5-2018

Characterization and Metabolic Engineering of Transcription Factors and Redox Dynamics in Candidate Consolidated Bioprocessing Biocatalysts

Kyle Benjamin Sander
University of Tennessee, ksande26@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Kyle Benjamin Sander entitled "Characterization and Metabolic Engineering of Transcription Factors and Redox Dynamics in Candidate Consolidated Bioprocessing Biocatalysts." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Chemical Engineering.

Brian H. Davison, Major Professor

We have read this dissertation and recommend its acceptance:

Eric T. Boder, Steven D. Brown, Paul D. Frymier

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Characterization and Metabolic Engineering of Transcription Factors and Redox Dynamics in Candidate Consolidated Bioprocessing Biocatalysts

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Kyle Benjamin Sander

May 2018

Abstract

This thesis studies the metabolic engineering of candidate consolidated bioprocessing biocatalyst microorganisms through targeting regulatory genes, with an emphasis on redox metabolism. Consolidated bioprocessing is the single-step hydrolysis and conversion of lignocellulosic material to biofuels. The biocatalysts considered are *Clostridium thermocellum* and *Caldicellulosiruptor bescii,* and the primary product of interest is ethanol. Both organisms are thermophilic anaerobic bacteria which encode and express genes that facilitate the deconstruction and solubilization of lignocellulose into fermentable carbohydrates. Furthermore, these organisms ferment these carbohydrates into ethanol, organic acids, as well as other fermentation products. We seek to improve redox metabolism and osmotolerance in these organsisms toward a biorefining objective goal of engineering a biocatalyst capable of facilitating economically viable consolidated bioprocessing.

Expression profiling, transcription factor regulon mapping, genetic engineering, and analytical fermentation were approaches employed to assay and understand which specific traits can be beneficially altered. The traits sought to be altered are characteristically complex, co-opting many cellular sub-processes to enable a molecular mechanism resulting in an observable trait. Such traits are notoriously difficult not only to understand, but to alter through classical metabolic engineering. Instead, the possibility of making system-wide changes through a minimal number of genetic alterations to methodically selected and/or screened regulatory genes was investigated.

Active redox-dependent systems were characterized in both bacteria, many of which are controlled by the global redox-state sensing transcription factor Rex. Eliminating Rex control over gene expression in *C. bescii* resulted in a more reduced intracellular redox state, and ultimately drives increased ethanol synthesis. A method for quantifying important redox metabolites intracellularly is also adopted and validated for use with *C. thermocellum*. This approach was extended to less characterized gene targets

and, arguably, even more complex traits. Screening of single-gene deletion mutants identified two strains of *C. bescii* showing phenotypic growth differences in elevated osmolarity conditions. One strain housed a deletion of the *fapR* gene, while the other a deletion of the *fruR*/*cra* gene. Characterizing these transcription factors and their regulons elucidates mechanisms which this organism uses to facilitate survival at elevated osmolarities. We are also able to construct genetic variants in *C. bescii* which are substantially more osmotolerant than native strains, highlighting the usefulness of these genes as targets and the applicability, and important considerations, of our metabolic engineering approach.

Table of Contents

List of Tables

Tables Embedded in Text

Tables in Appendix

List of Figures

Figures Embedded in Text

Figures in Appendix

List of Attachments

Attachment 1. Normalized probe hybridization intensity for methyl viologen and hydrogen peroxide exposed *C. thermocellum* DSM 1313. Filename: attachment_1_Sander.xlsx

Attachment 2. RNAseq differential expression for strain JWCB005Δ*rex* at early, mid, and late log phase of growth. Filename: attachment_2_Sander.xlsx

Attachment 3. RNAseq differential expression for strains JWCB005ΔB5X54_RS01260 and JWCB005ΔB5X54_RS06355 grown in elevated osmolarities of glycerol, sodium chloride, and glucose. Filename: attachment_3_Sander.xlsx

Chapter 1: Introduction and Motivation

Bioenergy needs in the transportation sector are currently met almost entirely with bioethanol [1]. Currently, the commodity fuel ethanol demand in the United States is met through a fermentation process with the fermentable carbohydrates originating from corn starch [2], making it a 'first-generation' biofuel. The production rate of fuel ethanol in the United States has climbed steadily since 2011, regularly hovering above 1 million gallons per day since the beginning of 2017 [3].

While bioethanol currently makes up a small portion of total transportation fuel needs in the United States, it will likely [4] constitute the bulk of the 36 billion gallons of total biofuels projected to be produced through 2022, as mandated by the Renewable Fuel Standard [5]. The Renewable Fuel Standard is a compulsory and integral part of larger legislation (the Energy Independence and Security Act of 2007) that aggregates and expresses the policy objectives in the United States to reduce our carbon footprint, become more energy independent and diverse, and economically bolster rural agricultural economies. As of 2016, only 312 million gallons of 'advanced' non-corn starch bioethanol was produced [5], far below the annual volume mandated for 2016 by the original Renewable Fuel Standard (~9 Billion gallons). It is unlikely that processes, abiotic or biotic, to generate 'drop-in' alternatives to ethanol will emerge to fill this gap, as these are subject to the same economic and technological constraints as bioethanol, and are far less technologically mature.

Though there is hope for replacing existing ethanol demand with more sustainable ethanol made using more advanced processes, the demand is unlikely to increase without the removal of the restrictions placed on its volumetric blending with gasoline, the so-called 'blend wall,' in the United States. More mature ethanol markets, such as Brazil which relies largely on ethanol to meet liquid transportation fuel demand, do not face this artificial ceiling on ethanol demand as their infrastructure is amenable to much higher ethanol blends.

Since 2007, when the Energy Independence and Security Act was enacted, the United States has dramatically increased efforts to recover domestic petroleum fuels, and the retail price of gasoline has

fallen nearly in half. These economic factors inhibit the development and deployment of sustainable alternative fuels. By 2075, under the International Energy Agency's '2 degree' scenario, biofuels are projected to make up ~30% of all transportation fuel energy in the world, while petroleum based fuels are projected to comprise < 10% [6]. No long-term transportation sector projections ever foresee electricity providing more than 50% of world-wide transportation sector energy [4]. These projections and facts illustrate the need for fuel ethanol. Furthermore, they negate the notion that bioethanol is a temporary 'stop gap' to other technologies, or that it is an 'optional' piece of the United States' sustainability strategy.

Consolidated bioprocessing (CBP), a process utilizing a single microbial biocatalyst for both deconstruction of lignocellulosic materials to soluble components and subsequent conversion to ethanol, is the least expensive biological route to ethanol production on the basis of capital, operating, and input costs [7]. Lignocellulosic material (such as high-yielding switchgrass) is the only bioenergy feedstock currently being considered that can potentially be grown on lands that cannot support food crop production [8]. While sustainably produced lignocellulosic material would make an ideal bioenergy feedstock, the biotechnological conversion of this material to fuel ethanol suffers from low product yields/productivities, and incomplete feedstock utilization. Increases in both of these metrics are needed to decrease the total costs ($/annual L of ethanol) of CBP produced ethanol. Abiotic processes (gasification, pyrolysis) for fuels using lignocellulosic material suffer the same cost constraints as CBP, and large scale deployment of these processes is hindered currently by prohibitively high capital costs [8]. As a result, most ethanol refineries, old and new, operate biological processes [9].

In addition to the technological challenges of the conversion process, another primary challenge to large-scale deployment of consolidated bioprocessing are the logistical challenges of obtaining sufficient lignocellulosic material to support a CBP process. Current research is addressing these challenges by exploring cost drivers of feedstock delivery such as optimizing the placement of biorefineries with

respect to their distance from places where feedstock material is grown [10] and utilizing a plurality of lignocellulosic materials to dampen the effects of supply disruptions [11].

The key to an effective CBP process is the biocatalyst, which must simultaneously degrade lignocellulose to their soluble, fermentable carbohydrate constituents, and ferment the carbohydrate fraction to make bioethanol. Two promising CBP biocatalysts are *Clostridium thermocellum* and *Caldicellulsiruptor bescii*. Both organisms are strict anaerobic thermophiles, culture conditions which inherently occlude contamination by common biotechnology contaminant microorganisms. *Caldicellulosiruptor* is one among thermophilic CBP candidates in that it can hydrolyze non-pretreated switchgrass and metabolize both the resulting pentose and hexose sugars simultaneously [12-14]. *Clostridium thermocellum* also hydrolyses lignocellulosic biomass, though does not naturally catabolize five-carbon sugar monomers. Both organisms encode a host of cellulolytic enzymes which facilitate direct solubilization of lignocellulosic biomass to their fermentable carbohydrate components [15]. Recently, strains *of C. bescii* have been engineered to produce ethanol, though at yields far below theoretical maximum [16]. *Clostridium thermocellum* produces ethanol naturally, though alongside other fermentation products. These native traits make these organisms ideal CBP candidate biocatalysts, though their native performance falls far short of objectives identified for cost-effective, industrial CBP [17, 18]. As such methods for genetically modifying these organisms were developed [19, 20], and efforts toward improving the bioperformance of these organisms has made great strides toward CBP performance objectives.

Previous and ongoing metabolic engineering these CBP candidate biocatalysts has largely targeted well-known enzymes belonging to well-understood pathways. Modulating carbon flux using this approach has afforded large gains in ethanol yield in CBP organisms [16, 21, 22], but is limited to the small number of pathways for which we have sufficient knowledge [9]. The diminishing returns realized from successive modifications using this approach suggest these efforts alone will not suffice to meet all

objectives for a commercial CBP catalyst. There are very few instances [23] of improving more complex behaviors in CBP biocatalysts, such as lignocellulosic biomass deconstruction (of which much of the mechanistic process is unknown [24-26]), or tolerance to stressors. Efforts to improve these CBP performance shortcomings rooted in more complex cellular processes, have not previously been undertaken, but remain as imminent obstacles to the industrial utilization of the organisms in a CBP setting.

Common to the works described in this thesis is the aim to increase the performance of these CBP biocatalysts, *Caldicellulosiruptor bescii* and *Clostridium thermocellum*, which we have endeavored to do so by exploring and metabolic engineering regulatory gene targets in these organisms. This is a proven strategy, particularly when faced with the task of improving performance metrics associated with complex and poorly understood traits, as we were. Transcription factors have been effectively used as metabolic engineering targets many times over towards attaining large bioprocessing performance gains. *Clostridium acetobutylicum* final ethanol titer increased ~200% upon deleting its rex redox-sensing transcription factor [27]. Over-expressing the FadR fatty acid metabolism regulator in *E. coli* increased total fatty acid production 5-7.5 fold, and was shown to be more effective than overexpressing individual enzymes responsible for fatty acid production [28]. Ethanol tolerance, but not productivity, was improved in *S. cerevisiae* by iteratively evolving and screening strains carrying mutated variants of the *spt15* transcription factor [29], showing that engineering transcription factors and regulators can have a large impact on performance metrics associated with complex and poorly understood cellular processes.

Facilitating more effective sugar-to-ethanol processes within these organisms will increase ethanol yield and productivity. Biological ethanol synthesis in both organisms requires the cofactor NADH in stoichiometric quantities to be provided by a host's native redox metabolism. Expressing an exogenous alcohol dehydrogenase shifted nearly 70% of product carbon flux toward ethanol without any

subsequent pathway optimization [16]. Low overall substrate conversion and low final ethanol product titer suggest further improving ethanol synthesis efficacy through engineering of redox systems in *C. bescii* is possible. The redox systems of *Clostridium theromcellum* have been studied extensively and redox modulation, either through end-product forcing [30], genetic modification [21, 31-33], or other alterations to growth conditions [34], effects the relative and absolute amount of ethanol synthesized. The native redox conditions, and the availability of NADH for ethanol synthesis, appear to be limiting ethanol production in *C. thermocellum* [35], as many of these genetic modifications which affect redox system components directly increase ethanol yield considerably. While improvements in ethanol synthesis have been realized in *C. thermocellum*, its redox systems, outside of glycolytic and fermentative pathways, remain largely unexplored. While these alterations did not suffice to achieve ethanol yield and productivity objectives, it was clear that redox-imbalance, as evidenced by overflow metabolite production [31, 34] and incomplete carbon balances, was still limiting ethanol synthesis.

In the second chapter, we studied redox dynamics with an emphasis on identifying previously unknown active redox nodes in *C. thermocellum* metabolism, and how they interact with overall redox metabolism to effect ethanol fermentation. We chose to focus on the transcriptomic response to redox perturbation as this would allow for a comprehensive view of redox dynamical systems in *C. thermocellum*, including the regulatory networks governing redox state response. Intracellular redox state was perturbed by adding known redox-active chemicals to steady-state growing cultures, and transcriptomic profiles were assessed during transient periods, as well as long after cultures had achieved perturbed steady state conditions. We perturbed growing cultures with methyl viologen, which served to decrease whole-cell measured redox potential, and, separately, with hydrogen peroxide, which increased whole-cell measured redox potential. Toward informing future redox-system metabolic engineering efforts, we elucidated active redox systems that were coordinately varying in response to redox conditions. Subsequent efforts have capitalized on these elucidations through the

elimination of a glutamine synthetase gene, which resulted in a 53% increase in ethanol yield and a 44% decrease in secreted soluble amino acids, an unwanted overflow metabolism product and carbon sink [36].

Global redox cofactor metabolites, NAD(H) and NADP(H), are charge carrier molecules through which many redox reactions are facilitated within the cell, including the final two steps of ethanol synthesis catalyzed by the AdhE enzyme in *C. thermocellum* [37], and in *C. bescii* engineered to produce ethanol [16]. Redox reactions can be facilitated by other recycled cofactor molecules and small proteins, though the most important redox charge carrier molecules in *C. thermocellum* and *C. bescii* are NAD(H), NADP(H) and the hydrogenase/PFOR-associated ferredoxin(s) [12, 34, 38-40]. NAD(H) and NADP(H) both form redox couples with their oxidized counterparts which have similar characteristic standard biological redox potentials (E_o'), while ferredoxin redox couple standard potentials are typically much lower [41]. As such these two typically catalyze different reactions within the cell metabolism. Measuring the intracellular concentrations of these molecules is imperative to gain a determining sense of intracellular redox conditions within a cell, and measuring the concentrations of these molecules separately allows for the ever-important estimation of thermodynamic forcing on important glycolytic and fermentative reactions. These measurements have historically proven to be challenging measurements to make and validate. Of similar importance is the measurement of energy-transferring adenylate cofactors; ATP, ADP and AMP. These molecules too have proven very difficult to measure. Intrinsic measures of state for these molecules (the adenylate energy charge for adenylates and redox potential for the nicotinamides) taken from the same organism grown in similar conditions can span the range of possible values. As these intrinsic parameters are known to be tightly controlled by the cells metabolism [42, 43], it is unlikely that this variation reliably represents biological systems being interrogated. Such wide variations, if they were indeed biological, in these fundamental intrinsic values would greatly disrupt cellular metabolic processes. It is more likely that many of these results are the

spurious consequences of experimentally and/or procedurally induced artifacts and do not represent the true biological state of the cells from which the measurements were taken. The reason these metabolites are notoriously difficult to reliably extract and quantify is that they are exceedingly labile, even under the most careful of handling and processing conditions [44, 45]. Previously used methods do not account for this and employ unnecessarily harsh chemicals and protocols, leaving these labile metabolites susceptible to degradation and/or conversion to their more stable conversion counterparts (NAD(P)H to NAD(P)+ and ATP to ADP and AMP). Toward obtaining robust and validated measured intracellular concentrations of these important intracellular metabolites, the second work of this thesis adapted an extraction and detection methodology previously developed and validated for use with *E. coli* cell biomass [45] to be used with the organisms of interest to this work and to the BioEnergy Science Center. This method was chosen as the basis for our adaptation primarily because of its careful attention to preserving labile metabolites, while also ensuring replete extraction of metabolites from cell biomass. Toward this end, we adapt and modified our method while paying close attention to one strong indicator of labile metabolite degradation; the adenylate energy charge (AEC). This parameter is tightly regulated in cellular tissues [42], and decreases in the most labile adenylate, ATP, are readily manifest in a decreases in the AEC. In achieving acceptably high, physiologically relevant AEC values, we arrive at an adapted method suitable for extracting and quantifying these metabolites from *C. thermocellum* and another thermophilic ethanologen often studied in co-culture with *C. thermocellum*, *Thermoanaerobacterium saccharolyticum*.

Toward studying the potential of engineering redox systems in these organisms, beyond what has already been accomplished in *C. thermocellum* through engineering catalytic enzymes, we sought to affect the redox-driven production of ethanol through modulating redox systems at the level of transcription, and demonstrate that redox state could be altered and ethanol bioproduction increased through the engineering of regulatory genes. We selected the *rex* gene as the regulatory target. The

Rex protein is a global transcription factor that regulates redox state within the cell. Its presence, structure, function, and regulon are largely conserved across gram positive bacteria [46]. Rex regulates gene expression in response to NADH/NAD+ levels within the cell. Binding of NADH disrupts Rex's ability to bind DNA and repress transcription. Rex in its apo form is free to bind to Rex-specific binding sites located in the promoter regions of genes and allosterically repress transcription. Rex has been shown to effectively regulate transcription in response to the NADH/NAD+ ratio, over total pool sizes many orders of magnitude [47], effectively regulating redox state rather than NADH concentration alone. Additionally, Rex regulon targets stretch beyond those genes which are NAD(H) dependent, and include genes which encode redox proteins dependent on other redox carrying cofactors [46, 48]. Combined, Rex indirectly serves as a master regulator of redox state in general for the cell. For this reason, the study of Rex and its regulon served two purposes. The first was to elucidate the, largely unknown, network of genes involved in redox state regulation and modulation within *C. bescii*. The second was to inform an engineering strategy aimed at poising redox systems in *C. bescii* toward more effective ethanol synthesis.

As a final and culminating effort in this work, we aimed to assess and leverage bioperformance impacts associated with ten different transcription factor targets in *C. bescii*. Little is known of the regulatory network space of *C. bescii*, and of the genus *Caldicellulosiruptor* in general. Due to this lack of previous knowledge, candidate gene targets were selected from primarily global differential expression profiles collected under growth on switchgrass (a candidate CBP feedstock) [49], and from a close relative *Caldicellulosiruptor saccharolyticus* (with which *C. bescii* shares 93% nucleotide similarity) while being sparged with hydrogen [50], a potent redox stress. Single-gene deletion mutants were generated for each of the ten candidate gene targets, and these strains were assayed for performance differences while grown under eight different CBP relevant conditions. Defined growth media and methods for a

variety of assay conditions [12, 51, 52], as well as a versatile and robust genetic system [20], developed and optimized previously for *C. bescii*, were used extensively in this work.

DNA transformation was initially made possible in *C. bescii* through identifying and mitigating the effects of a potent native restriction endonuclease [20]. Eliminating its activity enabled eletrotransformation of DNA into *C. bescii* cell biomass. Subsequently, methods were developed to delete genes and introduce genes onto the chromosome using a non-replicating suicide vector and homologous recombination [53]. A replicating shuttle vector [54], as well as an additional selection marker conferring resistance to kanamycin [55] were subsequently developed and are both used in these works.

From the traits screened, two genes were shown to effect osmotolerance. Elevated concentrations of osmolytes is one of the many cytological stresses biocatalysts encounter during CBP. For example, basal *C. bescii* media containing 50 g/L of ethanol (a nominal target final concentration for a cost-effective CBP process) contains >1,300 mOsm/L. Natively, *C. bescii* is relatively sensitive to higher osmolarities, displaying severe growth defects at 160 mM synthesized acetate [12], 160 mM added MOPS buffer [56], and ~200 mOsm/L from added NaCl [51], far below what is needed for industrial CBP. We again assessed the regulons of these two genes using global transcription profiling under high osmolarity loadings of different osmolytes, allowing us to predict the regulons of these two transcription factors and hypothesize the modes of action by which they effect osmotolerance. We are following up by combining these two mutations in a single strain to generate a highly osmotolerant strain of *C. bescii*.

The exploration and modulation of regulatory genes has not only broadened our knowledge of previously unknown redox systems in these organisms, but allowed for further bioproductivity improvements in these two organisms. It is shown herein that CBP bioperformance gains rooted in complex traits, such as the redox state of the cell and tolerance to elevated osmolarities, can be realized through exploration and metabolic engineering of regulatory gene targets in CBP organisms.

Chapter 2: *Clostridium thermocellum* DSM 1313 Transcriptional Responses to Redox Perturbation

Kyle Sander[1,2,3], Charlotte M. Wilson[2,4], Miguel Rodriguez, Jr. [2,4], Dawn M. Klingeman[2,4], Thomas Rydzak[2,4], Brian H. Davison[1,2,4], and Steven D. Brown[1,2,4,§].

[1]Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, Tennessee 37996, USA

[2]BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

[3]Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, Tennessee 37996, USA

[4]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, USA

[§]Corresponding author

The following work was published in the journal Biotechnology for Biofuels in 2015. It is presented here as it appears in the journal, formatted to fit applicable requirements for inclusion in this thesis. The citation for this work, as published is:

Sander, K., et Al., *Clostridium thermocellum DSM 1313 transcriptional responses to redox perturbation.* Biotechnol Biofuel, 2015. **8**(1): p. 211.

This work is the initial, preliminary exploration of redox metabolism in the thermophilic CBP candidate organisms studied throughout this thesis. In this work, we elaborate the cellular response to a discrete and defined redox stress in *C. thermocellum* giving initial insight into the coordinated responses employed when challenged with a changing redox environment. We were able to observe a difference in ethanol yield during the prescribed redox treatment, meeting a primary objective of observing active and dynamic redox systems which are closely linked to the synthesis of ethanol, and which impact the ethanol product yield. Furthermore, we observed these differences at the level of transcription, leading us to the realization that redox dynamics are affected by transcription-level changes, and not being enabled in other ways (e.g. metabolically or within the proteome for example). Observing these systems not only informed future metabolic engineering efforts by others, but led us to pursue redox-sensing transcription factors, and transcriptional regulatory machinery in general in future efforts to increase bioperformance in these CBP candidate organisms.

Kyle Sander led efforts in designing the study, forming the hypotheses to be addressed, designed and carried out all experiments, collected and analyzed samples and data, assisted in microarray cDNA preparation, hybridization and analysis, and drafted the manuscript and figures. Charlotte M. Wilson assisted in designing experiments, preparing cDNA, hybridizing cDNA and analyzing expression profiling

results, and assisted in editing the manuscript. Miguel Rodriguez Jr. assisted in experimental design, and quantified fermentation products in samples taken during experiments. Dawn M. Klingeman prepared cDNA from cell biomass samples, hybridized cDNA, and assisted in analyzing microarray results. Thomas Rydzak provided thoughtful, thorough, and scientifically critical reviews of the article prior to submission. Brian H. Davison and Steven D. Brown assisted in hypothesis formulation, experimental design, data analysis, and editing the manuscript for submission.

Abstract

Background

Clostridium thermocellum is a promising consolidated bioprocessing candidate organism capable of directly converting lignocellulosic biomass to ethanol. Current ethanol yields, productivities and growth inhibition are industrial deployment impediments for commodity fuel production by this bacterium. Redox imbalance under certain conditions and in engineered strains may contribute to incomplete substrate utilization and may direct fermentation products to undesirable overflow metabolites. Toward a better understanding of redox metabolism in *C. thermocellum*, we established continuous growth conditions and analyzed global gene expression during addition of two stress chemicals (methyl viologen and hydrogen peroxide) which changed the fermentation redox potential.

Results

The addition of methyl viologen to *C. thermocellum* DSM 1313 chemostat cultures caused an increase in ethanol and lactate yields. A lower fermenter redox potential was observed in response to methyl viologen exposure, which correlated with a decrease in cell yield and significant differential expression of 123 genes ($\log_2 > 1.5$ or $\log_2 < -1.5$, with a 5% false discovery rate). Expression levels decreased in four main redox active systems during methyl viologen exposure; the [NiFe] hydrogenase, sulfate transport and metabolism, ammonia assimilation (GS-GOGAT), and porphyrin/siroheme biosynthesis. Genes encoding sulfate transport and reduction and porphyrin/siroheme biosynthesis are co-located immediately downstream of a putative *iscR* regulatory gene, which may be a cis-regulatory element controlling expression of these genes. Other genes showing differential expression during methyl viologen exposure included transporters and transposases.

Conclusions

The differential expression results from this study support a role for *C. thermocellum* genes for sulfate transport/reduction, glutamate synthase-glutamine synthetase (the GS-GOGAT system), and porphyrin biosynthesis being involved in redox metabolism and homeostasis. This global profiling study provides gene targets for future studies to elucidate the relative contributions of prospective pathways for co-factor pool re-oxidation and *C. thermocellum* redox homeostasis.

Keywords

Clostridium thermocellum DSM 1313, microarray, transcriptomics, methyl viologen, chemostat, redox, sulfate, GS-GOGAT, hydrogenase.

Background

Clostridium thermocellum natively expresses enzymes to both deconstruct lignocellulosic biomass and ferment cellulose into ethanol, making it a candidate biocatalyst for consolidated bioprocessing (CBP). *C. thermocellum* hydrolyzes lignocellulosic biomass rapidly and efficiently using an elaborate enzyme system in the form of free and cell bound multi-enzyme cellulolytic complexes called cellulosomes [57, 58]. Though its hydrolysis machinery is among the fastest and most effective known [59], its fermentative metabolism results in yields and productivities too low for cost-effective industrial lignocellulosic ethanol production [58, 60].

In addition to ethanol, *C. thermocellum* natively produces acetate, lactate, formate and hydrogen. Efforts to eliminate these undesirable products [61-63], in conjunction with metabolic engineering of lignocellulosic substrates, have been met with higher ethanol yields [64]. Another limitation to ethanol yield and productivity is a presumed 'overflow' metabolism by which *C. thermocellum* makes a number of other products, a phenomenon that seems to be exacerbated when fermentative metabolic pathways that reoxidize redox cofactors are eliminated or when substrate loadings are relatively high [65-67]. *C. thermocellum* metabolism is affected by the addition of exogenous fermentation products [68] and

inhibitor chemicals [69] as well as other environmental perturbations. These responses were seen through altered end product distributions, O/R balances, and inhibited substrate uptake. Further, metabolic changes that originate at the level of sensing and transcription have been observed in response to different physical and chemical perturbations [70-73]. The mechanisms by which *C. thermocellum* senses, regulates and balances redox status remain poorly understood and a deeper understanding may inform future metabolic engineering efforts. The potential inability of various engineered and wild-type strains to sufficiently reoxidize redox cofactors and possible cellular redox imbalance from overly reduced cofactor pools is an area of biotechnological interest.

Redox metabolism has been studied in many organisms capable of carrying out a variety of redox reactions. One method to examine redox related metabolism is to observe gene expression responses to an altered redox environment [74, 75]. Other studies have employed comparative genomics approaches [76], *rex* regulatory gene deletion studies [77, 78] or high-throughput genetic approaches such as rapid transposon liquid enrichment sequencing (TnLE-seq) under stress conditions [79] to investigate physiological and regulatory responses. Such studies give insights into not only transcriptional responses, but also regulons and signaling responses to such environments. After methyl viologen exposure *Clostridium acetobutylicum* showed decreased expression of solvent producing genes, as well genes involved in sulfate and iron transport, while butanol synthesizing genes showed increased transcription concomitant with a much higher butanol/acetone ratio [75]. Fermentation and metabolite analysis of *Clostridium cellulolyticum* implicated high NADH/NAD^+ ratio, and low pyruvate:ferredoxin oxidoreductase activity as causing limited fermentative metabolism and production of overflow products [80, 81]. Hence, studies into redox metabolism have the potential to not only provide fundamental insights but also have potential to advance applied goals.

In this study, we established *C. thermocellum* steady state chemostat cultures and investigated redox processes after perturbing environmental conditions through separate additions of two redox-active

chemicals; methyl viologen and hydrogen peroxide. Fermentative changes were observed and transcriptional responses of *C. thermocellum* DSM 1313 were studied using DNA microarray analyses.

Results and Discussion

Preliminary Batch Experiments with Methyl Viologen

Preliminary batch growth indicated methyl viologen reduction *in vivo* required viable metabolically active cellular biomass. A change to blue coloration was used as an indication of methyl viologen reduction. To assess methyl viologen reduction, abiotic medium, autoclave killed cells (in medium or water), live growing cells and cell free spent medium, each containing 150 mg/L methyl viologen, were incubated in *C. thermocellum* growth conditions overnight. Blue coloration was only observed after incubation of cells from log phase or stationary phase cultures medium containing cellobiose and methyl viologen. Blue coloration was not observed after 5 days of incubation in fresh medium, spent medium, autoclaved cells (in medium or resuspended in water), cells (log phase or stationary phase) resuspended in water, after 5 days of incubation at 55°C. Preliminary batch fermentations were conducted to estimate the appropriate methyl viologen concentration to introduce into the carbon-limited chemostats (Figure A1). A decrease in growth rate of batch cultures was seen in cultures containing initial loadings of methyl viologen of 150 ug/mL and higher. Consistent with an earlier study [82], an increase in end-point ethanol productivity was observed in cultures containing methyl viologen (Figure A2) and at a final concentration of 150 mg/L was used in chemostat studies. Methyl viologen has been shown to inhibit *Clostridium butyricum* hydrogen production during glycerol fermentation [83]; however, in preliminary *C. thermocellum* cellobiose batch studies specific end-point H_2 productivity increased with methyl viologen.

Chemostat Response to Methyl Viologen Addition

Carbon limited (1.1 g/L cellobiose) chemostat grown cultures (0.1 hr^{-1}) were supplied with MTC medium containing methyl viologen (Figure 1). Time “0” for methyl viologen addition was the point when the

feed medium was switched from MTC medium to MTC medium containing methyl viologen (~88h after the experiment began). Addition of methyl viologen to the reactor lowered cell density (as measured by fermenter OD_{600}) (Figure 1). An increase in specific ethanol production was observed concomitant with methyl viologen exposure, lower cell densities and an approximate 50 mV decrease in redox potential (Figure 1 and 2). Throughout methyl viologen exposure, a total of 123 individual genes were significantly differentially expressed within at least one timepoint (defined as a $\log_2$ expression change of <-1.5 or >1.5 at a false discovery rate <0.05) (Table 1). After methyl viologen was flushed from the system, ethanol specific productivity returned to pre-exposure levels (Figure 2) and redox potential measurements returned to pre-exposure redox potential levels (Figure 1A). Production of acetate remained unchanged before and during methyl viologen addition. Acetate concentration and yield increased after methyl viologen exposure was completed and as methyl viologen was being diluted out of the chemostats. CO_2 and hydrogen production was not measured in the open chemostat system. Lactate began to be synthesized after 50 hours of methyl viologen exposure, increasing a specific productivity from 1.85 mM/OD_{600} to 3.55 mM/OD_{600}. Before and throughout methyl viologen addition, all cellobiose substrate was utilized as no residual cellobiose or glucose was detected by HPLC analysis of samples from the fermenters. Specific fermentation products (converted to reflect the amount of carbon in the products; mol C-equivalents/L/OD_{600}) increased throughout methyl viologen exposure. Ethanol and estimated CO_2 (mol C equivalents/L/OD_{600}) increased 17.6% and 15.8%, respectively, during 60 hours of increasing methyl viologen exposure (Figure 2). Increased alcohol concentrations have been observed after adding methyl viologen to cultures of *C. acetobutylicum,* which was attributed to decreased hydrogen production [84, 85]. This phenomenon was observed upon methyl viologen addition to *C. acetobutylicum* cultures and was attributed, in part, to increased transcription of butanol synthesis pathway genes [75]. In this study, transcription of ethanol synthesis genes was not increased significantly. Global transcriptional analysis suggests found two broad temporal expression patterns to

Table 1. Number of genes showing differential expression after beginning methyl viologen addition. Significant differential expression was determined to be genes showing $\log_2$ fold change relative to untreated controls of >1.5 or <-1.5 at a 5% false discovery rate.

	3 min.	**15 min.**	**35 min.**	**7 hr.**	**14 hr.**	**50 hr.**	**60 hr.**
Number of Genes Up-Regulated	0	14	1	3	1	21	36
Number of Genes Down-Regulated	20	37	3	8	1	40	47

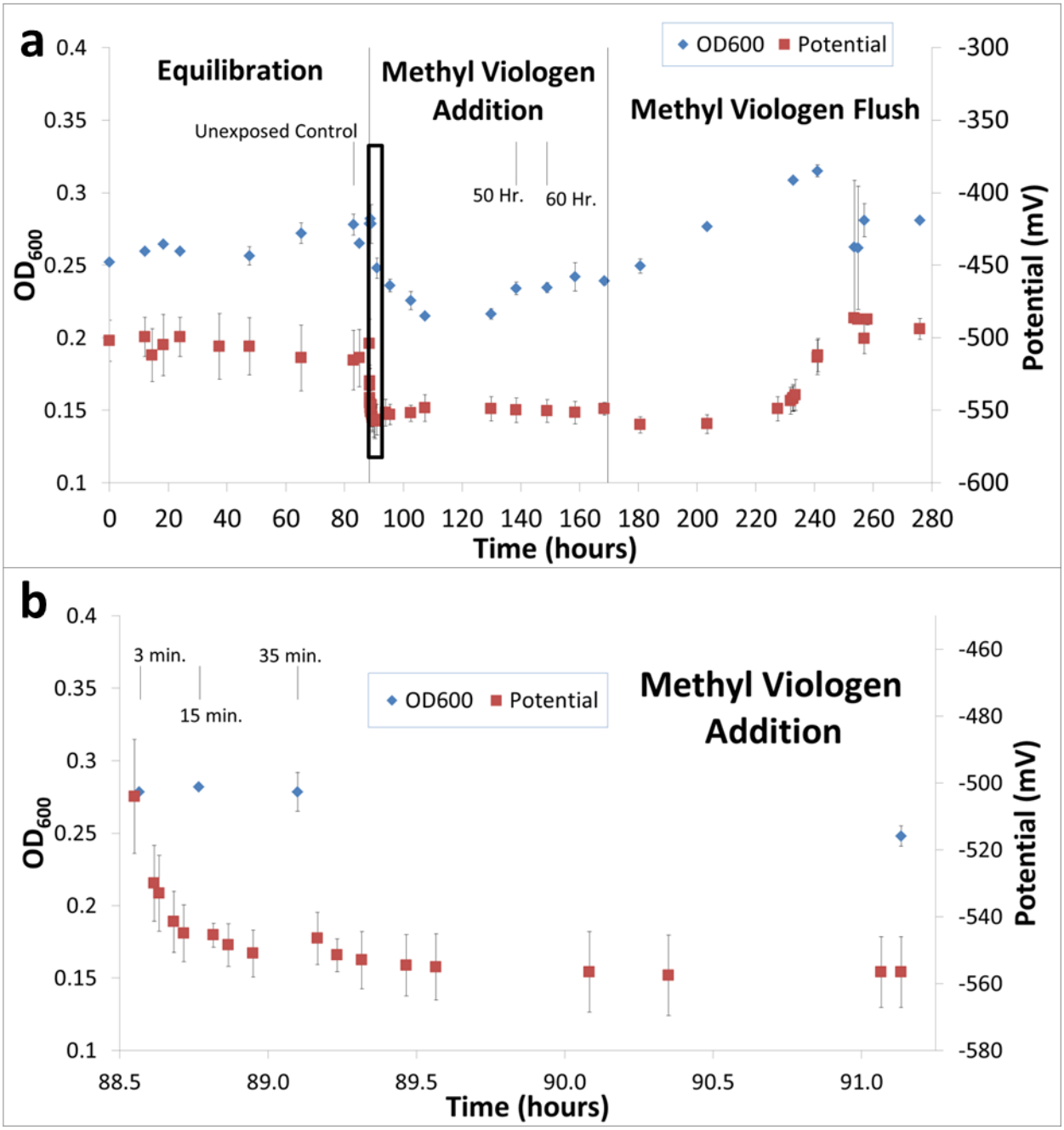


Figure 1. (A) OD_{600} and measured redox potential before, during and after methyl viologen addition. (B) Detailed view of boxed region indicated in Figure 1(A).

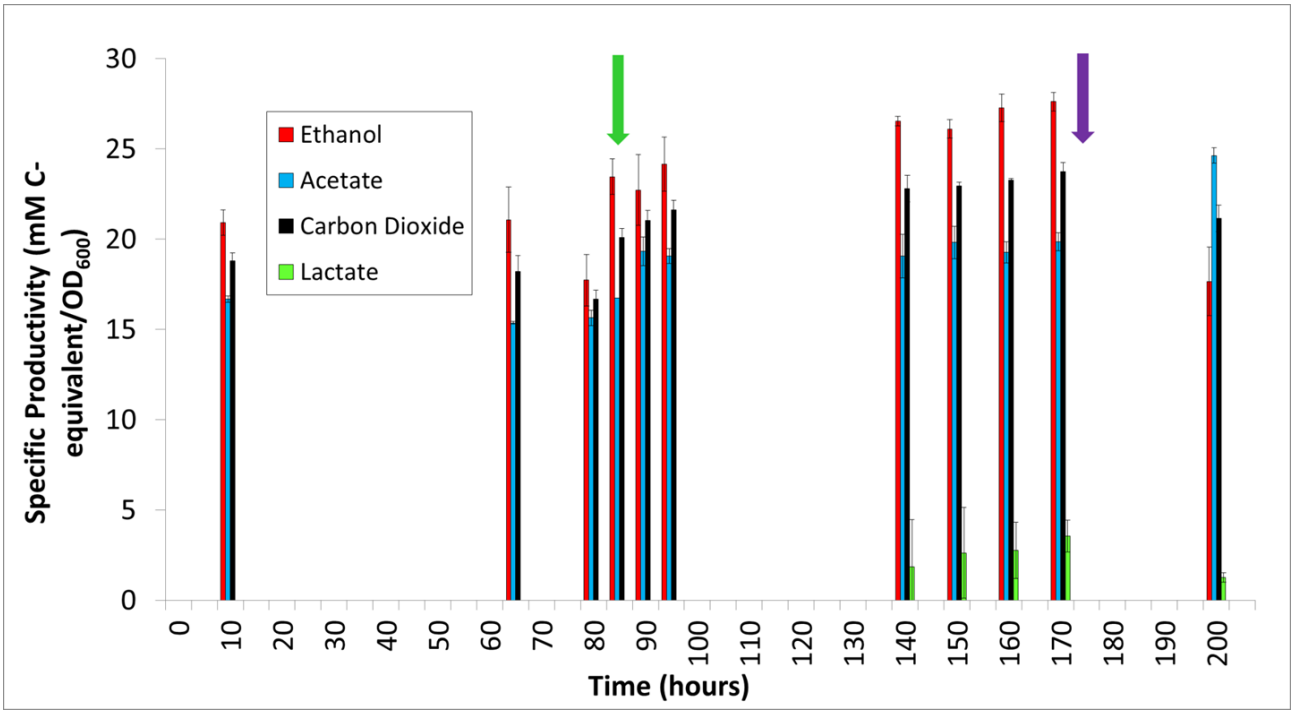


Figure 2. Fermentation productivity during addition of methyl viologen to chemostat culture. Specific fermentation products (converted to reflect the amount of carbon in the products (mol C-equivalents/L/OD_{600}) are reported in equivalent carbon mole basis (e.g., 1 mol ethanol = 2 mol C-equivalent). Productivity in carbon moles is normalized to OD600 to account for changing cell yields observed across the time culture was exposed to methyl viologen. CO2 productivity is estimated by assuming one mole of CO2 is produced for each mole of ethanol and each mole of acetate produced. Green arrow indicates when methyl viologen exposure began and purple arrow indicates time when methyl viologen flushing from the reactor began.

be present across genes which showed differential expression in at least one time-point taken during methyl viologen exposure. Indeed, differential expression at 15 minutes was very similar to that sampled at 50 and 60 hours. The first pattern observed was decreased expression at 15 minutes, followed by a return to pre-exposure expression, and finally decreased expression at 50 and 60 hours. We interpret this result to suggest an expression response occurring immediately after initial exposure to methyl viologen (while concentrations of methyl viologen in the fermenters are low) followed by a decay of this response to pre-exposure conditions. Beginning at 50 hours, expression levels return to levels similar to those seen at 15 minutes, but do so in a sustained fashion as those levels of expression are largely maintained for another 10 hours. It should be mentioned that, by 50 hours, concentrations of methyl viologen in the fermenters had essentially asymptoted to that of concentrations introduced in the feed carboy; 150 mg/L.

While many genes showed significant and appreciable differential expression, much of the expression differences appeared disparate and singular, or to occur in genes which had poor or no annotation. In an effort to piece together a system-wide view of expression differences with the objective of visualizing whole metabolic systems, normalized, log2 differential expression data were overlayed onto a metabolic map to further interrogate metabolic systems showing differential expression. From this, metabolic systems which showed, in an apparently coordinated fashion, significantly decreased transcription in many related loci were identified; those of glutamate/ammonia metabolism as well as those of sulfate transport and reduction

Decreased Transcription of Redox Active Pathways

Methyl viologen can occur in three reduction states, MV^0, MV^+, and MV^{2+}, with the redox potential of the most reduced form being similar to that of electron transfer proteins, flavodoxins and ferredoxins [86, 87]. Methyl viologen can be used in place of natural electron-acceptors and oxidoreductase proteins in nitrogenase [88], nitrate reductase [89] and other enzyme assays . In addition to these known

enzymatic interactions, methyl viologen has also been shown to interact directly with iron-sulfur clusters and, in some cases, degrade them [90]. Fe-S cluster degradation by methyl viologen and other redox-active compounds has been shown to initiate wide-ranging transcriptional changes through induction of the SoxRS transcription factor system [91]. Using mechanisms similar to these, it is conceivable that methyl viologen can interact with redox metabolism indirectly in *C. thermocellum* and, as such, our experimental system may be capturing transcriptional outcomes of these mechanisms in addition to direct interactions facilitated by enzymes and redox chemistry with redox-active intracellular metabolites.

Four redox active pathways showed decreased transcription at various time points after methyl viologen addition; sulfur transport and assimilation, ATP dependent GS-GOGAT ammonia assimilation, porphyrin and siroheme biosynthesis and the [NiFe] Fd dependent hydrogenase (Figures 1-3). All four systems show decreased transcription at 50 h (3,000 min)- 60h (3,600 min) after methyl viologen addition when estimated methyl viologen concentration in the reactor is greatest and the fermenter redox potential was the lowest. GS-GOGAT and some genes of the [NiFe] system and gene cluster show decreased transcription at three and 15 minutes after beginning methyl viologen addition.

Ammonia Assimilation

Previous studies have shown the malic enzyme is allosterically activated by ammonia for the conversion of malate to pyruvate and it is likely the primary carbon flux channel under certain growth conditions in *C. thermocellum* [92]. The putative DSM 1313 malic enzyme gene (Clo1313_1878) showed slightly greater transcript levels after methyl viologen treatment, but not at levels considered significant. Intracellular metabolites, such as ammonia, and enzymatic activities were not measured as part of present study. We observed genes encoding the GS-GOGAT system showed decreased transcription (Table 2). Both glutamine synthetase (GS) and glutamate synthase (GOGAT) showed decreased expression (log_2 differential expression of -1.3 to -2.4 relative to untreated controls) at 3 min and 15

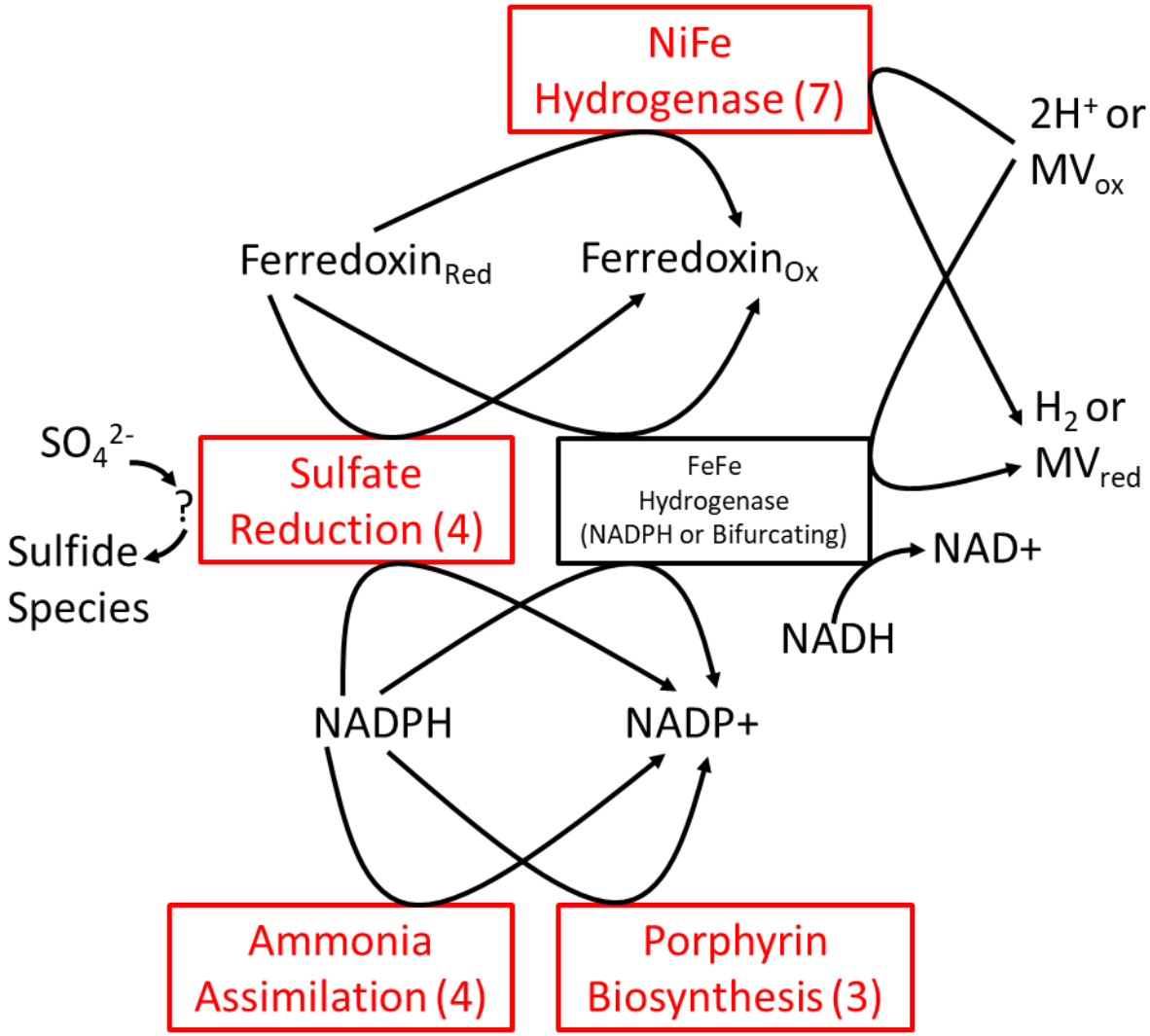


Figure 3. Global view of main of *C. thermocellum* DSM 1313 transcriptional responses to methyl viologen. Red indicates decreased transcription of genes in the indicated systems. Numbers in parenthesis are the number of genes in the indicated pathways that show decreased expression during methyl viologen addition.

minutes and also at 3000 and 3600 minutes (log_2 differential expression of -0.6 to -1.5 relative to untreated controls (Table 2)). Other genes annotated as glutamine synthetase (Clo1313_2038, Clo1313_2031, and Clo1313_1357) in KEGG and predicted to catalyze the same enzymatic reaction (E.C. 6.3.1.2) showed no differential transcription in this study. One other gene annotated as a glutamate synthase (Clo1313_1849) also did not show any differential transcription in this study. Recent reports suggest Clo1313_1849 actually encodes the NfnA subunit of NfnAB [69].

Clo1313_2036 and Clo1313_2035, the genes putatively encoding two subunits of glutamate synthase, are part of a gene cluster (Clo1313_2030 – Clo1313_2036) that showed similar expression behavior (Table 2). Clo1313_2034 is annotated as a 4Fe-4S ferredoxin iron-sulfur binding domain-containing protein. Top global protein-BLAST similarity scores are to iron-sulfur cluster containing ferredoxin in other strains of *C. thermocellum* and other *Clostridia*. Both the GS-GOGAT system and glutamate dehydrogenase are annotated as being NADPH dependent and, to our knowledge, ferredoxin dependent GS-GOGAT activity has not been assayed for in *C. thermocellum* and NADH dependent GOGAT activity was not found in crude cell lysates of *C. thermocellum* DSM 1273 [93]. Glutamate synthesis from glutamine using glutamate synthase may compete for reductant with other redox processes, including the reduction of methyl viologen. Glutamine biosynthesis also requires glutamate, ATP and ammonium. Reduced expression of the GS-GOGAT system may be the result of added demand for reductant and/or altered cofactor pool states introduced by the presence of oxidized methyl viologen, may reflect lower intracellular ATP levels and/or reflect less cellular demand for glutamine. Such altered states may trigger a transcriptional response toward preserving reductant and/or ATP at the expense of glutamine and glutamate production using this GS-GOGAT system.

Transcriptional regulation of ammonia transport, the GS-GOGAT system and glutamate dehydrogenase appears to be complex, multi-layered, and varies greatly between organisms [94, 95]. A transcriptional regulator of ammonia assimilation characterized in *C. acetobutylicum*, relies on expression of a *nitR*

Table 2. Differential expression for GS-GOGAT ammonia and [NiFe] hydrogenase genes. Values in bold indicate statistical significance using a 5% false discovery rate.

