other thermophilic or Gram-positive bacteria. The experimentally verified biochemical functions included the ferredoxin-dependent oxidoreductase GOR (25), a phosphoglycerate kinase (PGK) that uses either ADP or GDP as cofactors (see Data Set S2), a GTP-forming phosphoenolpyruvate carboxykinase (PCK) (see Data Set S2), as well as a unique combination of three redox-related enzymes that all use ferredoxin as an electron carrier: MBH, BF-H<sub>2</sub>ase, and BF-Nfn (9, 25, 32). In addition, the *C. bescii* model integrated utilization pathways of various simple and complex carbohydrates, including monosaccharides of both C5 (e.g., xylose and arabinose) and C6 (e.g., glucose, fructose, and galactose), disaccharides (e.g., lactose, maltose, and cellobiose), as well as a variety of plant polysaccharides, including both cellulose and hemicellulose (xylan). Functional roles of extracellular enzymes that contribute to polysaccharide degradation, including 15 glycosyl hydrolases (GHs) and three polysaccharide lyases (PLs), were manually curated and incorporated into the model. The extracellular GHs are responsible for the degradation of cellulose, xylan, mannans, and starch, while the PLs are required for pectin degradation. Uptake of oligo- and monosaccharides involves 18 ABC superfamily transporters encoded by 65 genes. The reconstructed carbohydrate catabolic pathways involve at least 77 presumably cytoplasmic enzymes, including 37 GHs. Specific transcriptional regulators controlling these carbohydrate utilization pathways were excluded from the model; their transcriptional regulons, including DNA-binding sites and carbohydrate utilization genes, and operons are presented in the accompanying paper (33). Compared to the existing models of related cellulose or hemicellulose-degrading species, the *C. bescii* model represented an expansion of 160 genes involved in the carbohydrate transport and catabolism (Table 1), the largest functional subsystem represented in the model (see Fig. S1).

**Model validation.** The *C. bescii* model was validated by comparing model simulations with experimental measurements of biomass and bioproduct yields in the wild-type and engineered strains. These included the wild-type strain DSM 6725 (WT) and the lactate dehydrogenase knockout strain MACB1034 ( $\Delta ldh$ ) (5, 15) (Table 2). The model was first validated based on the prediction of biomass and bioproduct yields when monosaccharides, specifically the glucose (C6) or fructose (C5), were used as sole carbon sources. Both WT and  $\Delta ldh$  strains were simulated in comparison to experimental measurements (see Data Set S2). A general consistency was observed between the model-predicted and the experimentally measured yields, with the WT model specifically matching the measurements of biomass yields and bioproduct (i.e., lactate, acetate, and pyruvate) productions in batch cultures of *C. bescii* (Fig. 1). Consistent with the experimental data, a slightly higher biomass flux was produced with fructose than with glucose in the  $\Delta ldh$  model, and higher biomass yields were observed in the WT strain than the  $\Delta ldh$  strain with either glucose or fructose as the carbon substrate (Fig. 1A). The modeled product yields also showed a high consistency with experimental measurements (Fig. 1B). As expected from experimental studies, the production of lactate was prevented in the  $\Delta ldh$  model, where the carbon flow was redirected to the production of acetate and pyruvate (see Data Set S2).

Additional validation of the *C. bescii* model was performed by comparing the carbon utilization phenotypes of the WT *C. bescii*, for both model predictions and experimental data (5, 34). A total of 27 carbon compounds were examined, including monosaccharides (C5 or C6), disaccharides, polysaccharides, small organic acids (e.g., acetate and lactate), and polyols (e.g., glycerol, erythritol, and xylitol). The model accurately

**TABLE 1** Comparison of the *C. bescii* model with existing models of plant-biomass degrading bacteria<sup>a</sup>

Parameter	<i>Caldicellulosiruptor bescii</i> DSM 6725	<i>Clostridium thermocellum</i> DSM 1313	<i>Clostridium thermocellum</i> ATCC 27405	<i>Thermoanaerobacterium saccharolyticum</i> JW/SL-YS485	<i>Clostridium cellulolyticum</i> H10
Physiology					
Growth temp, range (optimal) (°C)	42–90 (78)	40–60 (50–55)	50–68 (60)	45–70 (60)	25–45 (34)
Plant biomass	Cellulose and hemicellulose	Cellulose	Cellulose	Hemicellulose	Cellulose and hemicellulose
Genome					
Size (Mbp)	2.92	3.56	3.84	2.83	4.07
No. of CDS	2,565	2,939	3,218	2,407	3,334
GEM					
Model ID	GEM-iChes	iCB1655	iCh446	<i>T. saccharolyticum</i> model	iFS431
No. of genes	610	665	446	315	431
% CDS in model	24	23	14	13	13
% COG metabolic genes in model	87	80	53	34	42
No. of genes of carbohydrate transport and utilization	160	100	55	22	69
No. of metabolites	718	795	599	503	603
No. of reactions (exclude exchange)	714	795	607	515	584
No. of gene-associated reactions	661	689	458	461	513
% gene-associated reactions	93	87	75	90	88
Consistency (%)	99	98	ND	85	57
Stoichiometry consistency (%)	100	100	ND	99	100
Mass balance (%)	99	97	ND	93	0
Charge balance (%)	99	97	ND	94	0
Metabolite connectivity (%)	100	100	ND	100	99
Bounded flux in default medium (%)	93	96	ND	11	1
Reference(s)	This study	18, 24	17, 61	19, 62	20, 62

<sup>a</sup>Growth temperature indicates both the growth range and the optimal temperatures (in parentheses). The number of CDS indicates the count of protein-coding sequences in the genome. The consistency scores were calculated using the Memote software (v0.11.1) for standardized model assessment (31). ND, no data were available due to incompatibility of the model with Memote.



**TABLE 2** Summary of strains modeled in this study<sup>a</sup>

Index	Genotype	Parent strain	Strain ID	Reference	Simulation(s)
WT	Wild type		DSM 6725	5	Model validation
$\Delta ldh$	$\Delta pyrE \Delta ldh$	WT	MACB1034	14	Model validation
E1	$\Delta pyrE \Delta ldh adhE^+ rnf\_Na^+$	$\Delta ldh$ mutant	MACB1062	10	Model validation
E1M	$adhE^+ rnf\_Na^+$	WT	This study	This study	Minimal network analysis; basal strain for the testing of engineering designs

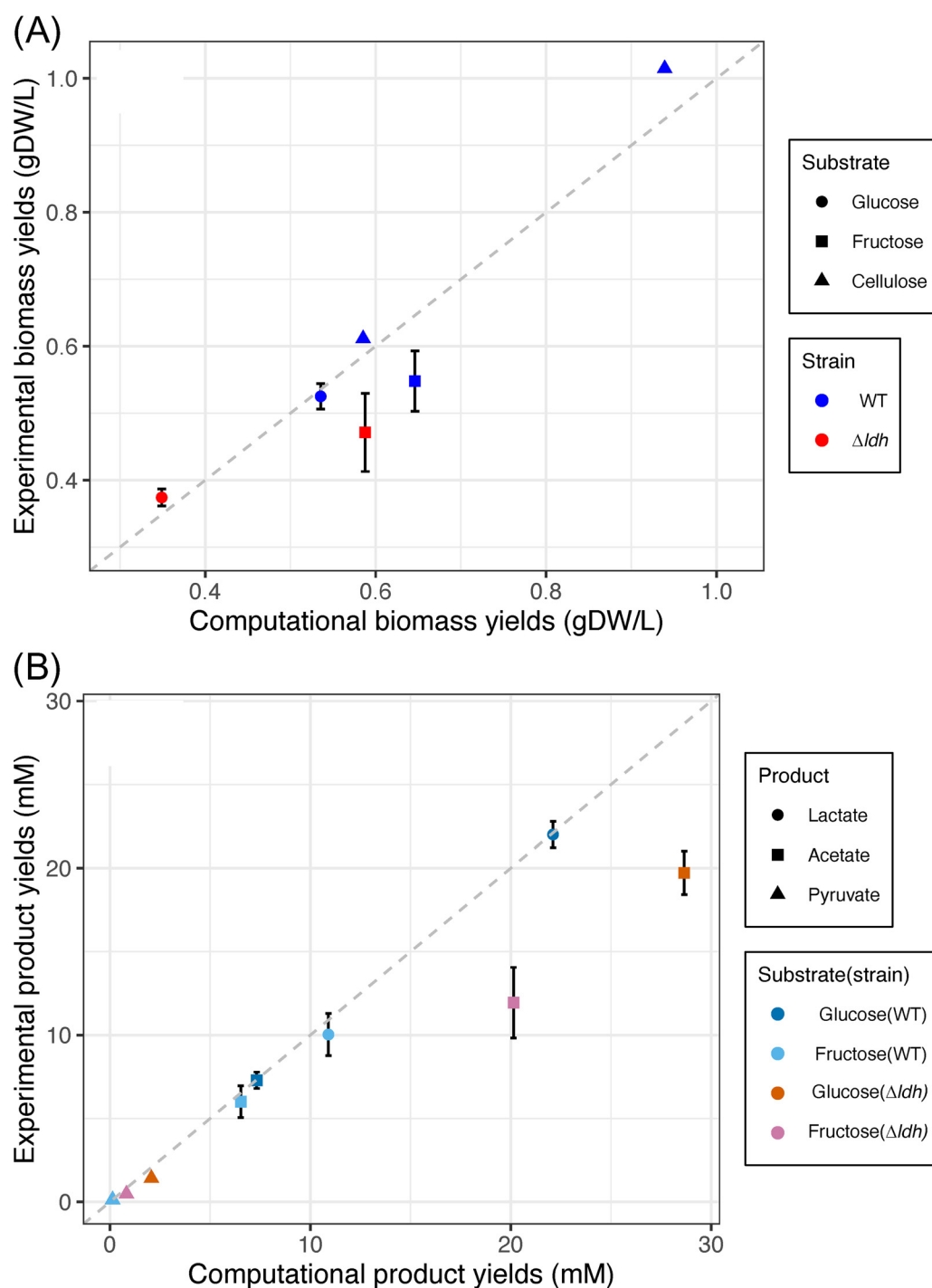
<sup>a</sup>The relevant enzymes are as follows: PyrE, orotate phosphoribosyltransferase; LDH, lactate dehydrogenase; AdhE, bifunctional acetaldehyde/alcohol dehydrogenase; Rnf\_Na, sodium ion-pumping reduced ferredoxin NAD oxidoreductase.

