University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

5-2016

In silico Driven Metabolic Engineering Towards Enhancing Biofuel and Biochemical Production

Richard Adam Thompson
University of Tennessee - Knoxville, rthomp46@utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

 Part of the Biochemical and Biomolecular Engineering Commons, Biological Engineering Commons, Computational Biology Commons, Molecular Biology Commons, and the Systems Biology Commons

Recommended Citation

Thompson, Richard Adam, "*In silico* Driven Metabolic Engineering Towards Enhancing Biofuel and Biochemical Production. " PhD diss., University of Tennessee, 2016.
https://trace.tennessee.edu/utk_graddiss/3755

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Richard Adam Thompson entitled "*In silico* Driven Metabolic Engineering Towards Enhancing Biofuel and Biochemical Production." 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 Energy Science and Engineering.

Cong T. Trinh, Major Professor

We have read this dissertation and recommend its acceptance:

Eric T. Boder, Adam M. Guss, Michael L. Simpson

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

In silico Driven Metabolic Engineering Towards Enhancing Biofuel and Biochemical Production

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Richard Adam Thompson

May 2016

© by Richard Adam Thompson, 2016
All Rights Reserved.

Dedicated to my parents

Richard and Kim

and my grandparents

Luis, Eleanor, Jim, and Manon

Acknowledgements

I would first like to thank my adviser, Dr. Cong Trinh, who as a recently hired assistant professor took a chance on a student from a new, outside department, and whose guidance and enthusiasm throughout my training has been instrumental. The opportunity he provided allowed me to pursue interesting research topics as well as exposed me to a wide range of techniques and ideas.

I would like to thank my committee members, Eric Boder, Adam Guss, and Mike Simpson, who over the years have given unwavering support and insightful perspectives. I am immensely grateful for the Bredesen Center for inviting me to pursue my doctorate at the University of Tennessee, particularly Lee Riedinger for his dedication to the program and students, as well as Wanda Davis, Jessica Garner, and Ben Allen for their invaluable assistance. Additionally, I am very grateful for the Department of Chemical Engineering at UT and the BioEnergy Science Center at ORNL for their continued support of my work.

Special acknowledgments are required for Donovan Layton and Kyle Sander, whose bench-top and bar-top conversations led to countless, mutually beneficial troubleshooting successes as well as enabled the articulation of some of the ideas presented here. I would also like to thank fellow Trinh lab compatriots for their insights, assistance, and friendship, in particular D.J. Conner, Sergio Garcia, Julie Hipp, Drew Kirkpatrick, Brian Mendoza, Dr. Paulo Avilo Neto, Dr. Narayan Niraula, Dr. Seunghyun Ryu, Kevin Spellman, Caleb Walker, Brandon Wilbanks, and Akshitha Yarrabothoula. I would also like to acknowledge many members of

BESC, in particular Dr. Steve Brown, Dr. Brian Davison, Joe Groom, Dr. Evert Holwerda, Prof. Lee Lynd, Dr. Intawat Nookaew, Dr. Dan Olson, Beth Papanek, Dr. Tom Rydzak, and Prof. Jan Westpheling, for their assistance at various stages in my training. At UT, I would like to thank Prof. Jeff Becker, Prof. Joe Bozell, Prof. Terry Hazen, Prof. Fu-min Minn, Alex Peeden, and Prof. Gary Sayler for access to to instrumentation, strain databases, and helpful information as I performed the experiments presented.

Finally, I would like to thank my family and friends for their unending love and support of my academic pursuits, without which I never would have accomplished this work. I'm especially grateful for my siblings Michelle and Kevin and my brother-in-law Austin, who keep me grounded, and my nieces Maddie and Harper, who never fail to make me smile. I'm also grateful for my aunts and uncles and cousins in the Cristobal and Thompson families who have always encouraged me to be myself. I'm also incredibly thankful for the network of close friends I've managed to maintain. Despite geographical distance and busy schedules I've always received encouragement from Greg Curtin, Melissa Dibben, Alex and Lizzie Donnelly, Nick Highers, Matt Hlavaty and Kailey Rhodes, Geoff Madrazo, Daniel and Jenna Mathewson, and Royce Munn.

Abstract

The development of a secure and sustainable energy economy is likely to require the production of fuels and commodity chemicals in a renewable manner. There has been renewed interest in biological commodity chemical production recently, in particular focusing on non-edible feedstocks. The fields of metabolic engineering and synthetic biology have arisen in the past 20 years to address the challenge of chemical production from biological feedstocks. Metabolic modeling is a powerful tool for studying the metabolism of an organism and predicting the effects of metabolic engineering strategies. Various techniques have been developed for modeling cellular metabolism, with the underlying principle of mass balance driving the analysis. In this dissertation, two industrially relevant organisms were examined for their potential to produce biofuels.

First, *Saccharomyces cerevisiae* was used to create biodiesel in the form of fatty acid ethyl esters (FAEEs) through expression of a heterologous acyl-transferase enzyme. Several genetic manipulations of lipid metabolic and / or degradation pathways were rationally chosen to enhance FAEE production, and then culture conditions were modified to enhance FAEE production further. The results were used to identify the rate-limiting step in FAEE production, and provide insight to further optimization of FAEE production.

Next, *Clostridium thermocellum*, a cellulolytic thermophile with great potential for consolidated bioprocessing but a weakly understood metabolism, was investigated for enhanced ethanol production. To accomplish the analysis, two models were

created for *C. thermocellum* metabolism. The core metabolic model was used with extensive fermentation data to elucidate kinetic bottlenecks hindering ethanol production. The genome scale metabolic model was constructed and tuned using extensive fermentation data as well, and the refined model was used to investigate complex cellular phenotypes with Flux Balance Analysis.

The work presented within provide a platform for continued study of *S. cerevisiae* and *C. thermocellum* for the production of valuable biofuels and biochemicals.

Table of Contents

1 Introduction **1**

1.1 Biotechnology for Fuels and Chemicals 1

1.1.1 Cellulosic Feedstocks 3

1.1.2 *Saccharomyces cerevisiae* as a platform species 5

1.1.3 *Clostridium thermocellum* as a platform species 6

1.2 Computational Modeling as a Tool for Metabolic Engineering 6

1.2.1 Underlying Principles 8

1.2.2 Metabolic Network Analysis 9

1.3 Outline of Dissertation 17

2 Materials and Methods **19**

2.1 Experimental Methods 19

2.1.1 *S. cerevisiae* Strain Development 19

2.1.2 *S. cerevisiae* Plasmid Construction 20

2.1.3 *S. cerevisiae* Growth Conditions 22

2.1.4 *C. thermocellum* Strain Development 22

2.1.5 *C. thermocellum* Growth Conditions 23

2.1.6 Analytical Methods 24

2.1.7 Bioreactor-linked Real-time Gas Measurement 26

2.1.8 Determination of Experimental Fluxes 27

2.2 *C. thermocellum* Metabolic Model Construction 28

2.2.1 Construction of Core Metabolic Model 28
2.2.2 Construction of Genome Scale Metabolic Model 30
2.3 Metabolic Modeling Techniques 36
2.3.1 Elementary Mode Analysis 36
2.3.2 Metabolic Flux Analysis 37
2.3.3 Metabolic Flux Ratio Analysis 38
2.3.4 Flux Balance Analysis 39
2.3.5 Flux Sum Analysis 40
2.3.6 Calculation of ATP Costs 40
2.3.7 Sampling of Flux Distributions 40

3 Enhancing Fatty Acid Ethyl Ester Production in *Saccharomyces cerevisiae* through Metabolic Engineering and Medium Optimization 42

3.1 Introduction 43
3.2 Results and Discussion 45
3.2.1 Metabolic engineering of *S. cerevisiae* to increase cytosolic acyl-CoA concentration 45
3.2.2 Manipulating Nitrogen Limiting Conditions for Enhanced FAEE Production 48
3.2.3 Deploying Codon Optimized Acyl-Transferase Genes for Enhanced FAEE Production 50
3.2.4 Identification of the Rate Limiting Step to Enhancing Acyl-CoA Pools 51
3.2.5 Increasing FAEE Production Under Fed-Batch Conditions . . 52
3.3 Conclusion 53

4 Elucidating Central Metabolic Redox Obstacles Hindering Ethanol Production in *Clostridium thermocellum* 55

4.1 Introduction 56

4.2 Results . . . 58
4.2.1 Quantitative analysis of cellular phenotypic states . . . 58
4.2.2 Metabolic model illustrates robust redox metabolism . . . 62
4.2.3 Metabolic model encompasses several phenotypes . . . 66
4.2.4 Elucidating redox bottlenecks hindering ethanol production . . . 67
4.2.5 Design of optimal genotype for high-yielding ethanol production 71
4.3 Discussion . . . 74

5 Examining Complex Phenotypes With A Genome Scale Metabolic Model of *Clostridium thermocellum* DSM 1313 Implementing an Adjustable Cellulosome 79
5.1 Introduction . . . 80
5.2 Results . . . 83
5.2.1 Model construction and comparison . . . 83
5.2.2 ATP Requirement for Growth on Cellobiose . . . 85
5.2.3 Additional ATP Requirement for Cellulosome Synthesis . . . 88
5.2.4 Application of GEM for Rational Strain Design . . . 90
5.2.5 Effect of Cellodextrin Length on Growth . . . 91
5.2.6 Effect of substrate and cell growth rate on bioenergetics of *C. thermocellum* . . . 93
5.3 Discussion . . . 100
5.3.1 Proposed Bioenergetic Regulatory Mechanism of *C. thermocellum* Fermentation . . . 101

Bibliography 107

Vita 151

List of Tables

2.1 *Saccharomyces cerevisiae* Strains and Plasmids 21

2.2 Comparison of ATP requirements for *C. thermocellum* and model organisms during anaerobic growth 34

4.1 Experimentally determined growth rates, fluxes, and yields for cellobiose-grown batch cultures 62

4.2 One-at-a-time sensitivity analysis for select reactions 63

4.3 Elementary mode analysis for various *C.thermocellum* strains. 72

5.1 Comparison of Clostridial Genome Scale Models 85

5.2 Experimental fluxes used for metabolic model constraints 86

5.3 Model-guided MCS strain designs for production of hydrogen, ethanol, and isobutanol. 91

List of Figures

1.1 Toy Network Analysis 10

2.1 Central metabolic network of *C. thermocellum* 29
2.2 *C. thermocellum* GEM Construction Flowchart and Reaction Classifications 33

3.1 Overview of *S. cerevisiae* pathways towards FAEE production from sugars 44
3.2 Initial Characterization of Engineered Yeast for FAEE Production . . 47
3.3 Characterization of Engineered Yeast Under Nitrogen Limitation . . . 49
3.4 Characterization of Engineered Yeast Under Alternative Feeding Conditions 52
3.5 Fermentation Profile of *S. cerevisiae* Strains 53

4.1 Representative fermentation profile and carbon balance for cellobiose-grown batch cultures. 59
4.2 Comparison of experimental and calculated fluxes. 60
4.3 Experimental yields and METAFoR analysis of cellobiose-grown batch cultures. 61
4.4 Parametric plots of normalized flux distributions through core model 65
4.5 *In silico* supplementation with exogenous electron sinks 68
4.6 Probing redox imbalance with exogenous electron sinks *in vivo* 69
4.7 Phenotypic spaces for various genotypes calculated by EMA 71

4.8 Application of MMF on *C. thermocellum* 74

5.1 Alternative Cellulosome Compositions 84
5.2 Tuning ATP Requirement for Cellobiose Growth 87
5.3 Tuning ATP Requirement for Cellulose Growth 89
5.4 Effect of Cellodextrin Length *in silico* and *in vivo* 92
5.5 Experimentally Observed Phenotypes and their Implementation as Model Constraints 94
5.6 Cofactor Turnover Predicted from Cellobiose and Cellulose Simulations 96
5.7 Calculated Flux Distributions Around PEP to Pyruvate Conversion . 98
5.8 Calculated Flux Distributions Around Pyruvate to Acetyl-CoA Conversion 99
5.9 Calculated Flux Distributions In Redox Metabolism 100
5.10 Proposed Mechanism of Ethanol Overflow Metabolism 102

Chapter 1

Introduction

Summary

This chapter is a review of many different aspects of biotechnology for the production of commodity chemicals, such as fuels. The chapter acts a funnel of information to steer the reader through relevant background information, steadily focusing in on the work presented throughout the dissertation. Herein, I present the current schools of thought for utilizing microbial species as cell factories, the reasoning behind the study of *Saccharomyces cerevisiae* and *Clostridium thermocellum*, and the powerful capabilities of computational modeling as a tool for studying metabolism and designing production strains.

1.1 Biotechnology for Fuels and Chemicals

Human societies have long utilized the microorganisms around them for the production of foods and beverages, such as bread, beer, yogurt, or kimchee, with evidence found as early as 7,000 BC (McGovern et al., 2004). Microbial fermentation produces the chemicals key to flavoring these foods and beverages, such as alcohols, organic acids, esters or lactones (Layton and Trinh, 2014). In the modern age, the traditional uses of microorganisms continue at large scale to feed an increasing population, and

as knowledge of chemistry and cellular metabolism increases, our use of these entities has increased as well. In fact, many of the products of microbial fermentation are of interest to many chemical industries (Rogers et al., 2013). One of the earliest non-food related industrial process utilizing microorganisms was the production of acetone during World War I, after shortages led to the search for alternative means of production. *Clostridium acetobutylicum* produces acetone, butanol, and ethanol from various sugars, and this ABE process was widely used until the 1950s when petroleum sources were economically favorable again (Gabriel, 1928; Jones and Woods, 1986; Ross, 1961; Taillefer and Sparling, 2016). Part of the reason for the return to petroleum is the branched nature of the ABE process limits the amount of each product, which hinders industrial potential. Currently, many industrial processes for synthesizing these chemicals rely on petroleum-based precursors. There is significant concern about the sustainability of petroleum feedstocks, however, such as uncertainty in supply due to political instability or the impact on the environment by increasing greenhouse gas concentrations in the atmosphere and contamination from spills (Demain et al., 2005). One way to address both concerns is to develop biotechnology to utilize renewable resources for producing the chemicals currently derived from petroleum (Demain, 2009).

The renewal of interest in biological alternatives to petroleum is facilitated by our understanding of biology, such as the development of high-throughput techniques for characterization and genetic manipulation (Olson et al., 2015). Naturally, this has led to a large industry dedicated to prospecting for useful biological entities, or genetic resources (Bull, 2004; Lanen and Shen, 2006). Regardless, for a biotechnological process to be considered as a proper petroleum replacement in energy or industrial applications, it should provide a net energy gain, have environmental benefits, be economically competitive, and be producible at large scale without reducing food supplies (Hill et al., 2006). To address the final point, the use of cellulosic biomass instead of starch or simple sugar-rich feedstocks is critical.

1.1.1 Cellulosic Feedstocks

Cellulose is the most abundant polymer in the biosphere, and from an economic standpoint, cellulosic biomass is less expensive than most energy sources at about $3 per gigajoule (GJ) (Lynd et al., 2008). Additionally, it is recognized that production of cellulosic crops, such as grasses or short rotation trees, could have more positive environmental attributes than production of corn, soy, or other annual row crops (Farrell et al., 2006; Greene et al., 2004; Hammerschlag, 2006). Given this information, future scenarios in which biological resources are converted to energy or commodity chemicals at sufficient scale primarily utilize cellulosic feedstocks (Somerville et al., 2010). One major challenge, however, is overcoming the recalcitrance barrier. Plants have been evolving for millions of years to resist degradation by outside microorganisms, and this resistance is largely responsible for the high costs of cellulose conversion (Himmel et al., 2007; Stephanopoulos, 2007). There are two main strategies for converting cellulosic biomass into desired chemicals: Simultaneous Saccharification and Fermentation (SSF) and Consolidated Bioprocessing (CBP).

Simultaneous Saccharification and Fermentation

In order to utilize cellulosic materials, many organisms need assistance in degrading the cellulose. The process called Simultaneous Saccharification and Fermentation (SSF) uses organisms which are efficient at producing chemicals of interest incubated with cellulolytic enzymes. First, the cellulosic biomass is pretreated to separate cellulose from other biopolymers such as lignin and, depending on the organism's capabilities, hemicellulose. Following pretreatment, the cellulose stream is added to a reactor containing fermentative organisms and cellulolytic enzymes, which are produced separately. This process has several problems, however, such as discrepancies between temperatures of optimal enzyme activity and microbial growth. Also prohibitive is the large cost of producing the cellulolytic enzymes.

Despite these shortcomings, SSF has been used for ethanol production using a wide variety of microorganisms, such as *Bacillus coagulans* (Ou et al., 2010), *Kluyveromyces marxianus* (Ballesteros et al., 2004; Kadar et al., 2004), *Kluyveromyces fragilis* (Hari Krishna et al., 2001), *Mucor indicus* (Karimi et al., 2006), *Rhizopus oryzae* (Karimi et al., 2006), various white rot fungi (Itoh et al., 2003), and *Saccharomyces cerevisiae* (Stenberg et al., 2000; Ohgren et al., 2007; Varga et al., 2004; Wyman et al., 1992). Lactic acid and butanol have also been produced via SSF (Anuradha et al., 1999; Qureshi et al., 2008).

Consolidated Bioprocessing

To alleviate the high cost of enzyme production, the CBP strategy has been developed, and is characterized by having enzymes be produced by the fermenting organism (Lynd et al., 2002). However, since an organism which can efficiently utilize cellulose *and* produce high product levels has not yet been discovered, there are two main strategies for developing a CBP microorganism. The first focuses on taking highly productive organisms and enabling them to grow on cellulose. The second focuses on taking organisms which can efficiently degrade cellulose and engineering them to produce higher yield and titers of product (Lynd et al., 2005; Olson et al., 2012).

Following the first strategy, several organisms have been used in CBP processes, such as lactate production in *Bacillus subtilis* (Zhang et al., 2011) or *Lactobacillus plantarum* (Okano et al., 2009), glutamate production in *Corynebacterium glutanicum* (Hyeon et al., 2011; Tsuchidate et al., 2011), three different biofuels from *Escherichia coli* (Bokinsky et al., 2011), ethanol production in *Saccharomyces cerevisiae* (van Zyl et al., 2007), and others (Hasunuma et al., 2013).

Following the second strategy, several organisms have been used for CBP, such as *Caldicellulosiruptor bescii* (Chung et al., 2014), *Cellvibrio japonicas* (Gardner and Keating, 2010), *Clostridium cellulolyticum* (Guedon et al., 2002; Li et al., 2012),

Clostridium thermocellum (Akinosho et al., 2014), and others (Hasunuma et al., 2013; Lynd et al., 2005; Olson et al., 2012).

One key difference in the application of these strategies is the difference in how cellulolytic organisms degrade cellulose, i.e. by using free enzymes or by using a cell-linked enzyme complex known as the cellulosome (Blumer-Schuette et al., 2014).

1.1.2 *Saccharomyces cerevisiae* as a platform species

Saccharomyces cerevisiae, or brewer's yeast, has long been used for conversion of grains to beer or bread, and because of its ubiquity has been used as a model organism to study eukaryotic genetics and metabolism. Thus, as many genetic tools are available and the knowledge of its physiology is significant, *S. cerevisiae* has been used for the synthesis of a large number of products across volume scales (Hong and Nielsen, 2012), e.g.

Biofuels such as ethanol (Guadalupe Medina et al., 2010), isobutanol (Chen et al., 2011a), and biodiesel in the form of FAEEs (Yu et al., 2012) or bisabolene (Peralta-Yahya et al., 2011)

Bulk Chemicals such as 1,2-propanediol (Lee and DaSilva, 2006), lactate (Skory, 2003), pyruvate (van Maris et al., 2004), glycerol (Hecker et al., 1990) and succinate (Raab et al., 2010)

Fine Chemicals such as β-carotene (Verwaal et al., 2007), amorphadiene (Farhi et al., 2011; Westfall et al., 2012), eicosapentaenoic acid (Tavares et al., 2011), resveratrol (Becker et al., 2003), vanillin (Brochado et al., 2010), and ascorbic acid (Sauer et al., 2004)

In this work, we focus on using simple sugars as a feedstock for *S. cerevisiae*, but the use of complex biomass as a substrate in a CBP process is currently the focus of many efforts (van Zyl et al., 2007).

1.1.3 *Clostridium thermocellum* as a platform species

One other microorganism of particular interest for CBP is *Clostridium thermocellum*. First isolated in 1926 (Viljoen et al., 1926) and recently re-evaluated phylogenetically as *Ruminiclostridium thermocellum* (Yutin and Galperin, 2013), this thermophilic, gram-positive fermicute grows on crystalline cellulose at one of the fastest observed rates (Lynd et al., 2002) at an optimal temperature of 60^{o}C.

Clostridium thermocellum degrades lignocellulosic biomass using a large, cell-linked cellulosome (Shoham et al., 1999) through an atypical glycolytic pathway ending in a branched fermentation profile capable of producing ethanol, acetate, hydrogen, formate, lactate, CO_2, and amino acids (McBee, 1954). Details of *C. thermocellum* catabolism and the complex system of redox balancing enzymes are outlined in the following chapters. To date, the lack of reliable genetic tools and the complexity of *C. thermocellum* metabolism has hindered metabolic engineering strategies towards producing a single or limited range of products.

Given the challenge of restructuring a branched metabolism with time-consuming genetic tools, it would be desirable to investigate the feasibility of engineering strategies systematically before attempting to construct modified strains. A way to satisfy that desire is to use metabolic modeling techniques on an accurate reconstruction of *C. thermocellum* metabolism.

1.2 Computational Modeling as a Tool for Metabolic Engineering

Modern advancements in DNA sequencing technology have led to the genome sequencing of many different organisms across all kingdoms of life. These sequences have been compiled into large databases (Pagani et al. (2012); Scheer et al. (2011); Schellenberger et al. (2010), `http://www.ncbi.nlm.nih.gov/`, `http://jgi.doe.gov/`, `https://www.ebi.ac.uk/`), and this wealth of data requires systematic means

of analysis. For instance, comparative genomics can be used to automatically construct draft metabolic networks in a high-throughput manner (Henry et al., 2010), and a standard protocol has been established to improve the accuracy of these draft metabolic networks (Thiele and Palsson, 2010). A reconstructed metabolic network provides a rational link between a genome sequence, the proteins encoded in the genome, and the reactions catalyzed by the proteins (Durot et al., 2009). This link allows one to directly study the relationships between cellular genotype and phenotype. Guided by mass conservation, many metabolic network analysis tools have been developed to understand how biochemical compounds, or metabolites, are produced, consumed, or transformed during the course of cellular metabolism (Lewis et al., 2012). More importantly, metabolic network analysis can be used to predict how cellular metabolism can change under genetic or environmental disturbances.

Metabolic network analysis can be categorized into dynamic or structural approaches. Dynamic metabolic network models can be useful in describing temporal behavior under different perturbations, and a general protocol has been developed for building and validating a dynamic genome scale model (Jamshidi and Palsson, 2008). Unfortunately, the implementation of dynamic modeling methods is hindered by the unavailability of kinetic parameters for the majority of enzymes within a metabolic network, despite attempts to estimate kinetic parameters under uncertainty (Flowers et al., 2013; Machado et al., 2012; Ramkrishna and Song, 2012; Song and Ramkrishna, 2012; Tan et al., 2011; Wang et al., 2004; Young et al., 2008).

Due to the lack of kinetic parameters, structural metabolic modeling has been widely used for analyzing cellular metabolism at a steady-state. In this work, given the context of interest in *C. thermocellum* and the level of uncertainty in its metabolism, the focus will be on structural modeling. Depending on the availability of experimental data and the scientific objective, different structural modeling techniques have been developed including metabolic flux analysis (MFA), flux balance analysis (FBA), and metabolic pathway analysis (MPA) (Lewis et al., 2012; Stephanopoulos et al., 1998; Trinh et al., 2009). Structural metabolic modeling

techniques share their underlying principles, yet differ in their efficacy in solving different metabolic engineering problems.

1.2.1 Underlying Principles

The primary assumption behind the analysis of metabolic networks is built upon the first principle of mass conservation of internal metabolites within a biological system (Reder, 1988; Schuster and Schuster, 1993; Stephanopoulos et al., 1998). A metabolic network can be defined by a single cell, a cell compartment, or network of cells in which material enters and leaves the system. The material is transformed through an intricate map of enzyme-catalyzed reactions. The reactions which transform metabolites are defined as internal, whereas reactions involving the transport of metabolites in and out are considered exchange reactions (Schilling et al., 2000a,b; Schuster and Hilgetag, 1994). The general equation to describe the mass conservation of metabolites within a system of defined volume is

$$\frac{d\mathbf{c}}{dt} = \mathbf{S} \cdot \mathbf{r} - \mu \cdot \mathbf{c}, \tag{1.1}$$

where $\mathbf{c}$ is an m x 1 column vector of internal metabolite concentrations (mmol/L), $\mathbf{S}$ is the m x n stoichiometric matrix of the network whose rows and columns represent metabolites and reactions, respectively, $\mathbf{r}$ is the n x 1 flux vector (mmol/L/hr), and μ is the dilution effect (1/hr), i.e., volume expansion during cell growth. Within a single cell, the contribution of volume change to the change in metabolite concentrations can be ignored since the dilution rate is much slower than reaction rates of enzymes (Stephanopoulos et al., 1998). During exponential growth, or during a stable continuous culture, the concentrations of internal metabolites can be assumed to not accumulate in the system (Fell, 1992; Stephanopoulos et al., 1998), and Equation 1.1 can be simplified to

$$\mathbf{S} \cdot \mathbf{r} = 0. \tag{1.2}$$

Enzymes, like everything, are governed by the laws of thermodynamics. Therefore, if a reaction i is irreversible (or far from equilibrium), it should be constrained to a positive flux by the inequality

$$r_i \geq 0. \tag{1.3}$$

These two constraint make up the framework for the aptly name Constraints Based Reconstruction and Analysis of metabolic networks.

1.2.2 Metabolic Network Analysis

Figure 1.1 shows how the problem can be formulated for an example toy network. This network consists of eleven reactions, two of which are reversible, and contains eleven metabolites, six of which are internal. Usually, Equation 1.2 is an underdetermined system where the number of metabolites is less than the number of reactions. Given the network structure and the availability of experimentally measured fluxes to constrain the system, three main techniques have been developed to solve the system of linear equations for the metabolic flux vector $\mathbf{r}$: metabolic flux analysis (MFA), flux balance analysis (FBA), and metabolic pathway analysis (MPA). These methods are distinct but can complement each other, and as the development of metabolic modeling techniques for rational strain design expands, they display a co-evolutionary trend (Machado and Herrgard, 2015). These methods are described below.

Metabolic Flux Analysis

The underdetermined nature of most metabolic networks dictates that $m < n$, and so there are n-m degrees of freedom if $m = rank(\mathbf{S})$. MFA can calculate $\mathbf{r}$ by measuring $n - m$ fluxes, decomposing $\mathbf{r}$ and $\mathbf{S}$ into measured and unmeasured components as follows:

$$\mathbf{r}_u = -\mathbf{S}_u^{-1} \cdot \mathbf{S}_m \cdot \mathbf{r}_m \tag{1.4}$$

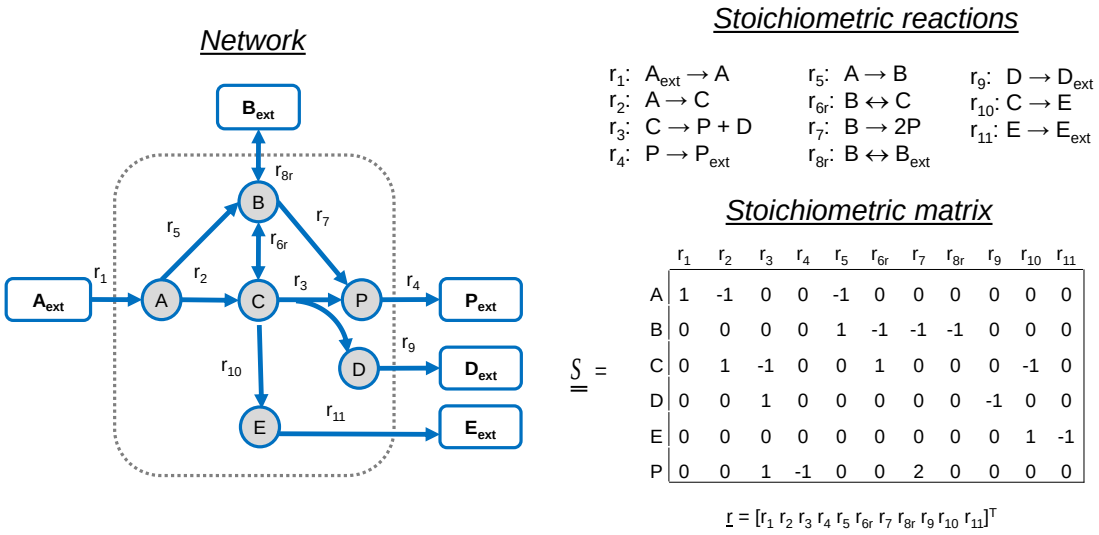

Figure 1.1: The toy network pictured on the left contains five internal metabolites (A, B, C, D, E, and P) linked through the internal reactions and transported into and out of the cell via exchange reactions. The stoichiometric matrix is constructed from the stoichiometric coefficients of each reaction. The rows correspond to internal metabolites and the columns to reactions.

MFA is set up to quantify the metabolic flux vector **r** corresponding to a phenotypical state of the cell under a given condition. It can be seen in Equation 1.4 that a large set of experimental data is required to solve for $\mathbf{r}_u$, especially for a large metabolic network, which can become experimentally infeasible. MFA usually handles small or simplified metabolic networks (Martnez et al., 2010), which may be lacking critical components.

Recently, ^{13}C-based MFA has been developed to calculate **r** using the ^{13}C labeling pattern of the protein-bound amino acids determined by either gas chromatography coupled mass spectroscopy (GC/MS) and / or nuclear magnetic resonance (NMR) (Christensen et al., 2002; Sauer, 2006; Wiechert et al., 2001; Zamboni et al., 2005, 2009). A more sophisticated technique, known as kinetic flux profiling, has recently been developed to calculate **r** by measuring the dynamic incorporation of labeled substrates, such as ^{13}C or ^{15}N, into intracellular metabolites. This

measurement can be subsequently used to calculate metabolic fluxes directly without relying on a simplified network (Yuan et al., 2006, 2008, 2010).

Flux Balance Analysis

Similar to MFA, FBA determines a metabolic flux vector $\mathbf{r}$ which represents a phenotypic state of a cell. A benefit of this method is that it can be applied to a genome-scale metabolic network where $m \ll n$ (Schilling et al., 1999). FBA is based on convex analysis, in particular the application of convex minimization in the context of the optimization theory branch of mathematics. The FBA method is characterized by the objective function(s), e.g. the maximization of cell growth, minimization of ATP usage, or the maximization of growth-coupled product synthesis (Feist and Palsson, 2010; Schuetz et al., 2007). Some reactions in $\mathbf{r}$ are constrained with experimentally measured or hypothetical rates, such as substrate uptake, product synthesis, nutrient availability (Feist and Palsson, 2008; Kauffman et al., 2003; Lewis et al., 2012; Price et al., 2004; Varma and Palsson, 1994). Additionally, thermodynamic constraints, regulatory constraints, or protein expression level (via transcriptomic and/or proteomic datasets) can also be applied to constrain the objective function (Chandrasekaran and Price, 2010; Jensen et al., 2011; Park et al., 2007). The general framework is

$$\begin{aligned} max \quad & \mathbf{c}^{\intercal}\mathbf{r} \\ s.t. \quad & \mathbf{S} \cdot \mathbf{r} = 0 \\ & r_{i,lb} \leq r_i \leq r_{i,ub} \quad \forall r_i \in \mathbf{r}, \end{aligned} \tag{1.5}$$

where $\mathbf{c}$ is the objective function vector, in which reactions involved in the objective are given a positive value and all other values are zero, $\mathbf{r}$ is the flux vector, and $\mathbf{S}$ is the stoichiometric matrix. The lower and upper boundaries for reaction r_i are $r_{i,lb}$ and $r_{i,ub}$, respectively.

Some challenges still apply when using FBA, e.g. i) to justify the assumption of objective functions used to predict the physiological state of the cell under a given condition (Schuster et al., 2008), and ii) account for all alternative solutions that may represent actual physiological state of the cell (Lee et al., 2000; Mahadevan and Schilling, 2003; Reed and Palsson, 2004). However, many techniques have been developed to assess these challenges. For instance, FBA has been altered to include genetic regulatory elements through methods such as regulatory FBA (rFBA) (Covert et al., 2001; Covert and Palsson, 2002, 2003), Gene Inactivity Moderated by Metabolism and Expression (GIMME) (Becker and Palsson, 2005), Steady-state regulatory FBA (SR-FBA) (Shlomi et al., 2007), and Probabilistic Regulation of Metabolism (PROM) (Chandrasekaran and Price, 2010).

An optimal solution from FBA is not necessarily, and almost never, unique. The variability in a flux distribution can be assessed by Flux Variability Analysis, where an objective is held constant (or within a minimal percentage of its optimal value) while the minimum and maximum feasible flux is calculated for all internal reactions (Mahadevan and Schilling, 2003). To look beyond the extreme flux limits at a certain objective, sampling methods can be employed to produce a set of flux distributions spanning the entire solution space. Methods for sampling include Monte Carlo-based "hit-and-run" algoritms (Smith, 1984; Wiback et al., 2004a; Lovasz, 1999), artificial centering "hit-and-run" (Kaufman and Smith, 1998; Megchelenbrink et al., 2014), and poling-based FBA (Binns et al., 2015).

FBA has also been built upon by assuming cells will minimize their adjustment of metabolism in response to experimental conditions (Ibarra et al., 2002) or gene knockouts (Segre and Vitkup, 2002; Shlomi et al., 2005). This paradigm has led the way for many algorithms implementing bilevel programming with the goal of rational strain design such as OptKnock (Burgard et al., 2003), RobustKnock (Pharkya et al., 2003), OptStrain (Pharkya et al., 2004), SimOptStrain (Kim et al., 2011), OptReg (Pharkya and Maranas, 2006), Genetic Design based on Local Search (GDLS) (Lun et al., 2009), Flux Scanning based on Enforced Objective Flux

(FSEOF) (Choi et al., 2010), OptOrf (Kim and Reed, 2010), OptForce (Ranganathan et al., 2010), BiMOMA (Kim et al., 2011), and MOMAKnock (Ren et al., 2013). FBA has also been coupled to genetic algorithms to ease the computational burden of searching the complex solution spaces of large networks in the methods OptGene (Patil et al., 2005) and Set-Based Evolutionary Algorithm and Simulated Annealing (SEA-SA) (Rocha et al., 2008).

Metabolic Pathway Analysis

Considering Equation 1.2 and Inequality 1.3 without an objective function leads to the realm of Metabolic Pathway Analysis (MPA). This formalism is capable of identifying all admissible metabolic flux vectors in a metabolic network, or the entire solution space. From convex analysis theory, solutions to the system belong to the convex polyhedral flux cone and are infinite (Rockafellar, 1970). To obtain a finite number of solutions which span the convex polyhedral cone, additional constraints can be applied.

Elementary Mode Analysis (EMA) uses a non-decomposability or genetic independence constraint to reduce the solutions of Equation 1.2 and Inequality 1.3 to a finite set of admissible flux vectors. An admissible flux vector $\mathbf{r}_a$ is called an elementary mode (EM) if there is no other admissible vector $\mathbf{r}_b$, when $\mathbf{r}_a \neq \mathbf{r}_b$ and $\mathbf{r}_b \neq 0$ such that the set of non-zero elements in $\mathbf{r}_b$ is a subset of the non-zero elements in $\mathbf{r}_a$ (Schuster and Hilgetag, 1994; Trinh et al., 2009). This constraint makes $\mathbf{r}$ contain a minimal set of enzymatic reactions which will be unique up to a scalar multiple, which implies that deletion of any reaction belonging to an EM will disrupt the entire network to no longer satisfy Equation 1.2 and Inequality 1.3.

Extreme Pathway Analysis (EPA) uses the same constraints as EMA but includes a requirement for systematic independence, which specifies that any extreme pathway (EP) cannot be expressed as a non-linear combination of two other EPs (Schilling et al., 2000a,b). The set of EPs for a network constitute the generating set of vectors which span the entire polyhedral convex cone. Additionally, all EPs form a subset

of the EMs (Klamt and Stelling, 2003; Papin et al., 2004). Taken together, EMA and EPA serve as a powerful tool for studying cellular metabolism by examining the entire solution space. However, the solution space will become compoundingly complex with the addition of dimensionality to the network, and the computational burden can be a limiting factor on the size of the network. Recently, significant effort has been given to compute EMs on large-scale networks (Centler et al., 2010; Hunt et al., 2014; Jevremovic et al., 2011; Kaleta et al., 2009).

The computational complexity does not hinder the applicability of MPA, as the ability to identify all feasible unique EMs inherent to a metabolic network is very useful. It has been used to systematically examine the metabolic capabilities of red blood cells (Wiback and Palsson, 2002), photosynthetic chloroplasts (Poolman et al., 2003), *Haemophilus influenzae* (Papin et al., 2002a), *Helicobacter pylori* (Price et al., 2002b), *Lactobacillus rhamnosus* (Poolman et al., 2004), *Cornynebacterium glutamicum* (Rajvanshi and Venkatesh, 2011), *Shewanella oneidensis* (Flynn et al., 2012), *Saccharomyces cerevisiae* (Schwartz and Kanehisa, 2006), and *Escherichia coli* (Carlson and Srienc, 2004b,a; Wiback et al., 2004b; Wlaschin et al., 2006).

Additionally, with the knowledge of the solution space of a network, one can explore metabolic network properties such as underlying structure (Ay and Kahveci, 2010; Papin et al., 2002b; Peres et al., 2011; Poolman et al., 2007; Price et al., 2003), network robustness (Larhlimi et al., 2011; Min et al., 2011; Stelling et al., 2002, 2004), network fragility (Behre et al., 2008; Wilhelm et al., 2004), potential phenotype discovery (Kaleta et al., 2011; Kenanov et al., 2010; Schauble et al., 2011), potential regulatory mechanisms (Cakir et al., 2004, 2007; Stelling et al., 2002; Wessely et al., 2011), and potential targets for therapeutic or pharmaceutical applications (Beuster et al., 2011).

