8-2016

Improving Metabolic Engineering and Characterization of *Clostridium thermocellum* for Improved Cellulosic Ethanol Production

Beth Alexandra Papanek
University of Tennessee, Knoxville, bpapanek@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Beth Alexandra Papanek entitled "Improving Metabolic Engineering and Characterization of *Clostridium thermocellum* for Improved Cellulosic Ethanol 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.

Adam M. Guss, Major Professor

We have read this dissertation and recommend its acceptance:

Michael L. Simpson, Brian H. Davison, Gary S. Sayler

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

Improving Metabolic Engineering and Characterization of *Clostridium thermocellum* for Improved Cellulosic Ethanol Production

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Beth Alexandra Papanek
August 2016

DEDICATION

To my mom, I would not be where I am or the person I am today without her love and support.

ACKNOWLEDGEMENTS

I would like to give special thanks to my advisor, Dr. Adam Guss for his continuous patience, help and support throughout my PhD process.

I would like to thank my fellow lab members, specifically those who were integral to the completion of this work. Without the help and support of Lauren Riley and Kaela O'Dell, this work would not have been possible.

I would like to thank Dr. Jan Westpheling and Dr. Bob Schmitz at the University of Georgia, and Dr. Dan Olson at Dartmouth University for their invaluable input, support and guidance.

I would also like to thank the BioEnergy Sciences Center. Without the financial support BESC provided, none of this work would have been possible.

ABSTRACT

Biofules are an important option for humanity to move away from its dependence on fossil fuels. Transitioning from food crops to lignocellulosic alternatives for the production of biofuels is equally important. Most commonly, biofuels are produced using a crop such as corn or soybeans to feed sugars to the yeast, *Saccharomyces cerevisiae* for the fermentation of ethanol. Lignocellulosic biofuel production would eliminate the need for food crops and transition to biomass such as switchgrass, poplar, or corn stover. Currently, lignocellulosic biofuel production is limited primarily because of the cost of converting the biomass to fermentable sugars than can then be metabolized by yeast. To overcome this barrier, a process must be employed that can convert lignocellulosic biomass directly to fuels and chemicals quickly and affordably. *Clostridium thermocellum* is one of the most promising candidates for the production of advanced biofuels because of its potential ability to convert cellulose directly to ethanol without the expensive addition of enzymes. Challenges to implementing *C. thermocellum* on an industrial scale still exist including side product formation, slow growth, limited titers, inhibition on high solids loadings, and a limited ability to perform genetic engineering. This thesis considers all of these concerns with *C. thermocellum* and attempts to systematically improve each characteristic to produce an industrially relevant strain of *C. thermocellum* for advanced biofuel production. Metabolic engineering is applied for the elimination of undesirable fermentation products. Laboratory evolution and medium supplementation are used to improve and understand the mechanisms that influence growth rate, and systematic approaches are used to improve transformation for more efficient genetic engineering of *C. thermocellum* in the future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Biofuels as Potential Fossil Fuel Replacements

The Industrial Revolution marked the start of the ubiquitous and voluminous consumption of fossil fuels (Vitousek et al., 1997). Since that time, steadily increasing standards of living have been built on the consumption of fossil fuels. However, the finite nature of these resources constrains equitable growth, instead favoring the development of nations that possess such resources or those that can afford their acquisition. Additionally, when consumed, fossil fuels release previously sequestered CO_2 into the atmosphere (Broecker et al., 1979). This newly released CO_2 is a primary cause of the greenhouse gas effect and the ensuing global climate change currently facing the planet (Solomon et al., 2009). These fundamental problems with the fossil fuels have spurred the need and desire for sustainable alternatives while maintaining the quality of life that many humans have come to enjoy.

The most near-term option to replace fossil fuels utilizes biomass for the production of biofuels. Biomass sequesters carbon temporarily as it grows, it grows continually, and far more regions of the world are capable of producing biomass than other fuel sources (Gustavsson et al., 1995). The term biomass refers to all forms of plant life, but not all are equally promising as alternative fuel sources. The most straightforward biomass application, referred to as first generation biofuel production, is processing corn or sugar cane for ethanol production or soybeans for biodiesel (de Vries et al., 2010) due to the simplicity of conversion to fuels. These methods, however, have been met with public resistance due to the reallocation of the limited market of food crops to an inedible product (Valentine et al., 2012). Pyrolysis and gasification are other options for generating fuels from biomass, but have challenges for industrial application. Pyrolysis processes result in a mixture of chemicals that are difficult to separate for the product of interest and this chemical mixture changes substantially with changes in the reaction conditions (Mohan et al., 2006). Gasification is primarily used for electricity generation, and not currently an option to produce liquid fuels (Bridgwater, 1995).

Second generation, or advanced biofuels use lignocellulosic biomass crops in the place of traditional food crops (Balan et al., 2013). Lignocellulosic plants are composed of lignin, cellulose, and hemicellulose and are not edible plant material for humans. Examples of these crops include, but are not limited to: corn stover, grasses, hard and softwood trees, and sugar cane bagasse. The disadvantage of using lignocellulosic crops for fuel production lies in their complex structure. Whereas corn, soy, and sugar cane have high amounts of free glucose available for direct fermentation by *Saccharomyces cerevisiae,* lignocellulosic crops' main sugar polymers, cellulose and hemicellulose, have

distinct chemical linkages that are more difficult to break compared to starch or sucrose. Unfortunately, none have freely available glucose molecules for direct fermentation (Petrou and Pappis, 2009). To overcome this problem, all polysaccharide components of the lignocellulosic plants must be broken down into fermentable components. Many different approaches exist for the degradation and separation of lignocellulosic biomass. Microbial degradation is the most promising and economical of the approaches and is thus the focus of the remainder of this work.

Consolidated Bioprocessing with *Clostridium thermocellum* for lignocellulosic biofuels

Two primary approaches exist to deconstruct lignocellulosic biomass into fermentable sugars: the first is deconstruction with chemicals and enzymes in order to release xylose and glucose molecules for traditional fermentation (Kumar et al., 2009). This approach allows *S. cerevisiae* to ferment the released sugars into ethanol. However, the chemicals and enzymes required for this method add substantial cost to the biofuel production process, and are therefore not currently an economical option (Klein-Marcuschamer et al., 2012) though it is being demonstrated in pilot plants.

The second approach is to employ a single organism or a consortium of microorganisms that produce enzymes, such as cellulases, to deconstruct lignocellulosic biomass as well as ferment the sugars to end products. This process, originally deemed Direct Microbial Conversion (Hogsett, et al., 1992) now commonly referred to as Consolidated Bioprocessing or CBP (Lynd et al., 2005), and has the potential to be substantially more economical because it eliminates the need for a separate tank and process for enzyme production (Lynd et al., 2002). Currently, there are no known organisms that can produce cellulolytic enzymes as well as ferment the resulting sugars to ethanol at a high yield and titer. Attempts have been made to engineer strains of *S. cerevisiae* to produce cellulolytic enzymes, but the enzymatic activity has not matched that of naturally cellulolytic organisms (Plessis et al., 2009; Yanase et al., 2010). A number of cellulolytic microorganisms could be engineered to perform CBP, and one of the most promising candidates for the industrial production of ethanol, and the focus of the remainder of this work, is *Clostridium thermocellum* (Olson et al., 2012) .

Clostridium thermocellum is a promising microbe for CBP with anaerobic, thermophilic, cellulolytic capabilities as well as a natural ability to produce ethanol (Demain et al., 2005). The presence of a large, extracellular enzyme complex known as the cellulosome makes it efficient at depolymerizing cellulose into fermentable sugars such as cellobiose (Gold and Martin, 2007). Another beneficial characteristic of *C. thermocellum* is its relative genetic tractability in

comparison to other cellulose-fermenting microorganisms (Guss et al., 2012). However, as with most naturally occurring microorganisms, *C. thermocellum* produces compounds other than ethanol during metabolism. Organic acids, excreted amino acids, and hydrogen are all found in the fermentation broth of a growing *C. thermocellum* culture (Ellis et al., 2012b; van der Veen et al., 2013). Each of these products limits achievable ethanol yield by diverting carbon and electron flux away from producing the desired ethanol end product. In order to compete with first generation biofuel processes, these side products must be eliminated and the yield of fermented ethanol must be improved. To accomplish this, genetic engineering of *C. thermocellum* must be done, making the genetic tractability of the organism essential for generating an industrially competitive strain.

A number of genetically engineered *C. thermocellum* DSM1313 mutant strains have been constructed that improve ethanol production by systematically eliminating fermentation side products. First of these targets was to delete the phosphotransacetylase (*pta)* to eliminate acetate production. The Δ*pta* strain eliminated acetate production but still showed wild type levels of ethanol production (Tripathi et al., 2010). Because the elimination of acetate didn't increase ethanol yields, *ldh* (lactate dehydrogenase) and Δ*pta* were deleted in the same strain (Argyros, 2011). In the resulting Δ*ldh* Δ*pta* strain, acetate and lactate were eliminated. The genetic changes were followed with laboratory evolution, and the final strain had ethanol yields 4.2-fold higher than wild type *C. thermocellum*. In another targeted engineering approach, deleting *hydG* and *ech* eliminated H2 production. The Δ*hydG* strain eliminated most H_2 accumulation by removing the hydrogenase maturation protein. This gene deletion eliminated the activity of three of the four hydrogenases present in *C. thermocellum.* Stacking Δ*hydG* and Δ*ech* eliminated all H_2 formation. When all H_2 was eliminated, ethanol yield doubled in comparison to wild type *C. thermocellum* (Biswas et al., 2015b). Formate production was eliminated in a Δ*pfl* strain as another attempt to redirect electron flux to ethanol. In Δ*pfl,* acetate production was decreased by 50%, but ethanol did not increase, and growth rate decreased by 2.9 fold compared with the wild type (Rydzak et al., 2015). Most recently, in strain AG553 (Δ*hydG*Δ*ldh*Δ*pfl*Δ*pta-ack*), almost all organic acid and hydrogen production has been eliminated through genetic modifications. As a result, strain AG553 showed an ethanol yield of 63% of the theoretical maximum, compared with 20% ethanol yield from wild type. This yield was constant when grown on both model substrates and pretreated biomass (Papanek et al., 2015). However, genetic elimination of many central metabolic pathways resulted in severe growth inhibition. This growth inhibition is problematic in an industrial setting as it slows the rate of ethanol production and therefore limits the volume of fuel that can be made at a biorefinery without increasing capital costs. Because of its slow growth, more work and additional strategies must be employed in order to improve the industrial relevance of engineered *C. thermocellum* strains.

Rational metabolic engineering has been successful in increasing the ethanol yields and titers of *C. thermocellum* DSM1313, but as explained, has still run into limitations in developing a completely optimized strain. Laboratory directed evolution is another strategy that is commonly employed to improve phenotypic qualities of microorganisms. Laboratory strain evolution is a beneficial tool for studying microorganisms. Growing a single culture results in 10-20 generations of a single organism (Herring et al., 2006). Rapid growth and frequent genomic mutations quickly create new genotypes and phenotypes. This is particularly useful in the case of complex phenotypes such as growth rate or inhibitor tolerance where there may not be enough information for targeted metabolic engineering (Dougherty and Arnold, 2009). In laboratory evolution, organisms are introduced to growth conditions under which certain mutant cells have a growth advantage (Portnoy et al., 2011). The purpose of this process is to select for the cells that are most capable of thriving under the conditions tested. By repeatedly transferring a small amount of cells to fresh media, the most fit culture is selected. After evolution, the culture is a mixture of genotypes in which some cells acquired beneficial genetic mutations while others did not. In order to correlate mutations with specific genotypes, pure cultures of the individual genotypes are necessary. Clean genetic backgrounds can be isolated from evolved cultures by plating the cultures and selecting single colonies. After selecting isolated colonies, advanced sequencing technology can rapidly determine the genotype and identify acquired mutations (Hall, 2007). Some examples of successfully using laboratory strain evolution include increasing ethanol tolerance in *C. thermocellum* (Shao et al., 2011; Shaw et al., 2015; Williams et al., 2007), evolving the citramalate synthase (CimA) from *Methananococcus jannashcii* for enhanced activity over a wider temperature range (Atsumi and Liao, 2008), and evolving *E. coli* for isobutanol tolerance and doing the genomic analysis to understand the evolution strategy (Atsumi et al., 2010). Both of these studies address improving a complicated phenotype by using laboratory evolution. Utilizing single colony isolation and resequencing is also an established protocol and has been used in *T. saccharolyticum* to identify a mutation in the *hfs* hydrogenase that correlates with ethanol production (Shaw et al., 2015). All of these studies are examples of utilizing laboratory evolution in place of targeted metabolic engineering for improving desired phenotypes.

In strain AG553, the biggest phenotypic drawback is the growth rate of the organism. Identifying the specific reason for slow growth in a microorganism is challenging, however. Metabolic bottlenecks can be caused by a variety of pathways and cofactors making growth rate optimization a challenging target for metabolic engineering Low growth rate is an ideal target for an evolution-based strain improvement strategy, which was employed with strain AG553 in Chapter 2. Although no additional external stress was applied for the purpose of its evolution, the culture was passaged daily, and the fastest growing cells were

enriched. Gradually the culture overcame the slow metabolism observed in the unevolved strain.

While both rational engineering and laboratory evolution improve ethanol yields and growth rate in *C. thermocellum* the strain is still not competitive on an industrial scale. Unfortunately, even after extensive research efforts, its ethanol yield remains too low compared to first generation ethanol production. Typical ethanol yields from yeast are 40-60 g/L (Siqueira et al., 2008), and CBP processes must be able to compete with these titers to have industrial relevance. The low yields are in part due to carbon still being diverted to a substantial amount of excreted amino acids, most notably valine and alanine, and organic acids such as pyruvate (Holwerda et al., 2014) that have not yet been targeted in the metabolic engineering efforts. These products remain as fermentation products of *C. thermocellum* due to the inefficiencies and slow rate of genetic engineering.

Genetic tools in *C. thermocellum*

A typical gene knockout in *C. thermocellum* DSM1313 takes 3-4 weeks if all of the steps are carried out without issue. This process becomes increasingly difficult as gene deletions are stacked and the phenotype of the organism is altered. Low transformation efficiency also prevents the use of other genetic engineering techniques such as integration of non-replicating plasmids and heterologous expression vectors. After encountering many of these problems while working to optimize *C. thermocellum,* we realized that better genetic tools would improve the efficiency and speed of engineering *C. thermocellum*.