Locus Tag	Gene Product	Metabolic Function	$log2\left(\frac{Expression\ (Time\ Exposed)}{Expression\ (Unexposed)}\right)$						
			3 min.	15 min.	35 min.	7 hr.	14 hr.	50 hr.	60 hr.
Clo1313_2303	glutamine synthetase catalytic region	GS-GOGAT Ammonia Fixation	**-2.2**	**-2.4**	**-0.9**	**1.0**	**1.2**	**-1.5**	**-1.3**
Clo1313_2036	glutamine amidotransferase class-II		**-1.3**	**-2.1**	-0.4	**1.6**	**1.5**	-0.6	-0.7
Clo1313_2035	ferredoxin-dependent glutamate synthase		**-2.0**	**-2.2**	**-0.6**	**1.2**	**1.2**	**-1.1**	**-0.9**
Clo1313_2034	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein		**-2.3**	**-2.4**	**-0.9**	**0.7**	**0.7**	**-1.5**	**-1.3**
Clo1313_0564	hydrogenase expression/formation protein HypE	NiFe Hydrogenase	**-0.4**	**-1.4**	**-0.5**	**-0.9**	**-0.7**	**-1.3**	**-1.7**
Clo1313_0565	hydrogenase expression/formation protein HypD		**-0.4**	**-0.6**	-0.1	**-0.3**	**-0.3**	**-0.7**	**-0.9**
Clo1313_0566	hydrogenase assembly chaperone hypC/hupF		**-0.5**	**-0.8**	-0.2	**-0.4**	**-0.3**	**-0.9**	**-1.2**
Clo1313_0567	(NiFe) hydrogenase maturation protein HypF		**-0.7**	**-0.9**	0.0	**-0.5**	-0.3	**-1.0**	**-1.0**
Clo1313_0568	hydrogenase accessory protein HypB		-0.1	**-0.5**	**0.5**	0.0	0.0	**-0.4**	**-0.9**
Clo1313_0569	hydrogenase expression/synthesis HypA		**-0.3**	**-0.9**	0.1	**-0.3**	-0.2	**-0.7**	**-1.1**
Clo1313_0570	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein		**-0.4**	**-1.2**	0.2	-0.2	-0.2	**-0.8**	**-1.3**

antiterminator protein and an antisense RNA [96]. Clo1313_2030 is annotated as a response regulator receiver and ANTAR domain protein. The Clo1313_2030 gene showed decreased expression mirroring that of the gene cluster containing the genes Clo1313_2030 through Clo1313_2036 and its product may provide an antiterminator role sensitive to ammonia concentrations similar to the one described in *C. acetobutylicum*. The AmtB and PII proteins contribute to ammonia transport and regulation in different systems [94]. Clo1313_2260 contains putative AmtB and PII domains is 51% and 81% similar to *C. acetobutylicum* ATCC 824 CA_C0682 and CA_C0681 at the protein level, respectively. In this study, Clo1313_2260 was not significantly differentially expressed.

Glutamate dehydrogenase (Clo1313_1847) catalyzes an alternative mechanism for synthesizing glutamate and assimilating ammonia and glutamate dehydrogenase (Clo1313_1847) was not differentially expressed in this study. Glutamate dehydrogenase activity was shown to be much higher than GOGAT in *C. thermocellum* DSM 1273 extracts, suggesting it is the predominant method for assimilating ammonia and generating glutamate [93]. However, these experiments were conducted at a maximum nitrogen concentration of 18 mM, supplied as ammonia and 0.2% yeast extract. MTC medium used in this experiment contained more inorganic nitrogen in the form of both urea (33.3 mM) and ammonia (as ammonium chloride at 28 mM). The medium used in this experiment and routinely for *C. thermocellum* growth contains a large excess of nitrogen source and it is possible to culture *C. thermocellum* in medium containing less nitrogen [97].

Glutamate is a precursor to many biosynthetic pathways, including the biosynthesis of porphyrin rings. Porphyrin ring biosynthesis also showed decreased transcription upon exposure to methyl viologen. Decreased demand for glutamate or increased availability may have led to decreased transcription of GS-GOGAT. Other putative KEGG pathways using glutamate as a precursor did not show differential expression. By contrast, increased transcription and translation of both glutamine synthetases (Cthe_0196, Cthe_1539) were observed after *C. themocellum* ethanol stress, which may have been

related to metabolite shuttling into carbamoyl-P, a precursor to pyrimidine, arginine and proline biosynthesis [72], or possibly glutamine synthetase and the ammonia assimilation pathways indirectly assist in reoxidizing redox cofactors.

Sulfate Transport and Metabolism

C. thermocellum strain ATCC 27405 has genes for and can assimilate sulfate [97]. As in strain ATCC 27405, sulfate transport genes (Clo1313_0114 – Clo1313_0117) and putative assimilatory sulfate reduction genes (Clo1313_0119, Clo1313_0120, and Clo1313_0124) are co-located on the DSM 1313 chromosome. In this study, these genes show similar expression patterns and lower expression levels under methyl viologen stress with $\log_2$ differential expression of -0.9 to -1.6 relative to untreated controls at 3000 and 3600 minutes (Table 3), which is consistent with methyl viologen-exposed *C. acetobutylicum* cells [75] and indicates potential similarity in the physiological and regulatory responses between these organisms. Adjacent to sulfur related genes, the *C. thermocellum* Clo1313_0107 gene encodes a putative transcriptional regulator IscR [70], and its expression increased following methyl viologen exposure (Table 3). Differential expression profiles for Clo1313_0107 (*iscR*) and porphyrin biosynthesis genes are similar when exposed to furfural or heat [70]. Further studies to generate and characterize a Clo1313_0107 deletion strain are required to elucidate its possible roles in sulfate uptake and metabolism, stress responses and gene regulatory networks. Because methyl viologen is a potential alternative electron sink to reoxidize reduced intracellular cofactors, *C. thermocellum* DSM 1313 could potentially repress genes involved in sulfate transport and reduction when exposed to methyl viologen in response to a decreased need for electron acceptors. Sulfate reduction is also ATP dependent and it may benefit *C. thermocellum* DSM 1313 energetically to decrease expression and/or activity of this pathway.

Porphyrin Biosynthesis

Many redox active enzymes are iron-sulfur containing proteins and siroheme is often a necessary redox active cofactor. The genes Clo1313_0108 – Clo1313_0113 putatively encode for proteins involved in porphyrin/siroheme biosynthesis and they are adjacent to genes involved in sulfate transport/reduction as well as genes potentially involved in regulation. The Clo1313_0109, Clo1313_0112 and Clo1313_0113 genes show differential expression at 3000 minutes after methyl viologen addition (log_2 differential expression of -0.7, -1.2 and -1.2 respectively, relative to untreated controls), and show the largest differential expression after 3600 minutes (log_2 differential expression of -1.2, -1.5 and -1.6 respectively, relative to untreated controls), when estimated methyl viologen concentration is the highest (Table 3). Three other putative porphyrin biosynthesis genes in this cluster do not meet criteria for significant differential expression. Clo1313_0124 is annotated as nitrite and sulfite reductase, predicted to contain an iron-sulfur/siroheme binding site, and also shows decreased transcription under methyl viologen exposure. Clo1313_0124 is annotated to be ferredoxin dependent. Expression of Clo1313_0109 was shown to increase upon exposure to *Populus* hydrolysate [73] and decreased expression 10, 30, and 60 minutes following exposure to increased ethanol concentration [72], which suggests, along with results from this study, that transcription of genes responsible for porphyrin biosynthesis is sensitive to cellular redox potential and/or the redox potential of the fermentation environment. With methyl viologen potentially providing an extra sink for cellular reductant through a process mediated by hydrogenases or occurring directly, there may be less overall demand for these proteins mediating redox reactions. Unbound heme and some of the metabolic intermediates of the tetrapyrrole biosynthesis pathway are known to be cytotoxic and can be recycled [98]. *Pseudomonas aeruginosa* appears to regulate early stages of heme biosynthesis in accordance with growth state and overall need for heme enzyme cofactors [99]. Many microorganisms have been shown to limit the amount of accumulated 5-aminolenulevulenate [98], suggesting it may be particularly cytotoxic and/or the precursor to a rate-

limiting step in tetrapyrrole biosynthesis. Thus, it may be advantageous for *C. thermocellum* DSM 1313 to regulate this pathway and Clo1313_0107 may play a role.

NiFe Hydrogenases

Methyl viologen has been shown to mediate reoxidation of reduced cellular species through interaction with hydrogenase proteins [100, 101]. *C. thermocellum* DSM 1313 encodes four hydrogenases; three [FeFe] hydrogenases and one [NiFe] hydrogenase. No differential expression was observed in genes encoding any of the three [FeFe] hydrogenase systems in response to methyl viologen exposure. It is thought that two of the three [FeFe] hydrogenases are bifurcating hydrogenases [1], necessarily requiring one mole of NADH and one mole of reduced ferredoxin to produce one mole of hydrogen. The third [FeFe] hydrogenase is thought to be NADPH dependent. It has been shown that the NADPH dependent hydrogenase activity was greater than either reduced ferredoxin or NADH dependent hydrogenase activity [102, 103]. Further, low amounts of Ech hydrogenase protein was quantified [1] and poor transcription of *ech* genes [104] was found using reverse transcriptase PCR from *C. thermocellum*. In contrast, we observe all genes in the *ech* hudrogenase gene cluster to be highly transcribed in the conditions used in this study, based on LOWESS normalized hybridization intensity. All but one of the genes encoding *ech,* an [NiFe]-containing hydrogenase (Clo1313_0564 – Clo1313_0570) show transient initial decreased transcription at 3 and 15 minutes (log_2 differential expression -0.3 to -1.2, relative to untreated controls) after beginning methyl viologen exposure, followed by a return to unchanged levels of transcription (significant log_2 differential expression not less than -0.4 or greater than 0.2, relative to untreated controls). Expression of this gene cluster again decreased at 3000 and 3600 minutes (log_2 differential expression -0.7 to -1.3, relative to untreated controls) after beginning methyl viologen exposure when the concentration of methyl viologen in the reactor is highest. All genes in this cluster appear to observe this temporal expression behavior (Table 3). It would be interesting to determine whether or not the [NiFe] hydrogenase is ferredoxin dependent

and biochemically characterize its relative contribution to hydrogen metabolism to better understand co-factor reoxidation capacity and dynamics in *C. thermocellum*.

C. thermocellum DSM 1313 has been reported to have overflow metabolic pathways, potentially related to overly reduced intracellular conditions and a need to re-oxidize redox cofactors [65, 67]. [NiFe] *ech* hydrogenase transcription does not appear to be transcriptionally linked to ATP generation or establishing/modulating the proton motive force in *C. thermocellum*, as there was no significant differential expression of the ATP synthase genes (Clo1313_2935 – Clo1313_2942) observed during addition of methyl viologen to the culture. In different systems [NiFe] hydrogenase is known to be controlled by a number of different mechanisms and responsive to many different environmental and physiological cues, such as H_2, O_2 and CO as well as the FnrT transcription factor [105, 106]. [NiFe] hydrogenases have also been linked to nitrogenase activity through electron transfer to membrane-bound Rnf [105]. *C. thermocellum* encodes genes with putative functions for N_2 fixation (Clo1313_2331, Clo1313_2332, and Clo1313_2339). Clo1313_2331 and Clo1313_2339 show significant decreased transcription at 3 minutes after methyl viologen addition and differential expression in an equivalent *C. thermocellum* ATCC 27405 gene, Cthe_1573 *(nifH)* showed increased expression 240 minutes after exogenous ethanol addition to batch culture [72]. *C. thermocellum* DSM 1237 was reported to show nitrogen fixing activity in-vitro, via reduction of acetylene, though doubt was cast as the activity was not seen to be repressed by added ammonia [93]. *C. thermocellum* diazotrophic growth was tested and under the conditions assayed nitrogen fixation was not detected [97]. *C. thermocellum* electron transfer requires greater study.

Transcription Differences in Other Systems

Transporters

Eight annotated transporters found to be significantly differentially expressed during at least one time point under methyl viologen treatment. A CorA family transporter annotated as a magnesium

Table 3. Differential expression information for genes Clo1313_0107 through Clo1313_0124. Values in bold indicate statistical significance using a 5% false discovery rate.

Locus Tag	Gene Product	Metabolic Function	$log2\left(\frac{Expression\ (Time\ Exposed)}{Expression\ (Unexposed)}\right)$						
			3 min.	**15 min.**	**35 min.**	**7 hr.**	**14 hr.**	**50 hr.**	**60 hr.**
Clo1313_0107	Transcriptional regulator, Rrf2 family	Putative IscR Transcription Factor	**1.1**	0.2	**2.4**	**2.1**	**1.9**	**1.3**	**1.7**
Clo1313_0109	precorrin-6X reductase	Porphyrin Biosynthesis	**-0.5**	**-0.4**	**0.2**	-0.2	-0.1	**-0.7**	**-1.2**
Clo1313_0112	delta-aminolevulinic acid dehydratase		**-0.8**	-0.5	-0.1	**-0.5**	-0.3	**-1.2**	**-1.5**
Clo1313_0113	glutamate-1-semialdehyde-2,1- aminomutase		**-0.9**	-0.6	0.1	-0.3	-0.3	**-1.2**	**-1.6**
Clo1313_0115	sulfate ABC transporter, inner membrane subunit CysT	Sulfate ABC Transporter	**-0.6**	-0.1	**0.5**	**0.4**	**0.4**	**-1.2**	**-1.3**
Clo1313_0116	sulfate ABC transporter, inner membrane subunit CysW		**-0.6**	0.1	**0.5**	0.4	0.4	**-1.0**	**-1.4**
Clo1313_0118	adenylylsulfate reductase, thioredoxin dependent	Sulfate Reduction	**-0.9**	-0.4	**0.5**	0.1	0.1	**-1.4**	**-1.6**
Clo1313_0124	nitrite and sulphite reductase 4Fe-4S region		**-1.4**	-0.2	0.2	0	-0.1	**-0.9**	**-1.5**

transporter showed relatively strong increased transcription at 3000 and 3600 minutes after methyl viologen addition. Five transporters annotated as being ATP binding showed decreased transcription at least one timepoint during methyl viologen exposure.

Sporulation

C. thermocellum strain ATCC 27405 sporulates at low frequencies (maximum ~7%) and not under certain conditions such as in response to low carbon or nitrogen environments [107], which confounds global population based analyses. In this study, three sporulation genes showed significant decreased transcription (false discovery rate < 0.05 and $\log_2$ differential expression less than -1.5 relative to untreated controls); Cthe_3070, Cthe_0044 and Cthe_2948. The initial sporulation developmental stages transcription factor, *spoOA* (Clo1313_1409), showed significant changes in expression at three and 15 minutes after beginning methyl viologen addition ($\log_2$ differential expression of 0.4 and -0.8, respectively). Two of the four histidine kinases (Clo1313_0286 and Clo1313_1973) previously identified as agents of sporulation control [108] showed increased differential transcription at 3000 ($\log_2$ 1.1 and 1.1, respectively) and 3600 minutes ($\log_2$ 1.0 and 1.5, respectively) after beginning methyl viologen exposure. Clo1313_1973 also showed relatively strong decreased transcription at 3 minutes (significant $\log_2$ differential expression of -1.5). In addition to spore formation, *C. thermocellum* (strain ATCC 27405) can develop an L-form morphology which is another non-growth cell state characterized by low metabolic activity with a spherical or pleomorphic morphology [107]. Spores and L-form cells were not observed together in a previous study [107]. Much remains to be elucidated regarding signal-transduction and regulatory cascades controlling these processes.

Transposases and Phage Associated Gene Expression

Two genes annotated as transposases (Clo1313_0662 and Clo1313_2686) showed increased transcription after 3600 minutes (significant $\log_2$ differential expression of 1.8 and 1.7, respectively). These two genes appear to be highly transcribed before and during exposure to methyl viologen

(LOWESS normalized expression intensity ranged from $\log_2$ 11.4 to 13.4). Transposons have been shown to interrupt and inactivate the *C. thermocellum cipA* genes [109]. *C. thermocellum* ATCC 27405 transcriptomic studies have shown differential expression for transposases under different growth and stress conditions [70, 72]. *C. thermocellum* strains differ in their phage and CRISPR content [110]. Increased expression of genes associated with a putative phage island (Clo1313_2379 – Clo1313_2409) was observed in one of two replicate samples taken after exposure to methyl viologen and immediately prior to beginning exposure to H_20_2, after 3600 minutes (60 hours) of re-equilibration growth on MTC not amended with either stressor chemical. The majority of genes from this phage island were found to show significant increased transcription in *C. thermocellum* ATCC 27405 after 30, 60 and 120 minutes of exposure to furfural and ethanol, with the largest transcriptional increases coming 30 minutes after exposure and dropping gradually at 60 and 120 minutes for both conditions [70, 72]. This gene region *C. thermocellum* DSM 1313 does not contain 'att sites' characteristic of functional lysogenic phage, though the closely related *C. thermocellum* ATCC 27405 does [110]. Genes for phosphate transport and regulation are adjacent to recombinase and phage genes in *C. thermocellum* ATCC 27405, which indicates possible mechanisms for horizontal gene transfer [70]. Observed strain genome differences and differential expression from this and prior studies suggest transposons, bacteriophage and CRISPR systems may have and may continue to play important roles in *C. thermocellum* evolution, although the functionality and implications for these systems have only begun to be investigated.

Hydrogen Peroxide Addition to Chemostat Culture

Compared to untreated fermentation redox potential (~-500 mV), fermenter redox sharply and briefly increased by approximately 100 mV initially and then remained approximately 25 mV higher than the control during prolonged H_2O_2 exposure. However, no genes showed significant differential expression ($\log_2$ fold change >1.5 or <-1.5 and a 5% false discovery rate) after chemostat hydrogen peroxide exposure and only a few gene comparisons showed significant differences (below fold threshold) after 3

minutes of H_2O_2 exposure. There was no apparent change in acetate, ethanol, lactate, or estimated CO_2 cell yields (moles C equivalent/OD_{600}). Furthermore, addition of H_2O_2 under the conditions mentioned previously had no prominent effect on fermenter OD_{600} (Figure A4).

It is likely the concentration of hydrogen peroxide introduced into the fermenter, though enough to alter the fermentation redox potential, was not enough to bring about detectable transcriptional-level changes. Redox couples within the cell, such as the GSH/GSSG couple [111], are able to act as redox buffers and provide cells protection against unfavorable environmental redox conditions while not necessarily requiring changes in transcription. It is conceivable that *C. thermocellum* could have used similar systems to modulate the hydrogen peroxide induced redox perturbation it was exposed to in this experiment, instead of systems requiring changes to its transcription profile. Additionally, hydrogen peroxide is a meta-stable chemical susceptible to degradation and chemical change under the conditions it was used in this experiment. A portion of the hydrogen peroxide may have degraded in the process steps leading up to its introduction into the fermenters. Though we recognize this possibility, we infer a change in redox potential after the addition of of hydrogen peroxide to the fermenters as evidence that the hydrogen peroxide treatment did have an oxidizing effect on the culture, though possibly not as large as anticipated.

Conclusions

We examined *C. thermocellum* DSM 1313 redox metabolism by analyzing the transcriptional response to gradual addition of methyl viologen to steady state cultures. Specific ethanol productivity increased steadily during methyl viologen addition, likely due to altered redox state, while OD_{600} dropped initially after methyl viologen addition and then recovered slightly and stabilized after 3000 minutes exposure. The redox potential of the fermentation was stable at -500 mV before methyl viologen addition and began dropping immediately after methyl viologen addition began and stabilized at -550 mV after approximately 100 minutes. We observed a number of redox active and ATP requiring systems showing

decreases in transcription as methyl viologen was added to chemostat cultures. Genes encoding sulfate transport and reduction, glutamate synthase-glutamine synthetases (the GS-GOGAT system), and portions of porphyrin biosynthesis showed differential expression in response to added methyl viologen, suggesting their involvement in mediating *C. thermocellum* redox homeostasis and energy metabolism. Genes encoding subunits and accessory proteins of the sole [NiFe] hydrogenase differentially expressed while those of [FeFe] hydrogenases did not. Other genes involved in transport, sporulation and transposons also showed differential expression upon exposure to methyl viologen. This global profiling study provides gene targets for future studies to elucidate the relative contributions of prospective pathways for co-factor pool re-oxidation and *C. thermocellum* redox homeostasis.

Methods

Strains, Media, and Materials

C. thermocellum DSM 1313 was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ) and grown in MTC medium essentially described previously [112]. Briefly, MTC was prepared as a mixture of five solutions and contained the following (final concentrations): Solution A; 1.1 g/L cellobiose, 2 mg/L resazurin. Solution B: 2.12 g/L $C_6H_7K_3O_8$, 1.25 g/L $C_6H_8O_7$-H_20, 1 g/L Na_2SO_4, 1 g/L KH_2PO_4, 2.5 g/L $NaHCO_3$. Solution C: 1.5 g/L NH_4Cl, 2 g/L CH_4N_2O. Solution D: 1 g/L $MgCl_2$-$6H_20$, 0.2 g/L $CaCl_2$-$2H_20$, 0.1 g/L $FeCl_2$-$2H_2O$, 1 g/L $C_3H_7NO_2S$-HCl-H_2O. Solution E: 20 mg/L Pyridoxamine dihydrochloride, 4 mg/L P-aminobenzoic acid, 2 mg/L D-biotin, 2 mg/L vitamin B12. Solution F: 0.5 mg/L $MnCl_2$-$4H_2O$, 0.5 mg/L $CoCl_2$-$6H_2O$, 0.2 mg/L $ZnCl_2$, 0.05 mg/L $CuCl_2$-$2H_2O$, 0.05 mg/L H_3BO_3, 0.05 mg/L Na_2MoO_4-$2H_2O$, 0.05 mg/L $NiCl_2$-$6H_2O$. Media prepared for bottle-based batch fermentations contained 5 g/L MOPS sodium salt. Media for chemostat cultivation did not contain MOPS sodium salt. All media were made anaerobic by sparging with N_2 gas.

Preliminary Batch Fermentations

C. thermocellum DSM 1313 was grown in MTC medium containing 1.1 g/L cellobiose. Cultures were inoculated into hungate tubes containing 10 mL of medium and initial headspace of 10% CO_2 (v/v), 5% H_2 (v/v), and the balance N_2. Methyl viologen was added to medium which was then pre-warmed overnight prior to inoculation. H_20_2 was added to pre-warmed medium immediately prior to inoculation. Cultures were inoculated with 1 mL of overnight grown culture and growth was monitored using a *Milton Roy Spectronic 21D* UV-Visible Spectrophotometer (Milton Roy Company). Soluble fermentation products were measured using HPLC (see HPLC analysis portion of methods) and headspace H_2 % was measured using an Agilent 6850 GC equipped with a thermal conductivity detector (TCD) for CO_2 and H_2 quantification (Agilent Technologies, USA).

In preliminary batch fermentations, the end of fermentation was determined based on a drop-in culture OD_{600}, common in *C. thermocellum* batch fermentations and thought to correspond with the onset of cell lysis. Samples for end-product determination were gathered immediately after the final OD600 reading was taken. End-products were normalized to the average maximum OD600 achieved during batch fermentation.

Chemostat Growth and Stress Application

C. thermocellum DSM 1313 was grown at 55°C in duplicate 1 L (total vessel capacity 1.3 L) chemostat cultures using water-jacketed BioFlo110 bioreactors (New Brunswick Scientific, Edison, NJ) with 1.1 g/L cellobiose as the carbon source in MTC medium, which was fed at a dilution rate of 0.1 hr^{-1}. Temperature, pH and agitation speed were monitored and controlled during fermentation. Culture pH was monitored using pH electrodes (Mettler-Toledo, Columbus, OH) and the pH control set point was maintained at 7.0 by automatic titration with 3 N KOH. Fermenter redox potential was measured using DPAS K8S 325 combination redox probes (Mettler-Toledo) attached to an independent signal transducer and readout (Sartorious Stedim Biotech, model # 8890354). Redox probes were checked for accuracy

using new redox probe calibration solutions (Ricca Chemical Company catalog #4330-16, Ricca Chemical Company catalog #9880-16, Orion Application Solution # 967901). Agitation was supplied by a central impeller with two paddled rotors maintained at 200 rpm and no additional baffles. Culture turbidity was measured taking optical density readings at 600 nm using a Genesys 20 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA).

Fermenters were inoculated using 100 mL of overnight cultures into 900 mL of neutral MTC previously sparged overnight with N_2 gas. After inoculation, N_2 sparging was stopped and for the duration of chemostat growth N_2 was flushed into the fermenter headspace. Chemostat steady state growth was defined as being at least 50 hours of continuous growth during which the culture OD_{600} fluctuated less than 5%. Individual chemical stressors were added to medium feed carboys at final concentrations of 10 mg/L H_20_2 or 150 mg/L methyl viologen and fed into the reactors with medium feed over 60 hours (6 retention times). Chemostat methyl viologen and H_20_2 concentrations were selected based preliminary batch experiments (Figure A1 and Figure A4a, respectively) and in each case, culture growth rates were not impacted by greater than 50% and final fermentation product concentrations (normalized to maximum OD_{600}) changed by >20%. Upon completion of each chemostat stress treatment, a 500 mL volume was withdrawn from each vessel and then reactors were re-filled in fed-batch mode to 1 L with only MTC medium and then chemostat operation was resumed. Remaining stressor was allowed to wash out for another 50 hours (5 retention times) and a re-equilibration period of 50 hours followed before the next stress application. Methyl viologen and hydrogen peroxide concentrations or chemical stability were not measured or assayed directly in the fermenters or feed carboys. Redox potential was measured and recorded in the fermenters to ensure stress chemical additions were changing the fermentation redox potential and the overall redox environment. Chemostat culture integrity was checked periodically by microscopy and PCR amplification and sequencing of the 16S rRNA gene using forward sequencing primer AGAGTTTGATCCTGGCTCAG and reverse sequencing primer

GGGCGGTGTGTACAAGG. Biofilm growth or excessive frothing was not observed or mitigated during chemostat growth.

Fermentation Product Analysis using High-Performance Liquid Chromatography (HPLC)

HPLC samples were collected by centrifugation at 13,000 rpm in a microcentrifuge and passed through a 0.22 µm filter, acidified with 11.6 mN H_2SO_4 and stored at 4°C until analyzed using a LaChrom Elite System (Hitachi High Technologies America, Inc., CA) fitted with a Aminex HPX 87H HPLC column (300 x 7.8mm) (Bio-Rad, Hercules, Dallas, TX) kept at 60°C and using a mobile phase of 5 mM H_2SO_4 with a flow rate of 0.5 mL/min for 35 min per sample. Eluted compounds were detected by a refractive index detector (Model L-2490) and quantified via retention time and peak areas. Standard curves were used to quantify peak areas from samples and each sample was injected at least twice.

RNA isolation, cDNA synthesis, Microarray Hybridization and Data Analysis

Briefly, 50 mL aliquots from chemostat cultures were harvested by centrifugation (8,000 rpm, 4°C, 4 min using a Sorvall RC5C plus centrifuge), the supernatant decanted and removed and the remaining pellets were quickly frozen in liquid nitrogen and stored at -80°C. Frozen cell pellets were resuspended in TRIzol reagent (Invitrogen, Carlsbad, CA), lysed by adding 1.5 mL of cell pellet/TRIzol mixture to 2 mL screw top tubes containing 800 mg of ashed glass beads (#11079101, Biospec products, Bartlesville, OK, USA) followed by bead beating for three 20 second cycles at 6,500 rpm using a Precellys-24 (Bertin Technologies, Montigny-le-Bretonneux, France). Total cellular RNA was purified using a QIAGEN RNeasy Mini kit according to the manufacturer's instructions, which included an on-column RNase-free DNase treatment to digest residual chromosomal DNA. RNA was quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE) and a Bioanalyzer 2100 (Agilent Technologies Inc., CA). Double-strand (ds) cDNA was generated from purified RNA using a ds-cDNA synthesis kit (Invitrogen Life Technologies, NY), which was subsequently labeled, hybridized and washed according to the

NimbleGen (Roche NimbleGen, IN) protocols as described previously [72]. *C. thermocellum* transcriptome profiles were generated using an established *C. thermocellum* strain 27405 DNA microarray platform that contains 5-7 unique probes per gene and with three technical replicates for each unique probe, as described previously [14-16]. *C. thermocellum* strains DSM1313 and ATCC27045 are closely related; with average nucleotide identities (ANIs) of 99.6 and 99.3 % in reciprocal genome comparisons, indicating the strains are very closely related. Strain ATCC 27405 has a putative high affinity phosphate transport system that DSM1313 lacks [16] and ATCC 27405 contains additional prophage and restriction-modification sequences [110]; hence the ATCC27405 DNA microarray was suitable to assess *C. thermocellum* strain DSM1313. Hybridizations were conducted using a 12-bay hybridization station (BioMicro Systems, Inc., UT). Microarrays were dried and then scanned with a Surescan high-resolution DNA microarray scanner (5 μm) (Agilent Technologies, CA), and the images were quantified using NimbleScan software (Roche NimbleGen). Raw data was $\log_2$ transformed and imported into the statistical analysis software JMP Genomics 6.0 software (SAS Institute, NC). Data were normalized together using a single round of the LOWESS normalization algorithm within JMP Genomics. Distribution analyses were conducted before and after normalization as a quality control step. An ANOVA was performed in JMP Genomics to determine differential expression levels between un-treated, equilibrated conditions immediately prior to stressor exposure and time points during stressor exposure using the False Discovery Rate (FDR) testing method ($p < 0.05$) and array slide used as a random variable. Microarray data have been deposited in the NCBI GEO database [accession number GSE71465] and can be queried at the following URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi (also see attachment 3).

The Pathway Tools software [113] was used to overlay differential expression data onto a *C. thermcellum* metabolic map. The 'Omics Viewer' function within the software was used along with the curated genome-inferred metabolic model of *C. thermocellum* as provided in the Pathway Tools

Pathway/Genome Database. Log2 normalized differential transcription as computed by differential expression analysis done using JMP genomics was imported and used to overlay the data.

Microarray Validation Using Real-Time Quantitative-PCR (RT-qPCR)

Microarray data were validated using RT-qPCR, as described previously [72]. Oligonucleotide sequences of the primers targeting five genes are listed in Table A1. Correlations between differential expression values obtained by RT-qPCR analysis and microarray analysis gave an R^2 value of 0.94 indicating expression values obtained by microarray analysis are of good quality.

Methyl Viologen Incubation Experiment

C. thermocellum DSM 1313 cells prepared under different conditions (washed and resuspended in distilled water, actively growing in medium, stationary phase in medium, autoclave killed, grown with and without resazurin) were incubated in 150 mg/L methyl viologen (Sigma Aldrich) to determine the origin of the reductant being used to reduce methyl viologen by growing cultures. To prepare live cell aliquots containing medium, 1 mL of log phase and stationary phase *C. thermocellum* DSM 1313 cells were aliquoted anaerobically into 1.5mL centrifuge tubes. To prepare live cell aliquots without medium, 1 mL of live cells (both log phase and stationary phase) were aliquoted and washed three times with and resuspended in anaerobic distilled water. To prepare spent medium aliquots, 1 mL of log phase cells were aliquoted, centrifuged at 14,000 rpm for 3 minutes and filtered with a 0.22 micron syringe filter. All preparations were conducted in duplicate. Methyl viologen was added to all preparations to a final concentration of 150 mg/L, incubated anaerobically at 55^{o}C for five days and visually inspected for blue coloration as an indication of reduced methyl viologen.

Chapter 3: Targeted Redox and Energy Cofactor Metabolomics in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*

Kyle Sander[1,2,5], Keiji G. Asano[4,5], Deepak Bhandari[4,5,*], Gary J. Van Berkel[4,5], Steven D. Brown[2,3,5], Brian Davison[1,2,3,5], Timothy J. Tschaplinski[3,5,+]

Affiliations:

1. Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN

2. Bredesen Center for Interdisciplinary Graduate Research and Education, University of Tennessee, Knoxville, TN

3. Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

4. Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

5. BioEnergy Sciences Center, Oak Ridge National Laboratory, Oak Ridge, TN

* - current affiliation: Centers for Disease Control and Prevention, Atlanta, GA

+ - corresponding author

The following work was published in the journal Biotechnology for Biofuels in 2017. It is presented here as it appears in the journal, formatted to fit applicable requirements for inclusion in this thesis. The citation for this work, as published is:

Sander, K., et Al., *Targeted redox and energy cofactor metabolomics in Clostridium thermocellum and Thermoanaerobacterium saccharolyticum.* Biotechnology for Biofuels, 2017. **10**(1): p. 270.

This work provides a method by which to directly measure redox, NAD(P)(H) and energy, A(T,D,M)P, metabolites from *C. thermocellum* and *T. saccharolyticum* cell biomass. Doing so allows for a reliable representation of redox state within the cell. Sampled, 'snapshot' representations of redox state allow for many robust and useful inferences to be made about a cells general metabolic state, metabolic abilities and hypothesize about what may be inhibiting cellular metabolism. While we did not utilize this method in subsequent work outlined in this thesis, the strain assessments we did in this validation work allowed us to confirm our hypotheses generated from the previous work; that *C. thermocellum* redox metabolism is dynamic, responds to metabolic engineering, and the reduced conditions brought on by fermentation, ultimately limit ethanol production and need to be alleviated. In the first chapter, this alleviation was accomplished by adding methyl viologen, which resulted in increased ethanol yields.

Kyle Sander led efforts in adapting, optimizing, and validating the method, assisted in designing experiments, assisted in conducting experiments, assisted in analyzing and interpreting data, and drafted the submitted manuscript. Deepak Bhandari, Keiji Asano, Timothy Tschaplinski, Gary Van Berkel, Brian Davison, and Steven D. Brown all assisted in designing experiments, interpreting data, and optimizing the adapted metabolite extraction and detection method. Deepak Bhandari and Keiji Asano conducted the liquid chromatography-tandem mass spectrometry metabolite detection used extensively throughout this manuscript. Keiji Asano assisted in drafting the manuscript, assisted in preparing figures, and assisted in conducting experiments.

Abstract

Background: *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum* are prominent candidate biocatalysts that, together, can enable the direct biotic conversion of lignocellulosic biomass to ethanol. The imbalance and suboptimal turnover rates of redox cofactors is currently hindering engineering efforts to achieve higher bioproductivity in both organisms. Measuring relevant intracellular cofactor concentrations will help understand redox state of these cofactors and help to identify a strategy to overcome these limitations, however, metabolomic determinations of these labile metabolites has historically proved challenging.

Results: Through our validations, we verified the handling and storage stability of these metabolites, and verified extraction matrices and extraction solvent were not suppressing mass spectrometry signals. We recovered adenylate energy charge ratios (a main quality indicator) above 0.82 for all extractions. NADH/NAD+ values of 0.26 and 0.04 for an *adhE* deficient strain of *C. thermocellum* and its parent, respectively, reflect the expected shift to a more reduced redox potential when a species lacks the ability to reoxidize NADH by synthesizing ethanol. This method failed to yield reliable results with *C. bescii* and poor-growing strains of *T. saccharolyticum.*

Conclusions: Our validated protocols demonstrate and validate the extraction and analysis of selected redox and energy related metabolites from two candidate consolidated bioprocessing biocatalysts, *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. This development and validation highlights the important, but often neglected, need to optimize and validate metabolomic protocols when adapting them to new cell or tissue types.

Keywords: *Clostridium thermocellum, Thermoanaerobecterium saccharolyticum*, redox, adenylate energy charge, targeted metabolomics

Background

Clostridium thermocellum is a promising consolidated bioprocessing candidate microorganism capable of enzymatically degrading lignocellulosic biomass and simultaneously converting soluble hydrolyzed sugars to ethanol. Metabolic characterization and engineering efforts have afforded large improvements in overall bioproductivity [114, 115], as well as engineering for heterologous production of isobutanol [116]. *Thermoanaerobacterium saccharolyticum* is a noted anaerobic, thermophilic ethanologen which has also been extensively studied and engineered [117]. While it does not possess the lignocellulolytic capability of *C. thermocellum*, its optimum temperature and pH compliment those of *C. thermocellum* and make it a well-suited co-culture counterpart. These two species of bacteria grown together have successfully produced 38 g/L ethanol in a fermentation initiated with 92 g/L of crystalline cellulose [115].

Previous metabolic engineering efforts toward improving ethanol bioproductivity in *C. thermocellum* and *T. saccharolyticum* have largely focused on carbon forcing [32, 115, 117]. Eliminating competing fermentation end products in these two organisms did not result in maximal ethanol yield on a carbon basis. In *C. thermocellum*, such efforts increased ethanol yield, but failed to decrease carbon flux to other unwanted products, such as amino acids [118], where pathway disruption would likely generate a conditional lethal strain.

The native redox metabolism has been the sole source of reductant enabling ethanol production in these two organisms as yield and overall productivity has improved, and overall conversion and substrate utilization has increased, and larger flux demands have been placed on cellular metabolism. *C. thermocellum* intracellular redox dynamics are unconventional and still being fully elucidated [35, 38, 114, 119]. Increasing expression of the *rnf* in *C. thermocellum* was able to increase ethanol yield 30% [35], while deleting the *nfnAB* genes in *T. saccharolyticum* caused a 30% decrease in ethanol yield [120].

Enabling the bifunctional alcohol dehydrogenases to accept both NADH and NADPH to facilitate ethanol conversion, rather than NADH alone, improved tolerance to ethanol [121], and increased ethanol yield by 37.5% and 73% in *C. thermocellum* and *T. saccharolyticum*, respectively [122]. It was identified through metabolic modeling that *C. thermocellum* does not re-oxidize reduced ferredoxin fast enough to support the fermentative metabolism, leading to metabolic stalling at the pyruvate to acetyl-CoA metabolic node [38], thus highlighting the large effect NADH-dependent ferredoxin re-oxidizing activity of Rnf has on metabolic flux and ethanol productivities. Rate limitation at the catabolic step of acetyl-CoA synthesis from pyruvate is further supported by a metabolomic pulse-chase study that used ^{13}C labeled cellobiose to show the unlabeled fraction of pyruvate decreased more slowly than the central glycolytic metabolites upstream of pyruvate [123]. This slower-than-expected depletion of unlabeled pyruvate may also be due to unlabeled CO_2-derived carbon being assimilated into pyruvate from the reversible activity of Pyruvate-Ferredoxin Oxidoreductase [124], a process which is also redox-driven and can impact ethanol productivity. These *in vivo* studies suggest it is intracellular redox state and redox-driven thermodynamic limitations of key metabolic reactions that is now limiting further improvements in yield and overall productivity of ethanol in these microorganisms. A clear and validated assessment of intracellular redox cofactors would help the mechanistic understanding of this limitation further and help identify strategies to increase redox-dependent metabolic flux toward the production of ethanol. Redox-centered metabolic engineering enabled *Yarrowia lipolytica* to produce fatty acid methyl esters at the highest yield and productivity achieved [125]. The performance metrics achieved simultaneously met final titer and productivity objectives (and falling just 4% shy of the yield objective) needed for cost effective production of Biological Renewable Diesel Blendstock [126]. Similarly, growth of *Pseudomonas putida* in a bioelectrochemical cell in media containing soluble redox mediators allowed it to produce 2-keto-gluconate at 90% of theoretical maximum yield [127].

Different metabolomic techniques used previously to estimate nicotinamide redox cofactors in *C. thermocellum* have given NADH/NAD+ ratios that span a large range [128, 129] and, because of the disparity, offer little metabolic insight beyond intra-experiment relative comparisons. As the relative concentrations of these two metabolites is a tightly regulated parameter [47], it is unlikely *C. thermocellum*, grown and sampled under similar conditions in these studies, is allowing the relative abundance of these metabolites to vary so much. Reliable and validated determination of NADH/NAD+ redox couples will assist in estimating reaction directionality and net flux ratios [130] of critical redox reactions in *C. thermocellum*. Intracellular concentrations, redox state, and adenylate energy charge can make metabolic models more accurate and representative, and elucidate energetic limitations in *C. thermocellum* metabolism. Furthermore, redox cofactor measurements can help understanding of cofactor requirements and interchangeability between charge-carrying species in *C. thermocellum*, and help identify pathways responsible for electron yield losses (in this case defined as electrons that are not being directed toward biomass or ethanol production).

Redox and energy metabolites are known to be chemically labile and susceptible to degradation under routine laboratory handling [131]. We hypothesize that the extraction and detection protocols being used are affecting reported measurements and need optimization and robust validation. Typically, upon adapting metabolomic methods developed for one microorganism for use in other microorganisms, a small number of validation experiments are done addressing a few concerns, but rarely are protocols validated comprehensively. We have identified the many common issues as critically affecting metabolite extractions from microorganisms and biological tissues in general, which sometimes go un-validated before they are adapted and employed.

Critical Aspects of Metabolomics Methods

Adenylate Energy Charge

Many studies use the adenylate energy charge (AEC) as an intrinsic efficacy indicator of metabolomic extraction and detections. The adenylate energy charge is known to be maintained between 0.80 – 0.95 in most cells [42]. This value is relatively static in growing microorganisms, particularly anaerobic organisms [131]. In facultative anaerobes the adenylate energy charge only undergoes small and transient changes, upon major shifts in growth state, such as shifting to anaerobic growth from aerobic growth [131]. The AEC is known to be tightly regulated, and is also sensed and responded to by elements of cell state-regulation [132, 133]. For these reasons, the AEC is often used as an indicator for overall cell well-being [134, 135], and a decreased AEC can be a proxy for the magnitude of stress induced from experimental treatments [136, 137].

ATP is known to be a particularly labile metabolite [131], as well as the most abundant of the three adenylate nucleotides used to calculate the adenylate energy charge. The ability to observe high and physiologically relevant adenylate charges in metabolomic datasets is a key indicator of adequately careful and reliable metabolite extraction and detection. A low adenylate energy charge may indicate that one or more processing steps could be degrading ATP, as well as other exceedingly labile metabolites. The regulated stability of the AEC, and the ability to detect adenosine cofactors alongside other metabolites, makes the AEC an ideal quality control indicator of metabolomic extractions from actively growing cells.

Solvent/Extraction and Quenching

While rapid and complete metabolic quenching is important to metabolomic extractions, an equal requisite is to quench cell metabolism and extract cellular metabolites in a way that preserves labile metabolites. Other ways to protect labile metabolites are through the introduction of chemical

protectants to the extraction protocols, such as redox and pH buffers. Protection of nicotinamide species with the use of chemical additives is specific to cells and tissues, whereby each cell/tissue type requires a specific protocol [138, 139]. Previous reports which quantify nicotinamide and adenosine cofactors show it is possible to preserve these labile species through minimal, cold handling alone, without the need for chemical protectants. The solvent mixture chosen was found to be a superior global metabolomic extraction solvent, developed with special consideration for extracting nucleotides [45]. This solvent mixture is amenable to global metabolomic profiling, and achieves metabolism quenching and extraction simultaneously [140, 141], important to minimizing sample handling.

Washing, Centrifugation, and Metabolite Leakage

Washing steps are often included in metabolomic quantification protocols of intracellular metabolites, to remove extracellular species and media components prior to extracting metabolites. Washing cells can cause metabolites to leech from the cells in substantial quantities [142-145]. We are unaware of any precedent to show washing of cells is necessary during fast-filtration metabolomic extractions. We have assessed media supernatant and spent culture supernatant for the metabolites of interest in this study, and found they were not present in either (data not shown). Some studies introduce a correction to metabolite concentrations by first attempting to quantify metabolite leeching, and then using these leakage yield losses to 'correct' metabolite quantifications [146, 147]. The amount of metabolite leakage may change as a function of experimental condition, cell growth state, as well as other parameters that are often experiment specific, necessitating careful quantification of leakage losses for each experiment. If the amount of leakage is large and variable, these yield corrections may not viably represent leakage across experiments and replicates.

Centrifugation steps can last on the order of ~minutes. If centrifugation is done before metabolic quenching, the metabolite profile can change, even at decreased temperatures [148]. If centrifugation

is done after metabolism quenching, metabolite degradation or metabolite leeching from cells may ensue.

With sub-second turnover rate of many reactions and degradation mechanisms involving these metabolites, quenching within the timescales of these reactions is preferred, as is offered by direct cooled solvent quenching. Studies have shown metabolites of upper glycolysis to have turnover rates of < 1 second, even at 0^{o}C [149-151]. Though most rapid quenching/extraction methods involve submerging cells + media directly into cooled extraction solvent, which can result in >20% leakage of some metabolites [152, 153].

Metabolite Mass Spectrometry Signal Suppression

The suppression of mass spectrometry signals of metabolites is often encountered in metabolomic protocols that do not separate cells from their spent media prior to metabolite extraction [154, 155]. The IDMS (Isotope Dilution Mass Spectrometry) method, or one of the many derivatives of this method [156], is used to check for and correct signal suppression in metabolomics. Labeled extracts used in IDMS themselves are subject to degradation from handling and storage. Metabolite degradation in IDMS standards which are spiked into sample extracts could incorrectly skew correction factors and lead to inaccurately corrected data. Labeled metabolites, particularly for global metabolomics, are typically produced by growing *Escherichia coli* on 100% labeled carbon substrate in minimal media, extracting those metabolites and spiking this extract into samples to be analyzed. This method of preparing labeled metabolites for IDMS results in metabolite pools that are incompletely labeled [45], requiring additional data corrections. This method introduces extensive data augmentation, an additional potential source of error. We are not aware of a published incidence of signal suppression in metabolomic studies employing fast-filtering and solvent extraction for the targeted subset of metabolites assayed for in this study. An alternative to IDMS, particularly amenable to targeted

metabolomics of a small number of metabolites, is to validate that there is not ion suppression of target metabolites occurring prior to analyzing experimental samples. It is important to re-affirm this upon introducing new or different experimental or sampling conditions.

Other Nicotinamide Metabolite Quantification Methods

Native In Vivo Fluorescence and Fluorescent Biosensor Detection of Pyrimidine Nucleotides

Nicotinamide cofactors natively fluoresce, and this fluorescence can be used to quantify them *in vivo*. The emission spectra for NADH and NADPH are similar (abs. 366 nm, emit. 460 nm) [157] and instantaneous fluorescent measurements cannot discriminate between the two species, nor can they discriminate between bound and free forms of these cofactors. The standard potential of NAD(H) and NADP(H) differs slightly [158], as does the intracellular concentration and often the relative ratios of the oxidized and reduced species and enzymes typically do not use both interchangeably. As such, these two charge carriers are not equivalent within the cell and combined measurements of both, as reported from chemical autofluorescence, are inappropriate when attempting to infer redox information about one or the other. Further complicating *in vivo* analysis of native forms is the fact that other biomolecules can interfere with fluorescent measurements, such as FAD and other flavins [159, 160]. The fluorescence decay properties of pyrimidine cofactors are different from each other and thus allow for their determination individually *in vivo* [161-163]. Coupling fluorescence decay analysis and spectral decoupling methods allows for the *in vivo* differentiation of free and protein-bound NADH [163, 164]. Fluorescence lifetime techniques require specialized equipment, cell preparations and techniques which would likely result in cells being in a state not representative of growth. These techniques are also not amenable to large numbers of samples and replicates, nor are they compatible with simultaneous determination of other metabolites through methods such as global metabolomic profiling.

Toward achieving NADH/NAD+ determinations during active growth states and increasing throughput and flexibility of analysis, abiotic and protein based biosensors have been developed to assay the *in vivo* redox potential of NADH/NAD+. Biotic biosensors have been developed to measure NADH/NAD+ redox state directly, largely leveraging the differential affinity of the Rex transcription factor for NAD+ and NADH [165-167], or indirectly, through the use of coupled reporter systems [168]. While these biological redox sensor systems can give measurements under a variety of growth states, they are vulnerable to interference from pH, other nucleotides/metabolites, temperature [169], and the exogenous redox potential [170]. Their use to quantitatively measure NADH/NAD+ requires careful control and calibration of many parameters which affect their performance [169] and, given the difficulties in calibrating and standardizing these biosensors for all possible conditions and interferences, measurements derived from these biosensors are usually reported as relative and differential after being normalized to an appropriate control. Furthermore, genetic biosensors must be genetically integrated and functionally validated for each adapted use, a particular challenge to metabolic investigations of non-model organisms whose heterologous genetic expression tools are still being developed [171]. Though of relevance to *C. thermocellum* and other biotechnologically relevant thermophiles, there is a class of NADH/NAD+ biosensors based on T-Rex, the Rex protein from *Thermus aquaticus* [172], a thermophile with an optimum growth temperature of 70^{o}C, though the biosensor itself has not been applied, tested or adapted at elevated temperatures.

Abiotic biosensors, based on activated surface chemistries synthesized specifically to record amperometric responses to oxidation of NADH extracted from cells. These devices assay NADH from biological extractions, which must be extracted/prepared, wherein doing so requires the same considerations as addressed when preparing extracts for LC-MS/MS. As with biotic biosensors, these devices are susceptible to interference from other biomolecules present in extracted matrices. Unlike biological *in situ* biosensors, assay conditions can be carefully controlled, allowing for calibration and

absolute quantification of NADH. The detection limit for NADH in these devices is similar to those reported for MS/MS methods and *in vivo* fluorescent methods (~20-160 nM) [173, 174].

Enzymatic Cycling Assay

Enzyme cycling assays are also commonly used to detect NADH and NAD+ [138, 139]. Extractions using ~1M acid or base (depending on the metabolite being assayed for) are commonly employed with these assays. This protocol is able to chemically stabilize and detect picomolar concentrations [138], which is well below the concentrations typically found in metabolite extractions. Extensive tissue specific requirements are typically required to preserve NADH and NAD+ from degradative ability of extraction matrices [138, 139].Extractions involving high concentrations of acid or base is destructive and not amenable to concomitant measurements of other metabolites. Aside from assessing recovery in 'blank' or matrix-laden extractions, there are few other options to assess metabolomic data quality with this method. This assay is not amenable to detecting NADH and NAD+ metabolites extracted in organic containing solvents and co-extraction of adenylate cofactors to determine the AEC is not possible.

In conducting cycling assays, unwanted nicotinamide species (NADH, NAD+) in each extraction are degraded away prior to quantifying the corresponding other species. While this was shown to occur to completion in pure solution [148], many cycling assay development and adaptation papers mention incomplete destruction of unwanted nicotinamides in the extractions, which then can interfere with the assay. Incomplete destruction and conversion of interfering species is difficult to detect and account for, even when assaying for recovery of exogenously added metabolites. When assaying for small quantities of metabolites, these interferences can have a large effect. Not only does fast filtering utilize quenching and extractions designed to preserve the native state of all metabolites extracted, but all metabolites are also analyzed for simultaneously, rather than separately from different extractions,

eliminating the possibility of overestimating the concentration of nicotinamide metabolites or the entire nicotinamide pool.