Metabolic Pathway Analysis for Rational Strain Design

Most pertinent for this work, MPA techniques have also been used for rational strain design. Conceptually speaking, from an entire set of EMs for a given metabolic

network, one can select a subset of EMs which provide the highest product yields. To enforce the operation of only the high-yielding EMs, a brute-force approach would be to delete all reactions whose fluxes are zero in the defined EM subsets. However, the number of knockout candidates can be significantly large given the fact that i) some reaction fluxes are inactive under a given growth condition considered in the analysis, e.g. anaerobic growth will not utilize oxygen-dependent respiration reactions, and ii) reactions belonging to a linear pathway will have zero flux if any reaction in the linear pathway is deleted. Thus, it is beneficial to find a minimum set of deleted reactions to reduce the entire set of EMs to a constrained, desirable phenotypic subset of EMs.

Minimal Metabolic Functionality (MMF) is an effective method for finding the constrained subset of EMs. The method is driven by the goal of finding a minimum set of deleted reactions which leads to overproduction of a target molecule coupled with cell growth (Trinh et al., 2009). First, EMA identifies all feasible pathways in a metabolic network. Next, reactions are chosen for deletion via manual inspection of the EM space following three guidelines: i) eliminate as many undesirable pathways as possible to reduce the entire set of EMs to a high-yield subset, ii) if EMs in the desirable subset allow cell growth, they must couple growth with high product yield, and iii) if an EM is in the desired subset but cell growth is not part of the EM, then the target product must be synthesized at theoretical maximum yield. This method proceeds in an iterative manner to constrain the phenotypic space of the network and has been used effectively to design and experimentally validate optimal *E. coli* strains to produce ethanol from hexoses and pentoses (Trinh et al., 2008), ethanol from glycerol (Trinh and Srienc, 2009), and higher alcohols from glucose (Trinh et al., 2011; Trinh, 2012).

Other algorithms have also been developed for rational design of strains which produce a chemical of interest in a more general manner than MMF. For instance, a Minimal Cut Set (MCS) is defined as a subset of reactions within the metabolic network that, once removed, force the flux of target reactions to zero (Klamt and Gilles, 2004). Generally speaking, this process can be accomplished by defining the

deletion task as the set of EMs to be deleted (Klamt, 2006) The MCS framework was improved to calculate constrained Minimal Cut Sets (cMCSs), which identify which MCSs are required to eliminate undesirable pathways (EMs) but also preserving high-yield pathways (Hädicke and Klamt, 2011). Each cMCS contains a unique and minimum set of deleted reactions that constrain the system with the predefined goal of product synthesis and/or side-product elimination. Alternatively, the Computational Approach for Strain Optimization Aiming at high Productivity (CASOP) method was developed with similar goals of designing optimal strains for high target chemical production (Hädicke and Klamt, 2010). CASOP is a systematic approach to score reactions which can be deleted, up-regulated, or down-regulated by finding a balance between high-yield EMs and high network robustness. Using two parameters, CASOP varies the distribution between growth and product synthesis while adjusting a weighting factor for each high-yielding EM.

Recently, the Systematic Multiple Enzyme Targeting (SMET) method was developed to couple dynamic modeling with EMA and cMCS to find kinetic bottlenecks in metabolic pathways (Flowers et al., 2013). First, a large ensemble of models with random kinetic parameters is generated and constrained to wild-type flux distribution. Next, cMCSs are calculated for a target chemical overproduction and a cMCS is applied to the ensemble of kinetic models. By studying the effect of genetic perturbations (e.g. knockout, down-regulation, up-regulation) on the system as a whole, kinetic bottlenecks hindering target chemical production can be inferred. Finally, EMA has been combined with optimization methods in the MODCELL method (Trinh et al., 2015). MODCELL seeks to find a genetic background which is the best chassis for the production of a wide-range of target chemicals, so that a single modular chassis genotype can be used in a plug-and-play manner with multiple production modules with minimal cell adaptation.

The EMA-based methods mentioned above have been used for designing genetic modification strategies for chemical synthesis in *Aspergillus niger* (Driouch et al., 2012), *Corynebacterium glutanicum* (Bartek et al., 2010; Chen et al., 2009; Kind

and Wittmann, 2011; Kromer et al., 2006; Neuner and Heinzle, 2011), *Eschericia coli* (Chen et al., 2010; Trinh et al., 2008; Trinh and Srienc, 2009; Liao et al., 1996; Unrean et al., 2010), *Klebsiellia pneumoniae* (Zhang and Xiu, 2009; Chen et al., 2011b), *Pseudomonas putida* and *Ralstonia eutropha* (Diniz et al., 2006), *Saccharomyces cerevisiae* (Carlson et al., 2002; Matsuda et al., 2011), and *Thermoanaerobacter saccharolyticum* (Unrean and Srienc, 2011, 2012).

1.3 Outline of Dissertation

Presented in this work is my efforts to synthesize FAEEs in *Saccharomyces cerevisiae* and to create a platform for the systematic study of *Clostridium thermocellum* DSM 1313 metabolism. This platform consists of two distinct yet intimately related metabolic models. In addition to constructing metabolic networks from the genome sequence (Feinberg et al., 2011), there are significant validation and refinement steps included. This platform is then used to examine experimentally observed phenotypes with a variety of techniques reviewed above. While the species and products differ, both directions utilize metabolic engineering techniques to elucidate bottlenecks in the production pathways.

This dissertation proceeds in the following manner:

Chapter 2 This chapter is a compilation of the methods used throughout the work. It includes descriptions of model creation and modeling methods, as well as details regarding experimental data acquisition and calculation of fermentation parameters.

Chapter 3 This chapter describes an effort to enhance fatty-acid ethyl esters in *Saccharomyces cerevisiae* using a combination of metabolic engineering and growth-condition optimization.

Chapter 4 This chapter describes the construction and validation of the core metabolic model. Using experimental data and metabolic flux analysis, we

are able to represent the phenotypes of multiple genotypes designed to perturb electron and carbon metabolism. Using predictions from the core model, we are then able to elucidate the steps in redox metabolism that hinder ethanol production.

Chapter 5 This chapter describes the construction and validation of the genome scale model. The meticulously refined GEM was tuned to accurately simulate growth on cellobiose and cellulose, accounting for the difference in cellulosome synthesis. Using this model, we examine several observed phenotypes presented throughout the literature and infer a large regulatory network which arises from growth on cellobiose.

Chapter 2

Materials and Methods

Summary

This chapter provides a compilation of methods used when performing the work presented throughout the dissertation. This includes both "wet-lab" experimental methods as well as *in silico* experimental methods.

2.1 Experimental Methods

2.1.1 *S. cerevisiae* Strain Development

The complete list of *S. cerevisiae* strains constructed for this work are listed in Table 2.1. Wild-type *S. cerevisiae* strain BY4741 and single gene knockouts were kindly provided by Prof. Jeff Becker (The University of Tennessee, Knoxville). To construct multiple gene knockout strains, the well-established *loxP*-cre recombinase system was employed. Gene deletion cassettes were amplified from plasmid pXP418 (Fang et al., 2011) using 40-50 base pair homology immediately upstream of the gene of interest. After transformation of linear PCR products by electroporation (Manivasakam and Schiestl, 1993), gene replacements were confirmed via colony PCR. Uracil (URA3) marker removal was accomplished by transforming the knockout strain with pBF3038

(Addgene) then growing in SC-Leu with 2% galactose as a carbon source to induce *cre* recombinase activity (Fang et al., 2011). URA3 marker removal was confirmed by selection on YPD+FOA plates (Boeke et al., 1984) and subsequent colony PCR. This gene disruption/marker removal process was repeated as needed (Hegemann and Heick, 2011). Final, confirmed *S. cerevisiae* strains were transformed with their respective plasmids via electroporation and selected on SC-URA plates. Colony PCR was used to confirm presence of heterologous genes before characterization.

2.1.2 *S. cerevisiae* Plasmid Construction

Table 2.1 shows a list of plasmids used in this study. The plasmid pRS426TEF (Mumberg et al., 1995) was used as a backbone for construction of the ethyl ester production plasmids. The acyltransferase gene *atfA* was amplified from *Acinetobacter* sp. ADP1 genomic DNA (Addgene) for constitutive expression using the TEF promoter. Sequences of *Acinetobacter* sp. ADP1 *atfA* and *Marinobacter hydrocarbonoclasticus* DSM 8798 MhATF1 codon-optimized for expression in *S. cerevisiae* were previously described (Shi et al., 2012) and synthesized as gBlock fragments (IDT Inc.). Expression plasmids were constructed using Gibson Assembly (Gibson et al., 2009; Gibson, 2011). All PCR products were generated using Phusion Hot Start II DNA polymerase (Thermo Inc.). The Gibson Assembly plasmid product was directly transformed into chemically-competent TOP10 cells (Invitrogen Inc.), selected with ampicillin, and extracted with a standard plasmid extraction kit (Zymo Research Inc.). The constructs were validated by colony PCR, digested fragment lengths, and sequencing (University of Tennessee, Knoxville Molecular Biology Resource Facility). All primers used for piece amplification and confirmation of gene deletions were synthesized by IDT Inc.

Table 2.1: *Saccharomyces cerevisiae* Strains and Plasmids

Strain Name	Genotype / Description	Source
BY4741	*S. cerevisiae MATa, ura3d0, his3-d200, leu2-d0, met15-d0*	ATCC
ScAT1001	BY4741 Δ*faa*2::kanMX	ATCC
ScAT1005	BY4741 Δ*pxa*2::kanMX	ATCC
ScAT1025	BY4741 Δ*acb*1::kanMX	ATCC
ScAT1031	BY4741 Δ*faa*2::kanMX Δ*pxa*2::*loxP* Δ*acb*1::*loxP*	This study
ScAT1033	BY4741 Δ*faa*2::kanMX Δ*pxa*2::*loxP*	This study
ScAT2000	BY4741 + pRS426TEF	This study
ScAT2200	BY4741+pAT01y	This study
ScAT2203	ScAT1001 + pAT01y	This study
ScAT2204	ScAT1005 + pAT01y	This study
ScAT2205	ScAT1025 + pAT01y	This study
ScAT2208	ScAT1033 + pAT01y	This study
ScAT2209	ScAT1031 + pAT01y	This study
ScAT2018	BY4741 + pAT07y	This study
ScAT2022	ScAT1031 + pAT07y	This study
ScAT2211	BY4741 + pAT27y	This study
ScAT2212	BY4741 + pAT28y	This study
ScAT2215	ScAT1031 + pAT27y	This study
ScAT2216	ScAT1031 + pAT28y	This study

Plasmid Name	Genotype / Description	Source
pAT01y	TEFp:*atfA*:CYCt, URA3	This study
pAT07y	GAL1p:*atfA*:CYCt, URA3	This study
pAT27y	TEFp:*atfA**:CYCt, URA3	This study
pAT28y	TEFp:*MhATF1**:CYCt, URA3	This study
pRS426TEF	TEFp::CYCt, URA3	Mumberg et al. (1995)
pXP418	loxP:URA3:loxP, TEFp::CYCt	Fang et al. (2011)
pBF3038	GAL1p:cre, LEU2	Fang et al. (2011)

2.1.3 *S. cerevisiae* Growth Conditions

Media

Laura Bertani complex medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) plus 100 mg/L ampicillin (if applicable) was used for DNA manipulation in *E. coli* TOP10. For *S. cerevisiae*, YPD complex medium containing 10 g/L yeast extract, 20 g/L peptone, and 20 g/L glucose was used to prepare competent cells before plasmid selection. Solid media was prepared as above with 20 g/L agar. Counter-selection for uracil marker removal in *S. cerevisiae* was conducted on plates containing 1 g/L 5-fluoro-orotic acid (FOA). For strain characterization, the defined SC-URA medium containing 6.7 g/L yeast nitrogen base (containing 5 g/L ammonium chloride), 2 g/L synthetic dropout mix without uracil, and 20 g/L glucose was used. All chemicals were purchased from Sigma.

Strain characterization

For the batch cultures, *S. cerevisiae* was characterized in 250 mL beveled flasks with 50 mL working volumes, shaking at 190 rpm at 30^{o}C. Nitrogen limitation studies were conducted in 50 mL working volumes with a starting C:N ratio of 50:1 (57 g/L glucose) in modified SC-URA containing only 1 g/L ammonium chloride, and a 10% dodecane overlay for FAEE capture was included. Fatty acid doping under nitrogen limitation studies followed the 50:1 C:N recipe above with the addition of 0.1% pentadecanoic acid and a dodecane overlay.

2.1.4 *C. thermocellum* Strain Development

C. thermocellum strains employed in this study were derived from the genetically tractable strain DSM 1313 with a Δhpt background, allowing for a two-stage selection counter-selection method using 8-azahypoxanthine (Argyros et al., 2011). *C. thermocellum* $\Delta hydG$ and $\Delta hydG$ Δech strains were previously described (Biswas

et al., 2015). The mutant Δ*hydG* Δ*pta-ack* was constructed from the Δ*hydG* strain using the plasmid pAMG498, similar to previous reports (Argyros et al. (2011); van der Veen et al. (2013)), but also deleting acetate kinase. Gene loci of *hpt*, *hydG*, *ech*, and *pta-ack* are Clo1313_2927, Clo1313_1571, Clo1313_0564-0575, and Clo1313_1185-1186, respectively. Strains were kindly provided by Adam M. Guss (Oak Ridge National Laboratory).

To eliminate pyruvate-formate lyase (PFL) activity and create an effective Δ*pfl* strain, the parent strain or Δ*hydG* Δ*ech* strain was cultured with 6 mM sodium hypophosphite (HPP), an analogue of formate and known PFL inhibitor (Rydzak et al. (2014); Unkrig et al. (1989)).

2.1.5 *C. thermocellum* Growth Conditions

C. thermocellum strains were cultured in a defined MTC medium containing (per liter): 5 g cellobiose, 2 g urea, 1.5 g ammonium chloride, 2 g sodium citrate tribasic dihydrate, 1.5 g citric acid monohydrate, 1 g sodium sulfate, 1 g potassium phosphate monobasic, 2.5 g sodium bicarbonate, 1 g magnesium chloride hexahydrate, 0.2 g calcium chloride dihydrate, 0.1 g iron (II) chloride tetrahydrate, 1 g L-cysteine HCl, 20 mg pyridoxamine dihydrochloride, 4 mg p-aminobenzoic acid, 2 mg biotin, 2 mg vitamin B12, 1.25 mg manganese chloride tetrahydrate, 0.5 mg cobalt chloride hexahydrate, 0.125 mg zinc chloride, 0.125 mg copper chloride dihydrate, 0.125 mg boric acid, 0.125 mg sodium molybdate dihydrate, and 0.125 mg nickel chloride hexahydrate (Ozkan et al. (2001); Zhang and Lynd (2003); Kridelbaugh et al. (2013)). Balch tube studies were conducted with 10 mL MTC plus the appropriate additional chemicals where noted. For bottle studies and inoculum preparation, 120 mL serum bottles were prepared with 90 mL MTC. Balch tube and serum bottle cultures also contained 5 g L^{-1} MOPS for buffering, starting at a pH adjusted to 7.4 with 5M KOH. Bottles and tubes were sealed with butyl rubber stoppers and aluminum seals,

while the head space contained a N_2:CO_2 (90:10) mixture. Cells were maintained at -80^{o}C prior to initial inoculation. Inocula were incubated in water baths set at 55^{o}C.

Bioreactor studies were conducted with a Biostat B+ (Sartorius Stedim US, Bohemia, NY) dual 1.5 L fermentation system at a working volume of 1 L MTC media with MOPS omitted. Prior to inoculation, freshly prepared media were autoclaved and sparged overnight with a N_2:CO_2 (90:10) gas mixture at 500 mL/min at 55^{o}C. Inoculation used 10% v/v mid-log phase cell culture to give a starting optical density of 0.05 in bioreactors. At inoculation, the gas flow was slowed to 150 mL/min, and the gas input was altered to flush the headspace only. The culture was maintained at pH 7.0 with 5 M KOH and 40% H3PO4, an agitation rate of 200 rpm, and a temperature of 55^{o}C. Batch bioreactor studies were performed at least in duplicate. For each run, triplicate samples were collected and analyzed at each time point. The fluxes, titers, and yields were reported as means ± standard deviation for all samples.

Optical density (OD) was measured via a spectrophotometer at 600 nm (Spectronic 200+, Thermo Fisher Scientific, Inc.) To establish a calibration curve between dry cell weight (DCW) and OD (i.e., 0.502 g DCW = 1 OD_{600}, R^2 ¿ 0.95), 80 mL culture samples at known optical densities were centrifuged and washed twice with 0.9% (w/v) NaCl in water. Washed cell pellets were then oven dried at 70^{o}C for at least three days, until DCW was stable. The calibration curve was constructed from at least two different cell samples for several different OD measurements. An elemental DCW composition of $C_4H_7O_2N$ (101 g mol^{-1}) was used to calculate DCW in moles (Rydzak et al., 2009).

2.1.6 Analytical Methods

Saccharomyces cerevisiae

Yeast growth was monitored by measuring optical density at 600 nm over time. Substrate and product concentrations were monitored by using the Shidmazu high performance liquid chromatorgraphy (HPLC) system. One mL aliquots of culture

were centrifuged, filtered at 0.2 micron, and ran on an Aminex 87H (Biorad Inc.) column using a refraction index detector to monitor glucose consumption with glycerol, acetate, and ethanol production (Trinh et al., 2008).

FAEE production in shake flask cultures was quantified via gas chromatography coupled with mass spectroscopy (GC/MS). A 100 L aliquot of the overlaying dodecane was removed from the culture and diluted 10-fold in ethyl acetate with ethyl pentadecanoate as an internal standard. Samples were run on an HP 6890 gas chromatograph equipped with a DB-5MS column (Agilent) using an HP 5973 mass spectrometer for detection. The GC method was ramped in all samples as follows: an initial temperature of 120^{o}C was maintained for 2 minutes, followed by a 30^{o}C min^{-1} increase to a final temperature of 250^{o}C, which was held for an additional 3 minutes. The solvent delay was 4.5 minutes to allow passage of dodecane. Product detection was accomplished using selected ion monitoring. The ions detected were 88.1 for all samples, as well as the principal ion for each compound being monitored (C10EE:200.3, C12EE: 228.2, C14EE: 256.3, C15EE: 270.3, C16:0EE: 284.3, C18:0EE: 312.3). The principal ion detection was altered as a function of time within the method to focus on the specific target compound at its specific retention time.

Clostridium thermocellum

Substrate and product concentrations were quantified by using a Shimadzu high pressure liquid chromatography (HPLC) system. One mL aliquots of culture were centrifuged, then the supernatants filtered at 0.2 micron and run on an Aminex 87H (Biorad Inc.) column using a refractive index detector (RID) to monitor cellobiose consumption with lactate, formate, acetate, pyruvate, and ethanol production by an established method (Trinh et al., 2008).

Gas chromatography coupled with mass spectrometry (GC/MS) was also used to quantify extracellular amino acids. At each measured time point during fermentation, 100 L aliquots of cell culture were centrifuged at 25,200 rcf for 5 minutes and

cell supernatants collected for secreted amino acid analysis. Amino acids were then derivatized using the EZ-faast™ Free/Physiological Amino Acid GC/MS kit (Model # KG0-7166, Phenomenex Inc., Torrence, CA, USA). The manufacturer's protocols for derivatization and GC/MS separation/analysis were followed, and the concentrations of secreted amino acids were calculated from calibration curves with standard mixtures provided by the kit as well as control mixtures made in-house.

2.1.7 Bioreactor-linked Real-time Gas Measurement

H_2 monitoring was accomplished in real-time using a universal gas analyzer (UGA-300) equipped with an RGA probe (SRS Inc., Sunnyvale, CA, USA) connected to the gas effluent stream from the bioreactor system. The UGA-300 was equipped with a multiple inlet valve for the concurrent sampling from both bioreactor vessels. Exhaust gases (mainly N_2, H_2, and CO_2) from the bioreactor vessels were monitored by in-line mass flow meters (Cole Palmer, Vernon Hills, IL) calibrated for a N_2:CO_2 mixture to allow for accurate calculation of H_2 concentration. The pressure within the reactor was maintained at 1.5 atm with a ball-type pressure relief valve in series with the effluent stream, downstream of the mass flow meter and before reaching the UGA-300. Accumulative H_2 ($n$$H_2$ in mole) were determined by the following formula:

$$n_{H2} = \int_0^t \frac{P \cdot v}{R \cdot T} \cdot X_{H2} \cdot dt \tag{2.1}$$

where t is the fermentation time (hr), P is the pressure of the reactor (Pa), v is the volumetric flow rate (m^3 hr^{-1}), R is the ideal gas constant (J mol^{-1} K^{-1}), T is temperature (K), and X_{H2} is the molar fraction of H_2.

Carbon dioxide production was not measured directly due to the presence of CO_2 in the influent gas causing fluctuations in the CO_2 signal from the UGA-300. The fluctuations were sufficient to disrupt an accurate measurement of CO_2 concentration. In these experiments, the CO_2 production was estimated from the concentrations of ethanol, acetate, formate, and cell mass at specific time points during the fermentation

by the formula:

$$[CO_2] \triangleq [ethanol] + [acetate] + 0.546[DCW] - [formate] \tag{2.2}$$

where square brackets represent concentrations of respective species. The scalar operator for cell mass represents the sum of the requirement of acetyl-CoA and α-ketoglutarate minus the requirement of oxaloacetate for cell mass synthesis (Deng et al., 2013). It should be noted that DCW and cell mass are used interchangeably.

2.1.8 Determination of Experimental Fluxes

Concentration profiles of 11 potentially prominent components of the fermentation broth (i.e., cellobiose, glucose, succinate, lactate, pyruvate, acetate, ethanol, formate, CO_2, H_2, and cell mass) were determined at each time point. Concentrations of all 20 major amino acids were measured during exponential growth and at the final time point, and the prominent species were included in flux analysis and carbon balance. Their fluxes were determined as follows:

$$r_p = \mu \cdot Y_{P/X} = \mu \cdot \frac{d[P]/dt}{d[X]/dt} \tag{2.3}$$

where r_p is the specific consumption/production rate (flux) of metabolite P (mmol g^{-1} hr^{-1}), μ is the specific growth rate (hr^{-1}), $Y_{X/P}$ is the yield of metabolite P per unit cell mass X, and brackets represent concentrations in mmol L^{-1}. Similarly, for chemostat cultures, the fluxes were calculated as follows:

$$r_p = D \cdot Y_{P/X} = D \cdot \frac{[P]_{out} - [P]_{in}}{[X]_{out} - [X]_{in}} \tag{2.4}$$

where D is the dilution rate (hr^{-1}). Where $[\cdot]_{in}$ and $[\cdot]_{out}$ represent the concentration of a given species for the feed and effluent, respectively.

2.2 *C. thermocellum* Metabolic Model Construction

2.2.1 Construction of Core Metabolic Model

A metabolic network was constructed for *C. thermocellum* which utilized cellobiose, a representative substrate for cellulosic biomass, as the sole carbon source. The network reconstruction was based on the publicly available curation of the DSM 1313 genome sequence (Feinberg et al., 2011) as well as an extensive literature review. The model accounts for the strict anaerobic metabolism of *C. thermocellum* which does not use the canonical forms of Embden-Meyerhof glycolysis (Zhou et al., 2013), the pentose phosphate pathway (Rydzak et al., 2012), or the Krebs (tricarboxylic acid, TCA) cycle (Roberts et al., 2010).

In addition, the export of major fermentative products (i.e., ethanol, acetate, lactate, formate, and H_2), CO_2, valine, and DCW were included in the model. The biomass term was constructed following a composition described previously (Gowen and Fong (2010); Roberts et al. (2010)), adjusting amino acid and G+C content for *C. thermocellum* DSM 1313. All cellular components are derived from 12 precursor metabolites (Neidhardt et al., 1990). Total ATP requirement for DCW synthesis (i.e., $Y_{ATP}^{max} = 42.7$ mmol ATP g^{-1} DCW) was estimated following the well-established protocol (Stouthamer, 1973). The maintenance energy requirement was modeled as a separate stoichiometric reaction.

Briefly, our model proceeds from cellobiose uptake via an ATP-binding cassette transporter and cleavage via cellobiose phosphorylase (Ng and Zeikus (1982); Nochur et al. (1992)). Glucose subunits are then processed through glycolysis until the phosphoenolpyruvate (PEP) step, where PEP can be converted to pyruvate through two branches: i) inter-conversion to oxaloacetate and malate through the "malate shunt," effectively transferring electrons from NADH to NADPH in a transhydrogenase type of mechanism, ii) decarboxylation of oxaloacetate to pyruvate

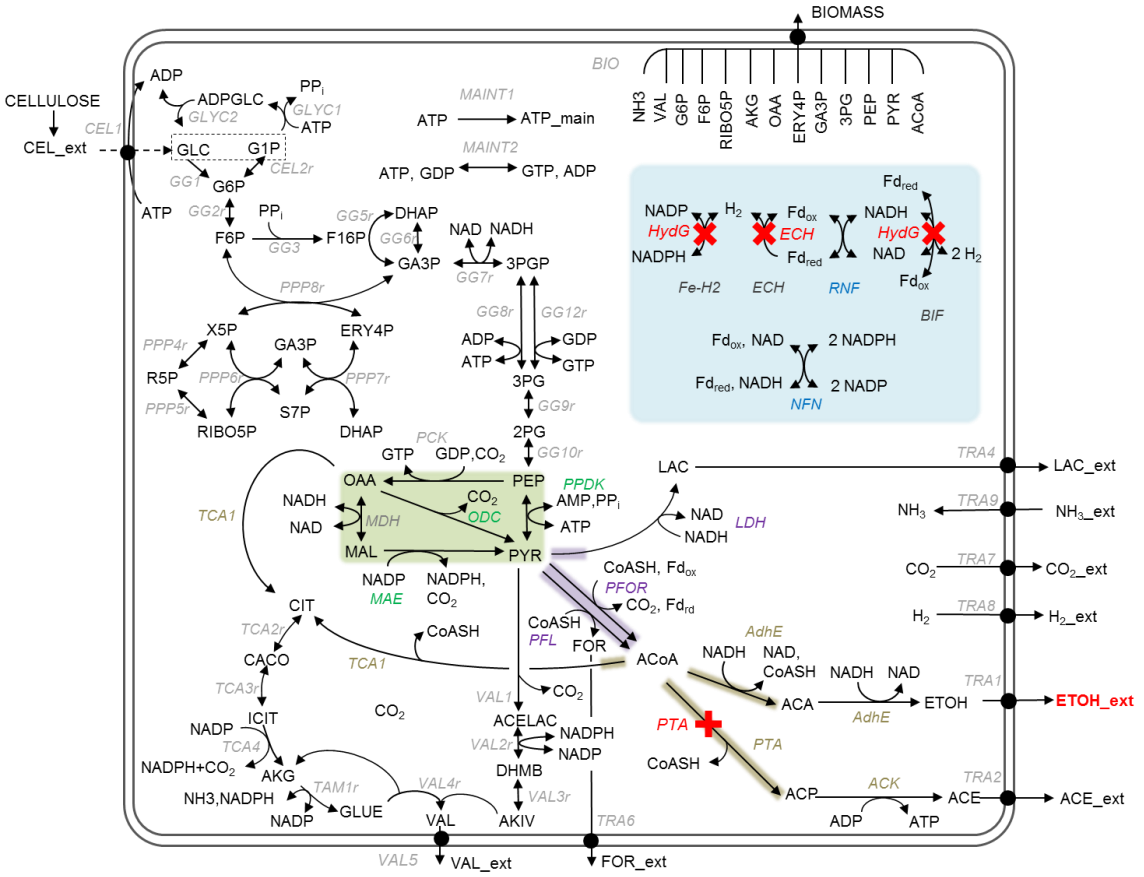

Figure 2.1: The central metabolic network of *C. thermocellum*. Genetic modifications investigated in this study are marked with a red X. Reactions which contribute to metabolite nodes of interest are highlighted. Pyruvate phosphate dikinase (PPDK), oxaloacetate decarboxylase (ODC), and malic enzyme (MAE) contribute to →PYR (green box). Pyruvate formate lyase (PFL), pyruvate ferredoxin oxidoreductase (PFOR), and lactate dehydrogenase (LDH) contribute to PYR→ (purple box). ACoA→ (brown box) consists of alcohol dehydrogenase (AdhE), phosphotransacetylase (PTA), and citrate synthase (TCA1). The redox pathway is highlighted in the blue box. Fd_{rd} → consists of the Ni-Fe energy conserving hydrogenase (ECH), the Rnf-type NADH:Fd oxidoreductase (RNF), the NADH-Fd_{rd}:$NADP^+$ oxidoreductase (NFN), and the NADH-Fd_{rd} bifurcating hydrogenase (BIF). The → H_2 node includes ECH, BIF, as well as the Fe-Fe NADPH-depenent hydrogenase (Fe-H2)

via oxaloacetate decarboxylase (ODC), and iii) direct conversion to pyruvate coupled with ATP production via pyruvate phosphate dikinase (PPDK). Enzymes in these pathways are transcribed (Riederer et al., 2011) and translated (Rydzak et al., 2012) according to recent reports, despite PPDK enzymatic activity not being confirmed in cell-free extracts (Zhou et al., 2013).

Redox metabolism of *C. thermocellum* is complex, with multiple potential pathways available for redox cofactor recycling (Carere et al., 2012). Our model incorporated detailed electron shuttling reactions, consistent with the most recent understanding (Rydzak et al., 2014). Most prominently represented in *C. thermocellum* is a complicated system which shuttles electrons between reduced ferredoxin, NADH, and NADPH using several different enzyme complexes including reduced ferredoxin:NAD(P) oxidoreductases such as RNF or NFN, an Fe-Ni energy conservation hydrogenase (ECH), and several Fe-Fe hydrogenases which are activated by HydG (Biswas et al., 2015). A visual representation of the core metabolic model can be seen in Figure 2.1 and a full list of reactions with stoichiometry is included in Appendix .

2.2.2 Construction of Genome Scale Metabolic Model

Construction of Draft Metabolic Network

The first draft of the metabolic network of *Clostridium thermocellum* DSM 1313 was constructed using the automatic reconstruction function *getKEGGModelForOrganism()* of the RAVEN Toolbox (Agren et al., 2013). This function compiled reactions from the KEGG database organism entry for *C. thermocellum* DSM1313 (T01933,*ctx*) with the complete set of coding sequences from the genome assembly (GenBank: CP002416.1) so that each reaction is linked to a specific protein encoded by the genome (Feinberg et al., 2011). This draft reconstruction contained gene-protein relationships (GPR), pathway information, and reaction stoichiometry.

Transport reactions were linked to genes by compiling a list of putative transporters and then curating reactions via manual inspection. The list was compiled by three methods: i) Using InterProScan 5 to find 169 putative transporters within the *C. thermocellum* DSM 1313 protein sequences, ii) Extracting annotated transporters from alternative Clostrial GEMs (Milne et al., 2011; Roberts et al., 2010; Salimi et al., 2010) and subjecting them to reciprocal blast hit (RBH) and *hmmscan* to determine similar genes in *C. thermocellum* DSM 1313. For RBH, we used $1e^{-50}$ and blast length of fifty amino acids as cut-offs. For *hmmscan*, our cutoff was also $1e^{-50}$, and iii) We used the Transporter Substrate Database (Zhao et al., 2011) to extract protein sequences for each substrate exchange reaction within the model and compared them to the genome sequence of DSM 1313 with a cutoff of 1 for *hmmscan* and $1e^{-50}$ for RBH.

Further, as it is known that *C. thermocellum* cofactor specificity in glycolysis is atypical (Zhou et al., 2013), we investigated reaction sets from the automatic reconstruction which only differed by cofactor choice, e.g. NADH versus NADPH or ATP versus GTP. Enzymes with available *in vitro* data were adjusted accordingly. For enzymes which the automatic reconstructed predicted multiple equivalent reactions with differing cofactors, the proteins linked to these reactions were analyzed for cofactor specificity using Cofactory (Geertz-Hansen et al., 2014). This software determines the specificity towards NAD, NADP, or FAD of proteins by predicting Rossman folds from primary sequence.

Refinement of Genome Scale Metabolic Model

The KEGG draft network contained many gaps in the central metabolism and was plagued by unrealistic predictions because it assumed that many reactions were reversible, which lead to thermodynamically infeasible pathways. To build a working GEM, the KEGG draft network was expanded and refined in the following manner:

i The central metabolic network recently reported (Thompson et al. (2015), Section 2.2.1) was manually built into the GEM structure, filling in gaps in glycolysis and redox metabolism which were not automatically included,

ii The dry cell weight composition presented for strain ATCC 27405 (Roberts et al., 2010) was adapted to reflect the differences in genomic content between strains as well as a stringent calculation of ATP requirements for biomass synthesis (Neidhardt et al., 1990; Stouthamer, 1973), then included the cell composition reaction into the network. The ATP required for DCW synthesis by cell component is given in Table 2.2,

iii Several artificial reactions to convert identical yet alternately described metabolites (e.g. β-D-Fructose 6-phosphate $\rightarrow$ D-Fructose 6-phosphate) were added within the network to close gaps between discrepancies,

iv The automatic gap filling function of the RAVEN Toolbox was used with the previously constructed *C. thermocellum* GEM *i*SR432 (Roberts et al., 2010) as a template to increase network connectivity,

v Reactions were added to fill gaps in sulfate utilization pathway, which is known to be utilized as a sole sulfur source (Kridelbaugh et al., 2013) as well as shikimate kinase, homoserine kinase, and spontaneous glutamate semialdehyde cyclization reactions to allow synthesis of all essential amino acids on minimal media, and

vi Each reaction in the network was manually inspected for appropriate reversibility, adjusting reactions outside of glycolysis and fermentation-based substrate-level phosphorylation to only consume ATP. This convention is commonplace (Thiele and Palsson, 2010) and it removed the cycles in the model which were incorrectly generating energy. During the manual curation process, extensive metadata was included for reactions and metabolites to allow for cross-linking between KEGG (Kanehisa and Goto, 2000), MetaCyc (Caspi et al., 2012),

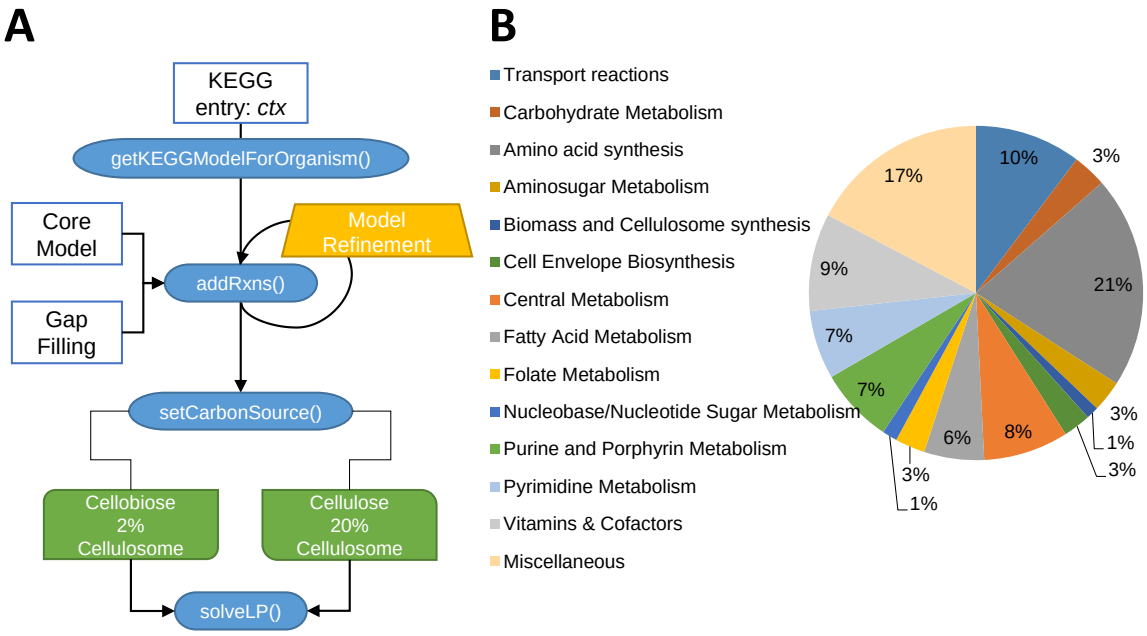

Figure 2.2: **(A)** Flowchart of model construction. **(B)** Distribution of *i*AT601 reactions belonging to KEGG pathways

MetaNetX (Ganter et al., 2013; Bernard et al., 2014), SEED (Overbeek et al., 2005), BRENDA (Scheer et al., 2011), and other databases.

The refined model has been deposited in the DOE KBase (https://narrative.kbase.us/narrative/ws.13674.obj.2). A workflow of the model construction process is given in Figure 2.2.