Some genetic tools have been developed previously for *C. thermocellum* DSM1313. For instance, plasmid origins of replication are needed for stable DNA maintenance in *C. thermocellum*. The pNW33N origin of replication was first used to demonstrate transformation in *C. thermocellum* and can replicate at the lower end of *C. thermocellum*'s temperature range. Additionally, computational modeling has been used to predict less stable origins that might be useful as temperature sensitive origins in *C. thermocellum* (Olson and Lynd, 2012a). Genes for positive and negative selection are also available for *C. thermocellum* DSM1313. The *pyrF* gene is involved in uracil biosynthesis, and the Δ*pyrF* strain was used for plasmid selection containing the *pyrF* gene (Tripathi, et al., 2010). *C. thermocellum* was also discovered to be sensitive to 8-azahypoxanthine (AZH), allowing *hpt* to act as a selecteable marker with AZH as counter selection (Argyros et al., 2011). This same work showed that *cat* conferred thiamphenicol resistance in *C. thermocellum* and is used as an antibiotic resistance gene. While these plasmid components allow for engineering in *C. thermocellum* DSM1313, few details are understood about their mechanisms. Additionally, the pBAS2 origin of replication from *Caldicellulosiruptor* was recently developed for use at *C.*

thermocellum's optimal growth temperature of 60°C (Groom et al., 2016). Acknowledging the slow development of genetic tools for the genus *Clostridia* also lead to an intron-based gene knock-out system (Heap et al., 2007). ClosTron is functional across species of *Clostridia*, and has been applied to *C. thermocellum* (Mohr et al., 2013) but the plasmid used to do the gene knockouts remains in the genome and makes subsequent genetic mutations difficult. The limitations in the existing genetic tool system for *C. thermocellum* leave room for significant improvements in addressing these challenges.

Successful introduction of foreign DNA has been achieved through electroporation protocols (Olson and Lynd, 2012b). Active Restriction-Methylation (RM) systems in *C. thermocellum* DSM1313 have been identified, and at least two known systems have been circumvented. The study by Guss, et. al (2012) definitively showed that the presence of *E. coli*'s Dam methylase substantially improved transformation efficiency while the Dcm methylase substantially inhibited transformation efficiency. This work showed improved transformation efficiencies, but final efficiencies were still not as high as transforming model organisms such as *E. coli*. This led to the hypothesis that a functioning RM system was still present in *C. thermocellum* DSM1313 limiting the transformation efficiency, which is further addressed in Chapter 3.

Expanding genetic tools in organisms beyond *C. thermocellum*

The metabolic engineering barriers described above are not only a problem when working with *C. thermocellum.* Despite a well-documented desire among researchers for a broader selection of genetically tractable microorganisms (Alper and Stephanopoulos, 2009; Keasling, 2010; Peralta-Yahya et al., 2012), there have been very few rational, systematic approaches to improving the genetic tractability of microorganisms as a whole. Most efforts have been specific to a single organism (Chung et al., 2012; Kita et al., 2013; Koksharova and Wolk, 2002) or for specific chemical production (Koller et al., 2010; Papoutsakis, 2008; Zheng et al., 2009) with no high-throughput methods that address the fundamental problems of genetically engineering new organisms.

The methods that have been applied to *C. thermocellum* DSM1313 are applicable to a broader set of interests than just Consolidated Bioprocessing (CBP) and ethanol production. In the last half century, microbial processes produced beneficial antibiotics, herbicides and insecticides, anticancer pharmaceuticals, as well as commodity chemicals (Rondon et al., 1999). The same principles that are being applied in *C. thermocellum* DSM1313 can be utilized in organisms that have not yet been successfully transformed, and that successfully addressing genetic engineering in a range of non-model

microorganisms will open the door to greater production of microbial commodities and chemicals. Biological production methods are more sustainable, environmentally benign alternatives to traditional fossil fuel manufacturing. This transition to microbial production will allow for the decoupling of economic growth from environmental destruction (Sheppard, 2011). In the conclusion, preliminary advances are discussed toward laying down a foundation for the rational engineering of new organisms.

C. thermocellum remains a promising organism for Consolidated Bioprocessing as it offers a more economical, sustainable approach to producing ethanol. Successful efforts to improve the industrial characteristics of the organism have been made, but work remains to be done. In order to make microbial cells efficient for the sustainable production of commodities and chemicals, metabolic engineering must be improved. This holds true for *C. thermocellum* as well as most other non-model microorganisms. If this improvement is successful, dependence on fossil fuels can be reduced while maintaining quality of life and economic development.

References

Alper, H., Stephanopoulos, G., 2009. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat Rev Microbiol. 7**,** 715-23.

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., Caiazza, N. C., 2011. High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microb. 77**,** 8288-8294.

Atsumi, S., Liao, J. C., 2008. Directed Evolution of Methanococcus jannaschii Citramalate Synthase for Biosynthesis of 1-Propanol and 1-Butanol by Escherichia coli. Appl Environ Microb. 74**,** 7802-7808.

Atsumi, S., Wu, T. Y., Machado, I. M. P., Huang, W. C., Chen, P. Y., Pellegrini, M., Liao, J. C., 2010. Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. Molecular Systems Biology. 6.

Balan, V., Chiaramonti, D., Kumar, S., 2013. Review of US and EU initiatives toward development, demonstration, and commercialization of lignocellulosic biofuels. Biofuels, Bioproducts and Biorefining. 7**,** 732-759.

Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., Guss, A. M., 2015. Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in Clostridium thermocellum. Biotechnology for biofuels. 8**,** 20.

Bridgwater, A. V., 1995. The technical and economic feasibility of biomass gasification for power generation. Fuel. 74**,** 631-653.

Broecker, W. S., Takahashi, T., Simpson, H. J., Peng, T.-H., 1979. Fate of Fossil Fuel Carbon Dioxide and the Global Carbon Budget. Science. 206**,** 409-418.

Chung, D., Farkas, J., Huddleston, J. R., Olivar, E., Westpheling, J., 2012. Methylation by a unique alpha-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. PloS one. 7**,** e43844.

de Vries, S. C., van de Ven, G. W. J., van Ittersum, M. K., Giller, K. E., 2010. Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. Biomass and Bioenergy. 34**,** 588-601.

Demain, A. L., Newcomb, M., Wu, J. H. D., 2005. Cellulase, clostridia, and ethanol. Microbiology and Molecular Biology Reviews. 69**,** 124-154.

Dougherty, M. J., Arnold, F. H., 2009. Directed evolution: new parts and optimized function. Curr Opin Biotechnol. 20**,** 486-91.

Ellis, L. D., Holwerda, E. K., Hogsett, D., Rogers, S., Shao, X. J., Tschaplinski, T., Thorne, P., Lynd, L. R., 2012. Closing the carbon balance for fermentation by Clostridium thermocellum (ATCC 27405). Bioresource technology. 103**,** 293-299.

Gold, N. D., Martin, V. J., 2007. Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis. J Bacteriol. 189, 6787-95.
Groom, J., Chung, D., Olson, D. G., Lynd, L. R., Guss, A. M., Westpheling, J., 2016. Promiscuous plasmid replication in thermophiles: Use of a novel hyperthermophilic replicon for genetic manipulation of Clostridium thermocellum at its optimum growth temperature. Metabolic Engineering Communications. 3, 30-38.
Guss, A. M., Olson, D. G., Caiazza, N. C., Lynd, L. R., 2012. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. Biotechnology for biofuels. 5, 30.
Gustavsson, L., Börjesson, P., Johansson, B., Svenningsson, P., 1995. Reducing CO2 emissions by substituting biomass for fossil fuels. Energy. 20, 1097-1113.
Hall, N., 2007. Advanced sequencing technologies and their wider impact in microbiology. Journal of Experimental Biology. 210, 1518-1525.
Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., Minton, N. P., 2007. The ClosTron: A universal gene knock-out system for the genus Clostridium. Journal of Microbiological Methods. 70, 452-464.
Herring, C. D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M. K., Joyce, A. R., Albert, T. J., Blattner, F. R., van den Boom, D., Cantor, C. R., Palsson, B. O., 2006. Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. Nat Genet. 38, 1406-12.
Hogsett, D. A., Ahn, H.-J., Bernardez, T. D., South, R., Lynd, L. R., 1992. Direct microbial conversion. Appl Biochem Biotech. 34, 527-541.
Holwerda, E. K., Thorne, P. G., Olson, D. G., Amador-Noguez, D., Engle, N. L., Tschaplinski, T. J., van Dijken, J. P., Lynd, L. R., 2014. The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading. Biotechnology for biofuels. 7, 155.
Keasling, J. D., 2010. Manufacturing molecules through metabolic engineering. Science. 330, 1355-8.
Kita, A., Iwasaki, Y., Sakai, S., Okuto, S., Takaoka, K., Suzuki, T., Yano, S., Sawayama, S., Tajima, T., Kato, J., Nishio, N., Murakami, K., Nakashimada, Y., 2013. Development of genetic transformation and heterologous expression system in carboxydotrophic thermophilic acetogen Moorella thermoacetica. J Biosci Bioeng. 115, 347-52.
Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., Blanch, H. W., 2012. The challenge of enzyme cost in the production of lignocellulosic biofuels. Biotechnology and bioengineering. 109, 1083-7.
Koksharova, O. A., Wolk, C. P., 2002. Genetic tools for cyanobacteria. Appl Microbiol Biotechnol. 58, 123-37.
Koller, M., Atlic, A., Dias, M., Reiterer, A., Braunegg, G., 2010. Microbial PHA Production from Waste Raw Materials. Microbiol Monogr. 14, 85-119.

Kumar, P., Barrett, D. M., Delwiche, M. J., Stroeve, P., 2009. Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production. Industrial & Engineering Chemistry Research. 48, 3713-3729.

Lynd, L. R., van Zyl, W. H., McBride, J. E., Laser, M., 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current opinion in biotechnology. 16, 577-83.

Lynd, L. R., Weimer, P. J., van Zyl, W. H., Pretorius, I. S., 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and molecular biology reviews. 66, 506-77.

Mohan, D., Pittman, C. U., Steele, P. H., 2006. Pyrolysis of Wood/Biomass for Bio-oil: A Critical Review. Energ Fuel. 20, 848-889.

Mohr, G., Hong, W., Zhang, J., Cui, G. Z., Yang, Y., Cui, Q., Liu, Y. J., Lambowitz, A. M., 2013. A targetron system for gene targeting in thermophiles and its application in Clostridium thermocellum. PloS one. 8, e69032.

Olson, D. G., Lynd, L. R., 2012a. Computational design and characterization of a temperature-sensitive plasmid replicon for gram positive thermophiles. Journal of Biological Engineering. 6, 1-10.

Olson, D. G., Lynd, L. R., 2012b. Transformation of *Clostridium thermocellum* by electroporation. Method Enzymol. 510, 317-30.

Olson, D. G., McBride, J. E., Shaw, A. J., Lynd, L. R., 2012. Recent progress in consolidated bioprocessing. Curr Opin Biotechnol. 23, 396-405.

Papanek, B., Biswas, R., Rydzak, T., Guss, A. M., 2015. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in Clostridium thermocellum. Metab Eng. 32, 49-54.

Papoutsakis, E. T., 2008. Engineering solventogenic clostridia. Curr Opin Biotechnol. 19, 420-9.

Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., Keasling, J. D., 2012. Microbial engineering for the production of advanced biofuels. Nature. 488, 320-8.

Petrou, E. C., Pappis, C. P., 2009. Biofuels: A Survey on Pros and Cons. Energ Fuel. 23, 1055-1066.

Plessis, L., Rose, S. H., Zyl, W. H., 2009. Exploring improved endoglucanase expression in Saccharomyces cerevisiae strains. Applied Microbiology and Biotechnology. 86, 1503-1511.

Portnoy, V. A., Bezdan, D., Zengler, K., 2011. Adaptive laboratory evolution — harnessing the power of biology for metabolic engineering. Current Opinion in Biotechnology. 22, 590-594.

Rondon, M. R., Goodman, R. M., Handelsman, J., 1999. The Earth's bounty: assessing and accessing soil microbial diversity. Trends Biotechnol. 17, 403-9.

Rydzak, T., Lynd, L. R., Guss, A. M., 2015. Elimination of formate production in Clostridium thermocellum. J Ind Microbiol Biotechnol. 42, 1263-72.

Shao, X., Raman, B., Zhu, M., Mielenz, J. R., Brown, S. D., Guss, A. M., Lynd, L. R., 2011. Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of *Clostridium thermocellum*. Applied Microbiology and Biotechnology. 92**,** 641-52.

Shaw, A. J., Miller, B. B., Rogers, S. R., Kenealy, W. R., Meola, A., Bhandiwad, A., Sillers, W. R., Shikhare, I., Hogsett, D. A., Herring, C. D., 2015. Anaerobic detoxification of acetic acid in a thermophilic ethanologen. Biotechnology for biofuels. 8**,** 1-12.

Sheppard, A., Gillespie, I, Hirsch M, Begley, C, 2011. Biosecurity and sustainability within the growing global bioeconomy. Current Opinion in Environmental Sustainability. 3**,** 4-10.

Siqueira, P. F., Karp, S. G., Carvalho, J. C., Sturm, W., Rodríguez-León, J. A., Tholozan, J.-L., Singhania, R. R., Pandey, A., Soccol, C. R., 2008. Production of bio-ethanol from soybean molasses by Saccharomyces cerevisiae at laboratory, pilot and industrial scales. Bioresource technology. 99**,** 8156-8163.

Solomon, S., Plattner, G.-K., Knutti, R., Friedlingstein, P., 2009. Irreversible climate change due to carbon dioxide emissions. Proceedings of the National Academy of Sciences. 106**,** 1704-1709.

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., 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 Microb. 76**,** 6591-9.

Valentine, J., Clifton-Brown, J., Hastings, A., Robson, P., Allison, G., Smith, P., 2012. Food vs. fuel: the use of land for lignocellulosic 'next generation' energy crops that minimize competition with primary food production. GCB Bioenergy. 4**,** 1-19.