represented the growth of *C. bescii* on 18 carbohydrates and its lack of growth on 8 other carbon sources, with only one inconsistency in the simulation of dextran utilization (see Fig. S2). While a complete dextran utilization pathway was identified by the *C. bescii* model, existing experimental data showed no growth when dextran was used as a sole carbon source. The simulation of *C. bescii* model for growth on polysaccharides was also validated with crystalline cellulose (Avicel PH-101) as the sole carbon source, using experimental data obtained from the WT strain of *C. bescii* (Fig. 1A). The prediction of biomass yields under a cellulose load of 10 and 50 g/liter agreed well with the experimental data (35). Specifically, a lower biomass yield was observed under a cellulose load of 50 g/liter, likely due to the low percentage of solubilized cellulose and the reduced glucose uptake and utilization under the high cellulose load (see Data Set S2). Overall, the *C. bescii* model was consistent with the experimental data in the prediction of biomass and bioproduct yields.

**Maximization of ethanol productions in *C. bescii*.** The validated *C. bescii* model was used to identify potential metabolic shifts resulting from the heterogeneous expression of the ethanol production gene *adhE*. To achieve this, random simulations of minimal metabolic networks were performed at three different levels of ethanol production: (i) no-ethanol, where the ethanol production flux was constrained to zero; (ii) half-maximum, where the ethanol production flux was constrained to 50% of the maximum; and (iii) maximum, where the ethanol production flux was constrained to the maximum of model predictions. Model simulations were performed with the E1M strain (Table 2), where AdhE was included to enable the synthesis of ethanol from acetyl coenzyme A (acetyl-CoA), and an Rnf\_Na was included based on an existing design of the ethanol-producing E1 strain (10) to evaluate its significance in redox balancing. While the original E1 strain included additional background mutations, such as the deletion of *ldh* and *pyrE*, these background mutations were omitted in the simulation of E1M to allow flexibility in the further engineering of ethanol-producing strains without requiring a specific genetic background.

The minimal metabolism simulations were performed using a modified DSMZ 516 medium with up to 11.8 g/liter cellulose as the sole carbon source, which corresponds to the experimentally measured consumption of cellulose in the E1 strain (10). Up to 1,000 minimal networks were identified for each of the three ethanol production conditions, while maintaining a minimal biomass production that is comparable to the experimentally measured biomass yields (0.41 g of dry weight/liter) under the same medium (see Data Set S3). Examining the occurrence of metabolic reactions among all minimal networks enabled the classification of reactions in the *C. bescii* model into three distinct sets: (i) a core-essential set, representing reactions required in all minimal networks of a simulation condition; (ii) a conditionally essential set, representing reactions required in some but not all minimal networks; and (iii) a nonessential set, representing reactions not used in any minimal networks. Convergence analysis of the minimal networks was performed under the no-ethanol, half-maximum, and maximum-ethanol simulations (see Materials and Methods).

Under all three conditions, the classification of core-essential, conditionally essential, and nonessential reaction sets converged before 1,000 rounds of simulations (see Fig. S3). With increasing production of ethanol, a slight increase was observed in the number of core-essential reactions (i.e., from 244 to 275 reactions in the no-ethanol to



**FIG 1** Experiment-based model validation using experimental measurements from WT and  $\Delta ldh$  using glucose and fructose as sole carbon sources (5, 15). (A) Growth yield validation; (B) product yield validation. A dashed line is included in each panel to mark the 1:1 comparison between experimental and computational data. Error bars represent the standard deviations among experimental replicates.

maximum-ethanol conditions), while a slight decrease was observed in the number of conditionally essential reactions (i.e., from 176 to 124 reactions). Several reactions in the conditionally essential set under no-ethanol simulation were identified in the core-essential set when ethanol production was required. These included AdhE and POR, where POR is required for the production of ethanol due to its essential roles in converting pyruvate to acetyl-CoA, a precursor required for ethanol production via AdhE

(see Data Set S3). The utilization of POR also led to the production and release of  $\text{CO}_2$ , along with the production of ethanol (Fig. 2).

Comparison of minimal networks under half-maximum and the maximum production of ethanol revealed a number of pathways that should be inhibited for the optimization of ethanol production. These included the release of  $\text{H}_2$  and the generation of other products, such as pyruvate, lactate, and acetoin (Fig. 2). In addition, 21 enzymatic functions were classified as core-essential when the ethanol production was constrained to the maximum, while they were conditionally essential under the half-maximum ethanol production (see Data Set S3). These included reactions in ATP/GTP production, amino acids metabolism, and pyrimidine metabolism. The reactions in amino acids and pyrimidine metabolism carried low fluxes and were mainly used for offsetting the overflow of GTP-forming phosphoenolpyruvate carboxykinase (PCK) reaction (Fig. 2). Several ATP-producing reactions, including the  $\text{F}_1\text{F}_0$  ATP synthase (ATPSYN) and the substrate-level phosphorylation pathway composed of phosphotransacetylase (PTA) and acetate kinase (AckA), were also classified as core-essential when ethanol production approaches the maximum. While the ATPSYN was required for the production of ATP, the substrate-level phosphorylation was mainly used for carbon overflow through the trace production of acetate. Finally, the redox reactions, involving MBH, BF- $\text{H}_2$ ase, and BF-Nfn, were also required as core-essential functions for maximizing the production of ethanol. According to model predictions, under maximum ethanol production, MBH carried positive metabolic fluxes that couple the oxidation of  $\text{Fd}_{\text{red}}$  with the production of  $\text{H}_2$ . Also, BF- $\text{H}_2$ ase carried negative metabolic fluxes that take the  $\text{H}_2$  produced from MBH into the generation of the reduced cofactors NADH and  $\text{Fd}_{\text{red}}$ , while BF-Nfn carries positive fluxes that facilitate the production of NADPH using reducing power from NADH and  $\text{Fd}_{\text{red}}$  (Fig. 2).

#### Impact of BF- $\text{H}_2$ ase reversibility and genetic backgrounds on ethanol production.

The minimal network analysis revealed the potential significance of coupling  $\text{H}_2$  production by MBH with the  $\text{H}_2$ -consuming direction of BF- $\text{H}_2$ ase in the conservation of reducing compounds for optimizing ethanol production. Such coupling might be feasible, as the BF- $\text{H}_2$ ase has been shown to readily consume  $\text{H}_2$  while reducing NAD and oxidized ferredoxin ( $\text{Fd}_{\text{ox}}$ ) *in vitro* (29). However, it is not clear whether the MBH and BF- $\text{H}_2$ ase coupling is feasible under physiological conditions. As a counter example, BF- $\text{H}_2$ ase has been determined as a primary source of  $\text{H}_2$  production rather than  $\text{H}_2$  consumption in *C. bescii* (9). This poses the question, does the BF- $\text{H}_2$ ase catalyze a reversible reaction *in vivo*? If not, what is the impact on ethanol production?