In vivo NMR

In vivo NMR has been used to detect intracellular metabolite concentrations in various microbes [175], including redox and energy cofactors [176]. *In vivo* determination does not require metabolites to be extracted from cell biomass prior to detection and quantification. The main drawback from NMR metabolomics is the relatively low detection limit, which is often many orders of magnitude above metabolite concentrations found in metabolomic extracts [177]. *In vivo* NMR metabolomic methods offset this limitation by detecting metabolites from highly concentrated material, *in situ* or as extracts from large amounts of cell biomass. *In vivo* intracellular adenylate cofactor determination of *C. thermocellum* [178, 179] used highly concentrated cells and, though the cells are metabolically active, the metabolic state of these cells may not represent the metabolic state of actively growing and fermenting cells. Metabolic or metabolomic inferences between the two cell states may be only tangential. *Ex vivo* NMR-based metabolomics circumvent low detection limit limitation by extracting metabolites from relatively large amounts of cell biomass [176].

In this study we conduct a series of experiments toward qualifying a protocol for the reliable simultaneous determination of NAD(H), NADP(H) and A(T,D,M)P. Toward adopting and optimizing a protocol originally developed for use with *E. coli* [45], we obtain intracellular energy and redox cofactor concentration measurements, as well as validation experiments which address common metabolomic concerns that introduce large artifacts in other metabolomic extraction and detection protocols; metabolite leakage, degradation, yield losses and mass spectrometer signal suppression. We use a solvent quenching/extraction of filtered cell biomass followed by direct determination of metabolites using LC-MS/MS modified to include a minimum number of processing steps, and occurring at or below

0°C in an anaerobic environment. We have omitted centrifugation and washing steps to avoid metabolite leakage and, because we observe no matrix induced mass spectrometry signal suppression, omit any signal correction methods (e.g. isotope-dilution mass spectrometry, standard additions) as well. These validations also bound the quantitative possibilities of our results and add confidence to the measurements. Similar validations might be used when adapting metabolomic methods to other cell or tissue types.

Results

The Adenylate Energy Charge and Metabolomic Protocol Efficacy

A key metric and indicator of metabolite extraction efficacy and quality typically referenced is the adenylate energy charge (AEC) recovered from observed metabolites, when possible. We followed this metric while adapting the protocol of [45] and augmenting it for use with *C. thermocellum* and *T. saccharolyticum*. Through our adaptation of this protocol, we aimed to alter and improve our protocols to attain and observe increasingly higher AEC values in metabolite extracts. We find the ability to preserve ATP generally indicates we were preserving other labile metabolites as well, such as NADH (Table 4)

Metabolite extraction efficacy increases with the AEC ratio and the AEC ratio is an appropriate quality control metric to use, if possible, when extracting metabolites. The AEC is an appropriate and important measure of metabolomic process quality and efficacy, as making changes that increased the AEC, we concomitantly saw that we could extract and preserve higher concentrations of the two most labile metabolites targeted in this study, ATP and NADH.

Table 4. Adenylate energy charge (AEC) improvements observed through protocol development.

Date	AEC	NADH (µM)	ATP (µM)	Protocol Improvements from Previous
1	0.411 +/- 0.017	0	0.246 +/- 0.031	Ethanol based solvent, aggressive sonication protocol, extraction temperatures reached ~50C
2	0.804 +/- 0.009	0.123 +/- 0.006	2.31 +/- 0.13	Fast-filtering extraction and aqueous/organic extraction solvent, adapted from [45] with modifications
3	0.91 +/- 0.01	1.17 +/- 0.06	3.61 +/- 0.02	Further improved handling, removed formic acid from extraction solvent

Metabolite Recovery Through a Single Extraction

Often in metabolomic studies, sample biomass is extracted multiple times [154, 180], presumably as a precautionary measure to ensure complete extraction. We aimed to determine if extracting biomass multiple times is necessary to extract redox and energy metabolites of interest from *C. thermocellum* and *T. saccharolyticum*. We extracted unwashed cell biomass entrained on a nylon filter into 2 mL of chilled extraction solvent. To extract cell biomass more than once, the cell-containing filter was washed with an additional 1 mL of fresh solvent, to prevent carryover, and then transferred to a fresh chilled 2 mL of extraction solvent. Within error, a single extraction of cell biomass is sufficient to extract metabolites from *C. thermocellum* using the protocol developed herein (Figure 4). Furthermore, the sample-to-solvent ratio is sufficient for metabolites to be extracted in a single extraction.

In doing this experiment, we were unable to detect all seven metabolites we were attempting to detect. To prepare extracts for LC-MS/MS analysis, extract from each of the four sequential extractions were combined as shown in Fig. 6, so as to make detection of incremental increases in subsequent extractions possible. In doing so, the extract concentration of all metabolites was ~1/4 of concentrations typically observed for *C. thermocellum*. Though this finding and suggestion is made on the basis of detecting five of the seven targeted metabolites in this experiment, the levels of corresponding cofactor pool counterpart metabolites which were detected are not varying in extracts of biomass extracted multiple times. As such, It is not likely that multiple extractions are necessary or will enable better metabolite recovery and detection.

Recovery Losses Through Handling in Fast Filtering and Collection

Through adding exogenous metabolites into the extraction solvent, and performing a ‘mock’ extraction (see methods) using a filter with no cell biomass entrained in it, we enumerated metabolite recovery

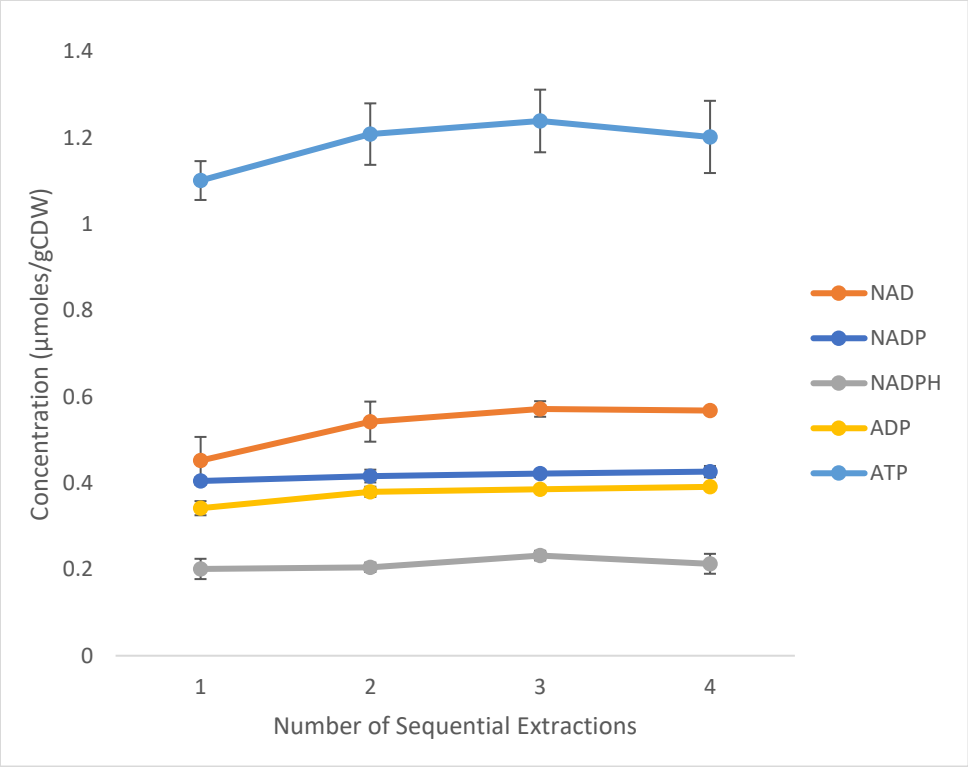


Figure 4. Cell biomass was extracted multiple times to determine if extracting biomass multiple times is necessary to recover all metabolites present in collected biomass. Using this protocol, extracting cell biomass once is sufficient for complete extraction and quantitation of metabolites. AMP and NADH were unable to be detected in this experiment.

losses which occur during a metabolite extraction. We assayed for yield losses using metabolite concentrations typical of those found in metabolite extracts of *C. thermocellum* and *T. saccharolyticum*.

ATP, ADP, NADPH and NADP incurred the largest recovery losses (Figure 5). Less than 10% of NADH and NAD+ were lost during sample handling. NADH and NAD+ loss differences are among the smallest of the seven redox and energy metabolites targeted in this study. The yield losses quantified here cannot account for large differences in NADH/NAD+ ratios observed in this study and others quantifying this parameter in *C. thermocellum* [128, 129]. Reduced nicotinamide cofactor losses cannot be accounted for in their oxidized counterparts and ATP does not appear to be hydrolyzing to ADP and AMP. Yield losses due to handling were observed in all metabolites, to varying degrees. We observed one of the largest recovery losses in ATP, though are still able to observe relatively high and physiologically relevant AEC ratios. Yield decreases across all seven metabolites suggests metabolites may have been lost to sorption to a surface or material contacted during the extraction protocol. One reason glass materials were chosen was to minimize such losses. The most likely source for these sorptive losses is the nylon filter. It may be prudent to further assess different filter material for their sorption properties and select filter media displaying appropriately low metabolite adsorption. As this mock extraction was carried out using filters containing no cell biomass, sorptive properties of filters may be different when they contain cell biomass. While these losses are non-trivial, it will serve as a basis and starting point for estimating true intracellular metabolite concentrations and inform future protocol improvements aimed at reducing these losses.

Storage Stability at -80°C.

These labile metabolites do undergo degradation at sub-zero temperatures [148]. To validate a typical storage protocol, an equimolar mixture of exogenous metabolites was prepared in fresh extraction solvent. Aliquots were frozen for prescribed lengths of time and metabolite concentration was analyzed

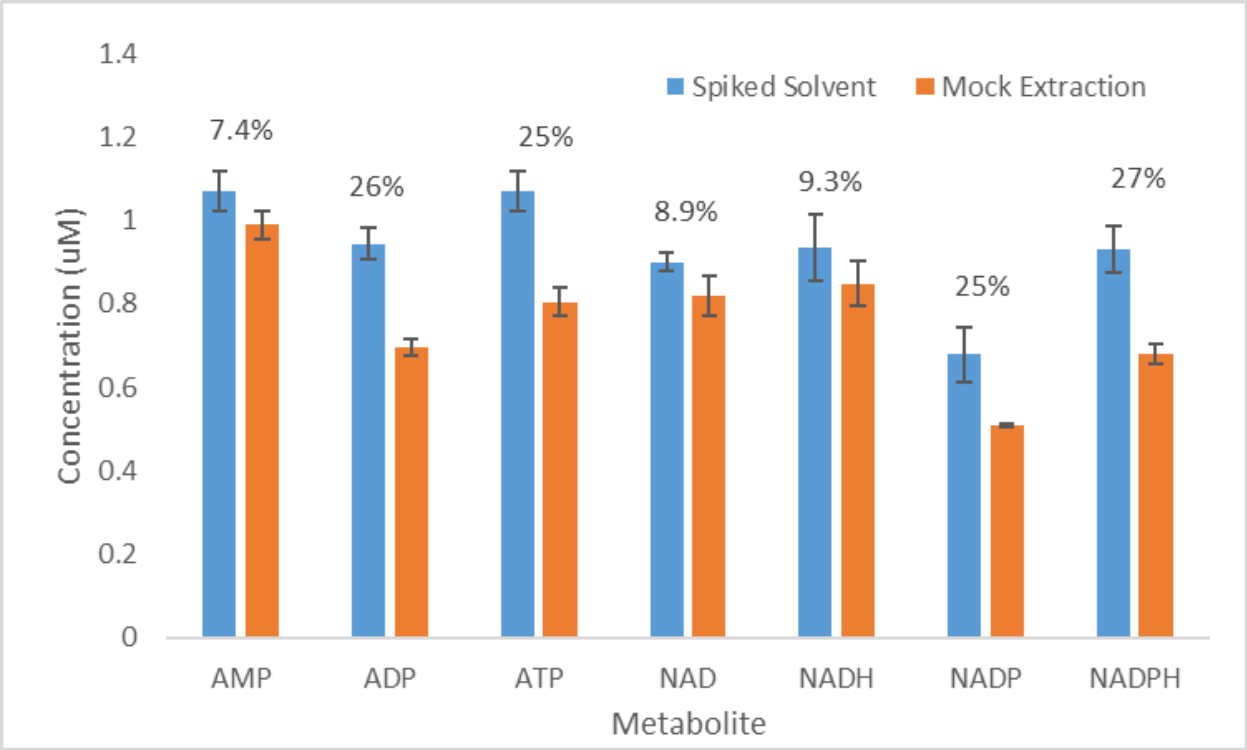


Figure 5. Extraction solvent containing exogenously added metabolites was used to conduct a 'mock-extraction' to assess metabolite losses due to handling. Blue bars indicate amount of metabolite quantified in solvent containing spiked metabolite. Orange bars indicate amount of solvent quantified in spiked solvent after one pass through a mock extraction. Noted above each metabolite is the percentage of each metabolite lost during mock-extractions relative to the amount present in the spiked solvent. A(T,D,M)P and NAD(P)(H) are susceptible to handling-related losses.

for stability over time. Each sample mixture was frozen and thawed once. Figure 6 shows metabolite concentrations of these seven metabolites, added exogenously, after being stored at -80°C for various lengths of time. All metabolites were stable when frozen at -80°C for up to five days.

No Matrix-Induced Signal Suppression of Targeted Metabolites

Signal suppression of the metabolites was assayed by first injecting either *C. thermocellum* cell extract or extraction solvent. A mixture of the seven metabolites of interest (0.25 µM each metabolite) were introduced into the stream of column eluent, creating a steady state mass spectrometry signal for each metabolite. The resulting combined signal was monitored for signal decrease at the expected retention time and *m/z* value corresponding to each metabolite. Small signal increases seen at the expected retention times of some trials appear as the result of metabolites present in the initial injected sample. Any signal decrease at the expected retention time and *m/z* value would indicate suppression of the metabolite signal by the cell extract matrix.

No signal decreases were seen at any of the expected retention times at any of the *m/z* signals in the presence of either *C. thermocellum* cell extract or extraction solvent, indicating that neither interferes with detected signals assayed for in this study (Figure 7). Instances of signal suppression were observed, but were found outside of the expected retention time, such as at ~2 minutes at *m/z* corresponding to NAD+. As cell extract matrix resulting from this extraction protocol does not produce any mass spectrometry signal interference for these seven metabolites, there is no need to correct for signal suppression.

Nicotinamide and Adenylate Cofactor Extractions from *C. thermocellum* and *T. saccharolyticum*

We have only validated other extraction aspects for *C. thermocellum,* however, we report intracellular metabolite concentrations for both *C. thermocellum* and *T. saccharolyticum* (Table 5). We have extracted and detected these redox and energy metabolites from a strain of *C. thermocellum* in which

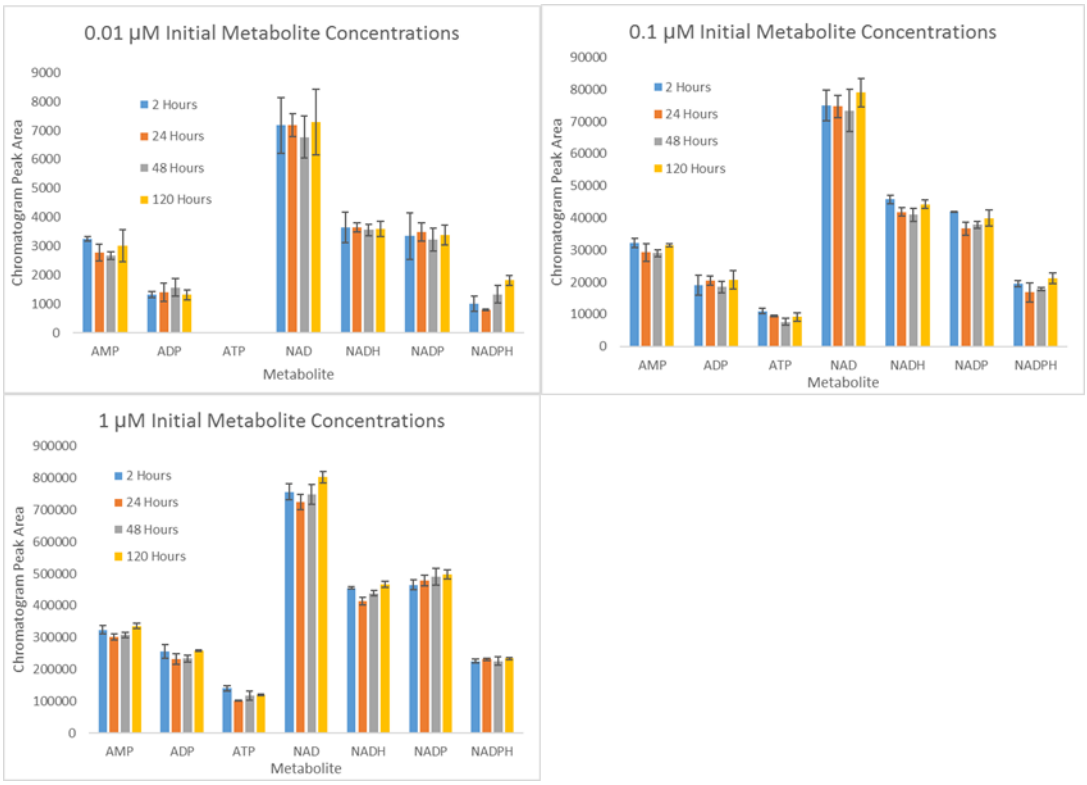


Figure 6. Storage stability of metabolites was assessed over 5 days at -80°C in extraction solvent at concentrations 0.01 µM to 1 µM. All metabolites appear stable under these storage conditions.

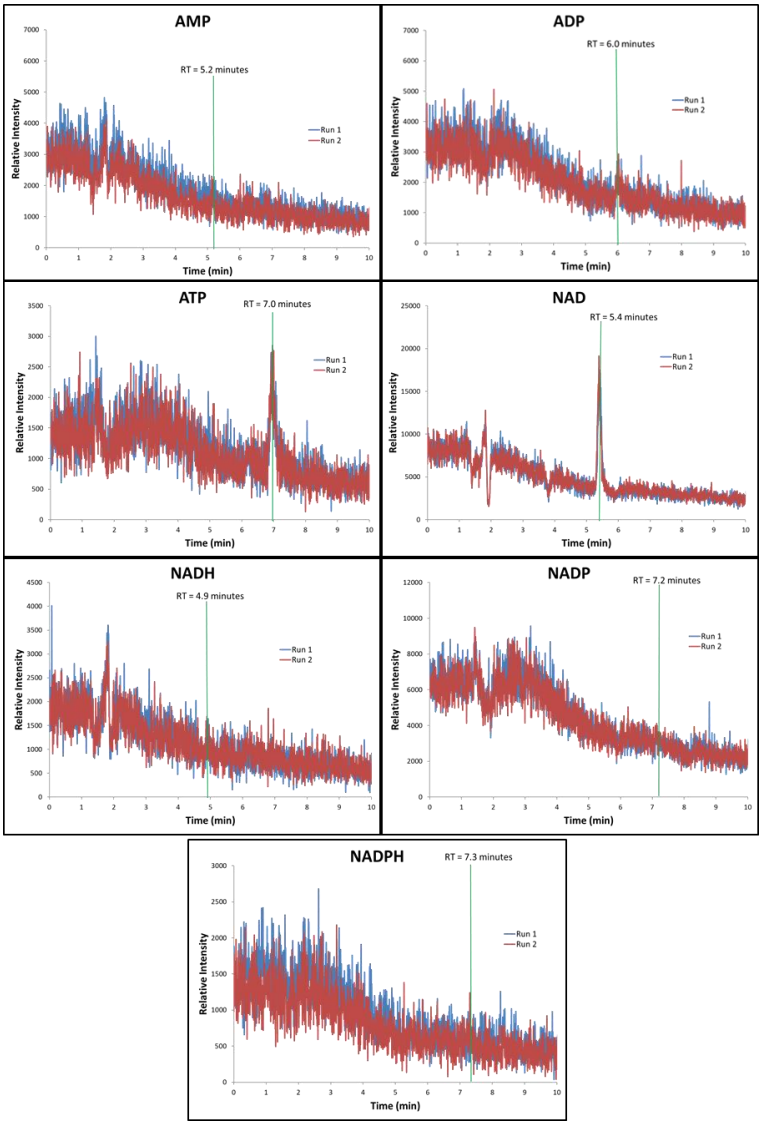


Figure 7: Mass-spectrometry signal suppression brought about by cell extract components were assessed as deflections in steady state metabolite signals (created by infusing a mixture of the seven metabolites of interest in this study into the chromatography column eluent). Predetermined retention times for each metabolite (indicated by green bars) were monitored for signal deflection, which would indicate signal suppression by the cellular extract matrix. No signal suppression was observed from extraction solvent or extraction matrix at expected retention times for metabolites. Each steady-state metabolite signal was assayed for signal suppression in the presence of cell extract twice. Both assays are shown overlaid (red and blue lines).

Table 5. Varying adenylate charge ratios observed across species highlights the need to develop protocols specific for each species/strain.

			Metabolite Concentrations (µmoles/g CDW)									
Species/Strain	**Genotype**	**Fermentation Capabilities**	**AMP**	**ADP**	**ATP**	**NAD**	**NADH**	**NADP**	**NADPH**	**NADH/ NAD+**	**NADPH/ NADP+**	**Adenylate Energy Charge**
Clostridium thermocellum												
LL345	Δhpt	wildtype	0.12±0.047	0.64±0.23	4.22±1.1	1.26±0.4	0.05±0.01	0.15±0.03	0.17±0.07	0.039683	1.13	0.91
LL1111	Δhpt ΔadhE	<5% of wt ethanol	0.53±0.14	1.87±0.54	5.64±1.14	1.83±0.42	0.48±0.23	0.49±0.04	0.71±0.15	0.262295	1.45	0.82
Thermoanaerobacterium saccharolyticum												
LL1025	wildtype	wildtype	0.16±0.078	0.88±0.21	4.4±1.26	2.41±0.51	0.06±0.01	0.52±0.1	0.09±0.02	0.024896	0.17	0.89

the bifunctional *adhE* gene has been removed [181], as well as its genetic parent strain. We find the Δ*adhE* strain to have a larger NADH/NAD+, as is expected without the function of the NADH-dependent enzyme. This relative difference in the NADH/NAD+ ratio has been shown before, though the magnitude of the ratios were much higher [129]. We also observe a much lower NADH/NAD+ ratio in *T. saccharolyticum* than was observed previously.

We also attempted to extract and detect metabolites from a Δ*adhE* strain of *T. saccharolyticum*, though the observed adenylate charge ratio was 0.69, a value too low for reliable metabolomic determination. This was likely due to the incompatibility between this poor growing strain [129] and this metabolomics protocol. As the strain grew poorly and unpredictably, it did not display a clear log phase of growth and it was difficult to discern the growth state the cells. Other studies have circumvented troubles brought about by this severe phenotype by adding yeast extract to growth media. We opted not to do this as previously, yeast-extract containing media had given interfering MS signals when assayed for previously (data not shown).

Contrary to previous findings [129], we find *T. saccharolyticum* to have a lower intracellular NADH/NAD+ ratio than *C. thermocellum*, and a much lower NADPH/NADP+ ratio. In agreement with these previous findings, we find *C. thermocellum* to have a much higher NADPH/NADP+ ratio than *T. saccharolyticum*. *T. saccharolyticum* is a noted natural ethanologen [117] and grows at a much lower optimum pH than *C. thermocellum*, suggesting it may employ far different membrane potential dynamics than *C. thermocellum*. *C. thermocellum* suffers from a large 'titer gap' [182], where it is tolerant to a far higher concentration of ethanol than it produces.

We also attempted to apply this protocol to extract and detect these metabolites from the lignocellulolytic thermophile *Caldicellulosiruptor bescii.* In our attempts, the AEC observed in metabolite extracts were 0.6-0.7, below acceptable values, likely due to the much lower total adenylate pool size

observed in *Caldicellulosiruptor* [146]. Though our protocol was very similar (with adjustments made to filter equivalent amount of cell biomass), we obtained very different results, further highlighting the need to optimize and validate metabolomic protocols for each cell type.

Discussion

The reliable metabolomic determination of labile metabolite detection requires careful considerations, beyond the considerations required of more stable metabolites. We have developed a protocol for extracting and detecting a subset of labile redox and energy metabolites, namely; ATP, ADP, AMP, NADH, NAD+, NADPH, and NADP from *C. thermocellum* and *T. saccharolyticum*. Throughout the development of this protocol, we achieved more reliable and higher quality metabolite extractions through minimizing the processing steps of our quenching and extractions, as well as ensuring cold, anaerobic culture handling up until the time samples were diluted and prepared for HPLC separation. Included in this study are a series of validations, meant to assess how various process steps can impact metabolite extraction yield at each process step. Data from these experiments can assess extraction efficacy, inform efforts to further improve this extraction protocol, or provide a format for adapting and optimizing this protocol for use in other species or cell types.

Adenylate Energy Charge and Reliable Quantifications

A hallmark of high quality metabolomic extractions is the ability to observe high and physiologically relevant adenylate energy charge ratios in metabolite extracts [45, 154]. The AEC is often cited, though briefly, as a sign that metabolites are being preserved in their physiological state [140, 154]. We used this ratio as the main indicator of quality, along with metabolite concentrations extracted and results of the various validations we did, to assess reliability of the targeted metabolomic protocol we have adapted for use in *C. thermocellum* and *T. saccharolyticum*. Based on relative increases in extraction yield, and instances of co-degradation (Table 4), we show that the ability to preserve ATP (and observe a

high adenylate energy charge) is an indicator that our extraction and detection protocol preserves other labile metabolites as well. While some studies indicate the need for acidic species, such as formic acid [140], to be present in the extraction solvent to reliably extract ATP, we found higher and more consistent AEC ratios in *C. thermocellum* and *T. saccharolyticum* using extraction solvents without formic acid. The addition of formic acid was originally empirically determined to increase adenylate detection when extracted from *E. coli [183]*, and was also suggested to aid in denaturing proteins [140], though no data was presented in support of this suggestion. The reason our empirical finding, that extractions are more effective when formic acid is omitted from the extraction solvent, differs from those made previously may be due to cell-wall structure differences between *E. coli* (gram-negative) and the organisms studied herein (both gram-positive). Gram-negative cell walls are generally considered more impervious than Gram-positive cell walls, making these organisms generally more resistant to antibiotics and able to support a chemically-isolated periplasmic space. Gram-positive cell membranes, while more structurally resistant to disruption, are typically more porous and susceptible to dyes and detergents. While these cell wall descriptions are generalities and do not always hold true, e.g. *C. thermocellum* can appear Gram-negative when subjected to a Gram stain [184], formic acid may not be necessary to extract these metabolites from the cellular matrix largely made up of gram-positive cell walls. Another important observation is that formic acid improved extraction of metabolites form cells grown aerobically, either in liquid culture [183] or grown on filter membranes supported on agar media [45]. *C. thermocellum* and *T. saccharolyticum* are both strict anaerobes. Another study employing a similar extraction protocol to measure intracellular metabolites from a strict anaerobe (*C. acetobutylicum*), also mentions using the same extraction solvent used in this study (40%/40%/20% acetonitrile/methanol/water) without the addition of formic acid [141], suggesting that growing the cells anaerobically or aerobically may determine metabolite extraction efficacy when formic acid is present in the extraction solvent.

Fast Filtering with Organic/Aqueous Solvent Precludes Matrix-Induced Ion Suppression and Simplifies Sample Handling and Analysis

We observe 9-27% yield loss during extraction handling steps of our seven targeted metabolites, with the highest losses coming from NADPH and NADP. Another study mentions not being able to extract NADPH in cold methanol, and achieved higher concentrations with perchloric acid [185], suggesting the use of organic solvents to extract NADP(H) may be suboptimal. Under analogous conditions (exogenously added cofactors, no extract matrix present) 14% of NAD+ and up to 17% of NADH was not recovered when using acid/base extraction and an enzymatic cycling assay [44]. They also mentioned observing higher yield losses of these metabolites when they were present at lower concentrations, and virtually no loss at higher concentrations, emphasizing the need to assay recovery using additions at metabolite concentrations expected in cell extracts. By comparison, we observed recovery losses of NADH and NAD+ of less than 10%. In another study, >95% of nicotinamides were recovered when coextracting exogenously added chemicals alongside *C. thermocellum* cell biomass [129], though it is unclear what concentration of exogenously added chemical was used and if the concentration used is representative of cell extract concentrations.

The present method is both sufficiently sensitive to detect ~10 nM quantities of metabolites in extracts, and is not chemically destructive. Not only does this protocol allow AEC monitoring, but also makes the protocol amenable to development of detection protocols of more metabolites are likely present in the extract [140]. Furthermore, there is no need for IDMS-based signal correction, as we found no evidence the mass spectrometry signal was being suppressed at the retention times for these seven metabolites. A similar method was used to assess intracellular metabolites in *C. acetobutylicum* [141], and appears to also have been adapted from the same protocol originally developed for *E. coli* [45]. They observed very high AEC values in log phase growing cells, though did not report any other validation experiments such as yield losses incurred at each process step.

To enable more quantitative metabolite determinations, which are imperative for making reliable thermodynamic inferences, accounting for the different types of yield loss can inform protocol improvement. Examples are losses due to leaching (or leakage), degradation, or sorptive losses, and losses during storage, and suppressed signals in detection. This is likely not an exhaustive list of yield losses, but accounts for the most documented sources of metabolite losses.

This fast-filtering protocol does not induce signal suppression during mass spectrometry detection. This is preferable as it does not require extensive sample alterations and data corrections, which both could be potential sources of error in measurements and add processing steps which might reduce extraction and detection reliability.

Redox Dynamics of Ethanol-Producing Anaerobic Thermophiles

To assess the performance and sensitivity of this protocol, we extracted metabolites from *C. thermocellum* Δ*adhE* strain (LL1111) as well as its parent (LL375). We also extracted metabolites from *T. saccharolyticum* Δ*adhE* strain (LL1076) as well as its parent (LL1025), though the AEC ratios obtained for the LL1076 strain were too low (0.69) to be considered reliable. We also tested this protocol with strains of *C. bescii*, though AEC ratios obtained for all strains were low (0.60 – 0.70), and are not discussed herein.

Two other studies [128, 129] have reported values for these metabolites in *C. thermocellum*, and have used protocols much different than the protocol developed and used in this study. We observe large differences in the NADH/NAD+ and NADPH/NADP+ ratios between all three studies, while all three studies reported similar intracellular concentrations for these metabolites. This suggests that the values obtained and reported are heavily influenced by the extraction and detection protocol used. [128] extracted metabolites from wildtype *C. thermocellum* as well as two strains that had been exposed/adapted to 3 g/L ethanol. Only in [128] was it possible to assess the AEC, as [129] used an

extraction protocol which does not preserve adenylate cofactors. Nicotinamide cofactors were recently reported from a suite of wildtype and engineered strains of *C. thermocellum* and *T. saccharolyticum* collected using a different extraction and analysis method [129]. Three of the strains analyzed were also assessed in the present study.

All but one AEC ratios calculated from reported values in [128] are below the physiological range of 0.8-0.95 for actively growing cells [42]. The WT and EA0 samples had observed adenylate charge ratios of 0.737 and 0.699 respectively, while the EA3 sample had an AEC of 0.873. The low AEC observed were likely due to the extraction protocol used; a multistep quenching and extraction followed by high pressure cell cracking, centrifugation, and filtering, all potential sources of degradation or extraction yield loss. We observed approximately the same concentration of adenylate and nicotinamide cofactors as was observed in this study using a quenching and extraction protocol with far fewer steps, and which can be completed in much less time overall, and in which cell metabolism is quenched within a few seconds rather ~10 minutes. [128] also grew *C. thermocellum* in a media containing 6 g/L yeast extract [186], which may have both altered metabolite states and, as we found in our work, may have been a source of MS signal suppression, justifying their use of stable isotope dilutions to correct for any signal suppression. Between this study and two other studies mentioned, reported values of NADH/NAD+ in wildtype, unperturbed, *C. thermocellum* strains range from 0.04 to 0.48. NADPH/NADP+ values range from 0.41 to 2.1. Redox couple ratios for the nicotinamides are reported as being much more reduced in the studies of [128, 129] than this study. Between these three studies, intracellular nicotinamide cofactor concentrations range from between 0.05 - 1.64; spanning two orders of magnitude. For comparison and context, NADH/NAD+ ratios change only two-fold when *C. acetobutylicum* shifts its metabolism from acidogenic to solventogenic [141], a major metabolic change as indicated by large shifts in intracellular metabolite profiles and AEC.

C. thermocellum uses a bifunctional alcohol dehydrogenase to produce most of the ethanol it produces [181]. This reaction is NADH dependent and the amount of ethanol produced changes in response to environmental and genetic changes [114, 187], suggesting a major determinant for ethanol production is the state of the NADH/NAD+ redox couple. As the Δ*adhE* strain does not have this enzymatic capability, relatively higher NADH/NAD+ ratios are expected in this strain, as observed in both this study and in [129].

Higher NADH/NAD+ in *C. thermocellum* than *T. saccharolyticum*

The NADH/NAD+ ratios observed in this study were approximately an order of magnitude smaller than those observed by [129]. The reason for this difference remains unclear. As many biochemical reactions can exist in a state near equilibrium, it is important to correctly determine the NADH/NAD+ ratio when making thermodynamic or directionality inferences, and a possible range spanning an order of magnitude for the same species grown in the same media and in similar conditions does not lend confidence to such inferences.

We observe a slightly higher NADH/NAD+ ratio in *C. thermocellum* when compared to mid log NADH/NAD+ ratio of *T. saccharolyticum*, though we observe relatively small intracellular NADH concentrations in both species. Higher NADH/NAD+ ratios were also observed previously in *T. saccharolyticum* relative to *C. thermocellum [129]*. A strong inverse relationship was observed between intracellular NADH/NAD+ and GAPDH activity in *C. acetobutylicum*, particularly in the range of NADH/NAD+ of 0 – 0.2 [188]. Another study demonstrates a direct link between GAPDH activity and flux through lower glycolysis in *Lactococcus lactis* [189]. This enzyme is part of the central glycolytic 'thermodynamic bottleneck' [130], a set of reactions with noticeably small free energy changes and operating at relatively low net flux ratios. This enzyme is likely operating very near equilibrium and relatively reduced GAP dehydrogenase activity, resulting from the observed higher NADH/NAD+, may

contribute ultimately to the disparity in ethanol productivity between *C. thermocellum* and *T. saccharolyticum*. NADH/NAD+ ratio are also driving reactions elsewhere in glycolysis of *C. thermocellum*, namely the Pyruvate:Ferredoxin Oxidoreductase (PFOR) in *C. thermocellum.* This enzyme has shown 'reverse' flux in *C. thermocellum*, fixing CO_2 and synthesizing formate during fermentative growth [124]. If the reaction catalyzed by PFOR is operating opposite to glycolytic flux, and has a relatively low $\Delta_rG^{o'}$ of -20 kcal/mol [190], it is reasonable that, given amenable concentrations, other glycolytic reactions might be operating similarly; at relatively low net flux ratios. In this way, the NADH/NAD+ redox couple would be heavily influencing the direction of flux for this reaction and potentially other reactions in which it participates.

Free vs. Bound Cofactors

It is only free redox cofactors that contribute to the reaction potential of those reactions which they participate in, and it is this potential we desire to estimate with quantitative, or near-quantitative estimations of metabolite concentrations. It is suggested that acetonitrile component in extraction solvents sufficiently denatures all proteins, thereby releasing would-be bound cofactors and making them available for extraction [191], though it is unclear whether this solvent mix is capable of releasing all enzyme bound cofactors [45]. It has been reported that a substantial portion of the nicotinamide cofactor pool is protein bound in mitochondria [164], while [140] found in *E. coli* that, globally, metabolites were largely in the free, unbound state within the cytosol. The whole-cell total NAD+/NADH ratio and the free/unbound NAD+/NADH portion of the pool differed by an order of magnitude in *Saccharomyces cerevisiae* and the thermodynamic potential of each of these pools would be interpreted differently [192], though total NADH/NAD+ and free NADH/NAD+ were found to approximate each other in human erythrocytes in a variety of growth states [193]. [194] used cytosolic lactate/pyruvate ratio as a proxy indicator of the free cytosolic NADH/NAD ratio, though the lactate/pyruvate ratio itself was labile. Fluorescent lifetime imaging and quantification measurements

(FLIM) can also discern free from bound NAD(H) and, thus, the free cytosolic NADH/NAD+ ratio [161], though the growth state of cells prepared for this technique may not approximate that of actively growing cells. It is unclear how well these estimates approximate those of the free, unbound portion of the pool and, thus, the thermodynamic driving force of these unbound pools.

Conclusions

To enable determining intracellular NAD(H), NADP(H), ATP, ADP and AMP concentrations in *C. thermocellum* and *T. saccharolyticum*, we have adapted and validated a cold solvent, fast-filtering protocol adapted based on a protocol developed for use with *E. coli* [45]. This protocol is validated on the basis of metabolite recovery, storage and handling stability, mass-spectrometry signal suppression, and the ability to recover physiologically relevant adenylate energy charge ratios in extractions. We compare our results with those of two similar studies utilizing different determination methods to quantify these metabolites in *C. thermocellum* and *T. saccharolyticum*. We find that our protocol recovers high adenylate energy charges and physiologically meaningful values for NADH/NAD+ and NADPH/NADP+ that are validated by other metabolomic data in the related literature. Due to tissue and extraction matrix specific needs, such validations can and should be used when adapting this and other metabolomic protocols for use in different cell and tissue types.

Methods

Strains, Media and Growth

All strains of *Clostridium thermocellum* (LL345 and LL1111) and *Thermoanaerobacterium saccharolyticum* (LL1025 and LL1076) used in this study were gifts of Lee Lynd (Dartmouth College) and his laboratory. Strain LL345 (Δ*hpt*) was used in all *C. thermocellum* metabolite extraction protocol validation experiments, unless otherwise listed. Strains of *C. thermocellum* were grown in MTC-5 media [32] and strains of *T. saccharolyticum* were grown in MTC-6 media [195]. Cultures were grown in 50 mL

aliquots in 135 mL serum bottles containing a starting gaseous headspace of 5% H_2, 10% CO_2 and the balance N_2. Cultures were grown to mid log phase at 55°C, shaking at 200 rpm. Cell growth was monitored by measuring OD_{600} measured in a Genesys 20 spectrophotometer (Thermo Fischer Scientific, Waltham, MA).

Metabolite Extractions

For Determination of Intracellular Metabolites

To determine intracellular metabolites, 5 mL of actively growing mid log phase cells were quickly aspirated and vacuum filtered onto Whatman Nylon Membrane 0.22 µm Filters (GE Healthcare Life Sciences, 7404-004). The filters were then submerged (with the filter face containing cell biomass 'down') into 2 mL of extraction solvent (Figure 8), consisting of 40% methanol (v) /40% acetonitrile (v)/20% water (v). The solvent was pre-chilled in a glass mini-petri dish (89000-300, VWR International, Radnor, PA) resting on top of an ice block which had been previously frozen at -80°C. The extraction solvent remained liquid throughout the extraction/submersion. Glass Pasteur pipettes (14672-380, VWR International, Radnor, PA) were used to collect extract and place extracts into pre-chilled silanized glass vials (MSCERT5000-S41W, Thermo Fischer Scientific, Waltham, MA). Glass vials were pre-chilled by placing them in pellet ice for ~20 minutes prior to adding extract to them. Extract was kept on pellet ice, in liquid form, and delivered for day-of LC-MS/MS analysis.

Metabolomic sampling and extractions were done in a Coy anaerobic chamber (Type B, Coy Laboratory Products, Grass Lake, MI). All glassware was brought into the anaerobic chamber 24 hours before metabolomic extractions to allow them to become anaerobic. Extraction solvent was prepared fresh for each extraction using HPLC grade solvents (water; WX0004-6, methanol; MX0488-6, acetonitrile; AX0142-6, VWR International, Radnor, PA). The solvent mixture was prepared, the headspace was sparged for 20 minutes with N_2 gas, and it was stored overnight at -20°C in the dark. Extraction solvent

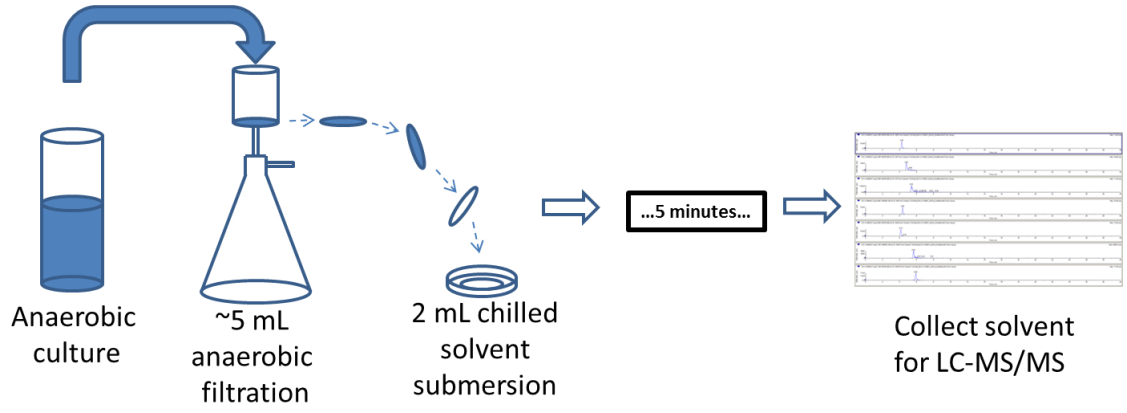


Figure 8. Diagram of the fast-filtering protocol used to extract and detect intracellular metabolites from cell biomass in this study.

was kept cold on pellet ice prior to use. Cell biomass from which metabolites were extracted (g CDW) was calculated using OD_{600} readings taken at the time of sampling and converted to cell dry weight using the conversion cited in [129].

For Determination of Metabolite Losses Due to Handling

To determine metabolite losses due to handling, a mixture containing 1.66 µM of each metabolite was prepared and chilled in glass petri dishes as described previously. A fresh filter was adhered to the filter extraction apparatus and wetted with anaerobic water that had been treated using a Barnstead Nanopure Analytical Ultrapure Water System (D11901, Thermo Fisher Scientific, Waltham, MA). This filter was then placed into 2 mL of pre-chilled extraction solvent containing metabolites and allowed to incubate for 5 minutes, to simulate handling steps used, and approximate interferences from dilution, adsorption, degradation, etc., encountered during a typical extraction. Extraction solvent from the petri dish containing solvent and the wetted filter was collected and measured. Extraction solvent containing the spiked in metabolite, but had not been used in the 'mock' extraction was also collected and metabolites quantified.

For Determination of Metabolite Yield Loss During Storage at -80°C

Metabolite mixtures containing 1 µM, 0.1 µM, and 0.01 µM of each metabolite was prepared in extraction solvent and aliquoted into silanized glass vials. The vials were frozen at -80°C. At each prescribed sampling time, one of the aliquots were thawed and analyzed for metabolite concentrations. Vials were thawed and analyzed via LC-MS/MS at 0, 24, 48 and 120 hours.

Extracting Cell Biomass Multiple Times to Determine Extraction Efficacy

To determine metabolite extraction efficacy and examine whether multiple extractions would afford more complete/quantitative extraction, *C. thermocellum* cell biomass was extracted as described above. After incubating the filter containing cell biomass for 5 minutes in pre-chilled solvent, the filter was

rinsed with an additional 1 mL of extraction solvent and placed into another glass petri dish containing 2 mL of fresh extraction solvent. The filter was rinsed and transferred two more times to fresh solvent, having the effect that the filter-laden cell biomass was exposed to fresh solvent four times sequentially. Samples were collected for LC-MS/MS analysis as shown in Figure 9.

LC-MS/MS of Intracellular Metabolites

LC-MS/MS analyses were performed using a Waters Aquity UPLC system coupled to either an ABSciex 4000 QTrap or ABSciex 5500 QTrap mass spectrometer equipped with a TurboIon Spray source. The mass spectrometer was operated in negative ion mode using multiple reaction monitoring (MRM). Chromatographic separation of metabolites was attained on a 150 mm x 2.1 mm ID, 5 µm SeQuant ZIC®*p*HILIC column (part number 1.50460.0001, Merck from VWR) using acetonitrile (mobile phase A) and 10 mM ammonium carbonate in 0.2% (v/v) aqueous ammonium hydroxide (mobile phase B). Metabolite elution was performed using a gradient from 80% A to 60% B over 15 min and holding at 60% B for 5 min and then to 80% A for a 10 minute equilibration period (30 min total run time) at a flow rate of 300 µL/min. Samples were diluted five-fold in 80/20 acetonitrile/water (v/v) and placed in an autosampler held at 4°C. Sample volume injected onto the column was 5 µL.

The mass spectrometer settings were as follows: IonSpray voltage -4.5 kV, curtain gas flow 20 (arb.), ion source gas 1 (nebulizer) flow 40 (arb.), ion source gas 2 (heating) flow 75 (arb.), nebulizing gas temperature 350°C. Ionization and collision cell parameters were optimized separately for each metabolite and are shown in table 6.

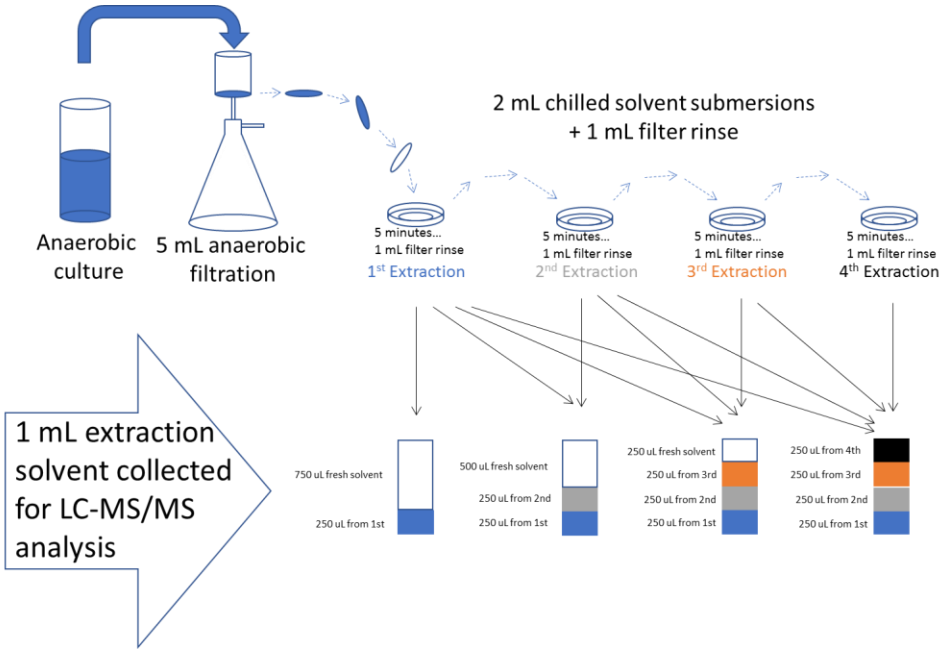


Figure 9. Diagram of method used to collect and aliquot cell extract from biomass that had been extracted multiple times.

Table 6. Ionization and collision cell parameters used to analyze metabolites in this study.

Metabolite	Product Ion (*m/z*)	Declustering Potential (DP) V	Collision Energy (CE) eV	Cell Exit Potential (CXP) V
AMP	79	-100	-60	-15
ADP	79	-105	-120	-15
ATP	79	-55	-100	-15
NAD	540.1	-70	-20	-10
NADH	79	-110	-120	-3
NADP	620.1	-60	-20	-10
NADPH	79	-110	-115	-5

Preparation of Calibration Curve

A concentrated stock solution (1 mM) of each metabolite standard was prepared in water. A concentrated mixture of metabolites (each 10 μM) was prepared by aliquoting the appropriate volume from each standard and diluting to a final volume of 5 mL in 80/20 acetonitrile/water (v/v). Serial dilutions were then made to obtain standard mixtures ranging from 0.01 μM to 1 μM. Five μL of each standard was injected onto the column. A linear calibration curve was generated by plotting the area response of the metabolite versus the concentration of the metabolite which was then used to determine the metabolite concentration in the cell extracts.

To determine yield loss of metabolites due to handling and storage, metabolite separation and analysis was done as described above, though analyzing either thawed or freshly prepared metabolite mixtures in place of a cell extract.

LC-MS/MS assessment of solvent and matrix induced signal suppression of targeted metabolites

To assess for signal suppression form cell extract matrix, chromatographic and mass spectrophotometric instruments were used as described above, with modifications. A mixture containing 0.25 μM of each metabolite was prepared in fresh extraction solvent. This mixture was infused into the elution stream exiting the chromatography column (5 μL/min standard mixture via syringe pump to 300 μL/min HPLC mobile phase flowrate), generating a steady state signal for each metabolite. Cell extract prepared, diluted, were injected and analyzed as stated above and signal suppression (indicated by deflections in the steady state signal of each metabolite) was assessed at the predetermined retention time for each metabolite.