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., Lynd, L. R., 2013. Characterization of *Clostridium thermocellum* strains with disrupted fermentation end-product pathways. J Ind Microbiol Biot. 40**,** 725-34.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., Melillo, J. M., 1997. Human Domination of Earth's Ecosystems. Science. 277**,** 494-499.

Yanase, S., Yamada, R., Kaneko, S., Noda, H., Hasunuma, T., Tanaka, T., Ogino, C., Fukuda, H., Kondo, A., 2010. Ethanol production from cellulosic materials using cellulase-expressing yeast. Biotechnology Journal. 5**,** 449-455.

Zheng, Y. N., Li, L. Z., Xian, M., Ma, Y. J., Yang, J. M., Xu, X., He, D. Z., 2009. Problems with the microbial production of butanol. J Ind Microbiol Biotechnol. 36**,** 1127-38.

CHAPTER I
ELIMINATION OF METABOLIC PATHWAYS TO ALL TRADITIONAL FERMENTATION PRODUCTS INCREASES ETHANOL YIELDS IN *CLOSTRIDIUM THERMOCELLUM*

A version of this chapter was originally published by Beth Papanek, Ranjita Biswas, Thomas Rydzak and Adam M Guss:

Beth Papanek, Ranjita Biswas, Thomas Rydzak, Adam Guss. “Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*.” *Metabolic Engineering* 32 (2015): 49-54.

Strain construction was done by Ranjita Biswas, Beth Papanek did all of the experimental strain characterization, Thomas Rydzak did the amino acid measurements and analysis. The manuscript was written and revised primarily by Beth Papanek and Adam Guss.

Abstract

Clostridium thermocellum has the natural ability to convert cellulose to ethanol, making it a promising candidate for consolidated bioprocessing (CBP) of cellulosic biomass to biofuels. To further improve its CBP capabilities, a mutant strain of *C. thermocellum* was constructed (strain AG553; *C. thermocellum* Δ*hpt* Δ*hydG* Δ*ldh* Δ*pfl* Δ*pta-ack*) to increase flux to ethanol by removing side product formation. Strain AG553 showed a two- to threefold increase in ethanol yield relative to the wild type on all substrates tested. On defined medium, strain AG553 exceeded 70% of theoretical ethanol yield on lower loadings of the model crystalline cellulose Avicel, effectively eliminating formate, acetate, and lactate production and reducing H_2 production by fivefold. On 5 g/L Avicel, strain AG553 reached an ethanol yield of 63.5% of the theoretical maximum compared with 19.9% by the wild type, and it showed similar yields on pretreated switchgrass and poplar. The elimination of organic acid production suggested that the strain might be capable of growth under higher substrate loadings in the absence of pH control. Final ethanol titer peaked at 73.4 mM in mutant AG553 on 20 g/L Avicel, at which point the pH decreased to a level that does not allow growth of *C. thermocellum,* likely due to CO2 accumulation. In comparison, the maximum titer of wild type *C. thermocellum* was 14.1 mM ethanol on 10 g/L Avicel. With the elimination of the metabolic pathways to all traditional fermentation products other than ethanol, AG553 is the best ethanol-yielding CBP strain to date and will serve as a platform strain for further metabolic engineering for the bioconversion of lignocellulosic biomass.

Introduction

There is a strong push in the biofuel industry to move away from traditional starch-based feedstocks that are currently used to produce the majority of the United States’ ethanol to more sustainable lignocellulosic crops (Klein-Marcuschamer and Blanch, 2015). This transition could reduce the biofuel industry’s dependence on food crops as well as reduce the environmental strain of growing the necessary crops (Brehmer et al., 2008). One enormous challenge in using lignocellulosic feedstocks is the cost-effective deconstruction of

recalcitrant plant cell walls to liberate the fermentable sugars (Himmel et al., 2007). Consolidated bioprocessing (CBP) is an approach to overcome this obstacle by using one or more organisms to degrade plant biomass in a single fermentation tank without the addition of commercial enzymes. CBP allows for the simultaneous hydrolysis of cellulose into soluble, metabolizable sugars, and fermentation of the resulting sugars to the product of interest (Lynd et al., 2005). *Clostridium thermocellum*, a thermophilic, anaerobic, cellulolytic bacterium, is one of the most promising microbial candidates for this process. The presence of a large membrane-associated enzyme complex called the cellulosome gives it the ability to hydrolyze lignocellulosic biomass directly to fermentable sugars (Lamed and Bayer, 1988). It is then able to convert the sugars to ethanol and other products such as lactate, acetate, formate, hydrogen, and excreted amino acids such as valine and alanine (Ellis et al., 2012a). Further, there is a proven and usable set of tools for genetically engineering the organism (Argyros et al., 2011; Guss et al., 2012; Olson and Lynd, 2012b; Tripathi et al., 2010), allowing for the rational manipulation of the *C. thermocellum* genome for improved phenotypes.

In addition to producing ethanol, *Clostridium thermocellum* naturally produces a range of organic acids and hydrogen. The generation of lactate, formate and acetate diverts carbon and electron flux away from ethanol and can quickly make fermentation conditions toxic, presumably due to pH, arresting growth and fermentation. To maximize ethanol yield from *C. thermocellum,* competing pathways must be eliminated. Previously, strains have been engineered to strategically eliminate one or two of these side products, including strains deficient in the production of (i) acetate via deletion of phosphotransacetylase (*pta*) (Tripathi et al., 2010), (ii) lactate via deletion of lactate dehydrogenase (*ldh*) and acetate via *pta* (Argyros et al., 2011), (iii) H_2 via deletion of hydrogenase maturase *hydG* to inactivate all three [FeFe] hydrogenases and *ech* to eliminate the [NiFe] Energy Converting Hydrogenase (Biswas et al., 2015a), and (iv) formate via deletion of pyruvate-formate lyase (*pfl*) (Rydzak et al., 2015a). Other mutant strains aimed at increasing ethanol production include deletion of lactate dehydrogenase in an ethanol tolerant strain of *C. thermocellum* (Biswas et al., 2014) and deletion of malate dehydrogenase in a strain of *C. thermocellum* that heterologously expressed a pyruvate kinase gene (Deng et al., 2013). However, these strains still produced organic acids and/or H_2 that can acidify the medium and divert carbon and electron flux away from ethanol, reducing ethanol titers and yields. We hypothesized that by combining gene deletions associated with product formation (Δ*hydG,* Δ*pfl,* Δ*pta-ack,* and Δ*ldh*) into a single strain*,* side product formation would be effectively eliminated, allowing greater carbon flux through the ethanol production pathway. Therefore, we stacked these mutations into a single strain and characterized the effect on both model substrates and pretreated plant biomass.

Materials and Methods

Growth Media

Escherichia coli strains were grown on LB medium supplemented with 12 μg ml^{-1} chloramphenicol as needed. For *C. thermocellum,* strains were grown either in modified DSM122 rich medium (Tripathi et al., 2010) or defined Medium for Thermophilic Clostridia (MTC) (Linville et al., 2013), with cellobiose, crystalline cellulose (Avicel), or pretreated biomass as the carbon source. To make MTC, Solution A was made in 162 mL serum bottles with cellobiose, Avicel PH105 or dilute-acid pretreated biomass as the carbon source. Dilute acid pretreatment was done previous to experimentation as explained in (Schell et al., 2003). Biomass was either poplar or switchgrass previously analyzed with quantitative saccharification from the batch described in (Wilson et al., 2013). The poplar sample was measured to have 646 mg glucose equivalents/g biomass ± 13.6 (other sugar composition included: 19.4 mg xylose equivalents/g biomass ± 1.6 and 1.6 mg arabinose equivalents/g biomass ± 0.2). The switchgrass sample was measured to have 522.5 mg glucose equivalents/g biomass ± 9.3 (other sugar composition included: 72.5 mg xylose equivalents/g biomass ± 0.4 and 7.1 mg arabinose equivalents/g biomass ± 1.0). After autoclave sterilization, serum bottles were degassed with filtered N_2 prior to addition of other components. Solutions B, C, E, and F for MTC were filter sterilized through a 0.22 μm filter and solutions D and M were autoclave sterilized. Sterile solutions were mixed in a sterile beaker and 5 mL of the mixed cocktail solution was added to each bottle containing Avicel as the carbon source. For bottles containing cellobiose or biomass as the carbon source, solution M was added to the cocktail mixture and 10 mL of mixed solution was added to each fermentation bottle. For both conditions, the final volume in the serum bottles was 50 mL. The bottles were then degassed with sterile-filtered N_2. When appropriate, *C. thermocellum* rich medium was supplemented with 12 μg ml^{-1} thiamphenicol (Tm; Sigma-Aldrich, Saint Louis, MO, USA), 50 μg ml^{-1} 5-fluoro-2'-deoxyuridine (FUdR; Sigma-Aldrich), or 500 μg ml^{-1} 8-azahypoxanthine (8AZH; Tokyo Chemical Industry, Co., Tokyo, Japan) during strain construction. Plates were solidified with 1.5% agar when appropriate.

Plasmid and Strain Construction

Plasmid pAMG498 (annotated, Genbank style sequence file, Supplemental File 1) for deletion of *C. thermocellum pta* and *ack* was constructed using yeast gap repair cloning in *Saccharomyces cerevisiae* InvSc1. Standard methods were used to construct *C. thermocellum* deletions (Olson and Lynd, 2012b). Briefly, plasmid DNA was isolated from a *dcm- E. coli* strain (Guss et al., 2012) and electroporated into *C. thermocellum* strains derived from strain DSM1313 with *hpt* deleted. Electroporated cells were plated on rich medium agar supplemented with Tm. Colonies were picked into liquid rich medium supplemented with Tm,

followed by plating dilutions in rich medium supplemented with Tm and FUdR. Colonies were re-streaked on rich medium agar plates supplemented with Tm and FUdR to ensure purity, picked into liquid rich medium supplemented with Tm, re-grown in the absence of Tm, and then plated in rich medium supplemented with 8AZH. Colonies were single colony purified, picked into liquid rich medium, and deletions were confirmed by PCR. Lactate dehydrogenase (Clo1313_1160; *ldh*) was deleted in *C. thermocellum Δhpt ΔhydG* (Biswas et al., 2015a) using plasmid pMU1777 (Argyros et al., 2011). Pyruvate-formate lyase (Clo1313_1717; *pflB*) and Pfl-activating enzyme (Clo1313_1716; *pflA*) were deleted in *C. thermocellum Δhpt ΔhydG Δldh* using plasmid pAMG281 (Rydzak et al., 2015a). Phosphotransacetylase (Clo1313_1185; *pta*) and acetate kinase (Clo1313_1186; *ack*) were deleted in *C. thermocellum Δhpt ΔhydG Δldh Δpfl* using plasmid pAMG498, resulting in *C. thermocellum* strain AG553 (*C. thermocellum Δhpt ΔhydG Δldh Δpfl Δpta-ack*).

Fermentation

Strains of *C. thermocellum* were revived from -80°C frozen stocks into MTC medium with 5 g/L cellobiose and grown at 51°C in a Coy anaerobic chamber (Coy Laboratory Products, Grass Lake, MI) until exponential phase (O.D. ~0.4), and 1 mL (2% inoculum) was added to each 162 ml serum bottle, resulting in a working volume of 51 mL. Bottles were incubated at 55°C with 100 rpm orbital shaking. Culture supernatants were sampled every 24 hours and fermentation products were measured by High Performance Liquid Chromatography (HPLC) until production ceased and fermentation product profiles stayed constant. All fermentations were performed twice with three biological replicates both times. The uninoculated MTC medium was found to contain substantial amounts of proline and threonine in addition to the added cysteine, so these three amino acids were removed from the total amino acid analysis.

Analytical Methods

Growth curve data were collected at OD_{600} in a BioTek Eon Microplate reader (BioTek Instruments Inc., Winooski, VT) inside a Coy anaerobic chamber at 55°C.

Fermentation products lactate, formate, acetate, and ethanol were measured using a Breeze HPLC system (Waters, Milford, MA) with an Aminex-HPX-87H column (Bio-Rad, Hercules, CA) and 5 mM sulfuric acid as the mobile phase as previously described (Rydzak et al., 2015a). Isobutanol formation was also measured using this method, but none was detected.

H_2 production was measured using an Agilent Technologies 6850 Series II Gas Chromatographer (Agilent Technologies, Santa Clara, CA) with a Carbonex 1010

PLOT (30.0 m x 530 µm I.D.; model Supelco 25467) column as previously described (Yee et al., 2014).

Excreted amino acids were measured using post-column derivatization with ninhydrin using an Aracus Amino Acid Analyzer (membraPure, Berlin, Germany) with a T111 Li-cation exchange column as previously described (Rydzak et al., 2015a).

Results

Construction of C. thermocellum Δhpt ΔhydG Δldh Δpfl Δpta-ack (Strain AG553 mutant

To eliminate H_2 and organic acid fermentation products, *C. thermocellum* strain Δ*hpt* Δ*hydG* was sequentially modified to delete lactate dehydrogenase (*ldh*), pyruvate-formate lyase (*pfl*), and phosphotransacetylase and acetate kinase (*pta-ack*) resulting in strain AG553, which is predicted to be deficient in the production of lactate, formate, and acetate. This strain also harbors a spontaneous point mutation in the bifunctional aldehyde/alcohol dehydrogenase *adhE* from strain *C. thermocellum* Δ*hpt* Δ*hydG* (Biswas et al., 2015a). The resulting D494G amino acid change expanded cofactor specificity of the alcohol dehydrogenase activity from using only NADH in wild type to using both NADH and NADPH in this strain. Efforts to delete the Ech hydrogenase to eliminate the remaining H_2 production (Biswas et al., 2015a) were unsuccessful. Growth was initially tested in minimal medium with cellobiose as the carbon source (Fig. 1-1).

C. thermocellum strain AG553 produces more ethanol on model substrates

Fermentation product formation by *C. thermocellum* AG553 mutant was initially tested on model substrates. When grown in defined medium with 5 g/L of the soluble disaccharide cellobiose as the carbon source, the mutant strain produced greater than two fold more ethanol than the wild type strain with no appreciable amounts of other fermentation products (Fig. 1-2A). The wild type strain, on the other hand, produced over 15 mM acetate with small amounts of lactate and formate as well. Final ethanol titer reached 32.8 mM for AG553 and 17.7 mM for wild type (56.1% and 30.3% theoretical yield), respectively. Wild type produced 4.2 mM total amino acids, while strain AG553 produced 2.2 mM (Fig. 1-2A). In both cases, valine was the most abundant amino acid produced (Fig. 1-3). H_2 production decreased approximately five-fold in strain AG553 relative to the wild type. Total carbon recovery on 5 g/L Cellobiose was 73.5% for wild type and 73.3% for AG553.