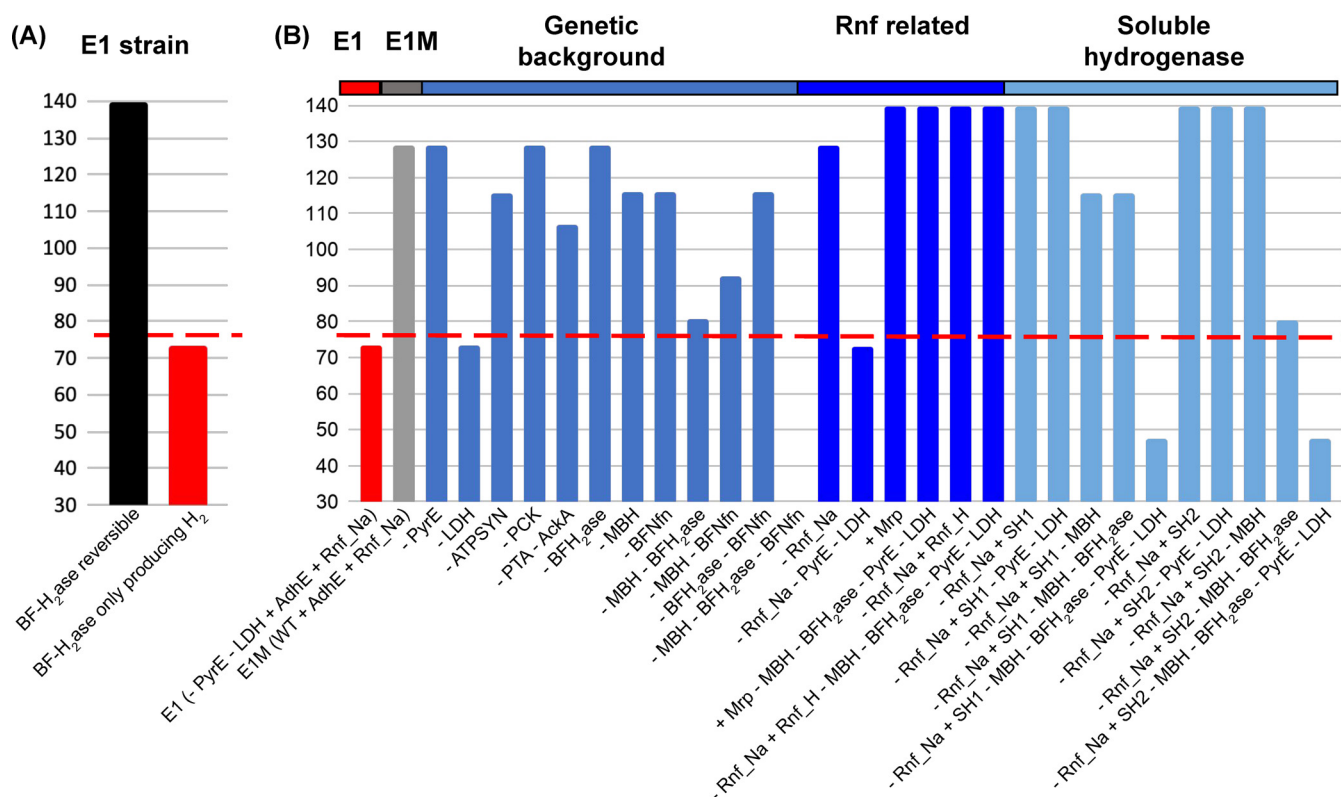
To address this question, comparisons were performed on the variability of metabolic fluxes between two distinct simulations: one constraining the BF- $\text{H}_2$ ase as a reversible reaction that is capable of carrying flux in both  $\text{H}_2$ -producing and  $\text{H}_2$ -consuming directions, and another constraining the BF- $\text{H}_2$ ase as an irreversible reaction that only carries flux in the  $\text{H}_2$ -producing direction. Simulations were performed on the engineered E1 strain (MACB1062), where ethanol production was optimized in the two BF- $\text{H}_2$ ase settings, followed by the simulation of flux variability for every reaction in the *C. bescii* model (see Data Set S4). When the E1 strain was modeled with a BF- $\text{H}_2$ ase that is  $\text{H}_2$ -producing only, it had an optimal ethanol production that was comparable to the experimental measurement (Fig. 3A). This was in contrast to the simulation of E1 when BF- $\text{H}_2$ ase is allowed to consume  $\text{H}_2$ , where the predicted ethanol production almost doubled the experimental measurements (Fig. 3A). Therefore, it is likely that the BF- $\text{H}_2$ ase in *C. bescii* could primarily carry out functions in the  $\text{H}_2$ -producing direction under physiological conditions.

The effect of a  $\text{H}_2$ -producing BF- $\text{H}_2$ ase on ethanol production was probed in a diverse range of genetic contexts (Fig. 3B). To better assess the influence of different genetic backgrounds, we used the E1M (Table 2), instead of E1, as the base strain for additional simulations (see Data Set S4). Two genetic backgrounds frequently used for *C. bescii* engineering were considered: one is the deletion of the *pyrE* gene, which is a feature of the stable parent strain for the engineering of ethanol production (15); the





msystems.asm.org 9



**FIG 3** Simulation of maximum ethanol productions for the evaluation of diverse engineering designs. (A) Simulation of ethanol production by the E1 strain with BF-H<sub>2</sub>ase as either reversible (black) or H<sub>2</sub>-producing only (red). (B) Simulation of engineering designs with the BF-H<sub>2</sub>ase constrained to the H<sub>2</sub>-producing direction. Strain E1M was proposed over the previous engineering design of E1 as a base strain. Further engineering designs were performed with insertion or deletion of genes on the E1M. The symbol “-” indicates the removal of a function, while the symbol “+” indicates the insertion of a function into the E1M model. The red dashed line indicates the experimentally measured ethanol production by the E1 strain.

other has a deletion of *ldh* gene, which is engineered to block lactate production (14). When the BF-H<sub>2</sub>ase was limited to the H<sub>2</sub>-producing direction, the deletion of *pyrE* had little influence on the ethanol production, while the *ldh* deletion caused a significant reduction of ethanol production (from 129 mM in the E1M base model to 73 mM in the *ldh*-deleted model) (Fig. 3B). In *C. bescii*, the gene that encodes LDH is also responsible for malate dehydrogenase (MDH) activity, which reversibly converts malate to oxaloacetate coupled to the reduction of NAD to NADH (see Data Set S2). Therefore, MDH function is also blocked in the *ldh* deletion mutant. Since MDH is one of the main sources of NADH when the BF-H<sub>2</sub>ase is in the H<sub>2</sub>-producing direction (see Data Set S4), the absence of MDH activity has a negative impact that overrides the benefit of removing LDH for enhancing ethanol production.

Other mutant strains were modeled by inhibiting combinations of the ATP/GTP production functions identified as essential for reaching the maximum production of ethanol, including ATPSYN, PCK, and AckA (Fig. 2). When BF-H<sub>2</sub>ase is constrained to the H<sub>2</sub>-producing direction, a 17% reduction in ethanol yield (from 129 to 107 mM) was observed when the acetate production was blocked, while no influence was observed in the PCK mutant (Fig. 3B). This is likely due to the requirement of an increased flux of carbon overflow in support of substrate-level phosphorylation when the redox balance between Fd<sub>red</sub> and NADH was shifted with the inhibition of the H<sub>2</sub>-consumption function of BF-H<sub>2</sub>ase (see Data Set S4). In contrast, the deletion of ATPSYN resulted in 10% reduction in the ethanol production when BF-H<sub>2</sub>ase was constrained as H<sub>2</sub>-producing only (Fig. 3B).

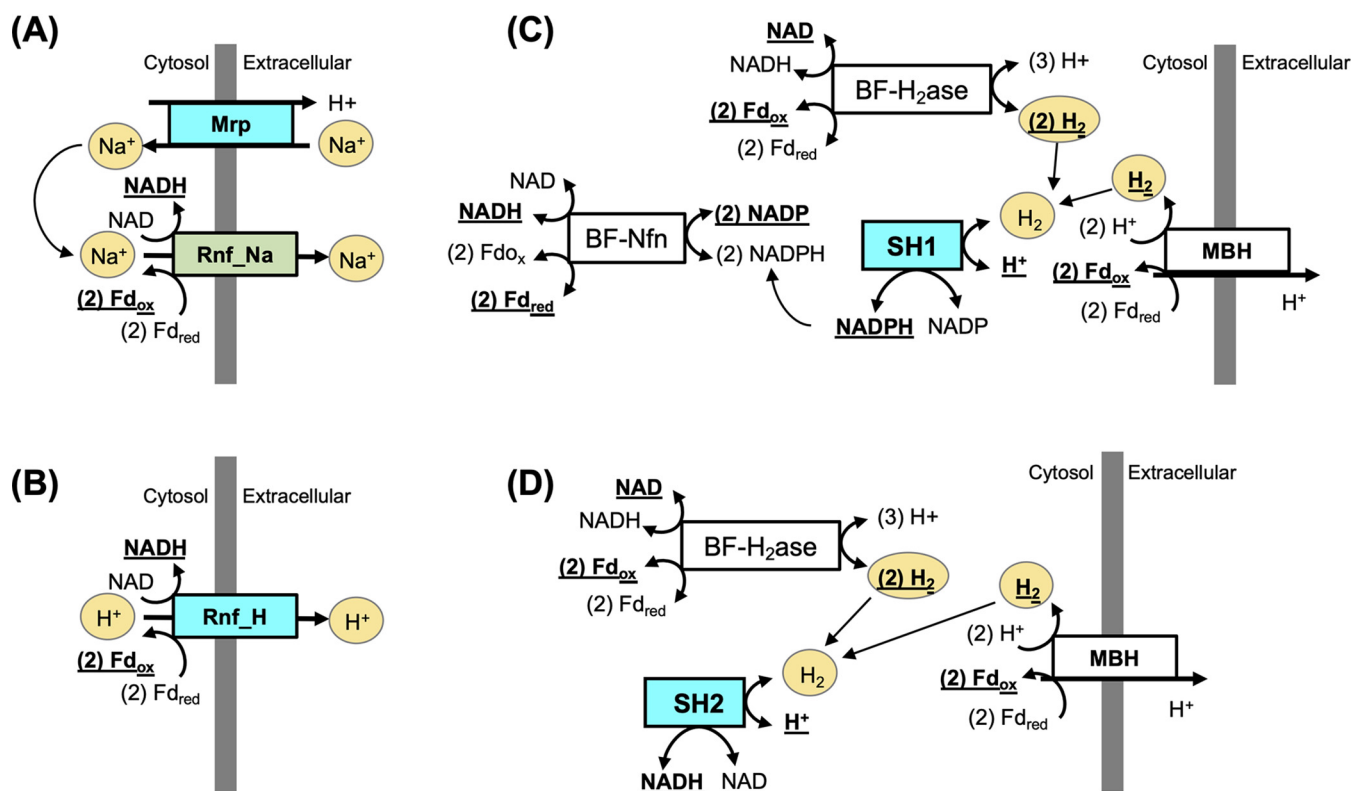
Additional mutants were introduced by inhibiting the redox related reactions, BF-H<sub>2</sub>ase, MBH, and BF-Nfn. The deletion of any combinations of the redox functions