Chapter 4: Rex in *Caldicellulosiruptor bescii*: Novel Regulon Members and its Effect on the Production of Ethanol and Overflow Metabolites

Authors: Kyle Sander[1,2,3], Daehwan Chung[3,4,+], Doug Hyatt[3,5], Janet Westpheling[3, 4], Dawn M. Klingeman[3,5], Miguel Rodriguez Jr. [3,5], Nancy L. Engle[3,5], Timothy J. Tschaplinski[3,5], Brian H. Davison[1,2,3,5], Steven D. Brown[2,3,5,# ,‡]

Affiliations

1. Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN

2. Bredesen Center for Interdisciplinary Graduate Research and Education, University of Tennessee, Knoxville, TN

3. BioEnergy Sciences Center, Oak Ridge National Laboratory, Oak Ridge, TN

4. Department of Genetics, University of Georgia, Athens, GA

5. Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

+ - Current Affiliation: National Renewable Energy Laboratory, Golden, CO

#- Corresponding Author: Please address all inquiries regarding this article to Steven D. Brown at brownsd@ornl.gov

This work, as presented, has been submitted for peer-review at the time of the publication of this thesis.

The work in this chapter investigates the hypothesis that redox metabolism can be altered through genetic modifications of a regulatory gene Rex; a redox-sensing global transcription factor in *C. bescii* responsible for regulating expression of some redox-active genes and maintaining redox state within the cell. Previous work in this thesis has established that redox metabolism and the redox state of the cell throughout fermentation are hindering ethanol synthesis, and that the redox state of these lignocellulolytic hyperthermophiles is likely dynamic and facile, susceptible to augmentation toward higher ethanol productivity. In this work we follow up on this hypothesis by broadly investigating the redox dynamics of *C. bescii*, through the study of this global transcription factor, and by assessing metabolite concentrations in strains shown to make markedly different amounts of ethanol. Further, we observe an increase ethanol productivity by eliminating Rex regulation of redox state in *C. bescii*. By alleviating Rex transcriptional control within the cell, the redox state of the cell is likely fluctuating and arriving at a new state whereby, after prolonged fermentation under nitrogen limitation, we see a 54% increase in final ethanol titer relative to the genetic parent strain containing a wildtype Rex gene.

Kyle Sander formulated hypotheses to be tested, designed and executed all experiments, generated genetic mutant strains, collected samples and data, analyzed and interpreted data, and drafted the manuscript. Daehwan Chung and Janet Westpheling graciously provided *C. bescii* genetic parent strains used in this study and assisted in generating mutant strains used in this study. Doug Hyatt performed Rex transcription factor binding site informatic prediction. Nancy Engle and Timothy J. Tschaplinski extracted metabolites from *C. bescii* biomass samples and quantified intracellular and extracellular metabolites. Dawn Klingeman assisted in cDNA generation and RNA-seq library generation. Miguel Rodriguez Jr. quantified fermentation products from supernatants by HPLC/RI detection. Brian H.

Davison and Steven D. Brown assisted in formulating hypotheses to be tested, designing experiments, interpreting data, as well as editing the manuscript.

Abstract

Rex is a global redox-sensing transcription factor that senses and responds to the intracellular [NADH]/[NAD^+] ratio, regulating its poise by regulating expression of a variety of genes in Gram-positive bacteria. We decipher and validate four new members of the Rex regulon in*Caldicellulosiruptor bescii*; a gene encoding a class V aminotransferase, the HydG FeFeHydrogenase maturation protein, a vaguely annotated oxidoreductase, and a gene encoding a hypothetical protein. Structural genes for the NiFe and FeFe hydrogenases, pyruvate:ferredoxin oxidoreductase, as well as the *rex* gene itself are also members of this regulon, as has been predicted in previous organisms. A *C. bescii rex* deletion strain constructed in an ethanol producing strain made 54% more ethanol (0.16 mM) than its genetic parent after 36 hours of fermentation, though only under nitrogen limited conditions. Metabolomic interrogation show this *rex* deficient ethanol-producing strain synthesizes other reduced overflow metabolism products likely in response to more reduced intracellular redox conditions and the accumulation of pyruvate. These results show ethanol production is strongly dependent on the native intracellular redox state in *C. bescii*, and highlight the promise of using this gene to metabolically engineer strains capable of producing ethanol at higher yield and final titer.

Importance

Redox metabolism drives the synthesis of bioethanol in *C. bescii* and other candidate consolidated bioprocessing biocatalysts, though the systems-level redox metabolism of this organism remains largely unexplored. In this study, new genomic, regulatory, and metabolic interconnections that govern *C. bescii* redox homeostasis are elucidated. Ethanol productivity differences between a *rex* deletion mutant and its parent strain constitutively expressing AdhE are explained in the context of metabolite fluxes, rather than direct control of expression by Rex, as is the case in other gram positive ethanologens. This work further expands knowledge into redox systems in *C. bescii*, the regulon of Rex, and will further inform metabolic and process engineering efforts to improve bioethanol production.

Introduction

Liquid transportation fuel demand is projected to increase through 2075 [6] and an inexpensive, reliable way to produce bioethanol from lignocellulosic feedstocks will be necessary to meet increased demand. Consolidated bioprocessing (CBP) is expected to be a cost-saving way for producing bioethanol [18], primarily because of its use of inexpensive lignocellulosic feedstocks and lower capital and operating costs due to biocatalysts that produce enzymes for the deconstruction and solubilization of feedstocks to soluble carbohydrates.

Caldicellulosiruptor bescii is a promising candidate biocatalyst for this single-step ethanol production process [196]. The genus *Caldicellulosiruptor* are anaerobic hyperthermophiles which can ferment a variety of organic substrates [14], and produce nearly theoretical amounts of hydrogen in order to recycle redox cofactors [197, 198]. It is capable of solubilizing lignocellulosic biomass through the activity of its suite of CAZymes [26, 199, 200], and fermenting many of the resulting soluble carbohydrates, producing primarily acetate, lactate, hydrogen, and CO_2. Some species can also produce ethanol, though this trait is not conserved across the genus. *C. bescii* does not natively produce ethanol as a fermentation product, but was recently engineered to produce ethanol directly from lignocellulosic biomass substrates by introducing an AdhE gene which is constitutively expressed [16, 56].

Ethanol synthesis in *C. bescii* relies on cofactors from the organism's native redox system, as is the case in other CBP organisms [31, 201]. These redox systems are plastic and subject to modulation through genetic modifications or by altering growth conditions. Eliminating lactate production in *C. bescii* increased overall hydrogen production [40], while eliminating the NiFe membrane-bound hydrogenase decreased ethanol yield in a strain expressing an exogenous bifunctional AdhE [39]. Another method of redox modulation by *Caldicellulosiruptor* is demonstrated by the closely related bacteria *Caldicellulosiruptor saccharolyticus. C. saccharolyticus* produces lactate upon sparging a continuously growing culture with hydrogen [50]. It was suggested that this results from the introduced hydrogen

inhibits hydrogen generation by hydrogenases, resulting in NADH-driven metabolic forcing of lactate dehydrogenase activity. While *Caldicellulosiruptor* species rely heavily on hydrogenases and fermentation to facilitate redox balance, other metabolic components also contribute to its redox metabolism. *C. bescii* strains expressing exogenous alcohol dehydrogenases with different redox cofactor requirements synthesized variable amounts of ethanol when grown under similar conditions [56]. These experiments demonstrate that a better understanding of cellular redox systems is needed to more effectively engineer this biocatalyst to produce ethanol more effectively.

Rex is a well characterized and conserved global redox responsive transcription factor [43] that was first characterized in *S. coelicolor* as a global repressor able to detect and regulate gene expression in response to the intracellular [NADH]/[NAD^+] couple [202]. In doing so, Rex not only regulates the poise of this redox couple, but the cellular redox state in general, as evidenced by its conserved transcriptional regulation of redox genes that are not NAD(H) dependent [43]. Conserved Rex regulon members include genes involved in energy conversion, redox metabolism, glycolytic and fermentation genes, and NAD biosynthesis [43]. Examples of genes where Rex regulation is found to be taxon specific and less conserved are those of hydrogenases, heme biosynthesis, sulfate reduction [203] and biofilm formation [204], solventogenic metabolism [48], nitrate and chlorate metabolism/resistance [205], and cytochrome biosynthesis [206].

Redox sensing and regulation by the transcriptional regulator Rex in the closely related species *Caldicellulosiruptor saccharolyticus* was partially inferred from a differential gene expression study of cells grown under hydrogen sparging [50]. Genes encoding elements of both the NiFe and FeFe hydrogenases were predicted to be under transcriptional control of Rex, as were the subunits of pyruvate:ferredoxin oxidoreductase. Lactate dehydrogenase is not predicted to be controlled by Rex in *C. saccharolyticus* [50], though it predicted to be controlled by Rex in other many other organisms [43]. Rex is predicted to be a global repressor in *C. saccharolyticus*, regulating expression of other regulatory

elements such as a histidine kinase, a CopG family transcription factor, the iron uptake regulator Fur, as well as the *rex* gene itself. Other regulatory targets of Rex and action of some regulatory elements in the regulon of *C. saccharolyticus* remain unknown and unexplored.

A consensus binding sequence for Rex was predicted in *C. saccharolyticus* [50] that is similar to predicted Rex binding sequences identified in bacteria from other genera [207]. The operator site in *C. saccharolyticus* and other organisms is an 18-20 bp palindrome sequence with the overall consensus sequence of TTGTGAANNNNTTCACAA [43]. The residues of the binding sequence important for Rex-DNA interaction have been shown to be its most conserved residues [47, 208-210].

Deleting the *rex* gene can have dramatic effects on intracellular redox state and the production of ethanol. A strain of *Clostridium acetobutylicum* containing a disrupted *rex* gene produced more ethanol and butanol after 60 hours than its parent strain or a rex deletion strain complemented with the *C. acetobutylicum rex* gene [211]. De-repression of *adhE* genes, which are transcriptionally controlled by Rex in *C. acetobutylicum* [48, 211], and increased NADH-dependent AdhE activity were shown to coincide with increases in ethanol and butanol production.

The goal of this study is to understand the redox metabolism in this unique and biotechnologically relevant species and genus, specifically as it pertains to ethanol synthesis, and we do so by studying the *rex* gene in *C. bescii*. A comprehensive understanding of redox metabolism in *C. bescii* will allow more effective engineering of redox systems to promote synthesis of ethanol at higher yield and productivity.

Results

Deletion of the *rex* Open Reading Frame

Two *rex* deletion mutants were generated for this study; one using strain JWCB005 [54] as the genetic parent strain, and the other using the ethanol-producing strain JWCB032 [16] as the parent strain. Strain JWCB005 contains a functional lactate dehydrogenase gene, while the lactate dehydrogenase gene in strain JWCB032 contains an active insertion element and the strain does not make lactate [212,

213]. Lactate production was not observed in our experiments with strain JWCB032 or the *rex* knockout derivative of this strain. Strain JWCB032 was engineered to produce ethanol as a fermentation product through heterologous expression of a *C. thermocellum adhE* gene, encoding an NADH-dependent bifunctional alcohol dehydrogenase [16]. Deletion strains were verified via PCR (Figure A6) and Sanger sequence analysis.

Bioinformatic Prediction of the Rex Regulon in *C. bescii*

Predicted promoter regions in the *C. bescii* genome (taken to be the 300 bp region upstream of all predicted open reading frames) were analyzed for the presence of a putative Rex transcription factor binding site using a *C. saccharolyticus* Rex consensus sequence [50]. Orthologous base pair matches were identified in *C. bescii* for 6 of the 11 binding sites previously identified for *C. saccharolyticus*. A total of 63 possible binding sites were identified in the *C. bescii* genome (Table A2) and scored based on their homology to the predicted *C. saccharolyticus* consensus Rex binding site sequence. We verified that all genes containing representative orthologs of those identified as Rex regulon members in *C. saccharolyticus* [50], as well as those binding sites identified for *C. saccharolyticus* in the RegPrecise database [207] were included in this list of 63 possible Rex binding sites.

Expression Profiling of JWCB005Δ*rex* and Selecting Transcription Factor Binding Sites for *In Vitro* Verification

Differential transcript expression was conducted with strain JWCB005Δ*rex* and its parent strain JWCB005 to identify genes and transcriptional units under Rex control. Samples were collected at early, mid and late log phase, (Figure A7a). Acetate synthesis was also monitored during experiments as indication of metabolic activity (Figure A7b). Within error, differences in growth or acetate production were not observed under the growth conditions tested. An average of 83.2% of RNA-seq reads that passed quality assurance aligned uniquely to the *C. bescii* genome (S.D. 1.3%), with an average per-sample genome coverage of 276X (S.D. 28). A Principle Component Analysis (PCA) showed two major

groupings of samples coinciding with the two different strains, and triplicate samples grouped together based on the different timepoints (Figure A8). Taken together, these summary analyses indicate that the RNA-seq data are of high quality.

Increased expression during at least one timepoint in JWCB005Δ*rex* relative to its parent strain was observed for 15 genes, while 9 showed decreased expression. Putative Rex binding sites were identified upstream of four genes based on differential expression analysis (ATHE_RS04105, ATHE_RS10680, ATHE_RS04720 and ATHE_RS05415). A binding site was also predicted upstream of ATHE_RS04125 (Figure 10), a putative transcriptional unit containing genes involved in tungstate transport potentially important to a tungsten containing oxidoreductase enzyme in *C. bescii* [214]. These and other binding sites from our *de novo* binding site predictions, sites predicted in the RegPrecise database entry for *C. saccharolyticus*, and those predicted in previous *C. saccharolyticus* analyses [50] for which there were *C. bescii* homologs were selected for validation. A summary of genes identified as being likely members of the Rex regulon in *C. bescii*, and binding targets were tested by electromobility shift assay (Table 7).

Electromobility Shift Assays to Test Rex Binding to Predicted *C. bescii* Binding Sites

In vitro binding assays allow analysis of recombinant Rex (rRex) binding to the predicted binding sites (Table 7) and suggest possible identification of systems under transcriptional control of Rex in *C. bescii* (figures 1-4). A series of operator sites that displayed an *in vitro* K_d between 10 and 50 nM were identified, similar to other *in vitro* Rex binding studies in other Gram-positive organisms [47, 48, 215]. Furthermore, operator sites giving these K_d values are those located in the promoter regions of genes shown to be conserved members of the Rex regulon [43].

Sequence specificity of *in vitro* binding reactions was checked by using a synthetic dsDNA probe in which

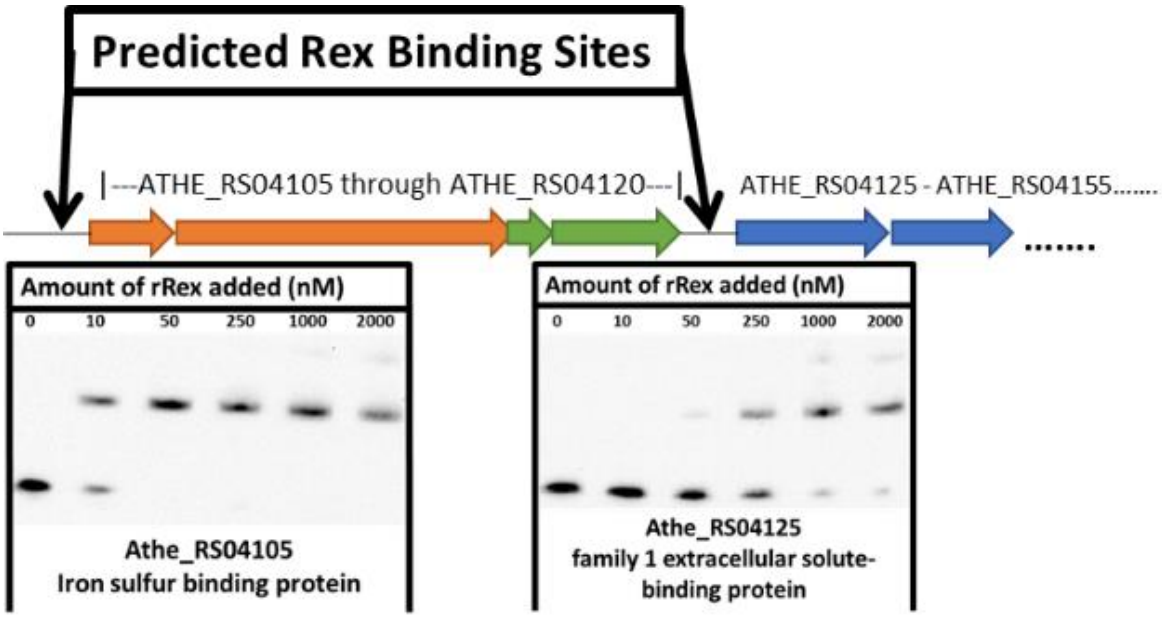


Figure 10. Electromobility shift assays of binding sites identified upstream of putative transcriptional units associated with a poorly annotated, though highly expressed, oxidoreductase gene. Rex represses a vaguely annotated, though highly expressed, tungstate-containing oxidoreductase gene [214]. rRex Binding upstream of ATHE_RS04125 was found to have a K_d much higher than other Rex binding sites, suggesting Rex does not bind to this site *in vivo*. Figure adapted from [214]. Gene color representations are as represented in [214]; Blue – tungstate transport, Green – pyranopterin biosynthesis, Orange - ferredoxin and 'XOR.'

Table 7. Rex operator binding sites chosen for *in vitro* binding validation from predicted Rex operator sites in the *C. bescii* genome. Sites were chosen from homologous sites identified in the RegPrecise [207] curated regulon, previously inferred Rex binding sites in *C. saccharolyticus* [50], and observed differential expression between JWCB005Δ*rex* and its parent strain (JWCB005). Grey shading indicates non-significant values. Black values indicate values which are significant (Benjamini-Hochburg adjusted p-value > 0.05). Putatively regulated transcriptional unit operon structure are reported as predicted in the DOOR database [216] and previous literature [198, 214].

Locus Tag Downstream of Operator Site	Predicted Transcription Unit	Distance from ATG (bp)	RegPrecise [207] Predicted Regulon	Predicted Regulon in [50]	Site Homology Prediction Score (This Study)	log_2 (JWCB005Δrex/JWCB005) of downstream gene		
						Early Log Phase	Mid Log Phase	Late Log Phase
ATHE_RS00825, Athe_0168	2 genes; CopG family transcriptional regulator, HydG hdyrogenase maturation protein	181	X	X	10.5	-0.21	-0.23	-0.27
ATHE_RS03255, Athe_0654	1 gene; Rex	41	X	X	9.25	N/A	N/A	N/A
ATHE_RS04105, Athe_0820	2 genes, Ferredoxin, 'XOR' oxidoreductase protein [214]	148			9.5	0.86	0.60	0.75
ATHE_RS04125, Athe_0824	3 genes; tungsten transport system [214]	274			8.75	0.47	0.38	0.77
ATHE_RS04390, Athe_0874	4 genes; subunits of pyruvate/ketoisovalerate:ferrodoxin oxidoreductase	112		X	9.5	-0.28	-0.73	-0.62
ATHE_RS04720, Athe_0942	2 genes; hypothetical protein, hypothetical protein	40			10	-1.46	-1.42	-0.83
ATHE_RS05415, Athe_1082	19 genes; Ech hydrogenase system	147	X	X	9	-0.43	-0.80	-1.13
ATHE_RS06475, Athe_1295	5 genes; Hyd hydrogenase system	39	X	X	10	-0.13	-0.35	-0.38
ATHE_RS10680, Athe_2126	3 genes; class V aminotransferase, phospoglycerate dehydrogenase (NADH), hypothetical protein	88			8.75	0.59	0.63	0.57
ATHE_RS11210, Athe_2226	1 gene; pyruvate carboxyltransferase (KEGG) or 2-isopropylmalate synthase LeuA (RefSeq)	169		X	8.75	0.08	0.12	0.22

the Rex operator site upstream of the rex gene were altered by introducing transverse mutations at conserved residues in the binding site. Assays using this probe showed Rex binding at a K_d ~500 nM, one order of magnitude greater than the K_d observed using an analogous probe containing the actual Rex binding site sequence, K_d ~50 nM (Figure A9), upstream of the Rex gene itself. Furthermore, probes of the same length but containing sequence found to be of low homology to the predicted *C. saccharolyticus* Rex consensus binding site were also used (Figure A10). These probes displayed no shift at concentrations of recombinant Rex protein of 2 µM.

NADH Binding by Rex in *C. bescii*

Rex protein binding to DNA with or without NADH was assayed *in vitro* (Figure 14). Binding was disrupted with as little as 5 µM of added NADH. NAD^+ had no effect on DNA binding of Rex at the concentrations assayed. The ability of *C. bescii* Rex to respond to varying [NADH]/[NAD^+] ratios at total cofactor pool concentrations of 1 µM and 100 µM was observed, suggesting that Rex in *C. bescii* operates in response to the relative [NADH]/[NAD^+] ratio rather than NADH concentration alone, as previously reported in other organisms [202].

Ethanol Productivity of a *rex* Deletion in an Ethanol Producing Strain of *C. bescii*

To test the direct impact of the Rex protein on ethanol production in *C. bescii*, we deleted the *rex* gene in an ethanol producing strain, JWCB032 [16]. JWCB032Δ*rex* showed no differences in growth profile or fermentation products relative to its parent after 48 hours when grown in LOD media (Figure A11). However, phenotypic differences were observed when these two strains were grown in LOD media augmented with $1/10^{th}$ the typical amount of ammonium carbonate (0.467 mM), the only source of soluble reduced nitrogen in LOD media [51] making this a nitrogen-limited growth condition. *C. bescii* actively grows for only 10-12 hours when grown in this condition (Figure 15a). Both strains grew to

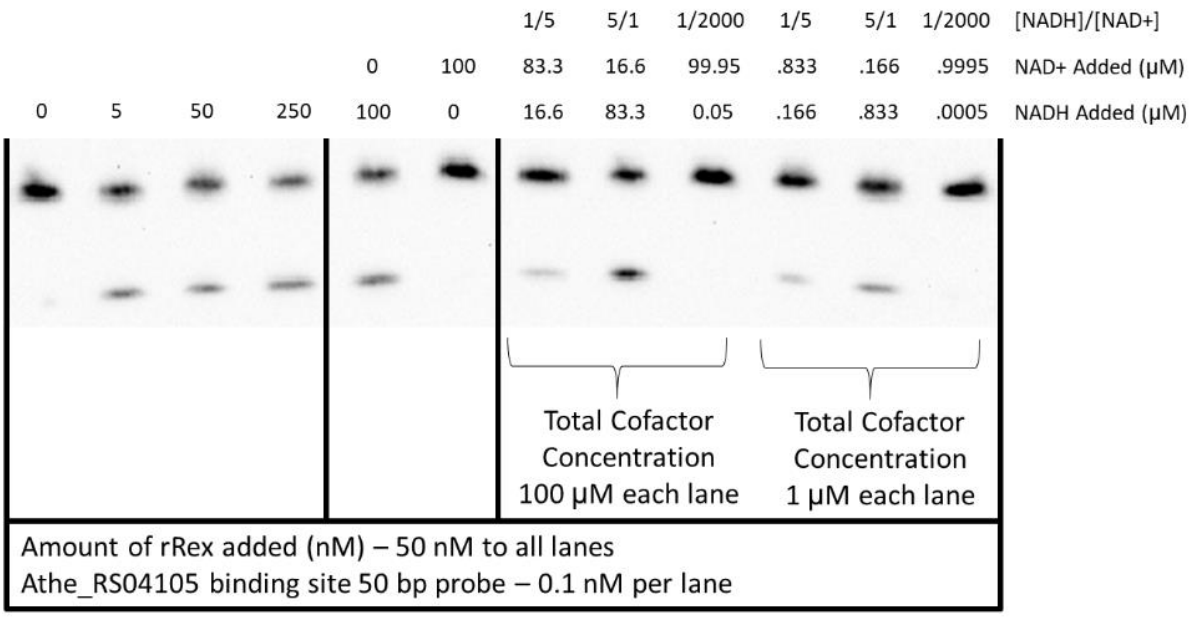


Figure 14. Electromobility shift assay showing DNA binding by Rex is disrupted by NADH and is sensitive to NADH/NAD^+ ratio across cofactor pool concentrations of 1µM to 100 µM.

similar turbidity under this growth condition, with the *rex* mutant exhibiting a lag phase about 5 hours longer than its parent strain (Figure 15). Differences in fermentation product profile appear after active growth ceases (>36 hours, Figure 15b). The *rex* knockout strain produced 54% more ethanol after 36 hours of fermentation than its parent strain (Figure 15), accounting for a 0.16 mM difference in final ethanol concentration between the two strains. The two strains produced similar amounts of acetate after 36 hours of fermentation. The relative difference in the amount of ethanol produced in JWCB032Δ*rex* relative to JWCB032 increased through 60 hours of fermentation.

To investigate ethanol production differences between the two strains under nitrogen-limiting conditions, intracellular and extracellular metabolomic profiles were generated after 36 hours of fermentation, when differences in ethanol concentrations were previously observed to be prominent (Figure 15). Succinate and 2-oxoglutarate, two TCA cycle metabolites, show a shift in TCA cycle carbon flux toward succinate production in JWCB032Δ*rex*. Lactate, glycerol/glycerol-3P, ethanol, hexadecanoate, and many amino acids concentrations were increased after 36 hours of fermentation in JWCB032Δ*rex*, particularly in the exometabolome. Mesaconate and citramalate were found in greater abundance in JWCB032Δ*rex*. Oxalomalate was found to be less abundant in, both in the pellet and the supernatant. Cystine was found in decreased abundance in JWCB032Δ*rex*, both intracellularly and in the supernatant.

Discussion

Expanding the Rex Regulon in *C. bescii*

The Rex regulon of the close *C. bescii* relative, *Caldicellulosiruptor saccharolyticus*, has been curated in the RegPrecise database [207]. RegPrecise predicts six Rex binding sites in *C. saccharolyticus*. Two of these sites are located upstream of *adhA* genes (Csac_0622 and Csac_0407) of which *C. bescii* does not have a homolog of. Homologs of the remaining four predicted Rex operator sites predicted in *C.*

saccharolyticus were found on the *C. bescii* genome (table 7). A study of gene expression in *C. saccharolyticus* under hydrogen sparging [50] found evidence to extend the putative Rex regulon of *C. saccharolyticus* beyond those genes predicted in RegPrecise. Our study validates Rex regulatory control of these previously predicted regulon members, as well as novel members of the Rex regulon not previously predicted; the HydG hydrogenase maturation factor collocated in a putative operon with a CopG family transcription factor, the 'XOR' gene [214], an unannotated highly expressed oxidoreductase gene, and a class V aminotransferase.

HydG is responsible for synthesizing a di-iron precursor to the H-cluster active site of the FeFe hydrogenase [217]. As the presence or absence of HydG protein can effectively regulate assembly of functional FeFe hydrogenase by modulating correct active site assembly [31], and the FeFe hydrogenase is a main hydrogen generation route in *C. bescii* [39], its regulation by Rex is consistent and expected. The regulon of this CopG transcription factor remains unknown and unpredicted.

Rex regulation of the 'XOR' gene is also to be expected given its presumed importance to redox metabolism in *C. bescii* [214]. While Rex is only known to sense the redox potential of the NADH/NAD^+ redox couple within the cell, it regulates redox metabolic reactions whose cofactors stretch beyond just this couple, exemplified in *C. bescii* by Rex regulation of the ferredoxin-dependent NiFe membrane bound hydrogenase (Figure 11) and of a likely ferredoxin-dependent pyruvate:ferredoxin oxidoreductase (Figure 12). While the specific function of this gene remains unknown, its relatively high expression level, unique reliance on tungsten, and coordinated expression with tungsten assimilation genes and a ferredoxin suggest its importance to redox metabolism in *C. bescii*.

Rex regulatory control of a class V aminotransferase in *C. bescii* is not well understood. A serine-pyruvate aminotransferase was identified as part of the Rex regulon of five genomes from *Thermotogales*, constituting an NADH-dependent step of a serine utilization pathway [43]. This Rex

regulated class V aminotransferase in *C. bescii* is annotated as an alanine-glyoxylate aminotransferase and no reference is given to its redox dependence. This gene is also collocated in a putative transcriptional unit with an NADH dependent 3-phosphoglycerate dehydrogenase, also a biochemical step in serine biosynthesis.

Ethanol-producing *rex*-deficient *C. bescii* produces more ethanol under nitrogen limiting conditions which extends fermentation

A *rex* deletion strain was generated using *C. bescii* strain JWCB032 [16] as the parent to investigate the effect Rex may have on ethanol production and overall redox metabolism. Differences in ethanol synthesis were observed when the cells were grown in nitrogen limiting conditions (Figure 15), though not in replete media, and only after 24 hours of fermentation. Under replete conditions, the pH of batch cultures of *C. bescii* drop to ~4.5 from an initial pH of 7.2 (data not shown) due mainly to the production of acetic acid. Growth and fermentation are not observed at or below these pH values. Growing the cells in nitrogen limiting conditions restricts active growth, the total amount of cell biomass synthesized, and acts to limit the total amount of acetate that is generated, allowing fermentation to continue long after active growth has ceased. A shift was observed in whole-cell redox state in *C. saccharolyticus* as it entered stationary phase, as indicated by an increase in lactate production and lower response when harvested cells were subjected to a poised amperometric cell [218]. This indicates redox mediated end-product shifting is occurring as cells enter stationary phase, likely favorable conditions for ethanol production as the exogenous AdhE of JWCB032 and native lactate dehydrogenase both rely on NADH to generate their respective products. *C. cellulolyticum* grown in nitrogen limited chemostat culture similarly showed an increase in fermentative flux toward ethanol as dilution rates increased [219]. Ethanol generation by heterologously expressed AdhE protein expressed in strain

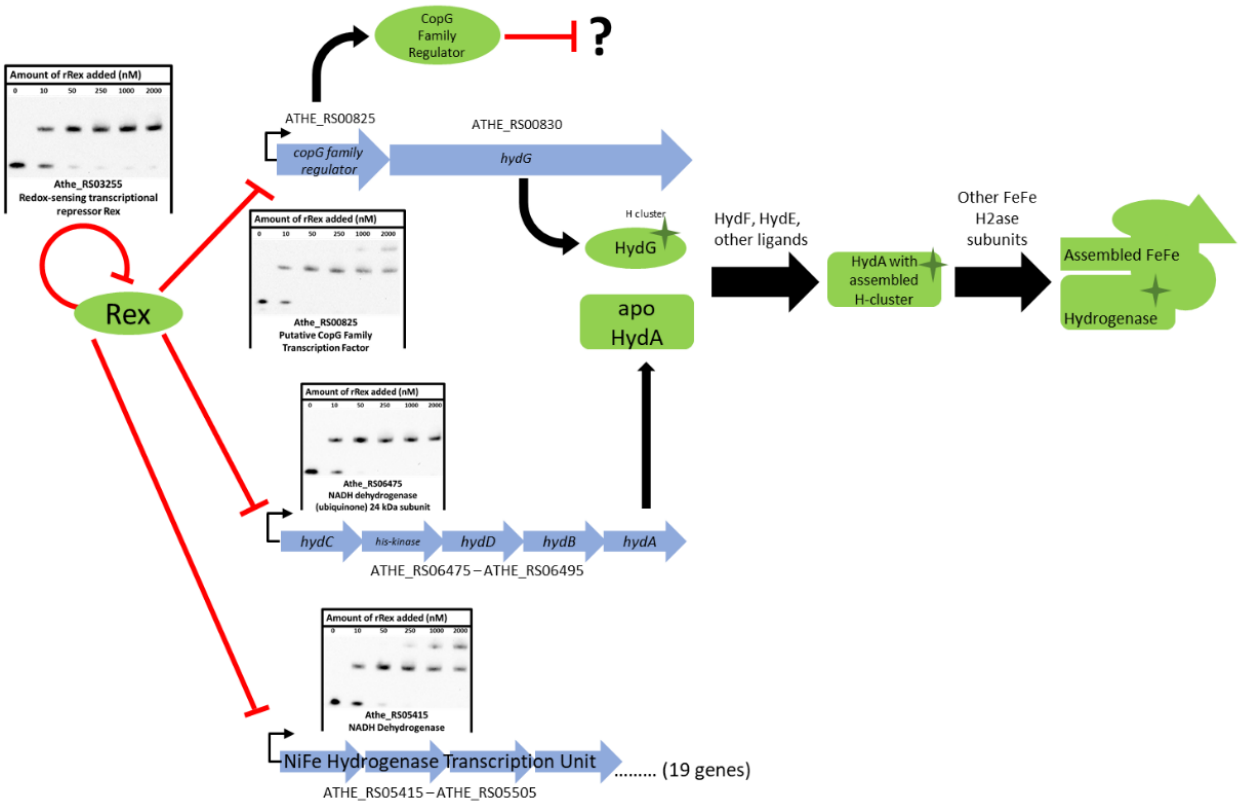


Figure 11. Electromobility shift assays and genomic orientations of Rex binding sites upstream of hydrogenase genes in *C. bescii*. Rex represses FeFe hydrogenase structural genes and *hydG*, a maturase necessary for active site assembly in FeFe hydrogenases. Rex represses expression of NiFe hydrogenase structural genes. Rex also auto-regulates itself. Genomic arrangement of *hyd* and *ech* genes inferred from homology to putative transcriptional units identified in *C. saccharolyticus* [198] and predicted transcriptional units identified using the DOOR prokaryote operon database [216]. Hyd subunit assembly scheme adapted from [217].

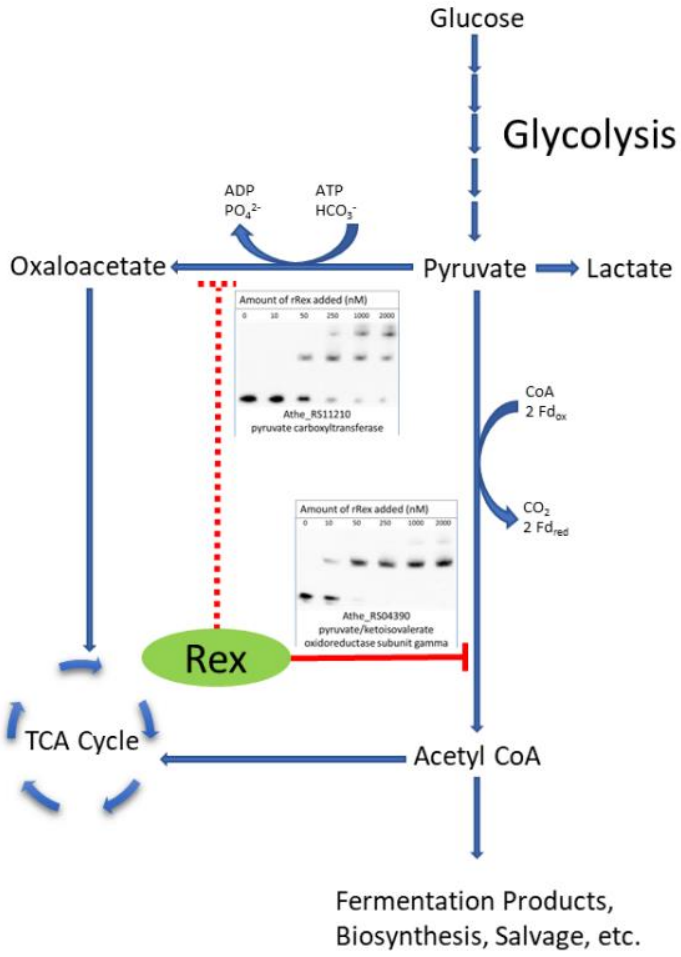


Figure 12. Electromobility shift assays of binding sites upstream of central glycolytic genes. Rex represses expression of ferredoxin-dependent pyruvate/ketoisovalerate oxidoreductase (Solid red line). Rex transcriptional control of pyruvate carboxyltransferase (ATHE_RS11210, also annotated as oxaloacetate decarboxylase), though remains doubtful because of the relatively high K_d observed *in vitro* (dashed red line).

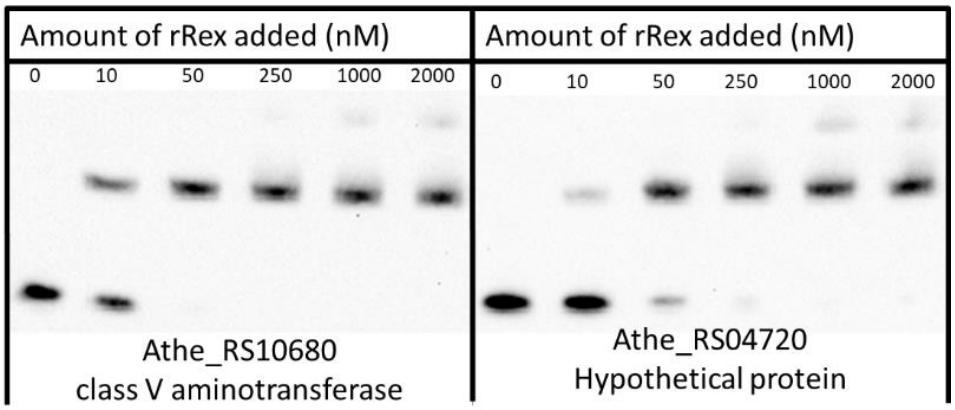


Figure 13. Electromobility shift assays of other predicted Rex binding sites whose role in redox metabolism, and *C. bescii* metabolism in general, is not well understood. Rex regulates expression of ATHE_RS10860, a class V aminotransferase as well as ATHE_RS04720, annotated as a hypothetical protein.

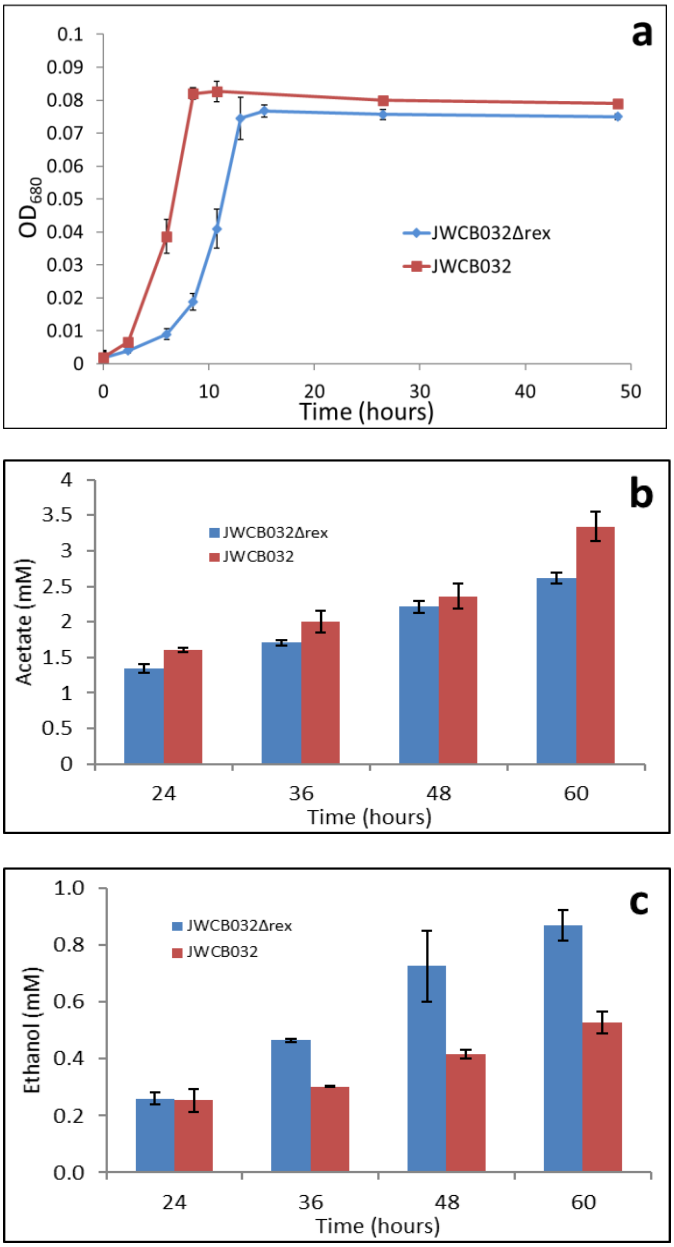


Figure 15. (a) Growth of ethanol producing JWCB032 and JWCB032Δ*rex* conducted in media containing 1/10th of typical concentration of ammonia. (b) Acetate and (c) ethanol produced by strains JWCB032Δ*rex* and JWCB032 showing a shift away from acetate and toward ethanol production after 36 hours of fermentation. Error bars represent one standard deviation of two culture replicates.

JWCB032 has been shown to be dependent on redox conditions in the native host. Perturbing redox metabolism in *C. thermocellum*, the native host of this AdhE enzyme, either through genetic modification or altering growth conditions, has predictably resulted in altered levels of the formation of ethanol and other reduced fermentation products [30, 31, 52, 121]. It is likely this AdhE enzyme is similarly sensitive to redox conditions when expressed in *C. bescii*.

Another possible explanation for elevated levels of ethanol in JWCB032Δ*rex* under nitrogen limitation would be an interaction between the *rex* regulator and nitrogen metabolism in *C. bescii,* which could be augmented in the JWCB032Δ*rex* strain deficient in this regulator. Metabolic and expression coupling between nitrogen metabolism and ethanol production has been studied in mutant strains of *C. thermocellum* [36], though no regulatory link involving Rex was found. No known link between nitrogen metabolism and ethanol production involving Rex in *C. bescii* is known to exist either.

To investigate why JWCB032Δ*rex* produced more ethanol than its parent strain, intracellular and extracellular metabolomics was conducted on this strain and its parent, JWCB032. Most metabolites found in increased abundance in strain JWCB032Δ*rex* require reduction reactions (with reductant being provided by NADH and NADPH in most cases) for their synthesis. Apart from the glycerol and glycerol-3P, inferred metabolic pathways [220] responsible for the synthesis of metabolites found in differential abundance originate at the pyruvate metabolic pathway node (Figure 16). The accumulation of pyruvate and the reduction of intracellular redox pools may support the synthesis of overflow metabolites.

Redox Buffer Systems

Increased concentrations of components of two known redox buffer systems were observed in JWCB032Δ*rex* cells; the cysteine-cystine couple [221] and the 2-oxoglutarate-glutamate couple [222, 223]. The reactions of both couples involve the interchange of one molecule of NAD(P)(H) and 2

electrons. The oxidized component of each couple was found in decreased concentration in JWCB032Δ*rex*, suggesting that these redox buffer systems are responding to more reduced redox conditions relative to those of strain JWCB032 by reducing the pools of these molecules. Cysteine is known to be one of the most oxidizing cellular redox buffer system [221]. Free cysteine has been shown to reduce iron in *E. coli*, generating hydroxyl radicals and causing DNA damage [224]. This is thought to be an evolutionary reason intracellular free cysteine levels are kept low in cells. Components of other cellular redox buffer systems were not detected in our metabolomic study of JWCB032 and JWCB032Δ*rex*. *C. bescii* redox conditions may be more readily augmented using the cysteine/cystine system rather than other redox buffering systems.

Differentially Abundant Metabolites in JWCB032Δ*rex*

Metabolites involved in glycerol metabolism, biosynthesis of the fatty acid hexadecanoate, five amino acids (isoleucine, phenylalanine, alanine, valine and leucine), TCA cycle metabolites, ethanol and acetate were found in increased abundance in JWCB032Δ*rex.* As the genes responsible for biosynthesis of these metabolites were not differentially expressed in strain JWCB005Δ*rex*, we hypothesize that these differences result from metabolite-driven flux. The metabolites glycerate, glycerol, and glycerol 3-phosphate were found in relatively increased abundance in strain JWCB032Δ*rex*, and two reactions involved in the synthesis of these metabolites (glyceraldehyde dehydrogenase and glycerol dehydrogenase) are redox dependent. There are no genes annotated in the *C. bescii* KEGG database entry [220] that encode catalytic enzymes for these reactions, nor are there any genes annotated with these functions in the recently re-annotated RefSeq annotations for the *C. bescii* DSM 6725 genome, though the presence of these molecules indicates these functions must be present *in vivo* in *C. bescii*. It is worth noting that *C. bescii* cannot grow in media containing glycerol as a sole source of carbon [14]. Increased abundance of amino acids, particularly branched chain amino acids, was also observed as

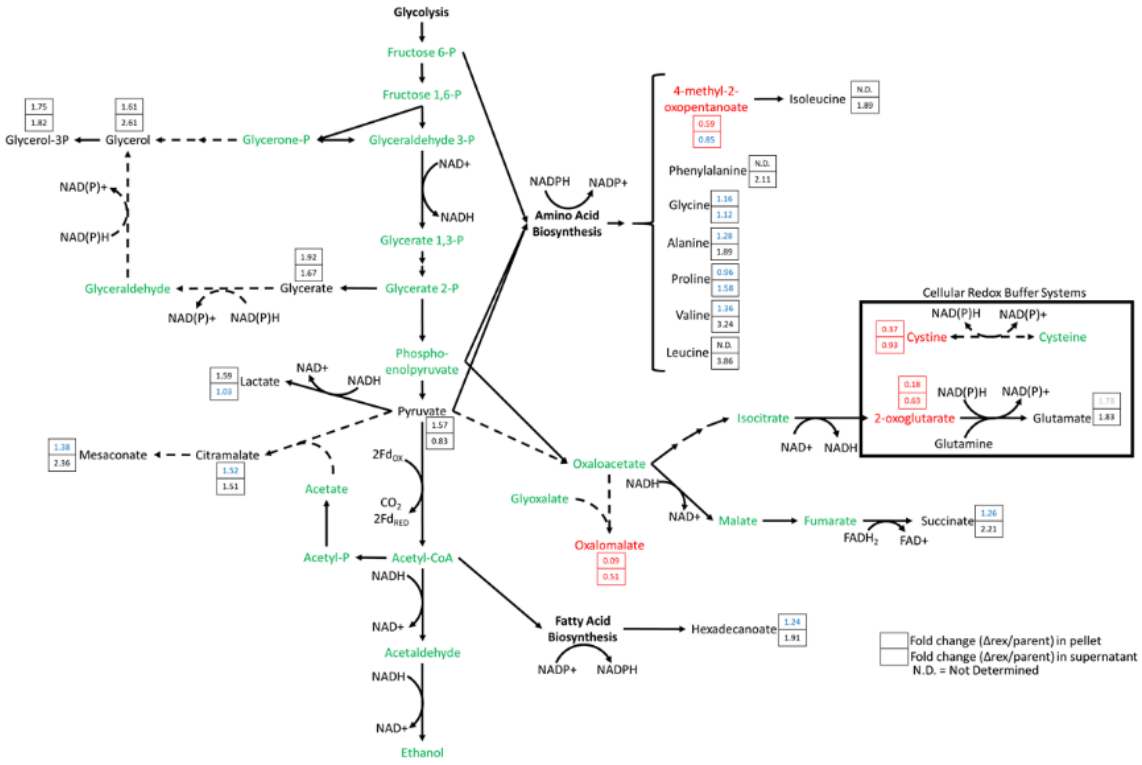


Figure 16: Differential metabolomic comparison of JWCB032Δ*rex* and its parent strain JWCB032. Metabolites indicate a metabolic shift toward reductive metabolic reactions in JWCB032Δ*rex*, indicative of more reduced intracellular redox status and possibly being driven by the accumulation of pyruvate. Solid lines indicate reactions annotated in KEGG for *C. bescii*, while dashed black lines indicate reactions not annotated in the *C. bescii* KEGG database entry. Metabolites labeled in green were either not detected or their fold-change differences were not statistically significant. Red colored metabolites showed significant decreased metabolites levels in JWCB032Δ*rex* (relative to its parent strain JWCB032) while black colored metabolites showed significantly increased metabolite levels. Black numbers indicate values that were found to be statistically significant, while blue numbers are not statistically significant. Significance was determined as p-value < 0.05 from a two-tailed students t-test (n = 4, equal variance assumed). Twelve replicate cultures were collected, and material from three cultures were combined to make one analytical replicate and four replicates per strain. Pellets and supernatants of four analytical replicates of each strain were thus analyzed, separately.

products of overflow metabolism of *C. thermocellum* [225, 226] in response to genetically induced redox perturbations and increased substrate loading. Metabolic flux of the TCA cycle and associated reactions was also redistributed toward reductive reactions, resulting in increased supernatant concentrations of succinate and glutamate and decreased supernatant and intracellular concentrations of oxalomalate and 2-oxoglutarate. Succinate production was increased 60% in *E. coli* by altering redox metabolism [227], highlighting the importance intracellular redox can have on the synthesis of succinate and TCA cycle flux.

According to the KEGG functional annotation [220] of the *C. bescii* genome, metabolites whose synthesis do not require redox reactions in reaction steps unique to their synthesis are mesaconate, citramalate, oxalomalate and acetate. Acetate and oxalomalate are competing pathways to metabolites whose synthesis pathways do depend on redox reactions, and the relatively decreased amount of these metabolites may be due to redox-driven flux through these competing pathways. Mesaconate can be synthesized from pyruvate or glutamate. Citramalate and mesaconate production are sequential metabolic steps, and because citramalate is also found in increased abundance, mesaconate is likely synthesized from pyruvate rather than from glutamate. KEGG does not annotate/assign genes for mesaconate synthesis from pyruvate or glutamate in *C. bescii*, though the most recent RefSeq annotation of the *C. bescii* DSM 6725 genome does contain an annotation for a gene encoding a citramalate synthase (ATHE_RS02505). No gene encoding a mesaconate hydratase, the enzyme synthesizing mesaconate from citramalate, could be identified. Neither of the two genes necessary for mesaconate synthesis from glutamate are annotated and protein BLAST homology searches for methylaspartate ammonia-lyase from *E. coli* and *Aspergillus oryzae* returned no genes with significant similarity. This further suggests mesaconate is being generated from citramalate as an intermediate and as an overflow metabolism product resulting from the accumulation of pyruvate.