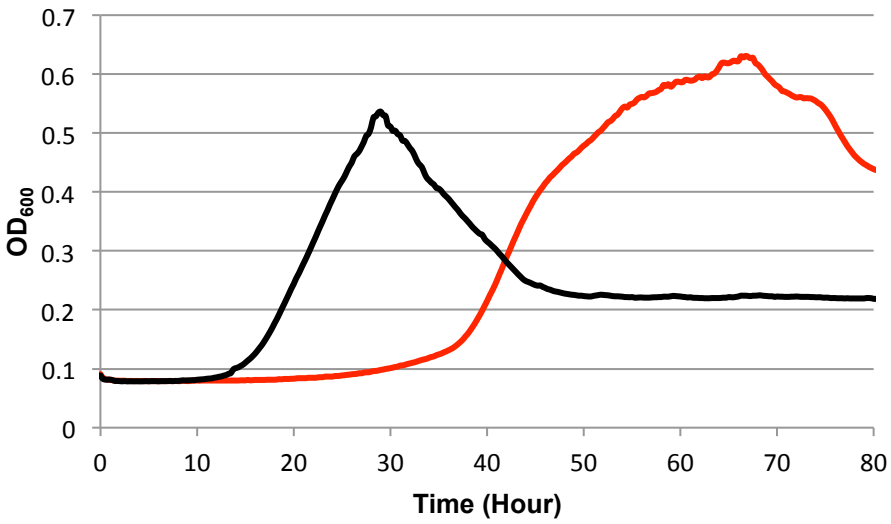

Figure 1-1 Growth curve of *C. thermocellum* mutant in defined media with 5 g/L cellobiose

The growth data was collected on a plate reader with measurements taken every 15 minutes. Lines are the average of triplicate cultures. Black line, wild type; Red line, *C. thermocellum* strain AG553.

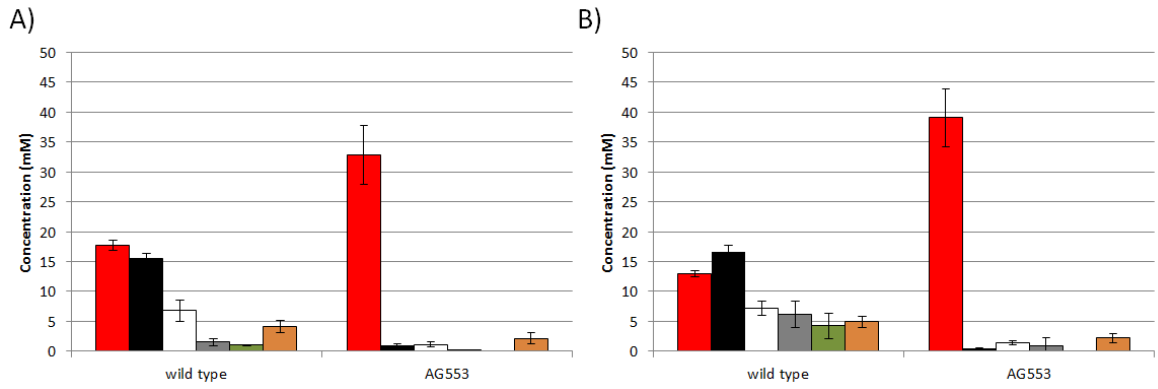

Figure 1-2 *C. thermocellum* mutant product formation on model substrates

Growth on 5 g/L (A) cellobiose and (B) crystalline cellulose Avicel in defined MTC media. Red, Ethanol; Black, Acetate; White, H_2 Grey, Lactate; Green, Formate; Orange, Excreted Amino Acids. H_2 concentration is reported in mmol H_2 in the gas phase per L liquid medium to facilitate comparisons.

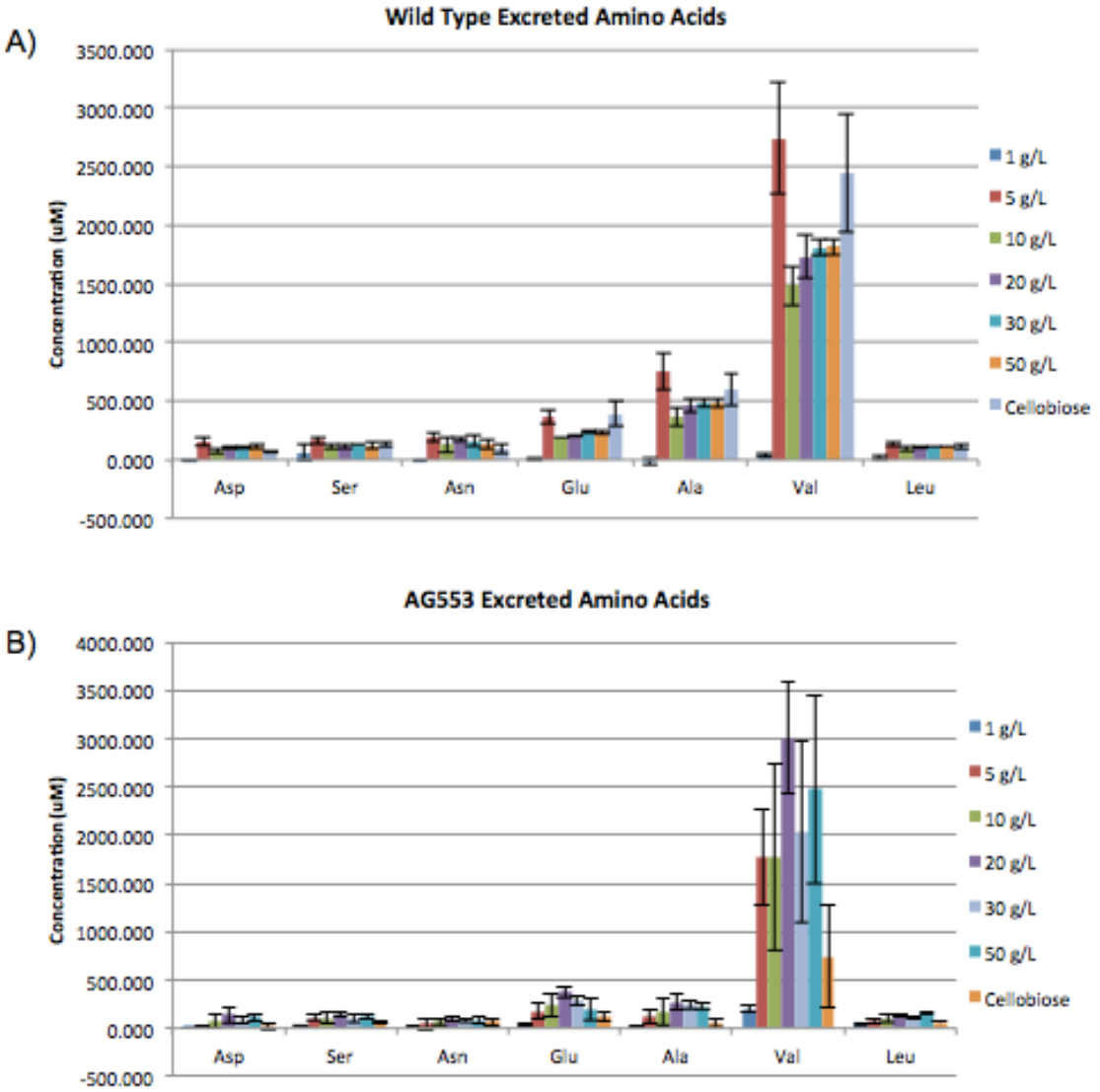

Figure 1-3 Meausred excreted amino acids at the end of fermentation on all loadings of Avicel and Cellobiose

A) End point fermentation measurements for Wild Type amino acids B) End point fermentation measurements for AG553 amino acids

While wild type *C. thermocellum* grew at a rate of 0.14 ± 0.03 hr^{-1}, strain AG553 experienced a substantially longer lag phase and had an initial growth rate of 0.03 ± 0.01 hr^{-1}. After reaching approximately an OD of 0.15, the growth rate increased to 0.13 ± 0.02 hr^{-1}. The observed lag phase was significantly longer for strain AG553. Wild type took 17.5 hours to double its initial OD while AG553 took 39 hours to double its initial OD.

On the model crystalline cellulose, Avicel PH105, mutant AG553 again produced three times more ethanol (39.1 mM, 63.5% theoretical yield) than the parent strain (12.3 mM, 19.9% theoretical yield), while the wild type strain produced roughly equal amounts of acetate and ethanol as its primary fermentation products (Fig. 1-2B). No significant accumulation of organic acids was found in AG553 fermentations. Similar to the cellobiose results, wild type and AG553 produced 4.9 mM and 2.3 mM amino acids, respectively (Fig. 1-2B), predominantly valine (Fig. 1-3). Again, H_2 levels decreased five-fold relative to wild type (Fig. 1-2B). Carbon recovery on 5 g/L Avicel was 80.3% for wild type and 82.3% for AG553.

C. thermocellum strain AG553 produces more ethanol on pretreated biomass

Washed, dilute-acid pretreated poplar and switchgrass were used to test the ability of strain AG553 to convert complex plant biomass to ethanol. The amount of available sugars present in this batch of biomass had been previously analyzed, with the pretreated poplar having 646.8 mg glucan/g of dry biomass and the pretreated switchgrass having 522.5 mg glucan/g of dry biomass (Wilson et al., 2013). Therefore, at 5 g/L loading, 3.2 g/L and 2.6 g/L of glucan were available from the poplar and switchgrass, respectively. Control fermentations containing an equal amount of crystalline cellulose (1032 mg glucan/g Avicel) were performed to allow direct comparison for cellulose bioconversion from the pretreated biomass. Therefore, 3.1 g/L and 2.5 g/L Avicel were used as controls for the poplar and switchgrass fermentations, respectively. Culture samples were then taken every 24 hours for HPLC analysis until product formation on biomass no longer increased. For the poplar experiment, fermentation was complete after 288 hours for wild type on poplar, 72 hours for wild type on Avicel, 288 hours for AG553 on poplar, and 96 hours for AG553 on Avicel. For the switchgrass experiment, fermentation was complete after 288 hours for wild type on switchgrass, 24 hours for wild type on Avicel, 288 hours for AG553 on switchgrass, and 96 hours for AG553 on Avicel. Strain AG553 produced 24.9 mM ethanol on Avicel and 23.8 mM ethanol on the pretreated poplar (Fig. 1-4A), representing 65.5%, and 62.6% respectively, of the theoretical yield of glucan to ethanol. In comparison, wild type *C. thermocellum* produced only 13.0 mM ethanol on Avicel (34.2% theoretical yield) and 13.4 mM on poplar (35.3% theoretical yield). These values represent 72.0% and 52.4% conversion of

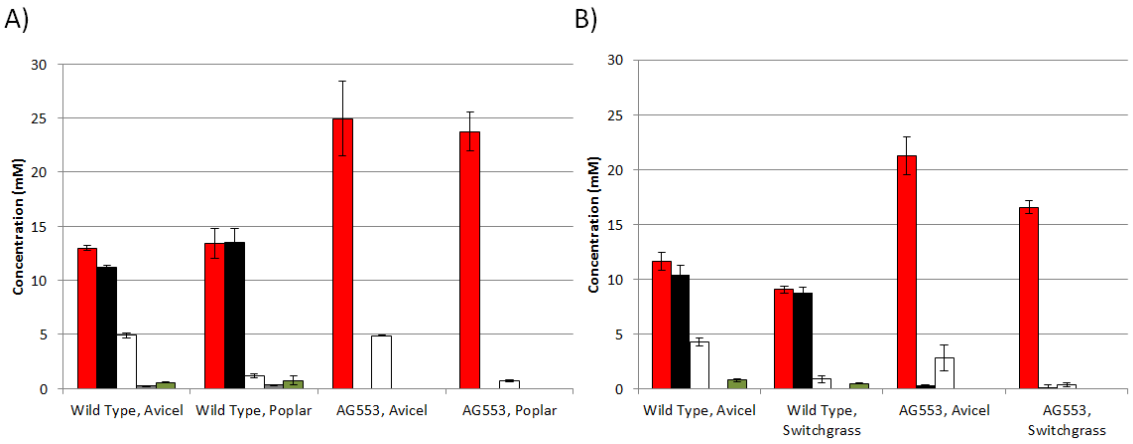

Figure 1-4 *C. thermocellum* mutant product formation on pretreated plant biomass

Avicel samples taken after 96 hours of fermentation and pretreated biomass samples taken after 288 hours of fermentation for both wild type *C. thermocellum* and *C. thermocellum* strain AG553 (A) Fermentation products from 3.2 g/L Avicel and 5 g/L washed, dilute acid pretreated poplar and (B) 2.6 g/L Avicel and 5 g/L washed, dilute acid pretreated switchgrass. Red, Ethanol; Black, Acetate; White, H_2; Grey, Lactate; Green, Formate.

available glucan to ethanol. In comparison, wild type *C. thermocellum* produced 11.7 mM and 9.1 mM ethanol on Avicel and switchgrass, respectively (Fig. 1-4B). This represents 39.2% yield of ethanol from the glucan from Avicel and 26% from switchgrass. Organic acid (lactate, formate, and acetate) production was essentially eliminated in strain AG553 from both poplar and switchgrass. Carbon recovery on poplar was 79.6% for wild type and 71.1% for AG553, while on switchgrass, the carbon recovery was 65.0% and 61.4% for wild type and AG553, respectively.