resulted in a reduction of ethanol yield (Fig. 3B), indicating the significant roles of redox balance in maximizing ethanol production. Among single mutants of these redox functions, the lowest ethanol production was observed with the deletion of MBH. This is likely due to the inhibition of proton pumping by MBH, which limits the ATPSYN-mediated ATP production using proton gradient. Indeed, the predicted ethanol yield of the MBH deletion mutant (116 mM) was comparable to the ATPSYN deletion mutant. In the MBH mutant, the ATPSYN flux was constrained to less than 2% of its minimum flux seen in the E1M base strain, while the flux of substrate-level phosphorylation (i.e., via AckA) was increased up to 50-fold. Both BF-H<sub>2</sub>ase and BF-Nfn were used in the Fd<sub>red</sub>-oxidizing direction in the MBH mutant, mediating redox balance by consuming the Fd<sub>red</sub> produced by POR (see Data Set S4). Among double mutants of the redox functions, the lowest ethanol production (81 mM) was observed when both MBH and BF-H<sub>2</sub>ase were removed (Fig. 3B). This is likely due to the added effect of reduced proton pumping (e.g., as seen in the single deletion of MBH) and a lack of redox mechanisms that couple the oxidation of Fd<sub>red</sub> (i.e., produced from POR) to the reduction of NAD; this highlights the significance of both ATP production and redox balance in enabling the maximum ethanol production of *C. bescii*. Further, the deletion of all three redox functions resulted in zero fluxes in both biomass yield and ethanol production, indicating the triple deletion of redox functions would be lethal (see Data Set S4).

**New engineering designs to enhance ethanol production.** As expected based on the above simulations, optimization of ethanol productions in *C. bescii* would require balancing ATP production and redox status and the selection of the appropriate genetic backgrounds for engineering designs. The reversibility of BF-H<sub>2</sub>ase also impacted redox reactions in central carbon metabolism: the NAD-reducing GAPDH function, while forming an alternative pathway to the ferredoxin reducing GOR when the BF-H<sub>2</sub>ase is reversible, was essential when BF-H<sub>2</sub>ase was constrained to only carry flux in H<sub>2</sub>-producing direction (see Fig. S4A and B). This is likely due to the high demand of NADH in the AdhE-dependent ethanol production and lack of a mechanism for coupling Fd<sub>red</sub> oxidation to the reduction of NAD when BF-H<sub>2</sub>ase functions solely in the H<sub>2</sub>-producing direction (Fig. 2). With the expectation of a primarily H<sub>2</sub>-producing BF-H<sub>2</sub>ase in *C. bescii* under physiological conditions (Fig. 3A), engineering strategies were explored with the H<sub>2</sub>-producing BF-H<sub>2</sub>ase version of the *C. bescii* model to achieve redox balancing and ethanol yield optimization.

A 77% increase in ethanol production is expected when comparing our original design of the ethanol-producing *C. bescii*, E1 (MACB1062) to a modeled E1M strain (Table 2), based on the observation that the *ldh* deletion inhibited ethanol production with a H<sub>2</sub>-producing BF-H<sub>2</sub>ase (Fig. 3B). Therefore, we chose E1M as the base strain to propose new engineering strategies. An additional feature in the design of the E1 strain included a sodium ion pumping oxidoreductase, Rnf\_Na, which was used with the goal of transferring the reducing power of Fd<sub>red</sub>, generated from GOR and POR, to the production of NADH. This initial strategy, however, had limited effect on the ethanol production. As shown in the model simulations, deletion of Rnf\_Na from either the E1M base model or the E1 engineered ethanol strain model had minimal influence on the maximum ethanol production (Fig. 3B). Upon closer examination, it was hypothesized that the potential lack of a sodium gradient under physiological conditions may limit the roles of Rnf\_Na in redox balancing in *C. bescii* (Fig. 2).

To test this hypothesis, two new engineering designs were proposed and tested in the *C. bescii* model. In the first design, a membrane-bound sodium-proton antiporter (Mrp) was inserted to the E1M strain to help establish a sodium gradient that can be used to drive the production of NADH by Rnf\_Na (Fig. 4A). In the second design, a proton- rather than sodium-driven reduced ferredoxin NAD oxidoreductase (Rnf\_H) was introduced to the E1M model in place of Rnf\_Na (Fig. 4B). Both engineering strategies were examined under a variation of genetic backgrounds (e.g., deletion of PyrE and LDH) and with the deletion of the H<sub>2</sub>-producing MBH and BF-H<sub>2</sub>ase (Fig. 3B). Optimization of ethanol production in both designs revealed a maximum production



**FIG 4** (A to D) Design of four potential engineering strategies predicted by the *C. bescii* model. Boxes with a white background indicate the native metabolic functions in *C. bescii*; boxes with a green background indicate insertion mutants present in existing strains of engineered *C. bescii*; boxes with a cyan background indicate the new engineering strategy proposed in this study. The products of each reaction are represented in boldface with underlines. Key intermediates of each engineering design are represented as ovals with a yellow background.

of 140 mM (see Data Set S4). The *pyrE* and *ldh* gene deletions were associated with only a minor reduction on the predicted maximum ethanol yield. The deletion of MBH and BF-H<sub>2</sub>ase had no influence on the maximum ethanol production for both engineering designs. Further examination of the flux variability indicated that the introduction of Rnf\_Na/Mrp or Rnf\_H served as an alternative to the essential proton-pumping function carried out by MBH. Therefore, a zero flux can be achieved for MBH and H<sub>2</sub> production, even when the BF-H<sub>2</sub>ase was constrained to carry only H<sub>2</sub>-producing fluxes (see Fig. S4C and D).

Besides Rnf functions, additional engineering strategies were explored to couple the H<sub>2</sub>-producing capacity of MBH and BF-H<sub>2</sub>ase with the generation of reduced end products. In a third engineering design, a soluble hydrogenase that is either NADH-producing (SH2) or NADPH-producing (SH1) was introduced in place of the Rnf\_Na in E1M. The SH2 and SH1 hydrogenases have been reported to consume H<sub>2</sub> under physiological conditions in hyperthermophilic archaea (36). Therefore, the introduction of SH2 and SH1 was proposed to transfer the H<sub>2</sub> production capability (i.e., via MBH and BF-H<sub>2</sub>ase) of *C. bescii* into the production of reducing equivalents for the AdhE mediated ethanol production (Fig. 4C and D). This engineering strategy was proven effective in achieving the maximum ethanol production (Fig. 3). Additional tests were similarly performed on the engineered SH2 and SH1 models against existing genetic backgrounds and redox function deletions. While the deletion of *pyrE* and *ldh* caused only a minor decrease in the maximum ethanol yield, the deletion of MBH and BF-H<sub>2</sub>ase reduced ethanol production by 17% (SH1) and 43% (SH2), and the deletion of MBH and BF-H<sub>2</sub>ase on the genetic background of *pyrE* and *ldh* deletions resulted in a 66% reduction in ethanol yield over the engineered SH2 and SH1 designs (see Data Set S4). Further inspection of the flux variability revealed potential mechanisms leading to the key roles of MBH and BF-H<sub>2</sub>ase in mediating electron flow for ethanol production

in the SH1 and SH2 designs. In the SH1 model, MBH is required for the production of proton gradients that drive ATP production via ATPSYN (see Fig. S4E). In the SH2 model,  $H_2$  production is driven by either MBH or BF- $H_2$ ase, and BF-Nfn has an essential role in mediating redox balance and facilitating NADPH productions (see Fig. S4F).

## DISCUSSION

The extremely thermophilic, cellulolytic, and hemicellulolytic bacterium, *C. bescii*, is a promising platform to produce desired bioproducts from nonpretreated plant biomass through consolidated bioprocessing. This study provides a first systems-level overview of the *C. bescii* metabolism through a complete GEM reconstruction and used GEM as a tool to inform new engineering designs. The simulation of carbohydrate degradation was enabled through the inclusion of 160 carbohydrate utilization genes (Table 1), including extracellular glycosyl hydrolases, polysaccharide lyases, and ABC superfamily transporters responsible for the uptake of oligo- and monosaccharides (described in detail in Rodionov et al. [33]). The reconstruction of carbohydrate utilization pathways was validated in the *C. bescii* GEM based on extensive validations using biochemical and physiological experimental data (Fig. 1; see also Data Set S2 in the supplemental material). Growth phenotypes were validated for 26 substrates, including C5/C6 monosaccharides, disaccharides, and polysaccharides (see Fig. S2).