Intracellular Redox Conditions and Pyruvate Accumulation Possibly Driving Metabolite Differences

Many reactions synthesizing differentially abundant metabolites are [NADH]/[NAD^+] dependent, and the observation of cellular redox buffer systems being active, further supports the hypothesis that metabolite differences observed in JWCB032Δ*rex* stem, in part, from redox-driven flux differences. Redox-driven overflow metabolism is observed in two other CBP organisms, *Clostridium thermocellum* [225] and *Clostridium cellulolyticum* [228]. Excess free amino acids are found in the supernatant of strains of *C. thermocellum* whose redox metabolism has been perturbed [225], and strains of *C. cellulolyticum* grown at high dilution rates [228]. Upon being challenged with these increased reductive loads, pyruvate is found in the supernatant of both species and both species synthesize lactate [225, 228]. *C. cellulolyticum* also shunts carbon at an earlier glycolytic node (Glucose 1-phosphate → Glucose 6-phosphate) toward glycogen and exopolysaccharide synthesis in response to increasing substrate loads [229]. Pyruvate accumulation and subsequent overflow metabolism has been observed and studied in detail in *C. thermocellum* [38, 225, 230, 231]. Accumulation of formate and hydrogen in *C. thermocellum* collectively restrict the re-oxidation of ferredoxin and limit its availability for pyruvate:ferredoxin oxidoreductase-enabled conversion of pyruvate to acetyl-CoA, causing accumulation of pyruvate and pyruvate-derived overflow metabolism products [38, 225, 232]. *C. cellulolyticum* growth was shown to be inhibited by high concentrations of NADH, which inhibits glyceraldehyde 3-phosphate dehydrogenase activity and limits glycolytic flux [81]. Substantial overflow metabolism and increased ethanol synthesis was observed at relatively lower [NADH]/[NAD^+] ratios in *C. cellulolyticum* [228], as to where strains of *C. thermocellum* demonstrating the most overflow metabolism and highest ethanol yields generally exhibit relatively higher [NADH]/[NAD^+] ratios [233] suggesting [NADH]/[NAD^+] ratio alone may not always determine flux directionality and flux toward overflow products. Overflow metabolism may be similarly metabolite and/or redox driven in *C. bescii*,

though direct measurement of intracellular redox conditions and/or validated metabolic modeling would further qualify this hypothesis.

Absence of Correlation Between the Rex Regulon, Differentially Expressed Genes and Differentially Abundant Metabolites

Curiously, genes downstream of some Rex binding sites did not display significant differential expression, such as ATHE_RS08255 and ATHE_RS04125, among others. Our differential expression experiment and metabolomic experiments were conducted under different growth conditions, at different sampling points, and used different strains, the absence of overlap between differentially expressed genes and differentially abundant metabolites suggests genes which synthesize the observed differentially abundant metabolites are either regulated by other transcription factors, possibly in addition to Rex, or the metabolite differences found in strain JWCB032Δ*rex* were the result of metabolite and/or redox driven flux. Of particular interest in this regard is the increased abundance of pyruvate observed in JWCB032Δ*rex* and verification of Rex expression control of genes encoding the pyruvate:ferredoxin oxidoreductase (Figure 12). One would expect de-repression of these genes (through deletion of the *rex* gene) to increase expression of this enzyme and more effectively convert pyruvate, though the opposite is observed in increased intracellular pyruvate accumulation, further suggesting metabolic differences observed are redox and metabolite driven.

Conclusions

In studying the rex gene in *Caldicellulosiruptor bescii*, we have found novel members of the Rex regulon not previously described. These include the 'XOR' gene [214], a vaguely annotated oxidoreductase gene that is highly expressed in *C. bescii*, a class V aminotransferase (ATHE_RS10680), a hypothetical protein (ATHE_RS04720), and the *hydG* gene contained in a putative transcriptional unit along with a CopG family transcription factor. The functional contribution of these genes to redox homeostasis, the

primary function of the Rex protein, remains unknonwn. A *rex* deletion mutant strain heterologously expressing a bifunctional alcohol dehydrogenase gene produced 54% more ethanol (a 0.16 mM increase in final titer) when fermentation continued for 36 hours (while limited for nitrogen). Metabolomic profiling shows differential abundance of reduced products and the shift of two known redox buffering systems toward their reduced counterparts, suggesting the elimination of the rex gene leads to a more reduced intracellular redox environment in stationary phase that in turn drives increased production of ethanol and other overflow metabolites.

Methods

Batch Growth and Fermentation

Caldicellulosiruptor bescii was grown in 50 mL culture volumes in sealed 135 mL serum bottles. The medium was comprised of components as described previously [51], with maltose being used as the primary carbon source, and ammonium chloride as the primary nitrogen source. Media were prepared, adjusted to pH of 6.8, and allowed to become anaerobic overnight through dissolved oxygen exchange in an anaerobic chamber containing an environment of 5% H_2, 10% CO_2 and the balance N_2. Cultures were grown at 75°C shaking at 200 rpm. Samples for cell growth were collected as 1 mL aliquots and measured for absorbance at OD_{680}. Nitrogen limited batch growth was conducted as outlined above with the only difference being the final concentration of ammonium chloride (the only source of reduced nitrogen in LOD media) of 0.467 mM rather than 4.67 mM as in standard LOD media.

pH Controlled Fermentation

Cells were cultured in 3L Applikon Ez-Control fermenters (Applikon Biotechnology, Delft, Netherlands) in a working culture volume of 1.5 L at a growth temperature of 75°C. Maltose, resazurin and an appropriate volume of water were added to assembled fermenters and autoclaved. Fermenters were cooled while sparging with N_2 gas. Upon cooling, other media components were added as pre-sterilized

stock solutions. The media was again heated and sparged with N_2 gas to ensure anaerobic conditions. Upon reaching 75°C, the pH was adjusted to 7.1 by sparging with an 80%/20% N_2/CO_2 gas mix. The pH was then aseptically checked using a second probe which had been calibrated with fresh pH buffers maintained at 75°C. Any differences were accounted for as the pH probe offset. Fermenters were inoculated to equivalent OD_{680} of 0.01-0.05 with batch grown cultures grown to mid-log phase. Stirring was maintained at 200 rpm without gas being sparged during growth, though the headspace outlet line was kept open to allow fermentation off-gasses to vent through a sterilized water trap.

Fermentation Product Analysis

Samples were collected from batch serum bottles or fermenters as 1 mL aliquots and centrifuged at maximum speed in a microcentrifuge for five minutes, followed by collecting and filtering (0.22µm) the supernatant. 250 µL of filtered supernatant was added to 1.75 uL of 2M H_2SO_4 and 20 µL of this mixture was injected onto a Biorad Aminex 87H column operated on a Hitachi LaChrom Ultra HPLC system (Hitachi High Technologies America, Dallas, TX). Chromatograms were collected on a Hitachi RI detector (part number L-2490).

Mutant Construction

C. bescii rex deletion mutants were constructed as described previously [234]. Strains generated for this study, necessary oligonucleotides, and plasmids used for strain generation are listed in table 8. Briefly, an integrating suicide vector was prepared containing 1000 bp of homology overlap to genomic sequence immediately 5' and 3' of the *rex* (ATHE_RS03255) coding sequence (Figure A6a). Plasmids were transformed into the respective genetic background, either strain JWCB005 [54] or strain JWCB032 [16], and transformants selected in LOD media containing no uracil to enforce prototrophic growth and plasmid integration. These cultures were subsequently plated on LOD media containing 40 uM uracil (to allow for a return to strain auxotrophy) and 6 mM 5-fluororotic acid (Oakwood Products Inc., Estill,

South Carolina) to screen and select for the double homologous recombination event. Deletions were confirmed by PCR (Figure A6b) and Sanger sequencing of the *rex* gene genomic locus.

RNA-seq Analysis

Samples were collected from fermenter-grown cultures of JWCB005Δ*rex* and its parent strain JWCB005 at early, mid and late log phase (Figure A7a). 25 mL culture aliquots were harvested, centrifuged at 20,000 x g in a Piramoon fixed angle FiberLite rotor (ThermoScientific, Waltham, MA) for 4 minutes at 4°C, decanted, snap frozen in liquid nitrogen and stored at -80°C. Total RNA was extracted by first incubating cell pellets in 250 µL of 20 mg/mL Lysozyme (Sigma Aldrich part number L-7651, St. Louis, MO) resuspended in SET buffer (50 mM Tris-HCl pH 8.0 50 mM EDTA, 20% w/v Sucrose) and incubated in a dry stationary bath at 37°C for eight minutes, vortexing briefly every two minutes. RNA was purified with a Qiagen RNEasy Kit according to manufacturer's protocol (Qiagen, Hilden, Germany). RNA concentration was quantified with a Nanodrop 1000 instrument (ThermoScientific, Waltham, MA) and RNA quality was assessed via RNA Integrity Numbers (RIN) obtained with an Agilent 2100 Bioanalyzer and corresponding RNAchip (Agilent Technologies, Santa Clara, CA). Ribosomal RNA was then depleted from total RNA samples with a RiboZero rRNA Removal Kit (Illumina Inc. San Diego, CA) following manufacturer's instructions. Next, cDNA was synthesized from RNA depleted of ribosomal RNA with a TruSeq Stranded mRNA Library Preparation Kit (Illumina Inc. San Diego, CA) following the manufacturers protocol. cDNA libraries were sequenced on an Illumina Hi-seq 2500 using v4 chemistry (Illumina Inc. San Diego, CA) and de-multiplexed as a sequencing service provided by The Genomic Services Lab at HudsonAlpha Institute for Biotechnology (HudsonAlpha, Huntsville, AL). Each sample library was sequenced on two different sequencing lanes and reads containing identical barcodes from each lane were combined for subsequent analysis. Reads obtained were scored for quality, trimmed, mapped, and the mapped reads counted with the corresponding functions in the CLC Genomics Workbench

Table 8. Primers, plasmids, and *C. bescii* strains generated and/or used in this study.

Primer Name	Sequence
pDCW88 gib assy backbone fwd	gtgcactctgacgctc
pDCW88 gib assy backbone rev	ggtaccaccagcctaac
pDCW88_athe_0654_up_fwd	tccaatgatcgaagttaggctggtggtaccatatcttcaattttgtccacagcag
pDCW88_athe_0654_up_rev	ttacataacgcattcatttcacctcaagtccttttctccccccttatcttcttttg
pDCW88_athe_0654_down_fwd	gacttgaggtgaaatgaatgc
pDCW88_athe_0654_down_rev	gttttcgttccactgagcgtcagagtgcacaacctttctaaattacttgcaacaag
upstm 5' flank fwd athe_0654_P3	agaatattgaagcgccgaac
dnstm 3' flank rev athe_0654_P3	gtggaaaaatcaccccagaa
internal fwd athe_0654_P3	gggtttggtcagcaaggata
internal rev athe_0654_P3	acccttaatcccaccttcaa
3' flank rev seq athe_0654_P3	tttgcaagatttgcgtaaga

Plasmid	Description	Reference
pETE01	Non-replicating suicide vector used to generate strains JWCB005Δ*rex* and JWCB032Δ*rex*	This study
pTXB1::*rex*	IPTG-inducible expression vector used to express recombinant Rex protein in *E. coli* T7 Express	This study

Strains	Description	Genotype	Reference
JWCB005	Genetic background strain used to generate JWCB005Δ*rex*, wildtype *ldh* locus	Δ*pyrFA* (*ura*$^-$/5-FOA^R)	[36]
JWCB005 Δ*rex*	Markerless rex deletion using strain JWCB005 as genetic parent	Δ*pyrFA* Δ*rex* (*ura*$^-$/5-FOA^R)	This study
JWCB032	Genetic background strain used to generate JWCB032Δ*rex*, *ldh* locus disrupted by ISCbe4, expressing genomically integrated bifunctional alcohol dehydrogenase (AdhE) from *C. thermocellum*	Δ*pyrFA ldh*::ISCbe4 Δ*cbeI*::Ps-layer Cthe-*adhE (ura*$^-$/5-FOA^R)	[4]
JWCB032 Δ*rex*	Markerless *rex* deletion using strain JWCB032 as genetic parent	Δ*pyrFA* Δ*rex ldh*::ISCbe4 Δ*cbeI*::Ps-layer Cthe-*adhE* (*ura*$^-$/5-FOA^R)	This study

version 8 using default settings for genome analysis of prokaryotes. Raw read counts for each coding sequence were used as input for differential expression analysis using the DEseq2 [235] package as part of the Bioconductor Suite in R. Genes were considered differentially expressed if they displayed differential normalized log_2-transformed read count abundance >0.5 or <-0.5 with a Benjamini-Hochburg adjusted p-value <0.05.

Metabolomic Profiling

Differential metabolomic profiling was conducted on strains JWCB032Δ*rex* and JWCB032 collected after 36 hours of batch serum-bottle growth in 50 mL culture volume of nitrogen limiting media, as described above. 50 mL replicates were collected, centrifuged and snap frozen as described in the sample collection section of RNA-seq analysis. Additionally, following centrifugation, the supernatants were aliquoted separately, snap frozen and stored at -80^{o}C for metabolomic analysis. Cell biomass was pooled from three 50 mL cultures to make one replicate for intracellular metabolite analysis, as was needed to attain sufficient signal intensity. Cell pellets and supernatants were analyzed for intracellular and extracellular metabolites, respectively, as described previously [225].

Transcription Factor Binding Site Prediction

The consensus Rex binding sequence identified in *C. saccharolyticus* [50] was used to seed a search genomic regions 300 bp upstream of every coding DNA sequence in the *C. bescii* genome. Binding sites were identified and scored based on similarity to this consensus binding site, yielding 63 total putative Rex binding sites across the *C. bescii* genome (Table A2). Homology scores for these sites ranged from 8.75 (least homologous) to 10.5 (most homologous).

Rex Protein Purification

Rex protein was purified by first expressing the *C. bescii rex* coding sequence on plasmid pTXB1 (New England Biolabs part number N6707S, Ipswitch, MA) upstream of the gyrase intein and chitin binding

domain (CBD), yielding plasmid pTXB1::*rex* (Table 8). This plasmid was transformed into T7 Express E. coli cells (New England Biolabs, Ipswitch, MA). Cells were grown, induced and harvested according to manufacturer's suggested instructions. Recombinant Rex (rRex) protein was purified from induced cell biomass according to protocols supplied with the IMPACT protein purification kit (New England Biolabs, Ipswitch, MA).

Electromobility Shift Assays

Electromobility shift assays were carried out to test Rex binding to predicted operator sites listed in table 7. Probes used were biotin-labeled 50-mers of double stranded DNA consisting of the 18 bp putative binding site centered on the probe and surrounding genomic sequence on either side to make 50 bp. Probes were ordered as complimentary oligonucleotides, with one oligonucleotide biotin labeled (Integrated DNA Technologies, Coralville, IA). All probes were annealed in a mixture containing 3 µL of 100 µM of each oligonucleotide, 3 µL of 10X polynucleotide kinase buffer (Roche part number 12579400, Roche Holding AG, Basel, Switzerland), 3 µL of 0.5 M NaCl, 18 µL of DEPC treated H_20. Mixtures were heated at 96^oC for 5 minutes in a Fisher Scientific Isotemp Stationary Bath (ThermoScientific, Waltham, MA) after which the heat block was removed from the bath and placed on a lab bench for 60 minutes to cool slowly. Probes were then purified using a Qiagen PCR clean-up kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

Electromobility shift assay reactions were conducted in 20 µL volumes and the following components were added in the following order: 1X EMSA binding buffer (LightShift Chemiluminescent EMSA Kit part number 20148, ThermoScientific, Waltham, MA), 25 ug dI-dC, recombinant Rex protein in various concentrations noted on figures, and 0.1 nM biotin-labeled 50 bp dsDNA probes. In reactions containing NADH or NAD^+, these components were added from freshly prepared stocks kept on ice prior to adding recombinant Rex protein but after adding other components. Reactions were incubated for 30 minutes

at room temperature after which 5 µL of EMSA loading buffer containing 15% w/v Ficoll (Sigma Aldrich, St. Louis, MO) and 0.4 % w/v Orange G (Sigma Aldrich, St. Louis, MO) was added to each reaction. 20 µL of the reaction mixture containing loading buffer was loaded onto a Novex 6% DNA Retardation Gel (Invitrogen # EC63652BOX, Life Technologies Corporation, Carlsbad, CA) and run in 0.5X TBE (Invitrogen # LC6675, Life Technologies Corporation, Carlsbad, CA) in an Invitrogen Novex Mini-Cell X Cell SureLock (Life Technologies Corporation, Carlsbad, CA) at 90 V for 45 minutes, or until Orange G from gel loading buffer had run ¾ of the length of the gel. After electrophoresis products were transferred onto Biodyne B Pre-Cut Modified Nylon 0.45 µm Membranes (Product # 77016 ThermoScientific Rockford, IL) using the Invitrogen X Cell Blot Module (catalog # EI9051) and Invitrogen Novex Mini-Cell X Cell SureLock (Life Technologies Corporation, Carlsbad, CA). Transfers were carried out in 0.5X TBE at 450 mA for 45 minutes, with the surrounding outer minicell volume packed with ice and filled with water to cool the blot module. Biotin labeled probes were detected on transfer membranes using the ThermoScientific Chemiluminescent Nucleic Acid Detection Module Kit (part number 89880, ThermoScientific, Rockford, IL) according to manufacturer's instructions.

Accession Numbers

Raw reads, processed data and experiment metadata were submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE102041 and can be queried at the following URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi (also see attachment 2)

Chapter 5: Understanding and Improving Osmotolerance in *Caldicellulosiruptor bescii* Through Metabolic Engineering of Selected Transcription Factors

Kyle B. Sander[1, 2, 3], Daehwan Chung[1,5,6], Dawn Klingeman[1,4], Miguel Rodriguez Jr.[1,4], Jason Whitham[1,4], , Brian H. Davison[1, 2, 3, 4], Jan Westpheling[1,5], Steven D. Brown[1, 2, 4,*]

1. BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN

2. Bredesen Center for Interdisciplinary Graduate Research and Education, University of Tennessee, Knoxville, TN

3. Department of Chemical and Biomolecular Engineering, University of Tennessee, Knoxville, TN

4. Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

5. Department of Genetics, University of Georgia, Athens, GA

6. National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO

* - Corresponding Author

This work is supported by the BioEnergy Science Center (BESC), which is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The manuscript has been authored by UT-Battelle, LLC, under contract no. DE-AC05-00OR22725 with the U.S. Department of Energy. The funders had no role in study design, data collection and interpretation, preparation of the manuscript, or the decision to submit the work for publication.

Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

In this final work of this thesis, we endeavor to effect performance through characterization and metabolic engineering of regulatory machinery — this time to affect the very complex trait of osmolarity resistance. This differs substantially from the third chapter because the cellular subsystems that are typically involved in survival under high osmolarity conditions are far more complex and numerous than those involved in ethanol fermentation. Furthermore, redox systems in *C. bescii* were largely either predicted previously, or could be confidently predicted from binding site assessments. Few of the systems which are responsible for survival under elevated osmolarity have been characterized in *C. bescii*. Therefore, neither the mechanisms that are responsible for *C. bescii's* relative sensitivity to osmolarity, nor the cellular mechanisms we are augmenting allowing for greater resistance to elevated osmolarity, were known before undertaking this work. Through this work, we show that engineering of regulatory machinery can best effect changes in systems with little to no *a priori* characterization. These strategies typically also require very few genetic modifications, but generate expression differences across the genome leading to desired phenotypes.

Kyle Sander formulated hypotheses to be tested, designed and conducted experiments, collected samples and data, analyzed and interpreted data, and drafted the manuscript. Daehwan Chung and Janet Westpheling assisted in generating strains needed for experiments. Dawn Klingeman led efforts to generate and sequence RNAseq libraries from collected samples. Miguel Rodriguez Jr. quantified fermentation products in samples. Jason Whitham led efforts to process raw RNAseq reads, quality check sequencing reads, trim, map, and count mapped reads, as well as helped to draft and edit the manuscript. Steven D. Brown and Brian H. Davison assisted in formulating hypotheses, designing experiments, interpreting data, and editing the manuscript.

Abstract

Through phenotype screening of 10 single-gene deletion mutants of regulatory genes in *C. bescii*, two genes important to growth in elevated osmolarities were identified. The first was the FapR fatty acid biosynthesis and malonyl-CoA metabolism repressor, which showed a more severe growth defect than its genetic parent strain when grown in high-osmolarity conditions introduced through the addition of ethanol, NaCl, glycerol, or glucose to growth media. The second was a single-gene deletion of the FruR/Cra carbon metabolism regulator, which conversely displayed a growth rate over three times higher than its genetic parent when grown in high-osmolarity media containing NaCl. Similar growth improvements were seen over the genetic parent strain in elevated levels of added ethanol, glycerol, or glucose as well. RNAseq analysis of mutant and parent strains, along previous predictions, were used to characterize the *C. bescii* FapR and FruR/Cra regulons. We find evidence of both local repression and novel global gene regulation by the FruR/Cra protein in *C. bescii*, though we are unable to determine if global gene expression differences are due directly or indirectly to the FruR/Cra protein. Genes exhibiting differential expression in a single-gene deletion of the *fapR*/*cra* gene include fructose metabolism genes collocated with the regulator, a transposase, sigma factor E, sigma factor G, two hypothetical proteins and a vaguely annotated oxidoreductase. We endeavored to generate a highly osmotolerant strain of *C. bescii* through overexpressing the FapR protein in a genetic background deficient of the *fapR*/*cra* gene, hypothesizing that such a strain will combine beneficial alterations of both phenotypes and result in a highly osmotolerant strain of *C. bescii*. This overexpression experiment is still in progress.

Introduction

Consolidated bioprocessing (CBP) is projected to be among the least expensive [236] processes being developed for the synthesis of ethanol from biomass. Critical to an effectve and economically viable CBP process is a biocatalyst organism capable of basal yield, productivity, and final titer metrics required by such a process [237, 238]. Currently no such organism exists and active efforts to engineer such an organism are underway. Each biocatalyst being engineered suffers from its own inherent limitations [12, 35, 239]. *Caldicellulosiruptor bescii*, a promising lignocellulolytic hyperthermophilic candidate CBP biocatalyst, is exceedingly sensitive to elevated osmolarities, showing pronounced growth defects at a fraction of the projected final osmolarities of CBP fermentations (>1000 mOsm/L) [12, 51, 56]. A large increase in the osmotolerance will be needed if *C. bescii* is to be used as a CBP biocatalyst.

Osmolarity stress in bacteria, in the context of bioprocessing, is typically brought about by elevated osmolarity in the extracellular environment exceeding that of the intracellular environment, causing intracellular influx of solutes and efflux of water. This decreases water activity inside of the cell and increases the intracellular concentration of potentially inhibitory solutes. Changes in growth environment osmolarity has also been shown to effect a number of other cellular behaviors and phenotypes, such as attachment [240], which metabolically interact with other cellular stress response systems, through pleiotropic regulators, such as carbon catabolite repression [241] and the expression of sigma factors [242].

Resistance to elevated osmolarity in bacteria is generally enabled through any of, or a combination of, many physiological responses. First, bacteria actively regulate their intracellular ionic environment to counter the charge-dissipating effect of ion intrusion into the cell toward maintaining cellular membrane potential [243]. Active transport processes often require chemical energy, and concomitantly ATP generating processes will be activated to supply ATP needed for active ion extrusion [244]. ATP

generation during osmostress has been detailed in *Synechosystsis sp.* PCC6803, where ATP generation is coupled to photosynthesis [245].

Some bacteria are also known to synthesize metabolites, such as trehalose or glutamate, which remain in the cellular cytoplasm to increase intracellular osmolarity and bring it closer to that of the environment outside the cell [243]. These synthesized counter-osmolytes typically do not interfere with cellular processes, and can also act as chaperones and promote correct protein folding and function [246].

Lastly, bacteria will adjust the fatty acid composition of their cell membranes, to increase membrane fluidity and decrease chemical permeability, making cells less susceptible to the physical effects of osmotic pressure [247]. An increase in anionic membrane lipids [248], and cardiolipin [249] typically accompanies an increase in extracellular osmolarity in bacteria and other cells. This shift in membrane lipid distribution is found to be beneficial to the function of osmotically driven membrane transport proteins, as well as mechanosensitive membrane proteins, both important to survival in elevated osmolarities [243]. Biophysical modeling suggests membrane lipid differences in the chemical structure of membranes of osmo-stressed cells simultaneously promote increased membrane fluidity and stability [250].

The response to osmotic stress, beyond observed effect on growth rate and product formation, has not been explicitly studied in the genus *Caldicellulosiruptor*, and it remains unclear which adaptation systems, if any, are utilized to counter the effects of elevated osmolarity. The effects of osmolarity on growth and fermentation productivity were empirically modeled in the closely related species *Caldicellulosiruptor saccharolyticus* [251]. It was found that the effect of elevated osmolarity was secondary to that of the inhibitory effect of dissolved hydrogen during continuous growth at moderate dilution rate. Hydrogen is able to super-saturate in this cultivation system [251], and the

Caldicellulosiruptor genus relies heavily on hydrogen production for ferredoxin turnover [252]. The growth conditions employed contained relatively low levels of dissolved osmolytes (e.g. 5 and 10 g/L initial glucose loading), and did not explicitly assess for osmotolerance across a broad range of osmolarities. It was found that decreases in growth and product evolution become prominent at 0.2 Osm/L, agreeing with an earlier finding that batch growth rates of *C. saccharolyticus* are reduced by 60% at 0.218 Osm/kg water [253].

It was found that adding >80 mM of the biological buffer MOPS (3-(N-morpholino)propanesulfonic acid), reduced the growth rate substantially [56]. After alleviating the ammonia limitation, it was observed that actively growing cells are able to continue growth in 520 mOsm/L, and able to initiate log-phase growth in media with a starting osmolarity >550 mOsm/L (mostly from dissolved glucose) after an extended lag phase [12], indicating strong growth inhibition. Interestingly, no growth was able to be initiated in spent media containing, among other osmolytes, 165 mM of organic acids produced during fermentation. This disparity suggests *C. bescii* is differentially resistant to different osmolytes. Ions capable of permeating, or being actively transported across, cell membranes are typically more cytotoxic than otherwise charge-neutral osmolytes, as they can disrupt and decrease membrane potential otherwise utilized to generate useful chemical potential energy for the cell [244].

A low osmolarity defined media was developed to alleviate the effects on growth present when *C. bescii* was grown in other growth medias [51] and this enabled the subsequent development of auxotrophy-based genetics in the species [20]. The osmolarity of this LOD (Low Osmolarity – Defined) media was half of that of previous media compositions, made possible by eliminating yeast extract, and decreasing the concentrations of macronutrients, among other changes. This resulted in better growth of *C. bescii* and a growth environment, whereby more osmolytes could be added before equivalent growth defects were observed [51].

Tolerance to osmolarity is a complex trait, as it involves the coordinated action of multiple cellular subsystems. Engineering more complex traits is difficult and cumbersome when modifying components of each system involved separately. Also, as effects of genetic modifications can be unpredictable, doing many successive modifications on a single strain increases the chances of encountering unintended effects [33] and detrimental unknown gene interactions. Success has been had engineering regulatory mechanisms of organisms, model and non-model alike, toward desired traits [254-256]. Modifying regulatory systems can potentially affect larger gene/metabolic space with fewer number of mutations than modifying enzymes. For example, modifying one global regulator can effect expression of many different operons charged with many different functions in a coordinated fashion [211]. We propose engineering regulatory machinery of *C. bescii* to effect osmotolerance by characterizing and engineering two regulatory gene targets we show to affect osmotolerance in *C. bescii*.

Results

Growth Phenotypes of Single-Gene Deletion Strains of Two Regulatory Proteins

Single-gene deletion strains of various regulatory genes were assessed for performance differences in CBP-related growth conditions (see methods). Regulatory gene targets were selected from genes belonging to regulatory gene families, and displayed differential expression in either or both of two previous differential gene expression studies [50, 257]. Also considered in selection of gene targets was the genes which were collocated with candidate regulatory gene targets, and the hypothesized regulatory actions of candidate gene targets. Single-gene deletion strains were generated for each of the 10 candidate regulatory genes (Table A5) and these strains, along with their genetic parent strains, were assayed in eight different CBP relevant growth conditions. Growth conditions included growth various soluble and insoluble carbon sources, including unpretreated, washed switchgrass, xylan, crystalline cellulose, glucose, or xylose. Other conditions represented stress conditions which can be

encountered during CBP; addition of methyl viologen to simulate a redox stress, added ethanol to simulate the stresses of fermentation product accumulation, and added NaCl to simulate the stress of elevated osmolarity.

Across all conditions screened for growth and fermentation phenotypes (Figures A14-A21), we found two single-gene deletion strains which both displayed prominent growth phenotypes when assayed under various elevated osmolarity conditions (Figures 17-20). When NaCl was added to growth media to make the total starting osmolarity 200 mOsm/L (calculated), strain JWCB005 ΔB5X54_RS01260 (Δ*cra*) displayed an estimated growth rate of 0.4, while its genetic parent displayed a similarly estimated growth rate of 0.12 (Figure 20). At a starting osmolarity of 400 mOsm/L (achieved by adding NaCl), strain JWCB005 ΔB5X54_RS01260 spent 15 hours in lag phase, while its genetic parent (JWCB005) spent 47 hours (Figure 18). Similar growth phenotypes, relative to the genetic parent strain, were observed in elevated osmolarity medias made with glycerol (Figure 17). Across increasing concentrations of added NaCl, strain JWCB005 ΔB5X54_RS06355 displayed more severe growth defects than its genetic parent strain (JWCB005) (Figure 18). Similar relative growth defects were observed when these strains were grown in LOD media containing 20 g/L added ethanol (Figure 19), when grown in 10 mL culture volumes in Balch tubes and 100% N_2 headspace. The relative growth differences exhibited by both strains in ethanol containing media could not be replicated in 50 mL cultures contained in 135 mL serum bottles and a headspace of 5% H_2, 10% CO_2, and 85% N_2.

Transcription Factor Binding Site Prediction

Using the Virtual Footprint tool associated with the PRODORIC database [258], we searched the *C. saccharolyticus* genome for individual FruR binding sites predicted in *Thermotoga* [259]. These

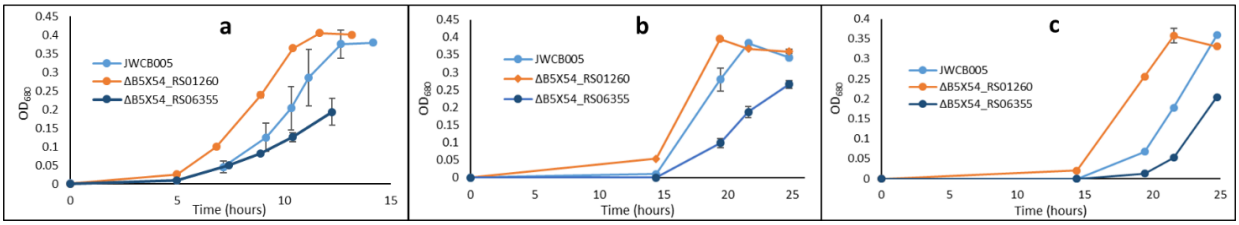


Figure 17. Growth phenotypes of strains ΔB5X54_RS01260 (Δ*cra*) and ΔB5X54_RS06355 (Δ*fapR*) when grown in increasing amounts of added glycerol. Glycerol was added to make total initial (calculated) media osmolarities of (a) 400 mOsm/L, (b) 500 mOsm/L, and (c) 600 mOsm/L. Batch growth was conducted in 50 mL culture in 135 mL serum bottles with 5% H_2/10% CO_2/85% N_2 as initial headspace. Strain JWCB005 is the genetic parent strain to both single gene deletion strains. Across the osmolyte concentrations assayed, strain ΔB5X54_RS01260 exhibits shorter lag phases while strain ΔB5X54_RS06355 exhibits a longer lag phase and lower apparent growth rates.

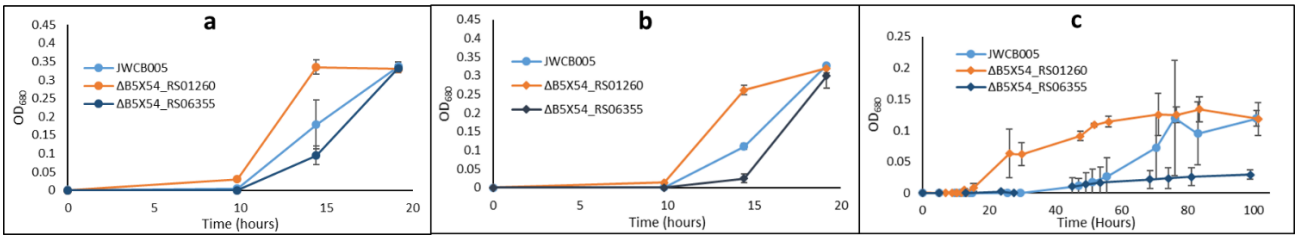


Figure 18. Growth phenotypes of strains ΔB5X54_RS01260 (Δ*cra*) and ΔB5X54_RS06355 (Δ*fapR*) when grown in increasing amounts of added NaCl. NaCl was added to make total initial (calculated) media osmolarities of (a) 200 mOsm/L, (b) 300 mOsm/L, and (c) 400 mOsm/L. Batch growth was conducted in 50 mL culture in 135 mL serum bottles with 5% H_2/10% CO_2/85% N_2 as initial headspace. Strain JWCB005 is the genetic parent strain to both single gene deletion strains. Across the osmolyte concentrations assayed, strain ΔB5X54_RS01260 exhibits shorter lag phases and higher apparent growth rates, while strain ΔB5X54_RS06355 exhibits longer lag phases and a lower apparent growth rate in 400 mOsm/L.

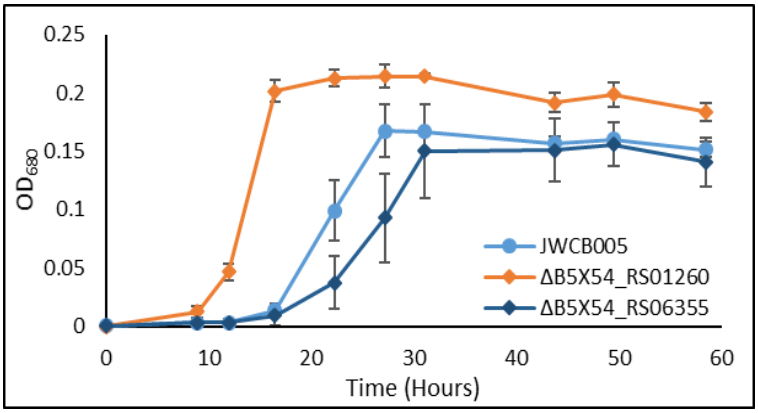


Figure 19. Growth phenotypes of ΔB5X54_RS01260 (Δ*cra*) and ΔB5X54_RS06355 (Δ*fapR*) when grown in liquid culture containing 20 g/L added ethanol. These growth experiments were conducted in 10 mL volumes in Balch tubes containing N_2 gas as initial headspace.

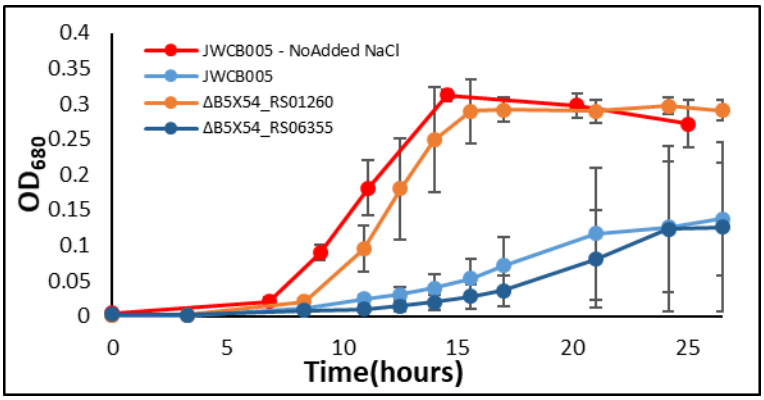


Figure 20. Growth phenotypes of strain ΔB5X54_RS01260 (Δ*cra*) and ΔB5X54_RS06355 (Δ*fapR*) when grown in liquid culture containing added NaCl to make calculated initial osmolarity of 200 mOsm/L. These growth experiments were conducted in 10 mL volumes in Balch tubes containing N_2 gas as initial headspace. Red line/markers indicate cultures grown containing no added NaCl.

searches did not find any putative binding sites. Manual inspection of the *Thermotoga* FruR binding sites predicted in RegPrecise [259] suggests that these predicted sites are about twice the length of a typical transcription factor binding site (~18-20 bp). These predicted sites may be two individual sites oriented adjacent to each other. Splitting each site in half yielded four 'half sites' with the regular expression consensus sequence (A|G)TCATAA(A|T)NNNNNAT(A|C)ANN. This consensus sequence is similar in length and conserved position sequence predicted for the FruR/Cra transcription factor in other organisms curated in the RegPrecise database [259] – particularly those genera in which consensus binding sites have been built using far more individual binding sites, such as *Bacillus* (17 predicted sites) and *Streptococcus* (34 predicted sites). The only putative site identified outside of coding regions when searching for this manually generated consensus sequence was a site located 46 bp immediately upstream of the the *fruR*/*cra* gene itself, presumably indicating the local negative regulation of itself and other collocated genes contained in the same putative transcriptional unit. Similarly, we were able to identify FapR binding sites in the *C. bescii* JWCB005 genome that are homologous to previously predicted FapR binding sites in *C. saccharolyticus* [50]; upstream of B5X54_RS06355 (a putative fatty acid biosynthesis transcription unit) and B5X54_RS07175 (a putative malonyl-CoA biosynthesis transcription unit).

Direct and Indirect Regulatory Actions of FapR and Cra in *C. bescii*

Toward a more complete assessment of the regulons of these two transcription factors, we conducted RNAseq analysis of these two strains, and their common genetic parent strain, under three different moderate elevated osmolarity conditions (added glycerol to total 500 mOsm/L, added glucose to total 300 mOsm/L, added glycerol to total 300 mOsm/L, see Figure A12). We assessed differentially expressed genes which showed common expression patterns in these three elevated osmolarity conditions, to avoid confounding osmolarity response with any that might be chemically specific to a particular osmolyte. Genes contained in putative operons encoding genes responsible for fatty acid

biosynthesis (B5X54_RS06360 - B5X54_RS06395) and malonyl-CoA biosynthesis (B5X54_RS07175 - B5X54_RS07190) displayed increased expression in strain JWCB005 ΔB5X54_RS06355 (Table 9).

In strain JWCB005 ΔB5X54_RS01260, increased expression of genes involved in fructose PTS as well as a phosphofructokinase (B5X54_RS01265 - B5X54_RS01290) (Table 10) was observed. Interestingly, we find these genes to not be significantly differentially expressed under elevated levels of glucose, suggesting these genes may be differentially regulated in response to glucose as compared to other osmolytes assayed. Additionally, two adjacent genes (a transposase, sigma factor E) were differentially expressed in strain JWCB005 ΔB5X54_RS01260 when grown in elevated levels of glucose and glycerol, but not NaCl, and a gene encoding sigma factor G (B5X54_RS04530) differentially expressed under all three elevated osmolarity conditions tested (Table 10).

We also find genes displaying decreased expression levels in strain JWCB005 ΔB5X54_RS01260. Three adjacent genes (B5X54_RS01990 - B5X54_RS02000) showed similar differential expression patterns. B5X54_RS01990 and B5X54_RS01995 are annotated as hypothetical proteins, and B5X54_RS02000 is annotated as a peptidase S8 protein. Similarly, a vaguely annotated oxidoreductase gene, B5X54_RS07480, showed increased expression in all three elevated osmolarity conditions.

Three genes exhibited decreased expression in both strains; B5X54_RS00980, B5X54_RS05305, and B5X54_RS05310 (Table 12). The latter two are annotated as hypothetical proteins, adjacent to each other and may likely be part of the same transcriptional unit. B5X54_RS00980 is annotated as an endo-1,4 beta xylanase. Expression differences in these two genes could be genomic artifacts acquired in strain JWCB005 before either of the single-gene deletion strains considered here were generated. As these genes showed similar expression patterns in two strains which displayed largely different growth phenotypes, it is further unlikely they are contributing to the observed phenotypes and were not considered further.

Table 9. Genes in strain JWCB005 ΔB5X54_RS06355 (Δ*fapR*) which exhibit increased expression when cells were cultured in elevated osmolarity conditions. Colors denote adjacent genes which may be co-expressed as part of the same transcriptional unit. Genes denoted in green are responsible for fatty acid biosynthesis. Genes denoted in red are responsible for the metabolism of malonyl-CoA, a fatty acid biosynthesis precursor and also a regulatory effector which binds the FapR transcription factor. All values are statistically significant (Wald test, Benjamini-Hochburg adjusted p-value < 0.05, n=3)

		$\log_2$(ΔB5X54_RS06355/JWCB005)		
Locus Tags	Annotation	Glycerol	Glucose	NaCl
B5X54_RS06360	phosphate acyltransferase	3.9	2.9	2.8
B5X54_RS06365	ketoacyl-ACP synthase III	3.8	2.3	2.2
B5X54_RS06370	malonyl CoA-acyl carrier protein transacylase	3.8	2.5	2.2
B5X54_RS06375	beta-ketoacyl-ACP reductase	3.7	2.2	2.1
B5X54_RS06380	acyl carrier protein	2.1	1.5	1.3
B5X54_RS06385	beta-ketoacyl-[acyl-carrier-protein] synthase II	2.8	1.8	1.7
B5X54_RS06390	ribonuclease III	2.2	1.3	1.3
B5X54_RS06395	radical SAM protein	2.3	1.3	1.5
B5X54_RS07175	methylmalonyl-CoA carboxyltransferase	3.6	1.6	1.9
B5X54_RS07180	hypothetical protein	3.6	1.8	2.2
B5X54_RS07185	acetyl-CoA carboxylase biotin carboxyl carrier protein subunit	3.6	1.8	2.1
B5X54_RS07190	oxaloacetate decarboxylase	3.7	1.7	2.0

Table 10. Genes in strain JWCB005 ΔB5X54_RS01260 (Δ*cra*) which exhibit increased expression when cells were cultured in elevated osmolarity conditions. Colors denote adjacent genes which may be co-expressed as part of the same transcriptional unit. Bold text denotes statistical significance (Wald test, Benjamini Hochburg adjusted p-value < 0.05, n=3), values that are not statistically significant are denoted in grey text.

		log_2(ΔB5X54_RS01260/JWCB005)		
Locus Tag	Annotation	Glycerol	Glucose	NaCl
B5X54_RS01265	1-phosphofructokinase	4.2	0.8	4.5
B5X54_RS01270	PTS fructose transporter subunit IIA	4.5	0.7	4.5
B5X54_RS01275	PTS fructose transporter subunit IIBC	4.8	0.6	4.3
B5X54_RS01280	HPr family phosphocarrier protein	4.1	0.8	3.7
B5X54_RS01285	phosphoenolpyruvate-protein phosphotransferase	4.5	0.6	4.0
B5X54_RS01290	Acetylesterase	2.4	0.4	2.7
B5X54_RS04520	transposase	1.4	1.8	0.6
B5X54_RS04525	sporulation sigma factor SigE	1.5	1.4	0.9
B5X54_RS04530	sporulation sigma factor SigG	1.3	1.2	1.0

Table 11. Genes in strain JWCB005 ΔB5X54_RS01260 (Δ*cra*) which exhibit decreased expression when cells were cultured in elevated osmolarity conditions. Colors denote adjacent genes which may be co-expressed as part of the same transcriptional unit. All values are statistically significant (Wald test, Benjamini-Hochburg adjusted p-value < 0.05, n=3)

		log_2(ΔB5X54_RS01260/JWCB005)		
Locus Tag	Annotation	Glycerol	Glucose	NaCl
B5X54_RS01990	hypothetical	-3.3	-2.2	-2.9
B5X54_RS01995	hypothetical	-3.5	-2.9	-3.2
B5X54_RS02000	peptidase S8	-3.8	-3.7	-3.5
B5X54_RS07480	oxidoreductase	-1.3	-2.2	-1.6

Table 12. Genes that were found to have similar differential expression in strain JWCB005 ΔB5X54_RS01260 (Δ*cra*) and JWCB005 ΔB5X54_RS06355 (Δ*fapR*), each compared to their common genetic parent strain JWCB005. Values are average differential log-fold changes observed for the indicated gene across three indicated high-osmolarity conditions (glucose, glycerol, sodium chloride). As these genes were similarly differentially expressed in strains showing different growth phenotypes, it is unlikely these genes are contributing to phenotype differences, and were not considered in this study. All values are statistically significant (Wald test, Benjamini-Hochburg adjusted p-value < 0.05, n=3)

Locus Tag	Annotation	Average $\log_2$(ΔB5X54_RS01260/JWCB005) across three osmolarity conditions	Average $\log_2$(ΔB5X54_RS06355/JWCB005) across three osmolarity conditions
B5X54_RS00980	endo-1,4-beta-xylanase	-4.1	-4.2
B5X54_RS05305	Hypothetical	-4.9	-2.5
B5X54_RS05310	Hypothetical	-4.3	-2.4

Additional genes showed significantly decreased expression in strain JWCB005 ΔB5X54_RS01260. These genes were distinct from other genes displaying differential expression in that their normalized read count values were near zero (Table 13). It is possible that this differential expression is artefactual carryover from genomic alterations which have occurred in the strain. One such example was considered further; B5X54_RS07480 - B5X54_RS07510. This locus was of interest as it contains many genes involved in uracil biosynthesis and sits adjacent to the *pyrF* gene (B5X54_RS07475) which is truncated and non-functional in this strain [54] – the truncation making uracil-auxotrophy based genetic modifications possible. PCR amplification of a region spanning the genomic region between genes B5X54_RS07485 and B5X54_RS07515 resulted in a single product of expected size when using genomic DNA from strain JWCB018 (an auxotrophic derivative of strain JWCB005) as template, though no PCR product could be amplified when using genomic DNA from JWCB005 ΔB5X54_RS01260 as PCR template (Figure A2). This suggests that differential expression results obtained from these genes in JWCB005 ΔB5X54_RS01260 are artifacts resulting from genomic differences between this strain and its genetic parent. Furthermore, repeated efforts to transform strain JWCB005 ΔB5X54_RS01260 using *pyrF* as a positive heterotrophic selection marker did not yield any transformants, suggesting additional genes (beyond just *pyrF*) necessary for uracil biosynthesis are missing or have been functionally disrupted in this strain. Differential expression from these genes, and other genes showing normalized read count values <5 were also omitted from analysis (Table 13), as these may too be genomic differences and not genuine expression differences.

In an effort to generate a strain more osmotolerant than JWCB005 ΔB5X54_RS01260, we will overexpress the *fapR* gene in strain JWCB005 ΔB5X54_RS01260. As a single-gene deletion of the *fapR* gene generated a strain, JWCB005 ΔB5X54_RS06355, sensitive to elevated concentrations of certain osmolytes, we hypothesize that overexpressing the *fapR* gene may increase osmotolerance and further improve growth characteristics in the presence of elevated osmolarities. We have generated the strains

Table 13. Genes found to have relatively low normalized expression values in strain JWCB005 ΔB5X54_RS01260 (Δ*cra*). These very low expression values may be artefactual, and instead the result of the loss of these genes in this strain. All differential expression log2 values are statistically significant (Wald test, Benjamini-Hochburg adjusted p-value < 0.05, n=3)

		log_2(ΔB5X54_RS01260/JWCB005)			Average Normalized Read Counts		
Locus Tag	Annotation	Glycerol	Glucose	NaCl	ΔB5X54_RS01260	JWCB005	ΔB5X54_RS06355
B5X54_RS04240	DNA-binding protein	-6.6	-6.4	-6.8	0.58	366	346
B5X54_RS04275	transcriptional regulator	-6.5	-7.5	-7.8	0.57	659	773
B5X54_RS04280	DNA primase	-6.9	-7.5	-8.3	0.57	824	968
B5X54_RS04285	hypothetical protein	-7.1	-7.5	-7.7	0.57	856	828
B5X54_RS04290	site-specific integrase	-6.6	-5.6	-6.2	0.12	300	277
B5X54_RS04295	tRNA-Pro	-5.6	-5.2	-5.7	1.6	212	220
B5X54_RS07485	dihydroorotate dehydrogenase	-6.6	-6.9	-6.8	0.59	358	383
B5X54_RS07490	histidine phosphatase family protein	-6.5	-6.6	-7.0	0.62	350	401
B5X54_RS07495	orotate phosphoribosyltransferase	-6.8	-6.9	-6.9	1	467	523
B5X54_RS07500	hypothetical protein	-6.6	-6.4	-6.6	0.39	290	304

necessary for this experiment and will conduct the comparative growth studies in the near future. In addition, we have expressed the B5X54_RS01260 gene in strain JWCB005 ΔB5X54_RS01260, generating a genetic complement in which we should see a phenotype reversion upon reintroducing the B5X54_RS00710 gene back into *C. bescii* if the osmolarity resistance phenotype is, in fact, due to the absence of the *fruR*/*cra* gene. Another strain hosting an expression vector without a corresponding expression cassette (an 'empty vector'), will serve as a control to both of the aforementioned strains.

Discussion

We found substantial growth phenotypes exhibited by two single-gene deletion strains of *C. bescii*. A single gene deletion of the *fapR*/*cra* gene showed considerable growth defect when grown in elevated levels of osmolarity (Figures 17-20) and a single-gene deletion of the *fruR* or *cra* gene grew substantially better than its genetic parent strain (on the basis of growth rate and time in lag phase) in elevated levels of added osmolarity (Figures 17-20). Neither strain displayed growth phenotypes when grown in unaltered, replete LOD media (Figure A15).

Osmolarity Stress Response and the Cellular Function of The FapR and FruR/Cra Transcription Factors in *C. bescii*

As similar relative growth phenotypes were observed in the presence of different osmolytes, we hypothesize these genes to be part of cellular processes that are a common response to osmolytes, and not to individual chemical stressors (e.g. ethanol stress, NaCl stress). The *fapR* gene, and associated regulon genes, was shown to be upregulated under hydrogen sparging in *C. saccharolyticus* [50]. Not surprisingly, the transcription levels of *fruR*/*cra* in *C. saccharolyticus* were increased when cells were grown on fructose [257], suggesting that this gene is in fact locally regulating genes contained within the same operon.