Increased ethanol titer in C. thermocellum AG553 using higher cellulose loadings

In addition to high yield, high titer ethanol production will also be important for industrial cellulosic biofuel production. We hypothesized that the lack of organic acid production would allow for higher ethanol titers to be achieved. We therefore examined the ability of strain AG553 to convert higher loadings of cellulose to ethanol using Avicel loadings of 1, 5, 10, 20, 30 and 50 g/L in serum bottles with defined medium. Samples were taken every 24 hours until product concentration stopped increasing, which was determined to be 48 hours for wild type C. thermocellum at all loadings and 144 hours for strain AG553 at all loadings except for 1g/L which was 72 hours. Wild type *C. thermocellum* produced approximately equimolar amounts of acetate and ethanol at all loadings (Fig. 1-5A), and fermentation profiles were similar at all loadings between 5 and 50 g/L, likely due to organic acid accumulation and the resulting pH dropping below 6 (Table 1-1). With strain AG553, on the other hand, high yields were attained at low cellulose loadings, with 77.9% and 58.0% of theoretical yield at 1 g/L and 5 g/L, respectively. As substrate loading increased, the overall yield decreased based on the total substrate provided (46.6%, 34.0%, 19.7%, and 11% theoretical yield for 10, 20, 30, and 50 g/L, respectively). The titer, on the other hand, reached a maximum of 73.7 mM at a loading of 20 g/L (Fig. 1-5B), which was not statistically different from the titer at 10, 30, or 50 g/L (P =.0.07, 0.10, and 0.08, respectively). All of these values were far below the ~5 g/L ethanol (108 mM) level at which wild type *C. thermocellum* begins to demonstrate slower growth (Herrero and Gomez, 1980). Interestingly, lactate was produced by strain AG553 at the higher cellulose loadings despite the fact that lactate dehydrogenase has been deleted. Fermentation of sugars to ethanol releases CO_2; therefore, we hypothesized that acidification of the medium from CO_2 accumulation in the serum bottles could also explain the growth and fermentation limitations of strain AG553 at higher cellulose loadings. Therefore, the pH of the fermentation bottles was measured at the end of the experiment, and despite producing far less organic acids, the final pH was substantially lower (Table 1-1) than the optimal pH of 6.8-7.1 for *C. thermocellum* (Garcia-Martinez et al., 1980; Mori, 1990).

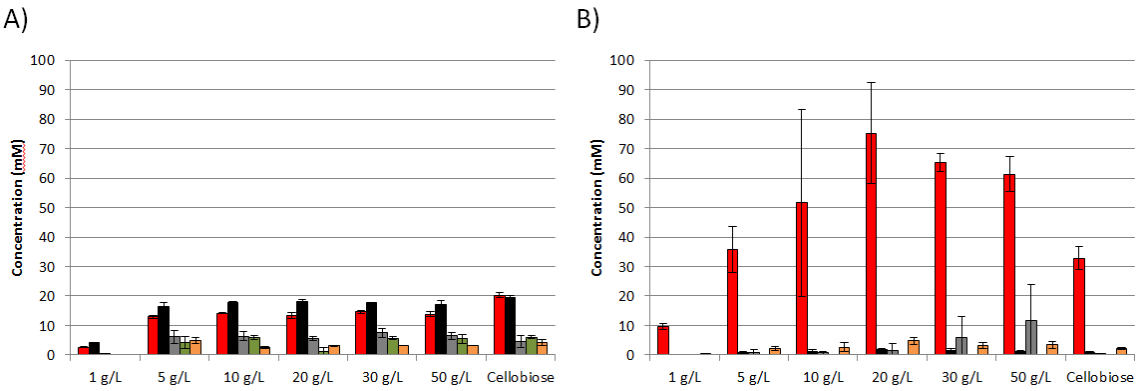

Figure 1-5 *C. thermocellum* mutant product formation on different loadings of Avicel PH105 in defined MTC medium

(A) Wild Type *C. thermocellum* (B) *C. thermocellum* strain AG553. Red, Ethanol; Black, Acetate; Grey, Lactate; Green, Formate; Orange, Excreted Amino Acids.

Table 1-1 Measured pH at the end of fermentation

Substrate Concentration	**Δ*hpt*, final pH**	**AG553, final pH**
0 g/L Avicel	7.99 ± 0.22	7.53 ± 0.02
1 g/L Avicel	7.26 ± 0.19	7.11 ± 0.02
5 g/L Avicel	6.13 ± 0.30	6.57 ± 0.09
10 g/L Avicel	6.07 ± 0.31	5.45 ± 0.15
20 g/L Avicel	6.07 ± 0.28	5.30 ± 0.05
30 g/L Avicel	6.06 ± 0.29	5.60 ± 0.42
50 g/L Avicel	6.11 ± 0.34	5.72 ± 0.19
5 g/L Cellobiose	6.19 ± 0.26	6.37 ± 0.02

Discussion

As no known organism is capable of both efficient lignocellulose deconstruction and industrially relevant yield, rate and titer of biofuel production, metabolic engineering is necessary to enable CBP. Previous studies have focused on one or two genetic modifications to increase flux to ethanol, and the capacity for organic acid production had always remained. By deleting the *ldh, hydG, pfl,* and *pta-ack* genes together, we were able to effectively eliminate lactate, acetate, and formate production and reduce H_2 production on model substrates as well as pretreated poplar and switchgrass. This resulted in a two to three-fold increase in ethanol yields when compared with the wild type strain to a maximum of ca. 70% of theoretical yield and decreased medium acidification. Previously, the highest reported yield was just under 64% theoretical (Biswas et al., 2015a) making strain AG553 capable of the highest ethanol yield reported in *C. thermocellum*. By generating a strain of *C. thermocellum* that does not produce substantial amounts of organic acids from cellulosic substrates, we have created a metabolic engineering platform for future engineering efforts to produce fuels and chemicals from lignocellulosic biomass.

Despite the elimination of the lactate dehydrogenase protein, when the mutant strain was challenged with higher loadings of crystalline cellulose, it exhibited some minor lactate production at later stages of fermentation (Fig. 4B). It has been previously reported that the malate dehydrogenase protein can exhibit promiscuous behavior producing lactate (Li et al., 2012). This could explain the small and sporadic lactate production observed especially under high cell stress fermentation conditions, and could indicate a metabolic bottleneck that still needs to be addressed. Though side products were removed and ethanol yields were improved, they are still values below those necessary for industrial implementation. Approximately the same amount of secreted amino acids was seen in strain AG553 as in the wild type. This is still a significant sink for carbon and electrons that could otherwise have gone to ethanol production. Future work will need to address excess amino acid production in *C. thermocellum*. Glycolysis in *C. thermocellum* converts glucose to pyruvate and generates two NAD(P)H. Conversion of pyruvate to acetyl-CoA generates two pairs of electrons in the form of reduced ferredoxin (Fd_{red}). The two pairs of electrons on NAD(P)H can be used to reduce one acetyl-CoA to ethanol via AdhE, which can only use NAD(P)H as an electron donor. In this scenario, only 50% of theoretical yield could be achieved. In this study, strain AG553 reached yields greater than 50% of theoretical, implying that *C. thermocellum* transfers electrons from Fd_{red} to NAD(P)H that can be used for reduction of acetyl-CoA to ethanol. While electron flux pathways in *C. thermocellum* are not fully elucidated, electron transfer from Fd_{red} to NAD(P)H can be catalyzed by Rnf (Clo1313_0061-0066; $Fd_{red} + NAD^+ + H^+_{in} \rightarrow Fd_{ox} + NADH + H^+_{out}$) or NfnAB (Clo1313_1848-1849; $Fd_{red} + NADH + 2\ NADP^+ + H^+_{in} \rightarrow Fd_{ox} + NAD^+ + 2\ NADPH$).This additional NADH can be generated by the conversion of reduced ferredoxin (Fd_{red}) by Rnf.

C. thermocellum strain AG553 produced approximately four-fold less H_2 due to the deletion of the gene encoding the FeFe hydrogenase maturase *hydG*. Previous work combining the *hydG* and the genes encoding the NiFe hydrogenase *ech* resulted in complete elimination of H_2 as a fermentation product (Biswas et al., 2015a). Ideally, strain AG553 would be further modified to remove *ech*, such that no H_2 is produced. However, we were not successful in creating this mutation, suggesting the possibility that it is essential under the conditions tested. This might not be surprising, as production of cell biomass results in the production of excess reducing equivalents, and production of a more reduced compound is needed in order to prevent a redox imbalance (Fuhrer and Sauer, 2009). In yeast, for example, this is remedied via glycerol production (Ansell et al., 1997); in *C. thermocellum*, H_2 production likely fulfills this role, especially in strain AG553, which lacks Pfl. Alternate redox-balancing strategies will likely be needed in the future to allow *ech* to be deleted.

For industrial ethanol production, it will be important to use inexpensive, renewable feedstocks. Model substrates such as cellobiose and crystalline Avicel are beneficial for the rapid testing of phenotypes in a research setting, but it is important to transition to complex plant biomass feedstocks such as poplar and switchgrass. In this work, strain AG553 had similar ethanol yields on biomass and model substrates such as Avicel, suggesting that the strain is not inhibited by the complexity of real-world biomass feedstocks. This makes the strain an excellent candidate for both continued research on lignocellulosic biofuel production as well as improving the lignocellulosic biofuel production process. Thus, *C. thermocellum* strain AG553 can serve as a platform strain for further organism and process optimization in the future

Conclusion

By combining deletions (Δ*hydG,* Δ*pfl,* Δ*pta-ack,* and Δ*ldh*) in *C. thermocellum*, we have achieved the highest ethanol yields reported to date on both model substrates and pretreated biomass with *C. thermocellum* strain AG553. The elimination of pathways for organic acid production increased the final ethanol titer achieved at higher cellulose loadings, but further work will be needed to improve ethanol yield and titer when challenged with higher amounts of substrate. Overall, *C. thermocellum* strain AG553 represents a new platform strain for future genetic engineering and process optimization for consolidated bioprocessing of lignocellulose to fuels and chemicals.

Abbreviations

8AZH, 8-azahypoxanthine; CBP, consolidated bioprocessing; FUdR, 5-fluoro-2'-deoxyuridine; Tm, thiamphenicol

References

Ansell, R., Granath, K., Hohmann, S., Thevelein, J. M., Adler, L., 1997. The two isoenzymes for yeast NAD(+)-dependent glycerol 3-phosphate dehydrogenase encoded by GPD1 and GPD2 have distinct roles in osmoadaptation and redox regulation. Embo J. 16**,** 2179-2187.

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., Caiazza, N. C., 2011. High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microb. 77**,** 8288-8294.

Biswas, R., Prabhu, S., Lynd, L. R., Guss, A. M., 2014. Increase in Ethanol Yield via Elimination of Lactate Production in an Ethanol-Tolerant Mutant of Clostridium thermocellum. PloS one. 9**,** e86389.

Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., Guss, A. M., 2015. Elimination of hydrogenase active site a ssembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum*. Biotechnol Biofuels. 8.

Brehmer, B., Bals, B., Sanders, J., Dale, B., 2008. Improving the corn-ethanol industry: studying protein separation techniques to obtain higher value-added product options for distillers grains. Biotechnology and bioengineering. 101**,** 49-61.

Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Joe Shaw, A., Lynd, L. R., 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. Metabolic engineering. 15**,** 151-8.

Ellis, L. D., Holwerda, E. K., Hogsett, D., Rogers, S., Shao, X., Tschaplinski, T., Thorne, P., Lynd, L. R., 2012. Closing the carbon balance for fermentation by *Clostridium thermocellum* (ATCC 27405). Bioresource technology. 103**,** 293-9.

Fuhrer, T., Sauer, U., 2009. Different biochemical mechanisms ensure network-wide balancing of reducing equivalents in microbial metabolism. J Bacteriol. 191**,** 2112-21.

Garcia-Martinez, D. V., Shinmyo, A., Madia, A., Demain, A. L., 1980. Studies on Cellulase Production by *Clostridium thermocellum*. Eur J Appl Microbiol. 9**,** 189-197.

Guss, A. M., Olson, D. G., Caiazza, N. C., Lynd, L. R., 2012. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. Biotechnology for biofuels. 5**,** 30.

Herrero, A. A., Gomez, R. F., 1980. Development of ethanol tolerance in *Clostridium thermocellum*: effect of growth temperature. Appl Environ Microb. 40**,** 571-7.

Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., Foust, T. D., 2007. Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science. 315**,** 804-7.

Klein-Marcuschamer, D., Blanch, H. W., 2015. Renewable fuels from biomass: Technical hurdles and economic assessment of biological routes. AIChE Journal. 61**,** 2689-2701.
Lamed, R., Bayer, E. A., 1988. The Cellulosome of *Clostridium thermocellum*. Adv Appl Microbiol. 33**,** 1-46.
Li, Y., Tschaplinski, T. J., Engle, N. L., Hamilton, C. Y., Rodriguez, M., Jr., Liao, J. C., Schadt, C. W., Guss, A. M., Yang, Y., 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**,** 2.
Linville, J. L., Rodriguez, M., Jr., Land, M., Syed, M. H., Engle, N. L., Tschaplinski, T. J., Mielenz, J. R., Cox, C. D., 2013. Industrial robustness: understanding the mechanism of tolerance for the Populus hydrolysate-tolerant mutant strain of *Clostridium thermocellum*. PloS one. 8**,** e78829.
Lynd, L. R., van Zyl, W. H., McBride, J. E., Laser, M., 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current opinion in biotechnology. 16**,** 577-83.
Mori, Y., 1990. Characterization of a Symbiotic Coculture of *Clostridium thermohydrosulfuricum* Ym3 and *Clostridium thermocellum* Ym4. Appl Environ Microb. 56**,** 37-42.
Olson, D. G., Lynd, L. R., 2012. Transformation of *Clostridium thermocellum* by electroporation. Method Enzymol. 510**,** 317-30.
Rydzak, T., Lynd, L., Guss, A., 2015. Elimination of formate production in *Clostridium thermocellum*. J Ind Microbiol Biot.
Schell, D. J., Farmer, J., Newman, M., McMillan, J. D., 2003. Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor - Investigation of yields, kinetics, and enzymatic digestibilities of solids. Appl Biochem Biotech. 105**,** 69-85.
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., 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 Microb. 76**,** 6591-9.
Wilson, C. M., Rodriguez, M., Jr., Johnson, C. M., Martin, S. L., Chu, T. M., Wolfinger, R. D., Hauser, L. J., Land, M. L., Klingeman, D. M., Syed, M. H., Ragauskas, A. J., Tschaplinski, T. J., Mielenz, J. R., Brown, S. D., 2013. Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass. Biotechnol Biofuels. 6**,** 179.
Yee, K. L., Rodriguez, M., Jr., Thompson, O. A., Fu, C., Wang, Z. Y., Davison, B. H., Mielenz, J. R., 2014. Consolidated bioprocessing of transgenic switchgrass by an engineered and evolved *Clostridium thermocellum* strain. Biotechnology for biofuels. 7**,** 75.