The *C. bescii* model was applied in a case study of ethanol production to reveal mechanisms related to ethanol yield optimization. Through a minimal network analysis, several core-essential functions were identified for optimizing ethanol production (Fig. 2). Of significance are POR, which provides acetyl-CoA as a precursor to the AdhE-based ethanol production pathway; ATPSYN, which is responsible for the production of ATP; and three enzymes involved in redox-related functions (i.e., MBH, BF- $H_2$ ase, and BF-Nfn) that mediate the interconversion of  $Fd_{red}$ , NADH, and NADPH. Interestingly, the model indicated that the coupling of MBH and the  $H_2$ -consuming direction of BF- $H_2$ ase is an efficient mechanism for producing NADH, which is required for ethanol synthesis using AdhE. In such a coupling, MBH drives the production of  $H_2$  while establishing a proton gradient that can be used by ATPSYN for the production of ATP, and BF- $H_2$ ase utilizes the  $H_2$  produced by MBH to drive the production of NADH along with a partial recovery of  $Fd_{red}$  (Fig. 2).

While the proposed redox coupling strategy between MBH and BF- $H_2$ ase seems effective, it relies on the BF- $H_2$ ase functioning in the  $H_2$ -consuming direction. Since BF- $H_2$ ase has been identified as a primary source of  $H_2$  production under physiological conditions (9), questions were raised about whether the coupling of MBH and BF- $H_2$ ase is feasible in engineered strains of *C. bescii*. By simulating the BF- $H_2$ ase either as reversible reactions or  $H_2$ -producing-only reactions, optimal ethanol production was predicted by the *C. bescii* model under individual or combined perturbations of the ATP/GTP production, redox-related, and genetic background functions (Fig. 3). These analyses further confirmed the significance of redox balance in facilitating maximum ethanol productions in *C. bescii*. More surprisingly, an LDH deletion mutant, which has been used in the design of the E1 strain of *C. bescii*, had limited ethanol production, especially when BF- $H_2$ ase has a  $H_2$ -producing function. This result corroborates with the experimentally measured ethanol production in the E1 strain, MACB1062 (10), and may serve as an additional indication that BF- $H_2$ ase primarily produces  $H_2$  in *C. bescii* (Fig. 3A).

The *C. bescii* model was also used to suggest new engineering strategies for optimizing ethanol yield. Modeling of the existing E1 strain of *C. bescii* showed that the Rnf\_Na function, originally introduced to facilitate the  $Fd_{red}$ -to-NADH conversion (10), was not effective at improving ethanol production due to the lack of a sodium ion gradient (Fig. 3; see also Data Set S4). To circumvent this problem, two new strategies were proposed based on the *C. bescii* model: (i) introduce a membrane bound sodium/proton antiporter, Mrp, to establish the sodium ion gradient required for Rnf\_Na function (Fig. 4A), and (ii) replace the sodium-translocating Rnf\_Na function with a proton-



translocating Rnf\_H function (Fig. 4B). According to the modeling results, both engineering strategies have the potential to reach maximum ethanol production of 140 mM using 11.8 g/liter cellulose as the sole carbon source. This almost doubles the experimentally measured yield (76 mM) obtained from existing designs of the AdhE-mediated ethanol-producing strains (Fig. 3B). The designs of Rnf\_Na/Mrp and Rnf\_H strains are also robust with the existing genetic backgrounds and independent of MBH and BF-H<sub>2</sub>ase coupling, providing alternative solutions in mediating proton pumping and redox balancing. Between the two strategies, the introduction of a Rnf\_H might be less useful for *C. bescii* because Rnf\_H homologs have been found only in mesophilic organisms with maximum growth temperatures near 40°C (37). In contrast, thermostable versions of Mrp are available, for example, from the hyperthermophilic archaea *Pyrococcus furiosus* ( $T_{\max}$  = 103°C) and *Thermococcus onnurineus* ( $T_{\max}$  = 90°C) (38, 39). Therefore, the introduction of an Mrp in the existing E1 design could be an optimal solution that would also enable future engineering of the H<sub>2</sub>-producing MBH and BF-H<sub>2</sub>ase functions into *C. bescii*.

Additional engineering strategies were also explored in the *C. bescii* model through the introduction of soluble hydrogenases, SH1 or SH2 (Fig. 3B and Fig. 4C and D). Candidates for the thermostable cytoplasmic SH1 and SH2 hydrogenases are found in the hyperthermophilic archaeal order *Thermococcales* (36, 40, 41). While carrying a similar capability to reach the maximum ethanol production and a similar robustness to existing genetic backgrounds such as in the Rnf\_Na/Mrp and Rnf\_H designs, the SH1 design is dependent on the proton pumping function of MBH; both SH1 and SH2 designs are dependent on the H<sub>2</sub>-producing functions of MBH or BF-H<sub>2</sub>ase (see Fig. S4E and F).

In summary, a genome-scale metabolic model of *C. bescii* was developed with the goal of optimizing metabolic engineering designs for the production of bio-based chemicals. Using ethanol productions as a case study, several key contributions to ATP production (i.e., ATPSYN), proton pumping (i.e., MBH), and redox processes (i.e., GOR, POR, MBH, BF-H<sub>2</sub>ase, and BF-Nfn) were highlighted that underlie the mechanisms for ethanol yield optimization. The mechanistic understanding acquired from modeling was also applied for the testing of new engineering designs that can be implemented by future experimental studies. The *C. bescii* model, as well as the computational procedures described in this study, will provide a solid foundation for guiding future engineering designs for the production of other desirable bioproducts.

## MATERIALS AND METHODS

**Culture conditions.** *C. bescii* WT and  $\Delta dh$  strains were grown under anaerobic conditions with 80:20 (vol/vol) nitrogen-carbon dioxide headspace at 75°C with shaking at 150 rpm. Monosaccharide model validation experiments were conducted in sealed 100 ml serum bottles with 50 ml of modified DSM 516 media, termed DG25 or DF25 (see Data Set S2). Cells for enzymatic activity assays were grown on glucose in triplicate in 2-liter bottles containing 1 liter of medium. Media were prepared as described previously (14) with the following modifications: growth substrates were adjusted from 5 g/liter to 25 mM and NH<sub>4</sub>Cl was increased from 6.16 to 24.11 mM. To quantify the correlation between biomass and protein yields, the *C. bescii* WT strain was grown in a 20-liter custom fermentor. Growth was maintained at 75°C with gas sparging with 80:20 (vol/vol) nitrogen-carbon dioxide at 1.5 liters/h and agitation by dual Rushton stack impellers at 150 rpm. The same modified medium was used with substrate loading (fructose) increased to 100 mM to avoid carbon-limiting growth. Media were prepared at pH 7.2 (room temperature) and allowed to acidify to pH 6.2 (measured at 75°C) during growth, whereupon the pH was maintained at 6.2 by the automated addition of 10% (wt/vol) sodium bicarbonate solution.

**Enzymatic activity assays.** For preparation of whole-cell extracts, 1-liter cultures were grown with 25 mM glucose as described above, to late exponential phase, and harvested by centrifugation at 6,000 × *g* for 10 min (Beckman Avanti J-30i JLA 10.500 rotor). Cell pellets were flash frozen with liquid nitrogen and stored at −80°C. The remaining steps of extract preparation were conducted anaerobically in a Coy chamber. Cell pellets were thawed and resuspended in 25 mM morpholinepropanesulfonic acid (pH 7.0), containing 1 mg/ml lysozyme at a ratio of 3 ml/g (wet weight) cells. Resuspended cells were incubated at room temperature for 10 min, followed by three 10-s rounds of sonication (Qsonica Q55, amplitude 40). For whole-cell extracts, lysates were washed with excess buffer over a regenerated cellulose centrifugation filter (10-kDa Amicon; EMD Millipore) to remove small molecules including redox cofactors. For cytosolic extracts, lysates were clarified by ultracentrifugation at 100,000 × *g* for 1 h (Beckman L90K ultracentrifuge 70.1Ti rotor), and then clarified lysates (S100) were also washed as described above. Both whole-cell and cytosolic extracts were sealed anaerobically and stored at −80°C until use.



Unless otherwise specified, all enzyme assays were performed at 75°C using the S100 fraction in either 3-ml cuvettes (2 ml reaction volume) sealed by silicone stoppers on an Agilent Technologies Cary 100 UV-Vis spectrophotometer with an Agilent Cary series temperature controller or by high-performance liquid chromatography (HPLC; Agilent 1260 Infinity, YMC Hydrosphere C<sub>18</sub> or Bio-Rad Aminex HPX-87H columns). NAD(P)(H)-dependent activities were measured by changes in absorbance at 340 nm ( $\epsilon = 6.22/\text{mM} \cdot \text{cm}$ ), and methyl viologen (MV)-dependent activities were measured at 600 nm ( $\epsilon = 13.70/\text{mM} \cdot \text{cm}$ ).