Implementation of the cellulosome into the Genome Scale Model

The cellulosome is the cellulose-degrading protein complex covalently bound to the surface of certain cellulolytic bacteria, such as *C. thermocellum*. The previous GEM of *C. thermocellum*, *i*SR432 (Gowen and Fong, 2010; Roberts et al., 2010), roughly included the cellulosome on top of the dry cell weight (DCW) reaction in a condition independent manner. However, it is well documented that the cellulosome fraction of total DCW changes from 2% to 20% when growing on cellobiose versus cellulose, respectively (Zhang and Lynd, 2005b), and the protein composition of the

Table 2.2: Comparison of ATP requirements for *C. thermocellum* and model organisms during anaerobic growth

ATP requirement (mmol ATP/g DCW)			
	C. thermocellum	*E. coli* Stouthamer (1973)	*S. cerevisiae* Verduyn et al. (1990)
Polysaccharides / Cell wall	7.16	2.05	4.63
Protein			
Amino acid synthesis	5.65	1.36	1.81
Polymerization	17.49	19.14	17.76
Lipid	0.52	0.14	0
RNA			
Nucleoside monophosphate formation	1.2	3.45	1.35
Polymerization	0.27	0.92	
mRNA turnover	0.3	1.39	0.71
DNA			
Deoxynucleoside monophosphate formation	0.48	0.86	
Polymerization	0.18	0.19	
Subtotal	33.23	29.5	26.26
Transport:			
Ammonium	8.43	4.24	6.29
Potassium	0.2	0.19	2.4
Phosphate	0.8	0.77	
Total ATP Required	42.66	34.71	34.95
Y_{ATP}^{Max} (g DCW / mol ATP)	23.44	28.81	28.61

cellulosome itself changes when growing on alternative substrates (Gold and Martin, 2007; Raman et al., 2009; Riederer et al., 2011; Wei et al., 2014; Wilson et al., 2013). Therefore, our cellulosome reaction was set up to allow for dynamic switching between modeling cellobiose versus cellulose-consuming growth conditions. This is an important distinction due to the increased ATP requirement for exporting more protein from the cell.

The fractional composition for cell dry weight was 0.5285 g Protein + 0.026 g DNA + 0.0655 g RNA + 0.076 g Lipid + 0.2242 g Cell Wall + 0.00494 g Solute Pool + 0.0304 g Total LTA → g Cell Dry Weight (CDW). The whole cell biomass term was adjusted depending on carbon source, specifically: 1 g CDW + 0.02 g Cellulosome term → Biomass for cellobiose cultures and 1 g CDW + 0.2 g Cellulosome term → Biomass for cellulose cultures (Roberts et al., 2010).

The composition of the cellulosome was initially set equivalent to the protein term. Using experimentally observed cellulosomal protein abundances, we altered the cellulosome composition systematically. First, a matrix A was created by counting the amino acids required to synthesize cellulosomal proteins. The entry $A_{i,j}$ corresponds to the number of amino acid i encoded by the sequence of protein j. Second, the abundances of each protein were condensed into a condition specific vector c normalized to *CipA* (Raman et al., 2009). The total amino acid count across all cellulosomal proteins for any condition can be calculated by $A \cdot c$. Finally, the condition specific amino acid count is converted to mmol/g cellulosome similar to the calculation of protein or DNA terms (Neidhardt et al., 1990).

To complement the adjustable cellulosome reaction, transport reactions were added for cellodextrin oligomers of length 3 to 6 glucose subunits, i.e. cellotriose (G3) to cellohexaose (G6). Glucose and cellobiose transport were included in the automatic reconstruction. It has been shown that *C. thermocellum* cellodextrin transporters prefer longer chain oligomers (Nataf et al., 2010). Further, there is a complex set of regulatory interactions where excess cellobiose represses cellulase activity (Zhang and Lynd, 2005b) and conversely, cellobiose uptake is inhibited by

the presence of G3 to G5 oligomers (Strobel et al., 1995). Upon entering the cell, cellodextrins are cleaved in a phosphorolytic manner (Zhang and Lynd, 2004) and so reactions were included to utilize G6 to G2 oligomers by a sequential chain-shortening pathway generating glucose-1-phosphate and a cellodextrin of length G(N-1). The final glucose residue in this depolymerization pathway is phosphorylated with ATP. This mechanism of transport and phosphorolytic cleavage costs 2 ATP per cellodextrin imported, regardless of length, and as such the ATP yield per glucose equivalent is higher when assimilating longer oligomers (Zhang and Lynd, 2005a), as demonstrated in Section 5.2.5.

It is difficult to obtain information regarding individual oligomer uptake rates *in vivo* so most studies report data in the units of mmol glucose equivalents / g dry cell weight (DCW)/ hr. In order to utilize this information as a constraint in the GEM, we implemented a flux ratio constraint (McAnulty et al., 2012) between the individual cellodextrin uptake reactions and the uptake rate of glucose equivalents as such

$$6r_{G6} + 5r_{G5} + 4r_{G4} + 3r_{G3} + 2r_{G2} + 1r_{G1} = r_{GluEq} \tag{2.5}$$

where $r_{G(N)}$ is the specific uptake rate of the cellodextrin of length N. This constraint maintains stoichiometric balance when using commonly reported experimental data to test the model.

2.3 Metabolic Modeling Techniques

2.3.1 Elementary Mode Analysis

From the construction of a *C. thermocellum* metabolic network, the mass balance for intracellular metabolites can be written as described in Section 1.2.2 under the steady-state assumption as:

$$\mathbf{S} \cdot \mathbf{r} = 0, \tag{2.6}$$

where **S** is the stoichiometric matrix whose rows and columns correspond to metabolites and reactions, respectively, and **r** is the intracellular flux vector. Irreversible reactions are subject to the constraint

$$r_i \geq 0. \tag{2.7}$$

Elementary mode analysis (EMA) was applied to calculate all possible solutions **r**, called elementary modes (EMs) that satisfy equation 2.6, inequality 2.7, and an additional non-decomposability constraint. The constraint states that for any **EM**1 and **EM**2, S(**EM**1) cannot be a subset of S(**EM**2) and vice versa where S(**EM**1) and S(**EM**2) are indices of non-zero fluxes of **EM**1 and **EM**2, respectively (Trinh et al., 2009). EMs were calculated for the metabolic network of *C. thermocellum* DSM 1313 using METATOOL (Pfeiffer et al. (1999); Kamp and Schuster (2006)).

To calculate cMCS for the GEM *i*AT601, we required cell growth be greater than 0.01 and specified a minimum product yield of 75% theoretical maximum using a recently develop regulatory cut set method (Mahadevan et al., 2015). All calculations were performed in MATLAB in the CellNetAnalyzer Toolbox (Klamt et al., 2007; Klamt and von Kamp, 2011).

2.3.2 Metabolic Flux Analysis

We calculated the metabolic flux distribution for *C. thermocellum* based on the measured fluxes of extracellular metabolites under a given condition. The flux distribution was determined by a non-negative, linear combination of EMs using Poolman et al.'s algorithm (Poolman et al., 2004), and briefly described below:

$$\mathbf{r} = \mathbf{EM} \cdot \mathbf{w} \tag{2.8}$$

where **r** is an nx1 flux distribution vector, **EM** is an nxk elementary mode matrix, and **w** ($\in \mathbb{R}^+$) in a kx1 weighting factor vector. By using the measured flux vector

$\mathbf{r}_m$ ($\mathbf{r}_m \subseteq \mathbf{r}$), $\mathbf{w}$ can be calculated as follows:

$$\mathbf{r}_m = \mathbf{EM}_m \cdot \mathbf{w} \tag{2.9}$$

$$\mathbf{w} = pinv(\mathbf{EM}_m) \cdot \mathbf{r}_m \tag{2.10}$$

where $pinv(\mathbf{EM}_m)$ is the mxk Moore-Penrose pseudo inverse matrix of $\mathbf{EM}_m$. The flux calculation errors were determined as follows:

$$error = \frac{\|\mathbf{r} - \mathbf{r}_m\|}{\|\mathbf{r}\|} \tag{2.11}$$

One-at-a-time sensitivity (OATS) analysis was performed by systematically removing an individual reaction from the EM space for a given genotype, calculating the flux distribution with the resulting EM space, and comparing the new calculated error to the error of the original flux distribution vector.

2.3.3 Metabolic Flux Ratio Analysis

To quantitatively differentiate flux distributions at key metabolite nodes in the central metabolic network upon perturbations, we employed the metabolic flux ratio (METAFoR) analysis (Sauer et al. (1999); Szyperski et al. (1999)). The metabolic flux ratio for an incoming flux(es) at node i ($\text{MFR}_{i\leftarrow}$) is defined as:

$$MFR_{i\leftarrow} = \frac{r_{j,i\leftarrow}}{\sum_k r_{k,i\leftarrow}} \tag{2.12}$$

where $r_{j,i\leftarrow}$ is the incoming flux(es) of node i through reaction j. $MFR_{i\rightarrow}$ can be defined similarly where the outgoing flux(es) of node $i(r_{j,i\rightarrow})$ is represented with the opposite arrow. Given the significant overflow metabolism in *C. thermocellum*, we are interested in carbon and electron fluxes directly to ethanol as well as upstream of ethanol synthesis. The key branch points in focus here are i) conversion of PEP to pyruvate ($\rightarrow$PYR) via pyruvate phosphate dikinase (PPDK), oxaloacetate

decarboxylase (ODC), and malic enzyme (MAE), with ODC and MAE jointly considered, ii) recycling reduced ferredoxin (Fd_{rd} $\rightarrow$) through the Ni-Fe energy conserving hydrogenase (ECH), the Rnf-type NADH:Fd oxidoreductase (RNF), the NADH-Fdrd:$NADP^+$ oxidoreductase (NFN), and the NADH-Fdrd bifurcating hydrogenase (BIF), iii) production of hydrogen ($\rightarrow$H2) through ECH and BIF as well as the Fe-Fe NADPH-dependent hydrogenase (Fe-H2), iv) consumption of pyruvate (PYR$\rightarrow$) through pyruvate-formate lyase (PFL), pyruvate:ferredoxin oxidoreductase (PFOR), and lactate dehydrogenase (LDH), and v) acetyl-CoA flux (ACoA$\rightarrow$) to alcohol dehydrogenase (AdhE), phosphotransacetylase (PTA), and citrate synthase (TCA1) for ethanol, acetate, and cell mass production, respectively.

2.3.4 Flux Balance Analysis

Flux Balance Analysis is a commonly used computational tool using stoichiometric and thermodynamic constraints to optimize a cellular objective, such as maximum cell growth (Varma and Palsson, 1994). FBA was calculated for *i*AT601 using the COBRA (Schellenberger et al., 2011) and RAVEN (Agren et al., 2013) toolboxes within the MATLAB environment (MathWorks, Natick, MA). The algorithm parameters were set to remove any Type III (internal) loops (Price et al., 2002a) within the solution to Equation 1.5. Changes in media recipe were implemented as bounds on nitrogen and sulfur sources. For MTC media (Holwerda et al., 2012), urea and ammonia were available as nitrogen source, while sulfate and cysteine were available as sulfur source. Cysteine uptake was bound at 0.5 mmol/gDCW/hr while all other species were unbound. For Low-carbon (LC) media (Holwerda et al., 2012), cysteine was the sole sulfur source while ammonia was the sole nitrogen source.

2.3.5 Flux Sum Analysis

Cofactor turnover was calculated using Flux-Sum Analysis (FSA) [90]. For any FBA solution, the flux-sum i, or turnover rate, of metabolite i can be represented by

$$\Phi_i = 0.5 \sum_j |s_{i,j} \cdot v_j| \tag{2.13}$$

for reactions j in which the metabolite participates.

2.3.6 Calculation of ATP Costs

To tune the ATP growth-associated maintenance requirement we performed a series of optimizations to fit experimental data from cellobiose-grown batch cultures. Initially, non-growth associated maintenance (NGAM) was set at 3.27 mmol ATP/gDCW/hr (Zhang and Lynd, 2005a), growth associated maintenance (GAM) was varied between 1 and 50 mmol ATP/gDCW/hr (Milne et al., 2011), and the cellulosome ATP requirement for synthesis was identical to the protein term of the biomass reaction (43.28 mmol ATP/g Protein/hr, Thompson et al. (2015)).

To tune the ATP cost of cellulosome synthesis, the coefficient for ATP in the cellulosome synthesis reaction was varied from 40 mmol ATP / g cellulosome / hr (equivalent to the lumped protein synthesis reaction) to 100 mmol ATP / g cellulosome / hr while optimizing for maximal growth and maintaining experimental constraints, similar to above. Tuning these additional ATP requirements allowed finding the best fit to experimentally observed growth rate. All subsequent simulations used these parameters (See Chapter 5).

2.3.7 Sampling of Flux Distributions

Investigation of the difference in ethanol production between culture conditions was performed by constraining ethanol and acetate production as a function of cell growth. In order to implement these constraints, we used a series of flux ratios. First, to

constrain the sum of ethanol and acetate yields, we calculated the following flux ratio

$$Y_{E/G} + Y_{A/G} = -2.9\mu + 1.9 \tag{2.14}$$

$$r_E + r_A = (-2.9\mu + 1.9) * r_{GluEq} \tag{2.15}$$

where the slope and intercept were obtained from the linear relationship in Figure 5.5A. Second, to constrain the ethanol to acetate (E:A) ratio, we similarly calculated the following flux ratio

$$r_E/r_A = m \ * \ \mu \ + \ b \tag{2.16}$$

where the slopes and intercepts for cellobiose and cellulose were obtained from the relationships in Figure 5.5B. Since the experimental data displayed some variance in this parameter, we wanted to ensure a complete representation of cell phenotypes. To accomplish this, we introduced approximately 20% noise into our constraint by randomly varying the slope and intercept.

Given the variability in reported E:A ratios, we chose to sample the phenotypic space of cellobiose and cellulose cultures given the constraints above in order to minimize bias between flux distributions. To perform the sampling, we first randomly generated 100 normally distributed values of between 0 and 0.3 (hr^{-1}) as well as 100 values for the glucose uptake rate between 5.5 and 7.5 mmol Glu Eq / g DCW / hr, the range seen across multiple datasets (See Chapter 5. For each growth rate and glucose uptake rate, the sum of ethanol and acetate yields and E:A ratio were calculated using Equations 2.14 and 2.16. All other fermentation products were unconstrained. The uniform sampling was performed using *optGpSampler* (Megchelenbrink et al., 2014) with a step size of 1000 for each randomly set growth rate. Retaining 1000 flux distributions at each growth rate gave a set of 100,000 flux distributions for comparison between cellobiose and cellulose simulations. Increasing the sample size to 500,000 did not significantly effect the results.

Chapter 3

Enhancing Fatty Acid Ethyl Ester Production in *Saccharomyces cerevisiae* through Metabolic Engineering and Medium Optimization

Summary

This chapter investigates the production of fatty-acid ethyl esters (FAEEs) in *S. cerevisiae* from glucose. The work combines rational metabolic engineering with culture optimization to enhance the production of FAEEs. Using exogenous fatty-acid feeding, the bottleneck in FAEE production was highlighted as a shortcoming in fatty acid synthesis which was assessed in a fed-batch manner to increase FAEE production. This work was published in Biotechnology and Bioengineering (Thompson and Trinh, 2014).

3.1 Introduction

Sustainable energy independence is a defining challenge of our time, and the production of transportation fuels from renewable resources is a promising contribution towards a solution. Biodiesels are an attractive target class of drop-in liquid transportation biofuels due to their physiochemical properties being similar to petrodiesels and their compatibility with existing infrastructure. Current commercial strategies for biodiesel production use plant-based oils which are converted to fatty acid methyl esters (FAMEs) via chemical trans-esterification (Fukuda et al., 2001), although the raw materials needed for this process are not suited for the current scale of our energy usage. An alternative promising strategy has been explored to produce fatty acid ethyl esters (FAEEs) within engineered microorganisms (Kalscheuer et al., 2007). Even though FAMEs and FAEEs are similar in their chemical and physical properties (Röttig et al., 2010), ethanol is more readily produced than methanol from low-cost lignocellulosic biomass feedstocks in a variety of industrial microorganisms, including *Saccharomyces cerevisiae* (Tsai et al., 2009). Thus, FAEEs are seen as an ideal biofuel target for synthesis directly from lignocellulosic, biomass-derived sugars (Figure 3.1).

It has been shown that expression of a bacterial acyl-transferase in *E. coli* (Kalscheuer et al., 2006; Steen et al., 2010) and *S. cerevisiae* (Kalscheuer et al., 2004; Yu et al., 2012) can lead to FAEEs production from the condensation of fatty acyl-CoAs and ethanol (Step 11, Figure 3.1). FAEE biosynthesis in *S. cerevisiae* takes place in the cytosol. The endogenous ethanol production pathway in *S. cerevisiae* is highly active, easily providing the high level of cytosolic ethanol needed for FAEEs biosynthesis, which has been reported as optimal at an ethanol concentration of 3% (Fan et al., 2013). To synthesize cytosolic acyl-CoAs, fatty acids are first synthesized *de novo* in the cytosol of *S. cerevisiae* from acetyl-CoA subunits by acetyl-CoA carboxylase (*acc1*) and the fatty acid synthesis (FAS) complex consisting of *fas1*/*fas2* (Step 8, Figure 3.1). Endogenously synthesized fatty acids are released as acyl-CoAs, the

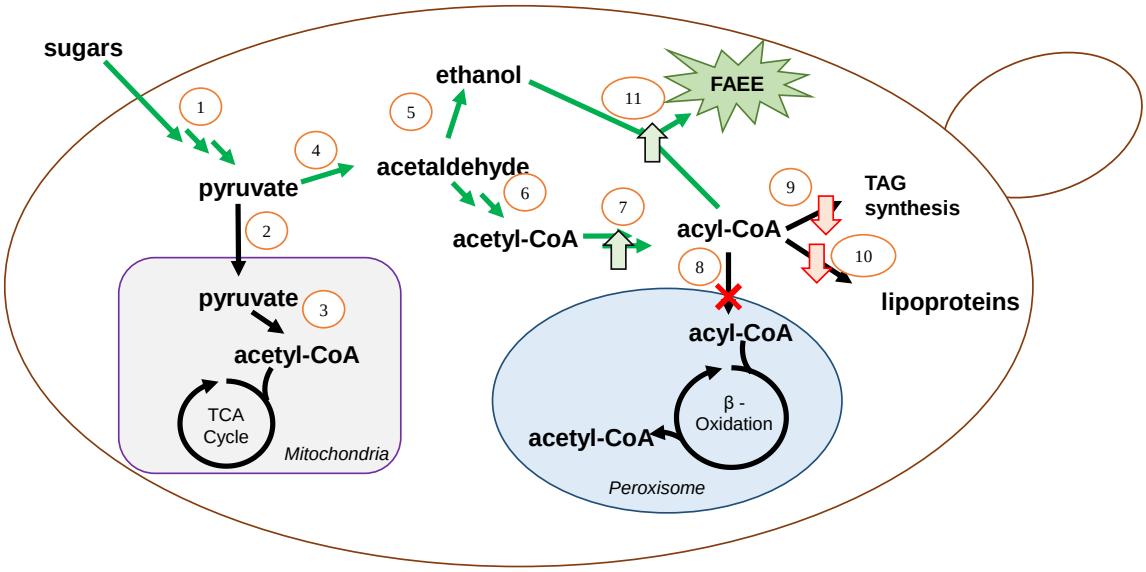

Figure 3.1: Overview of *S. cerevisiae* pathways towards FAEE production from sugars. The pathway of interest is outlined in green with competing pathways shown in black. Enzyme targets for metabolic engineering are accentuated with appropriate symbols. Enzyme and pathway names: (1) glycolysis/pentose phosphate pathway, (2) mitochondrial pyruvate carrier, (3) pyruvate dehydrogenase, (4) pyruvate decarboxylase, (5) alcohol dehydrogenase, (6) aldehyde dehydrogenase/ acetyl-CoA synthase, (7) fatty acid synthesis (FAS) complex, (8) peroxisomal acyl-CoA transporter, (9) lipid synthesis pathways, (10) protein acylation pathways, and (11) alcohol acyltransferase.

precursors for FAEEs biosynthesis (Tehlivets et al., 2007). Since the total lipid content of *S. cerevisiae* is mostly comprised of C16 and C18 fatty acids under normal growth conditions (Trotter, 2001), these fatty acyl-CoAs are suitable for biodiesel production (Westfall and Gardner, 2011). Thus, *S. cerevisiae* is a promising candidate for metabolic engineering to enhance FAEEs production from lignocellulosic biomass-derived fermentable sugars.

Recently, several strategies have been reported for metabolically engineering *S. cerevisiae* for enhanced FAEE production. One strategy was to manipulate the fatty acid biosynthesis pathway by overexpressing or otherwise altering the regulation of *acc1* and FAS proteins (Runguphan and Keasling, 2014; Shi et al., 2012). With this strategy, the highest titer of FAEEs produced directly from glucose were reported to

be about 5.4 mg/L. The other strategy aimed at eliminating the competing fatty acid pathways such as degradation through -oxidation or storage in the form of triacylglycerides (TAGs) or sterol esters (SEs) (Runguphan and Keasling, 2014; Valle-Rodrguez et al., 2014), with the highest titer reaching about 17 mg/L.

In this study, combined strategies were used to increase cytosolic acyl-CoA pools available for FAEEs biosynthesis by disrupting the acyl-CoA transport to the β-oxidation pathway and TAG biosynthesis. Next, nitrogen limited culture conditions were employed to push carbon towards fatty acyl-CoAs production. Codon optimized acyltransferase genes were also designed to pull the acyl-CoAs towards FAEEs biosynthesis. Taken altogether, the rate liming step that can be overcome to enhance FAEEs production in *S. cerevisiae* has been clarified.

3.2 Results and Discussion

3.2.1 Metabolic engineering of *S. cerevisiae* to increase cytosolic acyl-CoA concentration

The inherent nature of *S. cerevisiae* to produce large quantities of ethanol directed our metabolic engineering strategy towards increasing the availability of the other substrate in FAEEs synthesis, acyl-CoAs, in the cytosol by eliminating competing acyl-CoA utilization pathways. This is in contrast to other reports which aim to increase acyl-CoA concentration by overexpressing genes involved in fatty acid synthesis (Runguphan and Keasling, 2014; Shi et al., 2012). Our strategies include

i Disrupting the transportation of acyl-CoAs into the peroxisome to eliminate β-oxidation (Step 8, Figure 3.1), and

ii Deleting the acyl-CoA binding protein that shuttles acyl-CoAs for TAG or lipoprotein biosynthesis (Steps 9 and 10, Figure 3.1).

Effect of disrupting acyl-CoAs transportation into the peroxisome on FAEEs production

Abundantly available cytosolic acyl-CoAs are transported into the peroxisome for degradation. The import is accomplished by a heterodimer encoded by the genes *pxa1* and *pxa2* with the help of an acyl-CoA synthase encoded by *faa2*, and β-oxidation can be essentially halted in strains with mutations in *faa2* and either of the *pxa* genes (van Roermund et al., 2012). Therefore, to enhance the cytosolic acyl-CoAs, *faa2* and *pxa2* were sequentially deleted.

We characterized the wildtype (ScAT2200), *faa2* mutant (ScAT2203), *pxa2* mutant (ScAT2204), and *faa2*/*pxa2* double mutant (ScAT2208) carrying pAT01y in batch shake flasks for FAEE production (Figure **??**). After 48 hours, the wildtype ScAT2200 produced 1.4±0.2 mg/L FAEEs while the single mutants ScAT2203 and ScAT2204 produced 1.8±0.1 mg/L FAEEs (33% improvement) and 1.9±0.1 mg/L FAEEs (35% improvement), respectively. Double mutant ScAT2208 slightly outperformed the single mutants with 51% improvement compared to the wildtype.

Our strategy to disrupt acyl-CoA transportation into the peroxisome and hence the β-oxidation is successful in enhancing the FAEEs production. This strategy is different from the previous reports which remove β-oxidation activity from the cells by deleting the peroxisomally located acyl-CoA oxidase gene *pox1* (Hiltunen et al., 2003; Runguphan and Keasling, 2014; Valle-Rodrguez et al., 2014).

Effect of disrupting acyl-CoA binding protein on FAEE production

Within all studied eukaryotes there is an acyl-CoA binding protein located in the cytosol which shuttles newly synthesized or exogenously sourced acyl-CoA molecules to downstream processes such as triacylglyceride synthesis or β-oxidation (Black and DiRusso, 2007). To enhance the cytosolic acyl-CoAs, we deleted the gene *acb1* encoding for the acyl-CoA binding protein in the single mutant ScAT2205. Strain characterization in batch shake flasks shows that ScAT2205 produced 2.1±0.1 mg/L

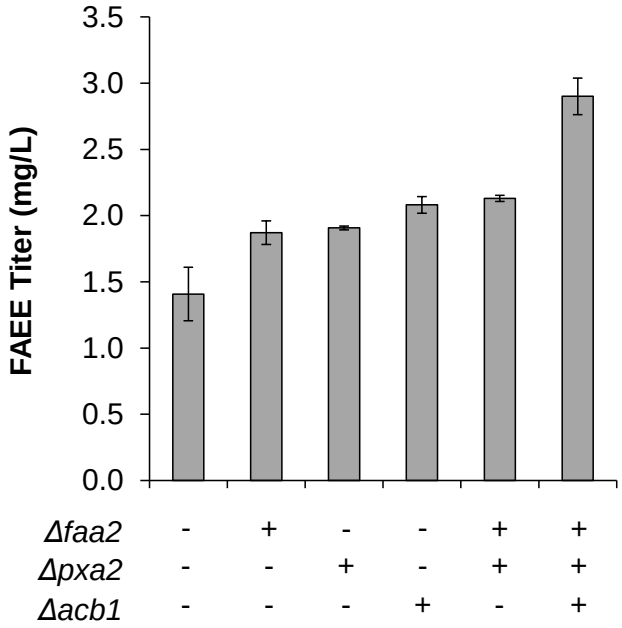

Figure 3.2: FAEE titer after 48 hours for strains ScAT2200, ScAT2203, ScAT2204, ScAT2205, ScAT2208, and ScAT2209 (from left to right). All strains contain pAT01y which constitutively expresses the acyl-transferase *atfA* from *Acinetobacter* sp. ADP1 and were grown under normal (non-nitrogen limiting) conditions.

FAEEs, about 48% improvement in comparison with the wildtype (Figure 3.2). This strain had similar performance as the double mutant ScAT2208.

Our strategy to delete *acb1* for enhanced FAEEs production is consistent with the observed phenotype of *acb1* mutant reported in the literature. Deleting *acb1* increased transcription levels of genes related to the fatty acid biosynthesis, such as *acc1*, *fas1*, and *fas2*, by roughly two folds, causing the accumulation of vesicles (Feddersen et al., 2007). In addition, with deletion of *acb1*, cytosolic levels of free acyl-CoA molecules increase up to 2.5 folds (Schjerling et al., 1996).

Synergistic effect of metabolic engineering strategies for FAEE production

The above two strategies were combined to further increase acyl-CoAs and hence FAEEs production by constructing the triple *faa2*/*pxa2*/*acb1* mutant ScAT2209.

This triple mutant produced 2.9±0.1 mg/L, accounting for 2.1 folds improvement over the wildtype under similar characterization conditions.

The increase in titers with an increase in genetic modifications is a clear indicator of cooperative enhancement between these strategies, and that removing competitive acyl-CoA consumption pathways is beneficial for enhanced FAEEs production. This level of endogenous FAEEs production, although comparable to other reports, is still lower than necessary for large scale, economically sustainable petroleum replacement. This fact prompted us to look for alternative approaches in culture conditions.

3.2.2 Manipulating Nitrogen Limiting Conditions for Enhanced FAEE Production

Changing the elemental balance of the culture media alone has been shown to significantly enhance lipid accumulation in *S. cerevisiae* (Kamisaka et al., 2007). Thus, we cultured our strains harboring *atfA* under nitrogen limiting conditions in an attempt to drive FAEE production. There was an immediate observable increase in the amount of FAEEs produced compared to previous culture conditions, corresponding to an over 3.5 fold increase in each strain investigated. Further, there was a dramatic improvement of roughly 3-fold in production in ScAT2209 over each of the other strains tested to a final titer of 11.7±0.7 mg/L (Figure 3.3A).

Most interesting was the fact that the wild-type strain carrying a blank plasmid, ScAT2000, was capable of producing statistically similar amounts of FAEEs to the *atfA* expressing strains without all of the *faa2*, *pxa2*, and *acb1* gene disruptions. This implies there is a synergistic effect of the two strategies we implemented for FAEE production under nitrogen limiting growth conditions. These results also indicate the presence of an endogenous method for producing FAEEs under nitrogen limiting conditions, which is not seen in previous reports that did not grow the yeast under nitrogen limitation.

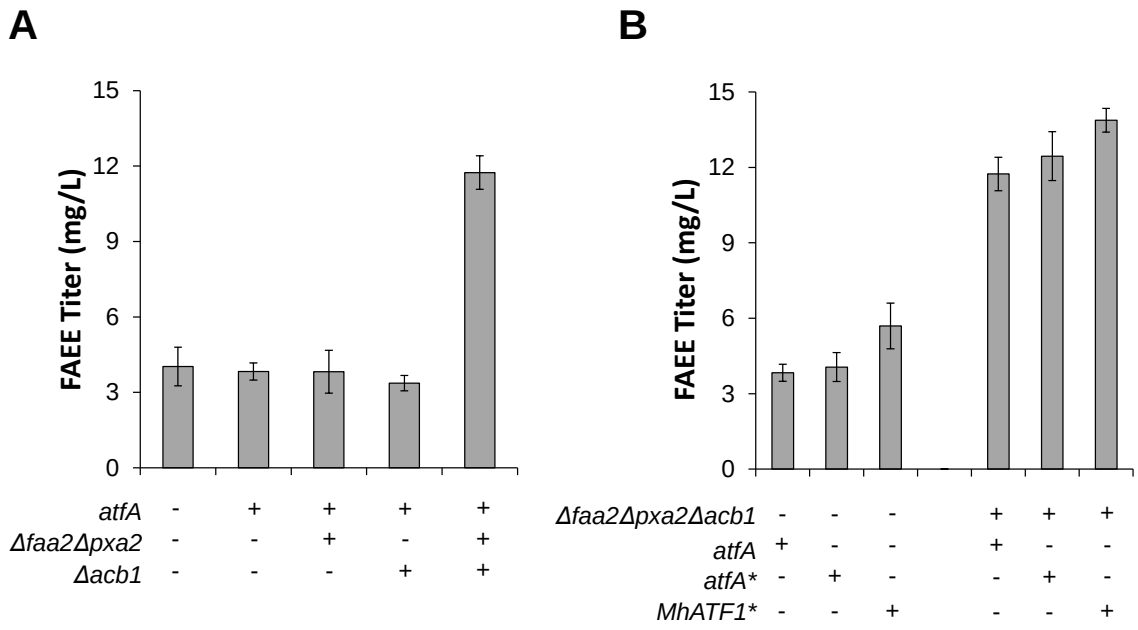

Figure 3.3: **(A)** Characterization of strains ScAT2000, ScAT2200, ScAT2205, ScAT2208, and ScAT2209 (from left to right) harboring a blank control plasmid or pAT01y cultured under nitrogen limiting conditions for FAEEs production after 48 hours. **(B)** Comparison of FAEE titers between strains expressing atfA or codon optimized genes *atfA** or *MhATF1** for both the wild-type background (BY4741) and the triple mutant background (ScAT1031) under nitrogen limiting conditions.

One possible explanation for this phenotype is the activity of the esterases encoded by *eeb1* and *eht1*, which have been known to synthesize medium chain length (C6-C10) FAEEs (Saerens et al., 2006). It is possible that the expression of these enzymes is enhanced during nitrogen limiting growth conditions, and this hypothesis is further strengthened when examining the chain length distribution of the products under the different growth conditions. The major products observed above and in previous studies were C16 and C18 FAEE, while these conditions led to a majority of FAEE products to be of length C10 and C12 (Figure 3.4 A). These results prompted two hypotheses that are explored below. First, *atfA* was the limiting step in FAEE biosynthesis, as it prefers longer chain alcohol and fatty acyl-CoA molecules as substrates (Barney et al., 2012). Second, acyl-CoAs were still limiting for FAEEs biosynthesis despite the efforts described so far.

3.2.3 Deploying Codon Optimized Acyl-Transferase Genes for Enhanced FAEE Production

To investigate if *atfA* activity was the limiting step in FAEEs biosynthesis, *S. cerevisiae* codon-optimized sequences of *atfA** and *MhATF1** from *Marinobacter hydrocarbonoclasticus* DSM 8798 were synthesized and incorporated into constitutive expression plasmids controlled by the TEF promoter to create plasmids pAT27y and pAT28y, respectively (Table 2.1). When cultured under nitrogen limiting conditions and with a 10% dodecane overlay, the wildtype strains ScAT2211 carrying *atfA** and ScAT2212 carrying *MhATF1** produced 4.0±0.6 and 6.0±0.8 mg/L FAEEs, respectively, after 96 hours (Figure 3.3B). Recalling a final titer of 3.8±0.3 mg/L for ScAT2200, this leads to a slight increase from ScAT2211, harboring the codon optimized *atfA**, and a 49% increase from ScAT2212, expressing *MhATF1**. These results are consistent with previous reports which used codon-optimized wax synthase genes and extensive genetic disruptions without nitrogen limitation (Runguphan and Keasling, 2014; Shi et al., 2012). They also show that the dramatic effect of culture conditions alone on FAEE production is comparable to other metabolic engineering strategies.

Furthermore, we combined codon-optimized genes, culture conditions, and our genetic disruption strategies. The non codon-optimized triple mutant strain ScAT2209 produced a final FAEE titer of 11.7±0.7 mg/L, where ScAT2215 and ScAT2216 produced 12.4±1.0 and 13.9±0.5 mg/L, respectively (Figure 3.3B). This shows expression of *MhATF1** in ScAT2216 gives an 18% increase in the top producing strain compared to ScAT2209, and a 2.4-fold improvement over the top producing strain with a wild-type background, ScAT2212. Taken all together, acyltransferases were limiting and FAEEs production can be significantly improved when codon-optimized genes *atfA** and *MhATF1** were implemented together with genetic disruption strategies and optimized culture conditions.

3.2.4 Identification of the Rate Limiting Step to Enhancing Acyl-CoA Pools

To investigate further if fatty acyl-CoA molecules are the limiting factor in FAEEs production, odd chain fatty acid (pentadecanoic acid) was added to the culture as a traceable marker for FAEEs (ethyl pentadecanoate) biosynthesis. The free fatty acid should be taken up in conjunction with glucose and converted to acyl-CoA upon entering the intracellular environment (Black and DiRusso, 2007). This should provide a steady stream of substrate for FAEEs production. In fact, we observe a steady increase in C15-FAEE over time in ScAT2216 (Figure 3.4A) with a maximum titer of 3.6±0.6 mg/L of C15 FAEE in ScAT2216. This corresponds to about 22% of total FAEE produced by ScAT2216. Interestingly, the percent composition of the various FAEE species changes compared to the studies without fatty acid addition (Figure 3.4B).

There is a clear decrease in medium chain length FAEEs, which can be explained by down-regulation of fatty acid synthesis complex proteins due to the addition of exogenous fatty acids (Chirala, 1992). However, the increase in C16- and C18-FAEEs compared to nitrogen limiting conditions without exogenous fatty acid supplementation cannot be fully explained in this manner. The addition of fatty acids into the media may be causing regulatory action by the cells so they increase the turnover of TAG lipid storage. The production of C15-FAEE under these conditions does not halt over time and gives further evidence to the hypothesis that acyl-CoA molecules are limiting in FAEEs biosynthesis. It is also worth noting that C15-FAEE is the only odd chain molecule we observe, giving additional confidence in our genotypes inability to degrade fatty acids via β-oxidation. This result indicates not only acyl transferase was limiting but also acyl-CoAs, hindering FAEE production.

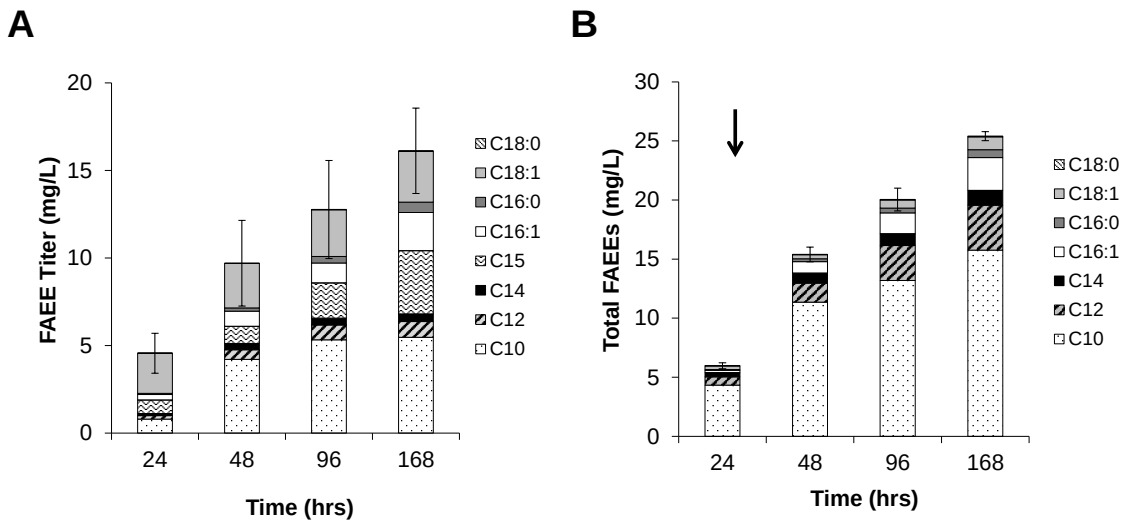

Figure 3.4: **(A)** Characterization of the strain ScAT2216 constitutively expressing *MhATF1** for FAEEs production in nitrogen limiting cultures with the addition of 0.1% pentadecanoic acid. **(B)** FAEE titers from fed-batch cultures of strain ScAT2216, constitutively expressing *MhATF1**. Additional glucose and salts were added after 24 hours, indicated by an arrow, as described in Section 2.1.3. FAEE composition is indicated by the legend.