FapR is a conserved transcription factor present in gram positive bacteria. Its regulon consists of the genes responsible for fatty acid biosynthesis as well as malonyl-CoA metabolism [260]. Malonyl-CoA also serves as the binding cofactor molecule for FapR. It has been shown previously in other organisms that altering this transcription factor can effect cell growth [261], and the composition of cellular membrane lipids [262], increasing the relative amount of long chain fatty acids in a strain of *B. subtilis* deficient in its *fapR* gene. Altering cellular membrane composition in bacteria effects a cells ability to withstand osmotic stress [263], suggesting deletion of this gene in *C. bescii* is resulting in similar membrane composition changes and affecting growth rate and survival.

The *fruR* gene was first characterized and shown to be a local regulator of fructose metabolism and a fructose-specific PTS, and was subsequently shown to have a regulon extending far beyond just genes it is collocated with [264-266]. The gene was aptly renamed Cra (for cyclic AMP independent repressor/activator). FruR was shown to both activate genes as well as repress genes [264], as well as work coordinately with other carbon regulatory genes to globally regulate carbon metabolism [267], for which we also see putative evidence of in *C. bescii*, as *fru* operon genes appear to be coordinately expressed differently in high concentrations of glucose (Table 10). This transcription factor regulates expression in response to binding fructose 1-phosphate [264, 268], a metabolic intermediate in fructose metabolism. It is known to be a global regulator of carbon metabolism, regulating genes involved in fermentation, glycolysis/gluconeogenesis, carbohydrate transport, and the TCA cycle [264, 267, 269, 270]. Though conserved homologs appear in many organisms [271-274], it is unclear whether it plays a global regulatory role in *C. bescii*, or if it is a local repressor of the fructose metabolism genes it is collocated with.

A FruR homolog with 70-80% operon-level sequence homology in Thermotoga [272] is predicted to be a local repressor of fructose metabolism genes, though this gene or its regulon have not been

biochemically or genetically evaluated. In addition to this local regulation, FruR/Cra may negatively regulate two stress-related sigma factors; Sigma E and Sigma G (Table 10). The regulons genes whose transcription is affected by these transcription factors in *C. bescii* remains unknown, though homologs of these two sigma factors are known to be active in stress-response [275, 276]. The RefSeq gene annotations of these two sigma factors annotate these genes as being involved in sporulation, though *C. bescii* is not known to form spores [14]. As both increased and decreased differential expression are observed in strain JWCB005 ΔB5X54_RS01260 (Table 10 and 11), this gene may have a pleiotropic regulatory role in *C. bsecii*. It is not possible at this time to determine if such pleiotropy would be the effect of direct or indirect regulatory action by this transcription factor.

Proposed Mechanism of FapR and Cra-Enabled Osmotolerance in *C. bescii*

We confirm that the FapR transcription factor does regulate fatty acid biosynthesis and malonyl-CoA metabolism genes in *C. bescii*, as predicted previously for the closely related species *C. saccharolyticus* [50]. As has been shown in other organisms, it is likely that the cell membrane lipid composition of *C. bescii* is altered in response to altering the *fapR* regulatory gene in *C. bescii*, directly affecting biophysical properties of the membrane. Such membrane composition changes may also be affecting the activity of osmo-responsive membrane proteins in *C. bescii*.

The role of FruR/Cra to osmoresistance in *C. bescii* remains more cryptic than that of FapR. Evidence collected in this work, and previously in *Caldicellulosiruptor*, offer little evidence as to what the osmolarity-specific mechanism that could be conferring the observed growth differences in strains containing a modified *fruR*/*cra* gene. From transcription factor binding site identification evidence, and differential expression evidence (Table 10 and 11), we postulate that FruR/Cra negatively regulates elements of fructose metabolism as well as fructose-specific PTS in *C. bescii*. Two other putative transcription units may be activated by FruR/Cra in *C. bescii* under elevated osmolarity conditions; one

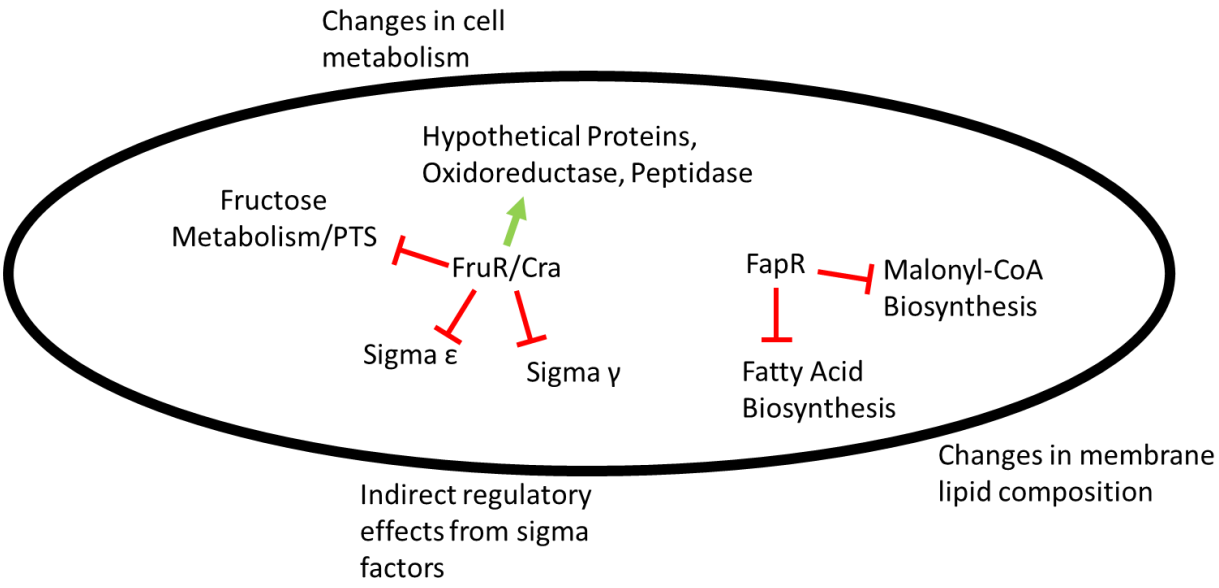


Figure 21. Proposed roles of the FruR/Cra and FapR transcription factors in *C. bescii*, and the hypothesized regulatory response of these transcription factors to increased osmolarity conditions.

containing a single gene generally annotated as an oxidoreductase. The other contains three genes; two annotated as hypothetical and one annotated as a peptidase.

Improving Osmoresistance in Consolidated Bioprocessing Through Regulatory Metabolic Engineering

Osmosensitivity is one of the defining needs in a candidate industrial consolidated bioprocessing organism. To make 50 g/L ethanol, a nominal target for consolidated bioprocessing, the final fermentation osmolarity will exceed that of seawater, which is ~1000 mOsm/L and a CBP biocatalyst will need to actively ferment at these osmolarities. The cellular response to osmotic shock involves a large number of cellular sub-processes [243], many of which are poorly understood in *C. bescii*. Engineering an osmoresistant strain using only rational metabolic engineering would, at present, be impossible without further characterizing osmolarity response systems in *C. bescii*. We instead pursued a strategy of identifying the metabolic systems important to survival in elevated osmolarities through screening transcription factors. Doing so allowed us to effect maximal gains in osmoresistance, by making very few genetic modifications. We show herein a strain with one transcription factor gene (*fruR*/*cra*) deleted had a growth rate over three times greater than its genetic parent (Figure 20). Furthermore, we identified a second transcription factor gene (*fapR*) which shows a substantial growth defect when grown in the presence of elevated osmolarities, suggesting yet another gene target through which further osmoresistance gains may be achieved. We aim to assess whether a strain carrying putatively beneficial alterations in both gene targets will, in fact, generate a highly osmotolerant strain.

Conclusions

Through screening of single-gene deletion mutants of selected regulatory genes, we identified two transcription factors important to osmoresistance in *C. bescii*, the FapR repressor, as well as a regulator which is most likely either the FruR local repressor of fructose metabolism or the global carbon metabolism regulator Cra. Differential expression and prediction of transcription factor binding sites

along the genome show the FapR gene regulates fatty acid biosynthesis genes as well as genes responsible malonyl-CoA metabolism, each organized into a functionally related putative transcriptional unit. The FruR transcription factor negatively regulates a functionally related transcription unit containing genes responsible for fructose metabolism and fructose-specific PTS. In addition, differential expression evidence also suggests it has a regulatory effect on other poorly-annotated genes. While the role of FapR in osmolarity and stress-related phenotypes is previously reported and conserved, the role of the *fruR*/*cra* gene in osmotolerance is unique and the mechanism of its influence on this phenotypic behavior remains cryptic. As these two gene targets both influence osmoresistance in *C. bescii*, we will engineer a strain which both loci are altered, with the hope of generating a highly osmotolerant strain of *C. bescii*. Doing so will enable *C. bescii* to grow more robustly in elevated osmolarity conditions in which its native growth and carbon conversion suffers. This approach has proven to effectively yielded the desired increase in resistance to osmolarity, with little *a priori* systems-level information about the mechanism responsible for this trait in *C. bescii*.

Methods

Screening Single Gene Deletion Mutants

Markerless deletions of 10 regulatory genes were generated in one of two strains of *C. bescii* (Table A14) using previously described methods [20, 53]. Single gene deletion strains were grown in eight different conditions and screened for growth or fermentation productivity phenotypes (Figures A14 – A21). Media was prepared as described previously [51]. Initial screening cultures utilizing soluble carbon sources (xylose, glucose, added ethanol, added NaCl, added methyl viologen) were all grown in Balch tubes containing 10 mL of media. Cultures containing insoluble carbon substrates (xylan, crystalline cellulose, and switchgrass) were grown in 50 mL of media in a 135 mL serum bottle. All medias contained 5 g/L of the respective carbon substrate. Cultures containing added methyl viologen, ethanol,

or NaCl contained 5 g/L glucose as the primary source of carbon. Methyl viologen was utilized at a concentration of 150 mg/L, ethanol was added to a final concentration 20 g/L, and NaCl was utilized to make media with an initial calculated total osmolarity of 200 mOsm/L. All media osmolarities were calculated based on accounting for media osmolarity of LOD media reported previously [51], and adding osmolytes as dry chemical sufficient to make each media the desired osmolarity. All screening cultures contained a headspace of 100% N_2. Growth and fermentation products were assayed as previously described [52]. Growth was monitored in cultures containing xylan or crystalline cellulose as the primary source of carbon by assaying total culture protein. 300 µL of culture was added to 100 µL of 500 mM NaOH and incubated at 100°C for 10 minutes. Incubated samples were then centrifuged (13,000 x g, 10 minutes). Aliquots of the supernatant were analyzed for total protein using the Coomassie (Bradford) Protein Assay Kit (ThermoFisher Scientific catalog #23200) according to manufacturer's instructions.

Subsequent osmolarity stressor assays (Figures 17 and 18), as well as batch growth for RNAseq analysis (Figure A12), were done in 50 mL culture volumes, in elevated osmolarity conditions indicated, in 135 mL serum bottles containing a headspace of ~5% H_2,10% CO_2, and 85% N_2.

RNAseq Analysis of Single Gene Deletion Mutants Under Elevated Osmolarity Conditions

RNAseq analysis was conducted on strains grown to mid-log phase in three different elevated osmolarity conditions; added NaCl to a total initial osmolarity of 300 mOsm/L, added glycerol to make a total initial osmolarity of 500 mOsm/L, and added glucose to a total initial osmolarity of 300 mOsm/L. All cultures were grown to mid-log phase (Figure A12) at which point sacrificial 50 mL cultures were collected, centrifuged, and cell biomass pellets were snap frozen in liquid nitrogen. Growth was monitored in parallel-grown triplicate cultures (Figure A12). RNA was extracted, and RNAseq libraries were synthesized and sequenced as described previously [277].

Fastq files were downloaded from HudsonAlpha's servers and verified for integrity by computing their checksum. Read quality was checked using FastQC (v. 0.11.5) [278]. RNAseq reads were trimmed the using Trimmomatic (v. 0.33) MAXINFO method with target length and strictness parameters set to 40 and 0.8, respectively [279]. Reads were mapped to the genome of *Caldicellulosiruptor bescii* JWCB005 (GCF_900166995.1, last modified 2017/03/09) using Bowtie2 (v. 2.2.9) with the same parameters as the very-sensitive preset option expect the number of mismatches was set to 1 [280], and reads were counted with HTSeq (v. 0.6.1p1) [281]. Differential expression was computed using the DEseq2 R package [278] using default parameters. Differentially expressed genes were determined to be those exhibiting a statistically-significant $\log_2$ fold change of >1 or <-1, normalized average read counts above 50, and individual strain normalized read counts > 5 (Tables 9 – 13). Raw, unprocessed RNAseq reads as well as differential expression data for all genes were deposited into the Gene Expression Omnibus database as accession number GSE107393 and can be queried at the following URL: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi (also see attachment 3).

Transcription Factor Binding Site Prediction

Transcription factor binding sites were predicted for the FruR/Cra transcription factor using the *C. bescii* JWCB005 genome (NCBI Accession # PRJEB19583). A regular expression consensus binding site was adapted from *Thermotoga* FruR predicted binding sites in RegPrecise [259]; the sequence (A|G)TCATAA(A|T)NNNNNAT(A|C)ANN. This sequence was used to seed a transcription factor binding site sequence search using the Virtual Footprint Regulon Analysis tool [282]. The above-mentioned sequence was entered as 'Regular Expression' subpattern type and the maximum distance to gene (maximum promoter length) was set to 350 bp. The match properties 'Ignore Match Orientation' and 'Remove Redundant (Palindromic) Matches' were selected.

Assessment of Genome Differences Between JWCB005/JWCB018 and JWCB005 ΔB5X54_RS01260

Primers were designed to amplify a genomic region spanning B5X54_RS07480 - B5X54_RS07510. A primer was designed to anneal to the B5X54_RS07480 coding region (pyr fwd screen, Table A1), and 3' to the B5X54_RS07510 coding DNA sequence (pyr rev screen, Table A12). Genomic DNA was extracted from the respective strains using a Zymo Research Quick-DNA Miniprep Kit (Catalog # D3006, Zymo Research Corp., Irvine, CA) according to manufacturer's instructions, and used as template in PCR amplification carried out using Phusion 2X Master Mix according to manufacturer's instructions (Catalog # M0531S, New England Biolabs, Ipswich, MA).

C. bescii Cra/FruR Protein Purification

We attempted to express and purify recombinant *C. bescii* FruR/Cra protein using the IMPACT Protein Purification System (New England Biolabs, Ipswich, MA), with the B5X54_RS01260 coding DNA sequence cloned into the 5' region upstream of the Mxe Gyrase Cleavage Intein and Chitin Binding Domain present on plasmid pTXB1. It was subsequently found that the native protein sequence encoded by B5X54_RS01260 becomes insoluble upon DTT-induced cleavage from the fusion affinity binding peptides. Special considerations were necessary to enhance solubility of the *E. coli* Cra protein upon recombinant expression and purification as well [264]. Problems encountered with protein insolubility during purification of *E. coli* Cra protein were partially solved by overexpressing the GroES and GroEL chaperone proteins simultaneously with the Cra protein [264]. Subsequent efforts to increase recombinant *C.bescii FruR/Cra* solubility were attempted (increasing additive concentrations, modifying regimes to elute protein from affinity resin, eluting protein in partially denaturing conditions), though none were successful.

Chapter 6: Summary and Future Directions

Herein we examine the effect that metabolic engineering using regulatory gene targets can have on performance of candidate consolidated bioprocessing organisms. We strive to understand if the engineering of gene targets, implying the engineering of regulatory mechanisms within the cell, would stand to be a promising and effective method to further effect growth and fermentation in organisms. Previous rational and classical engineering strategies (carbon forcing, electron forcing) had all but been exhausted [31, 33, 39, 40] in these organisms, and performance metrics (such as ethanol product yield, ethanol tolerance, and lignocellulose substrate conversion) have not been met fully at this time.

We are not investigating whether such effects are possible to attain, as regulatory engineering was demonstrated previously in various organisms [78, 284, 285] and shown to have dramatic effects on desired traits in these organisms. Instead, this thesis sought to engineer for phenotypes unique to these organisms; the unique processes we wish to enable them to perform, and unique environments we will challenge them with. We chose to undertake this thesis question knowing that only a very scant amount of information existed about the systems we would engineer, another stark difference from previous regulatory engineering efforts.

Redox metabolism had been identified as the primary bottleneck in *C. thermocellum* fermentation [21]. Similarly, a large relative shift in carbon flux was brought about in *C. bescii* through the unoptimized introduction of an NADH-dependent alcohol dehydrogenase gene [16], and a large shift in electron metabolism through the elimination of NADH-dependent lactate dehydrogenase [40]. These findings led us to postulate that fermentation performance gains would be realized through engineering the redox systems of these organisms. Coincidentally, both of these organisms contain actively-expressed *rex* genes, suggesting that this global redox-state regulator [46] was maintaining redox state within these cells. The study and engineering of the *rex* gene an ideal way to both comprehensively

understand redox dynamics in these organisms, and engineer these systems to ultimately improve the redox-dependent synthesis of ethanol.

Chapters 2, 3, and 4 in this thesis address redox dynamics and the *rex* gene in these organisms. We identify previously unknown active redox-mediating systems at play in these organisms which are acting as unwanted, yield-decreasing, electron and carbon sinks.

These organisms will ferment industrially in a media that contains a primary carbon substrate that is largely insoluble. The insights and engineering efforts that stem from them will be relevant to those identified in these experiments, which are almost entirely carried out in soluble carbon substrate media, as both are ultimately potent reductive stresses. One likely difference will be the dynamic profile of redox stress experienced by these organisms in these two different scenarios; soluble vs. insoluble carbon substrates. Early-stage fermentation conditions are characterized by much less osmotic pressure and less carbon forcing on fermentative glycolysis in insoluble substrate fermentations, though later stages of these two fermentation scenarios will more closely resemble each other as soluble sugars tend to accumulate as fermentation slows. The different dynamic profiles of redox stress will need to be considered when engineering efforts aimed more specifically at insoluble substrate utilization.

Identification of the active redox systems has informed ongoing metabolic engineering efforts to streamline and optimize redox-dependent flux in these organisms, including the generation of a glutamine synthetase deletion mutant which exhibited 53% higher ethanol yield and 43% less amino acid overflow metabolism products [36]. Further, we identify a robust and validated method for quantifying redox-charge carriers, the currency of intracellular redox, within a cell. We confirm that engineering the *rex* gene does have the intended effect on cellular redox state and ethanol synthesis, and there are not additional significant insulating levels of regulation (transcription or metabolic) working to maintain the redox-state in these organisms. Another important finding of chapter three are the key regulon members, both previously predicted and those identified here for the first time, which

constitute the network nodes with which the *rex* regulator interacts. Every biological cell must maintain its redox state, though there is much variety and disparity in the way different cells and tissues do so, and *C. bescii* is no exception to this general observation. Its redox state regulation contains elements mirrored in other species, as well as many elements not identified elsewhere.

It appears that the Rex regulator is active during particular metabolic states in these organisms during fermentation, and not during other metabolic states. This is evidenced by the ethanol fermentation phenotype in a *rex*-deficient *C. bescii* only occurring long after active growth had ceased, and the absence of *rex*-regulator (which is negatively autoregulated) differential expression under oxidizing conditions in *C. thermocellum*. We currently do not understand these metabolic states, and, in particular, the differences between them. The fermentation shift observed in *C. bescii* is reminiscent of the solventogenic shift that occurs in mesophilic species of *Clostridia [141]*, and similarly occurs inverse to active growth. An important future direction, as it relates directly to ethanol productivity, would be a regulatory and mechanistic exploration of this phenomenon. As this appears outwardly similar to *Clostridial* solventogenic shifting, the path and template followed for metabolically understanding this shift should be pursued in *Caldicellulosiruptor*. Such understanding will afford metabolic engineering of ethanol synthesis in *C. bescii* that is more sophisticated and more amenable with native metabolism.

The fifth chapter explores the same regulatory engineering approach, though this time by addressing a multitude of candidate targets and bioperformance traits in combinations factorially. In contrast to the knowledge-driven approach in engineering redox systems, this approach was primarily data-driven. It became a data driven approach because of the lack of *a priori* knowledge, and not expressly because of an abundance of pre-existing data, which we also did not possess at the time. The data we did have (in the form of two differential expression profiles), and distantly inferable hypotheses from past work in *Caldicellulosiruptor*, led us to a set of candidate gene targets. Screening of single-gene deletion mutants led us fortuitously to two gene targets effecting osmolarity tolerance, one of the most potent growth

inhibitions identified in *C. bescii*. Furthermore, no mechanistic understanding existed about resistance to osmolarity in *Caldicellulosiruptor*, and thus there was little hope of improving this performance metric by way of rational and classical metabolic engineering. We have quantified the benefits of altering these two loci, individually, by way of growth differences in a selection of different osmolytes representing different chemical and metabolic characteristics (neutral and ionic chemicals, metabolically inert, interacting, and catabolized osmolytes). Further, identifying regulon members of these regulatory genes, as was done previously with *rex*, allowed us a first glance into osmoresistance mechanisms in *C. bescii* and to hypothesize about the dynamic systems which are responsible for survival at elevated osmolarities in *C. bescii*. Through a rudimentary description of the effects these genes can have on osmotolerance, we have made effective headway into a very difficult metabolic engineering problem and growth limitation using very little data and very few genetic modifications.

In this work we have deciphered instances where engineering using regulatory genes as targets can be effective. Often in non-model organisms, predicted or experimentally verified information about regulatory networks is scant or altogether missing. Even for relatively well characterized regulators like Rex, regulon information is predicted for a portion of the regulon, such as in the case of *C. bescii*, or altogether absent, as is the case for *C. thermocellum*. Surprisingly, it is in these cases where regulatory engineering can be effective. Though it may not be known exactly which systems are being affected or in what way, effects of altering regulatory systems (or lack thereof) are immediately apparent and can thus be quickly screened, given appropriate methods. Once a regulatory target is identified as important, it can then be further characterized (to whatever extent is needed or possible) and related engineering strategies can be improved and refined.

Strains with engineered regulatory machinery rarely display unwanted or severely detrimental phenotypes, as is common when engineering catalytic enzymes or small portions of individual pathways. The disruption of pathways through metabolic enzyme engineering can cause catastrophic effects to cell

metabolism which the existing metabolism may not be able to effectively deal with, such as large unpredicted and unwanted shifts in metabolic flux, severe growth defects, and suboptimal performance in a desired trait. It is likely that these effects are rarely encountered in regulatory engineering as metabolic systems are left intact. Further, transcription units are often redundantly or multiply regulated by more than one regulatory element, meaning the elimination of any one regulator will likely not completely abolish regulation of expression of genes within its regulon. Thus, regulatory engineering can be an effective way to affect phenotypes while not introducing other severely detrimental, unwanted, and unpredicted effects. Annotation information about some regulatory genes is missing, vague, or inaccurate in non-model organisms. Many of the genes screened for phenotypes in this work were vaguely or mis-annotated, such as the *cra* gene itself. Also, hypothetical, unannotated, or vaguely annotated genes were found to be active in the *C. thermocellum* response to redox stress and regulated components of the Rex and Cra regulons in *C. bescii*. None the less, other information was relied upon to both select genes for screening and subsequently select them for further engineering. Most other forms of metabolic engineering is completely reliant on annotation information, which is largely derived from homology, and pathway information, which is both incomplete and also derived from homology. When working with non-model organisms, homology derived methods are often of limited use due to the large evolutionary distances between model organisms from which homology seed information is derived and the organisms of interest. Our empirical, data-driven methods used in this work can serve as effective indicators of bioperformance. While we were able to identify, and even engineer regulatory targets, poorly annotated genes identified in regulons amount to dead ends. Without reliable gene annotations, we cannot further proceed generating testable hypotheses which involve these genes, nor can we compile a comprehensive understanding of how these systems work. Both of these limitations severely hinder our ability to engineer or understand these systems. Homology based methods, while useful,

have inherent biases and provide no insight into the function of these genes. Gene function inference methods which are independent of homology are desperately needed to close this information gap. Many regulatory genes act globally across the genome, regulating many individual metabolic systems simultaneously. Complex traits, by definition, are those which involve the simultaneous action of many metabolic systems. As such, impacting bioperformance associated with complex traits is an ideal application of regulatory engineering. We show in this work the effectiveness of this approach in identifying and utilizing two genes to effect osmolarity tolerance, a complex trait. As metabolic engineering and synthetic biology move toward ever-more difficult tasks, such as bioremediation in unforgiving environments or curing complex metabolic disorders, the prevalence of complex traits and the need to effectively understand and impact them will be necessary. The approach outlined in this work, of deciphering and altering regulatory networks associated with these traits, will help to both understand and engineer complex traits.

One future direction worth pursuing is the engineering of a strain whereby these identified regulatory gene targets are both augmented in a single strain toward affording more robust growth in elevated osmolarities. This experiment, along with a genetic complementation analysis of the *fruR*/*cra* deficient strain, would further elucidate the importance of the *cra* gene to osmotolerance and the potential of combining beneficial mutations in both genes to generate a highly osmotolerant strain of *C. bescii*.

Another important future direction will be a more concise and comprehensive understanding of the function of the FruR/Cra transcription factor in *C. bescii*. The single-gene deletion strain exhibits a pronounced phenotype, and this gene stands as one of the most promising growth-enhancing genetic determinants in *C. bescii*. Understanding the mechanisms by which this is occurring is essential to tuning and optimizing the use of this gene as a metabolic engineering target.

Through this work we have found regulatory engineering to be an effective approach when the trait being engineered involves many cellular sub-systems which act simultaneously to produce a phenotype

and interact with each other in a complex way. Such traits are difficult to effect through other rational engineering methods, as the dynamics of these highly interconnected systems is typically lacking. This factorial approach to regulatory engineering to be useful when there is little pre-existing knowledge and/or data to formulate reliable hypotheses from.

References

1. Rydzak, T., et al., *Proteomic analysis of Clostridium thermocellum Core Metabolism: Relative protein expression profiles and growth phase-dependent changes in protein expression.* BMC Microbiology, 2012. **12**(1): p. 214.
2. EIA, *Biofuels: Ethanol and Biodiesel Explained.* 2017.
3. 2, E.-. *Weekly U.S. Oxygenate Plant Production of Fuel Ethanol.* 2017.
4. Lynd, L.R. and C.H. de Brito Cruz, *Make Way for Ethanol.* Science, 2010. **330**(6008): p. 1176-1176.
5. Agency, U.S.E.P., *Proposed Renewable Fuel Standards for 2017, and the Biomass-Based Diesel Volume for 2018.* 2016.
6. Fulton, L.M., et al., *The need for biofuels as part of a low carbon energy future.* Biofuel Bioprod Biorefin, 2015. **9**(5): p. 476-483.
7. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update.* Curr Opin Biotechnol., 2005. **16**.
8. Lynd, L.R., et al., *Bioenergy and African transformation.* Biotechnology for Biofuels, 2015. **8**(1): p. 1-18.
9. Davison, B.H., et al., *The impact of biotechnological advances on the future of US bioenergy.* Biofuels, Bioproducts and Biorefining, 2015. **9**(5): p. 454-467.
10. Miao, Z., et al., *Lignocellulosic biomass feedstock transportation alternatives, logistics, equipment configurations, and modeling.* Biofuels, Bioproducts and Biorefining, 2012. **6**(3): p. 351-362.
11. Oke, M.A., M.S.M. Annuar, and K. Simarani, *Mixed Feedstock Approach to Lignocellulosic Ethanol Production—Prospects and Limitations.* BioEnergy Research, 2016. **9**(4): p. 1189-1203.
12. Basen, M., et al., *Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium Caldicellulosiruptor bescii.* Bioresource Technology, 2014. **152**: p. 384-392.
13. Kataeva, I., et al., *Carbohydrate and lignin are simultaneously solubilized from unpretreated switchgrass by microbial action at high temperature.* Energy & Environmental Science, 2013. **6**(7): p. 2186-2195.
14. Hamilton-Brehm, S.D., et al., *Caldicellulosiruptor obsidiansis sp. nov., an Anaerobic, Extremely Thermophilic, Cellulolytic Bacterium Isolated from Obsidian Pool, Yellowstone National Park.* Appl Environ Microbiol, 2010. **76**(4): p. 1014-1020.
15. Blumer-Schuette, S.E., et al., *Thermophilic lignocellulose deconstruction.* FEMS Microbiol Rev, 2013. **38**.
16. Chung, D., et al., *Direct conversion of plant biomass to ethanol by engineered Caldicellulosiruptor bescii.* Proc Natl Acad Sci U S A, 2014. **111**(24): p. 8931-8936.
17. Jin, M., et al., *Consolidated bioprocessing (CBP) of AFEX™-pretreated corn stover for ethanol production using Clostridium phytofermentans at a high solids loading.* Biotechnology and Bioengineering, 2012. **109**(8): p. 1929-1936.
18. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update.* Curr Opin Biotechnol, 2005. **16**(5): p. 577-583.
19. Olson, D.G. and L.R. Lynd, *Transformation of Clostridium thermocellum by electroporation.* Methods Enzymol, 2012. **510**: p. 317-30.
20. Chung, D., J. Farkas, and J. Westpheling, *Overcoming restriction as a barrier to DNA transformation in Caldicellulosiruptor species results in efficient marker replacement.* Biotechnol Biofuel, 2013. **6**(1): p. 82.
21. Biswas, R., et al., *Increase in Ethanol Yield via Elimination of Lactate Production in an Ethanol-Tolerant Mutant of Clostridium thermocellum.* PLoS One, 2014. **9**(2): p. e86389.

22. Papanek, B., et al., *Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in Clostridium thermocellum.* Metabolic Engineering, 2015. **32**: p. 49-54.
23. Brown, S.D., et al., *Mutant alcohol dehydrogenase leads to improved ethanol tolerance in Clostridium thermocellum.* Proc Natl Acad Sci U S A, 2011. **108**(33): p. 13752-7.
24. Alahuhta, M., et al., *The structure and mode of action of Caldicellulosiruptor bescii family 3 pectate lyase in biomass deconstruction.* Acta Crystallographica Section D, 2013. **69**(4): p. 534-539.
25. Chung, D., et al., *Expression of the Acidothermus cellulolyticus E1 endoglucanase in Caldicellulosiruptor bescii enhances its ability to deconstruct crystalline cellulose.* Biotechnology for Biofuels, 2015. **8**(1): p. 113.
26. Brunecky, R., et al., *Revealing Nature's Cellulase Diversity: The Digestion Mechanism of Caldicellulosiruptor bescii CelA.* Science, 2013. **342**(6165): p. 1513-1516.
27. Wietzke, M. and H. Bahl, *The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in Clostridium acetobutylicum.* Appl Microbiol Biotechnol, 2012. **96**(3): p. 749 - 761.
28. Zhang, F., et al., *Enhancing fatty acid production by the expression of the regulatory transcription factor FadR.* Metabolic Engineering, 2012. **14**(6): p. 653-660.
29. Alper, H. and G. Stephanopoulos, *Global transcription machinery engineering: A new approach for improving cellular phenotype.* Metabolic Engineering, 2007. **9**(3): p. 258-267.
30. Rydzak, T., et al., *End-product induced metabolic shifts in Clostridium thermocellum ATCC 27405.* Appl Microbiol Biotechnol, 2011. **92**(1): p. 199-209.
31. Biswas, R., et al., *Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in Clostridium thermocellum.* Biotechnol Biofuel, 2015. **8**: p. 20.
32. Rydzak, T., L.R. Lynd, and A.M. Guss, *Elimination of formate production in Clostridium thermocellum.* J Ind Microbiol Biotechnol, 2015. **42**(9): p. 1263-1272.
33. Papanek, B., et al., *Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in Clostridium thermocellum.* Metab Eng, 2015. **32**.
34. Holwerda, E., et al., *The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading.* Biotechnol Biofuels, 2014. **7**.
35. Lo, J., et al., *Engineering electron metabolism to increase ethanol production in Clostridium thermocellum.* Metab Eng, 2017. **39**: p. 71-79.
36. Rydzak, T., et al., *Deletion of Type I glutamine synthetase deregulates nitrogen metabolism and increases ethanol production in Clostridium thermocellum.* Metab Eng, 2017. **41**: p. 182-191.
37. Lo, J., et al., *The bifunctional alcohol and aldehyde dehydrogenase gene, adhE, is necessary for ethanol production in Clostridium thermocellum and Thermoanaerobacterium saccharolyticum.* J Bacteriol, 2015. **197**.
38. Thompson, R.A., et al., *Elucidating central metabolic redox obstacles hindering ethanol production in Clostridium thermocellum.* Metab Eng, 2015. **32**: p. 207-219.
39. Cha, M., D. Chung, and J. Westpheling, *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, 2015: p. 1-9.
40. Cha, M., et al., *Metabolic engineering of Caldicellulosiruptor bescii yields increased hydrogen production from lignocellulosic biomass.* Biotechnol Biofuel, 2013. **6**(1): p. 85.
41. Buckel, W. and R.K. Thauer, *Energy conservation via electron bifurcating ferredoxin reduction and proton/Na+ translocating ferredoxin oxidation.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2013. **1827**(2): p. 94-113.
42. Berg, J.M., J.L. Tymoczko, and L. Stryer, *Biochemistry*. Vol. 6. 2002, New York: W. H. Freeman and Company.

43. Ravcheev, D.A., et al., *Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex.* J Bacteriol, 2012. **194**(5): p. 1145-57.
44. Wimpenny, J.W.T. and A. Firth, *Levels of Nicotinamide Adenine Dinucleotide and Reduced Nicotinamide Adenine Dinucleotide in Facultative Bacteria and the Effect of Oxygen.* J Bacteriol, 1972. **111**(1): p. 24-32.
45. Bennett, B.D., et al., *Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach.* Nat Protocol, 2008. **3**(8): p. 1299-311.
46. Ravcheev, D.A., et al., *Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex.* J Bacteriol, 2012. **194**.
47. Brekasis, D. and M.S. Paget, *A novel sensor of NADH/NAD$^+$ redox poise in Streptomyces coelicolor A3(2).* EMBO J, 2003. **22**(18): p. 4856-65.
48. Zhang, L., et al., *Redox-responsive repressor Rex modulates alcohol production and oxidative stress tolerance in Clostridium acetobutylicum.* J Bacteriol, 2014. **196**(22): p. 3949-63.
49. Zurawski, J.V., et al., *Comparative Analysis of Extremely Thermophilic Caldicellulosiruptor Species Reveals Common and Unique Cellular Strategies for Plant Biomass Utilization.* Applied and Environmental Microbiology, 2015. **81**(20): p. 7159-7170.
50. Bielen, A.A.M., et al., *A thermophile under pressure: Transcriptional analysis of the response of Caldicellulosiruptor saccharolyticus to different H_2 partial pressures.* Int J Hydrogen Energy, 2013. **38**(4): p. 1837-1849.
51. Farkas, J., et al., *Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by Caldicellulosiruptor bescii.* J Ind Microbiol Biotechnol, 2013. **40**(1): p. 41-49.
52. Sander, K., et al., *Clostridium thermocellum DSM 1313 transcriptional responses to redox perturbation.* Biotechnol Biofuel, 2015. **8**(1): p. 211.
53. Chung, D., et al., *Methylation by a Unique α-class N4-Cytosine Methyltransferase Is Required for DNA Transformation of <italic>Caldicellulosiruptor bescii</italic> DSM6725.* PLoS One, 2012. **7**(8): p. e43844.
54. Chung, D., et al., *Construction of a Stable Replicating Shuttle Vector for Caldicellulosiruptor Species: Use for Extending Genetic Methodologies to Other Members of This Genus.* PLoS ONE, 2013. **8**(5): p. e62881.
55. Lipscomb, G.L., et al., *A Highly Thermostable Kanamycin Resistance Marker Expands the Tool Kit for Genetic Manipulation of Caldicellulosiruptor bescii.* Applied and Environmental Microbiology, 2016. **82**(14): p. 4421-4428.
56. Chung, D., et al., *Cellulosic ethanol production via consolidated bioprocessing at 75°C by engineered Caldicellulosiruptor bescii.* Biotechnol Biofuel, 2015. **8**(1): p. 163.
57. Blumer-Schuette, S.E., et al., *Thermophilic lignocellulose deconstruction.* FEMS Microbiol. Rev., 2013. **38**: p. 393–448.
58. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update.* Curr. Opin. Biotechnol., 2005. **16**(5): p. 577-583.
59. Shoham, Y., R. Lamed, and E.A. Bayer, *The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides.* Trends Microbiol, 1999. **7**(7): p. 275-81.
60. Lynd, L.R., H.E. Grethlein, and R.H. Wolkin, *Fermentation of cellulosic substrates in batch and continuous culture by Clostridium thermocellum.* Appl. Environ. Microbiol., 1989. **55**(12): p. 3131-3139.
61. Biswas, R., et al., *Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in Clostridium thermocellum.* Biotechnol Biofuels, 2015. **8**(1): p. 20.
62. Argyros, D.A., et al., *High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes.* Appl. Environ. Microbiol., 2011. **77**: p. 8288-8294.

63. Tripathi, S.A., et al., *Development of pyrF-based genetic system for targeted gene deletion in Clostridium thermocellum and creation of a pta mutant.* Appl. Environ. Microbiol., 2010. **76**: p. 6591-9.
64. Yee, K.L., et al., *Consolidated bioprocessing of transgenic switchgrass by an engineered and evolved Clostridium thermocellum strain.* Biotechnol Biofuels, 2014. **7**: p. 75.
65. Deng, Y., et al., *Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in Clostridium thermocellum.* Metabolic Eng., 2013. **15**: p. 151-158.
66. Holwerda, E., et al., *The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading.* Biotechnol. Biofuels, 2014. **7**(1): p. 155.
67. van der Veen, D., et al., *Characterization of Clostridium thermocellum strains with disrupted fermentation end-product pathways.* J Ind Microbiol Biotechnol, 2013. **40**: p. 725-734.
68. Rydzak, T., et al., *End-product induced metabolic shifts in Clostridium thermocellum ATCC 27405.* Appl. Microbiol. Biotechnol., 2011. **92**(1): p. 199-209.
69. Rydzak, T., et al., *Insights into electron flux through manipulation of fermentation conditions and assessment of protein expression profiles in Clostridium thermocellum.* Appl Microbiol Biotechnol, 2014. **98**(14): p. 6497-6510.
70. Wilson, C.M., et al., *Global transcriptome analysis of Clostridium thermocellum ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass.* Biotechnol. Biofuels, 2013. **6**: p. 179.
71. Wilson, C.M., et al., *Clostridium thermocellum transcriptomic profiles after exposure to furfural or heat stress.* Biotechnol. Biofuels, 2013. **6**: p. 131.
72. Yang, S., et al., *Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress.* BMC Genomics, 2012. **13**(1): p. 336.
73. Linville, J.L., et al., *Transcriptomic analysis of Clostridium thermocellum Populus hydrolysate-tolerant mutant strain shows increased cellular efficiency in response to Populus hydrolysate compared to the wild type strain.* BMC Microbiol, 2014. **14**(1): p. 215.
74. Beliaev, A.S., et al., *Global transcriptome analysis of Shewanella oneidensis MR-1 exposed to different terminal electron acceptors.* J Bacteriol, 2005. **187**(20): p. 7138-7145.
75. Honicke, D., et al., *Global transcriptional changes of Clostridium acetobutylicum cultures with increased butanol: acetone ratios.* New Biotechnol, 2012. **29**(4): p. 485 - 493.
76. Ravcheev, D.A., et al., *Transcriptional regulation of central carbon and energy metabolism in bacteria by redox-responsive repressor Rex.* J. Bacteriol., 2012. **194**: p. 1145-1157.
77. Christensen, G.A., et al., *Rex (encoded by DVU_0916) in Desulfovibrio vulgaris Hildenborough is a repressor of sulfate adenylyl transferase and is Regulated by NADH.* J. Bacteriol., 2015. **197**: p. 29-39.
78. Wietzke, M. and H. Bahl, *The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in Clostridium acetobutylicum.* Appl. Microbiol. Biotechnol., 2012. **96**(3): p. 749-761.
79. Korte, H., et al., *Genetic basis for nitrate resistance in Desulfovibrio strains.* Frontiers Microbiol., 2014. **5**: p. 153.
80. Desvaux, M., *Unravelling carbon metabolism in snaerobic cellulolytic bacteria.* Biotechnol Prog, 2006. **22**(5): p. 1229-1238.
81. Payot, S., et al., *Metabolism of cellobiose by Clostridium cellulolyticum growing in continuous culture: evidence for decreased NADH reoxidation as a factor limiting growth.* Microbiol, 1998. **144**(2): p. 375-384.
82. Li, H.-F., et al., *Metabolic control of Clostridium thermocellum via inhibition of hydrogenase activity and the glucose transport rate.* Appl. Microbiol. Biotechnol., 2012. **93**(4): p. 1777-1784.

83. Reimann, A., H. Biebl, and W.D. Deckwer, *Influence of iron, phosphate and methyl viologen on glycerol fermentation of Clostridium butyricum.* Appl. Microbiol. Biotechnol., 1996. **45**(1-2): p. 47-50.
84. Rao, G. and R. Mutharasan, *Alcohol production by Clostridium acetobutylicum induced by methyl viologen.* Biotechnol. Lett., 1986. **8**(12): p. 893-896.
85. Rao, G. and R. Mutharasan, *Altered electron flow in continuous cultures of Clostridium acetobutylicum induced by viologen dyes.* Appl Environ Microbiol, 1987. **53**(6): p. 1232-1235.
86. Thorneley, R.N.F., *A convenient electrochemical preparation of reduced methyl viologen and a kinetic study of the reaction with oxygen using an anaerobic stopped-flow apparatus.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1974. **333**(3): p. 487-496.
87. Tsibris, J.C.M. and R.W. Woody, *Structural studies of iron-sulfur proteins.* Coordination Chem. Rev., 1970. **5**(4): p. 417-458.
88. Ware, D.A., *Nitrogenase of Klebsiella pneumoniae: interaction with viologen dyes as measured by acetylene reduction.* Biochem. J., 1972. **130**(1): p. 301-302.
89. Guerrero, M.G., et al., *Preparation and characterization of a soluble nitrate reductase from Azotobacter chroococcum.* Arch. Mikrobiol. , 1973. **91**: p. 287-304.
90. Dietrich, L.E.P., et al., *The phenazine pyocyanin is a terminal signalling factor in the quorum sensing network of Pseudomonas aeruginosa.* Molecular Microbiology, 2006. **61**(5): p. 1308-1321.
91. Gu, M. and J.A. Imlay, *The SoxRS response of Escherichia coli is directly activated by redox-cycling drugs rather than by superoxide.* Molecular Microbiology, 2011. **79**(5): p. 1136-1150.
92. Taillefer, M., et al., *Reassesment of the transhydrogenase 'malate shunt' in Clostridium thermocellum ATCC 27405 through kinetic characterization of malic enzyme and malate dehydrogenase.* Appl. Environ. Microbiol., 2015. **81**: p. 2423-32.
93. Bogdahn, M. and D. Kleiner, *Inorganic nitrogen metabolism in two cellulose-degrading Clostridia.* Arch Microbiol, 1986. **145**(2): p. 159-161.
94. Pengpeng, W. and Z. Tan, *Ammonia assimilation in rumen bacteria: a review.* Anim Biotechnol, 2013. **24**(2): p. 107-28.
95. Fedorova, K.P., et al., *The role of AmtB, GlnK and glutamine synthetase in regulation of of transcription factor TnrA in Bacillus subtilis.* Cell and Tissue Biology, 2013. **7**(3): p. 297-301.
96. Fierro-Monti, I.P., S.J. Reid, and D.R. Woods, *Differential expression of a Clostridium acetobutylicum antisense RNA: implications for regulation of glutamine synthetase.* J Bacteriol, 1992. **174**(23): p. 7642-7.
97. Kridelbaugh, D.M., et al., *Nitrogen and sulfur requirements for Clostridium thermocellum and Caldicellulosiruptor bescii on cellulosic substrates in minimal nutrient media.* Bioresource Technol, 2013. **130**(0): p. 125-135.
98. Schobert, M. and D. Jahn, *Regulation of heme biosynthesis in non-phototrophic bacteria.* J Mol Microbiol Biotechnol, 2002. **4**(3): p. 287-94.
99. Hungerer, C., et al., *Regulation of the hemA gene during 5-aminolevulinic acid formation in Pseudomonas aeruginosa.* J Bacteriol, 1995. **177**(6): p. 1435-1443.
100. Erbes, D.L. and R.H. Burris, *The kinetics of methyl viologen oxidation and reduction by the hydrogenase from Clostridium pasteurianium.* Biochimica et Biophysica Acta (BBA) - Enzymology, 1978. **525**(1): p. 45-54.
101. Okura, I., K.-i. Nakamura, and S. Nakamura, *Kinetics of methyl viologen reduction by hydrogen catalyzed by hydrogenase from Desulfovibrio vulgaris.* J Inorg Biochem, 1981. **14**(2): p. 155-161.
102. Rydzak, T., et al., *Growth phase-dependant enzyme profile of pyruvate catabolism and end-product formation in Clostridium thermocellum ATCC 27405.* J. Biotechnol., 2009. **140**(3-4): p. 169-75.

103. Carere, C., et al., *Role of transcription and enzyme activities in redistribution of carbon and electron flux in response to N_2 and H_2 sparging of open-batch cultures of Clostridium thermocellum ATCC 27405.* Appl. Microbiol. Biotechnol., 2014. **98**: p. 2829-2840.
104. Carere, C.R., et al., *Pyruvate catabolism and hydrogen synthesis pathway genes of Clostridium thermocellum ATCC 27405.* Ind. J. Microbiol., 2008. **48**(2): p. 252-266.
105. Vignais, P.M., *Regulation of hydrogenase gene expression*, in *The Purple Phototrophic Bacteria*, C.N. Hunter, et al., Editors. 2009, Springer Netherlands. p. 743-757.
106. Rey, F.E., Y. Oda, and C.S. Harwood, *Regulation of uptake hydrogenase and effects of hydrogen utilization on gene expression in Rhodopseudomonas palustris.* J. Bacteriol., 2006. **188**(17): p. 6143-52.
107. Mearls, E.B., J.A. Izquierdo, and L.R. Lynd, *Formation and characterization of non-growth states in Clostridium thermocellum: spores and L-forms.* BMC Microbiol., 2012. **12**(1): p. 180.
108. Mearls, E.B. and L.R. Lynd, *The identification of four histidine kinases that influence sporulation in Clostridium thermocellum.* Anaerobe, 2014. **28**: p. 109-19.
109. Zverlov, V.V., et al., *Mutations in the scaffoldin gene, cipA, of Clostridium thermocellum with impaired cellulosome formation and cellulose hydrolysis: Insertions of a new transposable element, IS1447, and implications for cellulase synergism on crystalline cellulose.* J Bacteriol, 2008. **190**(12): p. 4321-4327.
110. Brown, S.D., et al., *Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of Clostridium autoethanogenum and analysis of CRISPR systems in industrial relevant Clostridia.* Biotechnol. Biofuels, 2014. **7**: p. 40.
111. and, D.R. and J. Beckwith, *Roles of Thiol-Redox Pathways in Bacteria.* Annual Review of Microbiology, 2001. **55**(1): p. 21-48.
112. Zhang, Y. and L.R. Lynd, *Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an enzyme-linked immunosorbent assay-based method with application to Clostridium thermocellum batch cultures.* Anal Chem, 2002. **75**(2): p. 219-227.
113. Karp, P.D., et al., *Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology.* Briefings in Bioinformatics, 2010. **11**(1): p. 40-79.
114. Biswas, R., et al., *Elimination of hydrogenase active site assembly blocks H_2 production and increases ethanol yield in Clostridium thermocellum.* Biotechnol Biofuel, 2015. **8**.
115. Argyros, D.A., et al., *High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes.* Appl Environ Microbiol, 2011. **77**(23): p. 8288-8294.
116. Lin, P.P., et al., *Consolidated bioprocessing of cellulose to isobutanol using Clostridium thermocellum.* Metab Eng, 2015. **31**: p. 44-52.
117. Herring, C.D., et al., *Strain and bioprocess improvement of a thermophilic anaerobe for the production of ethanol from wood.* Biotechol Biofuel, 2016. **9**(1): p. 125.
118. Papanek, B., et al., *Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in Clostridium thermocellum.* Metab Eng, 2015. **32**: p. 49-54.
119. Sander, K., et al., *Clostridium thermocellum DSM 1313 transcriptional responses to redox perturbation.* Biotechnol Biofuel, 2015. **8**(1): p. 1-14.
120. Lo, J., et al., *Deletion of nfnAB in Thermoanaerobacterium saccharolyticum and its effect on metabolism.* J Bacteriol, 2015.
121. Brown, S.D., et al., *Mutant alcohol dehydrogenase leads to improved ethanol tolerance in Clostridium thermocellum.* Proc Nat Acad Sci USA, 2011. **108**(33): p. 13752-13757.
122. Zheng, T., et al., *Cofactor Specificity of the Bifunctional Alcohol and Aldehyde Dehydrogenase (AdhE) in Wild-Type and Mutant Clostridium thermocellum and Thermoanaerobacterium saccharolyticum.* J Bacteriol, 2015. **197**(15): p. 2610-2619.