Pyruvate oxidoreductase (POR), indole-3-pyruvate oxidoreductase (IOR), 2-ketoisovalerate oxidoreductase (VOR), and 2-ketoglutarate oxidoreductase (KGOR) assays were performed as described previously using 1 mM MV, 5 mM pyruvate, 2.5 mM indole-3-pyruvate, 5 mM 2-ketoisovalerate, and 5 mM 2-ketoglutarate substrates, respectively (42). Malate dehydrogenase (MDH) and lactate dehydrogenase (LDH) activities were measured by the NADH-dependent oxidation of oxaloacetate to malate with 1 mM NADH and 2 mM oxaloacetate, as described previously (43). Pyruvate kinase (PK) reactions contained phosphoenolpyruvate (0.5 mM) with 0.5 mM ADP or GDP (no activity observed with GDP). PK assays were linked to pyruvate oxidoreductase by monitoring the pyruvate-dependent reduction of MV (1 mM) and contained 50 mM EPPS buffer (pH 7.0), 5 mM MgCl<sub>2</sub>, 5 mM KCl, 0.1 mM CoA, and 50  $\mu\text{M}$  thiamine pyrophosphate (TPP) in total volume of 2 ml. Pyruvate phosphate dikinase (PPDK) assays mimicked PK assays, with AMP and GMP replacing ADP and GDP (no activity observed with GMP). Phosphoglycerate kinase (PGK) was measured discontinuously in the reverse direction with 3-phosphoglycerate and ATP or GTP, measuring nucleotide hydrolysis by HPLC, with minor modifications to the described protocol (44). Briefly, S100 extract was incubated at 70°C with 2.5 mM MgCl<sub>2</sub>, 2 mM 3-phosphoglycerate, and 100  $\mu\text{M}$  ATP or GTP in 50 mM phosphate (pH 7.0), in a total volume of 1 ml. After incubation, 0.1 ml of the reaction mixture was removed from the reaction vial at 1-min intervals up to 5 min, cooled on ice to halt the reaction, and injected into the HPLC (hydrosphere C<sub>18</sub> column). Nucleoside monophosphate kinase (NMPK) and nucleoside diphosphate kinase (NDPK) were measured discontinuously by HPLC (hydrosphere C<sub>18</sub> column) by measuring the appearance and disappearance of ATP, ADP, AMP, GTP, GDP, and GMP after incubation for 5 min at 70°C with cell extracts and MgCl<sub>2</sub> in 50 mM phosphate buffer (pH 7) (45, 46).

To avoid complications from the native LDH and MDH activities, the following assays were performed using extracts prepared from the  $\Delta ldh$  strain instead of the WT. Malic enzyme (ME) was assayed in both forward and reverse directions as described previously, with both NADH and NADPH (47). Phosphoenolpyruvate carboxykinase (PCK) was measured in the forward direction with both GDP and ADP by linking oxaloacetate formation with NADH oxidation through the addition of L-MDH from pig heart (Roche, Mannheim, Germany), as described previously (48). Activity was determined discontinuously by measuring oxaloacetate formation by HPLC (Aminex HPX-87H column). No ADP linked PCK activity was detected. Phosphoenolpyruvate carboxylase (PEPC) was measured as described previously, in the manner for PCK described, but without the addition of ADP or GDP (49). Oxaloacetate decarboxylase (OADC) activity was measured in S100 extracts by linking pyruvate formation with NADH oxidation through the addition of L-LDH from rabbit muscle (Roche), as described previously (50). Pyruvate carboxylase (PC) was assayed in S100 similarly to PEPC, by linking oxaloacetate formation from pyruvate to NADH oxidation with pig heart L-MDH (Roche) (51).

**Genome-scale metabolic modeling.** The genome-scale metabolic model of *C. bescii* was reconstructed using the complete genome of strain DSM 6725 (NCBI accession no. NC\_012034.1). Manual curations were performed by combining homology searches to public databases, such as KEGG (52), EggNOG (53), UniProt (54), Pfam (55), BioCyc (55, 56), and TCDB (57), a subsystem-based comparative genomics approach for the curation and analysis of carbon utilization pathways (Rodionov et al. [33]), existing literature on the *C. bescii* physiology and metabolic function, and enzymatic assays performed as a part of this study.

The *C. bescii* model was represented in a YAML format following conventions defined by the PSAMM software package (58, 59). All metabolic simulations were performed with PSAMM version 1.0 using the IBM ILOG CPLEX Optimizer version 12.9.0.0 and Python version 3.7. A biomass equation was formulated based on the composition of cell mass in *C. bescii*, including DNA, RNA, protein, lipids, carbohydrates, ions, trace metabolites, lipoteichoic acids, and cell wall components, with the stoichiometry of the overall biomass equation reflecting the gram composition of individual components in 1 g of cell dry weight (see Data Set S1). Individual synthesis equations were used to represent the biosynthesis of macromolecules and the assembly of ion and trace metabolites, with stoichiometry calibrated to represent the millimolar concentrations of basic building blocks in 1 g of each macromolecule.

Metabolic simulations were performed by representing experimental media settings used for the validation of growth and product yield simulations (see Data Set S2). The overall stoichiometry consistency, formula and charge balance of *C. bescii* model were validated by *masscheck*, *formulacheck*, and *chargecheck* implementation in PSAMM (58, 59), where the biomass-related reactions, compound sinks, exchange reactions, and reactions containing compounds with undefined formula ("R" or "X" group presents in formula) were excluded from the formula and charge check and instead manually inspected to ensure the proper formulation. Additional consistency check of the *C. bescii* model was performed with the standardized genome-scale metabolic model testing software, Memote (31), and compared to other published models (Table 1).

**Model validation.** Two strains of *C. bescii* were used for model validations performed in this study: the wild-type strain DSM 6725 (WT) and the lactate dehydrogenase knockout mutant strain MACB1034 ( $\Delta ldh$ ) (5, 15) (Table 2). A linear conversion was applied to convert the experimental measurements of protein yields to the cell dry weight yields (see Fig. S5 in the supplemental material, and the cell density measurements were converted to cell dry weights based on an estimated weight of  $8.2 \times 10^{-10}$  mg/cell (60). The WT and  $\Delta ldh$  strains were simulated to examine the growth and bio-based chemical production

using monosaccharides (glucose and fructose) as the sole carbon source, and the WT strain was simulated to examine the growth of *C. bescii* using cellulose (Fig. 1). Simulation of biomass yield was performed with the PSAMM *fba* function for the WT strain and with the *moma2* implementation in the PSAMM *genedelete* function for the  $\Delta dh$  strain. Simulation of metabolic products (e.g., lactate, acetate, and pyruvate) were calibrated based on the ratios of their relative concentrations, and the simulation of growth using cellulose was performed by optimizing the biomass yield while calibrating the production of carbon products (e.g., lactate, acetate, and solubilized glucose) based on experimental data (see Data Set S2). Finally, the simulation of growth/no-growth phenotype was performed using the basal medium of DG25 while replacing glucose to a corresponding sole carbon source, and the model predictions were compared to existing growth data of *C. bescii* (5, 34). Detailed simulation settings were provided in Data Set S2.

**Model-based simulation of metabolic engineering strategies for ethanol production.** A modified DSMZ 516 medium (10, 14) was used for the simulation of engineering strategies for ethanol production, with crystalline cellulose (Avicel PH-101) as the sole carbon source. The minimal network analysis was performed using the PSAMM *randomsparse* function. Three sets of 1,000 *randomsparse* simulations were performed with ethanol production constrained to 0% (no-ethanol), 50% (half-maximum), and 99.99% (maximum) of the predicted maximum. The selection of 99.99% as a threshold for the maximum ethanol simulation was to account for potential uncertainties in solving linear programming models using the CPLEX optimizer. A convergence test was performed for each set of simulations to confirm that the classification of core-essential, conditionally essential, and nonessential reactions were stabilized after 1,000 random simulations (see Fig. S3). A detailed description of model configurations for the minimal network analysis was provided in Data Set S3. The evaluation of different engineering designs was similarly performed by configuring the model with corresponding knock-in/knockout functions while optimizing the production of ethanol (see Data Set S4). Flux ranges under diverse ethanol production thresholds was performed using the PSAMM *fva* function for different engineering designs while setting the BF-H<sub>2</sub>ase as either unconstrained or constrained to the H<sub>2</sub>-producing condition. For each simulation 100 steps were taken by varying the ethanol production constraints from zero to the maximum, while the flux variability was examined for the key metabolic reactions in the *C. bescii* model (see Fig. S4).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**DATA SET S1**, XLSX file, 0.3 MB.

**DATA SET S2**, XLSX file, 0.5 MB.

**DATA SET S3**, XLSX file, 0.4 MB.

**DATA SET S4**, XLSX file, 0.4 MB.

**FIG S1**, TIF file, 0.5 MB.

**FIG S2**, PDF file, 0.04 MB.

**FIG S3**, TIF file, 0.8 MB.

**FIG S4**, PDF file, 1.5 MB.

**FIG S5**, TIF file, 1.1 MB.

**TABLE S1**, PDF file, 0.02 MB.

## ACKNOWLEDGMENTS

We thank Keith Dufault-Thompson and Jing Wang for their help on scripting and software developments in this study.

Conceptualization: Y.Z., D.A.R., R.M.K., and M.W.W.A.; research design: W.Z. and Y.Z.; model curation: K.Z., W.Z., D.A.R., and Y.Z.; data analysis: W.Z., Y.Z., and K.Z.; wet lab experiments: G.M.R., D.N.N., J.C., R.G.B., T.N.N.T., R.M.K., and M.W.W.A. W.Z. and Y.Z. wrote the manuscript. All authors reviewed the manuscript.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program, under award DE-SC0019391.