CHAPTER II
IMPROVED GROWTH OF *CLOSTRIDIUM THERMOCELLUM* *ΔHYDG ΔLDH ΔPFL ΔPTA-ACK VIA MEDIUM SUPPLEMENTATION WITH FORMATE AND STRAIN EVOLUTION*

A portion of this chapter was originally published by Liang Tian, Beth Papanek, Daniel G. Olson, Thomas Rydzak, Evert K. Holwerda, Tianyong Zheng, Jilai Zhou, Marybeth Maloney, Nannan Jiang, Richard J. Giannone, Robert L. Hettich, Adam M. Guss, and Lee R. Lynd:

Tian, L, Papanek, B, Olson, DG, Rydzak, T, Holwerda, EK, Zheng, T, Zhou, J, Maloney, MB, Jiang, N, Giannone, RJ, Hettich, RL, Guss, AM, Lynd, LR. "Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum." Biotechnology for Biofuels* (2016) Accepted.

The remaining work from this chapter is in preparation by Beth Papanek, Kaela O'Dell, Punita Manga, Richard J. Giannone, Dawn M. Klingeman, Robert L. Hettich, Steven D. Brown, and Adam M. Guss:

Papanek, B, O'Dell, K, Manga, P, Giannone, RJ, Klingeman, DM, Hettich, RL, Brown, SD, Guss, AM "Omics study of improved growth in *Clostridium thermocellum* Δ*hydG*Δ*ldh*Δ*pfl*Δ*pta-ack* via medium supplementation with formate."

Strain evolution was done by Nannan Jiang, single colony isolation and genomic DNA preparation was done by Beth Papanek the Joint Genome Institute performed the strain resequencing, growth and fermentation data was collected by Kaela O'Dell, omics cultures were grown and prepared by Beth Papanek, Dawn Klingeman prepared RNA and ran RNA sequencing samples, Richard Giannone prepared and ran proteomics samples. Omics analysis and manuscript preparation was done by Beth Papanek and Adam Guss.

Abstract

Clostridium thermocellum is a potentially useful organism for the production of lignocellulosic biofuels because of its ability to directly deconstruct cellulose and convert it to ethanol in a process called Consolidated Bioprocessing (CBP). In order to achieve higher yields and titers of ethanol produced by *C. thermocellum,* previous work used gene deletions to remove fermentation side products (Δ*hpt* Δ*hydG* Δ*ldh* Δ*pfl* Δ*pta-ack*), generating strain AG553. This strain showed improved ethanol yield relative to the wild type strain, but also exhibited slowed growth. In order to improve the growth rate of the organism, we used serial transfer to fresh medium daily for ~1500 generations in two separate lineages. This laboratory evolution successfully improved the growth rate from 0.13 ± 0.01 hr^{-1} in AG553 to 0.20 ± 0.02 hr^{-1} in Lineage 1 and 0.21 ± 0.03 hr^{-1} in Lineage 2 making the evolved strains more similar to the wild type growth rate of 0.25 ± 0.04 hr^{-1}. In order to understand the improvement in growth rate, single colonies were isolated and resequenced, and one single colony (strain AG601) was selected for further investigation. Medium supplementation with formate was recently found to improve the growth rate of *C.*

thermocellum Δ*pfl*, and we found this to be true for strains AG553 and AG601 as well. To understand the impact of formate supplementation on wild type, AG553, and AG601, proteomics and RNA sequencing analyses were performed with and without the addition of 2 mM formate. The stress response functions were dramatically more present in the original mutant strain, and to a lesser extent, in AG601. Even after the addition of formate, stress response genes remained upregulated, suggesting that other issues such as redox imbalances likely explain the stress. Expression of C1 metabolism genes varied with formate addition, indicating that the primary benefit of added formate is the supply of C1-units for biosynthesis. Overall, the combination of strain evolution and media supplementation successfully improved the growth of *C. thermocellum* while maintaining high ethanol yield.

Introduction

Alternative forms of energy, ranging from solar and wind power to nuclear fusion are being investigated as potential replacements to end human dependence on fossil fuels. One of the most promising of these options is the use of microorganisms for the production of fuels and chemicals that are traditionally derived from fossil fuels (Demain, 2009). An important consideration when implementing microbial systems is their ability to produce the fuels and chemicals renewably. Consolidated Bioprocessing (CBP), or using microbial systems directly to deconstruct lignocellulosic plant material to valuable compounds without additional enzymes is an example of using a more sustainable process (Olson et al., 2012).

Clostridium thermocellum is a gram positive, thermophilic Firmicute that utilizes a multi-enzyme complex called the cellulosome to naturally depolymerize cellulose and hemicellulose (Bayer et al., 1994). Its cellulolytic capabilities are the reason for its consideration as a candidate organism for CBP (Akinosho et al., 2014). Though *C. thermocellum* naturally degrades lignocellulosic biomass, it produces a number of fermentation products including lactate, formate, acetate, ethanol, hydrogen, and amino acids (Ellis et al., 2012a; Holwerda et al., 2014). Metabolic engineering is a crucial step in controlling carbon and electron flux to only the product of interest, and substantial effort has gone to engineering *C. thermocellum* for the optimization of ethanol production (Argyros et al., 2011; Biswas et al., 2014; Biswas et al., 2015b; Deng et al., 2013; Papanek et al., 2015; Rydzak et al., 2015b; Tripathi et al., 2010). Unfortunately, engineering the central metabolism of *C. thermocellum* has, in many cases, resulted in slower or more inhibited growth phenotypes (Biswas et al., 2015b; Papanek et al., 2015; Rydzak et al., 2015b). This phenomenon was observed in strain AG553 (*C. thermocellum* Δ*hpt* Δ*hydG* Δ*ldh* Δ*pfl* Δ*pta-ack*), in which metabolic pathways to acetate, lactate, formate and most H_2 were deleted. AG553 only exhibited a

growth rate of 0.13 hr^{-1} compared with a growth rate of 0.25 hr^{-1} for the wild type strain. When applied in an industrial setting, both the ethanol yield and the rate at which that ethanol is produced are crucial.

Previously, the pyruvate:formate lyase (*pfl*) was deleted in an effort to improve ethanol yields achieved by *C. thermocellum*. The deletion of *pfl* tripled the generation time of the strain in comparison to the wild type (Rydzak et al., 2015b). When *C. thermocellum* Δ*pfl* was grown in medium supplemented with formate, the generation time improved from approximately 6 hours to 3.4 hours. A previous study in *Staphylococcus aureus* had shown that supplementing the growth medium with formate after the deletion of *pfl* restored much of the growth profile (Leibig et al., 2011). The elimination of formate production was proposed to inhibit growth due to the disruption of C1 metabolism. Formate acts as a donor for formylation of tetrahydrofolate (THF), which in turn acts as a precursor for purine, formylmethionine, methioinine and S-adenosyl-methionine synthesis. Because Δ*pfl* is part of the genetic background in strain AG553, it stood to reason that formate addition could benefit the growth of this strain similarly.

Strain evolution is a useful technique in microorganisms because of their fast generation times, large population sizes, and the ability to freeze and store a certain point in the evolutionary process for future investigation (Elena and Lenski, 2003). Adaptive Laboratory Evolution allows specific phenotypes to be connected or correlated to specific genetic mutations (Conrad et al., 2011). This approach is exceptionally useful for complex phenotypes where rational targets for improvement might not exist such as toxin tolerance, or growth rate. Subsequent genome resequencing then allows causative mutations to be correlated with the improved phenotype. These benefits applied directly to the inhibited growth seen in *C. thermocellum* AG553. To improve the growth rate, evolutionary strategies and growth medium supplementation were applied to AG553, the highest ethanol yielding strain of *C. thermocellum.* Genomics, transcriptomics and proteomics were implemented in order to elucidate the impact both evolution and medium supplementation had on AG553 as well as its evolutionary descendant, AG601.

Materials and Methods

Growth Media

C. thermocellum was grown in modified DSM122 rich medium referred to as CTFUD (Olson and Lynd, 2012b) or defined Media for Thermophilc Clostridia (MTC) (Hogsett, 1995) modified to reduce urea concentration and using filter sterilization for all components instead of autoclaving. This modified defined media is referred to as MTC_5 (Rydzak et al., 2015b). Where indicated, MTC_5 was supplemented with 2 mM sodium formate.

Evolution and Isolation of Colonies

C. thermocellum strain AG553 was inoculated into two separate cultures of 5 mL CTFUD. Each culture was diluted approximately 1000-fold into 5 mL fresh CTFUD medium approximately once per day for 150 transfers, generating two separate lineages of evolved strain AG553. Frozen stocks were stored at -80°C after every ten passages.

Individual members of these lineages were isolated from both culture #30 and #150 for each lineage (Table 2). Approximately 5 μL of each growing culture was streaked on a 1.5% CTFUD agar plate. Colonies were visible and distinct after two days and single colonies were selected into 5 mL CTFUD medium.

Genome Resequencing

Genomic DNA was isolated using a ZR Fungal/Bacterial DNA Mini Prep™ kit (Zymo Research, Orange CA) and sent to the Joint Genome Institute (Walnut Creek, CA) for resequencing. DNA library preparation for Illumina sequencing was performed using 100ng of sample DNA that was sheared to 500bp using a Covaris LE220 focused-ultrasonicator. The sheared DNA fragments were size selected by double-SPRI and then using a Kapa Biosystems' library preparation kit the selected fragments were end-repaired, A-tailed, and ligated with Illumina compatible sequencing adaptors from IDT containing a unique molecular index barcode for each sample library. The prepared libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then multiplexed into a pool of 10 libraries for sequencing, and the pool was loaded and sequenced on the Illumina MiSeq sequencing platform utilizing a MiSeq Reagent Kit, v2, 300 cycle, following a 2x150 indexed run recipe.

These reads were aligned to the reference genome using BWA (Li and Durbin, 2009) and putative SNPs and small indels were called using samtools and mpileup (Li et al., 2009). Putative structural variants were called using a combination of BreakDancer (Chen et al., 2009) (filtered to quality 90+), CNVnator (Abyzov et al., 2011) and Pindel (Ye et al., 2009). The summary of these results and identified mutations were then consolidated and provided by JGI.

Cultures for Proteomics and RNA Sequencing

Cultures of *C. thermocellum* were revived from -80°C freezer stocks into 5 mL MTC_5 at 51°C in a Coy Anaerobic Chamber (Coy Laboratory Products, Grass Lake, MI) and grown to mid exponential phase (O.D. ~0.4). 2 mL of the growing culture was added to 100 mL MTC_5 in a 298 mL stoppered bottle (Chemglass Life Sciences, Vineland, NJ) to a working volume of 102 mL. Cultures were

grown to exponential phase (O.D. ~0.27) at which point they were rapidly transferred to two 50 mL conical tubes and centrifuged at 8,000 rpm for five minutes. The supernatant was removed and the cultures were immediately flash frozen in liquid N_2. Frozen cell pellets were held at -80°C until proteomics or RNA sequencing analysis was performed.

Results

Strain evolution improved growth rate of C. thermocellum AG553

C. thermocellum strain AG553 grows much more slowly than wild type, presumably due to metabolic bottlenecks or imbalances. To improve the growth rate, the strain was serially transferred approximately daily such that the fastest growing cells would begin to dominate the culture. Each transfer consisted of a 1000-fold dilution, or approximately 10 generations. The cultures were checked for their average growth rates after every thirty transfers (Figure 2-1). The average growth rate improved over time for both lineages, with a large increase occurring for lineage #1 at the 30^{th} transfer. After 150 passages (~1500 generations), the growth rate approached that of wild type (Figure 2-1).

The nature of this type of laboratory evolution results in the cultures being a mixture of different genetically distinct *C. thermocellum* cells. This diversity of genotypes makes understanding of the causative mutation(s) difficult and further genetic manipulation impossible. Therefore, the isolation of individual genetic background strains was necessary, and individual cells from the populations were isolated via single colony purification. Colonies were isolated from each lineage from early in the evolution (transfer #30) and from the end (transfer #150) (Table 2-1).

Resequencing of Isolated Single Colonies

To compare the genetic changes of the individual isolates, genome resequencing was performed with Illumina sequencing on the original AG553 strain and 9 isolates (Table 2-1). Nineteen total mutations were found among the resequenced strains. These mutations were strongly divided based on which of the two lineages was the source of the strain. Only one of the nineteen was found in both lineage one and lineage two strains. Of the resequenced strains, AG601 was identified as unique and of possible interest for improved growth. This led us to use AG601 as a representative “fast-growing” strain of *C. thermocellum* for further investigation. AG601 possessed nine of the nineteen mutations, and had two mutations unique to only that strain including the non-synonymous SNP found in Clo1313_0637 (Sporulation transcriptional activator, Spo0A) and the non-coding SNP found in the intergenic region between

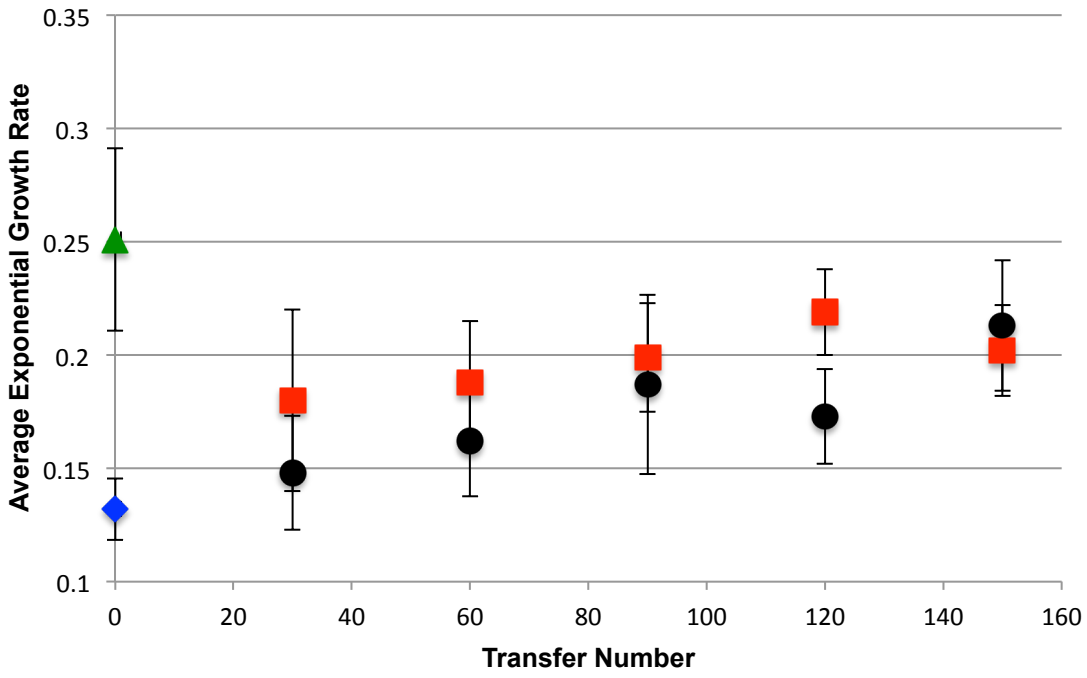

Figure 2-1 Average growth rate at different transfers during strain evolution

AG553 was transferred daily into fresh CTFUD in two separate lineages. Cultures were tested for growth rate every thirty transfers. Red squares, Lineage 1; Black circles, Lineage 2; Green triangle, Wild Type; Blue diamond, AG553 before evolution.