3.2.5 Increasing FAEE Production Under Fed-Batch Conditions

FAEE production in each of the studies above sharply increased in the first 48 hours of culture, but only slightly climbed for the remainder of culture time. Time-course data of cell growth shows that all glucose was consumed by the cells after 24 hours (Figure 3.5). These data indicate that acyl-CoAs are being generated directly from glucose as opposed to fatty acid recycling pathways. Therefore, to further increase FAEE production, we employed a fed batch culture strategy. For our study we replenished glucose and minimal ammonium after 24 hours to reach a total C:N ratio of 100:1 (see Section 2.1.3). In ScAT2216, the titer was increased more than twice to a maximum titer of 25.4±0.4 mg/L FAEEs (Figure 3.4B). These results culminate an increase of over 17 fold in endogenous FAEE production from our initial

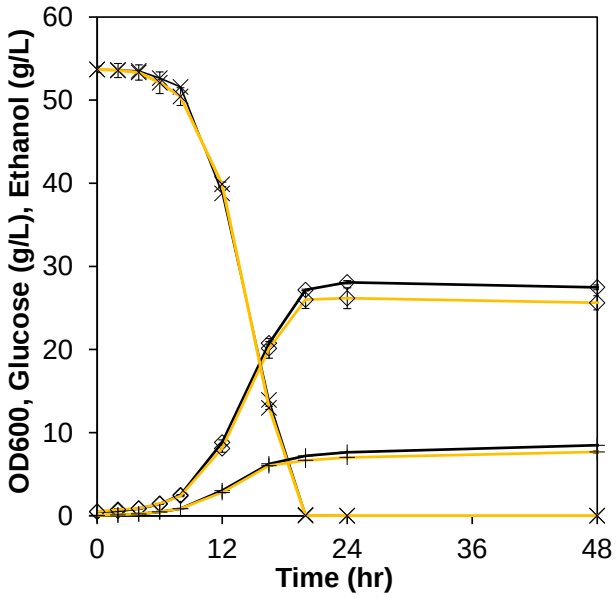

Figure 3.5: Glucose consumption (X), ethanol production (diamonds), and cell growth (+) for the wild-type ScAT2200 (black) and the triple mutant ScAT2209 (yellow) cultured in Nitrogen limited conditions.

characterizations, as well as a roughly 40% increase in titer compared to the highest previously reported results (Valle-Rodrguez et al., 2014).

3.3 Conclusion

In this chapter, it is shown that a genetic disruption strategy aimed at diminishing various acyl-CoA consumption pathways is sufficient to increase FAEEs production compared to the wild-type in *S. cerevisiae* expressing a bacterial acyl-transferase. Further, we have demonstrated that our metabolic engineering strategies can be enhanced dramatically by altering culture conditions, reaching a maximum titer of over 25 mg/L FAEEs, a 40% improvement over previous reports, and a 17 fold improvement over our initial strain characterization. Very recently, it was reported that elimination of oxidation and TAG synthesis can lead to an increase in FAEEs production to comparable levels of about 17 mg/L (Valle-Rodrguez et al., 2014). This

strategy, if combined with our strategy of *acb1* elimination, should further increase FAEEs production.

Looking forward, we can immediately see ways to further improve FAEEs titer. In all of our studies, ethanol is produced at a high yield (0.49±0.02 g/g glucose). This carbon flux away from fatty acid production is clearly detrimental to FAEE yield, and so a balance must be reached between ethanol and acyl-CoA flux to optimize FAEE production. This level of metabolic engineering in *S. cerevisiae* will not be trivial but it is arguably necessary to reach levels seen in other production systems (Zhang et al., 2012). We expect that altering transcription of acyltransferase genes by manipulating promoter strength (Blazeck et al., 2012) can be instrumental in this effort.

It is also known that atfA can catalyze the synthesis of triacylglycerides (TAGs) from a diacylglyceride and an acyl-CoA molecule (Stoveken et al., 2005) and recently has been shown to lead to an overproduction of TAGs in *E. coli* (Janßen and Steinbüchel, 2014). This adds additional complications for optimizing FAEEs production and implies that yield of FAEEs can be diminished by side reactions of the acyltransferase. These reports lead to questions regarding the applicability of protein engineering in this case to facilitate higher production of the desired target. Significant efforts are being made to understand the critical domains (Villa et al., 2013) and residues (Barney et al., 2013) for this class of enzyme, and so engineered acyltransferases are to be implemented soon.

Chapter 4

Elucidating Central Metabolic Redox Obstacles Hindering Ethanol Production in *Clostridium thermocellum*

Summary

In this chapter, a predictive stoichiometric metabolic model for *C. thermocellum* is presented which incorporates the current state of understanding. Particular attention is paid to cofactor specificity in the atypical glycolytic enzymes and the complex energy, redox, and fermentative pathways with the goal of aiding metabolic engineering efforts. We validated the model's capability to encompass experimentally observed phenotypes for the parent strain and derived mutants designed for significant perturbation of redox and energy pathways. Metabolic flux distributions revealed significant alterations in key metabolic branch points (e.g., phosphoenolpyruvate, pyruvate, acetyl-CoA, and cofactor nodes) in engineered strains for channeling electron and carbon fluxes for enhanced ethanol synthesis, with the best performing

strain doubling ethanol yield and titer compared to the parent strain. *In silico* predictions of a redox-imbalanced genotype incapable of growth were confirmed *in vivo*, and a mutant strain was used as a platform to probe redox bottlenecks in the central metabolism that hinder efficient ethanol production. The results highlight the robustness of the redox metabolism of *C. thermocellum* and the necessity of streamlined electron flux from reduced ferredoxin to NAD(P)H for high ethanol production. The model was further used to design a metabolic engineering strategy to phenotypically constrain *C. thermocellum* to achieve high ethanol yields while requiring minimal genetic manipulations. The model can be applied to design *C. thermocellum* as a platform microbe for consolidated bioprocessing to produce ethanol and other reduced metabolites. This work was published in Metabolic Engineering (Thompson et al., 2015)

4.1 Introduction

Sustainable energy production is a defining challenge for modern society, and achieving this goal will depend on the ability to produce fuels and chemicals from renewable sources. Consolidated bioprocessing (CBP) (Demain, 2009) has the appeal of combining the steps of hydrolytic enzyme production, feedstock solubilization, and fermentation of multiple sugars into a single step (van Zyl et al., 2007), with the goal of maximizing cost-efficiency of lignocellulose conversion (Lynd et al., 2008).

C. thermocellum is an anaerobic, thermophilic, gram-positive bacterium which has been considered for CBP due to its strong capability to ferment lignocellulosic feedstocks quickly while producing ethanol and hydrogen (Lynd et al. (2005); Ragauskas et al. (2014)). However, wild-type *C. thermocellum* produces a wide range of fermentative products (e.g. acetate, formate, lactate) as well as amino acids, which limits the yield of ethanol (Ellis et al., 2012). Also, the wild-type cannot grow at the high concentrations of ethanol that are industrially necessary without extensive directed evolution (Herrero and Gomez (1980); Brown et al. (2011); Shao et al.

(2011)). Progress in engineering *C. thermocellum* for increased ethanol production has been made in several ways: isolating a high ethanol-tolerant mutant via adaptive evolution (Williams et al., 2007), elimination of competing carbon fermentative pathways (e.g., acetate (Tripathi et al., 2010), lactate (Biswas et al., 2014), and both lactate and acetate (van der Veen et al., 2013) with directed evolution), expressing an exogenous pyruvate kinase from *Thermoanaerobater saccharolyticum* and deleting malic enzyme (Deng et al., 2013), as well as eliminating a major competitor for electrons, H_2 production (Biswas et al., 2015). None of these strategies, however, has been successful in reaching the goal of 90% conversion of fermentable sugars to ethanol as seen in non-cellulolytic mesophilic organisms such as *Saccharomyces cerevisiae* (Kuyper et al., 2005) and *Escherichia coli* (Trinh et al. (2008); Yomano et al. (1998)).

An important obstacle to metabolic engineering of *C. thermocellum* for enhanced biofuel production is an incomplete knowledge of its metabolism. Recent studies suggest that *C. thermocellum* possesses an atypical central metabolism. For instance, the glycolysis does not have an annotated pyruvate kinase (Deng et al., 2013). Fructose-6-phosphate kinase uses pyrophosphate (PPi) instead of ATP, and GTP can substitute for ATP in several reactions which might have important implications for metabolism (Zhou et al., 2013). The pentose phosphate pathway lacks several enzymes in the oxidative branch which implies the need for alternative pathways for NADPH synthesis (Rydzak et al., 2012). The Krebs cycle is also incomplete with the lack of succinate dehydrogenase (Roberts et al., 2010). In addition, *C. thermocellum* has many transhydrogenase reactions involving ferredoxin, NADH, and / or NADPH.

In this work, the current understanding of *C. thermocellum* metabolism was consolidated into a predictive metabolic network model used for metabolic flux quantification and rational strain design. We characterized engineered *C. thermocellum* strains designed and constructed with significant perturbations of energy and redox pathways to validate the model. Redox perturbation involved partial and full disruption of H_2 production while energy perturbation disrupted acetate production.

Their unique phenotypes were analyzed by quantitative metabolic flux distributions. Next, we used the model to elucidate the key bottleneck in the redox metabolism of *C. thermocellum* hindering high-yield ethanol production. Finally, the validated model was used as an input for elementary mode analysis (Trinh et al., 2009) to design an optimal genotype for efficient ethanol production in *C. thermocellum*.

4.2 Results

4.2.1 Quantitative analysis of cellular phenotypic states

First characterized was the *C. thermocellum* strain DSM 1313 Δ*hpt* (henceforth denoted "parent strain). Similar to literature reports, the parent strain produced a mixture of fermentative products and amino acids (with valine as the most dominant) (Figure 4.1A) with a good carbon recovery (Figure 4.1B, Table 4.1). Due to the mixed fermentation, the ethanol yield (0.16 ± 0.01 g ethanol g^{-1} cellobiose) was low, as expected.

From the time-dependent concentration data, we extracted experimental fluxes for measured metabolites (Table 4.1), and calculated the metabolic flux distribution of the parent strain. Linear regression between the experimental and calculated fluxes gave an excellent fit ($R^2 = 0.96$, Figure 4.2), indicating that the calculated flux distribution vector is a reasonable assessment of the central metabolism of *C. thermocellum*. In the subsequent sections, flux distributions of various *C. thermocellum* strains were calculated for quantitative analysis of their cellular physiological states, and were used to identify reaction bottlenecks hindering efficient ethanol production. For an in depth summary of the construction of the core metabolic model, see Section 2.2.1.

We used METAFoR analysis to examine the resulting flux distribution at key nodes more closely. Flux from pyruvate is determined directly by monitoring formate and lactate production. Additionally, the fate of acetyl-CoA in our model is

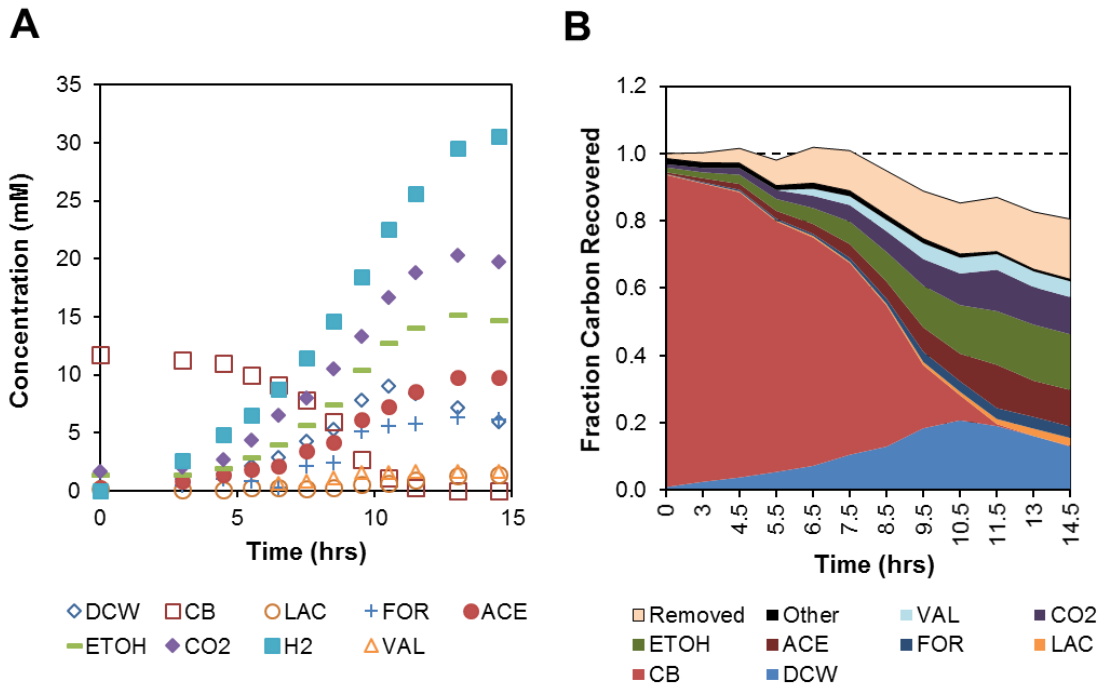

Figure 4.1: (A) A concentration profile for the major metabolic end-products of the parent strain in a representative batch fermentation. **(B)** A representative carbon-balance applied throughout a parent strain batch fermentation. "Removed" indicates carbon removed during sampling while "other" consists of trace amino acids, mostly alanine and isoleucine. Abbreviations: CDW, cell dried weight; CB, cellobiose; LAC, lactate; FOR, formate; ACE, acetate; ETOH, ethanol; CO2, carbon dioxide; H2, hydrogen; and VAL, valine

determined by fermentation products and cell mass. However, the flux to pyruvate from PEP and flux through the redox pathways are only constrained by H_2 production which might not be effectively constrained and require additional experimental evidence (e.g., enzyme activities and proteomics). The best fit of experimental fluxes onto the EM space predicts that over 99% of flux to pyruvate proceeds through ODC and MAE combined, with the small remainder proceeding through PPDK (Figure 4.3B). One-at-a-time sensitivity (OATS) analysis shows no effect when the malate shunt is deleted (Table 4.2), but this calculated flux distribution is consistent with high enzymatic activity in crude cell extracts (Zhou et al., 2013).

For H_2 flux, the best fit predicts over 90% proceeds through the ECH hydrogenase. This observation makes sense because, under high gas flow rate that was used in

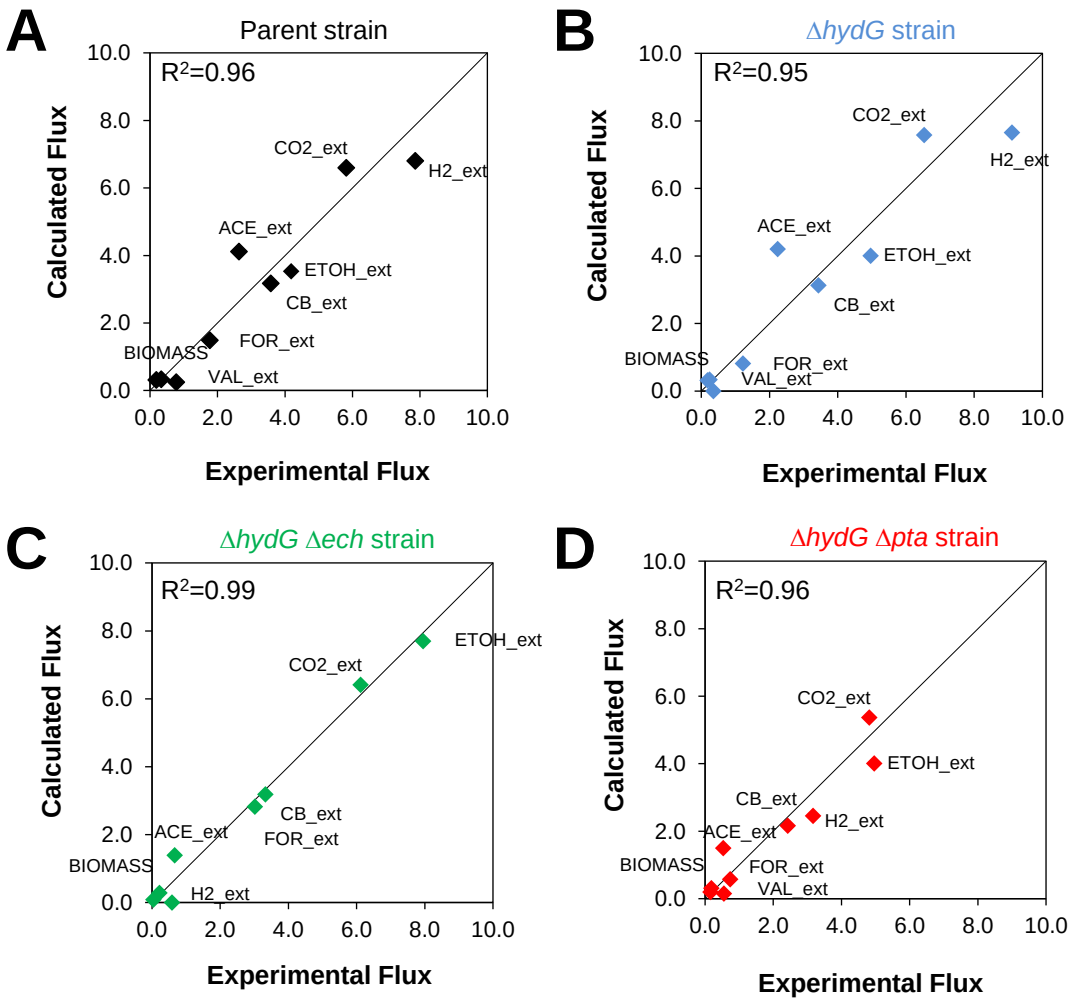

Figure 4.2: Comparison between the experimental and calculated fluxes for the parent strain **(A)**, Δ*hydG* strain **(B)**, Δ*hydG*Δ*ech* strain **(C)**, and Δ*hydG*Δ*pta* strain **(D)**.

the experiments, the soluble H_2 was low and hence ECH was probably the major hydrogenase used in a thermodynamically favorable reaction. Since the network has many local redundancies around H_2 to allow for meaningful stoichiometric analysis for the parent strain, we also performed the OATS analysis to investigate the network structure's ability to give equivalent solutions. The results show that individual deletions of ECH, RNF, NFN, Fe-H2, or BIF have no effect on the error of the experimental data fitting (Table 4.2). This suggests that for the wildtype the

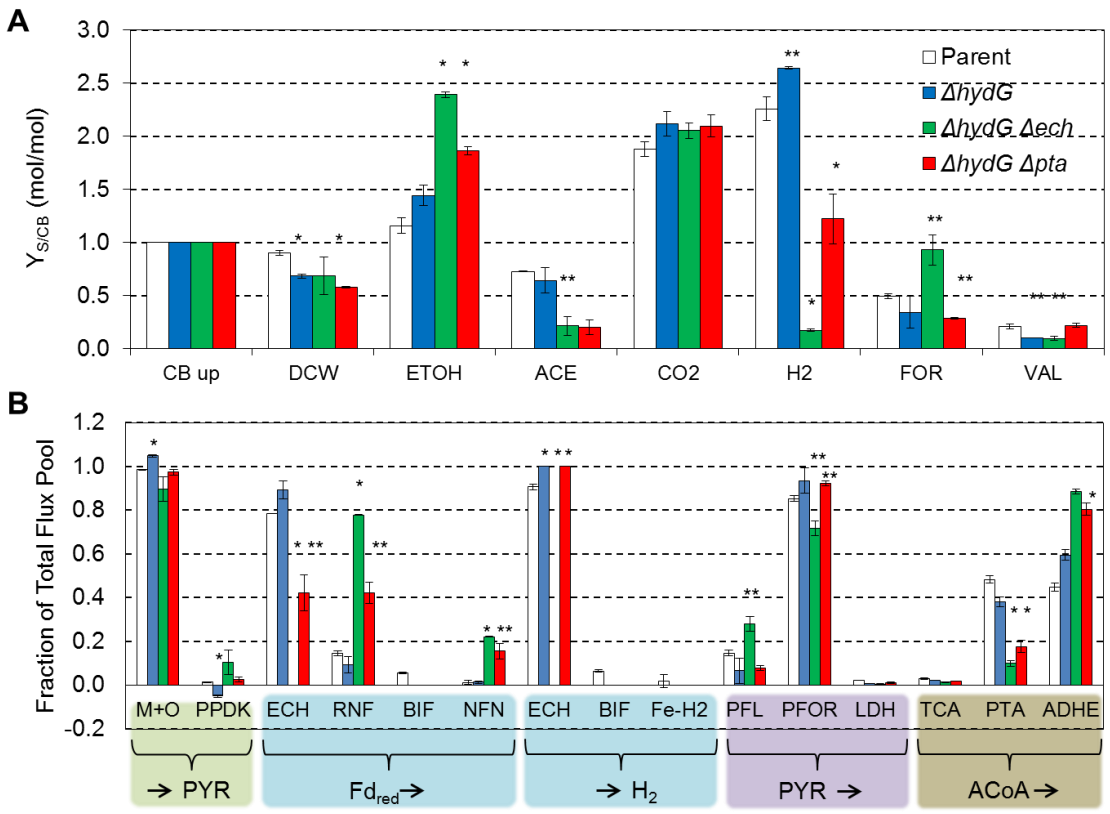

Figure 4.3: **(A)** Experimental yields of major fermentative products on cellobiose for different genotypes. **(B)** Fraction of contribution of fluxes through specific metabolic nodes via METAFoR analysis for different genotypes. The groups are based on the metabolic nodes to which they contribute (Methods). Abbreviations: M+O, malic enzyme and oxaloacetate decarboxylase are collectively considered; PPDK, pyruvate phosphate dikinase; ECH, Ni-Fe Ech-type hydrogenase; RNF, NADH:Fd oxidoreductase; NFN, NADH-Fdrd:$NADP^+$ oxidoreductase. BIF, bifurcating hydrogenase; Fe-H2, NADPH dependent Fe-Fe hydrogenase; PFL, pyruvate formate lyase; PFOR, Pyruvate:Fd oxidoreductase; LDH, lactate dehydrogenase; TCA, citrate synthase (towards cell mass); PTA, acetate synthesis; and ADHE, ethanol synthesis. In both panels, student t-test significance compared to the parent strain is reported by asterisks, i.e., * for p-value < 0.01, and ** for p-value < 0.05

Table 4.1: Experimentally determined growth rates, metabolic fluxes, and yields per cellobiose for selected fermentative products for the parent, Δ*hydG*, Δ*hydG* Δ*ech*, and Δ*hydG* Δ*pta-ack* strains. Experimental data were obtained through pH controlled bioreactors. n.d. − Not detected

Fermentation parameters	Parent	Δ*hydG*	Δ*hydG* Δ*ech*	Δ*hydG* Δ*pta*
μ (hr^{-1})	0.33±0.01	0.24±0.02	0.22±0.02	0.16±0.02
Ethanol yield (g g^{-1})	0.16±0.01	0.19±0.01	0.32±0.04	0.25±0.01
Ethanol titer (g L^{-1})	0.63±0.03	0.79±0.01	1.27±0.05	1.01±0.07
% Carbon Recovery	90.5±1.6	79.3±1.6	89.9±9.2	83.5±4.1
Experimental flux (mmol g^{-1} DCW hr^{-1})				
r_{CBup}	3.58±0.16	3.43±0.25	3.33±0.62	2.42±0.05
r_{ETH}	4.19±0.10	4.96±0.24	7.95±1.37	4.65±0.25
r_{ACE}	2.63±0.86	2.24±0.57	0.67±0.17	0.53±0.21
r_{CO2}	6.82±0.01	7.25±0.11	6.80±1.01	5.29±0.50
r_{FOR}	1.77±0.01	1.22±0.60	3.01±0.56	0.74±0.06
r_{H2}	7.86±0.38	9.11±0.84	0.58±0.06	3.17±0.84
r_{VAL}	0.78±0.12	0.35±0.01	0.31±0.01	0.55±0.07
r_{LAC}	0.18±0.01	n.d.	0.05±0.01	0.03±0.01

redundancies in the redox pathway can compensate for individual gene deletions, and that genetically perturbing the system is necessary to obtain additional insights.

4.2.2 Metabolic model illustrates robust redox metabolism

Effect of partial hydrogen deletion

We characterized previously constructed strains with disrupted H_2 production in order to elucidate the effect of redox perturbation on *C. thermocellum* metabolism and ethanol production (Biswas et al., 2015). The first characterized mutant was the Δ*hydG* strain containing a disruption of the maturase enzyme complex, which inactivates the three Fe-Fe hydrogenases in *C. thermocellum* (Feinberg et al. (2011); Mulder et al. (2011)) (Figure 2.1). This strain also acquired a spontaneous mutation in the bifunctional alcohol/aldehyde dehydrogenase *adhE*, which allowed it to not only utilize NADH as an electron donor for acetaldehyde reduction to ethanol, but also NADPH (Biswas et al., 2015). Deletion of *hydG* caused a 40% drop in the

Table 4.2: One-at-a-time sensitivity (OATS) analysis for select reactions. "n/a" designates reactions which were removed prior to sensitivity analysis due to genotype.

		% Change in fitting error upon reaction removal			
Nodes of interest	Reactions	Parent	Δ*hydG*	Δ*hydG* Δ*ech*	Δ*hydG* Δ*pta*
→PYR	MAE+ODC	0	0	0	0
	PPDK	0	0	0	0
PYR→	PFOR	228	163	583	288
	PFL	28	4	245	10
Electrons→	ECH	0	170	n/a	113
	Fe-H2	0	n/a	n/a	n/a
	BIF	0	n/a	n/a	n/a
	RNF	0	0	32	0
	NFN	0	0	0	0
Products→	Cell Growth	16	11	31	13
	$ETOH_{out}$	114	71	764	246
	ACE_{out}	286	176	226	123
	$CO2_{out}$	309	221	717	374
	FOR_{out}	28	4	245	10
	$H2_{out}$	267	170	n/a	113

specific growth rate compared to the parent strain, and a lower yield of cell mass on cellobiose (Table 4.1). While the cellobiose uptake flux was minimally affected, the ethanol flux and yield on cellobiose increased 18% and 24%, respectively, and this slight increase in ethanol production was accompanied by a corresponding drop in acetate production (Figure 4.3), presumably due to greater electron availability for ethanol synthesis while lowering the acetyl-CoA pool used for acetate production.

Most interestingly, H_2 yield increased ∼17% compared to the parent strain (Figure 4.3A), which resulted in a very limited increase in ethanol production. This phenotype was unexpected and conflicts with previous reports of serum bottle cultures of the Δ*hydG* strain having a significant drop in H_2 yield (Biswas et al., 2015). This discrepancy can be explained by culture conditions, as we used a venting

mechanism to measure gas concentrations in real-time and controlled pH, which prevents accumulation of H_2 and likely increases flux through the ECH hydrogenase.

To gain insights into the redox metabolism of *C. thermocellum*, we examined the flux distributions of the Δ*hydG* strain, which matched well with the experimental data ($R^2 = 0.99$, Figure 4.4). A parametric plot of normalized flux distributions with respect to the cellobiose uptake fluxes shows no significant difference between the parent and Δ*hydG* strains ($R^2 = 0.98$, Figure 4.4A). Outliers of this plot demonstrate the discrepancies of fluxes between the parent strain and the mutants. The only significant outlier was the NADPH-dependent AdhE* reaction, which is known not to be active in the parent strain (Lamed and Zeikus, 1980). Additionally, METAFoR analysis suggests that the malate shunt was still the major pathway to channel the carbon flux from PEP to pyruvate, and the fate of acetyl-CoA was minimally affected by the deletion of *hydG* (Figure 4.3B). METAFoR analysis also shows that H_2 production can be compensated by the major ECH hydrogenase activity for effective electron shuttling in the absence of HydG-related enzyme activity. OATS analysis shows that removal of ECH from the Δ*hydG* network leads to a dramatic 170% change in error (Table 4.2).

Effect of complete deletion of hydrogen production

To characterize a dramatic push of electron flux towards enhanced ethanol production, we investigated the dual deletion Δ*hydG* Δ*ech* strain. The results show this mutant produced little H_2 and enhanced ethanol yield by 2.1 fold, while having similar growth characteristics as the Δ*hydG* strain (Table 4.1). Similar to the phenotype of the Δ*hydG* strain, the increase in ethanol production of the Δ*hydG* Δ*ech* strain was accompanied by decreased acetate production. However, one striking phenotype was that the formate yield, not the lactate yield, increased two-fold (Figure 4.3A).

To elucidate how the Δ*hydG* Δ*ech* strain changed its metabolism to balance redox reactions, we calculated and analyzed the metabolic flux distribution (Supplementary Table S2). The distribution matched well with the experimental data ($R^2 = 0.99$).

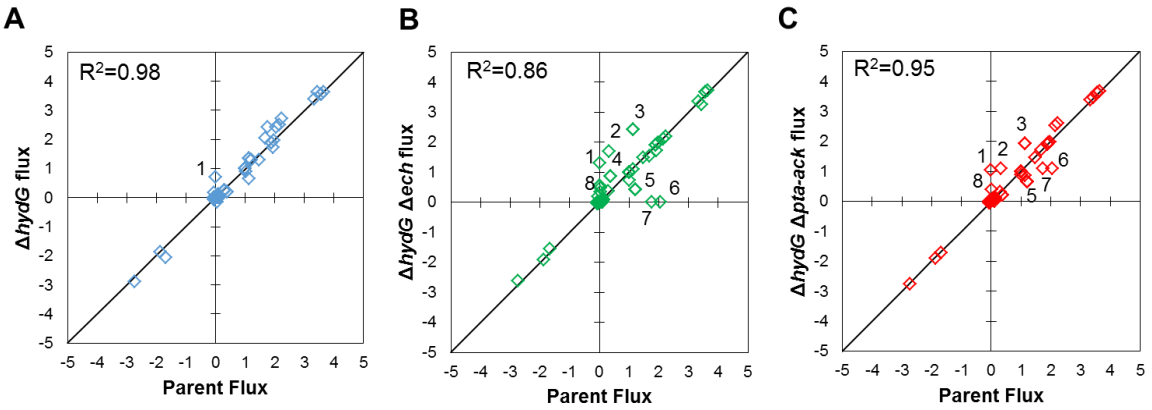

Figure 4.4: Parametric plots of flux distributions between **(A)** the parent and Δ*hydG* strains, **(B)** the parent and Δ*hydG* Δ*ech* strains, and **(C)** the parent and Δ*hydG* Δ*pta* strains. Reaction outliers are labeled as (1) NAPDH-dependent adhE* activity, (2) RNF, (3) ethanol export, (4) PFL, (5) acetate export, (6) ECH, (7) hydrogen export, and (8) NFN

Figure 4.4B shows the significant discrepancies between the parent and the Δ*hydG* Δ*ech* strain ($R^2 = 0.82$). The outliers in this plot can be classified into two distinct groups, one associated with both carbon and electron fluxes (i.e., NADPH-dependent AdhE*, total ethanol production, PTA, and PFL) and the other associated with only electron fluxes (i.e., RNF, NFN, ECH, and total H_2 production). METAFoR analysis on the flux distribution of the Δ*hydG* Δ*ech* strain further revealed the dramatic restructuring of redox metabolism. The calculated share of Fd_{rd} through the RNF reaction increased four-fold compared to the parent strain, whereas the flux through the NFN reaction, which was practically inactive in the parent strain, increased to account for 20% of the Fd_{rd} recycling (Figure 4.3B). This increase in NAD(P)H availability allowed for ethanol production to increase to 80% of the total flux through acetyl-CoA, and was matched by a drop in acetate production. Interestingly, ethanol yield only reached 59% of the theoretical limit when hydrogenase activities were diminished.

In summary, the complete disruption of H_2 production provided an effective means to increase flux of reducing equivalents to ethanol. This significant perturbation might

have caused accumulation of reducing equivalents and hence triggered a regulatory mechanism for redox balancing where cells diverted pyruvate to acetyl-CoA through the redox neutral PFL reaction instead of the reduced ferredoxin-generating PFOR reaction. This undesirable loss of reducing equivalents lessened the potential ethanol yield. The effect of this bifurcation at the pyruvate node on redox balance and ethanol production is further investigated in Section 4.2.4.

4.2.3 Metabolic model encompasses several phenotypes

As an additional assessment of our models predictive capability, we constructed and characterized the Δ*hydG* Δ*pta* strain which has perturbations designed to route carbon and electrons toward ethanol production with reduced production of H_2 and acetate. The results show that the Δ*hydG* Δ*pta* strain achieved a 61% increase in the ethanol yield (Figure 4.3A) while its growth rate decreased compared to the parent strain and the aforementioned Δ*hydG* strain by 50% and 30%, respectively (Table 4.1). This decreasing trend carried over to the cellobiose uptake rate, acetate production rate, and H_2 production rate (Table 4.1). The decreased growth rate and cell mass yield, with respect to the parent strain, could be attributed to the lack of ATP synthesis via substrate level phosphorylation through the pta-ack reaction pathway. In the Δ*hydG* Δ*pta* strain, the lack of acetate production helped increase carbon flux from the acetyl-CoA node towards ethanol production, while also increasing flux of reducing equivalents (Figure 4.3A). The need to balance carbon and electron fluxes to ethanol also translated to a 72% decrease in formate production, compared to the parent strain.

Like the aforementioned mutants, the calculated flux distribution matched experimental data well ($R^2 = 0.96$), and when parametrically compared to the parent strain, one can quickly see a similar pattern of outlier reactions as seen in the Δ*hydG* Δ*ech* strain (Figure 4.4C). The most prominent outliers focused again on carbon fluxes (i.e., NADPH-dependent AdhE*, total ethanol production, and PTA) and the

electron fluxes (e.g., RNF, NFN, ECH, and total H_2 production). To elaborate on this point, METAFoR analysis (Figure 4.3B) shows how the Δ*hydG* Δ*pta* strain recycled Fd_{rd} equivalently through the ECH and RNF reactions, as opposed to the parent strain predominately utilizing ECH. Furthermore, the NFN flux increased to about 14% of the total reduced ferredoxin recycling in the Δ*hydG* Δ*pta* strain. The increase in RNF and NFN fluxes provided the NAD(P)H necessary for ethanol production, and as expected the share of acetyl-CoA going to ethanol production increased over 60% compared to the parent strain (Figure 4.3B).

Taken all together, the model elucidates how *C. thermocellum* adjusts electron and carbon fluxes when its redox metabolism is perturbed towards ethanol production. These results give confidence in the model structure and any further phenotypic predictions.

4.2.4 Elucidating redox bottlenecks hindering ethanol production

While the energy and redox perturbations enhanced ethanol flux, the ethanol yield was still low, only reaching 60% of the theoretical limit. This result indicates that the capacity for ethanol flux might still be limiting ethanol production. Since *C. thermocellum* employs multiple metabolic control valves, such as flux bifurcation at the pyruvate node via PFL and PFOR and/or distribution of electron fluxes via NFN, RNF, BIF, and the Fe-H2 reactions, we used EMA to simulate additional redox-disrupted genotypes to identify key bottlenecks hindering high-yielding ethanol production.

PFL is vital for redox balancing when hydrogenases are inactive

Relative to the parent strain, a 70% increase in the flux to formate in the Δ*hydG* Δ*ech* mutant suggests that accumulation of Fd_{rd} results in a decrease in PFOR flux and increase in PFL flux. The importance of formate production is also apparent in

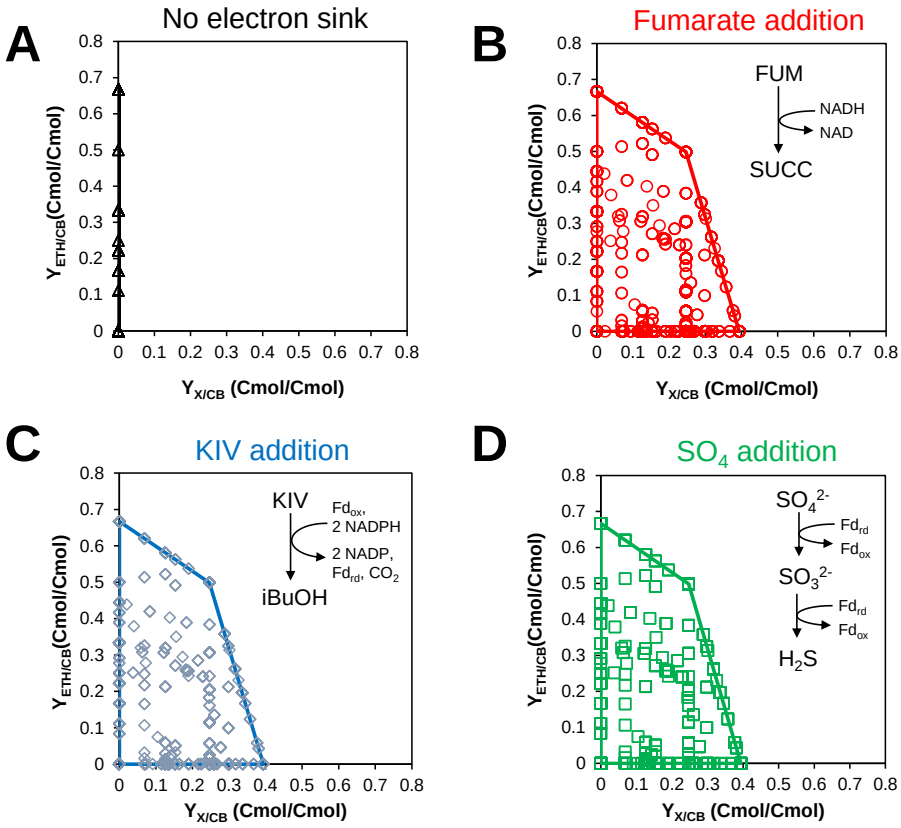

Figure 4.5: The 2-D phenotypic spaces of ethanol versus cell mass yields on cellobiose for $\Delta hydG$ Δech Δpfl strains without **(A)** and with addition of different external electron acceptors including fumarate **(B)**, keto-isovalerate (KIV) **(C)**, and sulfate **(D)**.

the OATS analysis for the $\Delta hydG$ Δech network (Table 4.2). To further examine this observation, an additional deletion of *pfl* in the $\Delta hydG$ Δech mutant would have an effect on potential phenotypes.