## REFERENCES

- DeLong JP, Burger O. 2015. Socio-economic instability and the scaling of energy use with population size. *PLoS One* 10:e0130547. <https://doi.org/10.1371/journal.pone.0130547>.
- Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. 2008. How biotech can transform biofuels. *Nat Biotechnol* 26:169–172. <https://doi.org/10.1038/nbt0208-169>.
- Yang S-J, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doepke C, Davis M, Westpheling J, Adams MWW. 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. *Appl Environ Microbiol* 75:4762–4769. <https://doi.org/10.1128/AEM.00236-09>.
- Blumer-Schuette SE, Kataeva I, Westpheling J, Adams MW, Kelly RM. 2008. Extremely thermophilic microorganisms for biomass conversion: status

- and prospects. *Curr Opin Biotechnol* 19:210–217. <https://doi.org/10.1016/j.copbio.2008.04.007>.
5. Yang S-J, Kataeva I, Wiegel J, Yin Y, Dam P, Xu Y, Westpheling J, Adams MWW. 2010. Classification of “*Anaerocellum thermophilum*” strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. *Int J Syst Evol Microbiol* 60:2011–2015. <https://doi.org/10.1099/ijs.0.017731-0>.
  6. Zeldes BM, Keller MW, Loder AJ, Straub CT, Adams MWW, Kelly RM. 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Front Microbiol* 6:1209. <https://doi.org/10.3389/fmicb.2015.01209>.
  7. Chung D, Cha M, Guss AM, Westpheling J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proc Natl Acad Sci U S A* 111:8931–8936. <https://doi.org/10.1073/pnas.1402210111>.
  8. Chung D, Cha M, Snyder EN, Elkins JG, Guss AM, Westpheling J. 2015. Cellulosic ethanol production via consolidated bioprocessing at 75°C by engineered *Caldicellulosiruptor bescii*. *Biotechnol Biofuels* 8:163. <https://doi.org/10.1186/s13068-015-0346-4>.
  9. Cha M, Chung D, Westpheling J. 2016. Deletion of a gene cluster for [Ni-Fe] hydrogenase maturation in the anaerobic hyperthermophilic bacterium *Caldicellulosiruptor bescii* identifies its role in hydrogen metabolism. *Appl Microbiol Biotechnol* 100:1823–1831. <https://doi.org/10.1007/s00253-015-7025-z>.
  10. Williams-Rhaesa AM, Rubinstein GM, Scott IM, Lipscomb GL, li Poole FL, Kelly RM, Adams MWW. 2018. Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. *Metab Eng Commun* 7:e00073. <https://doi.org/10.1016/j.mec.2018.e00073>.
  11. Sander K, Chung D, Hyatt D, Westpheling J, Klingeman DM, Rodriguez M, Engle NL, Tschaplinski TJ, Davison BH, Brown SD. 2019. Rex in *Caldicellulosiruptor bescii*: novel regulon members and its effect on the production of ethanol and overflow metabolites. *Microbiologyopen* 8:e00639. <https://doi.org/10.1002/mbo3.639>.
  12. Straub CT, Bing RG, Otten JK, Keller LM, Zeldes BM, Adams MWW, Kelly RM. 2020. Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. *Biotechnol Bioeng* 117:3799–3808. <https://doi.org/10.1002/bit.27529>.
  13. Chung D, Cha M, Farkas J, Westpheling J. 2013. Construction of a stable replicating shuttle vector for *Caldicellulosiruptor* species: use for extending genetic methodologies to other members of this genus. *PLoS One* 8:e62881. <https://doi.org/10.1371/journal.pone.0062881>.
  14. Lipscomb GL, Conway JM, Blumer-Schuetz SE, Kelly RM, Adams MWW. 2016. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. *Appl Environ Microbiol* 82:4421–4428. <https://doi.org/10.1128/AEM.00570-16>.
  15. Williams-Rhaesa AM, Poole FL, Dinsmore JT, Lipscomb GL, Rubinstein GM, Scott IM, Conway JM, Lee LL, Khatibi PA, Kelly RM, Adams MWW. 2017. Genome stability in engineered strains of the extremely thermophilic lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. *Appl Environ Microbiol* 83:e00444-17. <https://doi.org/10.1128/AEM.00444-17>.
  16. Rubinstein GM, Lipscomb GL, Williams-Rhaesa AM, Schut GJ, Kelly RM, Adams MWW. 2020. Engineering the cellulolytic extreme thermophile *Caldicellulosiruptor bescii* to reduce carboxylic acids to alcohols using plant biomass as the energy source. *J Ind Microbiol Biotechnol* 47:585–597. <https://doi.org/10.1007/s10295-020-02299-z>.
  17. Dash S, Khodayari A, Zhou J, Holwerda EK, Olson DG, Lynd LR, Maranas CD. 2017. Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. *Biotechnol Biofuels* 10:108. <https://doi.org/10.1186/s13068-017-0792-2>.
  18. García S, Thompson RA, Giannone RJ, Dash S, Maranas CD, Trinh CT. 2020. Development of a genome-scale metabolic model of *Clostridium thermocellum* and its applications for integration of multi-omics datasets and computational strain design. *Front Bioeng Biotechnol* 8:772. <https://doi.org/10.3389/fbioe.2020.00772>.
  19. Currie DH, Raman B, Gowen CM, Tschaplinski TJ, Land ML, Brown SD, Covalla SF, Klingeman DM, Yang ZK, Engle NL, Johnson CM, Rodriguez M, Joe Shaw A, Kenealy WR, Lynd LR, Fong SS, Mielenz JR, Davison BH, Hogsett DA, Herring CD. 2015. Genome-scale resources for *Thermoanaerobacterium saccharolyticum*. *BMC Syst Biol* 9:1–15. <https://doi.org/10.1186/s12918-015-0159-x>.
  20. Salimi F, Zhuang K, Mahadevan R. 2010. Genome-scale metabolic modeling of a clostridial coculture for consolidated bioprocessing. *Biotechnol J* 5:726–738. <https://doi.org/10.1002/biot.201000159>.
  21. Anwar Z, Gulfranz M, Irshad M. 2014. Agro-industrial lignocellulosic biomass a key to unlock the future bio-energy: a brief review. *J Radiat Res Appl Sci* 7:163–173. <https://doi.org/10.1016/j.jrras.2014.02.003>.
  22. Freier D, Mothershed CP, Wiegel J. 1988. Characterization of *Clostridium thermocellum* JW20. *Appl Environ Microbiol* 54:204–211. <https://doi.org/10.1128/AEM.54.1.204-211.1988>.
  23. Currie DH, Guss AM, Herring CD, Giannone RJ, Johnson CM, Lankford PK, Brown SD, Hettich RL, Lynd LR. 2014. Profile of secreted hydrolases, associated proteins, and SlpA in *Thermoanaerobacterium saccharolyticum* during the degradation of hemicellulose. *Appl Environ Microbiol* 80:5001–5011. <https://doi.org/10.1128/AEM.00998-14>.
  24. Stainthorpe AC, Williams RAD. 1988. Isolation and properties of *Clostridium thermocellum* from Icelandic hot springs. *Int J Syst Evol Microbiol* 38:119–121. <https://doi.org/10.1099/00207713-38-1-119>.
  25. Scott IM, Rubinstein GM, Poole FL, Lipscomb GL, Schut GJ, Williams-Rhaesa AM, Stevenson DM, Amador-Noguez D, Kelly RM, Adams MWW. 2019. The thermophilic biomass-degrading bacterium *Caldicellulosiruptor bescii* utilizes two enzymes to oxidize glyceraldehyde 3-phosphate during glycolysis. *J Biol Chem* 294:9995–10005. <https://doi.org/10.1074/jbc.RA118.007120>.
  26. Cha M, Chung D, Elkins JG, Guss AM, Westpheling J. 2013. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. *Biotechnol Biofuels* 6:85. <https://doi.org/10.1186/1754-6834-6-85>.
  27. Lee LL, Crosby JR, Rubinstein GM, Laemthong T, Bing RG, Straub CT, Adams MWW, Kelly RM. 2020. The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in “Caldi World.” *Extremophiles* 24:1–15. <https://doi.org/10.1007/s00792-019-01116-5>.
  28. Buckel W, Thauer RK. 2018. Flavin-based electron bifurcation, a new mechanism of biological energy coupling. *Chem Rev* 118:3862–3886. <https://doi.org/10.1021/acs.chemrev.7b00707>.
  29. Schut GJ, Adams MWW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* 191:4451–4457. <https://doi.org/10.1128/JB.01582-08>.
  30. Wang S, Huang H, Moll J, Thauer RK. 2010. NADP<sup>+</sup> reduction with reduced ferredoxin and NADP<sup>+</sup> reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*. *J Bacteriol* 192:5115–5123. <https://doi.org/10.1128/JB.00612-10>.
  31. Lieven C, Beber ME, Olivier BG, Bergmann FT, Ataman M, Babaei P, Bartell JA, Blank LM, Chauhan S, Correia K, Diener C, Dräger A, Ebert BE, Edirisinghe JN, Faria JP, Feist AM, Fengos G, Fleming RMT, García-Jiménez B, Hatzimanikatis V, van Helvoirt W, Henry CS, Hermjakob H, Herrgård MJ, Kaafarani A, Kim HU, King Z, Klamt S, Klipp E, Koehorst JJ, König M, Lakshmanan M, Lee D-Y, Lee SY, Lee S, Lewis NE, Liu F, Ma H, Machado D, Mahadevan R, Maia P, Mardinoglu A, Medlock GL, Monk JM, Nielsen J, Nielsen LK, Nogales J, Nookaew I, Palsson BO, Papin JA, et al. 2020. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* 38:272–276. <https://doi.org/10.1038/s41587-020-0446-y>.
  32. Nguyen DMN, Schut GJ, Zadovnyy OA, Tokmina-Lukaszewska M, Poudel S, Lipscomb GL, Adams LA, Dinsmore JT, Nixon WJ, Boyd ES, Bothner B, Peters JW, Adams MWW. 2017. Two functional distinct NADP-dependent ferredoxin oxidoreductases maintain the primary redox balance of *Pyrococcus furiosus*. *J Biol Chem* 292:14603–14616. <https://doi.org/10.1074/jbc.M117.794172>.
  33. Rodionov DA, Rodionova IA, Rodionov VA, Arzamasov AA, Zhang K, Rubinstein GM, Tanwee TNN, Bing RG, Crosby JR, Nookaew I, Basen M, Brown SD, Wilson CM, Klingeman DM, Poole FL, II, Zhang Y, Kelly RM, Adams MWW. 2021. Transcriptional regulation of plant biomass degradation and carbohydrate utilization genes in the extreme thermophile *Caldicellulosiruptor bescii*. *mSystems* 6:e01345-20. <https://doi.org/10.1128/mSystems.01345-20>.
  34. Ingvadottir EM, Scully SM, Orlýgsson J. 2017. Evaluation of the genus of *Caldicellulosiruptor* for production of 1,2-propanediol from methylpentoses. *Anaerobe* 47:86–88. <https://doi.org/10.1016/j.anaerobe.2017.04.015>.
  35. Basen M, Rhaesa AM, Kataeva I, Prybol CJ, Scott IM, Poole FL, Adams MWW. 2014. Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium *Caldicellulosiruptor bescii*. *Bioresour Technol* 152:384–392. <https://doi.org/10.1016/j.biortech.2013.11.024>.
  36. Wu C-H, McTernan PM, Walter ME, Adams MWW. 2015. Production and application of a soluble hydrogenase from *Pyrococcus furiosus*. *Archaea* 2015:912582. <https://doi.org/10.1155/2015/912582>.

37. Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR. 2012. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD<sup>+</sup> oxidoreductase essential for autotrophic growth. *mBio* 4:e00406-12. <https://doi.org/10.1128/mBio.00406-12>.
38. McTernan PM, Chandrayan SK, Wu C-H, Vaccaro BJ, Lancaster WA, Adams MWW. 2015. Engineering the respiratory membrane-bound hydrogenase of the hyperthermophilic archaeon *Pyrococcus furiosus* and characterization of the catalytically active cytoplasmic subcomplex. *Protein Eng Des Sel* 28:1–8. <https://doi.org/10.1093/protein/gzu051>.
39. Ito M, Morino M, Krulwich TA. 2017. Mrp antiporters have important roles in diverse bacteria and archaea. *Front Microbiol* 8:2325. <https://doi.org/10.3389/fmicb.2017.02325>.
40. Lipscomb GL, Schut GJ, Scott RA, Adams MWW. 2017. SurR is a master regulator of the primary electron flow pathways in the order *Thermococcales*. *Mol Microbiol* 104:869–881. <https://doi.org/10.1111/mmi.13668>.
41. Schut GJ, Nixon WJ, Lipscomb GL, Scott RA, Adams MWW. 2012. Mutational analyses of the enzymes involved in the metabolism of hydrogen by the hyperthermophilic archaeon *Pyrococcus furiosus*. *Front Microbiol* 3:163.
42. Schut GJ, Menon AL, Adams MWW. 2001. 2-Keto acid oxidoreductases from *Pyrococcus furiosus* and *Thermococcus litoralis*. *Methods Enzymol* 331:144–158.
43. Tayeh MA, Madigan MT. 1987. Malate dehydrogenase in phototrophic purple bacteria: purification, molecular weight, and quaternary structure. *J Bacteriol* 169:4196–4202. <https://doi.org/10.1128/jb.169.9.4196-4202.1987>.
44. García-Díaz M, Canales J, Sillero MA, Sillero A, Cameselle JC. 1989. Phosphoglycerate kinase from yeast synthesizes guanosine 5'-tetraphosphate. *Biochem Int* 19:1253–1264.
45. Shi X-Z, Zhao X-F, Wang J-X. 2008. Molecular cloning and analysis of function of nucleoside diphosphate kinase (NDPK) from the scallop *Chlamys farreri*. *Biochemistry (Mosc)* 73:686–692. <https://doi.org/10.1134/s0006297908060096>.
46. Kandeel M, Kitade Y. 2008. Molecular characterization, heterologous expression and kinetic analysis of recombinant plasmodium falciparum thymidylate kinase. *J Biochem* 144:245–250. <https://doi.org/10.1093/jb/mvn062>.
47. Suye S, Yokoyama S. 1985. NADPH production from NADP<sup>+</sup> using malic enzyme of *Achromobacter parvulus* IFO-13182. *Enzyme Microb Technol* 7:418–424. [https://doi.org/10.1016/0141-0229\(85\)90040-7](https://doi.org/10.1016/0141-0229(85)90040-7).
48. Petrescu I, Bojan O, Saied M, Bârzu O, Schmidt FH, Kühnle HF. 1979. Determination of phosphoenolpyruvate carboxykinase activity with deoxyguanosine 5'-diphosphate as nucleotide substrate. *Anal Biochem* 96:279–281. [https://doi.org/10.1016/0003-2697\(79\)90582-7](https://doi.org/10.1016/0003-2697(79)90582-7).
49. Meyer CR, Rustin P, Wedding RT. 1988. A simple and accurate spectrophotometric assay for phosphoenolpyruvate carboxylase activity. *Plant Physiol* 86:325–328. <https://doi.org/10.1104/pp.86.2.325>.
50. Dimroth P. 1981. Characterization of a membrane-bound biotin-containing enzyme: oxaloacetate decarboxylase from *Klebsiella aerogenes*. *Eur J Biochem* 115:353–358. <https://doi.org/10.1111/j.1432-1033.1981.tb05245.x>.
51. Payne J, Morris JG. 1969. Pyruvate carboxylase in *Rhodospseudomonas spheroides*. *Microbiology* 59:97–101. <https://doi.org/10.1099/00221287-59-1-97>.
52. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases, and drugs. *Nucleic Acids Res* 45:D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
53. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 47:D309–D314. <https://doi.org/10.1093/nar/gky1085>.
54. UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–D515. <https://doi.org/10.1093/nar/gky1049>.
55. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>.
56. Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, Ong WK, Paley SM, Subhraveti P. 2019. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform* 20:1085–1093. <https://doi.org/10.1093/bib/bbx085>.
57. Saier MH, Jr, Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The Transporter Classification Database (TCDB): recent advances. *Nucleic Acids Res* 44:D372–D379. <https://doi.org/10.1093/nar/gkv1103>.
58. Steffensen JL, Dufault-Thompson K, Zhang Y. 2016. PSAMM: a Portable System for the Analysis of Metabolic Models. *PLoS Comput Biol* 12:e1004732. <https://doi.org/10.1371/journal.pcbi.1004732>.
59. Dufault-Thompson K, Steffensen JL, Zhang Y. 2018. Using PSAMM for the curation and analysis of genome-scale metabolic models. *Methods Mol Biol* 1716:131–150. [https://doi.org/10.1007/978-1-4939-7528-0\\_6](https://doi.org/10.1007/978-1-4939-7528-0_6).
60. Zurawski JV, Conway JM, Lee LL, Simpson HJ, Izquierdo JA, Blumer-Schuette S, Nookaew I, Adams MWW, Kelly RM. 2015. Comparative analysis of extremely thermophilic *Caldicellulosiruptor* species reveals common and unique cellular strategies for plant biomass utilization. *Appl Environ Microbiol* 81:7159–7170. <https://doi.org/10.1128/AEM.01622-15>.
61. McBEE RH. 1954. The characteristics of *Clostridium thermocellum*. *J Bacteriol* 67:505–506. <https://doi.org/10.1128/JB.67.4.505-506.1954>.
62. Lee Y-E, Jain MK, Lee C, Gregory Zeikus J. 1993. Taxonomic distinction of saccharolytic thermophilic anaerobes: description of *Thermoanaerobacterium xylanolyticum* gen. nov., sp. nov., and *Thermoanaerobacterium saccharolyticum* gen. nov., sp. nov.; reclassification of *Thermoanaerobium brockii*, *Clostridium thermosulfurogenes*, and *Clostridium thermohydrosulfuricum* E100-69 as *Thermoanaerobacter brockii* comb. nov., *Thermoanaerobacterium thermosulfurigenes* comb. nov., and *Thermoanaerobacter thermohydrosulfuricus* comb. nov., respectively; and transfer of *Clostridium thermohydrosulfuricum* 39E to *Thermoanaerobacter ethanolicus*. *Int J Syst Evol Microbiol* 43:41–51.