123. Olson, D.G., et al., *Glycolysis without pyruvate kinase in Clostridium thermocellum.* Metab Eng, 2017.
124. Xiong, W., et al., *CO_2-fixing one-carbon metabolism in a cellulose-degrading bacterium Clostridium thermocellum.* Proc Natl Acad Sci U S A, 2016. **113**(46): p. 13180-13185.
125. Qiao, K., et al., *Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism.* Nat Biotech, 2017. **35**: p. 173-177.
126. Davis, R., et al., *Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbons: Dilute-Acid and Enzymatic Deconstruction of Biomass to Sugars and Biological Conversion of Sugars to Hydrocarbons*. 2013.
127. Lai, B., et al., *Anoxic metabolism and biochemical production in Pseudomonas putida F1 driven by a bioelectrochemical system.* Biotechnol Biofuel, 2016. **9**(1): p. 39.
128. Cui, J., et al., *Separation and Quantification of Water-Soluble Cellular Metabolites in Clostridium thermocellum using Liquid Chromatography-Isotope Dilution Tandem Mass Spectrometry.* Anal Lett, 2013. **46**(17): p. 2767-2786.
129. Beri, D., et al., *Nicotinamide cofactor ratios in engineered strains of Clostridium thermocellum and Thermoanaerobacterium saccharolyticum.* FEMS Microbiol Lett, 2016.
130. Flamholz, A., et al., *Glycolytic strategy as a tradeoff between energy yield and protein cost.* Proc Natl Acad Sci U S A, 2013. **110**(24): p. 10039-10044.
131. Pradet, A. and P. Raymond, *Adenine Nucleotide Ratios and Adenylate Energy Charge in Energy Metabolism.* Annu Rev Plant Physiol, 1983. **34**(1): p. 199-224.
132. Oakhill, J.S., et al., *AMPK Is a Direct Adenylate Charge-Regulated Protein Kinase.* Science, 2011. **332**(6036): p. 1433-1435.
133. Atkinson, D.E., *Energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers.* Biochem, 1968. **7**(11): p. 4030-4034.
134. Karl, D.M. and O. Holm-Hansen, *Methodology and measurement of adenylate energy charge ratios in environmental samples.* Mar Biol, 1978. **48**(2): p. 185-197.
135. Wiebe, W.J. and K. Bancroft, *Use of the adenylate energy charge ratio to measure growth state of natural microbial communities.* Proc Natl Acad Sci U S A, 1975. **72**(6): p. 2112-2115.
136. Chapman, A.G., L. Fall, and D.E. Atkinson, *Adenylate energy charge in Escherichia coli during growth and starvation.* J Bacteriol, 1971. **108**.
137. Erskine, J.M. and M.S. Koch, *Sulfide effects on Thalassia testudinum carbon balance and adenylate energy charge.* Aquat Bot, 2000. **67**(4): p. 275-285.
138. Lowry, O.H., et al., *The Measurement of Pyridine Nucleotides by Enzymatic Cycling.* J Biol Chem, 1961. **236**(10): p. 2746-2755.
139. Lilius, E.-M., V.-M. Multanen, and V. Toivonen, *Quantitative extraction and estimation of intracellular nicotinamide nucleotides of Escherichia coli.* Anal Biochem, 1979. **99**(1): p. 22-27.
140. Bennett, B.D., et al., *Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli.* Nat Chem Biol, 2009. **5**(8): p. 593-599.
141. Amador-Noguez, D., et al., *Metabolome Remodeling during the Acidogenic-Solventogenic Transition in Clostridium acetobutylicum.* Appl Environ Microbiol, 2011. **77**(22): p. 7984-7997.
142. Bolten, C.J., et al., *Sampling for metabolome analysis of microorganisms.* Anal Chem, 2007. **79**.
143. Japelt, K.B., J.H. Christensen, and S.G. Villas-Boas, *Metabolic fingerprinting of Lactobacillus paracasei: the optimal quenching strategy.* Microb Cell Fact, 2015. **14**: p. 132.
144. Link, H., B. Anselment, and D. Weuster-Botz, *Leakage of adenylates during cold methanol/glycerol quenching of Escherichia coli.* Metabolomics, 2008. **4**(3): p. 240-247.
145. Meyer, H., M. Liebeke, and M. Lalk, *A protocol for the investigation of the intracellular Staphylococcus aureus metabolome.* Anal Biochem, 2010. **401**(2): p. 250-259.

146. Bielen, A.A.M., et al., *Pyrophosphate as a central energy carrier in the hydrogen-producing extremely thermophilic Caldicellulosiruptor saccharolyticus.* FEMS Microbiol Lett, 2010. **307**(1): p. 48-54.
147. Meyer, C.L. and E.T. Papoutsakis, *Increased levels of ATP and NADH are associated with increased solvent production in continuous cultures of Clostridium acetobutylicum.* Appl Microbiol Biotechnol, 1989. **30**(5): p. 450-459.
148. Lowry, O.H., J.V. Passonneau, and M.K. Rock, *The Stability of Pyridine Nucleotides.* J Biol Chem, 1961. **236**(10): p. 2756-2759.
149. Wellerdiek, M., et al., *Metabolic quenching of Corynebacterium glutamicum: efficiency of methods and impact of cold shock.* Bioprocess Biosyst Eng, 2009. **32**(5): p. 581-592.
150. Koning, W.d. and K.v. Dam, *A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH.* Anal Biochem, 1992. **204**(1): p. 118-123.
151. van Gulik, W.M., *Fast sampling for quantitative microbial metabolomics.* Curr Opin Biotechnol, 2010. **21**(1): p. 27-34.
152. Canelas, A.B., et al., *Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics.* Anal Chem, 2009. **81**(17): p. 7379-7389.
153. Taymaz-Nikerel, H., et al., *Development and application of a differential method for reliable metabolome analysis in Escherichia coli.* Anal Biochem, 2009. **386**(1): p. 9-19.
154. McCloskey, D., et al., *A model-driven quantitative metabolomics analysis of aerobic and anaerobic metabolism in E. coli K-12 MG1655 that is biochemically and thermodynamically consistent.* Biotechnol Bioeng, 2014. **111**(4): p. 803-815.
155. Lu, H., et al., *Integrated isotope-assisted metabolomics and ^{13}C metabolic flux analysis reveals metabolic flux redistribution for high glucoamylase production by Aspergillus niger.* Microb Cell Fact, 2015. **14**(1): p. 147.
156. Wu, L., et al., *Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C-labeled cell extracts as internal standards.* Anal Biochem, 2005. **336**(2): p. 164-171.
157. Chance, B. and B. Thorell, *Localization and Kinetics of Reduced Pyridine Nucleotide in Living Cells by Microfluorometry.* Journal of Biological Chemistry, 1959. **234**(11): p. 3044-3050.
158. Thauer, R.K., K. Jungermann, and K. Decker, *Energy conservation in chemotrophic anaerobic bacteria.* Bacterol Rev, 1977. **41**(1): p. 100-180.
159. Huang, S., A.A. Heikal, and W.W. Webb, *Two-Photon Fluorescence Spectroscopy and Microscopy of NAD(P)H and Flavoprotein.* Biophys J, 2002. **82**(5): p. 2811-2825.
160. Ince, C., J.M.C.C. Coremans, and H.A. Bruining, *In Vivo NADH Fluorescence*, in *Oxygen Transport to Tissue XIV*, W. Erdmann and D.F. Bruley, Editors. 1992, Springer US: Boston, MA. p. 277-296.
161. Lakowicz, J.R., et al., *Fluorescence lifetime imaging of free and protein-bound NADH.* Proc Natl Acad Sci U S A, 1992. **89**(4): p. 1271-1275.
162. Mayevsky, A. and E. Barbiro-Michaely, *Use of NADH fluorescence to determine mitochondrial function in vivo.* Int J Biochem Cell Biol, 2009. **41**(10): p. 1977-1988.
163. Rehman, A.U., et al., *Fluorescence quenching of free and bound NADH in HeLa cells determined by hyperspectral imaging and unmixing of cell autofluorescence.* Biomed Opt Express, 2017. **8**(3): p. 1488-1498.
164. Kasimova, M.R., et al., *The free NADH concentration is kept constant in plant mitochondria under different metabolic conditions.* Plant Cell, 2006. **18**(3): p. 688-98.
165. Zhao, Y., et al., *SoNar, a Highly Responsive NAD+/NADH Sensor, Allows High-Throughput Metabolic Screening of Anti-tumor Agents.* Cell Metab, 2015. **21**(5): p. 777-789.

166. Hung, Yin P., et al., *Imaging Cytosolic NADH-NAD+ Redox State with a Genetically Encoded Fluorescent Biosensor.* Cell Metab, 2011. **14**(4): p. 545-554.
167. Zhao, Y. and Y. Yang, *Frex and FrexH: Indicators of metabolic states in living cells.* Bioeng Bugs, 2012. **3**(3): p. 181-188.
168. Knudsen, J.D., M. Carlquist, and M. Gorwa-Grauslund, *NADH-dependent biosensor in Saccharomyces cerevisiae: principle and validation at the single cell level.* AMB Express, 2014. **4**(1): p. 81.
169. Moussa, R., et al., *An evaluation of genetically encoded FRET-based biosensors for quantitative metabolite analyses in vivo.* J Biotechnol, 2014. **191**: p. 250-259.
170. Tian, X., et al., *The effect of redox environment on l-lactic acid production by Lactobacillus paracasei—A proof by genetically encoded in vivo NADH biosensor.* Proc Biochem, 2015. **50**(12): p. 2029-2034.
171. Scully, S.M. and J. Orlygsson, *Recent Advances in Genetic Engineering of Thermophilic Ethanol Producing Bacteria*, in *Engineering of Microorganisms for the Production of Chemicals and Biofuels from Renewable Resources*, G. Gosset, Editor. 2017, Springer International Publishing: Cham. p. 1-29.
172. Bilan, D.S., et al., *Genetically encoded fluorescent indicator for imaging NAD+/NADH ratio changes in different cellular compartments.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2014. **1840**(3): p. 951-957.
173. Akhtar, M.H., et al., *Sensitive NADH detection in a tumorigenic cell line using a nano-biosensor based on the organic complex formation.* Biosens Bioelectron, 2016. **85**: p. 488-495.
174. Han, S., et al., *Synergistic effect of pyrroloquinoline quinone and graphene nano-interface for facile fabrication of sensitive NADH biosensor.* Biosens Bioelectron, 2017. **89, Part 1**: p. 422-429.
175. Shanks, J.V., *In situ NMR systems.* Curr Issues Mol Biol, 2001. **3**(1): p. 15-26.
176. Nagana Gowda, G.A., et al., *Simultaneous Analysis of Major Coenzymes of Cellular Redox Reactions and Energy Using Ex Vivo ^{1}H NMR Spectroscopy.* Anal Chem, 2016. **88**(9): p. 4817-4824.
177. Markley, J.L., et al., *The future of NMR-based metabolomics.* Curr Opin Biotechnol., 2017. **43**: p. 34-40.
178. Tolman, C.J., S. Kanodia, and M.F. Roberts, *^{31}P and ^{13}C NMR analyses of the energy metabolism of the thermophilic anaerobe Clostridium thermocellum.* J Biol Chem, 1987. **262**(23): p. 11088-96.
179. Nochur, S.V., A.L. Demain, and M.F. Roberts, *Carbohydrate utilization by Clostridium thermocellum: Importance of internal pH in regulating growth.* Enz Microb Technol, 1992. **14**(5): p. 338-349.
180. Muhamadali, H., et al., *Metabolic Profiling of Geobacter sulfurreducens during Industrial Bioprocess Scale-Up.* Appl Environ Microbiol, 2015. **81**(10): p. 3288-3298.
181. Lo, J., et al., *The bifunctional alcohol and aldehyde dehydrogenase gene, adhE, is necessary for ethanol production in Clostridium thermocellum and Thermoanaerobacterium saccharolyticum.* J Bacteriol, 2015.
182. Olson, D.G., et al., *Recent progress in consolidated bioprocessing.* Curr Opin Biotechnol, 2012. **23**(3): p. 396-405.
183. Rabinowitz, J.D. and E. Kimball, *Acidic Acetonitrile for Cellular Metabolome Extraction from Escherichia coli.* Anal. Chem., 2007. **79**: p. 6167-6173.
184. Garrity, G., et al., *Bergey's Manual of Systematic Bacteriology. Volume 3. The Firmicutes*. 2010.
185. Park, C., et al., *Metabolic Profiling of Klebsiella oxytoca: Evaluation of Methods for Extraction of Intracellular Metabolites Using UPLC/Q-TOF-MS.* Appl Biochem Biotechnol, 2012. **167**(3): p. 425-438.

186. Johnson, E.A., A. Madia, and A.L. Demain, *Chemically Defined Minimal Medium for Growth of the Anaerobic Cellulolytic Thermophile Clostridium thermocellum.* Appl Environ Microbiol, 1981. **41**(4): p. 1060-1062.
187. Rydzak, T., et al., *End-product induced metabolic shifts in Clostridium thermocellum ATCC 27405.* Appl Environ Microbiol, 2011. **92**(1): p. 199.
188. Girbal, L. and P. Soucaille, *Regulation of Clostridium acetobutylicum metabolism as revealed by mixed-substrate steady-state continuous cultures: role of NADH/NAD ratio and ATP pool.* J Bacteriol, 1994. **176**(21): p. 6433-6438.
189. Even, S., et al., *Pyruvate Metabolism in Lactococcus lactis Is Dependent upon Glyceraldehyde-3-phosphate Dehydrogenase Activity.* Metab Eng, 1999. **1**(3): p. 198-205.
190. Caspi, R., et al., *The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.* Nucleic Acids Research, 2014. **42**(D1): p. D459-D471.
191. Sporty, J.L., et al., *Single sample extraction protocol for the quantification of NAD and NADH redox states in Saccharomyces cerevisiae.* J Sep Sci, 2008. **31**(18): p. 3202-11.
192. Canelas, A.B., W.M. van Gulik, and J.J. Heijnen, *Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions.* Biotechnol Bioeng, 2008. **100**(4): p. 734-743.
193. Marshall, W.E. and A. Omachi, *Measured and calculated NAD+/NADH ratios in human erythrocytes.* Biochim Biophys Acta, 1974. **354**(1): p. 1-10.
194. Sun, F., et al., *Biochemical Issues in Estimation of Cytosolic Free NAD/NADH Ratio.* PLoS ONE, 2012. **7**(5): p. e34525.
195. Hogsett, D.A.L., *Cellulose hydrolysis and fermentation by Clostridium thermocellum for the production of ethanol*. 1995, Dartmouth College.
196. Yang, S.-J., et al., *Classification of 'Anaerocellum thermophilum' strain DSM 6725 as Caldicellulosiruptor bescii sp. nov.* Int J Syst Evol Microbiol, 2010. **60**(9): p. 2011-2015.
197. Bielen, A., et al., *Biohydrogen Production by the Thermophilic Bacterium Caldicellulosiruptor saccharolyticus: Current Status and Perspectives.* Life, 2013. **3**(1): p. 52.
198. van de Werken, H.J.G., et al., *Hydrogenomics of the extremely thermophilic bacterium Caldicellulosiruptor saccharolyticus.* Appl Environ Microbiol, 2008. **74**.
199. Blumer-Schuette, S.E., et al., *Discrete and Structurally Unique Proteins (Tāpirins) Mediate Attachment of Extremely Thermophilic Caldicellulosiruptor Species to Cellulose.* Journal of Biological Chemistry, 2015. **290**(17): p. 10645-10656.
200. Blumer-Schuette, S.E., et al., *Caldicellulosiruptor Core and Pangenomes Reveal Determinants for Noncellulosomal Thermophilic Deconstruction of Plant Biomass.* J Bacteriol, 2012. **194**(15): p. 4015-4028.
201. Li, Y., et al., *Combined inactivation of the Clostridium cellulolyticum lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations.* Biotechnol Biofuel, 2012. **5**(1): p. 2.
202. Sickmier, E.A., et al., *X-ray structure of a Rex-family repressor/NADH complex insights into the mechanism of redox sensing.* Structure, 2005. **13**(1): p. 43-54.
203. Christensen, G.A., et al., *Rex (Encoded by DVU_0916) in Desulfovibrio vulgaris Hildenborough Is a Repressor of Sulfate Adenylyl Transferase and Is Regulated by NADH.* J Bacteriol, 2015. **197**(1): p. 29-39.
204. Bitoun, J.P., et al., *Transcriptional repressor Rex is involved in regulation of oxidative stress response and biofilm formation by Streptococcus mutans.* FEMS Microbiol Lett, 2011. **320**(2): p. 110-7.
205. Carlson, H.K., et al., *Mechanisms of direct inhibition of the respiratory sulfate-reduction pathway by (per)chlorate and nitrate.* ISME J, 2015. **9**(6): p. 1295-305.

206. Larsson, J.T., A. Rogstam, and C. von Wachenfeldt, *Coordinated patterns of cytochrome bd and lactate dehydrogenase expression in Bacillus subtilis.* Microbiol, 2005. **151**(Pt 10): p. 3323-35.
207. Novichkov, P.S., et al., *RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria.* BMC Genomics, 2013. **14**: p. 745.
208. McLaughlin, K.J., et al., *Structural Basis for NADH/NAD$^+$ Redox Sensing by a Rex Family Repressor.* Mol Cell, 2010. **38**(4): p. 563-575.
209. Pagels, M., et al., *Redox sensing by a Rex-family repressor is involved in the regulation of anaerobic gene expression in Staphylococcus aureus.* Mol Microbiol, 2010. **76**(5): p. 1142-1161.
210. Pei, J., et al., *The mechanism for regulating ethanol fermentation by redox levels in Thermoanaerobacter ethanolicus.* Metab Eng, 2011. **13**(2): p. 186-93.
211. Wietzke, M. and H. Bahl, *The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in Clostridium acetobutylicum.* Appl Microbiol Biotechnol, 2012.
212. Cha, M., et al., *Isolation and bioinformatic analysis of a novel transposable element, ISCbe4, from the hyperthermophilic bacterium Caldicellulosiruptor bescii.* J Ind Microbiol Biotechnol, 2013. **40**(12): p. 1443-1448.
213. Chung, D., J. Farkas, and J. Westpheling, *Detection of a novel active transposable element in Caldicellulosiruptor hydrothermalis and a new search for elements in this genus.* J Ind Microbiol Biotechnol, 2013. **40**(5): p. 517-521.
214. Scott, I.M., et al., *A New Class of Tungsten-Containing Oxidoreductase in the Genus of the Plant Biomass-Degrading, Thermophilic Bacteria Caldicellulosiruptor.* J Appl Environ Microbiol, 2015.
215. Wang, E., et al., *Structure and functional properties of the Bacillus subtilis transcriptional repressor Rex.* Mol Microbiol, 2008. **69**(2): p. 466-478.
216. Mao, X., et al., *DOOR 2.0: presenting operons and their functions through dynamic and integrated views.* Nucleic Acids Res, 2014. **42**(Database issue): p. D654-9.
217. Kuchenreuther, J.M., et al., *The HydG Enzyme Generates an Fe(CO)$_2$(CN) Synthon in Assembly of the FeFe Hydrogenase H-Cluster.* Science, 2014. **343**(6169): p. 424-427.
218. Kostesha, N., et al., *Probing the redox metabolism in the strictly anaerobic, extremely thermophilic, hydrogen-producing Caldicellulosiruptor saccharolyticus using amperometry.* Extremophiles, 2011. **15**(1): p. 77-87.
219. Desvaux, M. and H. Petitdemange, *Flux Analysis of the Metabolism of Clostridium cellulolyticum Grown in Cellulose-Fed Continuous Culture on a Chemically Defined Medium under Ammonium-Limited Conditions.* Appl Environ Microbiol, 2001. **67**(9): p. 3846-3851.
220. Kanehisa, M., et al., *KEGG as a reference resource for gene and protein annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D457-62.
221. Banerjee, R., *Redox outside the Box: Linking Extracellular Redox Remodeling with Intracellular Redox Metabolism.* J Biol Chem, 2012. **287**(7): p. 4397-4402.
222. Ballester-Tomás, L., et al., *Redox engineering by ectopic expression of glutamate dehydrogenase genes links NADPH availability and NADH oxidation with cold growth in Saccharomyces cerevisiae.* Microbial Cell Factories, 2015. **14**(1): p. 100.
223. Mailloux, R.J., et al., *α-Ketoglutarate Dehydrogenase and Glutamate Dehydrogenase Work in Tandem To Modulate the Antioxidant α-Ketoglutarate during Oxidative Stress in Pseudomonas fluorescens.* Journal of Bacteriology, 2009. **191**(12): p. 3804-3810.
224. Park, S. and J.A. Imlay, *High Levels of Intracellular Cysteine Promote Oxidative DNA Damage by Driving the Fenton Reaction.* J Bacteriol, 2003. **185**(6): p. 1942-1950.
225. Holwerda, E.K., et al., *The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading.* Biotechnol Biofuel, 2014. **7**(1): p. 155.
226. Biswas, R., et al., *Improved growth rate in Clostridium thermocellum hydrogenase mutant via perturbed sulfur metabolism.* Biotechnol Biofuel, 2017. **10**(1): p. 6.

227. Singh, A., et al., *Manipulating redox and ATP balancing for improved production of succinate in E. coli.* Metab Eng, 2011. **13**(1): p. 76-81.
228. Guedon, E., et al., *Carbon and Electron Flow in Clostridium cellulolyticum Grown in Chemostat Culture on Synthetic Medium.* J Bacteriol, 1999. **181**(10): p. 3262-3269.
229. Guedon, E., M. Desvaux, and H. Petitdemange, *Kinetic analysis of Clostridium cellulolyticum carbohydrate metabolism: importance of glucose 1-phosphate and glucose 6-phosphate branch points for distribution of carbon fluxes inside and outside cells as revealed by steady-state continuous culture.* J Bacteriol, 2000. **182**(7): p. 2010-7.
230. Olson, D.G., et al., *Glycolysis without pyruvate kinase in Clostridium thermocellum.* Metab Eng, 2017. **39**: p. 169-180.
231. Deng, Y., et al., *Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in Clostridium thermocellum.* Metab Eng, 2013. **15**: p. 151-8.
232. Thompson, R.A. and C.T. Trinh, *Overflow Metabolism and Growth Cessation in Clostridium thermocellum DSM1313 during High Cellulose Loading Fermentations.* Biotechnol Bioeng, 2017: p. n/a-n/a.
233. Beri, D., et al., *Nicotinamide cofactor ratios in engineered strains of Clostridium thermocellum and Thermoanaerobacterium saccharolyticum.* FEMS Microbiol Lett, 2016. **363**(11).
234. Chung, D., et al., *Methylation by a Unique α-class N4-Cytosine Methyltransferase Is Required for DNA Transformation of Caldicellulosiruptor bescii DSM6725.* PLOS ONE, 2012. **7**(8): p. e43844.
235. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
236. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update.* Current Opinion in Biotechnology, 2005. **16**(5): p. 577-583.
237. Lynd, L.R., *OVERVIEW AND EVALUATION OF FUEL ETHANOL FROM CELLULOSIC BIOMASS: Technology, Economics, the Environment, and Policy.* Annual Review of Energy and the Environment, 1996. **21**(1): p. 403-465.
238. Lynd, L.R., et al., *Microbial Cellulose Utilization: Fundamentals and Biotechnology.* Microbiology and Molecular Biology Reviews, 2002. **66**(3): p. 506-577.
239. Linville, J.L., et al., *Industrial robustness: understanding the mechanism of tolerance for the Populus hydrolysate-tolerant mutant strain of Clostridium thermocellum.* PLoS ONE, 2013. **8**.
240. Waligora, A.-J., et al., *Clostridium difficile Cell Attachment Is Modified by Environmental Factors.* Applied and Environmental Microbiology, 1999. **65**(9): p. 4234-4238.
241. Balsalobre, C., et al., *Alterations in protein expression caused by the hha mutation in Escherichia coli: influence of growth medium osmolarity.* J Bacteriol, 1999. **181**(10): p. 3018-24.
242. Kanesaki, Y., et al., *Salt stress and hyperosmotic stress regulate the expression of different sets of genes in Synechocystis sp. PCC 6803.* Biochem Biophys Res Commun, 2002. **290**(1): p. 339-48.
243. Wood, J.M., *Bacterial responses to osmotic challenges.* The Journal of General Physiology, 2015.
244. Wood, J.M., *Osmosensing by Bacteria: Signals and Membrane-Based Sensors.* Microbiology and Molecular Biology Reviews, 1999. **63**(1): p. 230-262.
245. Hagemann, M., et al., *Flavodoxin accumulation contributes to enhanced cyclic electron flow around photosystem I in salt-stressed cells of Synechocystis sp. strain PCC 6803.* Physiologia Plantarum, 1999. **105**(4): p. 670-678.
246. Roberts, M.F., *Organic compatible solutes of halotolerant and halophilic microorganisms.* Saline Systems, 2005. **1**: p. 5-5.
247. Los, D.A. and N. Murata, *Membrane fluidity and its roles in the perception of environmental signals.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2004. **1666**(1–2): p. 142-157.

248. Romantsov, T. and J.M. Wood, *Contributions of Membrane Lipids to Bacterial Cell Homeostasis upon Osmotic Challenge*, in *Biogenesis of Fatty Acids, Lipids and Membranes*, O. Geiger, Editor. 2016, Springer International Publishing: Cham. p. 1-22.
249. Lopez, C.S., et al., *Variations of the envelope composition of Bacillus subtilis during growth in hyperosmotic medium.* Curr Microbiol, 1998. **36**(1): p. 55-61.
250. Poger, D. and A.E. Mark, *A Ring to Rule Them All: The Effect of Cyclopropane Fatty Acids on the Fluidity of Lipid Bilayers.* The Journal of Physical Chemistry B, 2015. **119**(17): p. 5487-5495.
251. Ljunggren, M., et al., *A kinetic model for quantitative evaluation of the effect of hydrogen and osmolarity on hydrogen production by Caldicellulosiruptor saccharolyticus.* Biotechnology for Biofuels, 2011. **4**(1): p. 31.
252. van Niel, E.W., P.A. Claassen, and A.J. Stams, *Substrate and product inhibition of hydrogen production by the extreme thermophile, Caldicellulsiruptor saccharolyticus.* Biotechnol Bioeng, 2003. **81**.
253. Willquist, K., P.A.M. Claassen, and E.W.J. van Niel, *Evaluation of the influence of CO2 on hydrogen production by Caldicellulosiruptor saccharolyticus.* International Journal of Hydrogen Energy, 2009. **34**(11): p. 4718-4726.
254. Farmer, W.R. and J.C. Liao, *Improving lycopene production in Escherichia coli by engineering metabolic control.* Nat Biotechnol, 2000. **18**(5): p. 533-7.
255. Lian, J., R. Chao, and H. Zhao, *Metabolic engineering of a Saccharomyces cerevisiae strain capable of simultaneously utilizing glucose and galactose to produce enantiopure (2R,3R)-butanediol.* Metabolic Engineering, 2014. **23**: p. 92-99.
256. Toyoda, K. and M. Inui, *Regulons of global transcription factors in Corynebacterium glutamicum.* Applied Microbiology and Biotechnology, 2015. **100**(1): p. 45-60.
257. VanFossen, A.L., et al., *Carbohydrate Utilization Patterns for the Extremely Thermophilic Bacterium Caldicellulosiruptor saccharolyticus Reveal Broad Growth Substrate Preferences.* Applied and Environmental Microbiology, 2009. **75**(24): p. 7718-7724.
258. Munch, R., *Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes.* Bioinformatics, 2005. **21**(22): p. 4187 - 4189.
259. Novichkov, P.S., et al., *RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in bacteria.* BMC Genomics, 2013. **14**(1): p. 1-12.
260. Schujman, G.E., et al., *Structural basis of lipid biosynthesis regulation in Gram-positive bacteria.* Embo j, 2006. **25**(17): p. 4074-83.
261. Liu, D., et al., *Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor–Actuator.* ACS Synthetic Biology, 2015. **4**(2): p. 132-140.
262. Schujman, G.E., et al., *FapR, a Bacterial Transcription Factor Involved in Global Regulation of Membrane Lipid Biosynthesis.* Developmental Cell, 2003. **4**(5): p. 663-672.
263. Guillot, A., D. Obis, and M.-Y. Mistou, *Fatty acid membrane composition and activation of glycine-betaine transport in Lactococcus lactis subjected to osmotic stress.* International Journal of Food Microbiology, 2000. **55**(1): p. 47-51.
264. Ramseier, T.M., et al., *In Vitro Binding of the Pleiotropic Transcriptional Regulatory Protein, FruR, to the fru, pps, ace, pts and icd Operons of Escherichia coli and Salmonella typhimurium.* Journal of Molecular Biology, 1993. **234**(1): p. 28-44.
265. Ramseier, T.M., et al., *The global regulatory protein FruR modulates the direction of carbon flow in Escherichia coli.* Molecular Microbiology, 1995. **16**(6): p. 1157-1169.
266. Saier, J.M.H., *Cyclic AMP-independent catabolite repression in bacteria.* FEMS Microbiology Letters, 1996. **138**(2-3): p. 97-103.
267. Ryu, S., et al., *Effect of the FruR Regulator on Transcription of the pts Operon in Escherichia coli.* Journal of Biological Chemistry, 1995. **270**(6): p. 2489-2496.

268. Chavarría, M., et al., *Fructose 1-phosphate is the one and only physiological effector of the Cra (FruR) regulator of Pseudomonas putida.* FEBS Open Bio, 2014. **4**(1): p. 377-386.
269. Bledig, S.A., T.M. Ramseier, and M.H. Saier, *FruR mediates catabolite activation of pyruvate kinase (pykF) gene expression in Escherichia coli.* Journal of Bacteriology, 1996. **178**(1): p. 280-3.
270. Cortay, J.C., et al., *In vitro asymmetric binding of the pleiotropic regulatory protein, FruR, to the ace operator controlling glyoxylate shunt enzyme synthesis.* Journal of Biological Chemistry, 1994. **269**(21): p. 14885-14891.
271. Barrière, C., et al., *Fructose Utilization in Lactococcus lactis as a Model for Low-GC Gram-Positive Bacteria: Its Regulator, Signal, and DNA-Binding Site.* Journal of Bacteriology, 2005. **187**(11): p. 3752-3761.
272. Rodionov, D., et al., *Transcriptional regulation of the carbohydrate utilization network in Thermotoga maritima.* Frontiers in Microbiology, 2013. **4**(244).
273. Chin, A.M., B.U. Feucht, and M.H. Saier, *Evidence for regulation of gluconeogenesis by the fructose phosphotransferase system in Salmonella typhimurium.* Journal of Bacteriology, 1987. **169**(2): p. 897-899.
274. Gaurivaud, P., et al., *Characterization of FruR as a putative activator of the fructose operon of Spiroplasma citri.* FEMS Microbiology Letters, 2001. **198**(1): p. 73-78.
275. De Las Peñas, A., L. Connolly, and C.A. Gross, *SigmaE is an essential sigma factor in Escherichia coli.* Journal of Bacteriology, 1997. **179**(21): p. 6862-6864.
276. Lee, J.-H., D.E. Geiman, and W.R. Bishai, *Role of Stress Response Sigma Factor SigG in Mycobacterium tuberculosis.* Journal of Bacteriology, 2008. **190**(3): p. 1128-1133.
277. Sander, K.B., et al., *Rex in Caldicellulosiruptor bescii: Novel Regulon Members and its Effect on the Production of Ethanol and Overflow Metabolites.* Submitted to Applied and Environmenal Microbiology, 2017.
278. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biology, 2010. **11**(10): p. R106.
279. Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-2120.
280. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature methods, 2012. **9**(4): p. 357-359.
281. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.
282. Münch, R., et al., *Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes.* Bioinformatics, 2005. **21**(22): p. 4187-4189.

Appendix

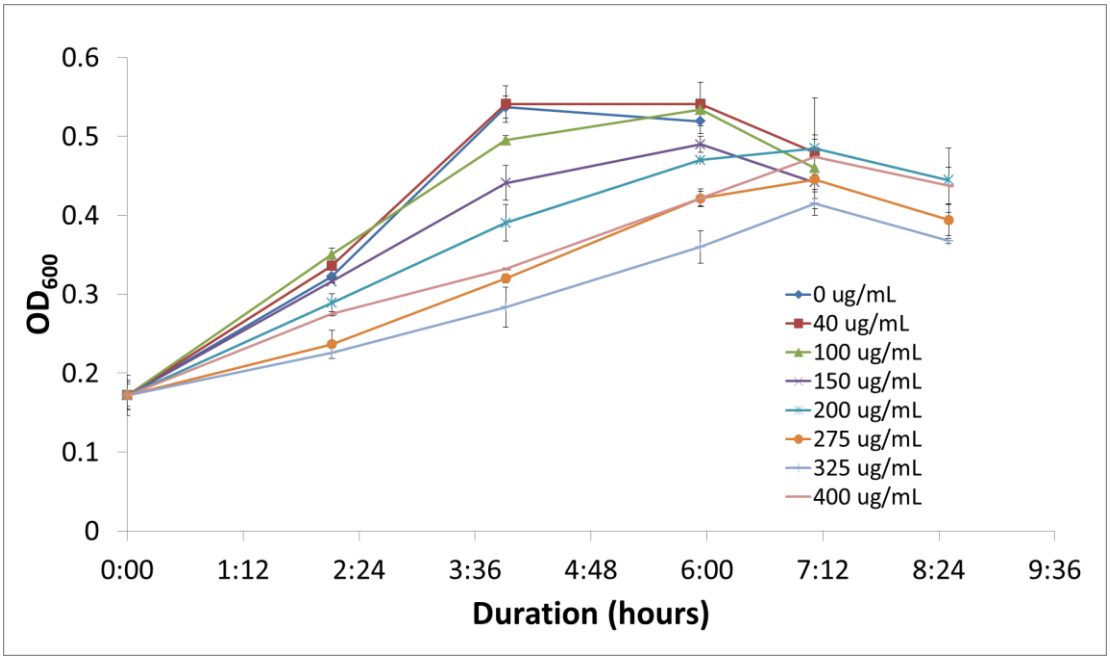


Figure A1. Batch fermentation performance under methyl viologen and hydrogen peroxide initial loadings. Adjusted OD_{600} of batch cultures grown at various initial methyl viologen concentrations. OD_{600} values were adjusted to account for initial blue coloration due to methyl viologen reduction. Cultures were grown in MTC media containing 1.1 g/L cellobiose.

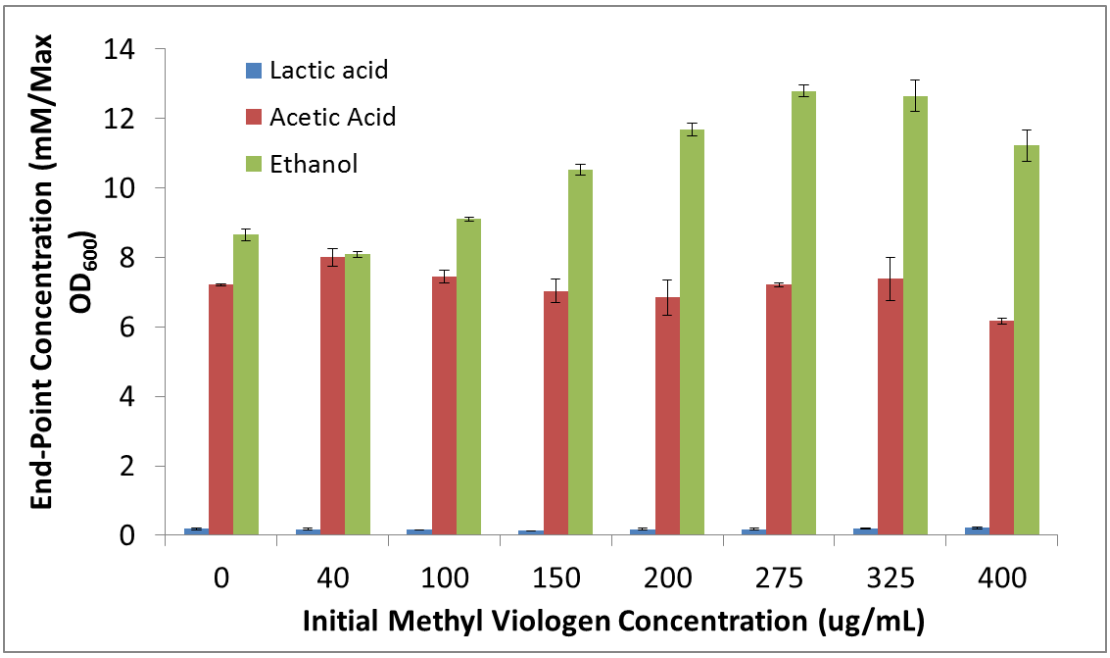


Figure A2. Specific end-point fermentation products at different initial methyl viologen concentrations, normalized to maximum OD_{600} achieved. Cultures were grown in MTC media containing 1.1 g/L cellobiose.

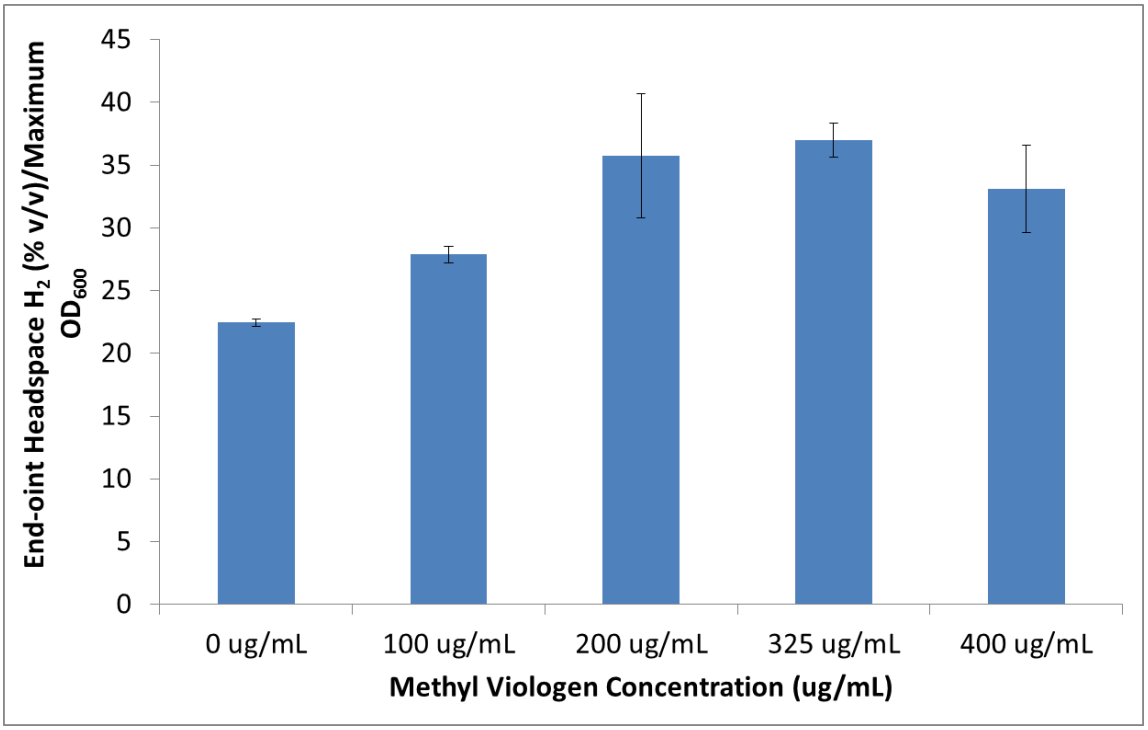


Figure A3. Specific end-point headspace H_2 at different initial methyl viologen concentrations, normalized to maximum OD_{600} achieved. Cultures were grown in MTC media containing 1.1 g/L cellobiose. Initial headspace % (v/v) H_2 was 5%.

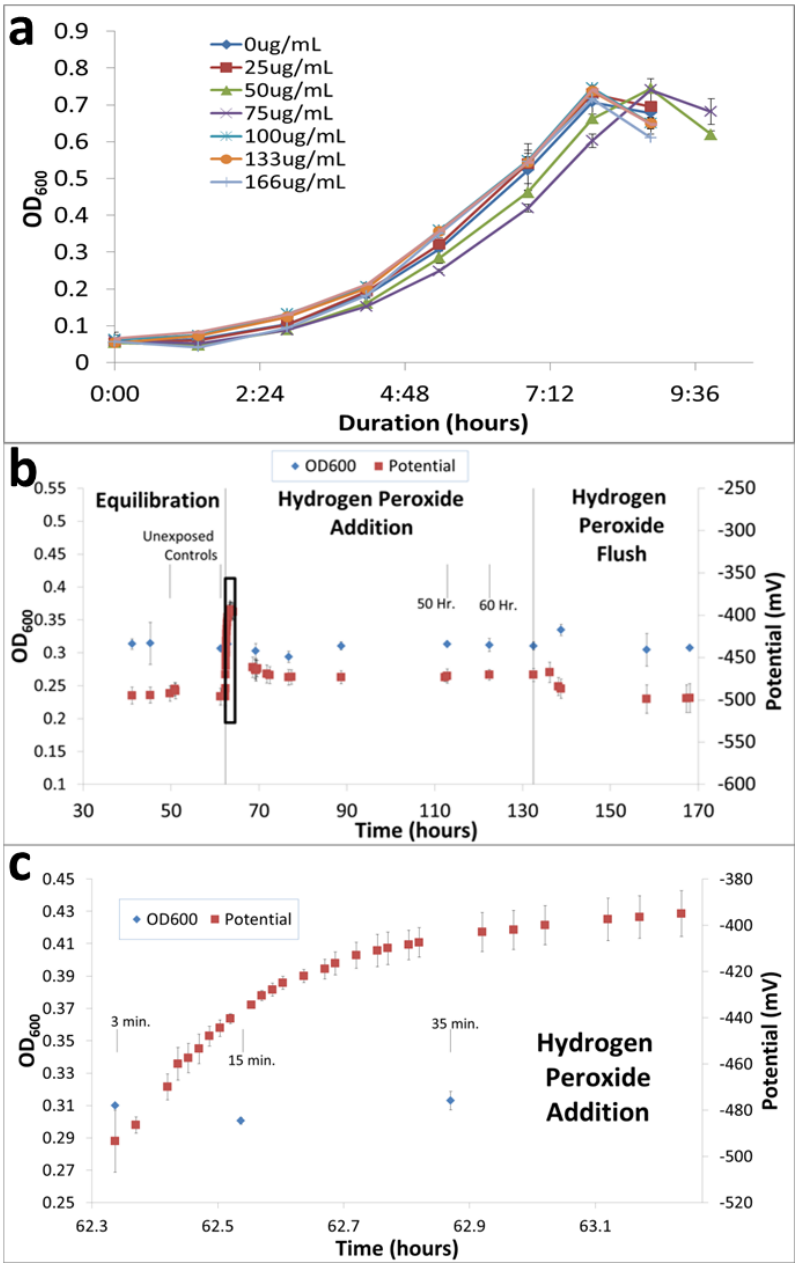


Figure A4. (A) Adjusted OD_{600} of batch cultures grown at various initial hydrogen peroxide concentrations. Cultures were grown in MTC media containing 1.1 g/L cellobiose; (B) Chemostat OD_{600} and measured redox potential before, during and after hydrogen peroxide addition; (C) Detailed view of boxed region indicated in panel (B).

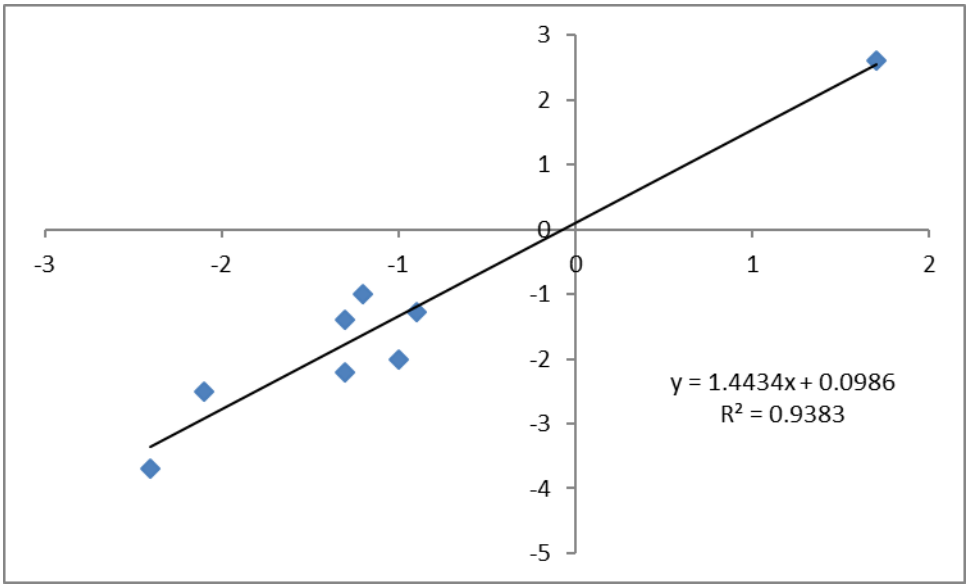


Figure A5. qPCR validation of expression differences of selected genes of interest.

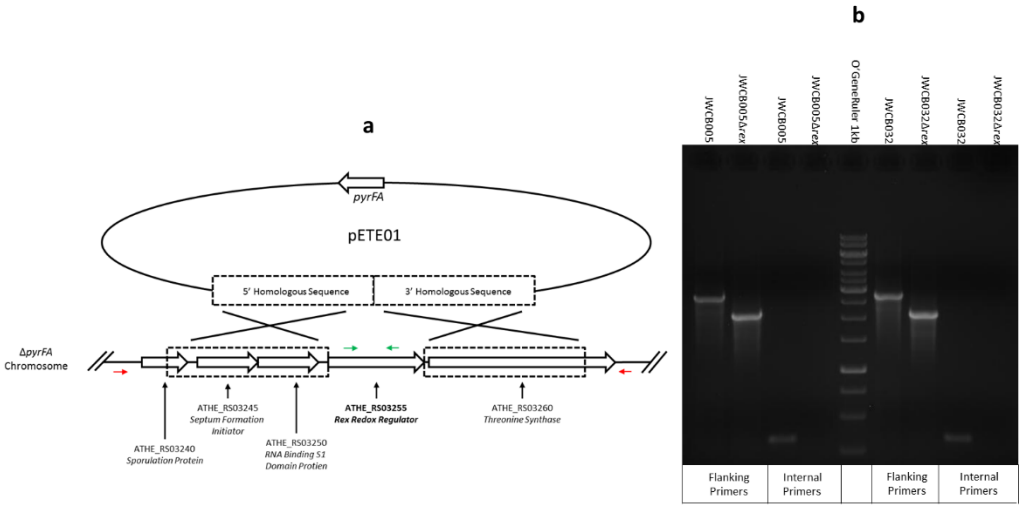


Figure A6. (a) Chromosomal integration and recombination scheme for constructing Δ*rex* mutants of JWCB005 and JWBCB032. Red arrows distal to homology region indicate primer binding sites for flanking primers used to screen knockout strains. Green arrows inside ATHE_RS03255 coding DNA sequence indicate primers binding internal to *rex* gene used to screen for the presence of *rex* coding sequence DNA. (b) PCR confirmation of Δ*rex* in strains JWCB005 and JWCB032 using primers which anneal within the deleted CDS (internal primers) and primers which anneal outside of the 5' and 3' flanking regions of homology (flanking primers). Expected *rex* wt amplicon using flanking primers is 2825 bp and expected Δ*rex* amplicon using flanking primers is 2171 bp. Expected amplicon length of PCR product using internal primers is 348 bp.

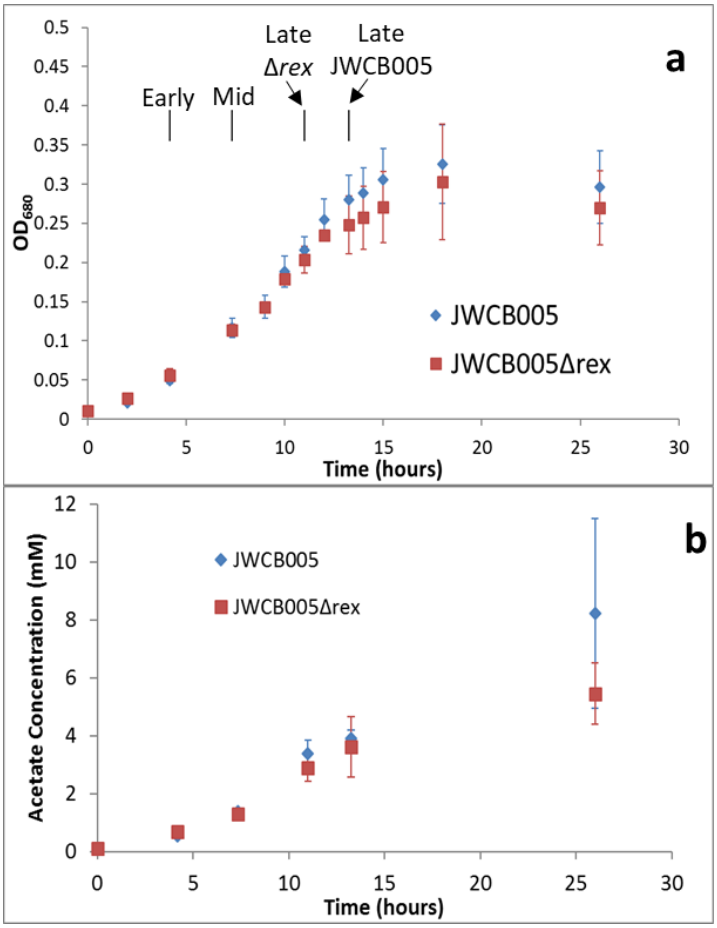


Figure A7. (a) Growth profile and (b) supernatant acetate concentration of JWCB005Δ*rex* strain and parent strain JWCB005. Labels indicate when samples were collected for expression profiling. No lactate was detected during growth.