In silico results, via EMA, show that the $\Delta hydG$ Δech Δpfl strain had zero EMs that supported cell growth (Figure 4.5A). To date, this strain has not been constructed. However, hypophosphite (HPP) is a known PFL inhibitor (Rydzak et al., 2014) and can be used to experimentally eliminate PFL activity. Therefore, we tested this genotype experimentally by growing the $\Delta hydG$ Δech strain in the presence of 6 mM HPP. While HPP had no effect on the growth of the parent strain, we were able to confirm the models prediction of growth inhibition when HydG, ECH,

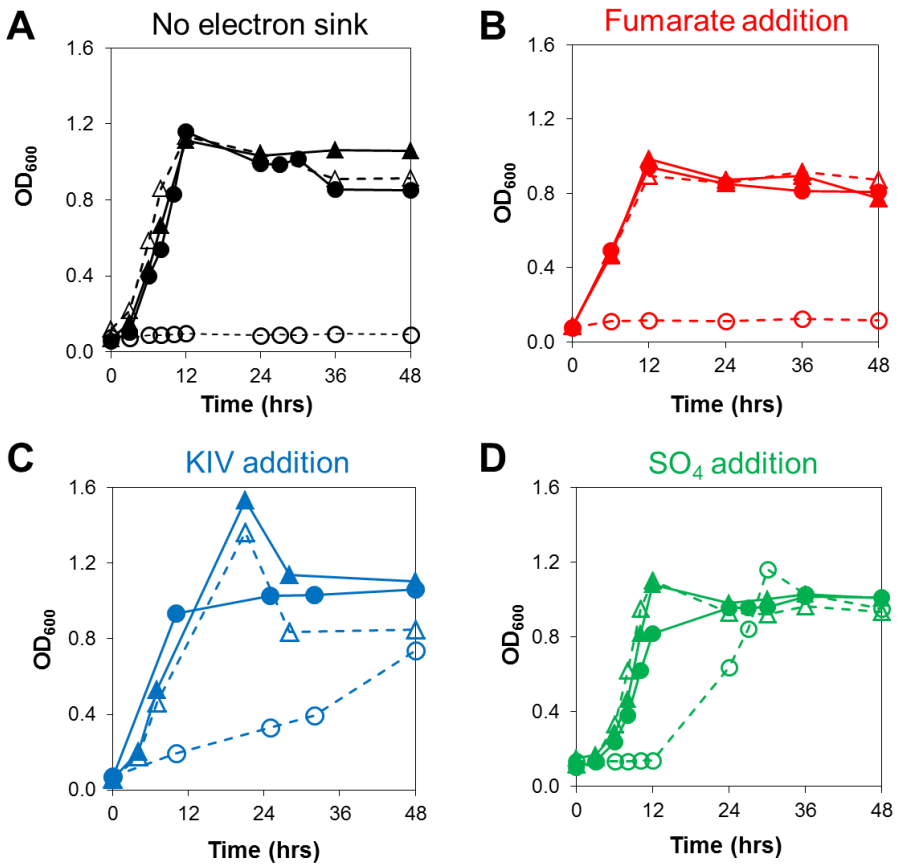

Figure 4.6: Growth characteristics of parent strain (triangles) and $\Delta hydG$ Δech (circles) in MTC media (filled symbols) or MTC with the PFL inhibitor hypophosphite (open symbols). To investigate redox bottlenecks, no additional electron sink **(A)**, 20 mM fumarate **(B)**, 20 mM 2-ketoisovalerate **(C)**, or 2 g/L total sulfate **(D)** were included in the medium to probe NADH, NAD(P)H, and Fd_{rd}, respectively.

and PFL were inactivated (Figure 4.6A). These results suggest that without the redox neutral relief valve of PFL, the lack of H_2 production causes a redox imbalance within the metabolism of *C. thermocellum* that eliminates cell growth.

Probing redox bottlenecks with external electron acceptors

To investigate if growth could be recovered when PFL is inhibited in the $\Delta hydG$ Δech strain, we simulated the addition of alternative electron consumption pathways into the model. The electron sinks were chosen to test the accumulation of the major

electron carrying species: 1) fumarate was tested for NADH recycling due to the presence of a fumarate reductase domain (Clo1313_3018) and a hypothetical protein with similarity to succinate dehydrogenase (Clo1313_2640); 2) 2-keto-isovalerate (KIV) was tested for NAD(P)H recycling due to experimental evidence of isobutanol production in *C. thermocellum* (Holwerda et al., 2014); and 3) sulfate was tested for ferredoxin recycling as at least one of the steps to sulfate reduction is ferredoxin dependent (Feinberg et al., 2011). *In silico* results demonstrate that addition of any single electron sink pathway into the Δ*hydG* Δ*ech* Δ*pfl* metabolic network restored the phenotypic space to the same area and yield ranges as the parent or Δ*hydG* strain (Figure 4.5). To test these phenotypes experimentally, we investigated the growth behavior of the parent versus Δ*hydG* Δ*ech* strains in the presence of various electron sinks, with and without inhibition of PFL by HPP.

First, the parent strain (used here as a positive control) could grow with or without the HPP inhibition regardless of supply of external electron acceptors (Figure 4.6A), as expected. For the Δ*hydG* Δ*ech* strain, exogenous addition of 40 mM fumarate was unable to restore growth with PFL inhibition (Figure 4.6B). This result indicates that i) there is not an accumulation of NADH and/or ii) the annotated fumarate reductase was not active. It should be noted that the level of fumarate addition was non-toxic because the parent strain and hydG ech strains grew under conditions without HPP. Furthermore, addition of 40 mM KIV led to slight growth (Figure 4.6C) and isobutanol production after two days. This result indicates that accumulation of NAD(P)H could be occurring.

Interestingly, the normal levels of sulfate in MTC media (1 g L^{-1} Na_2SO_4) were not sufficient to allow growth of Δ*hydG* Δ*ech* Δ*pfl*. However, addition of extra sulfate (2 g L^{-1} Na_2SO_4 total) into the culture medium was able to bring the maximum OD_{600} to levels comparable to the control (Figure 4.6D) after a lag phase. This result suggests that reduced ferredoxin is the species most likely accumulating, causing the lethal electron imbalance.

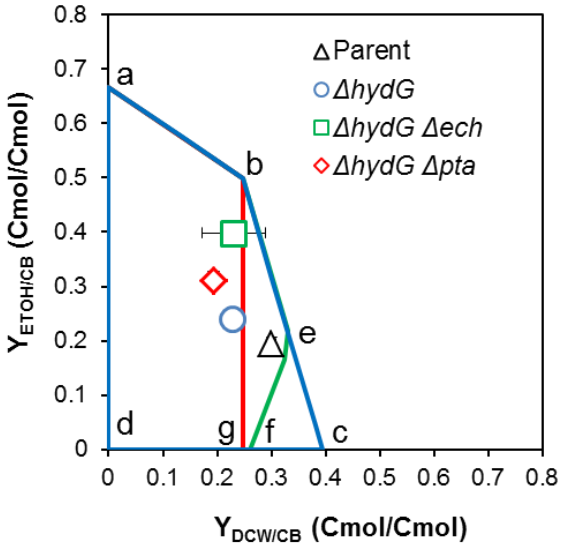

Figure 4.7: The 2-D phenotypic spaces of ethanol versus cell mass yields on cellobiose for various genotypes. The phenotypic spaces (or convex hull) are encompassed by the areas abcd for the parent and Δ*hydG* strains, abefd for Δ*hydG* Δ*ech* strain, and abgd for Δ*hydG* Δ*pta* strain. Symbols within these areas are experimental yields.

4.2.5 Design of optimal genotype for high-yielding ethanol production

With the trained metabolic model, EMA was applied to analyze the network structure and give insights for rational strain design (Trinh et al., 2009).

Additional modifications are necessary to constrain phenotype

To explain why the above metabolic engineering strategies were not sufficient to achieve high-ethanol yields, we used EMA to qualitatively assess phenotypic spaces of the parent, Δ*hydG*, Δ*hydG* Δ*ech*, and Δ*hydG* Δ*pta*, solely based on the metabolic network structure. Table 4.3 summarizes the EM properties of each genotype while Figure 4.7 compares the phenotypic spaces of the parent strain with the three mutant strains, focusing on the 2-D projection of the ethanol versus cell mass yields on cellobiose. This representation is commonly used to qualitatively compare which

Table 4.3: Elementary mode analysis for various *C.thermocellum* strains.

Network characteristics	Parent	Δ*hydG*	Δ*hydG* Δ*ech*	Δ*hydG* Δ*pta*
Total EMs	5202	2258	1148	816
ETH EMs	1982	1357	807	499
BIO EMs	3858	1478	718	558
Both EMs	1143	842	470	334
Range ETH (Cmol/Cmol)	0.00-0.67	0.00-0.67	0.00-0.67	0.00-0.67
Range BIO (Cmol/Cmol)	0.00-0.39	0.00-0.39	0.00-0.33	0.00-0.25

phenotypes a metabolic network can feasibly attain given its stoichiometric structure. By definition, the experimental phenotypes of the parent or any knock-out mutant must fall within the phenotypic space predicted by the model unless new reactions are added to the network; otherwise, there is at least one error present in the model structure.

When examining the presented strains, an immediately noticeable result is the similarity of the phenotypic spaces. The Δ*hydG* strain had the same space boundaries as the parent strain (the area abcd, Figure 4.7), while the Δ*hydG* Δ*ech* strain shows a slightly contracted space (the area abefd) with regards to the cell mass yield. Since cellobiose and dry cell weight have equivalent oxidation states, the contracted space can be attributed to the lack of carbon-free electron disposal in the form of hydrogen production requiring the synthesis of reduced carbon compounds such as ethanol and lactate to balance redox. This flux of carbon to other reduced products at low ethanol yield limits the yield of dry cell weight. The Δ*hydG* Δ*pta* strain exhibits a truncated boundary on the cell mass yield axis (the area abgd) which can be attributed to a lower ATP availability for cell synthesis due to a lack of substrate-level phosphorylation by *pta-ack*.

Most importantly, for each genotype, the admissible range of ethanol still stretched from 0.00 to 0.67 (Cmol/Cmol). This indicates that there are accessible phenotypic states for these strains that do not produce any ethanol; the existence of the undesired EMs in these strains helps explain why they did not reach high ethanol yields

experimentally (Table 4.3). It should be noted that each phenotypic space represents the metabolic capabilities of each genotype at quasi steady state. Under a given operating condition, each genotype will occupy a point within its own space, i.e., symbols in Figure 4.7 designate the phenotypic states of the parent strain and its mutants under our characterized conditions.

Metabolic model predicts feasibility of high-ethanol yielding phenotype

We investigated the ability of the central metabolic model to describe a high ethanol yielding ($> 75\%$ theoretical maximum) phenotype by performing the Minimal Metabolic Functionality (MMF) algorithm (Trinh et al., 2008). The MMF method constrains the phenotypic space by iteratively performing gene deletions *in silico*, aiming to reduce the total number of elementary modes and the accessible range of the desired product. When running the MMF algorithm, we opted to start with the $\Delta hydG$ Δech genotype, since an increase in ethanol yield was observed in that strain, the dual NADH/NADPH activity has been shown for ethanol synthesis, and genetic manipulation in *C. thermocellum* is a bottleneck in strain design (Tripathi et al. (2010); Mohr et al. (2013)).

Figure 4.8 presents the stepwise progress of the MMF algorithm on the number of EMs describing the genotype, as well as the range of ethanol yields. While simulating a $\Delta hydG$ Δech $\Delta pta\text{-}ack$ strain reduced the total number of elementary modes by about half, and simulating $\Delta hydG$ Δech $\Delta pta\text{-}ack$ Δldh reduced the total EMs by another 40%, the range of ethanol yield was not constrained (Figure 4.8). However, simulating $\Delta hydG$ Δech $\Delta pta\text{-}ack$ Δldh Δval_{out} brought the ethanol yield range to 0.50-0.67 (Cmol ETOH/Cmol CB) for non-growth associated production and 0.50-0.62 (Cmol ETOH/Cmol CB) for growth-associated ethanol production, while reducing the total number of EMs again by over half. This result invites the hypothesis that central metabolism can be constrained to a high-yielding phenotypic space with only three additional modifications. The ability to delete or overexpress genes in *C. thermocellum* is a significant challenge in its own right (Guss et al. (2012);

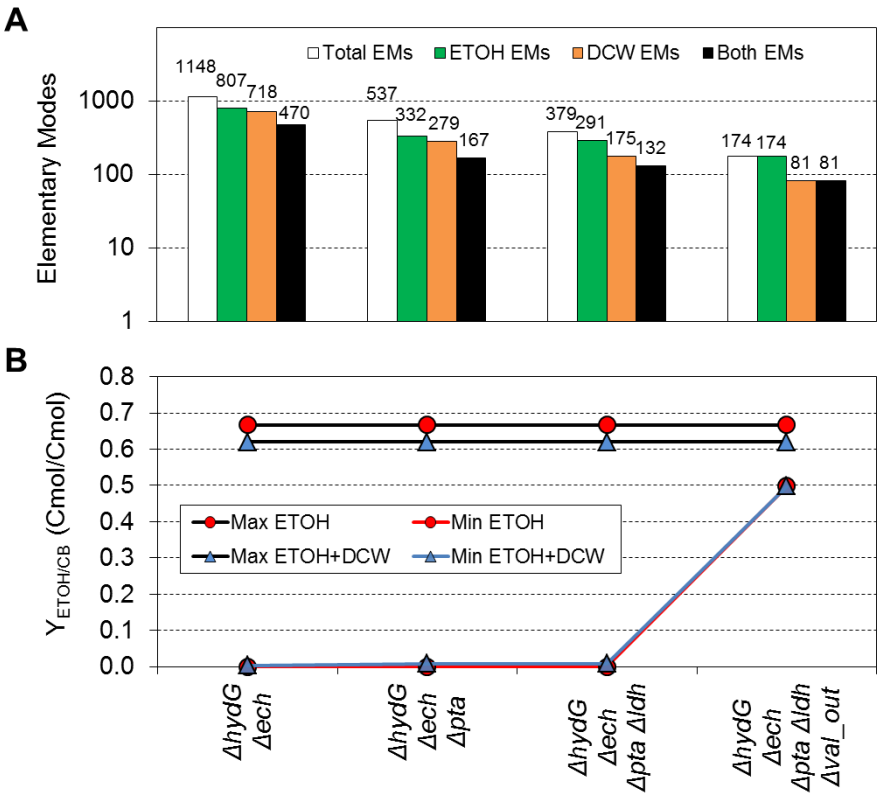

Figure 4.8: Application of MMF algorithm to identify the optimal genotype of *C. thermocellum* for high-yield ethanol production. **A**) Effect of reaction deletions on the total EMs, ethanol-producing EMs, cell mass-producing EMs, and coupled ethanol- and cell mass-producing EMs. **B**) Effect of reaction deletions on minimum and maximum yields of ethanol and cell mass on cellobiose.

Mohr et al. (2013)), and as such the construction of the optimal strain predicted here is beyond the scope of this work.

4.3 Discussion

This chapter presented the development and validation of a predictive, stoichiometric, central metabolic model of *C. thermocellum* for flux quantification to elucidate cellular phenotypes and for rational strain design based on EMA. The model was able to illuminate how *C. thermocellum* dramatically shifts its cofactor recycling processes when exposed to significant energy and redox perturbations. In each of the mutants

presented, a change from the parent strain central metabolism led to an increase in ethanol production. The proposed mechanism is an increase in conversion of reduced ferredoxin to NAD(P)H through NFN and RNF.

This study reveals the major driving force in ethanol production is a need to balance redox metabolism, as opposed to the push of carbon flux. At the same time, the diverse routes for regeneration of oxidized electron carriers offer many different avenues for producing reduced products other than ethanol. For instance, in the Δ*hydG* strain, the results suggest that the reduced ferredoxin positively affects ECH flux since specific H_2 production increased without a significant change in the production of ethanol. Additionally, the Δ*hydG* Δ*ech* and Δ*hydG* Δ*pta* strains had similar acetate yields, but the Δ*hydG* Δ*ech* strain had a 28% increase in ethanol yield compared to the Δ*hydG* Δ*pta* strain (Figure 4.3A).

The need to balance redox metabolism as a driving force is also intuitive when considering acetyl-CoA is made available by both PFOR and PFL reactions as well as the fact that the sum of PFOR and PFL fluxes is not significantly altered by genotype. This suggests AdhE* activity is not acetyl-CoA limited. The experimental data suggests that regenerating reduced ferredoxin solely through PFOR is not sufficient to match the rate of cellobiose uptake and glycolysis. To compensate, pyruvate consumption is diverted from the stalling PFOR reaction towards the redox neutral PFL reaction (Figure 4.3B), which is consistent with literature reports of conditions with disrupted H_2 synthesis (Rydzak et al. (2014); Biswas et al. (2015)). The decrease in flux through PFOR can be attributed to rate-limitation at either i) reduced ferredoxin recycling via NFN or RNF, or ii) the push of NAD(P)H to ethanol synthesis (i.e. AdhE* activity).

It is worth noting the difference in H_2 production in the Δ*hydG* strain in serum bottles (Biswas et al., 2015) and the bioreactor studies presented here could be linked to inhibition of ECH activity and an increase in the activities of NFN and RNF. The calculated flux distribution for the Δ*hydG* Δ*ech* strain shows that NFN and RNF become highly active to compensate for diminished hydrogenase activity. Therefore,

it is feasible to propose that the higher partial pressure of H_2 in serum bottle cultures increases the availability of NAD(P)H in the $\Delta hydG$ strain. Indeed, it has been shown that artificially increasing H_2 in the culture atmosphere leads to increased ethanol production (Rydzak et al., 2011). The hypothesis that H_2 inhibition is lower for ECH than the bifurcating hydrogenase or Fe-Fe hydrogenases could explain the higher H_2 yield in the parent strain serum bottle studies, since all hydrogenases should be actively present. In contrast, the only way that the $\Delta hydG$ Δech strain can recycle reduced ferredoxin is to produce NAD(P)H via RNF or NFN.

To further highlight the importance of redox balance, inhibiting PFL activity in the $\Delta hydG$ Δech strain causes growth to cease. This phenotype was predicted by our model and then confirmed in batch cultures. This effective genotype serves as a promising experimental condition to probe bottlenecks in redox cofactor recycling by use of exogenous electron sinks. We anticipated that the addition of fumarate would serve as a NADH probe and restore redox balance through succinate production. However, the fumarate study was inconclusive as the inability to restore growth could be due to a lack of fumarate reductase expression, lack of a fumarate transport protein, or an incorrect annotation of the fumarate reductase protein.

The inconclusive results from the fumarate study led to the use of KIV as a NAD(P)H probe. This gave more promising results, as growth was slowly able to recover, cellobiose was consumed in the process, and isobutanol was detected. The slow growth recovery clearly shows additional reducing equivalents can be recycled via KIV conversion to isobutanol. The slow rate of growth likely reflects a kinetic bottleneck in the conversion of reduced ferredoxin to NAD(P)H and/or low activity of KIV reductase, although isobutylaldehyde reduction kinetics or toxicity cannot be ruled out.

The final electron sink, sulfate, was designed to probe the accumulation of reduced ferredoxin. After a lag phase, growth was recovered to a rate comparable to controls without HPP. The lag phase could be due to an adjustment in gene expression levels for greater sulfate reduction. These results allude to reduced ferredoxin

being the main accumulating redox cofactor, which implies that re-oxidation of ferredoxin is a rate limiting step in the electron recycling pathway of *C. thermocellum*. Future experiments overexpressing these potential bottleneck enzymes should enhance ethanol production.

The application of the MMF method to *C. thermocellum* revealed that elimination of H_2 production coupled with the removal of acetate and lactate synthesis as well as valine secretion effectively constrains the network to a high yielding ethanol phenotype (Figure 4.8). It is worthwhile again to note that PFL is not included in this set of reaction deletions, as the model predicts that removing H_2 and formate production does not allow cell growth. While the MMF method seeks to find all reaction deletion sets that have the minimum size (i.e., cardinality) and exclude reactions downstream of a linear pathway (Trinh et al., 2009), it is possible to identify other unique sets with larger sizes by using the cMCS tool to achieve the same high yielding ethanol phenotype (Hädicke and Klamt, 2011). Due to many possible candidate design strains generated by the cMCS method (>5000 for our network), great care must be taken to screen out those that are not biologically relevant. Given the redox bottlenecks in reduced ferredoxin recycling, as previously discussed, simply deleting the abovementioned genes will probably not immediately facilitate high ethanol productivity. Presumably, a directed evolutionary approach to select for faster growth or, more proactively, the intentional overexpression of NFN, RNF, and AdhE*, will be necessary for the designed strain to overcome metabolic obstacles and reach the desired, high ethanol yielding phenotype.

In summary, this study accents the hypothesis that constricting electron flux contributes more than carbon redirection to ethanol production in *C. thermocellum*, and identifies bottlenecks in the redox metabolism hindering high ethanol yields. Rational strain design based on the MMF method and *in vivo* characterizations predict increasing flux through NFN, RNF, and/or AdhE* while eliminating acetate, lactate, H_2, and valine production is the most promising strategy for optimizing the

production of ethanol. We envision that our trained metabolic model will be used to engineer *C. thermocellum* as a CBP platform to produce other reduced metabolites.

Chapter 5

Examining Complex Phenotypes With A Genome Scale Metabolic Model of *Clostridium thermocellum* DSM 1313 Implementing an Adjustable Cellulosome

Summary

This chapter describes the construction and refinement of a genome scale model for *Clostridium thermocellum*. The meticulously refined GEM was tuned to accurately simulate growth on cellobiose and cellulose in non-carbon limited conditions, accounting for the difference in cellulosome synthesis. Using this model, I examine several observed phenotypes presented throughout the literature and infer a large regulatory network which arises from growth on cellobiose. Finally, the model is

used with a third-party algorithm to predict optimal genotypes for the production of ethanol, hydrogen, and isobutanol. This work has been submitted to Biotechnology for Biofuels.

5.1 Introduction

For a sustainable energy economy, the necessity of producing fuels and chemicals from renewable feedstocks is well-acknowledged, and the use of bio-based resources is a promising route for significantly lowering the carbon footprint of liquid transportation fuels (Demain, 2009). To produce economically favorable biofuels, consolidated bioprocessing (CBP) is attractive as it uses specialized micro-organisms for direct conversion of lignocellulosic biomass into target chemicals in a single step (Lynd et al., 2005, 2008; van Zyl et al., 2007).

Of particular interest for CBP is the gram-positive thermophile *Clostridium thermocellum*, which exhibits a high growth rate on cellulose (Demain et al., 2005; Lynd et al., 2002) and can endogenously produce the biofuels ethanol (Lamed and Zeikus, 1980), hydrogen (Levin et al., 2006), and isobutanol (Holwerda et al., 2014). These desirable phenotypes are feasible because *C. thermocellum* possesses a large, organized, extracellular cellulosome (Bayer et al., 1983; Lamed et al., 1983) which is highly efficient at degrading lignocellulosic materials (Shoham et al., 1999) as well as an intricate, robust system of branched catabolic pathways that recycle reduced ferredoxin and NAD(P)H in order to facilitate a fast glycolytic rate and energy generation (Rydzak et al., 2009). This branched metabolism, however, makes production of a single product such as ethanol in *C. thermocellum* quite challenging.

Recently, there has been extensive work towards engineering C. thermocellum for increased ethanol production, e.g.

i Elimination of acetate production (Tripathi et al., 2010)

ii Elimination of lactate production (Biswas et al., 2014)

iii Elimination of both acetate and lactate production (van der Veen et al., 2013)

iv Elimination of hydrogen production (Biswas et al., 2015)

v Elimination of formate production (Rydzak et al., 2015)

vi Elimination of most traditional fermentation products (Papanek et al., 2015)

vii Elimination of malic enzyme activity while expressing an endogenous pyruvate kinase (Deng et al., 2013)

Despite these efforts, ethanol yield is still below industrially relevant levels. In the best performing strain a yield above 70% theoretical maximum has only been demonstrated at low substrate loadings (Papanek et al., 2015), and ethanol yield drops when substrate concentrations are increased (Brener and Johnson, 1984; Holwerda et al., 2014). These reports open many questions into the robustness of *C. thermocellum* redox metabolism and how regulatory mechanisms lead to the observed phenotypes in both cellobiose and cellulose grown cultures.

Constraint-based genome-scale metabolic modeling is rapidly becoming a standard tool for investigating cellular metabolism. The information contained in a genome sequence is redefined as a series of mass and charge balanced reactions in a genome-scale metabolic model (GEM). When coupled with thermodynamic constraints, metabolic constraints (e.g., substrate uptake rates and/or product consumption rates), and a cellular objective, GEM analysis can determine metabolic flux distributions, i.e. cellular phenotypes, under specified growth conditions. A repertoire of metabolic pathway analysis tools based on flux balance analysis and elementary mode analysis have recently been developed to analyze these GEMs and extensively reviewed (Feist and Palsson, 2008; Lewis et al., 2012; Senger et al., 2014; Simeonidis and Price, 2015). A *C. thermocellum* GEM *i*SR432 has been constructed previously (Roberts et al., 2010) and used as a scaffold for transcriptomic constraints (Gowen and Fong, 2010). While useful, several recent results highlight several limitations of *i*SR432 but this model has several shortcomings, e.g.

i There have been many advancements in the knowledge of *C. thermocellum* atypical glycolysis (Zhou et al., 2013), pentose phosphate pathway (Rydzak et al., 2012), and redox metabolism redundancies (Carere et al., 2014; Rydzak et al., 2014) which were not included in the original model

ii The model was constructed for the strain ATCC 27405, but DSM 1313 (Feinberg et al., 2011) is the genetically tractable parent strain used in metabolic engineering strategies (Olson and Lynd, 2012)

iii The model includes a generic cellulosome term but it is not variable with respect to carbon source, which has been shown to vary substantially (Zhang and Lynd, 2005b)

iv The model does not accurately predict certain cellular phenotypes like ethanol production (Roberts et al., 2010)

In this work, we constructed a new GEM for DSM 1313 from the KEGG database and manually curated the model with the most current knowledge of C. thermocellum metabolism (Thompson et al., 2015). We next refined the GEM using several sets of high-quality batch fermentation data for cell growth on various carbon sources, i.e., cellobiose and cellulose. This is accomplished by first tuning the energetic requirements for growth on cellobiose, then finding the additional ATP cost of producing the cellulosome for growth on cellulose. With this validated model, we investigated a series of interesting observations presented in the literature. First, we reproduced the difference in cell yield with respect to cellodextrin length, a direct consequence of the phosphorolytic sugar assimilation mechanism of *C. thermocellum* (Zhang and Lynd, 2004, 2005a). Next, we elucidated a regulatory mechanism to explain why cultures growing on cellulose do not reach the ethanol yields of cultures growing on cellobiose. We further demonstrated the robust energy and redox metabolism of *C. thermocellum* that enables it to dramatically adjust to environmental growth perturbations. Finally, we used the model to predict metabolic

engineering strategies to enhance the production of the desirable biofuels ethanol, hydrogen, and isobutanol for future experimental study.

5.2 Results

5.2.1 Model construction and comparison

Following the construction process outlined in the METHODS section, we obtained the *C. thermocellum* DSM1313 GEM, named *i*AT601 following convention (Reed et al., 2003; Roberts et al., 2010). This new model presents a significant improvement from the existing *C. thermocellum* ATCC 27405 GEM (Roberts et al., 2010) by incorporating very recently expanded knowledge of *C. thermocellum* metabolism. In particular, we updated the cofactor specificity of glycolytic enzymes (Zhou et al., 2013) based on *in vitro* protein characterization as well as performed Cofactory analysis (Geertz-Hansen et al., 2014) to resolve cofactor specificity when *in vitro* data was unavailable (see METHODS). We manually curated the intricate carbon overflow and redox metabolisms with recently acquired knowledge (Holwerda et al., 2014; Thompson et al., 2015).

Importantly, we also built the GEM *i*AT601 to account for the composition and synthesis cost of the cellulosome because *C. thermocellum* is known to alter cellulase expression when cultured on different carbon sources (e.g., cellobiose, cellulose, switchgrass, etc.) (Raman et al., 2009; Zhang and Lynd, 2005b). To construct the cellulosome term for the GEM *i*AT601, we compiled both the protein and amino acid distributions for the cellulosomes experimentally measured for growth on different carbon substrates (Raman et al., 2009). While protein compositions of the cellulosomes (e.g., hydrolyases, scaffodins, dockerins, etc.) significantly changed for growth on different substrates, amino acid compositions of these cellulosomes remained relatively similar 5.1. FBA simulations using amino acid compositions of various cellulosomes and maximum growth objective gave similar values of predicted

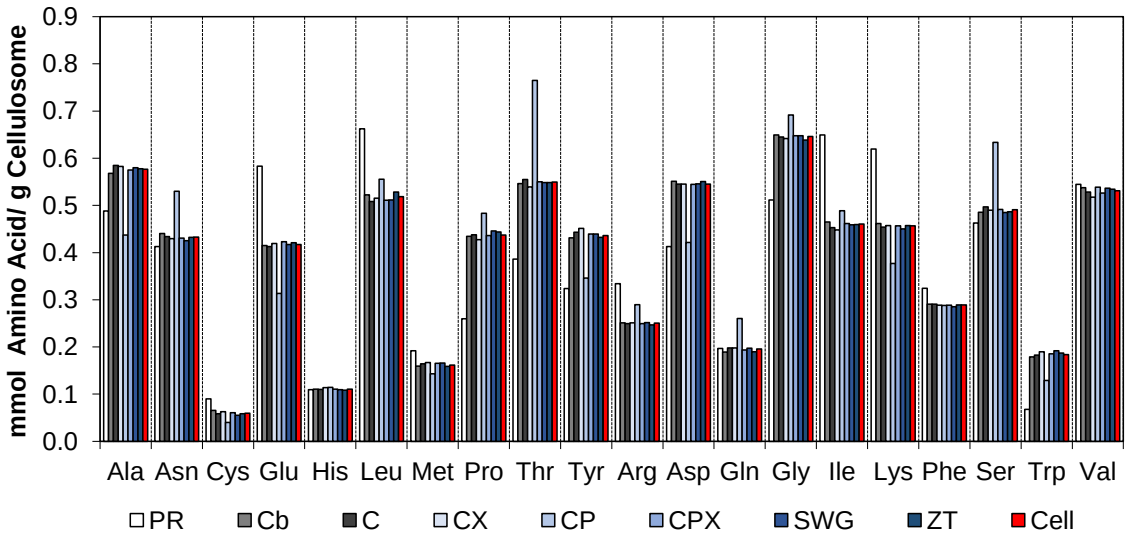

Figure 5.1: Distribution of amino acids per unit cellulosome under different conditions. PR, generic protein term from Dry Cell Weight approximation. Cb, cellobiose. C, crystalline cellulose (avicel). CX, cellulose + xylan. CP, cellulose + pectin. CPX, cellulose + pectin + xylan. SWG, pretreated switchgrass. ZT, amorphous cellulose (Z-Trim). Cell, median values across all other amino acid distributions and the values used in our cellulosome term. Data adapted from (Raman et al., 2009)

optimal growth within a range of approximately 0.2% (st. dev. = 0.34e-4) of each other. Thus, we used the median amino acid requirement across the different culture conditions for the cellulosome term in the GEM *i*AT601. It is interesting to observe that *C. thermocellum* could dynamically modulate protein components of cellulosomes to control its biomass degradation machinery under environmental perturbations (e.g., different growth rates (Dror et al., 2003) or various types of cellulosic substrates (Nataf et al., 2010; Zhang and Lynd, 2005b)), amino acid distributions of cellulosomes remained relatively constant, suggesting an evolutionary influence on the dynamic degradation machinery. Fundamental knowledge of the regulatory mechanisms controlling the finer dynamics of the machinery is lacking and waiting to be explored.

Table 5.1: Comparison of the GEM attributes among various Clostridial species. Abbreviations: *C. acetob, Clostridium acetobutylicum; C. beij, Clostridium beijerinckii; C. cellulolyt, Clostridium cellulolyticum; and C. therm, Clostridium thermocellum.*

	*i*CAC490	*i*CM926	*i*SF431	*i*SR432	*i*AT601
	C. acet.	*C. beij.*	*C. cell.*	*C. therm.*	*C. therm.*
Strain	DSM 824	NCIMB 8052	H10	ATCC 27405	DSM 1313
ORFs	4017	5100	3575	3238	3033
Reactions	794	938	621	577	869
Transport	120	68	45	73	110
Included Genes	490	925	431	432	552
Metabolites	707	821	603	525	903
Updated in	2012	2012	2010	2010	2015
Reference	McAnulty et al. 2012	Milne et al. 2011	Salimi et al. 2010	Roberts et al. 2010	This work

Overall, the GEM *i*AT601 contains 872 reactions, 904 metabolites, and 601 genes. Included in the model are 114 transport and exchange reactions for the 57 extracellular metabolites. This represents a 51%, 72%, and 39% increase in reactions, metabolites, and genes, respectively, over the existing GEM of *C. thermocellum* ATCC 27405 (Roberts et al., 2010). For additional comparison, *i*AT601 covers 19% of the annotated ORFs in the DSM 1313 genome, which is higher coverage than the average 13% for other Clostridial GEMs (Table 5.1). The GEM *i*AT601 encompasses all major metabolic pathways of *C. thermocellum*, and the numbers of reactions within different KEGG pathways are summarized in Figure 2.2.

5.2.2 ATP Requirement for Growth on Cellobiose

After construction, we proceeded to train the model using pH-controlled batch fermentation data collected for the wild-type DSM1313 grown on MTC medium with either cellulose or cellobiose as a carbon source (Holwerda et al., 2012; Thompson et al., 2015). Table 5.2 presents the experimental fluxes used to constrain the model, and for all simulations, a non-growth associated maintenance (NGAM) cost of 3.27

Table 5.2: Experimental fluxes used for metabolic model constraints

Experimental flux (mmol species/gDCW/hr)	Cellobiose	Cellulose
$r_{GluEqup}$	7.16±0.16	6.39±0.08
r_{ETH}	4.19±0.10	2.63±0.03
r_{ACE}	2.63±0.86	3.40±0.15
r_{CO2}	5.82±0.02	n.d.
r_{FOR}	1.77±0.01	1.38±0.01
r_{H2}	7.86±0.38	n.d.
r_{VAL}	0.78±0.12	n.d.
r_{LAC}	0.18±0.01	0.00±0.0
μ (hr^{-1})	0.33±0.01	0.31±0.01

mmol ATP/gDCW/hr was used (Zhang and Lynd, 2005a). We first investigated the models growth predictions with the cellodextrin uptake rate as a sole flux constraint. For cellobiose-grown simulations, the model did not predict any ethanol production under maximum growth conditions. Figure 5.2 shows the predicted phenotype for all major fermentation products under this initial condition, and immediately noticeable is that maximizing cell growth correlated with an overestimation of acetate production, presumably due to the additional ATP produced via substrate level phosphorylation via the phophotransacetylase acetate kinase pathway. While the acetate overestimation was expectedly coupled with the over-predicted formate production to balance the redox state of the cell in silico, these simulated results were clearly not consistent with the experimentally observed phenotype. Since the model predicted faster growth than observed experimentally, it is clear that the growth-associated maintenance (GAM) cost must be refined.

While it is straightforward to calculate the ATP required to synthesize 1 g of dry cell weight (Neidhardt et al., 1990; Stouthamer, 1973), the extra requirement for GAM (i.e., for regulation of cellular osmotic level, protein secretion, and flagellar motion) is less straightforward and is normally calculated with substrate-limited chemostat experiments (Pirt, 1965). Since the GAM is typically condition dependent (Tempest and Neijssel, 1984) and industrially relevant conditions are not

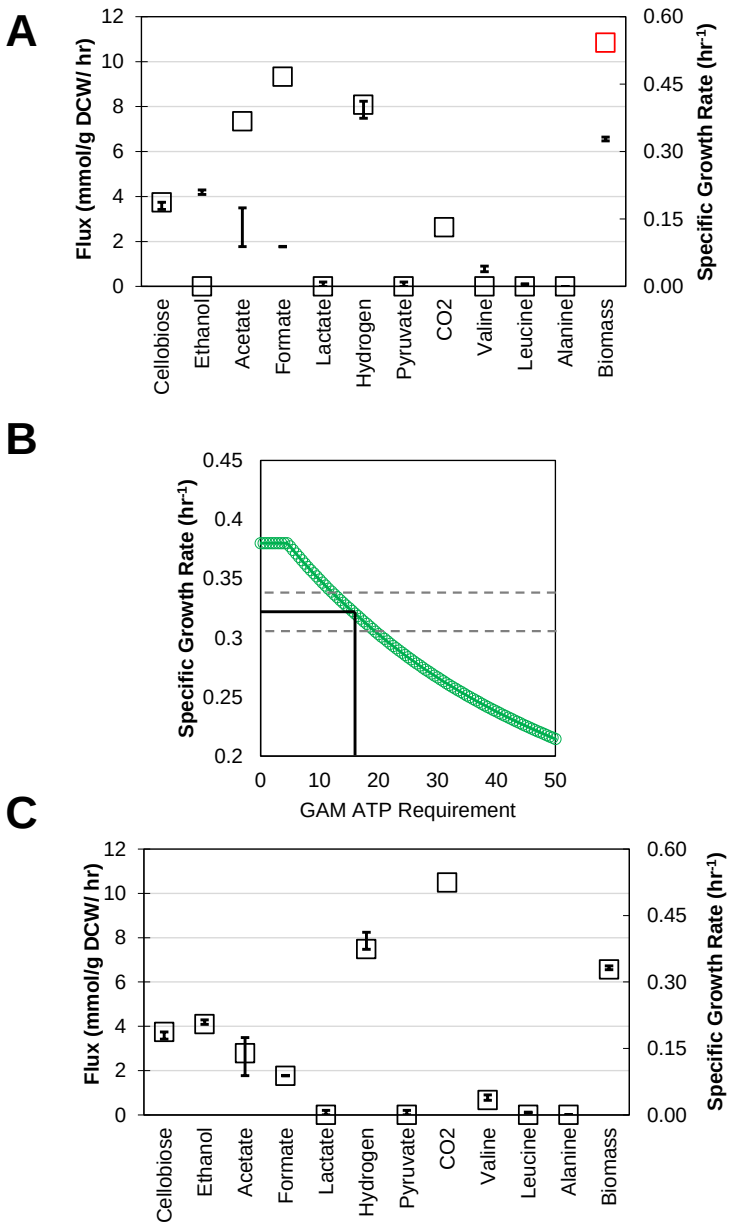

Figure 5.2: (A) Comparison of experimental and simulated metabolic fluxes for optimal growth of *C. thermocellum* on cellobiose without tuned growth-associated maintenance (GAM) ATP requirement. FBA simulation used only experimentally determined cellobiose uptake as a constraint. (B) Identification of best-fit GAM ATP requirement. The model energy balance is tuned by altering GAM ATP requirement and optimizing growth rate with specified fermentation constraints. Dotted lines frame the experimentally observed growth rate range, while solid lines illustrate the average observed growth rate and the best-fit GAM ATP requirement. (C) Comparison of experimental and simulated metabolic fluxes for optimal growth of *C. thermocellum* on cellobiose with tuned GAM ATP requirement.

carbon-limited (Humbird et al., 2010), we used the model to estimate an appropriate GAM coefficient. To find the GAM, we set experimentally measured fluxes as constraints (Table 5.2) and varied an ATP requirement in addition to DCW synthesis while optimizing cell growth. A value of 13.5 mmol ATP/gDCW/hr was found to best fit growth on cellobiose in batch conditions (Figure 5.2B). When we applied all experimentally obtained flux constraints and this tuned GAM coefficient then maximized growth, the models predictions accounting for the GAM requirement matched well with the experimental fermentation profile (Figure 5.2C).