Clo1313_0932 and Clo1313_0933. While there were no mutations in catabolism-related genes that could immediately explain the phenotypic change observed in the evolved lineages, there were several that seemed to be associated with transcriptional processes. As the representative strain, was used for further study into the mechanisms of improved growth phenotypes through comparison with wild type and AG553.

Formate addition increases growth rate of mutant strains

As it was observed in the Rydzak, et. al (2015) study, the addition of 2 mM formate to the growth medium of the Δ*pfl* restored the growth rate to that of the wild type strain, possibly by providing C1 building blocks for biosynthesis. Because *pfl* is deleted in strain AG553, the impact of adding formate to the growth medium was tested (Figure 2-2). The growth rate of both AG553 and AG601 increased significantly with the addition of formate, while no statistical difference was seen in the growth of wild type. Fermentation product formation of all three strains was also tested with and without the addition of 2 mM formate (Table 2-2). No difference was seen in the fermentation profiles of all three strains when formate is added.

Genotype and formate addition both impact mRNA and protein abundance

To further understand the observed differences between the wild type, AG553, and AG601, both proteomics and RNA sequencing were performed with and without 2 mM formate supplemented in the medium. The RNA sequencing data was analyzed first to understand the transcriptomic response of the different strains of *C. thermocellum* with and without added formate. All differential expression is reported as a $\log_2$ fold change. Expression differences were required to meet a cutoff of ± 1.0 to be considered for analysis.

In wild type *C. thermocellum*, only two genes were differentially expressed upon addition of formate to the growth medium. Clo1313_0111 and Clo1313_0113 were both moderately downregulated (-1.1 and -1.2 $\log_2$ fold change, respectively) with the addition of formate. Both of these genes are predicted to be involved in siroheme biosynthesis.

Comparison of strain AG553 to the wild type strain with no formate added can help elucidate the transcriptomic response to the deletion of the four central metabolism genes. Unsurprisingly, the deletion of these genes seemed to have a large effect on gene expression, under these conditions; 1184 genes were differentially expressed relative to wild type. Of these, 625 showed increased expression while 559 decreased. Many of these differentially expressed genes were involved in the stress response, with sporulation, shock response, and cell protection genes being both the most commonly upregulated and some of the most highly upregulated. Among the downregulated genes, Clo1313_0108 through Clo1313_0115, a cluster of genes involved in C1 metabolism

Table 2-1 List of all strains isolated from evolution of AG553 and mutations found from resequencing single colonies.

X's indicate the strains where each mutation is found in reference to the unevolved AG553 sequence.

		Lineage	1						2				
		Transfer	30	150					30	150			
Type of Mutation	Base Pair	Effect	AG599	AG600	AG601	AG602	AG603	AG604	AG605	AG606	AG607	Gene Locus	Gene Name
SNP	963	Non-synonymous								X		Clo1331_0001	Chromosomal replicator initiator protein DnaA
SNP	510359	Non-synonymous		X	X	X	X	X				Clo1313_0460	50S ribosomal protein L6
SNP	724432	Non-synonymous			X							Clo1313_0637	Sporulation transcriptional activator, Spo0A
SNP	880814	Non-synonymous	X	X	X	X	X	X				Clo1313_0764	CRISPR-associated protein
SNP	1090041	Non-coding			X							Between Clo1313_0932 & Clo1313_0933	Intergenic region between protein HSP-70 cofactor & chaperone protein DnaK
Deletion	1101762	Frameshift						X				Clo1313_0945	ArgC quorum sensing receptor; signal transduction histidine kinase regulating citrate/malate metabolism
SNP	1115704	Non-synonymous								X	X	Clo1313_0955	RNA-metabolizing metalla-beta-lactamase
SNP	1569773	Non-synonymous									X	Clo1313_1323	DNA Primase, DnaG
SNP	1705320	Synonymous		X								Clo1313_1447	tRNA delta(2)-isopentenylpyrophosphate transferase
SNP	2135443	Non-synonymous	X	X	X	X	X	X				Clo1313_1824	Uncharacterized ABC transporter
Deletion	2172122	Frameshift								X		Clo1313_1857	XRE family transcriptional regulator
Deletion	2175943	Non-coding							X	X	X	Between Clo1313_1861 & Clo1313_1862	Intergenic region between hypothetical protein & thioredoxin
SNP	2207637	Non-synonymous		X	X	X	X	X				Clo1313_1891	Uridine kinase
Insertion	2343123	Non-coding								X		Between Clo1313_1989 & Clo1313_1990	Intergenic region between VTC domain-containing protein & spore coat protein CotH

Table 2-1 Continued

		Lineage	1						2				
		Transfer	30	150					30	150			
Type of Mutation	Base Pair	Effect	AG599	AG600	AG601	AG602	AG603	AG604	AG605	AG606	AG607	Gene Locus	Gene Name
SNP	2732187	Non-coding		X	X	X	X	X				Between Clo1313_2323 & Clo1313_2324	Intergenic region between ABC transporter & AsnC family transcriptional regulator
SNP	2876908	Non-synonymous									X	Clo1313_2462	AAA+ ATPase; Putative ClnC protease
Deletion	3099115	Frameshift		X	X	X	X	X				Clo1313_2633	Nucleotidyltransferase; mannose-1-phosphate guanyltransferase
SNP	3189147	Non-synonymous								X	X	Clo1313_2711	Hypothetical protein; DUF324
SNP	3327209	Non-synonymous	X	X	X	X	X	X			X	Clo1313_2832	Rho transcription termination

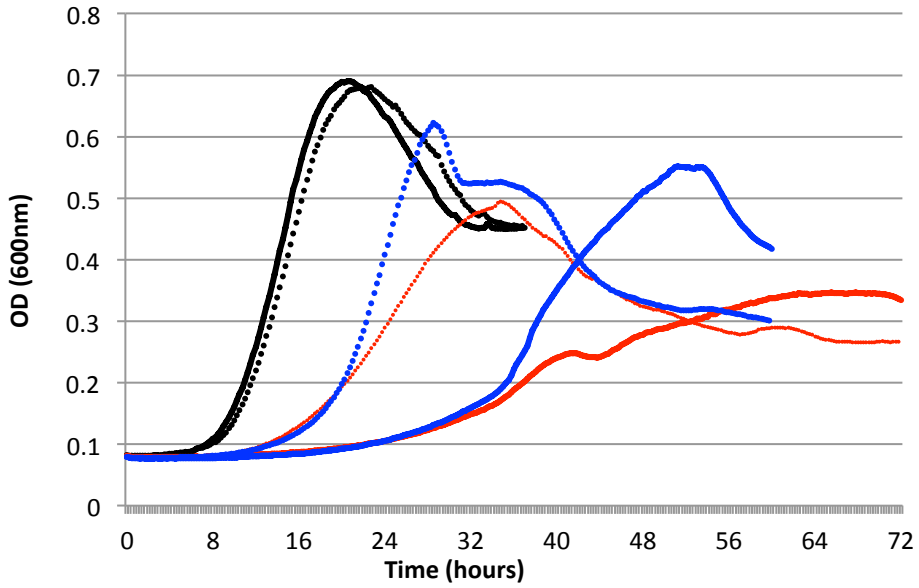

Figure 2-2 Growth data for Wild Type, AG553, and AG601 with and without formate supplementation

Growth curves show the average of three cultures sampled every 15 minutes. Black, WT; Red, AG553; Blue, AG601. For all three the dotted line shows with formate addition.

Table 2-2 End point fermentation data for Wild Type, AG553, and AG601 with and without formate addition

Strain	Condition	Ethanol	Acetate	Formate	Lactate	Excreted Amino Acids
Wild Type	MTC_5	16.6 ±1.2mM	16.0 ±3.2mM	7.0 ± 2.9mM	0.3 ± 0.3mM	1.9 ±0.1mM
	MTC_5 + 2mM Formate	16.2 ±2.1mM	16.2 ±5.4mM	8.8 ± 3.8mM	0.3 ± 0.4mM	1.9 ±0.1mM
AG553	MTC_5	32.0 ±2.5mM	0	0	0	1.9 ±0.0mM
	MTC_5 + 2mM Formate	30.8 ±2.7mM	0	1.5 ± 0	0	2.6 ±0.0mM
AG601	MTC_5	33.5 ±2.8mM	0	0	0	1.5 ±0.1mM
	MTC_5 + 2mM Formate	29.5 ±1.9mM	0	2.7 ± 0.2mM	0	1.7 ±0.1mM

(Figure 2-3), show some of the greatest differences. Looking at the categorization of the differentially expressed genes, general processes showed some strong trends as well. In AG553, 59 genes in the Translation, Ribosomal Structure, and Biogenesis category were downregulated, while 50 genes in the category for Signal Transduction Mechanisms were also downregulated. Other downregulated categories included Intracellular Trafficking and Secretion (12 genes down) and Cell Motility and Secretion (19 genes down).

To identify the effects of formate addition on the unevolved AG553, the RNA sequencing profiles of AG553 with and without formate supplementation were compared. After formate addition, 351 genes were differentially expressed. Of these, 198 were upregulated and 153 downregulated. Several C1 metabolism genes were again among the differentially expressed. A series of four genes (Clo1313_1009 – Clo1313_1013) involved in converting PRPP from the Pentose Phosphate Pathway to aminoimidazole ribotide for thiamine metabolism were all down regulated upon addition of formate. A series of five genes (Clo1313_0109 – Clo1313_0112) involved in siroheme biosynthesis were all significantly upregulated with the addition of formate in AG553 (Figure 2-3). Genes involved in cell stress and shock were also still very prevalent in the RNA sequencing data. Both transcriptional and translational processes showed upregulation with the addition of formate to the AG553 strain. This was supported by the COG categorization. Categories that showed a strong upregulation trend included Translation, Ribosomal Structure & Biogenesis (37 genes), Signal Transduction Mechanisms (24 genes), Cell Motility and Secretion (14 genes). In addition, Amino Acid Transport and Metabolism, Coenzyme Metabolism, and Translation all showed an overall trend toward upregulation after formate was added.

To help identify the mechanisms for improved growth after evolution of AG553, the evolved strain AG601 was compared with AG553 with no formate added. Strain AG601 showed 905 differentially expressed genes compared with AG553 with 374 upregulated after evolution. While differences in the COG categories were somewhat less clear than in previous comparisons, Energy Production and Conversion, Nucleotide Transport and Metabolism, and Posttranslational Modification, Protein Turnover, Chaperones all showed a trend toward upregulation.

In strain AG601 the impact of formate addition was far less dramatic than in strain AG553. Only 234 genes were seen to be differentially expressed, with 117 of these upregulated. Both transcriptomic comparisons of strain AG601 as well as the resequencing results imply that rather than *C. thermocellum* AG553 acquiring mutations in a narrow set of metabolic pathways, mutations resulted in a more widespread, general genetic upregulation to eliminate metabolic bottlenecks experienced after the deletion of central metabolism genes.

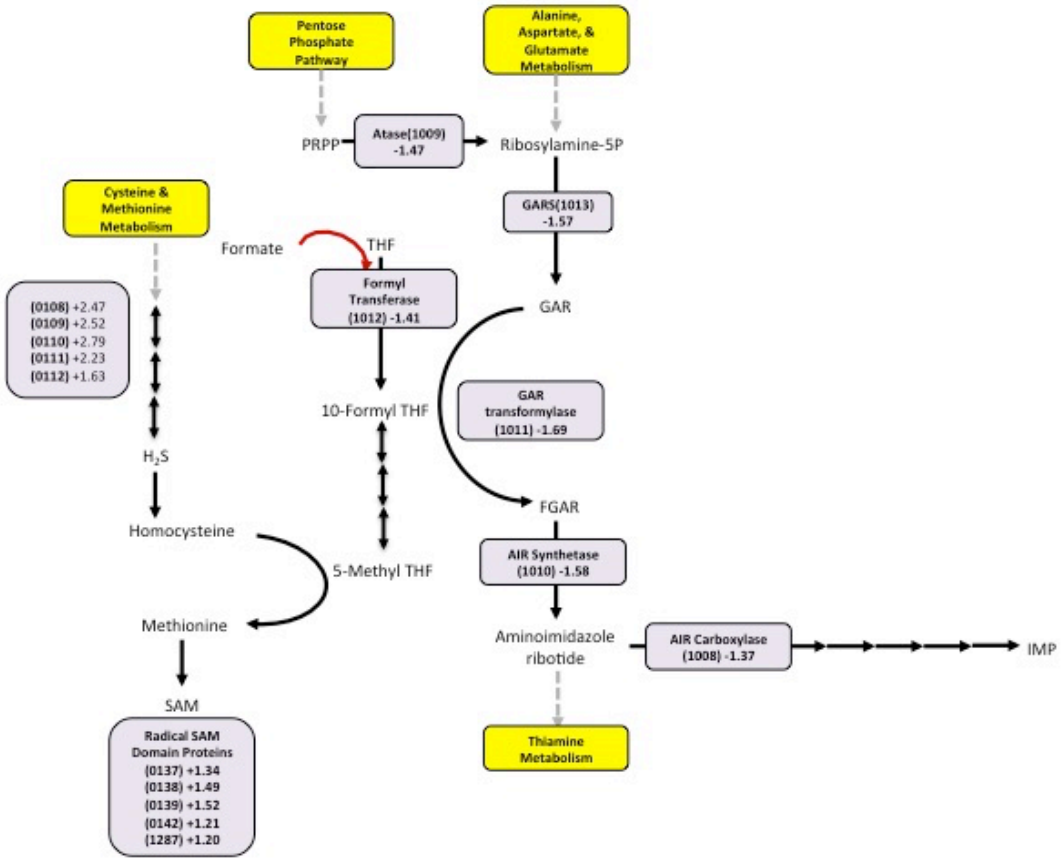

Figure 2-3 Summary of C1 metabolism changes observed in AG553 Formate vs. AG553 No Formate

Gene numbers are named in parentheses and numbers associated indicate the $\log_2$ fold change seen in the RNA sequencing data.