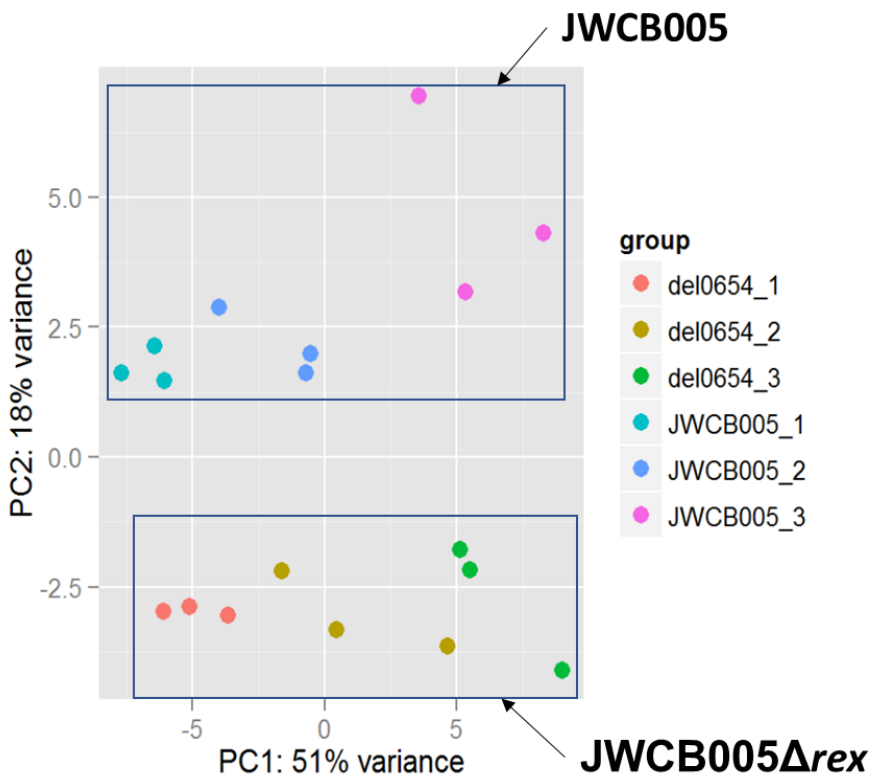


Figure A8. Principal component analysis of normalized mapped RNA-seq read-counts from strains. Replicates of JWCB005 are labeled as 'JWCB005_X,' where X is the replicate number and replicates of JWCB005Δ*rex* are labeled as 'del0654_X' in the same manner.

Figure A9. Electromobility shift assay using ATHE_RS03255 probe containing a mutated Rex binding site showing *in vitro* EMSA assay is sequence specific for Rex operator sites.

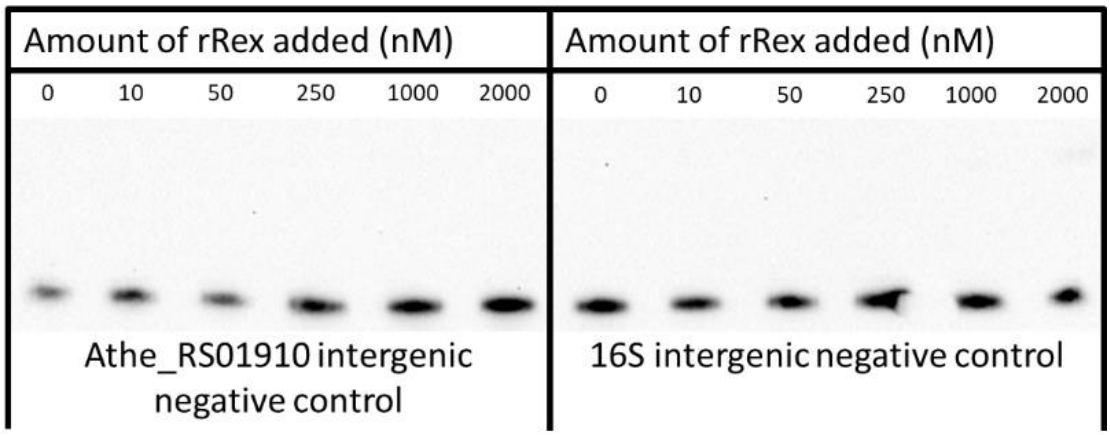


Figure A10. Electromobility shift assay reactions using probes containing non-homologous sequence to Rex binding sites showing Rex binding is sequence specific for previously predicted operator sites.

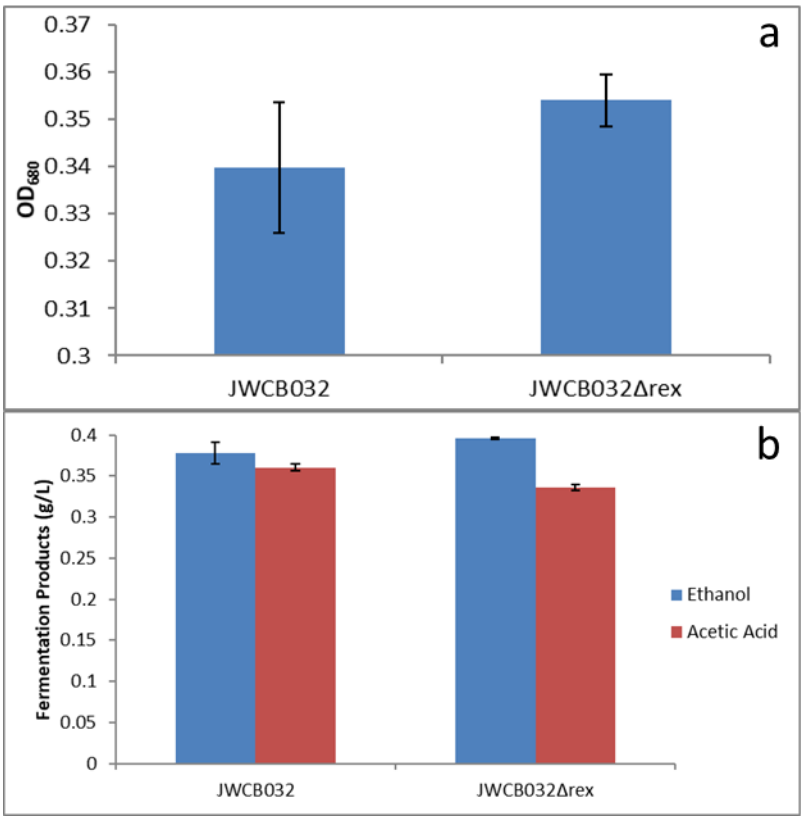
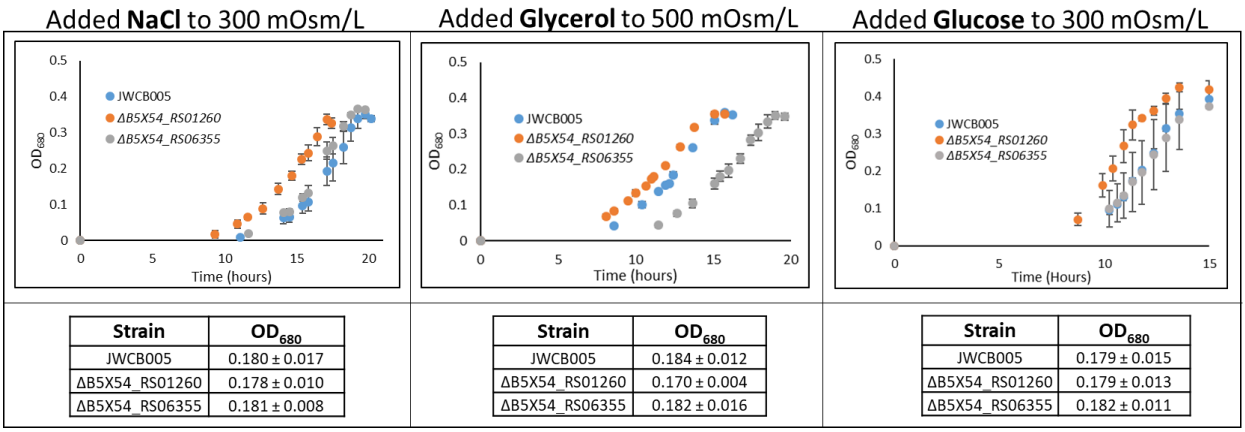


Figure A11. (a) OD_{680} observed after 48 hours of growth in replete LOD media. (b) Fermentation products observed after 48 hours of growth in replete LOD media.



Strain	OD_{680}
JWCB005	0.180 ± 0.017
ΔB5X54_RS01260	0.178 ± 0.010
ΔB5X54_RS06355	0.181 ± 0.008

Strain	OD_{680}
JWCB005	0.184 ± 0.012
ΔB5X54_RS01260	0.170 ± 0.004
ΔB5X54_RS06355	0.182 ± 0.016

Strain	OD_{680}
JWCB005	0.179 ± 0.015
ΔB5X54_RS01260	0.179 ± 0.013
ΔB5X54_RS06355	0.182 ± 0.011

Figure A12. Growth profiles and OD_{680} of cultures when samples were collected for RNAseq analysis for strains JWCB005 ΔB5X54_RS01260, JWCB005 ΔB5X54_RS06355, and JWCB005 (genetic parent strain to both single-gene deletion strains).

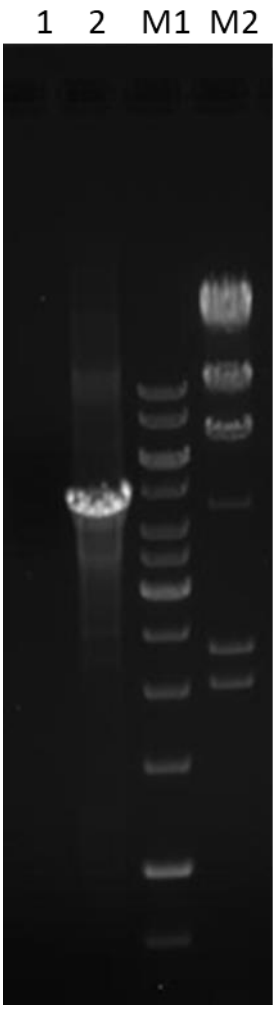


Figure A13. PCR of the genomic region spanning B5X54_RS07485 - B5X54_RS07500 failed to yield a PCR product when using genomic DNA from strain JWCB005 Δ B5X54_RS01260 and primers designed based on the JWCB005 genome, while yielding a product of expected size when using genomic DNA from strain JWCB018 (a derivative of strain JWCB005) as template for PCR. Expected size for PCR product based on JWCB005 genome is 4841 bp. Lane 1 is PCR reaction using JWCB005 ΔB5X54_RS01260 genomic DNA as template. Lane 2 is PCR reaction using JWCB018 genomic DNA as template. M1 is O'gene 10 kb ladder (ThermoFisher Scientific catalog # SM1163). M2 is Lambda DNA/HindIII Marker (ThermoFisher Scientific catalog # SM0102).

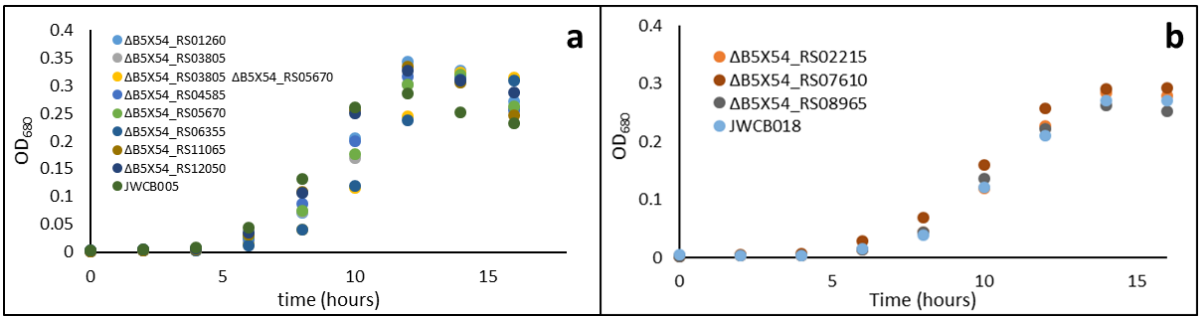


Figure A14. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L xylose as the primary source of carbon. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

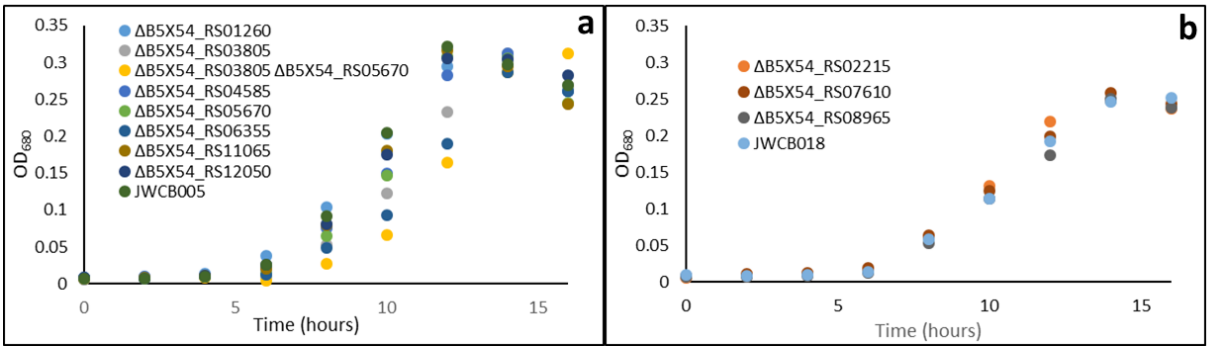


Figure A15. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L glucose as the primary source of carbon. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

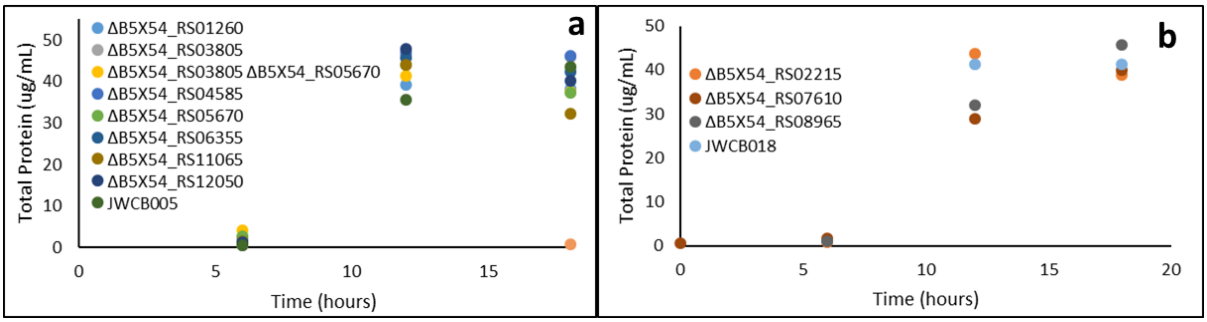


Figure A16. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L xylan as the primary source of carbon. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

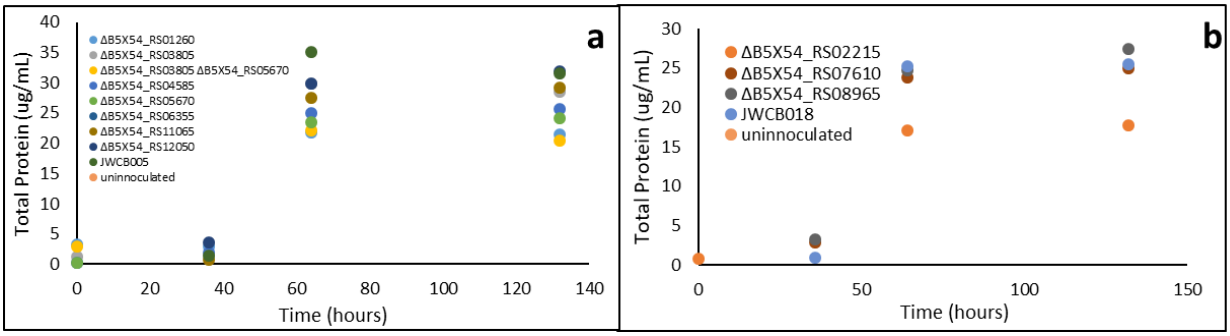


Figure A17. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L crystalline cellulose as the primary source of carbon. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

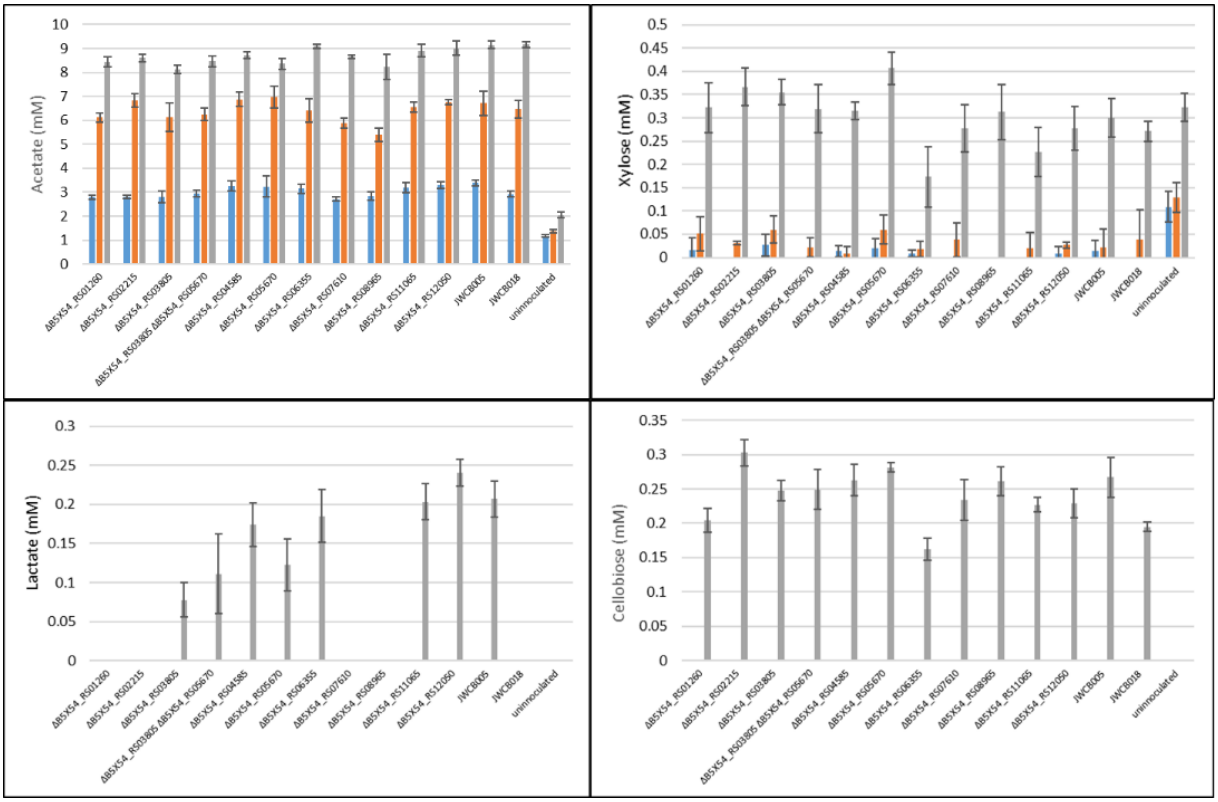


Figure A18. Single-gene deletion mutant screen for fermentation product phenotypes in media containing 5 g/L washed, unpretreated cellulose as the primary source of carbon. Supernatant fermentation product concentrations were collected and assayed after 24 hours (blue bars), 48 hours (orange bars), and 120 hours (grey bars) of fermentation.

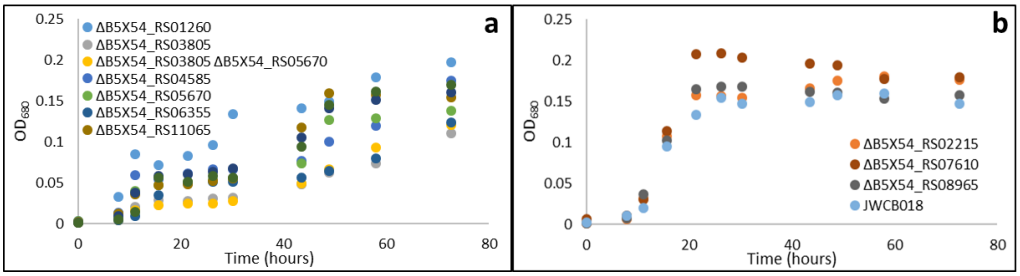


Figure A19. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L glucose as the primary source of carbon and 150 mg/L added methyl viologen. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

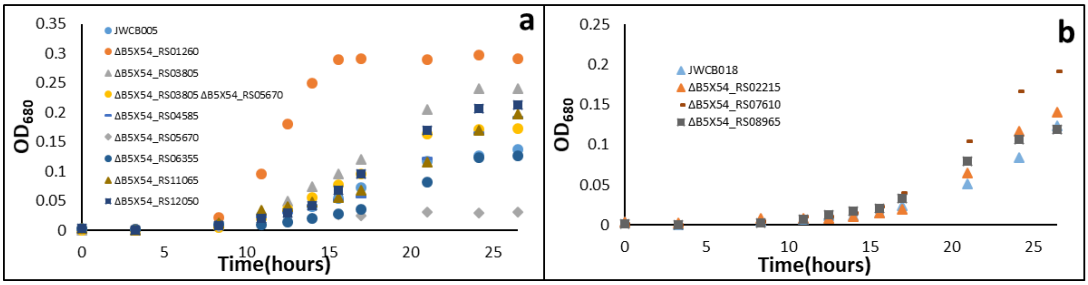


Figure A20. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L glucose as the primary source of carbon and added NaCl to an initial calculated media osmolarity of 200 mOsm/L. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

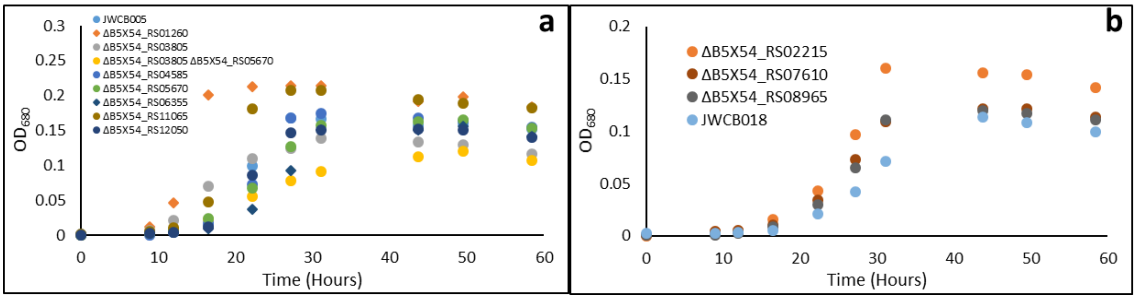


Figure A21. Single-gene deletion mutant screen for growth phenotypes in media containing 5 g/L glucose as the primary source of carbon and 20 g/L added ethanol. (a) denotes strains generated using JWCB005 as the genetic parent strain while (b) denotes strains generated using JWCB018 as the genetic parent strain.

Table A1. Primers used in qPCR validation of gene expression differences of selected genes.

Locus Tag	*C. thermocellum* DSM 1313 Annotations	F Primer Sequence	R Primer Sequence
Cthe_0197	glutamine amidotransferase class-II	tgacggtgatgtggctaaag	tcccttggtctttttcgttg
Cthe_1539	glutamine synthetase catalytic region	ctcattttcctggtggctgt	ggctccaagtctgtggtcat
Cthe_2524	IscR	CCGAATCATACCTGGAGCAG	GAAGGGTCGTCTGCAAGAAC
Cthe_3016	(NiFe) hydrogenase maturation protein HypF	tccctatgacaggcacaaca	atatggtttgggcgtgaaag
Cthe_3019	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	cgatggagcataaaccgact	aacgacggagcaggatacat

Table A2. These predicted operator sites were taken as the complete list of all possible Rex binding sites in the *C. bescii* genome. Additional evidence from Rex binding site predictions in *C. saccharolyticus* and differential expression were used to narrow down this list to likely putative Rex operator sites in *C. bescii*. * indicates locus tags not represented in the most recent RefSeq reannotation of the *C. bescii* genome. Locus tags from the previous generation annotation are provided instead.

Locus	Annotation	Strand	Begin	End	Binding Site Sequence	Length	Score	Distance to ATG
ATHE_RS00445	hypothetical protein	+	112399	112761	ATGTATAAAATCTAACTT	18	9	72
ATHE_RS00820	glycosyl transferase family 2	-	206185	207282	TTGTTAATATTATAACAT	18	10	209
ATHE_RS00825	CopG family transcriptional regulator	+	207653	207898	ATGTTATAATATTAACAA	18	10.5	181
ATHE_RS00970	AbrB family transcriptional regulator	-	244120	244389	AAGTATTTATAATAACAT	18	9	106
ATHE_RS01085	hypothetical protein	+	267966	268904	TTGTAGACTATTTTACAA	18	8.75	273
ATHE_RS01090	hypothetical protein	-	269094	269807	TTGTTATGTAAATAACTT	18	9	95
Athe_0262*	#N/A	-	312811	313146	ATGTTTTAAATATTACCT	18	8.75	41
ATHE_RS01280	hypothetical protein	-	313300	314661	TTGTATATAAATTACCAT	18	9.25	129
ATHE_RS01695	hypothetical protein	+	403042	403926	TTGAAAATAATATAACAG	18	9	196
ATHE_RS02010	hypothetical protein	+	469694	470419	TTGAATTTAAAATATCAA	18	8.75	76
ATHE_RS02075	glyceraldehyde 3-phosphate reductase	-	480265	481251	TTGTTGAGATTTTAACTA	18	9	275
ATHE_RS02085	hypothetical protein	+	483441	483914	TTGCAATTATAATAACAA	18	9.5	281

Table A2 Continued

Locus	Annotation	Strand	Begin	End	Binding Site Sequence	Length	Score	Distance to ATG
ATHE_RS02420	hypothetical protein	-	560115	560873	TTGAATTTAAAATATCAA	18	8.75	74
ATHE_RS02730	hypothetical protein	-	621256	621510	TTGATTTTATTATATCAA	18	8.75	177
ATHE_RS02735	glucokinase	+	621716	622666	TTGTTTTCATAATATCAA	18	9.25	178
ATHE_RS02955	sugar ABC transporter permease	+	671125	672066	TTGAAAAAGTTTTAACAA	18	9	230
ATHE_RS02990	excisionase	-	680874	681317	TTGTTTATAATATAACTA	18	10	59
ATHE_RS03055	hypothetical protein	-	696222	696917	TTGTATTTTTATTAGCTT	18	8.75	159
ATHE_RS03075	xylose repressor	+	700500	701699	TTGCTTTAGTATTAACAA	18	9	290
ATHE_RS03110	preprotein translocase subunit SecD	+	709214	710461	TTGTTGACTTTTTATCAA	18	8.75	119
ATHE_RS03255	transcriptional regulator	+	737116	737769	ATGTCAAAAAAATATCAA	18	9.25	41
ATHE_RS03280	transcription elongation factor GreA	+	742642	743130	TTGATTTATAATTGACAA	18	8.75	115
ATHE_RS03355	hypothetical protein	+	760255	761088	TGGTTTTAATACTAACAA	18	9	196
ATHE_RS03475	sorbitol-6-phosphate dehydrogenase	-	782881	783660	ATGTTAAAACAATATCAA	18	9.25	36
ATHE_RS03785	pyrophosphatase	+	844362	846497	TTGTCAAATTATTGACAA	18	9.25	49
ATHE_RS03830	peptidase M16	+	854249	855523	TTGTGAGATTTATAACAA	18	9.5	189
ATHE_RS04100	hypothetical protein	-	914722	915471	TCGTTTAAAATTTAACAA	18	9.5	148

Table A2 Continued

Locus	Annotation	Strand	Begin	End	Binding Site Sequence	Length	Score	Distance to ATG
ATHE_RS04125	tungsten ABC transporter permease	+	919029	919847	TTGTAAATCTTCTGACAA	18	8.75	274
ATHE_RS04390	pyruvate synthase	+	973763	974338	ATGTTGATTTTTTAACAT	18	9.5	112
ATHE_RS04625	hypothetical protein	+	1025101	1026303	TTGATAATTTTATAACAA	18	9.5	42
ATHE_RS04720	hypothetical protein	+	1044772	1045617	CTGTAAAAAATTTAACAA	18	10	40
ATHE_RS04760	hypothetical protein	-	1053958	1054299	TTGTAATTTTTATACCAT	18	9.25	206
ATHE_RS04765	hypothetical protein	+	1054588	1055535	ATGGTATAAAAATTACAA	18	8.75	102
ATHE_RS04830	hypothetical protein	-	1070216	1070944	ATGTATTAGAAATTACAC	18	8.75	103
ATHE_RS05410	hydrolase Cof	-	1175967	1176821	AAGTTAATTTGTTAACAA	18	9	84
ATHE_RS05415	NADH dehydrogenase	+	1177033	1178907	TTGTTAACAAATTAACTT	18	9	147
ATHE_RS06125	methyltransferase	+	1316442	1316996	GTGTTTAAAATTTACCAA	18	9.25	115
ATHE_RS06335	30S ribosomal protein S21	+	1359037	1359213	AGGTTAATAATTTAACAC	18	9	287
ATHE_RS06475	NADH dehydrogenase	+	1385700	1386182	ATGTTAAATTTCTAACAA	18	10	39
ATHE_RS06655	NAD kinase	+	1419393	1420178	ATGTATTTGTTGTAGCAA	18	8.75	204
ATHE_RS07090	hypothetical protein	+	1504465	1504998	ATGTTTTTATAGTAGCAA	18	9.25	115
ATHE_RS07400	hypothetical protein	+	1565517	1566143	TTGTTTCTATTATACCA	18	8.75	243

Table A2 Continued

Locus	Annotation	Strand	Begin	End	Binding Site Sequence	Length	Score	Distance to ATG
ATHE_RS08595	histidinol phosphate phosphatase	+	1801188	1801988	ATGTCAATAAAATATCAT	18	8.75	67
ATHE_RS08810	MFS transporter	+	1827609	1828790	TTGATAAATAAATAACCA	18	9	216
ATHE_RS10330	histidine kinase	-	2171070	2172833	ATGTTATTATGTCACAA	18	9.25	231
ATHE_RS10335	alpha-glucosidase	-	2172903	2175221	TTGTATTGATTATAGCAA	18	9.25	155
ATHE_RS10680	class V aminotransferase	-	2260691	2261851	TTGTCAAATTATTGACAT	18	8.75	88
ATHE_RS10870	flagellar hook protein FlgE	-	2293272	2294738	ATGATTTAAAATTGACAA	18	8.75	135
ATHE_RS11210	2-isopropylmalate synthase	+	2359000	2360358	ATGTTATCACAATATCAA	18	8.75	169
ATHE_RS11390	phosphohydrolase	-	2398473	2399756	TTGCATTTAATTTTACAA	18	8.75	178
Athe_2305*	#N/A	-	2441838	2442488	TTGTTAAAAAATTGACCA	18	9.25	73
ATHE_RS11720	malate transporter	-	2465100	2466050	ATGCAAAAAATTTGACAA	18	8.75	43
ATHE_RS11890	hypothetical protein	-	2505938	2507032	TTGTTTAATAATTTACTG	18	8.75	167
ATHE_RS11905	membrane protein	+	2510839	2511846	TTGTAAAGTATTTCACAC	18	8.75	147
ATHE_RS11980	sugar ABC transporter substrate-binding protein	-	2529621	2531081	ATGTATTGTTTTTATCAT	18	8.75	98
ATHE_RS12440	XRE family transcriptional regulator	-	2605848	2606216	TTGCTTTTTTGTTAACAA	18	9	70
ATHE_RS12590	hypothetical protein	-	2627991	2628185	ATGTTGATATTATAGCAG	18	8.75	128

Table A2 Continued

Locus	Annotation	Strand	Begin	End	Binding Site Sequence	Length	Score	Distance to ATG
ATHE_RS12835	AbrB family transcriptional regulator	+	2663972	2664253	TTGCATTTAATTTTACAA	18	8.75	98
ATHE_RS13540	CRISPR-associated protein	-	2822434	2823759	ATGTTATATTGTTAACAA	18	10	167
ATHE_RS13740	hypothetical protein	+	2869336	2871690	TTGAAATTAATATAACTA	18	9	65

Table A3. Primers used in this study

Primer Name	Primer Sequence	Purpose
upstm 5' flank fwd athe_1015_P3	TGGGCTACTCCTTTGAGGAA	genomic screening primer flanking the B5X54_RS05670 gene
dnstm 3' flank rev athe_1015_P3	CAATTGAATTTCCCGACCTG	genomic screening primer flanking the B5X54_RS05670 gene
internal rev athe_1015_P3	GCTCCTGGTTGAAACCAAAC	genomic screening primer binding internal to the B5X54_RS05670 gene
upstm 5' flank fwd athe_1669_P3	GGGATATCGCTATAAGATGAAGG	genomic screening primer flanking the B5X54_RS08965 gene
dnstm 3' flank rev athe_1669_P3	TTTCGGAATATAATTGTGCATGA	genomic screening primer flanking the B5X54_RS08965 gene
internal rev athe_1669_P3	TTCACGAAAAGAAGGAGACCA	genomic screening primer binding internal to the B5X54_RS08965 gene
upstm 5' flank fwd athe_1152_P3	CGTCAAAAAGAGCCAAGAGG	genomic screening primer flanking the B5X54_RS06355 gene
dnstm 3' flank rev athe_1152_P3	TCAGGGGGTGAAAATGGTAA	genomic screening primer flanking the B5X54_RS06355 gene
internal rev athe_1152_P3	AAATTTGCCCCTGAAAACCT	genomic screening primer binding internal to the B5X54_RS06355 gene
upstm 5' flank fwd athe_2264_P3	GCATAACTCCTAATGAATGTCCAA	genomic screening primer flanking the B5X54_RS12050 gene
dnstm 3' flank rev athe_2264_P3	GAGGTGCTGCGTATGTTCTG	genomic screening primer flanking the B5X54_RS12050 gene
internal rev athe_2264_P3	GCTGCAGCAGACCTGTACG	genomic screening primer binding internal to the B5X54_RS12050 gene
upstm 5' flank fwd athe_0146_P3	ACCTGCCACCACAACTACAA	genomic screening primer flanking the B5X54_RS01260 gene
dnstm 3' flank rev athe_0146_P3	GCACCCTTTGATCCCATCG	genomic screening primer flanking the B5X54_RS01260 gene
internal rev athe_0146_P3	CAAGCCTTTCCACAGGACAT	genomic screening primer binding internal to the B5X54_RS01260 gene
upstm 5' flank fwd athe_0343_P3	GAGAGTTCTGTGGGGCAAAA	genomic screening primer flanking the B5X54_RS02215 gene
dnstm 3' flank rev athe_0343_P3	ATCTTGTGAACGCCAAGGTC	genomic screening primer flanking the B5X54_RS02215 gene
internal rev athe_0343_P3	GGCTGTGCAACTGTTGTGAT	genomic screening primer binding internal to the B5X54_RS02215 gene
upstm 5' flank fwd athe_0799_P3	CGGTGATGTAACCACTCCAA	genomic screening primer flanking the B5X54_RS04585 gene
dnstm 3' flank rev athe_0799_P3	TTCCAATACCGCTTTTAAATGCCT	genomic screening primer flanking the B5X54_RS04585 gene
internal rev athe_0799_P3	TAATCTTGACTGCCGTGCATC	genomic screening primer binding internal to the B5X54_RS04585 gene

Table A3 Continued

Primer Name	Primer Sequence	Purpose
dnstm 3' flank rev athe_2074_P3	AACACATGGGAGGAGACCAG	genomic screening primer flanking the B5X54_RS11065 gene
internal rev athe_2074_P3	GAAGTGAATGAGATTGCAAGAGGC	genomic screening primer binding internal to the B5X54_RS11065 gene
upstm 5' flank fwd athe_1407_P3	CTGCTGCTCTTGCTCTCCTT	genomic screening primer flanking the B5X54_RS07610 gene
dnstm 3' flank rev athe_1407_P3	CAAAGTTGGAACAATGATTGAGG	genomic screening primer flanking the B5X54_RS07610 gene
internal rev athe_1407_P3	GGGCAGGAAGTTGAAAAACA	genomic screening primer binding internal to the B5X54_RS07610 gene
pyr fwd screen	CTTGAAAATCCAGGGGTTGA	pyrF region to screen pyr biosyntheis region of gDNA of delAthe_0146 mutant
pyr rev screen	GGTCAGTTTTCCCTTGGACA	pyrF region to screen pyr biosyntheis region of gDNA of delAthe_0146 mutant
pJGW03 fwd insert check seq	Gccgcatctgagagtt	screening insert site on pJGW03-based integrating suicide vectors
pJGW03 rev insert check seq	Ctacggaaggagctgtg	screening insert site on pJGW03-based integrating suicide vectors
pDCW88 gib assy backbone fwd	Gtgcactctgacgctc	amplify pJGW03 based integrating suicide vector backbone
pDCW88 gib assy backbone rev	Ggtaccaccagcctaac	amplify pJGW03 based integrating suicide vector backbone
pDCW88_athe_1152_up_fwd	tccaatgatcgaagttaggctggtggtaccAATCTATAAGGTAAAGGACCTTGTTG	amplify upstream region of the B5X54_RS06355 gene
pDCW88_athe_1152_up_rev	AATCCTCATATTTTACTACACTCCCATTTCAATTTTATCACCATGATTTAATATTTGA	amplify upstream region of the B5X54_RS06355 gene
pDCW88_athe_1152_down_fwd	GAAATGGGAGTGTAGTAAAATATGAG	amplify downstream region of the B5X54_RS06355 gene
pDCW88_athe_1152_down_rev	Gttttcgttccactgagcgtcagagtgcacgccatctgagcaaactgc	amplify downstream region of the B5X54_RS06355 gene
pDCW88_athe_1015_up_fwd	tccaatgatcgaagttaggctggtggtaccATAAAATTTAAAGAGGCAGCTTTAGC	amplify upstream region of the B5X54_RS05670 gene
pDCW88_athe_1015_up_rev	AAGCGCAGATGGCGGTAAATTTTTGAAAAGatctttccacctcatttttctg	amplify upstream region of the B5X54_RS05670 gene
pDCW88_athe_1015_down_fwd	CTTTTCAAAAATTTACCGCC	amplify downstream region of the B5X54_RS05670 gene
pDCW88_athe_1015_down_rev	gttttcgttccactgagcgtcagagtgcacAATACCAGAGCCAACACCTG	amplify downstream region of the B5X54_RS05670 gene

Table A3 Continued

Primer Name	Primer Sequence	Purpose
pDCW88_athe_0343_up_rev	acaatggttttaaaataaaaaataacctTCTTTACACCCCCAGATACTTTTAC	amplify upstream region of the B5X54_RS02215 gene
pDCW88_athe_0343_down_fwd	GAAGGTTATTTTTTATTTTAAAACCATTG	amplify downstream region of the B5X54_RS02215 gene
pDCW88_athe_0343_down_rev	gttttcgttccactgagcgtcagagtgcacGCCTTTTTGTGGGCTATATTTAAC	amplify downstream region of the B5X54_RS02215 gene
pDCW88_athe_2074_up_fwd	tgatcgaagttaggctggtggtaccGAAAGATAAGACTTATTTTTCATTATAGAAACTG	amplify upstream region of the B5X54_RS11065 gene
pDCW88_athe_2074_up_rev	aactaataaaaggttattatacaaaagaggTTTTATTGGGTTTGAGATGATTTG	amplify upstream region of the B5X54_RS11065 gene
pDCW88_athe_2074_down_fwd	CCTCTTTTGTATAATAACCTTTTATTAGTTTA	amplify downstream region of the B5X54_RS11065 gene
pDCW88_athe_2074_down_rev	gttttcgttccactgagcgtcagagtgcacGGATTTAGTTCTACACCCAAAATAG	amplify downstream region of the B5X54_RS11065 gene
pDCW88_athe_2264_up_fwd	ccaatgatcgaagttaggctggtggtaccAAACTCATATAATTTTTCACTCATTTTTTC	amplify upstream region of the B5X54_RS12050 gene
pDCW88_athe_2264_up_rev	AAAATTATATTTGAGGAGGTTTGGTAGGCTatgaactgagatgtatcttaaaaagc	amplify upstream region of the B5X54_RS12050 gene
pDCW88_athe_2264_down_fwd	AGCCTACCAAACCTCCTCA	amplify downstream region of the B5X54_RS12050 gene
pDCW88_athe_2264_down_rev	gttttcgttccactgagcgtcagagtgcacTGTGTACTATGTTGGTAGTAACGAAA	amplify downstream region of the B5X54_RS12050 gene
pDCW88_athe_1407_up_fwd	tccaatgatcgaagttaggctggtggtaccCTCCTCATTTACACCCATAACA	amplify upstream region of the B5X54_RS07610 gene
pDCW88_athe_1407_up_rev	TAAAATTGTCCCGCTGGAGAATGAGAGAGTaaggtaagaggcaaaaacactg	amplify upstream region of the B5X54_RS07610 gene
pDCW88_athe_1407_down_fwd	ACTCTCTCATTCTCCAGCG	amplify downstream region of the B5X54_RS07610 gene
pDCW88_athe_1407_down_rev	gttttcgttccactgagcgtcagagtgcacATGACTTAACACAGATGACTTTTGG	amplify downstream region of the B5X54_RS07610 gene
pDCW88_athe_0799_up_fwd	tccaatgatcgaagttaggctggtggtaccAATTGGAGAAAAACTGTCTTCAAAG	amplify upstream region of the B5X54_RS04585 gene

Table A3 Continued

Primer Name	Primer Sequence	Purpose
pDCW88_athe_0799_down_fwd	AAAAGAGGAGTAATGAAAAATGGTC	amplify downstream region of the B5X54_RS04585 gene
pDCW88_athe_0799_down_rev	gttttcgttccactgagcgtcagagtgcacCTTCATATCTGCCACTTTCAAG	amplify downstream region of the B5X54_RS04585 gene
pDCW88_athe_1669_up_fwd	tccaatgatcgaagttaggctggtggtaccTTGTTCTAATGTAAGATCTGAAAACAATC	amplify upstream region of the B5X54_RS08965 gene
pDCW88_athe_1669_up_rev	tcagatgagcagaaatgaggtatataaaagGATGGTTAAAGATGAGATTGTAAAGG	amplify upstream region of the B5X54_RS08965 gene
pDCW88_athe_1669_down_fwd	CTTTTATATACCTCATTTCTGCTCATC	amplify downstream region of the B5X54_RS08965 gene
pDCW88_athe_1669_down_rev	tttcgttccactgagcgtcagagtgcacTATAAAGAGTTAAAAAGAGGAGATTAAAGAT	amplify downstream region of the B5X54_RS08965 gene
pDCW88_athe_0146_up_fwd	tccaatgatcgaagttaggctggtggtaccCGACTTTGTATACAACCCATCTTC	amplify upstream region of the B5X54_RS01260 gene
pDCW88_athe_0146_up_rev	caatttgatcccaccttgaaatttaaaattTTTATTCACCTTTCGTTTATGC	amplify upstream region of the B5X54_RS01260 gene
pDCW88_athe_0146_down_fwd	AATTTTAAATTTCAAGGTGGGATC	amplify downstream region of the B5X54_RS01260 gene
pDCW88_athe_0146_down_rev	gttttcgttccactgagcgtcagagtgcacTTTCGTTTTAGCATTCTTATAAGCTC	amplify downstream region of the B5X54_RS01260 gene
fwd amp_pDCW173::0146_overlap	GAGTATGAAAAACTTAGTGTGAAGGTCATCACGTAAgagggtgagattgattctcac	amplifying replicating shuttle vector backbone containing B5X54_RS01260 gene
rev amp_pDCW173::0146_overlap	CTGCGCAATTCTGCTTTTTCTTTCCTCTGCAAACATaactactcaccaaacctccttg	amplifying replicating shuttle vector backbone containing B5X54_RS01260 gene
fwd athe_0146::pDCW173 insert	aaatcatacaaggaggtttggtgagtagttATGTTTGCAGAGGAAAGAAAA	amplifying B5X54_RS01260 gene for expression insert
rev athe_0146::pDCW173 insert	taaaagagggtgagaatcaatctcacctcTTACGTGATGACCTTCACACTA	amplifying B5X54_RS01260 gene for expression insert
fwd amp_pDCW173::1152_overlap	GTTGCATTAAATGAAGAGACTAATTCAGCGCAATAGgagggtgagattgattctcac	amplifying replicating shuttle vector backbone containing B5X54_RS06355 gene
rev amp_pDCW173::1152_overlap	AAGTCTGTGTCTTTCTTTCCTTGATAGTTTTGCCATaactactcaccaaacctccttg	amplifying replicating shuttle vector backbone containing B5X54_RS06355 gene

Table A3 Continued

Primer Name	Primer Sequence	Purpose
rev athe_1152::pDCW173 insert	taaaagagggtgagaatcaatctcacc ctcCTATTGCGCTGAATTAGTCT CTT	amplifying B5X54_RS06355 gene for expression insert
Forward (Gblk_fwd) - Kan HT	ATGAAAGGACCTATAATTATG	amplifying high-temperature kanamycin marker for autonomously replicating expression vectors
Reverse (Gblk_rev) - Kan HT	TCAAAATGGTATTCTTTTG	amplifying high-temperature kanamycin marker for autonomously replicating expression vectors
gb_bbR_173iCbhtk	TCTTCTCTAGTCATAATTATAG GTCCTTTCATtctagagaccatcctt tctatg	amplifying expression vectors to insert kanamycin marker in place of existing pyrF gene
gb_bbF_07_cbhtk_2	AGTAGATGTTAGCAAAAGAAT ACCATTTTGAgatatgtaacggtga acagttg	amplifying expression vectors to insert kanamycin marker in place of existing pyrF gene
fwd_kaninsert_07_XbaI	AACTTTCTACATAGAAAGGAT GGTCTCTAGatgaaaggacctataa ttatgactagag	amplifying high-temperature kanamycin marker for autonomously replicating expression vectors
rev_kaninsert_07_NdeI	CGTTACATATCAAAGGGAAAA CTGTCCATAtcaaaatggtattctttt gctaacatc	amplifying high-temperature kanamycin marker for autonomously replicating expression vectors
fwd_mkrscreen_pJGW07	atccgttgatcttcctgcat	screening marker site on autonomously replicating expression vectors
rev_mkrscreen_pJGW07	ctcacgcaaaaacaacgaac	screening marker site on autonomously replicating expression vectors
fwd amp173_Mkr::Kan overlap	AGTAGATGTTAGCAAAAGAAT ACCATTTTGAcatcatcatcatcatc actaacc	amplifying expression vector backbones to insert high-temperature kanamycin marker
rev amp173_Mkr::Kan overlap	TCTTCTCTAGTCATAATTATAG GTCCTTTCATtctagagaccatcctt tctatg	amplifying expression vector backbones to insert high-temperature kanamycin marker

Table A4. Plasmids used in this study

Plasmid	Purpose
pB5X54_RS06355_del	Non-replicating vector used to delete the B5X54_RS06355 gene
pB5X54_RS05670_del	Non-replicating vector used to delete the B5X54_RS05670 gene
pB5X54_RS02215_del	Non-replicating vector used to delete the B5X54_RS02215 gene
pB5X54_RS11065_del	Non-replicating vector used to delete the B5X54_RS11065 gene
pB5X54_RS12050_del	Non-replicating vector used to delete the B5X54_RS12050 gene
pB5X54_RS07610_del	Non-replicating vector used to delete the B5X54_RS07610 gene
pB5X54_RS04585_del	Non-replicating vector used to delete the B5X54_RS04585 gene
pB5X54_RS08965_del	Non-replicating vector used to delete the B5X54_RS08965 gene
pB5X54_RS01260_del	Non-replicating vector used to delete the B5X54_RS01260 gene

Table A5. Strains used in this study

Strain	Genotype	Reference
JWCB005 ΔB5X54_RS06355	ΔB5X54_RS06355 Δ*pyrFA*	This Study
JWCB005 ΔB5X54_RS03805	ΔB5X54_RS03805 Δ*pyrFA*	This Study
JWCB005 ΔB5X54_RS05670	ΔB5X54_RS05670 Δ*pyrFA*	This Study
JWCB018 ΔB5X54_RS02215	ΔB5X54_RS02215 Δ*pyrFA* Δ*cbeI*	This Study
JWCB005 ΔB5X54_RS11065	ΔB5X54_RS11065 Δ*pyrFA*	This Study
JWCB005 ΔB5X54_RS12050	ΔB5X54_RS12050 Δ*pyrFA*	This Study
JWCB018 ΔB5X54_RS07610	ΔB5X54_RS07610 Δ*pyrFA* Δ*cbeI*	This Study
JWCB005 ΔB5X54_RS04585	ΔB5X54_RS04585 Δ*pyrFA*	This Study
JWCB018 ΔB5X54_RS08965	ΔB5X54_RS08965 Δ*pyrFA* Δ*cbeI*	This Study
JWCB005 ΔB5X54_RS01260	ΔB5X54_RS01260 Δ*pyrFA*	This Study
JWCB005	Δ*pyrFA*	[54]
JWCB018	Δ*pyrFA* Δ*cbeI*	[20]

Vita

The author Kyle Sander grew up in Beaverton, OR (a suburb of Portland, OR) and graduated from Oregon State University in 2007 with a B.S. in Chemical Engineering. He spent one year as an environmental engineering and process engineering co-op at Georgia-Pacific, LLC. In Toledo, OR. He returned to Oregon State University to pursue an M.S. in Biological and Ecological Engineering, awarded in 2010. The promise of being a part of a newly formed interdisciplinary graduate program (Bredesen Center for Interdisciplinary Research and Graduate Education), and the hope of conducting metabolic engineering research at Oak Ridge National Laboratory brought him to the University of Tennessee, Knoxville for PhD studies in the Department of Chemical and Biomolecular Engineering. Kyle will continue exploring the genomic and metabolic landscape of nature's organisms toward more effective and sustainable manufacturing of products using biological factories, better understanding of the earth and the way it's processes interact with biology, and exploring better and safer ways to interact with our own genetic and metabolic architecture, and that of the microorganisms that inhabit our bodies. Biology contains some of the most complicated systems man has ever happened upon and Kyle will be part of the continued effort to better understand and utilize them.