5.2.3 Additional ATP Requirement for Cellulosome Synthesis

We next performed *in silico* analysis of *C. thermocellum* growth on cellulose. It has been shown experimentally that the cellulosome is no longer suppressed as when *C. thermocellum* grows on cellobiose (Zhang and Lynd, 2005b), and so we increased the percent of dry cell weight attributed to the cellulosome and applied experimentally measured flux constraints for simulation (Table 5.2). Using the previously calculated GAM value for growth on cellulose, however, still returned poor cell growth predictions when specific fermentation rates were included as a constraint (Figure 5.3A). Since the cellulosome is a large, extracellular enzyme complex, the discrepancy between the models prediction and experimental observation was likely due to not accounting for an increased ATP demand for cellulosome synthesis and secretion.

To further train the GEM *i*AT601, we set the GAM and NGAM as described above while similarly increasing the ATP requirement for cellulosome synthesis and secretion to simulate maximum growth rates. We found that an ATP cost of 57 mmol ATP/g cellulosome/hr was the best fit to wild-type growth on cellulose (Figure 5.3B). This corresponds to 14 mmol ATP/g cellulosome/hr greater than what is required for the cell protein synthesis. Given that the cellulosome represents a greater proportion

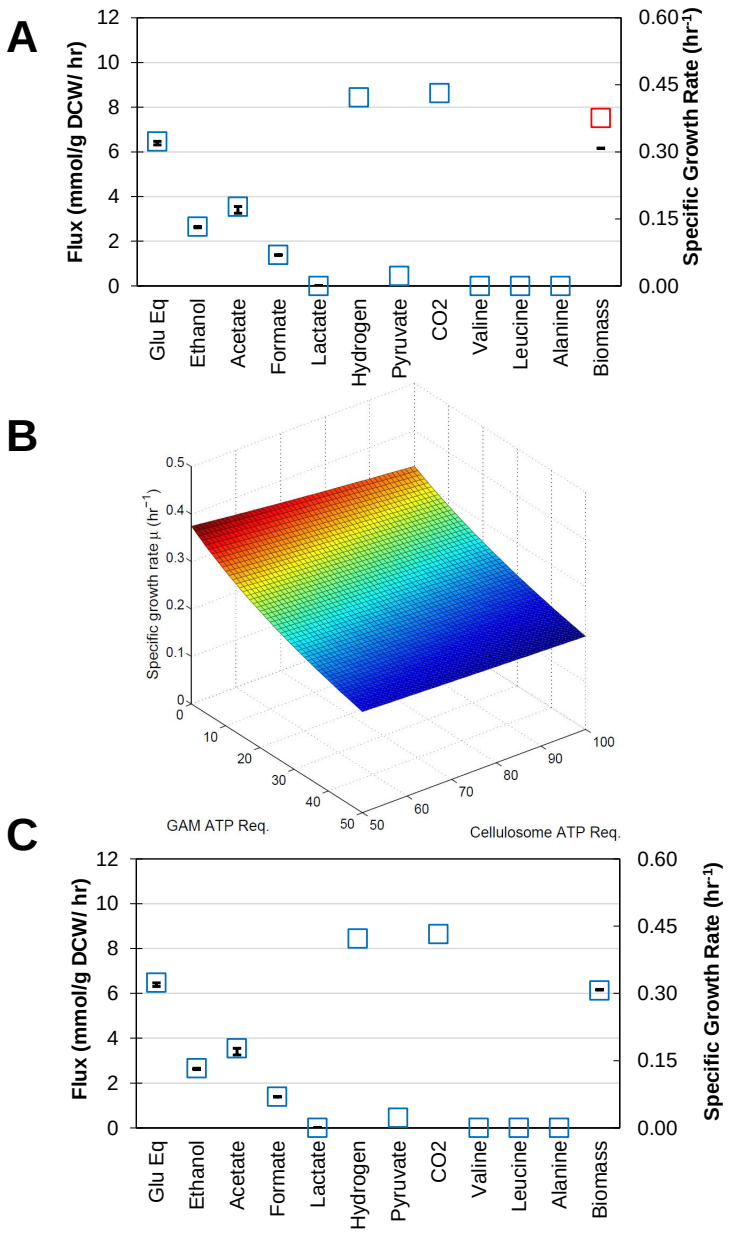

Figure 5.3: (A) Comparison of experimental and simulated metabolic fluxes for optimal growth of *C. thermocellum* on cellulose without ATP requirement for cellulosome synthesis. FBA simulation used experimentally measured fermentation product fluxes and the calculated GAM ATP requirement as constraints. (B) Identification of best-fit ATP requirement for cellulosome biosynthesis. (C) Comparison of experimental and simulated metabolic fluxes for optimal growth of *C. thermocellum* on cellulose without ATP requirement for cellulosome synthesis. FBA simulation used all experimental flux values as well as the best fit for GAM and cellulosome synthesis ATP requirements as constraints.

of the dry cell weight for growth on cellulose than cellobiose (Zhang and Lynd, 2005b), this ATP cost is an equivalent overall increase of 1.14 and 11.4 mmol ATP/g DCW/hr for cellobiose and cellulose simulations, respectively. By applying the ATP cost and fermentation rates as constraints, simulations of cell growth on cellulose matched very well with experimental data (Figure 5.3C). Reapplying the cellulosome ATP cost to the previous cellobiose simulations did not alter the results outside of the experimentally observed flux ranges, and so for all further studies the GAM and cellulosome ATP coefficients are fixed at these values.

5.2.4 Application of GEM for Rational Strain Design

One important application of the tuned GEM *i*AT601 is to guide strain engineering for enhanced production of chemicals of interest. For instance, the constrained minimal cut set (cMCS) method (Hädicke and Klamt, 2011) can be used identify all feasible genotype variants with minimum metabolic functionalities tailored for production of specific chemicals (Trinh et al., 2008). While ethanol is a valuable product, there is also interest in using *C. thermocellum* to produce isobutanol (Holwerda et al., 2014; Lin et al., 2015) or hydrogen (Islam et al., 2006; Levin et al., 2006). Using the cMCS method altered for genome scale models (von Kamp and Klamt, 2014), we investigated the feasibility of strain design for the production of ethanol, hydrogen, and isobutanol (Table 5.3). Based on the tuned GEM *i*AT601, we founded 67 unique cut sets of size 6 and 185 cut sets of 7 that could produce high ethanol yields while tightly coupling with cell growth. As anticipated, many of the highly represented reactions are associated with central metabolism, redox metabolism in particular. These results are indicative of the level of redundancy within *C. thermocellum* redox metabolism and provide perspective on some of the shortcomings of previously reported metabolic engineering strategies. Further mining of these strategies is needed to design an informed metabolic engineering effort.

Table 5.3: Model-guided MCS strain designs for production of hydrogen, ethanol, and isobutanol.

Product	# Modifications	# Strain Designs
Ethanol	6	67
	7	185
Hydrogen	4	12
	5	221
	6	1105
	7	4816
Isobutanol	7	28

For hydrogen production, we found many solutions, including 12 intervention strategies of size 4 and 4,816 strategies of size 7. The presence of strategies of size 4 implies fewer modifications are needed for high hydrogen production compared to ethanol, which requires a minimum of 6 modifications. Finally, for isobutanol we only found 28 strategies of size 7, hinting that high isobutanol production in *C. thermocellum* will be a challenge due to greater modifications required. Many of these metabolic engineering strategies are not trivial, and are expected to be useful in guiding experimental implementations.

5.2.5 Effect of Cellodextrin Length on Growth

We next employed the GEM *i*AT601 to validate interesting cellular phenotypes of *C. thermocellum*. It has been experimentally shown that *C. thermocellum* could uptake up to cellohexaose (G6) and preferentially assimilate longer cellodextrins with an average carbon length of $\tilde{4}.2$, where the DCW yield increased with the increasing length of cellodextrins supplied as a carbon source (Strobel, 1995; Strobel et al., 1995; Zhang and Lynd, 2005a).

To investigate the effect of assimilating various (G2-G6) cellodextrins and glucose (G1) on cell yields, we set the glucose-equivalents uptake flux at a constant 6.5 mmol/g DCW/hr while altering the sole carbohydrate species available. To allow for direct comparison with experimental results (Strobel, 1995; Strobel et al., 1995),

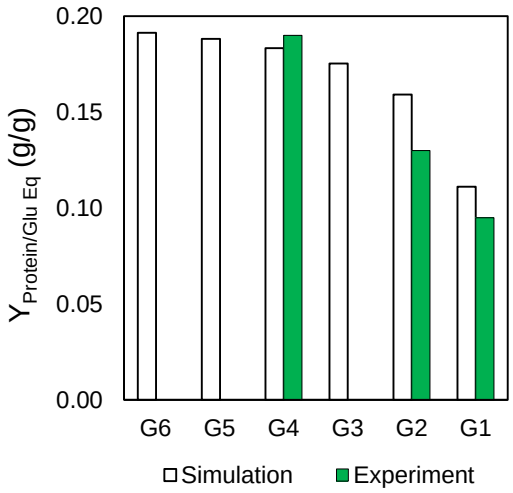

Figure 5.4: Comparison of effect of cellodextrin lengths on yield of cell protein per glucose equivalent (g/g) during simulation with *i*AT601 and values reported in the literature (Strobel, 1995). For culture simulations, a fixed glucose equivalent uptake rate of 6.5 mmol/gDCW/h was used. Cellodextrins of length N are shown as GN.

the simulation results are presented as yield of protein per glucose (g/g), where protein yield was calculated as the sum of fluxes to cellulosome and cell protein production (g proteinaceous component/g DCW/hr) divided by the glucose-equivalents uptake flux (g glucose equivalents/g DCW/hr). Our simulation shows that the maximum protein yield obtainable with G4 was 95% of that obtainable on G6 while yields on G3, G2, and G1 dropped to around 92%, 83%, and 58%, respectively, of the maximum under experimental conditions tested (Figure 5.4). This drop in maximum protein yields with respect to shorter cellodextrins matched well with the experimental data (Figure 5.4)(Strobel, 1995; Strobel et al., 1995); the trend clearly explained the calculated bioenergetic benefit to assimilation of longer cellodextrins (Zhang and Lynd, 2005a). This result establishes confidence in the overall bioenergetic constraints within the model for downstream analysis.

5.2.6 Effect of substrate and cell growth rate on bioenergetics of *C. thermocellum*

Extensive compilation of fermentation data for cell growth in comparable conditions (i.e. equivalent media recipes and available substrate) on cellobiose and cellulose (Avicel) in both batch and continuous cultures under different growth (or dilution) rates revealed several unique and interesting phenotypes regarding bioenergetics of *C. thermocellum* (Holwerda et al., 2012, 2014; Thompson et al., 2015; Zhang and Lynd, 2005b,a). For instance, the ethanol to acetate (E:A) ratio is a commonly used indicator of bioenergetic balance in a given metabolic state of an anaerobic cell culture, where ethanol production is primarily tied to redox balance and acetate production is coupled with ATP synthesis. The experimentally observed E:A ratio differs substantially when wild-type *C. thermocellum* grew on various cellulosic substates (cellobiose versus cellulose) under various growth rates (Figure 5.5). Specifically, *C. thermocellum* could reach an E:A ratio upwards of 2 for growth on cellobiose, while the ratio never crested 1 for growth on cellulose. While the E:A ratios highly depend on types of cellulosic substrates used, the sum of ethanol and acetate yields inversely correlates with the growth (or dilution) rates (Figure 5.5A) regardless of cellulosic substrates used. This correlation makes biological sense as a higher growth rate will require higher amounts of acetyl-CoA for DCW precursor synthesis and less will be available for ethanol or acetate synthesis. This also means that the correlation can serve as an ideal global constraint on bioenergetics of *C. thermocellum* and is employed for simulation in this study.

To better understand the bioenergetics of *C. thermocellum* when growing on different substrates under various growth rates, we sampled flux distributions based on experimental constraints followed by detailed analysis of the key cellular processes resulting in the observed trends of E:A ratios and sum of ethanol and acetate yields. Sampling is a common technique for examining a network structure to compare

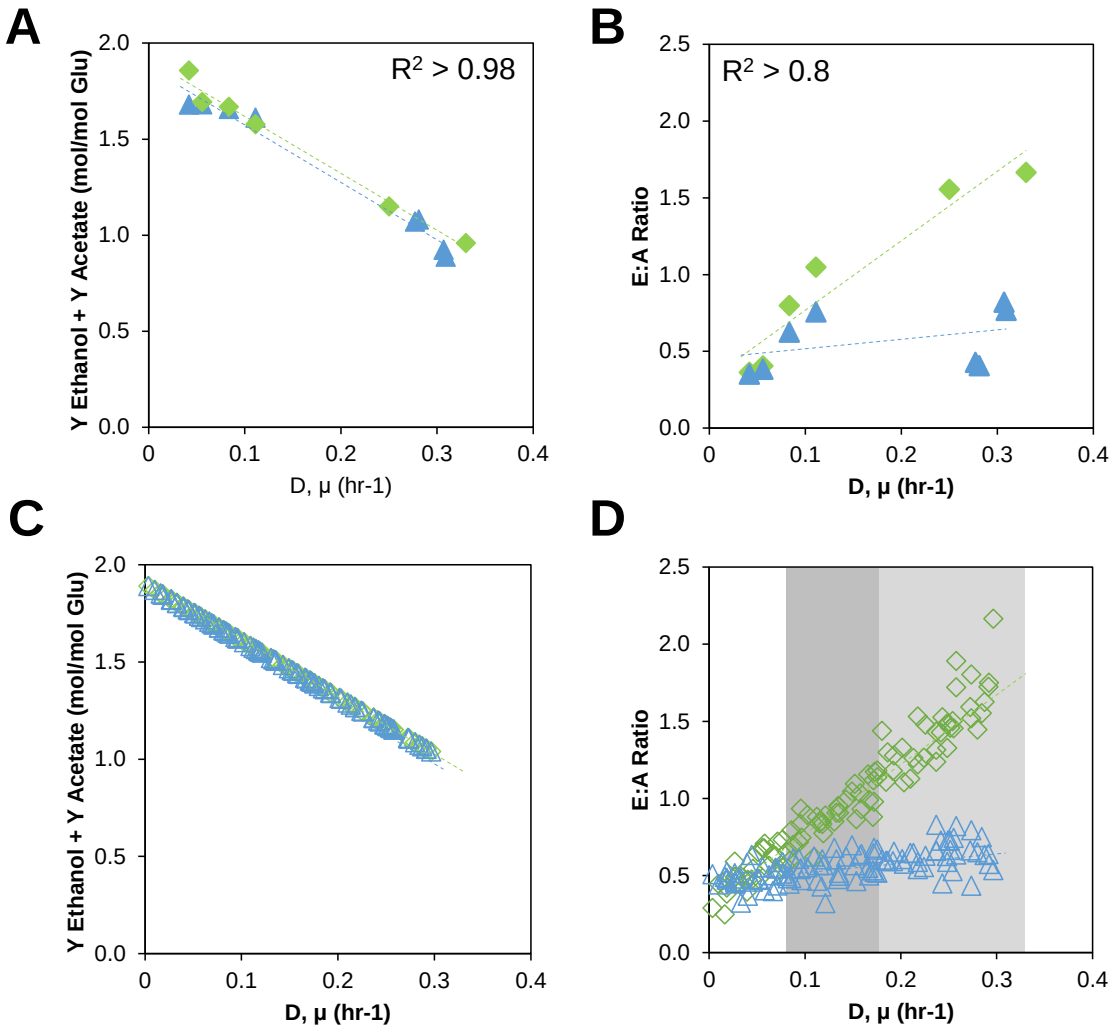

Figure 5.5: (A) Experimental data compilation of sum of ethanol and acetate yields from literature. (B) Experimental data compilation of E:A ratios for cellulose (triangles) or cellobiose (diamonds). (C) *In silico* implementation of the sum of ethanol and acetate yields. (D) *In silico* implementation of E:A ratios. The shaded regions outline the points within the low growth, medium growth, and high growth sets. Symbols: cellulose (triangles) or cellobiose (diamonds).

differences in conditions (Almaas et al., 2004; Schellenberger and Palsson, 2009) and/or infer regulatory elements (Bordel et al., 2010).

We set a tight constraint on the sum of ethanol and acetate yields with respect to growth rates (Figure 5.5C). We also introduced a noise level of 20% to the E:A ratio at a given growth rate to account for variability among the E:A ratio parameters (Figure 5.5D). The sum of yields and E:A ratios are considered jointly as the observed constraints below. For all sampling runs, the glucose equivalent uptake rates were randomly varied between the experimentally observed range of 5.0 7.5 mmol glucose equivalents/g DCW/hr, but set equal for both cellobiose and cellulose cultures. This setup allowed us to obtain 100,000 individual yet comparable flux distributions across a range of growth rates that sufficiently covered the observed variance for both cellobiose and cellulose simulations. The distributions in fermentation products were distinct for each carbon source at different growth rates. Increasing the number of sampling points to 500,000 did not have a significant effect on the flux trends, and so we are confident that these distributions are representative of cellular metabolism. The sampled flux distributions were analyzed to understand the metabolic differences which lead to the observed phenotypic differences between carbon sources.

Global Redox and Energy Cofactor Turnover

From the calculated flux distributions, we analyzed the turnover rates of the key metabolites ATP, GTP, pyrophosphate (PP_i), reduced ferredoxin (Fd_{rd}), NADH, and NADPH across our sample sets to elucidate how redox and energy metabolism is modulated to exhibit the observed trends of E:A ratios and the sum of ethanol and acetate yields across various environmental perturbations (e.g., cellulosic substrates and growth rates). It should be noted that a turnover rate of a metabolite determines how frequent that metabolite is biologically transformed and recycled at a given steady state and does not inherently give insight into the metabolite concentration within the cell. The result shows that the turnover rates of ATP, GTP, PP_i, and NADPH increased steadily with an increase in growth rates (Figure 5.6) as

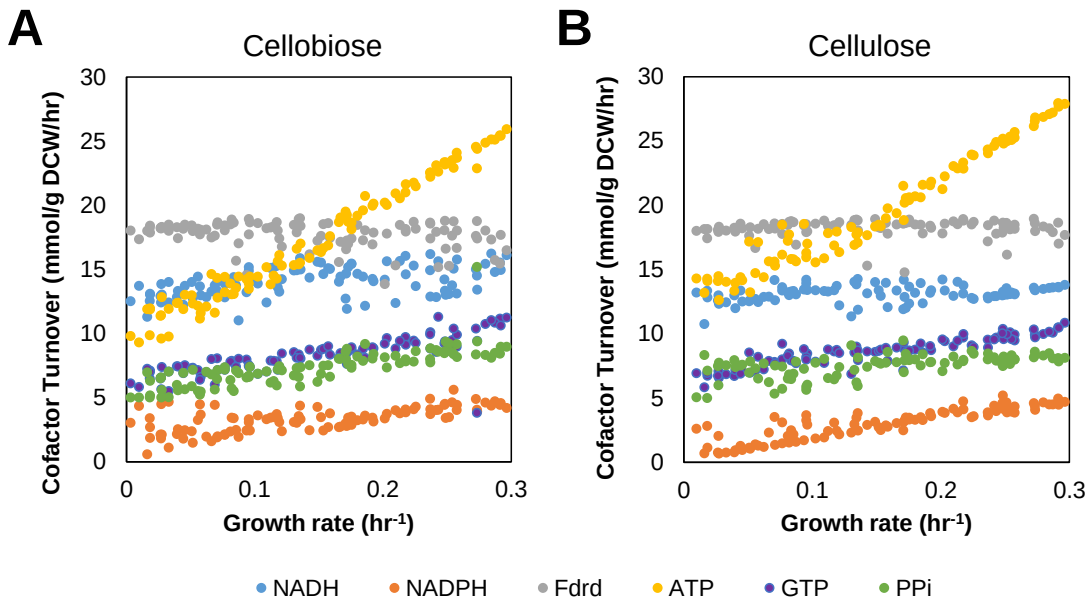

Figure 5.6: Average cofactor turnover with respect to growth rate for simulations using cellobiose (A) and cellulose (B).

expected because the synthesis of biomass requires these cofactors. The ATP turnover rate increased more sharply for cellulose cultures which could be attributed to the additional burden of cellulosome synthesis as well as requirement of acetate biosynthesis. The ATP trend matched well with previously reported experimental evidence (Zhang and Lynd, 2005a).

We further analyzed the turnover rates of NADH and Fd_{rd} to illuminate the experimentally observed phenotypes. For growth on cellobiose, NADH turnover rates slightly increased as specific growth rates increased (Figure 5.6A), which well correlated with the enhanced ethanol fluxes leading to higher E:A ratios observed experimentally. In contrast, the decrease in Fd_{rd} turnover manifested with a general decrease in hydrogen production, providing more electrons available for ethanol biosynthesis. For growth on cellulose, the levels of NADH and Fd_{rd} turnover rates were fairly level across growth rates, clearly explaining lower ethanol production and hence lower E:A ratios as observed experimentally.

Central Carbon Metabolism

We next examined the effect of the observed constraints on several key reactions of central carbon metabolism from phosphoenolpyruvate (PEP) to pyruvate to acetyl-CoA. For the conversion of PEP to pyruvate, simulations of both carbon sources predicted substantial flux through phosphoenolpyruvate carboxykinase (PEPCK), which is the first step in the malate shunt. The PEPCK activity increased with the increasing growth rates for both carbon sources even though the cellobiose simulations had a much tighter distribution (Figure 5.7A,B). Direct conversion to pyruvate through pyruvate:pyrophosphate dikinase (PPDK), however, remained fairly constant for cellulose simulations but increased for cellobiose simulations with increasing growth rates (Figure 5.7C,D). Regardless of cellulosic substrates and growth rates, PPDK fluxes were much lower than PEPCK fluxes. This simulation result clearly highlights the significant role of the PEPCK-dependent malate shunt on bioenergetics of *C. thermocellum* by generating energy in terms of GTP and producing NADPH from NADH, both of which are required for biomass synthesis and affect the experimentally observed ethanol production.

For the conversion of pyruvate to acetyl-CoA, the simulations for both carbon sources predicted fluxes through pyruvate:ferredoxin oxidoreductase (PFOR) were relatively constant under different growth rates, although the distribution was much wider for cellulose simulations (Figure 5.8A,B). This might hint at less metabolic flexibility in the PFOR reaction when growing on cellobiose. Fluxes through pyruvate:formate lyase (PFL) were lower than PFOR fluxes in both conditions across growth rates (Figure 5.8C,D), but the cellobiose cultures were predicted to have higher PFL fluxes than cellulose cultures. The latter implies that ethanol production might be limiting in cellobiose cultures because PFL is known to function as a metabolic valve to relieve redox imbalance (Thompson et al., 2015).

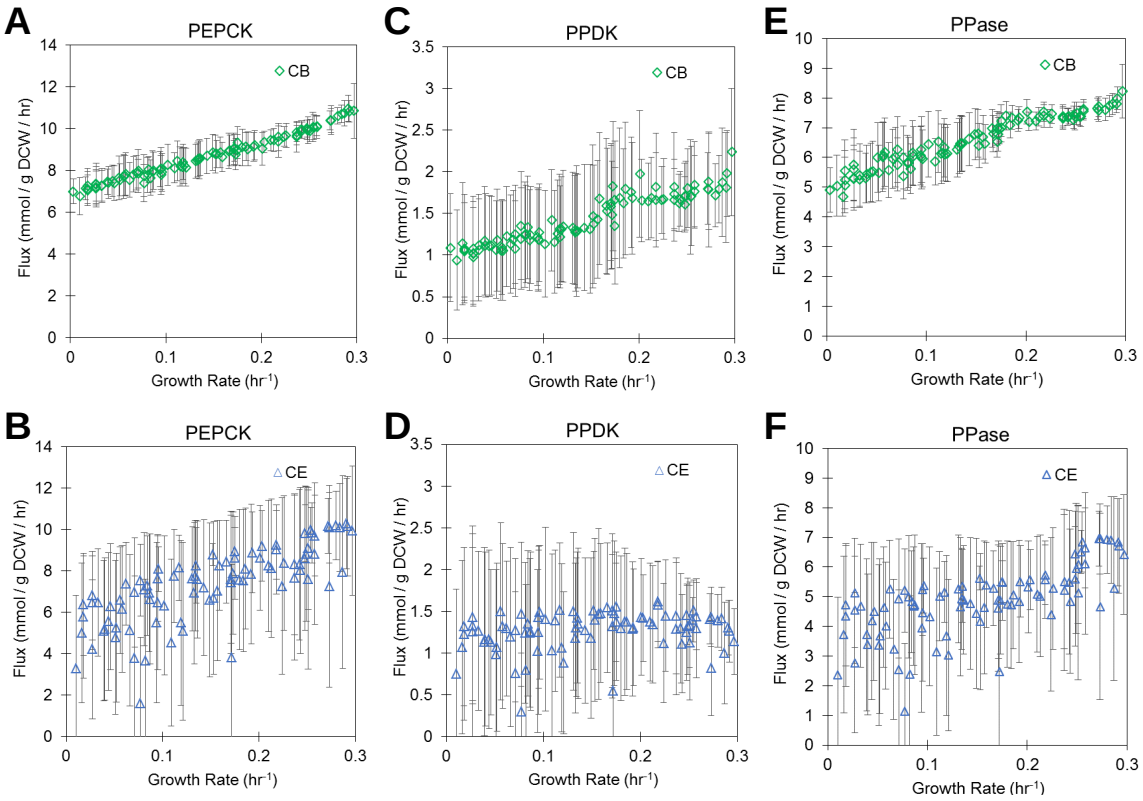

Figure 5.7: Average flux through key reactions involved in converting PEP to pyruvate as well as the membrane bound pyrophosphatase. Upper panels in green diamonds are cellobiose simulations, lower panels in blue triangles are cellulose simulations.

Redox Metabolism

We further examined the effect of the observed constraints on individual redox reactions. Regardless of cellulosic substrates, hydrogen production through the [NiFe] energy-conserving hydrogenase (ECH) increased almost exponentially as the growth rate increased, especially for E:A > 1 (or $\mu > 0.18$ hr^{-1}) (Figure 5.9A,B). In contrast, hydrogen production through the bifurcating hydrogenase (BIF) dropped significantly for cellobiose with an increase in growth rates, yet it remained fairly consistent across growth rates for cellulose (Figure 5.9C,D). This translates to a decrease in hydrogen production in cellobiose cultures while hydrogen production remains fairly constant in cellulose cultures. The conversion of reduced ferredoxin to NADH

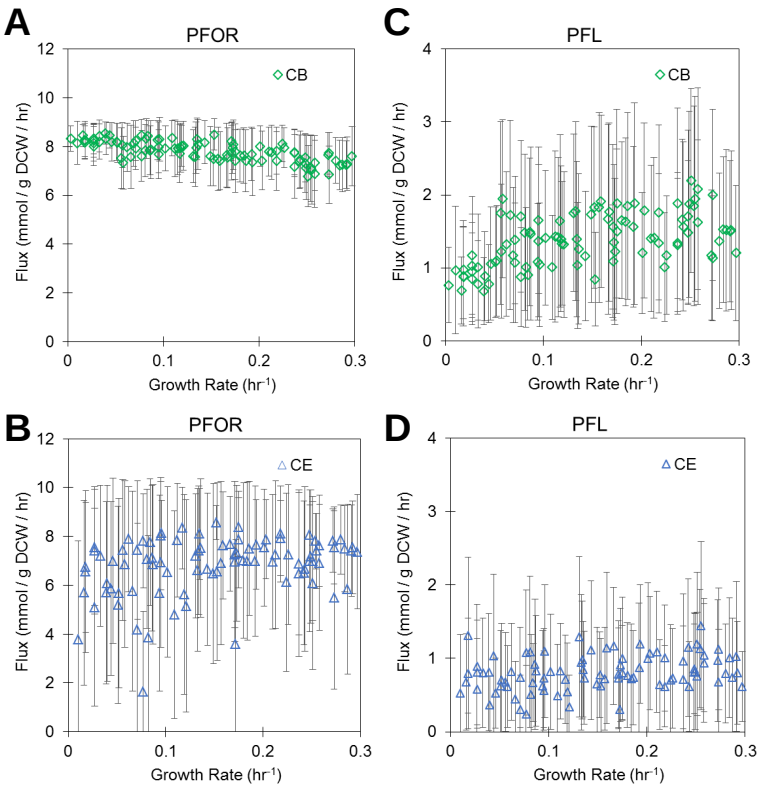

Figure 5.8: Average flux through reactions involved in synthesizing acetyl-CoA from pyruvate. Upper panels in green diamonds are cellobiose simulations, lower panels in blue triangles are cellulose simulations.

through reduced ferredoxin:NADH oxidoreductase (RNF) was significantly greater for cellobiose simulations (Figure 5.9E,F), particularly at high growth rates where cellulose simulations did not use RNF at all. Interestingly, the flux through RNF was parabolic in shape on cellobiose with increasing growth rate with an inflection point occurring with E:A $\tilde{1}$. Additionally, the flux through NADH dependent reduced ferredoxin:NADP+ oxidoreductase (NFN) steadily increased with growth rate, because NADPH preferred for synthesis of cell precursors (Alberts et al., 2002), (Figure 5.9G,H).

The observed constraints are anticipated to increase the requirement for NADH in cellobiose cultures and the simulations corroborate this expectation, in particular by increasing the RNF flux and decreasing the BIF flux on cellobiose. Taken

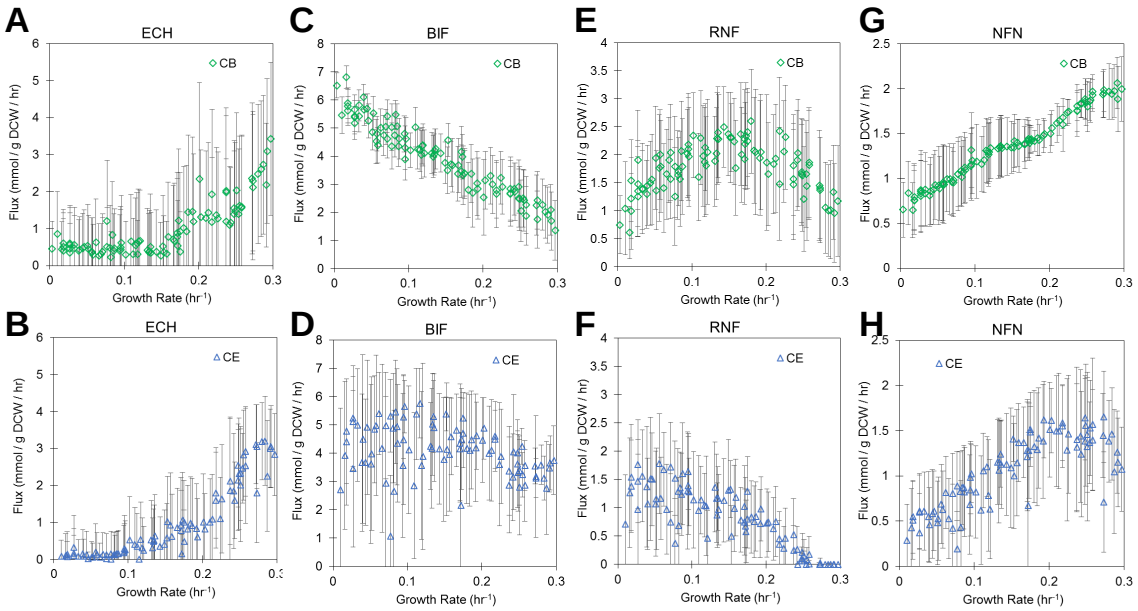

Figure 5.9: Average flux through important redox related reactions. Upper panels in green diamonds are cellobiose simulations, lower panels in blue triangles are cellulose simulations.

altogether, these results illustrate how *C. thermocellum* restructures its metabolism during growth on different carbon sources.

5.3 Discussion

In this work, we have constructed the novel genome-scale model (GEM) of *C. thermocellum* DSM1313 *i*AT601. After extensive refinement with literature reports, we calculated the ATP requirements for growth-associated maintenance and cellulosome synthesis by fitting experimental data. With this model, we explored complex cellular phenotypes and model-guided strain design strategies for producing valuable chemicals. It is important to consider cellular phenotypes under different conditions in order to broadly understand and predict cellular behavior.

In particular, the cascade of carbon from PEP to pyruvate to acetyl-CoA provides key precursors for DCW synthesis, and when coupled to the complex redox

metabolism of *C. thermocellum*, consists of alternative means of generating energy and shuttling electrons (Thompson et al., 2015). Examining these reactions at various growth rates and on different substrates is an effective way to explore the bioenergetics of *C. thermocellum*. Generally, the flux distributions from cellobiose simulations were not as broadly distributed as flux distributions from cellulose simulations. This tightening of flux distributions implies

i tighter regulatory mechanisms are imposed during growth on cellobiose than on cellulose at these key metabolic nodes, and/ or

ii that bioenergetic constraints on carbon and electron flow limit the metabolic flexibility during growth on cellobiose

5.3.1 Proposed Bioenergetic Regulatory Mechanism of *C. thermocellum* Fermentation

Taken altogether, we can use the simulation results presented along with literature reports to propose a mechanism which explains the metabolic differences between cellobiose and cellulose cultures of *C. thermocellum*. Four key, interrelated motifs can help to shed light on this mechanism: Motif 1 energy modulation via acetate production, Motif 2 redox metabolism, Motif 3 regulation of PEP to pyruvate conversion, and Motif 4 PFL-dependent redox relief valve.

Motif 1

Energy modulation via acetate production is one of the critical motifs regulating bioenergetics of *C. thermocellum*. Acetate production during growth on cellobiose drops because less PTA-ACK activity would be necessary to generate the required ATP for cellulosome synthesis, which has been shown experimentally (Zhang and Lynd, 2005a). This is highlighted by the cofactor turnover in our simulations, where ATP turnover is higher for cellulose than cellobiose simulations (Figure 5.6). It has

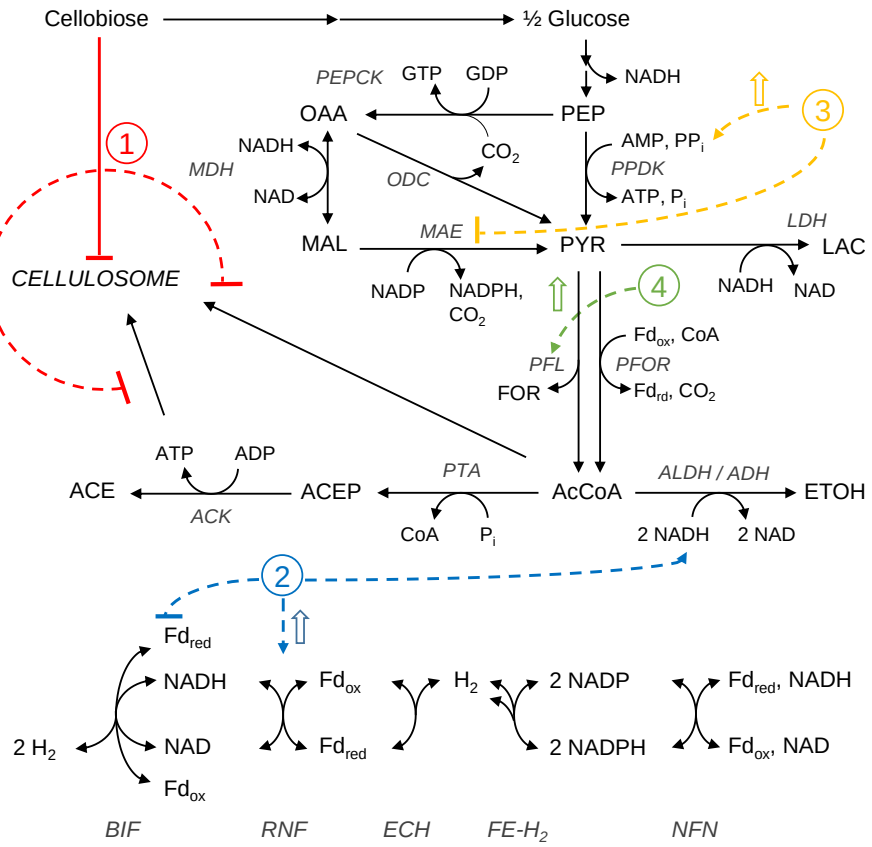

Figure 5.10: Proposed mechanism of bioenergetics influencing *C. thermocellum* during growth on cellulosic substrates. Motif 1: Cellobiose inhibits cellulosome production, and the lower ATP requirement reduces the need for PTA-ACK. Motif 2: RNF activity is upregulated and/or BIF activity is downregulated to convert more reduced ferredoxin to NADH. Motif 3: Increased RNF activity can be used to synthesize PP_i, which alters the conversion of PEP to pyruvate. Motif 4: PFL acts as a redox relief valve and is likely activated by a redox imbalance.

been experimentally observed that cellobiose inhibits synthesis of the cellulosome at both enzymatic (Zhang and Lynd, 2005b) and transcriptomic levels (**?**) (Figure 5.10), however real-time PCR has shown little difference in PTA-ACK expression between cellulose and cellobiose cultures (Stevenson and Weimer, 2005) which suggests cellobiose is not a direct regulator of acetate synthesis. Alternatively, elimination of hydrogen production leads to diminished acetate production in *C. thermocellum*, and it has been proposed that electron perturbations are more influential than PTA-ACK perturbations on ethanol production (Thompson et al., 2015).

Motif 2

Redox metabolism of *C. thermocellum* is very robust and a critical motif in controlling cellular bioenergetics. To account for high E:A ratios for growth on cellobiose, the cell must have an ample supply of NADH. This is facilitated by an increase in RNF and decrease in BIF activities on cellobiose as observed in the simulations (Figures 5.9, 5.10). An increase in RNF flux will also limit the NADH available for hydrogen synthesis by BIF. RNF is expressed during batch growth on cellulose (Raman et al., 2011), and in chemostats expression is significantly higher for cellobiose cultures than for cellulose cultures across growth rates (Riederer et al., 2011). Thus, it is feasible that cellobiose can activate RNF expression, in an opposite phenotype to cellulosome synthesis. Further, the low flux through RNF in all simulations suggests that RNF may be limiting ethanol production by throttling NADH generation.

Motif 3

The motif of PEP to pyruvate conversion in *C. thermocellum* enables it to efficiently regulate cellular bioenergetics and carbon flux. Our simulations provide evidence to a link between RNF and PP_i by ways of the conversion of PEP to pyruvate. The RNF protein complex is embedded in the membrane and couples proton export to ferredoxin oxidation to form NADH. The generated proton motive force can be used to drive ATP and/or PP_i synthesis (Raman et al., 2011; Zhou et al., 2013), which is consistent with the higher PPase flux in cellobiose culture simulations (Figure 5.7E,F). Experimentally, it has been shown with *in vitro* purified enzyme assays that a high concentration of PP_i can enhance PPDK activity and inhibit MAE activity (Taillefer et al., 2015). It is interesting that the allosteric control of PP_i on PPDK and MAE was not included in the model, yet the constraints on ethanol and acetate manifest into an observed increase in PPDK activity in cellobiose simulations (Figure 5.10). An additional consequence of more PPDK flux means less NADH is converted to

NADPH through the malate shunt, and the additional NADH can then go towards ethanol production.

Motif 4

PFL has been described as a redox relief valve in *C. thermocellum* (Thompson et al., 2015). For context, as described above, RNF produces NADH and oxidizes ferredoxin. It has been suggested previously that RNF is the major bottleneck in ethanol production, and the limiting capacity of RNF causes an accumulation of reduced ferredoxin which then leads to an increase in PFL flux (Thompson et al., 2015). This idea comes from experimental evidence where suppressing hydrogen production via chemical inhibitors or genetic manipulations leads to an increase in formate production on cellobiose, and the PFL reaction has been described as an overflow reaction used to generate acetyl-CoA from pyruvate without generating reduced cofactors (Rydzak et al., 2011, 2014). Interestingly, PFL and its activating enzyme are highly expressed across multiple conditions (Rydzak et al., 2012; Raman et al., 2011; Wei et al., 2014), even when no formate production is observed. Expression of PFL without formate production implies a redox related, possibly allosteric, regulatory mechanism (Figure 5.10). While formate production was not reported in most of the training data, in our simulations the cellobiose set has a consistently higher PFL flux (Figure 5.8C,D). The importance of PFL in the production of acetyl-CoA from pyruvate can also be seen when eliminating PFL activity through chemical inhibition or genetic manipulation, which is shown to increase lactate production more than ethanol production (Rydzak et al., 2014, 2015). Under conditions of redox stress, it would be more beneficial to produce ethanol and consume 2 NADH than to produce lactate and consume 1, particularly if PFL is used to generate acetyl-CoA without producing reduced cofactors. However, if PFL activity is not possible, *C. thermocellum* cannot completely balance carbon and redox cofactors to produce ethanol and cell growth, which stalls the conversion of pyruvate to acetyl-CoA and leads to lactate production.

Relationship between Ethanol Production and Cellulose Degradation

Similar to PFL, alcohol dehydrogenases (ADHs) responsible for ethanol production are also seen to be highly expressed under multiple conditions, although ethanol production varies (Rydzak et al., 2012; Raman et al., 2011). The availability of reduced ferredoxin or NADH could then feasibly activate ethanol and formate production as an overflow to relieve redox stress. The availability of NADH as the main activator of ethanol synthesis also makes sense when considering that addition of methyl viologen to cellobiose chemostats led to an increase in ethanol production without significant increase in transcription of ethanol synthesis genes (Sander et al., 2015). More evidence to the overflow behavior of ethanol and formate production can be seen when growing cells in continuous cultures with a lower cellodextrin feed concentration. Under low substrate conditions in rich media, ethanol and formate production are very low, and acetate is the major fermentation product, regardless of using cellobiose or cellulose as a carbon source (Stevenson and Weimer, 2005).

The results presented above offer some interesting suggestions, and to properly understand why cellobiose cultures produce so much more ethanol, it is useful to think of *C. thermocellum* in its native environment, i.e. degrading complex biomass in a co-culture in soil (Viljoen et al., 1926; Akinosho et al., 2014), where there is not likely to be a substantial cellobiose concentration. In fact, in designed co-cultures, *C. thermocellum* prefers to make acetate and hydrogen if these products can be consumed by its cohabitant (Weimer and Zeikus, 1977). Isolated growth on cellobiose, however, can be considered a perturbation away from the native environment since high concentrations of cellobiose have been shown to repress cellulosome synthesis (Zhang and Lynd, 2005b). This repression is key to the increased ethanol production on cellobiose, because by lowering the cellulosome burden by 10 fold (i.e. from 20% DCW to 2% DCW), our calculations above estimate that the cell needs to produce about 11 mmol ATP / g DCW less on cellobiose.

Without the ATP burden of cellulosome synthesis, less flux through PTA-ACK is needed, and this triggers a dramatic restructuring of carbon and electron fluxes to maintain the rates of glycolysis examined here. As a result, there is an increase in RNF flux, which leads to more PPDK flux. Both of these reactions can enhance the supply of NADH and lead to the observed overflow of ethanol production on cellobiose. As RNF reaches its maximum capacity, PFL flux increases to balance the need for acetyl-CoA and the redox state of the cell.

Generally, these results indicate that for growth on cellulose, a high glycolytic flux and sufficient conversion of reduced ferredoxin to NADH (e.g., by eliminating hydrogen production or overexpressing RNF) will be critical for high ethanol production. A significant level of control is necessary to accomplish this goal, although it is still unclear at this point how exactly the proposed motifs are controlled with respect to each other, or what additional regulatory elements might be active with, or instead of, the proposed mechanism above. There are still many questions regarding the bioenergetic control mechanisms which balance carbon and electron fluxes in *C. thermocellum*. However, these questions can be addressed as more OMICs and fermentation datasets become available and are integrated.

Bibliography

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., and Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol*, 9(3):e1002980. 30, 39

Akinosho, H., Yee, K., Close, D., and Ragauskas, A. (2014). The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Frontiers in Chemistry*, 2. 5, 105

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002). *Molecular Biology of the Cell*, volume 4th Ed. Garland Science, New York. 99

Almaas, E., Kovacs, B., Vicsek, T., Oltvai, Z. N., and Barabasi, A. L. (2004). Global organization of metabolic fluxes in the bacterium escherichia coli. *Nature*, 427(6977):839–843. 10.1038/nature02289. 95

Anuradha, R., Suresh, A. K., and Venkatesh, K. V. (1999). Simultaneous saccharification and fermentation of starch to lactic acid. *Process Biochemistry*, 35(34):367–375. 4

Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R., Feinberg, L. F., Olson, D. G., Foden, J. M., Miller, B. B., Lynd, L. R., Hogsett, D. A., and Caiazza, N. C. (2011). High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Applied and Environmental Microbiology*, 77(23):8288–8294. 22, 23

Ay, F. and Kahveci, T. (2010). Functional similarities of reaction sets in metabolic pathways. pages 102–111. 14

Ballesteros, M., Oliva, J., Negro, M., Manzanares, P., and Ballesteros, I. (2004). Ethanol from lignocellulosic materials by a simultaneous saccharification and fermentation process (SSF) with *Kluyveromyces marxianus* cect 10875. *Process Biochemistry*, 39(12):1843–1848. 4

Barney, B. M., Mann, R. L., and Ohlert, J. M. (2013). Identification of a residue affecting fatty alcohol selectivity in wax ester synthase. *Appl. Environ. Microbiol.*, 79(1):396–399. 54

Barney, B. M., Wahlen, B. D., Garner, E., Wei, J., and Seefeldt, L. C. (2012). Differences in substrate specificities of five bacterial wax ester synthases. *Appl. Environ. Microbiol.*, 78(16):5734–5745. 49

Bartek, T., Blombach, B., Zoennchen, E., Makus, P., Lang, S., Eikmanns, B. J., and Oldiges, M. (2010). Importance of NADPH supply for improved L-valine formation in *Corynebacterium glutamicum*. *Biotechnology Progress*, 26(2):361–371. 16

Bayer, E. A., Kenig, R., and Lamed, R. (1983). Adherence of *Clostridium thermocellum* to cellulose. *Journal of Bacteriology*, 156(2):818–827. 80

Becker, J. V., Armstrong, G. O., van der Merwe, M. J., Lambrechts, M. G., Vivier, M. A., and Pretorius, I. S. (2003). Metabolic engineering of *Saccharomyces cerevisiae* for the synthesis of the wine-related antioxidant resveratrol. *FEMS Yeast Res*, 4(1):79–85. 5

Becker, S. A. and Palsson, B. Ø. (2005). Genome-scale reconstruction of the metabolic network in *Staphylococcus aureus* N315: an initial draft to the two-dimensional annotation. *BMC Microbiology*, 5(1):1–12. 12

Behre, J., Wilhelm, T., von Kamp, A., Ruppin, E., and Schuster, S. (2008). Structural robustness of metabolic networks with respect to multiple knockouts. *Journal of theoretical biology*, 252(3):433–441. 14

Bernard, T., Bridge, A., Morgat, A., Moretti, S., Xenarios, I., and Pagni, M. (2014). Reconciliation of metabolites and biochemical reactions for metabolic networks. *Briefings in Bioinformatics*, 15(1):123–135. 10.1093/bib/bbs058. 33

Beuster, G., Zarse, K., Kaleta, C., Thierbach, R., Kiehntopf, M., Steinberg, P., Schuster, S., and Ristow, M. (2011). Inhibition of alanine aminotransferase *in silico* and *in vivo* promotes mitochondrial metabolism to impair malignant growth. *Journal of Biological Chemistry*, 286(25):22323–22330. 14

Binns, M., de Atauri, P., Vlysidis, A., Cascante, M., and Theodoropoulos, C. (2015). Sampling with poling-based flux balance analysis: optimal versus sub-optimal flux space analysis of *Actinobacillus succinogenes*. *BMC Bioinformatics*, 16(1):1–19. 12

Biswas, R., Prabhu, S., Lynd, L. R., and Guss, A. M. (2014). Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of *Clostridium thermocellum*. *PLoS ONE*, 9(2):e86389. 57, 80

Biswas, R., Zheng, T., Olson, D., Lynd, L., and Guss, A. (2015). Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum*. *Biotechnology for Biofuels*, 8(1):20. 22, 30, 57, 62, 63, 75, 81

Black, P. N. and DiRusso, C. C. (2007). Yeast acyl-CoA synthetases at the crossroads of fatty acid metabolism and regulation. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1771(3):286–298. 46, 51

Blazeck, J., Garg, R., Reed, B., and Alper, H. S. (2012). Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering*, 109(11):2884–2895. 54

Blumer-Schuette, S. E., Brown, S. D., Sander, K. B., Bayer, E. A., Kataeva, I., Zurawski, J. V., Conway, J. M., Adams, M. W., and Kelly, R. M. (2014).

Thermophilic lignocellulose deconstruction. *FEMS Microbiol Rev*, 38(3):393–448. 5

Boeke, J. D., Croute, F., and Fink, G. R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular and General Genetics MGG*, 197(2):345–346. 20

Bokinsky, G., Peralta-Yahya, P. P., George, A., Holmes, B. M., Steen, E. J., Dietrich, J., Soon Lee, T., Tullman-Ercek, D., Voigt, C. A., Simmons, B. A., and Keasling, J. D. (2011). Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 108(50):19949–19954. 4

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol*, 6(7):e1000859. 95

Brener, D. and Johnson, B. (1984). Relationship between substrate concentration and fermentation product ratios in *Clostridium thermocellum* cultures. *Appl Environ Microbiol*, 47(5):1126 – 1129. 81

Brochado, A. R., Matos, C., Moller, B. L., Hansen, J., Mortensen, U. H., and Patil, K. R. (2010). Improved vanillin production in baker's yeast through *in silico* design. *Microb Cell Fact*, 9:84. 5

Brown, S. D., Guss, A. M., Karpinets, T. V., Parks, J. M., Smolin, N., Yang, S., Land, M. L., Klingeman, D. M., Bhandiwad, A., Rodriguez, Miguel, J., Raman, B., Shao, X., Mielenz, J. R., Smith, J. C., Keller, M., and Lynd, L. R. (2011). Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33):13752–13757. 56

Bull, A. T. (2004). Biotechnology, the art of exploiting biology. In *Microbial Diversity and Bioprospecting*, pages 3–10. American Society of Microbiology. 2

Burgard, A. P., Pharkya, P., and Maranas, C. D. (2003). OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657. 12

Cakir, T., Kirdar, B., Onsan, Z. I., Ulgen, K. O., and Nielsen, J. (2007). Effect of carbon source perturbations on transcriptional regulation of metabolic fluxes in *Saccharomyces cerevisiae*. *BMC systems biology*, 1(Journal Article):18. 14

Cakir, T., Kirdar, B., and Ulgen, K. O. (2004). Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. *Biotechnology and bioengineering*, 86(3):251–260. 14

Carere, C., Rydzak, T., Cicek, N., Levin, D., and Sparling, R. (2014). Role of transcription and enzyme activities in redistribution of carbon and electron flux in response to N2 and H2 sparging of open-batch cultures of *Clostridium thermocellum* ATCC 27405. *Applied Microbiology and Biotechnology*, 98(6):2829–2840. 82

Carere, C., Rydzak, T., Verbeke, T., Cicek, N., Levin, D., and Sparling, R. (2012). Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H2 and ethanol-producing bacteria. *BMC Microbiology*, 12(1):295. 30

Carlson, R., Fell, D., and Srienc, F. (2002). Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnology and Bioengineering*, 79(2):121–134. 17

Carlson, R. and Srienc, F. (2004a). Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: creation of overall flux states. *Biotech Bioeng*, 86(2):149–162. 14

Carlson, R. and Srienc, F. (2004b). Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Identification of reactions. *Biotech Bioeng*, 85(1):1–18. 14

Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Pujar, A., Shearer, A. G., Travers, M., Weerasinghe, D., Zhang, P., and Karp, P. D. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*, 40(Database issue):D742–53. 32

Centler, F., Kaleta, C., Speroni di Fenizio, P., and Dittrich, P. (2010). A parallel algorithm to compute chemical organizations in biological networks. *Bioinformatics*, 26(14):1788–1789. 14

Chandrasekaran, S. and Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proceedings of the National Academy of Sciences*, 107(41):17845–17850. 11, 12

Chen, N., Du, J., Liu, H., and Xu, Q. Y. (2009). Elementary mode analysis and metabolic flux analysis of L-glutamate biosynthesis by *Corynebacterium glutamicum*. *Annals of Microbiology*, 59(2):317–322. 16

Chen, X., Nielsen, K., Borodina, I., Kielland-Brandt, M., and Karhumaa, K. (2011a). Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. *Biotechnology for Biofuels*, 4(1):21. 5

Chen, Z., Liu, H., Zhang, J., and Liu, D. (2010). Elementary mode analysis for the rational design of efficient succinate conversion from glycerol by *Escherichia coli*. *Journal of biomedicine and biotechnology*, 2010:518743. 17

Chen, Z., Liu, H. J., and Liu, D. H. (2011b). Metabolic pathway analysis of 1,3-propanediol production with a genetically modified *Klebsiella pneumoniae* by overexpressing an endogenous NADPH-dependent alcohol dehydrogenase. *Biochemical Engineering Journal*, 54(3):151–157. 17

Chirala, S. S. (1992). Coordinated regulation and inositol-mediated and fatty acid-mediated repression of fatty acid synthase genes in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 89(21):10232–10236. 51

Choi, H. S., Lee, S. Y., Kim, T. Y., and Woo, H. M. (2010). *In silico* identification of gene amplification targets for improvement of lycopene production. *Applied and Environmental Microbiology*, 76(10):3097–3105. 13

Christensen, B., Karoly Gombert, A., and Nielsen, J. (2002). Analysis of flux estimates based on 13C-labelling experiments. *European Journal of Biochemistry*, 269(11):2795–2800. 10

Chung, D., Cha, M., Guss, A. M., and Westpheling, J. (2014). Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. *Proceedings of the National Academy of Sciences*, 111(24):8931–8936. 4

Covert, M. W. and Palsson, B. O. (2002). Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *Journal of Biological Chemistry*, 277(31):28058–28064. 12

Covert, M. W. and Palsson, B. O. (2003). Constraints-based models: Regulation of gene expression reduces the steady-state solution space. *Journal of Theoretical Biology*, 221(3):309 – 325. 12

Covert, M. W., Schilling, C. H., and Palsson, B. O. (2001). Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1):73 – 88. 12

Demain, A. L. (2009). Biosolutions to the energy problem. *J Ind Microbiol Biotechnol*, 36(3):319–32. 2, 56, 80

Demain, A. L., Newcomb, M., and Wu, J. H. (2005). Cellulase, Clostridia, and ethanol. *Microbiology and molecular biology reviews : MMBR*, 69(1):124–154. 2, 80

Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Shaw, A. J., and Lynd, L. R. (2013). Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. *Metabolic Engineering*, 15:151–158. 27, 57, 81

Diniz, S. C., Voss, I. V., and Steinbuchel, A. (2006). Optimization of cyanophycin production in recombinant strains of *Pseudomonas putida* and *Ralstonia eutropha* employing elementary mode analysis and statistical experimental design. *Biotechnology and bioengineering*, 93(4):698–717. 17

Driouch, H., Melzer, G., and Wittmann, C. (2012). Integration of *in vivo* and *in silico* metabolic fluxes for improvement of recombinant protein production. *Metabolic Engineering*, 14(1):47–58. 16

Dror, T. W., Morag, E., Rolider, A., Bayer, E. A., Lamed, R., and Shoham, Y. (2003). Regulation of the cellulosomal celS (cel48A) gene of *Clostridium thermocellum* is growth rate dependent. *Journal of Bacteriology*, 185(10):3042–3048. 84

Durot, M., Bourguignon, P.-Y., and Schachter, V. (2009). Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*, 33(1):164–190. 7

Ellis, L. D., Holwerda, E. K., Hogsett, D., Rogers, S., Shao, X., Tschaplinski, T., Thorne, P., and Lynd, L. R. (2012). Closing the carbon balance for fermentation by *Clostridium thermocellum* atcc 27405. *Bioresource Technology*, 103(1):293–299. 56

Fan, L., Liu, J., Nie, K., Liu, L., Wang, F., Tan, T., and Deng, L. (2013). Synthesis of medium chain length fatty acid ethyl esters in engineered *Escherichia coli* using endogenously produced medium chain fatty acids. *Enzyme and Microbial Technology*, 53(2):128–133. 43

Fang, F., Salmon, K., Shen, M. W. Y., Aeling, K. A., Ito, E., Irwin, B., Tran, U. P. C., Hatfield, G. W., Da Silva, N. A., and Sandmeyer, S. (2011). A vector set for systematic metabolic engineering in *Saccharomyces cerevisiae*. *Yeast*, 28(2):123–136. 19, 20, 21

Farhi, M., Marhevka, E., Masci, T., Marcos, E., Eyal, Y., Ovadis, M., Abeliovich, H., and Vainstein, A. (2011). Harnessing yeast subcellular compartments for the production of plant terpenoids. *Metab Eng*, 13(5):474–81. 5

Farrell, A., Plevin, R., Turner, B., Jones, A., O'Hare, M., and Kammen, D. (2006). Ethanol can contribute to energy and environmental goals. *Science*, 311(5760):506 – 508. 3

Feddersen, S., Neergaard, T. B. F., Knudsen, J., and Faergeman, N. J. (2007). Transcriptional regulation of phospholipid biosynthesis is linked to fatty acid metabolism by an acyl-CoA-binding-protein-dependent mechanism in *Saccharomyces cerevisiae*. *Biochemical Journal*, 407:219–230. 47

Feinberg, L., Foden, J., Barrett, T., Davenport, K. W., Bruce, D., Detter, C., Tapia, R., Han, C., Lapidus, A., Lucas, S., Cheng, J.-F., Pitluck, S., Woyke, T., Ivanova, N., Mikhailova, N., Land, M., Hauser, L., Argyros, D. A., Goodwin, L., Hogsett, D., and Caiazza, N. (2011). Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM 1313. *Journal of Bacteriology*, 193(11):2906–2907. 17, 28, 30, 62, 70, 82

Feist, A. and Palsson, B. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nat Biotech*, 26(6):659 – 667. 11, 81

Feist, A. M. and Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349. 11

Fell, D. A. (1992). Metabolic control analysis: a survey of its theoretical and experimental development. *The Biochemical journal*, 286 (Pt 2)(Pt 2):313–330. 8

Flowers, D., Thompson, R. A., Birdwell, D., Wang, T., and Trinh, C. T. (2013). SMET: Systematic Multiple Enzyme Targeting a method to rationally design optimal strains for target chemical overproduction. *Biotechnology journal*, 8(5):605–618. 7, 16

Flynn, C. M., Hunt, K. A., Gralnick, J. A., and Srienc, F. (2012). Construction and elementary mode analysis of a metabolic model for *Shewanella oneidensis* MR-1. *Bio Systems*, 107(2):120–8. 14

Fukuda, H., Kondo, A., and Noda, H. (2001). Biodiesel fuel production by transesterification of oils. *Journal of Bioscience and Bioengineering*, 92(5):405–416. 43

Gabriel, C. L. (1928). Butanol fermentation process1. *Industrial and Engineering Chemistry*, 20(10):1063–1067. 2

Ganter, M., Bernard, T., Moretti, S., Stelling, J., and Pagni, M. (2013). MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics*, 29(6):815–816. 10.1093/bioinformatics/btt036. 33

Gardner, J. G. and Keating, D. H. (2010). Requirement of the type II secretion system for utilization of cellulosic substrates by *Cellvibrio japonicus*. *Applied and*

Environmental Microbiology, 76(15):5079–5087. Cited By :16 Export Date: 18 March 2016. 4

Geertz-Hansen, H. M., Blom, N., Feist, A. M., Brunak, S., and Petersen, T. N. (2014). Cofactory: sequence-based prediction of cofactor specificity of rossmann folds. *Proteins*, 82(9):1819–28. 31, 83

Gibson, D., Young, L., Chuang, R., Venter, J., Hutchison, C., and Smith, H. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, 6:343 – 345. 20

Gibson, D. G. (2011). Enzymatic assembly of overlapping DNA fragments. *Methods Enzymol*, 498:349–61. 20

Gold, N. and Martin, V. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J Bacteriol*, 189(19):6787 – 6795. 35

Gowen, C. M. and Fong, S. S. (2010). Genome-scale metabolic model integrated with RNAseq data to identify metabolic states of *Clostridium thermocellum*. *Biotechnology Journal*, 5(7):759–767. 28, 33, 81

Greene, N., Celik, F., Dale, B., Jackson, M., Jayawardhana, K., Jin, H., Larson, E., Laser, M., Lynd, L., MackKenzie, D., Mark, J., McBride, J., McLaughlin, S., and Saccardi, D. (2004). Growing energy: How biofuels can help end america's oil dependence. 3

Guadalupe Medina, V., Almering, M. J., van Maris, A. J., and Pronk, J. T. (2010). Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor. *Appl Environ Microbiol*, 76(1):190–5. 5

Guedon, E., Desvaux, M., and Petitdemange, H. (2002). Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic engineering. *Applied and*

Environmental Microbiology, 68(1):53–58. Cited By :78 Export Date: 18 March 2016. 4

Guss, A. M., Olson, D. G., Caiazza, N. C., and Lynd, L. R. (2012). Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. *Biotechnology for Biofuels*, 5. 73

Hädicke, O. and Klamt, S. (2010). CASOP: A computational approach for strain optimization aiming at high productivity. *J Biotechnol*, 147(2):88–101. 16

Hädicke, O. and Klamt, S. (2011). Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Engineering*, 13(2):204–213. 16, 77, 90

Hammerschlag, R. (2006). Ethanol's energy return on investment: A survey of the literature 1990-present. *Environmental Science & Technology*, 40(6):1744–1750. 3

Hari Krishna, S., Janardhan Reddy, T., and Chowdary, G. V. (2001). Simultaneous saccharification and fermentation of lignocellulosic wastes to ethanol using a thermotolerant yeast. *Bioresource Technology*, 77(2):193–196. 4

Hasunuma, T., Okazaki, F., Okai, N., Hara, K. Y., Ishii, J., and Kondo, A. (2013). A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. *Bioresource Technology*, 135:513–522. 4, 5

Hecker, D., Bisping, B., and Rehm, H. J. (1990). Continuous glycerol production by the sulphite process with immobilized cells of *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology*, 32(6):627–632. 5

Hegemann, J. H. and Heick, S. B. (2011). *Delete and Repeat: A Comprehensive Toolkit for Sequential Gene Knockout in the Budding Yeast Saccharomyces cerevisiae*, volume 765 of *Methods in Molecular Biology*, book section 12, pages 189–206. Humana Press. 20

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotech*, 28(9):977–982. 7

Herrero, A. A. and Gomez, R. F. (1980). Development of ethanol tolerance in *Clostridium thermocellum*: effect of growth temperature. *Applied and Environmental Microbiology*, 40(3):571–577. 56

Hill, J., Nelson, E., Tilman, D., Polasky, S., and Tiffany, D. (2006). Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. *Proceedings of the National Academy of Sciences*, 103(30):11206–11210. 2

Hiltunen, J. K., Mursula, A. M., Rottensteiner, H., Wierenga, R. K., Kastaniotis, A. J., and Gurvitz, A. (2003). The biochemistry of peroxisomal β-oxidation in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews*, 27(1):35–64. 46

Himmel, M., Ding, S., Johnson, D., Adney, W., Nimlos, M., Brady, J., and Foust, T. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 315(5813):804 – 807. 3

Holwerda, E., Thorne, P., Olson, D., Amador-Noguez, D., Engle, N., Tschaplinski, T., van Dijken, J., and Lynd, L. (2014). The exometabolome of *Clostridium thermocellum* reveals overflow metabolism at high cellulose loading. *Biotechnology for Biofuels*, 7(1):155. 70, 80, 81, 83, 90, 93

Holwerda, E. K., Hirst, K. D., and Lynd, L. R. (2012). A defined growth medium with very low background carbon for culturing *Clostridium thermocellum*. *Journal of Industrial Microbiology & Biotechnology*, 39(6):943–947. 39, 85, 93

Hong, K.-K. and Nielsen, J. (2012). Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cellular and Molecular Life Sciences*, 69(16):2671–2690. 5

Humbird, D., Mohagheghi, A., Dowe, N., and Schell, D. J. (2010). Economic impact of total solids loading on enzymatic hydrolysis of dilute acid pretreated corn stover. *Biotechnology Progress*, 26(5):1245–1251. 88

Hunt, K. A., Folsom, J. P., Taffs, R. L., and Carlson, R. P. (2014). Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, 30(11):1569–1578. 14

Hyeon, J. E., Jeon, W. J., Whang, S. Y., and Han, S. O. (2011). Production of minicellulosomes for the enhanced hydrolysis of cellulosic substrates by recombinant *Corynebacterium glutamicum*. *Enzyme and Microbial Technology*, 48(4-5):371–377. 4

Ibarra, R. U., Edwards, J. S., and Palsson, B. O. (2002). *Escherichia coli* K12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 420(6912):186–189. 10.1038/nature01149. 12

Islam, R., Cicek, N., Sparling, R., and Levin, D. (2006). Effect of substrate loading on hydrogen production during anaerobic fermentation by *Clostridium thermocellum* 27405. *Appl Microbiol Biotechnol*, 72(3):576 – 583. 90

Itoh, H., Wada, M., Honda, Y., Kuwahara, M., and Watanabe, T. (2003). Bioorganosolve pretreatments for simultaneous saccharification and fermentation of beech wood by ethanolysis and white rot fungi. *Journal of Biotechnology*, 103(3):273–280. 4

Jamshidi, N. and Palsson, B. O. (2008). Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*, 4. 7

Janßen, H. J. and Steinbüchel, A. (2014). Production of triacylglycerols in *Escherichia coli* by deletion of the diacylglycerol kinase gene and heterologous overexpression of *atfA* from *Acinetobacter baylyi* ADP1. *Applied Microbiology and Biotechnology*, pages 1–12. 54

Jensen, P., Lutz, K., and Papin, J. (2011). TIGER: Toolbox for Integrating Genome-scale metabolic models, Expression data, and transcriptional Regulatory networks. *BMC Systems Biology*, 5(1):147. 11

Jevremovic, D., Trinh, C. T., Srienc, F., Sosa, C. P., and Boley, D. (2011). Parallelization of nullspace algorithm for the computation of metabolic pathways. *PARALLEL COMPUTING*, 37(6-7):261–278. 14

Jones, D. T. and Woods, D. R. (1986). Acetone-butanol fermentation revisited. *Microbiol Rev*, 50(4):484–524. 2

Kadar, Z., Szengyel, Z., and Reczey, K. (2004). Simultaneous saccharification and fermentation (SSF) of industrial wastes for the production of ethanol. *Industrial Crops and Products*, 20(1):103–110. 4

Kaleta, C., de Figueiredo, L. F., and Schuster, S. (2009). Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Research*, 19(10):1872–1883. 14

Kaleta, C., de Figueiredo, L. F., Werner, S., Guthke, R., Ristow, M., and Schuster, S. (2011). *In silico* evidence for gluconeogenesis from fatty acids in humans. *Plos Computational Biology*, 7(7). 14

Kalscheuer, R., Luftmann, H., and Steinbuchel, A. (2004). Synthesis of novel lipids in *Saccharomyces cerevisiae* by heterologous expression of an unspecific bacterial acyltransferase. *Applied and Environmental Microbiology*, 70(12):7119–7125. 43

Kalscheuer, R., Stolting, T., and Steinbuchel, A. (2006). Microdiesel: Escherichia coli engineered for fuel production. *Microbiol Sgm*, 152:2529 – 2536. 43

Kalscheuer, R., Stoveken, T., and Steinbuchel, A. (2007). Engineered microorganisms for sustainable production of diesel fuel and other oleochemicals from renewable plant biomass. *Intern Sugar J*, 109(1297):16 – 19. 43

Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K., and Uemura, H. (2007). DGA1 (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the $\Delta snf2$ disruptant of *Saccharomyces cerevisiae*. *Biochem J*, 408(1):61–68. 48

Kamp, A. v. and Schuster, S. (2006). Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):1930–1931. 37

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 28(1):27–30. 32

Karimi, K., Emtiazi, G., and Taherzadeh, M. J. (2006). Ethanol production from dilute-acid pretreated rice straw by simultaneous saccharification and fermentation with *Mucor indicus*, *Rhizopus oryzae*, and *Saccharomyces cerevisiae*. *Enzyme and Microbial Technology*, 40(1):138–144. 4

Kauffman, K. J., Prakash, P., and Edwards, J. S. (2003). Advances in flux balance analysis. *Current opinion in biotechnology*, 14(5):491–496. 11

Kaufman, D. E. and Smith, R. L. (1998). Direction choice for accelerated convergence in hit-and-run sampling. *Oper Res*, 46. 12

Kenanov, D., Kaleta, C., Petzold, A., Hoischen, C., Diekmann, S., Siddiqui, R. A., and Schuster, S. (2010). Theoretical study of lipid biosynthesis in wild-type *Escherichia coli* and in a protoplast-type L-form using elementary flux mode analysis. *FEBS Journal*, 277(4):1023–1034. 14

Kim, J. and Reed, J. (2010). OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Systems Biology*, 4(1):53. 13

Kim, J., Reed, J. L., and Maravelias, C. T. (2011). Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS ONE*, 6(9):e24162. 12, 13

Kind, S. and Wittmann, C. (2011). Bio-based production of the platform chemical 1,5-diaminopentane. *Applied Microbiology and Biotechnology*, 91(5):1287–1296. 16

Klamt, S. (2006). Generalized concept of minimal cut sets in biochemical networks. *Bio Systems*, 83(2-3):233–247. 16

Klamt, S. and Gilles, E. D. (2004). Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234. 15

Klamt, S., Saez-Rodriguez, J., and Gilles, E. D. (2007). Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC SYSTEMS BIOLOGY*, 1. 37

Klamt, S. and Stelling, J. (2003). Two approaches for metabolic pathway analysis? *Trends in biotechnology*, 21(2):64–69. 14

Klamt, S. and von Kamp, A. (2011). An application programming interface for CellNetAnalyzer. *Biosystems*, 105(2):162–168. 37

Kridelbaugh, D. M., Nelson, J., Engle, N. L., Tschaplinski, T. J., and Graham, D. E. (2013). Nitrogen and sulfur requirements for *Clostridium thermocellum* and *Caldicellulosiruptor bescii* on cellulosic substrates in minimal nutrient media. *Bioresource Technology*, 130(0):125–135. 23, 32

Kromer, J. O., Wittmann, C., Schroder, H., and Heinzle, E. (2006). Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metabolic Engineering*, 8(4):353–369. 17

Kuyper, M., Toirkens, M. J., Diderich, J. A., Winkler, A. A., van Dijken, J. P., and Pronk, J. T. (2005). Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain. *FEMS Yeast Research*, 5(10):925–934. 57

Lamed, R., Setter, E., and Bayer, E. A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology*, 156(2):828–836. 80

Lamed, R. and Zeikus, J. G. (1980). Ethanol production by thermophilic bacteria: Relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. *Journal of Bacteriology*, 144(2):569–578. 64, 80

Lanen, S. G. V. and Shen, B. (2006). Microbial genomics for the improvement of natural product discovery. *Current Opinion in Microbiology*, 9(3):252 – 260. Ecology and industrial microbiology / Techniques. 2

Larhlimi, A., Blachon, S., Selbig, J., and Nikoloski, Z. (2011). Robustness of metabolic networks: A review of existing definitions. *Biosystems*, 106(1):1–8. 14

Layton, D. S. and Trinh, C. T. (2014). Engineering modular ester fermentative pathways in *Escherichia coli*. *Metabolic Engineering*, 26:77–88. 1

Lee, S., Phalakornkule, C., Domach, M. M., and Grossmann, I. E. (2000). Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers & Chemical Engineering*, 24(2-7):711–716. 12

Lee, W. and DaSilva, N. A. (2006). Application of sequential integration for metabolic engineering of 1,2-propanediol production in yeast. *Metab Eng*, 8(1):58–65. 5

Levin, D. B., Islam, R., Cicek, N., and Sparling, R. (2006). Hydrogen production by *Clostridium thermocellum* 27405 from cellulosic biomass substrates. *International Journal of Hydrogen Energy*, 31(11):1496–1503. 80, 90

Lewis, N. E., Nagarajan, H., and Palsson, B. O. (2012). Constraining the metabolic genotypephenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Micro*, 10(4):291–305. 7, 11, 81

Li, Y., Tschaplinski, T. J., Engle, N. L., Hamilton, C. Y., Rodriguez, M., Liao, J. C., Schadt, C. W., Guss, A. M., Yang, Y., and Graham, D. E. (2012). Combined inactivation of the *Clostridium cellulolyticum* lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations. *Biotechnology for Biofuels*, 5(1):1–13. 4

Liao, J. C., Hou, S. Y., and Chao, Y. P. (1996). Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol Bioeng*, 52(Journal Article):129–140. 17

Lin, P. P., Mi, L., Morioka, A. H., Yoshino, K. M., Konishi, S., Xu, S. C., Papanek, B. A., Riley, L. A., Guss, A. M., and Liao, J. C. (2015). Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metab Eng*, 31:44–52. 90

Lovasz, L. (1999). Hit-and-run mixes fast. *Math Program*, 86. 12

Lun, D. S., Rockwell, G., Guido, N. J., Baym, M., Kelner, J. A., Berger, B., Galagan, J. E., and Church, G. M. (2009). Large-scale identification of genetic design strategies using local search. *Mol Syst Biol*, 5:296. 12

Lynd, L., Weimer, P., van Zyl, W., and Pretorius, I. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev*, 66(3):506 – 577. 4, 6, 80

Lynd, L., Zyl, W., McBride, J., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol*, 16(5):577 – 583. 4, 5, 56, 80

Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., Himmel, M., Keller, M., McMillan, J. D., Sheehan, J., and Wyman, C. E. (2008). How biotech can transform biofuels. *Nature biotechnology*, 26(2):169–172. 3, 56, 80

Machado, D., Costa, R. S., Ferreira, E. C., Rocha, I., and Tidor, B. (2012). Exploring the gap between dynamic and constraint-based models of metabolism. *Metabolic Engineering*, 14(2):112–119. 7

Machado, D. and Herrgard, M. J. (2015). Co-evolution of strain design methods based on flux balance and elementary mode analysis. *Metabolic Engineering Communications*, 2:85 – 92. 9

Mahadevan, R. and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4):264–276. 12

Mahadevan, R., von Kamp, A., and Klamt, S. (2015). Genome-scale strain designs based on regulatory minimal cut sets. *Bioinformatics*. 37

Manivasakam, P. and Schiestl, R. H. (1993). High efficiency transformation of *Saccharomyces cerevisiae* by electroporation. *Nucleic Acids Research*, 21(18):4414–4415. 19

Martnez, I., Bennett, G. N., and San, K.-Y. (2010). Metabolic impact of the level of aeration during cell growth on anaerobic succinate production by an engineered *Escherichia coli* strain. *Metabolic Engineering*, 12(6):499–509. 10

Matsuda, F., Furusawa, C., Kondo, T., Ishii, J., Shimizu, H., and Kondo, A. (2011). Engineering strategy of yeast metabolism for higher alcohol production. *Microbial Cell Factories*, 10(1):70. 17

McAnulty, M. J., Yen, J. Y., Freedman, B. G., and Senger, R. S. (2012). Genome-scale modeling using flux ratio constraints to enable metabolic engineering of Clostridial metabolism *in silico*. *BMC Syst Biol*, 6:42. 36

McBee, R. (1954). The characteristics of *Clostridium thermocellum*. *J Bacteriol*, 67(4):505 – 506. 6

McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nuez, A., Butrym, E. D., Richards, M. P., Wang, C.-s., Cheng, G., Zhao, Z., and Wang, C. (2004). Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17593–17598. 1

Megchelenbrink, W., Huynen, M., and Marchiori, E. (2014). *optGpSampler*: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS ONE*, 9(2):e86587. 12, 41

Milne, C., Eddy, J., Raju, R., Ardekani, S., Kim, P.-J., Senger, R., Jin, Y.-S., Blaschek, H., and Price, N. (2011). Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* NCIMB 8052. *BMC Systems Biology*, 5(1):130. 31, 40

Min, Y., Jin, X., Chen, M., Pan, Z., Ge, Y., and Chang, J. (2011). Pathway knockout and redundancy in metabolic networks. *Journal of Theoretical Biology*, 270(1):63–69. 14

Mohr, G., Hong, W., Zhang, J., Cui, G.-z., Yang, Y., Cui, Q., Liu, Y.-j., and Lambowitz, A. M. (2013). A targetron system for gene targeting in thermophiles and its application in *Clostridium thermocellum*. *PLoS ONE*, 8(7):e69032. 73, 74

Mulder, D. W., Shepard, E. M., Meuser, J. E., Joshi, N., King, P. W., Posewitz, M. C., Broderick, J. B., and Peters, J. W. (2011). Insights into [FeFe]-hydrogenase structure, mechanism, and maturation. *Structure*, 19(8):1038–1052. 62

Mumberg, D., Mller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, 156(1):119–122. 20, 21

Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E. A., Sonenshein, A. L., and Shoham, Y. (2010). *Clostridium thermocellum* cellulosomal

genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proceedings of the National Academy of Sciences*, 107(43):18646–18651. 35, 84

Neidhardt, F., Ingraham, J., and Schaechter, M. (1990). *Physiology of the bacterial cell: a molecular approach*. Sinauer Associates. 28, 32, 35, 86

Neuner, A. and Heinzle, E. (2011). Mixed glucose and lactate uptake by *Corynebacterium glutamicum* through metabolic engineering. *Biotechnology journal*, 6(3):318–329. 17

Ng, T. K. and Zeikus, J. G. (1982). Differential metabolism of cellobiose and glucose by *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. *Journal of Bacteriology*, 150(3):1391–1399. 28

Nochur, S. V., Jacobson, G. R., Roberts, M. F., and Demain, A. L. (1992). Mode of sugar phosphorylation in *Clostridium thermocellum*. *Applied Biochemistry and Biotechnology*, 33(1):33–41. 28

Ohgren, K., Bura, R., Lesnicki, G., Saddler, J., and Zacchi, G. (2007). A comparison between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using steam-pretreated corn stover. *Process Biochemistry*, 42(5):834–839. 4

Okano, K., Tanaka, T., Ogino, C., Fukuda, H., and Kondo, A. (2009). Biotechnological production of enantiomeric pure lactic acid from renewable resources: recent achievements, perspectives, and limits. *Applied Microbiology and Biotechnology*, 85(3):413–423. 4

Olson, D. G. and Lynd, L. R. (2012). Transformation of *Clostridium thermocellum* by electroporation. *Methods in enzymology*, pages 317–330. 82

Olson, D. G., McBride, J. E., Shaw, A. J., and Lynd, L. R. (2012). Recent progress in consolidated bioprocessing. *Current Opinion in Biotechnology*, 23(3):396–405. 4, 5

Olson, D. G., Sparling, R., and Lynd, L. R. (2015). Ethanol production by engineered thermophiles. *Curr Opin Biotechnol*, 33:130–41. 2

Ou, M. S., Ingram, L. O., and Shanmugam, K. T. (2010). L(+)-lactic acid production from non-food carbohydrates by thermotolerant *Bacillus coagulans*. *Journal of Industrial Microbiology & Biotechnology*, 38(5):599–605. 4

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crcy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rckert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., and Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33(17):5691–5702. 33

Ozkan, M., Desai, S. G., Zhang, Y., Stevenson, D. M., Beane, J., White, E. A., Guerinot, M. L., and Lynd, L. R. (2001). Characterization of 13 newly isolated strains of anaerobic, cellulolytic, thermophilic bacteria. *Journal of Industrial Microbiology and Biotechnology*, 27(5):275. 23

Pagani, I., Liolios, K., Jansson, J., Chen, I.-M. A., Smirnova, T., Nosrat, B., Markowitz, V. M., and Kyrpides, N. C. (2012). The Genomes OnLine Database (GOLD) v.4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Research*, 40(D1):D571–D579. 6

Papanek, B., Biswas, R., Rydzak, T., and Guss, A. M. (2015). Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. *Metabolic Engineering*, 32:49–54. 81

Papin, J. A., Price, N. D., Edwards, J. S., and Palsson, B. B. O. (2002a). The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *Journal of theoretical biology*, 215(1):67–82. 14

Papin, J. A., Price, N. D., and Palsson, B. O. (2002b). Extreme pathway lengths and reaction participation in genome-scale metabolic networks. *Genome research*, 12(12):1889–1900. 14

Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., and Palsson, B. O. (2004). Comparison of network-based pathway analysis methods. *Trends in biotechnology*, 22(8):400–405. 14

Park, J. H., Lee, K. H., Kim, T. Y., and Lee, S. Y. (2007). Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. *Proceedings of the National Academy of Sciences*, 104(19):7797–7802. 11

Patil, K., Rocha, I., Forster, J., and Nielsen, J. (2005). Evolutionary programming as a platform for *in silico* metabolic engineering. *BMC Bioinformatics*, 6(1):308. 13

Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., and Lee, T. S. (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nat Commun*, 2:483. 5

Peres, S., Vallee, F., Beurton-Aimar, M., and Mazat, J. P. (2011). ACoM: A classification method for elementary flux modes based on motif finding. *Biosystems*, 103(3):410–419. 14

Pfeiffer, T., Sanchez-Valdenebro, I., Nuno, J. C., Montero, F., and Schuster, S. (1999). METATOOL: for studying metabolic networks. *Bioinformatics*, 15(3):251–257. 37

Pharkya, P., Burgard, A., and Maranas, C. (2004). OptStrain: a computational framework for redesign of microbial production systems. *Genome Res*, 14:2367 – 2376. 12

Pharkya, P., Burgard, A. P., and Maranas, C. D. (2003). Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnology and Bioengineering*, 84(7):887–899. 12

Pharkya, P. and Maranas, C. D. (2006). An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metabolic Engineering*, 8(1):1–13. 12

Pirt, S. (1965). The maintenance energy of bacteria in growing cultures. *Proceedings of the Royal Society of London B: Biological Sciences*, 163(991):224–231. 86

Poolman, M. G., Fell, D. A., and Raines, C. A. (2003). Elementary modes analysis of photosynthate metabolism in the chloroplast stroma. *European journal of biochemistry / FEBS*, 270(3):430–439. 14

Poolman, M. G., Sebu, C., Pidcock, M. K., and Fell, D. A. (2007). Modular decomposition of metabolic systems via null-space analysis. *Journal of theoretical biology*, 249(4):691–705. 14

Poolman, M. G., Venkatesh, K. V., Pidcock, M. K., and Fell, D. A. (2004). A method for the determination of flux in elementary modes, and its application to *Lactobacillus rhamnosus*. *Biotechnology and bioengineering*, 88(5):601–612. 14, 37

Price, N. D., Famili, I., Beard, D. A., and Palsson, B. O. (2002a). Extreme pathways and kirchhoff's second law. *Biophys J*, 83(5):2879–82. 39

Price, N. D., Papin, J. A., and Palsson, B. O. (2002b). Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis. *Genome research*, 12(5):760–769. 14

Price, N. D., Reed, J. L., and Palsson, B. O. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature reviews.Microbiology*, 2(11):886–897. 11

Price, N. D., Reed, J. L., Papin, J. A., Famili, I., and Palsson, B. O. (2003). Analysis of metabolic capabilities using singular value decomposition of extreme pathway matrices. *Biophysical journal*, 84(2 Pt 1):794–804. 14

Qureshi, N., Saha, B. C., Hector, R. E., Hughes, S. R., and Cotta, M. A. (2008). Butanol production from wheat straw by simultaneous saccharification and fermentation using *Clostridium beijerinckii*: Part Ibatch fermentation. *Biomass and Bioenergy*, 32(2):168–175. 4

Raab, A. M., Gebhardt, G., Bolotina, N., Weuster-Botz, D., and Lang, C. (2010). Metabolic engineering of *Saccharomyces cerevisiae* for the biotechnological production of succinic acid. *Metab Eng*, 12(6):518–25. 5

Ragauskas, A., Akinosho, H., Yee, K., and Close, D. (2014). The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. *Frontiers in Chemistry*, 2. 56

Rajvanshi, M. and Venkatesh, K. V. (2011). Phenotypic characterization of *Corynebacterium glutamicum* under osmotic stress conditions using elementary mode analysis. *Journal of Industrial Microbiology and Biotechnology*, 38(9):1345–1357. 14

Raman, B., McKeown, C. K., Rodriguez, M., J., Brown, S. D., and Mielenz, J. R. (2011). Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. *BMC Microbiol*, 11:134. 103, 104, 105

Raman, B., Pan, C., Hurst, G. B., Rodriguez, Miguel, J., McKeown, C. K., Lankford, P. K., Samatova, N. F., and Mielenz, J. R. (2009). Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: A quantitative proteomic analysis. *PLoS ONE*, 4(4):e5271. 35, 83, 84

Ramkrishna, D. and Song, H.-S. (2012). Dynamic models of metabolism: Review of the cybernetic approach. *AIChE Journal*, 58(4):986–997. 7

Ranganathan, S., Suthers, P. F., and Maranas, C. D. (2010). OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol*, 6(4):1–11. 13

Reder, C. (1988). Metabolic control theory: A structural approach. *Journal of Theoretical Biology*, 135(2):175–201. 8

Reed, J. and Palsson, B. (2004). Genome-Scale *in silico* models of *E. coli* have multiple equivalent phenotypic states: Assessment of correlated reaction subsets that comprise network states. *Genome Research*, 14(9):1797 – 1805. 12

Reed, J., Vo, T., Schilling, C., and Palsson, B. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol*, 4(9):R54. 83

Ren, S., Zeng, B., and Qian, X. (2013). Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. *BMC Bioinformatics*, 14(Suppl 2):S17–S17. 13

Riederer, A., Takasuka, T. E., Makino, S.-i., Stevenson, D. M., Bukhman, Y. V., Elsen, N. L., and Fox, B. G. (2011). Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Applied and Environmental Microbiology*, 77(4):1243–1253. 30, 35, 103

Roberts, S., Gowen, C., Brooks, J. P., and Fong, S. (2010). Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Systems Biology*, 4(1):31. 28, 31, 32, 33, 35, 57, 81, 82, 83, 85

Rocha, M., Maia, P., Mendes, R., Pinto, J. P., Ferreira, E. C., Nielsen, J., Patil, K. R., and Rocha, I. (2008). Natural computation meta-heuristics for the *in silico* optimization of microbial strains. *BMC Bioinformatics*, 9(1):1–16. 13

Rockafellar, R. T. (1970). *Convex analysis*, volume 28 of *Princeton mathematical series.* Princeton University Press, Princeton, N.J. 13

Rogers, P., Chen, J.-S., and Zidwick, M. J. (2013). *Organic Acid and Solvent Production: Acetic, Lactic, Gluconic, Succinic, and Polyhydroxyalkanoic Acids*, pages 3–75. Springer Berlin Heidelberg, Berlin, Heidelberg. 2

Ross, D. (1961). The acetone-butanol fermentation. *Prog Ind Microbiol*, 3:71–90. 2

Röttig, A., Wenning, L., Brker, D., and Steinbchel, A. (2010). Fatty acid alkyl esters: perspectives for production of alternative biofuels. *Applied Microbiology and Biotechnology*, 85(6):1713–1733. 43

Runguphan, W. and Keasling, J. D. (2014). Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metabolic Engineering*, 21(0):103–113. 44, 45, 46, 50

Rydzak, T., Grigoryan, M., Cunningham, Z., Krokhin, O., Ezzati, P., Cicek, N., Levin, D., Wilkins, J., and Sparling, R. (2014). Insights into electron flux through manipulation of fermentation conditions and assessment of protein expression profiles in *Clostridium thermocellum*. *Applied Microbiology and Biotechnology*, pages 1–14. 23, 30, 68, 75, 82, 104

Rydzak, T., Levin, D., Cicek, N., and Sparling, R. (2009). Growth phase-dependant enzyme profile of pyruvate catabolism and end-product formation in *Clostridium thermocellum* ATCC 27405. *J Biotechnol*, 140(3-4):169 – 175. 24, 80

Rydzak, T., Levin, D. B., Cicek, N., and Sparling, R. (2011). End-product induced metabolic shifts in *Clostridium thermocellum* ATCC 27405. *Appl Microbiol Biotechnol*, 92(1):199–209. 76, 104

Rydzak, T., Lynd, L., and Guss, A. (2015). Elimination of formate production in *Clostridium thermocellum*. *Journal of Industrial Microbiology & Biotechnology*, 42(9):1263–1272. 81, 104

Rydzak, T., McQueen, P. D., Krokhin, O. V., Spicer, V., Ezzati, P., Dwivedi, R. C., Shamshurin, D., Levin, D. B., Wilkins, J. A., and Sparling, R. (2012). Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. *BMC Microbiol*, 12:214. 28, 30, 57, 82, 104, 105

Saerens, S. M. G., Verstrepen, K. J., Van Laere, S. D. M., Voet, A. R. D., Van Dijck, P., Delvaux, F. R., and Thevelein, J. M. (2006). The *Saccharomyces cerevisiae* EHT1 and EEB1 genes encode novel enzymes with medium-chain fatty acid ethyl ester synthesis and hydrolysis capacity. *Journal of Biological Chemistry*, 281(7):4446–4456. 49

Salimi, F., Zhuang, K., and Mahadevan, R. (2010). Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. *Biotechnology Journal*, 5(7):726–738. 31

Sander, K., Wilson, C. M., Rodriguez, M., Klingeman, D. M., Rydzak, T., Davison, B. H., and Brown, S. D. (2015). *Clostridium thermocellum* DSM 1313 transcriptional responses to redox perturbation. *Biotechnology for Biofuels*, 8:211. 105

Sauer, M., Branduardi, P., Valli, M., and Porro, D. (2004). Production of l-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*. *Appl Environ Microbiol*, 70(10):6086–91. 5

Sauer, U. (2006). Metabolic networks in motion: 13C-based flux analysis. *Mol Syst Biol*, 2. 10

Sauer, U., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wthrich, K., and Bailey, J. E. (1999). Metabolic Flux Ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *Journal of Bacteriology*, 181(21):6679–6688. 38

Schauble, S., Heiland, I., Voytsekh, O., Mittag, M., and Schuster, S. (2011). Predicting the physiological role of circadian metabolic regulation in the green alga *Chlamydomonas reinhardtii*. *PLoS ONE*, 6(8):e23026. 14

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Shngen, C., Stelzer, M., Thiele, J., and Schomburg, D. (2011). BRENDA, the enzyme information system in 2011. *Nucleic Acids Research*, 39(suppl 1):D670–D676. 6, 33

Schellenberger, J. and Palsson, B. O. (2009). Use of randomized sampling for analysis of metabolic networks. *J Biol Chem*, 284. 95

Schellenberger, J., Park, J., Conrad, T., and Palsson, B. (2010). BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1):213. 6

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmanian, S., Kang, J., Hyduke, D. R., and Palsson, B. O. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox v2.0. *Nat Protoc*, 6(9):1290–1307. 39

Schilling, C. H., Edwards, J. S., Letscher, D., and Palsson, B. O. (2000a). Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol Bioeng*, 71(Journal Article):286–306. 8, 13

Schilling, C. H., Edwards, J. S., and Palsson, B. O. (1999). Toward metabolic phenomics: Analysis of genomic data using flux balances. *Biotechnology Progress*, 15(3):288–295. 11

Schilling, C. H., Letscher, D., and Palsson, B. O. (2000b). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 203(3):229–48. 8, 13

Schjerling, C. K., Hummel, R., Hansen, J. K., Brsting, C., Mikkelsen, J. M., Kristiansen, K., and Knudsen, J. (1996). Disruption of the gene encoding the acyl-coa-binding protein (ACB1) perturbs acyl-coa metabolism in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 271(37):22514–22521. 47

Schuetz, R., Kuepfer, L., and Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol*, 3(Journal Article). 11

Schuster, R. and Schuster, S. (1993). Refined algorithm and computer program calculating all non-negative fluxes admissible in steady states of biochemical reaction systems with and without some fluxes rates fixed. *CABIOS*, 9(1):79–85. 8

Schuster, S. and Hilgetag, S. (1994). On elementary flux modes in biochemical reaction systems at steady state. *J Biol Syst*, 2(2):165–182. 8, 13

Schuster, S., Pfeiffer, T., and Fell, D. A. (2008). Is maximization of molar yield in metabolic networks favoured by evolution? *Journal of Theoretical Biology*, 252(3):497–504. 12

Schwartz, J.-M. and Kanehisa, M. (2006). Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis. *BMC Bioinformatics*, 7(1):186. 14

Segre, D. and Vitkup, D. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*, 99:15112 – 15117. 12

Senger, R. S., Yen, J. Y., and Fong, S. S. (2014). A review of genome-scale metabolic flux modeling of anaerobiosis in biotechnology. *Current Opinion in Chemical Engineering*, 6:33–42. 81

Shao, X., Raman, B., Zhu, M., Mielenz, J., Brown, S., Guss, A., and Lynd, L. (2011). Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of *Clostridium thermocellum*. *Applied Microbiology and Biotechnology*, 92(3):641–652. 56

Shi, S., Valle-Rodriguez, J. O., Khoomrung, S., Siewers, V., and Nielsen, J. (2012). Functional expression and characterization of five wax ester synthases in *Saccharomyces cerevisiae* and their utility for biodiesel production. *Biotechnology for Biofuels*, 5(7):(24 February 2012). 44, 45, 50

Shlomi, T., Berkman, O., and Ruppin, E. (2005). Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*, 102(21):7695–700. 12

Shlomi, T., Eisenberg, Y., Sharan, R., and Ruppin, E. (2007). A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*, 3. 12

Shoham, Y., Lamed, R., and Bayer, E. A. (1999). The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends in Microbiology*, 7(7):275–281. 6, 80

Simeonidis, E. and Price, N. D. (2015). Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology & Biotechnology*, 42(3):327–338. 81

Skory, C. D. (2003). Lactic acid production by *Saccharomyces cerevisiae* expressing a *Rhizopus oryzae* lactate dehydrogenase gene. *Journal of Industrial Microbiology and Biotechnology*, 30(1):22–27. 5

Smith, R. L. (1984). Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Oper Res*, 32. 12

Somerville, C., Youngs, H., Taylor, C., Davis, S. C., and Long, S. P. (2010). Feedstocks for lignocellulosic biofuels. *Science*, 329(5993):790–792. 3

Song, H.-S. and Ramkrishna, D. (2012). Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metabolic Engineering*, 14(2):69–80. 7

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., del Cardayre, S. B., and Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, 463(7280):559–562. 43

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002). Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420. 14

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F. J., and Doyle, J. (2004). Robustness of cellular functions. *Cell*, 118(6):675–685. 14

Stenberg, K., Bollk, M., Rczey, K., Galbe, M., and Zacchi, G. (2000). Effect of substrate and cellulase concentration on simultaneous saccharification and fermentation of steam-pretreated softwood for ethanol production. *Biotechnology and Bioengineering*, 68(2):204–210. 4

Stephanopoulos, G. (2007). Challenges in engineering microbes for biofuels production. *Science*, 315(5813):801–804. 3

Stephanopoulos, G., Aristidou, A. A., and Nielsen, J. H. (1998). *Metabolic engineering : principles and methodologies*. Academic Press, San Diego. 7, 8

Stevenson, D. M. and Weimer, P. J. (2005). Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in

continuous culture. *Applied and Environmental Microbiology*, 71(8):4672–4678. 102, 105

Stouthamer, A. H. (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek*, 39(1):545–565. 28, 32, 34, 86

Stoveken, T., Kalscheuer, R., Malkus, U., Reichelt, R., and Steinbuchel, A. (2005). The Wax Ester Synthase/Acyl Coenzyme A:Diacylglycerol Acyltransferase from *Acinetobacter* sp. strain ADP1: Characterization of a Novel Type of Acyltransferase. *J. Bacteriol.*, 187(4):1369–1376. 54

Strobel, H. J. (1995). Growth of the thermophilic bacterium *Clostridium thermocellum* in continuous culture. *Current Microbiology*, 31(4):210–214. 91, 92

Strobel, H. J., Caldwell, F. C., and Dawson, K. A. (1995). Carbohydrate transport by the anaerobic thermophile *Clostridium thermocellum* LQRI. *Applied and Environmental Microbiology*, 61(11):4012–4015. 36, 91, 92

Szyperski, T., Glaser, R. W., Hochuli, M., Fiaux, J., Sauer, U., Bailey, J. E., and Wuthrich, K. (1999). Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional ^{13}C labeling and two-dimensional NMR spectroscopy. *Metabolic engineering*, 1(3):189–197. 38

Taillefer, M., Rydzak, T., Levin, D. B., Oresnik, I. J., and Sparling, R. (2015). Reassessment of the transhydrogenase/malate shunt pathway in *Clostridium thermocellum* ATCC 27405 through kinetic characterization of malic enzyme and malate dehydrogenase. *Applied and Environmental Microbiology*, 81(7):2423–2432. 103

Taillefer, M. and Sparling, R. (2016). *Glycolysis as the Central Core of Fermentation*, pages 1–23. Springer Berlin Heidelberg, Berlin, Heidelberg. 2

Tan, Y., Lafontaine Rivera, J. G., Contador, C. A., Asenjo, J. A., and Liao, J. C. (2011). Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metabolic Engineering*, 13(1):60–75. 7

Tavares, S., Grotkjaer, T., Obsen, T., Haslam, R. P., Napier, J. A., and Gunnarsson, N. (2011). Metabolic engineering of *Saccharomyces cerevisiae* for production of eicosapentaenoic acid, using a novel $\delta 5$ -desaturase from *Paramecium tetraurelia*. *Appl Environ Microbiol*, 77(5):1854–61. 5

Tehlivets, O., Scheuringer, K., and Kohlwein, S. D. (2007). Fatty acid synthesis and elongation in yeast. *Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids*, 1771(3):255–270. 44

Tempest, D. W. and Neijssel, O. M. (1984). The status of Yatp and maintenance energy as biologically interpretable phenomena. *Annual Reviews in Microbiology*, 38(1):459–513. 86

Thiele, I. and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protocols*, 5(1):93–121. 7, 32

Thompson, R. A., Layton, D. S., Guss, A. M., Olson, D. G., Lynd, L. R., and Trinh, C. T. (2015). Elucidating central metabolic redox obstacles hindering ethanol production in *Clostridium thermocellum*. *Metabolic Engineering*, 32:207–219. 32, 40, 56, 82, 83, 85, 93, 97, 101, 102, 104

Thompson, R. A. and Trinh, C. T. (2014). Enhancing fatty acid ethyl ester production in *Saccharomyces cerevisiae* through metabolic engineering and medium optimization. *Biotechnology and bioengineering*, 111(11):2200–2208. 42

Trinh, C. T. (2012). Elucidating and optimizing *E. coli* metabolisms for obligate anaerobic butanol and isobutanol production. *Appl Microbiol Biotechnol*. 15

Trinh, C. T., Li, J., Blanch, H. W., and Clark, D. S. (2011). Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol. *Appl Environ Microbiol*, 77(14):4894–4904. 15

Trinh, C. T., Liu, Y., and Conner, D. J. (2015). Rational design of efficient modular cells. *Metabolic Engineering*, 32:220 – 231. 16

Trinh, C. T. and Srienc, F. (2009). Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl Environ Microbiol*, 75:6696–6705. 15, 17

Trinh, C. T., Unrean, P., and Srienc, F. (2008). Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Applied and Environmental Microbiology*, 74:3634–3643. 15, 17, 25, 57, 73, 90

Trinh, C. T., Wlaschin, A., and Srienc, F. (2009). Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, 81(5):813–826. 7, 13, 15, 37, 58, 71, 77

Tripathi, S. A., Olson, D. G., Argyros, D. A., Miller, B. B., Barrett, T. F., Murphy, D. M., McCool, J. D., Warner, A. K., Rajgarhia, V. B., Lynd, L. R., Hogsett, D. A., and Caiazza, N. C. (2010). Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. *Appl. Environ. Microbiol.*, 76(19):6591–6599. 57, 73, 80

Trotter, P. J. (2001). The genetics of fatty acid metabolism in *Saccharomyces cerevisiae*. *Annual Review of Nutrition*, 21(1):97–119. 44

Tsai, S.-L., Oh, J., Singh, S., Chen, R., and Chen, W. (2009). Functional assembly of minicellulosomes on the *Saccharomyces cerevisiae* cell surface for cellulose hydrolysis and ethanol production. *Applied and Environmental Microbiology*, 75(19):6087–6093. 43

Tsuchidate, T., Tateno, T., Okai, N., Tanaka, T., Ogino, C., and Kondo, A. (2011). Glutamate production from *beta*-glucan using endoglucanase-secreting *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology*, 90(3):895–901. 4

Unkrig, V., Neugebauer, F. A., and Knappe, J. (1989). The free radical of pyruvate formate-lyase. *European Journal of Biochemistry*, 184(3):723–728. 23

Unrean, P. and Srienc, F. (2011). Metabolic networks evolve towards states of maximum entropy production. *Metabolic Engineering*, 13(6):666–673. 17

Unrean, P. and Srienc, F. (2012). Predicting the adaptive evolution of *Thermoanaerobacterium saccharolyticum*. *Journal of Biotechnology*, 158(4):259–66. 17

Unrean, P., Trinh, C. T., and Srienc, F. (2010). Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. *Metab Eng*, 12(2):112–122. 17

Valle-Rodrguez, J. O., Shi, S., Siewers, V., and Nielsen, J. (2014). Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid ethyl esters, an advanced biofuel, by eliminating non-essential fatty acid utilization pathways. *Applied Energy*, 115(0):226–232. 45, 46, 53

van der Veen, D., Lo, J., Brown, S. D., Johnson, C. M., Tschaplinski, T. J., Martin, M., Engle, N. L., van den Berg, R. A., Argyros, A. D., Caiazza, N. C., Guss, A. M., and Lynd, L. R. (2013). Characterization of *Clostridium thermocellum* strains with disrupted fermentation end-product pathways. *Journal of Industrial Microbiology and Biotechnology*, 40(7):725–734. 23, 57, 81

van Maris, A. J., Geertman, J. M., Vermeulen, A., Groothuizen, M. K., Winkler, A. A., Piper, M. D., van Dijken, J. P., and Pronk, J. T. (2004). Directed evolution of pyruvate decarboxylase-negative *Saccharomyces cerevisiae*, yielding

a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl Environ Microbiol*, 70(1):159–66. 5

van Roermund, C. W. T., IJlst, L., Majczak, W., Waterham, H. R., Folkerts, H., Wanders, R. J. A., and Hellingwerf, K. J. (2012). Peroxisomal fatty acid uptake mechanism in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*, 287(24):20144–20153. 46

van Zyl, W. H., Lynd, L. R., den Haan, R., and McBride, J. E. (2007). Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. *Advances in Biochemical Engineering/Biotechnology*, 108(Journal Article):205–235. 4, 5, 56, 80

Varga, E., Klinke, H. B., Rczey, K., and Thomsen, A. B. (2004). High solid simultaneous saccharification and fermentation of wet oxidized corn stover to ethanol. *Biotechnology and Bioengineering*, 88(5):567–574. 4

Varma, A. and Palsson, B. (1994). Metabolic Flux Balancing: Basic concepts, scientific and practical use. *Nat Biotech*, 12(10):994 – 998. 11, 39

Verduyn, C., Postma, E., Scheffers, W. A., and van Dijken, J. P. (1990). Energetics of *Saccharomyces cerevisiae* in anaerobic glucose-limited chemostat cultures. *Microbiology*, 136(3):405–412. 34

Verwaal, R., Wang, J., Meijnen, J. P., Visser, H., Sandmann, G., van den Berg, J. A., and van Ooyen, A. J. (2007). High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl Environ Microbiol*, 73(13):4342–50. 5

Viljoen, J. A., Fred, E. B., and Peterson, W. H. (1926). The fermentation of cellulose by thermophilic bacteria. *The Journal of Agricultural Science*, 16(01):1–17. 6, 105

Villa, J. A., Cabezas, M., de la Cruz, F., and Moncalin, G. (2013). Identification of folding domains and sequence motifs critical for WS/DGAT acyltransferase

activity, probed by limited proteolysis and mutagenesis. *Applied and Environmental Microbiology*. 54

von Kamp, A. and Klamt, S. (2014). Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput Biol*, 10(1):e1003378. 90

Wang, L., Birol, I., and Hatzimanikatis, V. (2004). Metabolic control analysis under uncertainty: Framework development and case studies. *Biophysical Journal*, 87(6):3750–3763. 7

Wei, H., Fu, Y., Magnusson, L., Baker, J. O., Maness, P.-C., Xu, Q., Yang, S., Bowersox, A., Bogorad, I., Wang, W., Tucker, M. P., Himmel, M. E., and Ding, S.-Y. (2014). Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. *Frontiers in Microbiology*, 5(142). 35, 104

Weimer, P. J. and Zeikus, J. G. (1977). Fermentation of cellulose and cellobiose by *Clostridium thermocellum* in the absence of *Methanobacterium thermoautotrophicum*. *Applied and Environmental Microbiology*, 33(2):289–297. 105

Wessely, F., Bartl, M., Guthke, R., Li, P., Schuster, S., and Kaleta, C. (2011). Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. *Mol Syst Biol*, 7. 14

Westfall, P. J. and Gardner, T. S. (2011). Industrial fermentation of renewable diesel fuels. *Current Opinion in Biotechnology*, 22(3):344–350. 44

Westfall, P. J., Pitera, D. J., Lenihan, J. R., Eng, D., Woolard, F. X., Regentin, R., Horning, T., Tsuruta, H., Melis, D. J., Owens, A., Fickes, S., Diola, D., Benjamin, K. R., Keasling, J. D., Leavell, M. D., McPhee, D. J., Renninger, N. S., Newman, J. D., and Paddon, C. J. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci U S A*, 109(3):E111–8. 5

Wiback, S. J., Famili, I., Greenberg, H. J., and Palsson, B. O. (2004a). Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J Theor Biol*, 228. 12

Wiback, S. J., Mahadevan, R., and Palsson, B. O. (2004b). Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: the *Escherichia coli* spectrum. *Biotechnology and bioengineering*, 86(3):317–331. 14

Wiback, S. J. and Palsson, B. O. (2002). Extreme pathway analysis of human red blood cell metabolism. *Biophysical journal*, 83(2):808–818. 14

Wiechert, W., Mllney, M., Petersen, S., and de Graaf, A. A. (2001). A universal framework for 13C Metabolic Flux Analysis. *Metabolic Engineering*, 3(3):265–283. 10

Wilhelm, T., Behre, J., and Schuster, S. (2004). Analysis of structural robustness of metabolic networks. *Systems biology*, 1(1):114–120. 14

Williams, T., Combs, J., Lynn, B., and Strobel, H. (2007). Proteomic profile changes in membranes of ethanol-tolerant *Clostridium thermocellum*. *Applied Microbiology and Biotechnology*, 74(2):422–432. 57

Wilson, C., Rodriguez, M., Johnson, C., Martin, S., Chu, T., Wolfinger, R., Hauser, L., Land, M., Klingeman, D., Syed, M., Ragauskas, A., Tschaplinski, T., Mielenz, J., and Brown, S. (2013). Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated *Populus* and switchgrass. *Biotechnology for Biofuels*, 6(1):179. 35

Wlaschin, A. P., Trinh, C. T., Carlson, R., and Srienc, F. (2006). The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. *Metabolic engineering*, 8(4):338–352. 14

Wyman, C. E., Spindler, D. D., and Grohmann, K. (1992). Simultaneous saccharification and fermentation of several lignocellulosic feedstocks to fuel ethanol. *Biomass and Bioenergy*, 3(5):301–307. 4

Yomano, L. P., York, S. W., and Ingram, L. O. (1998). Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production. *Journal of industrial microbiology and biotechnology*, 20(2):132–138. 57

Young, J. D., Henne, K. L., Morgan, J. A., Konopka, A. E., and Ramkrishna, D. (2008). Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. *Biotechnology and Bioengineering*, 100(3):542–559. 7

Yu, K. O., Jung, J., Kim, S. W., Park, C. H., and Han, S. O. (2012). Synthesis of FAEEs from glycerol in engineered *Saccharomyces cerevisiae* using endogenously produced ethanol by heterologous expression of an unspecific bacterial acyltransferase. *Biotechnology and Bioengineering*, 109(1):110–115. 5, 43

Yuan, J., Bennett, B. D., and Rabinowitz, J. D. (2008). Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nat. Protocols*, 3(8):1328–1340. 11

Yuan, J., Fowler, W. U., Kimball, E., Lu, W., and Rabinowitz, J. D. (2006). Kinetic flux profiling of nitrogen assimilation in *Escherichia coli*. *Nat Chem Biol*, 2(10):529–530. 11

Yuan, Y., Hoon Yang, T., and Heinzle, E. (2010). 13C metabolic flux analysis for larger scale cultivation using gas chromatography-combustion-isotope ratio mass spectrometry. *Metabolic Engineering*, 12(4):392–400. 11

Yutin, N. and Galperin, M. Y. (2013). A genomic update on Clostridial phylogeny: Gram-negative spore formers and other misplaced Clostridia. *Environ Microbiol*, 15(10):2631–41. 6

Zamboni, N., Fendt, S. M., Ruhl, M., and Sauer, U. (2009). 13C-based metabolic flux analysis. *Nature protocols*, 4(6):878–892. 10

Zamboni, N., Fischer, E., and Sauer, U. (2005). FiatFlux–a software for metabolic flux analysis from 13C-glucose experiments. *BMC Bioinformatics*, 6(Journal Article):209. 10

Zhang, F., Carothers, J. M., and Keasling, J. D. (2012). Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotech*, 30(4):354–359. 54

Zhang, Q. and Xiu, Z. (2009). Metabolic pathway analysis of glycerol metabolism in *Klebsiella pneumoniae* incorporating oxygen regulatory system. *Biotechnology Progress*, 25(1):103–115. 17

Zhang, X. Z., Sathitsuksanoh, N., Zhu, Z., and Percival Zhang, Y. H. (2011). One-step production of lactate from cellulose as the sole carbon source without any other organic nutrient by recombinant cellulolytic *Bacillus subtilis*. *Metabolic Engineering*, 13(4):364–372. 4

Zhang, Y. and Lynd, L. (2003). 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*, 75(2):219 – 227. 23

Zhang, Y.-H. P. and Lynd, L. R. (2004). Kinetics and relative importance of phosphorolytic and hydrolytic cleavage of cellodextrins and cellobiose in cell extracts of *Clostridium thermocellum*. *Applied and Environmental Microbiology*, 70(3):1563–1569. 36, 82

Zhang, Y.-H. P. and Lynd, L. R. (2005a). Cellulose utilization by *Clostridium thermocellum*: Bioenergetics and hydrolysis product assimilation. *Proceedings of*

the National Academy of Sciences of the United States of America, 102(20):7321–7325. 36, 40, 82, 86, 91, 92, 93, 96, 101

Zhang, Y.-H. P. and Lynd, L. R. (2005b). Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. *Journal of Bacteriology*, 187(1):99–106. 33, 35, 82, 83, 84, 88, 90, 93, 102, 105

Zhao, M., Chen, Y., Qu, D., and Qu, H. (2011). TSdb: A database of transporter substrates linking metabolic pathways and transporter systems on a genome scale via their shared substrates. *Science China Life Sciences*, 54(1):60–64. 31

Zhou, J., Olson, D. G., Argyros, D. A., Deng, Y., van Gulik, W. M., van Dijken, J. P., and Lynd, L. R. (2013). Atypical glycolysis in *Clostridium thermocellum*. *Applied and Environmental Microbiology*, 79(9):3000–3008. 28, 30, 31, 57, 59, 82, 83, 103

Vita

Adam Thompson was born to Richard and Kim Thompson on September 16, 1985 in Marietta, Georgia. Adam graduated from Cherokee High School in 2004. He attended Georgia Institute of Technology, earning a Bachelor's of Science degree in 2008 as a member of the inaugural class of Biochemistry graduates. He continued studies at Georgia State University while supported by a graduate assistantship, and earned a Master's of Science degree in Chemistry in 2011. During this time, he worked with Dr. Markus Germann studying the biophysical properties of synthetic nucleotides designed for potential gene therapies. He plans to continue his career in the field of metabolic engineering within his interests of commodity chemical production, agriculture, and health.