The trends in the proteomics data supported the trends observed in the RNA sequencing data. Though the fold changes in the proteomics data were generally lower than those seen in the corresponding RNA sequencing, the two sets of data were correlated for all of the conditions tested (Figure 2-4). The congruence between the transcriptomic and proteomic data further supports the correlations described above.

Discussion

For industrial biofuel production, strain performance will be critical; including not only is the yield and titer of ethanol, but also the rate at which that ethanol is produced. Though strain AG553 is amongst the highest ethanol-yielding strains of *C. thermocellum* published to date, it exhibited a severely reduced growth rate compared to the wild type. Serial transfer of strain AG553 into fresh medium improved the growth rate of both lineages over time. After 150 transfers, the strains exhibited a growth rate close to that of the wild type strain while still producing ethanol as its only major fermentation product. This increased growth rate combined with the higher ethanol yields brings *C. thermocellum* closer to having the attributes needed for industrial application.

Although multiple evolved single colonies were isolated from the initial mixed cultures, there were a relatively limited number of genetic changes seen in the resequencing results. Only 14 SNPs, 1 insertion and 4 deletions were seen across the nine separate evolved strains that were resequenced. Of these mutations, there were even fewer that provided direct insight into potential mechanisms for improved growth. Mutations in Clo1313_0637 (the Spo0A homolog) and Clo1313_2832 (the Rho transcriptional termination factor) of AG601 were possible exceptions to this trend. Spo0A is a conserved master regulator of the sporulation process (Paredes et al., 2005). While most spore-forming bacteria have a single homolog of *spo0A*, *C. thermocellum* has two, Clo1313_0637 and Clo1313_1409, which share 56% amino acid identity with each other. The Clo1313_1409 protein is 58.3% identical to the well-characterized *Bacillus subtilis* Spo0A, whereas Clo1313_0637 is 49% identical. Previous deletion of Clo1313_1409 completely eliminated sporulation in *C. thermocellum* (Mearls et al., 2012), suggesting that it might be the primary controller of sporulation. However, a mutant strain of *C. thermocellum* ATCC27405 that was evolved to be tolerant to poplar hydrolysate was found to have a mutation in the second *spo0A* homolog Cthe3087 (Linville et al., 2013) (99% amino acid identity to Clo1313_0637), suggesting that the second Spo0A may also play a role in coordinating a stress response such as sporulation. This mutation could allow the cell to continue to metabolize sugar and grow even under stressful conditions. Further investigation into the role of Clo1313_0637 may help elucidate its function. Another potentially important mutation in AG601 is in the termination factor *rho* (Clo1313_2832). While this gene is typically

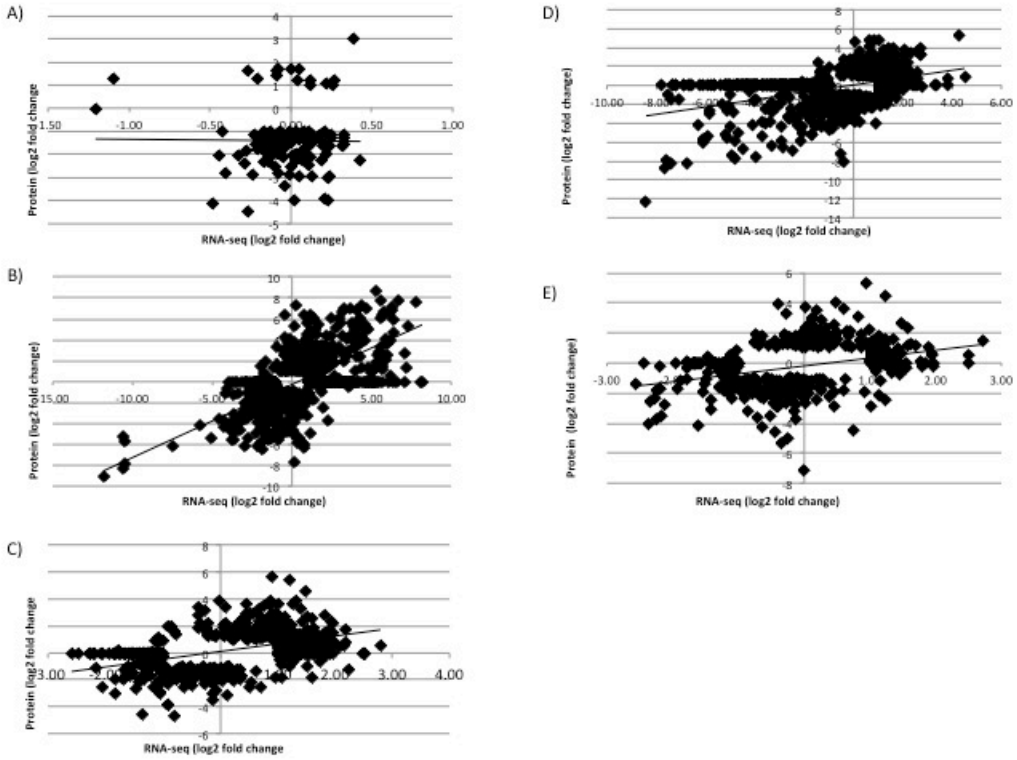

Figure 2-4 Proteomics log2 fold changes vs RNA sequencing log2 fold changes

A) Wild type +2mM formate vs no formate B) AG553 no formate vs Wild Type no formate C) AG553 +2mM formate vs AG553 no formate D) AG601 no formate vs AG553 no formate E) AG601 +2mM formate vs AG601 no formate

essential for growth, SNPs in *rho* in *E. coli* helped to confer resistance to ethanol by enhancing transcriptional read-through and altering gene expression (Haft et al., 2014). A similar mechanism may be occurring in AG601.

Formate addition improved the growth rate of strains of *C. thermocellum* containing the *Δpfl* mutation. Exogenous formate did not, however, alter the fermentation end product profiles in any of the three strains studied suggesting that formate plays a crucial role in a different aspect of metabolism. In addition to being a fermentation product, formate acts as a C1 donor group for biosynthesis of serine, purines, methionine, and formyl-methionine (Thauer et al., 1972). Upon the addition of formate, both the RNA sequencing and the proteomics response data sets show changes in many of the genes associated with C1 metabolism, providing further evidence that it is this portion of metabolism most affected by the presence or absence of formate.

Formate clearly plays a crucial role in amino acid and serine metabolism, and one might hypothesize that eliminating formate production in the cell was causing the observed stress response. Interestingly, when formate was added to the AG553 culture, many of the same stress response genes continued to be upregulated. This implies that there is a strong stress response even when biosynthetic needs are met, that exogenous formate compensates for the *pfl* deletion, and that deficiency in C1 metabolism is not the only or primary source or stress in the cell. The continued presence of stress response genes is more likely to be caused by a redox imbalance rather than the presumed lack of metabolic building blocks.

Evolution and isolation of AG601 produced a strain with greatly improved the growth rate compared to the unevolved mutant strain AG553. Taking into consideration the genomic resequencing, the RNA sequencing, and the proteomics data, it seems likely that this improvement was primarily the cell finding ways to adapt to the stress of its disrupted central metabolism. Though some stress response genes were upregulated in AG601, they were fewer and less dramatically upregulated than in the unevolved AG553. AG601 was the only strain of the evolved isolates that contained the mutation in *spo*0A, likely interrupting the sporulation process, and conferring the ability to grow more rapidly than its unmutated counterparts. Laboratory evolution was successfully applied to *C. thermocellum* for growth rate improvement and mechanistic understanding of that improvement.

The use of genomics, proteomics, and transcriptomics allowed us insight into the causes and responses to a faster growth rate in *C. thermocellum.* Both evolution and medium supplementation were successful in restoring the mutant strain's growth rate to nearly that of the wild type strain. While individual genes cannot be correlated to this change, stress response and finding a way to

compensate for the C1 metabolic pathways seem to play a critical role in improving growth.

Conclusion

Though the high ethanol yield achieved in AG553 was an important step toward making an industrial CBP organism, the slow growth phenotype is a severe limitation. By utilizing strain evolution, growth was improved, and combining several omics technologies, we were able to more fully understand how growth can be improved. This will help inform further metabolic engineering efforts to create a better industrial strain of *C. thermocellum*.

References

Abyzov, A., Urban, A. E., Snyder, M., Gerstein, M., 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 21**,** 974-84.

Akinosho, H., Yee, K., Close, D., Ragauskas, A., 2014. The emergence of Clostridium thermocellum as a high utility candidate for consolidated bioprocessing applications. Frontiers in chemistry. 2**,** 66.

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., Caiazza, N. C., 2011. High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microb. 77**,** 8288-8294.

Bayer, E. A., Morag, E., Lamed, R., 1994. The cellulosome--a treasure-trove for biotechnology. Trends Biotechnol. 12**,** 379-86.

Biswas, R., Prabhu, S., Lynd, L. R., Guss, A. M., 2014. Increase in Ethanol Yield via Elimination of Lactate Production in an Ethanol-Tolerant Mutant of Clostridium thermocellum. PloS one. 9**,** e86389.

Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., Guss, A. M., 2015. Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in Clostridium thermocellum. Biotechnology for biofuels. 8**,** 20.

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L., Mardis, E. R., 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 6**,** 677-81.

Conrad, T. M., Lewis, N. E., Palsson, B. Ø., 2011. Microbial laboratory evolution in the era of genome-scale science. Molecular Systems Biology. 7.

Demain, A. L., 2009. Biosolutions to the energy problem. J Ind Microbiol Biotechnol. 36**,** 319-32.

Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Joe Shaw, A., Lynd, L. R., 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. Metabolic engineering. 15**,** 151-8.

Elena, S. F., Lenski, R. E., 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet. 4**,** 457-469.

Ellis, L. D., Holwerda, E. K., Hogsett, D., Rogers, S., Shao, X., Tschaplinski, T., Thorne, P., Lynd, L. R., 2012. Closing the carbon balance for fermentation by *Clostridium thermocellum* (ATCC 27405). Bioresource technology. 103**,** 293-9.

Haft, R. J., Keating, D. H., Schwaegler, T., Schwalbach, M. S., Vinokur, J., Tremaine, M., Peters, J. M., Kotlajich, M. V., Pohlmann, E. L., Ong, I. M., Grass, J. A., Kiley, P. J., Landick, R., 2014. Correcting direct effects of

ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. Proc Natl Acad Sci U S A. 111, E2576-85.
Hogsett, D., Cellulose hydrolysis and fermentation by *Clostridium thermocellum* for the production of ethanol. Thayer School of Engineering. Dartmouth College, Hanover, NH, 1995.
Holwerda, E. K., Thorne, P. G., Olson, D. G., Amador-Noguez, D., Engle, N. L., Tschaplinski, T. J., van Dijken, J. P., Lynd, L. R., 2014. The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading. Biotechnology for biofuels. 7, 155.
Leibig, M., Liebeke, M., Mader, D., Lalk, M., Peschel, A., Gotz, F., 2011. Pyruvate formate lyase acts as a formate supplier for metabolic processes during anaerobiosis in Staphylococcus aureus. J Bacteriol. 193, 952-62.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 25, 1754-60.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., Genome Project Data Processing, S., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25, 2078-9.
Linville, J. L., Rodriguez, M., Jr., Land, M., Syed, M. H., Engle, N. L., Tschaplinski, T. J., Mielenz, J. R., Cox, C. D., 2013. Industrial robustness: understanding the mechanism of tolerance for the Populus hydrolysate-tolerant mutant strain of *Clostridium thermocellum*. PloS one. 8, e78829.
Mearls, E. B., Izquierdo, J. A., Lynd, L. R., 2012. Formation and characterization of non-growth states in *Clostridium thermocellum*: spores and L-forms. BMC microbiology. 12, 180.
Olson, D. G., Lynd, L. R., 2012. Transformation of *Clostridium thermocellum* by electroporation. Method Enzymol. 510, 317-30.
Olson, D. G., McBride, J. E., Shaw, A. J., Lynd, L. R., 2012. Recent progress in consolidated bioprocessing. Curr Opin Biotechnol. 23, 396-405.
Papanek, B., Biswas, R., Rydzak, T., Guss, A. M., 2015. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in Clostridium thermocellum. Metab Eng. 32, 49-54.
Paredes, C. J., Alsaker, K. V., Papoutsakis, E. T., 2005. A comparative genomic view of clostridial sporulation and physiology. Nat Rev Microbiol. 3, 969-78.
Rydzak, T., Lynd, L. R., Guss, A. M., 2015. Elimination of formate production in Clostridium thermocellum. J Ind Microbiol Biotechnol. 42, 1263-72.
Thauer, R. K., Kirchniawy, F. H., Jungermann, K. A., 1972. Properties and function of the pyruvate-formate-lyase reaction in clostridiae. Eur J Biochem. 27, 282-90.
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., 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 Microb. 76**,** 6591-9.
Ye, K., Schulz, M. H., Long, Q., Apweiler, R., Ning, Z., 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 25**,** 2865-71.

CHAPTER III
OVERCOMING RESTRICTION FOR IMPROVED TRANSFORMATION EFFICIENCY IN *CLOSTRIDIUM THERMOCELLUM* DSM 1313

The work in this chapter is in preparation by Beth Papanek, Lauren A. Riley, Daniel G. Olson, Lee R. Lynd, Jan Westpheling, and Adam M. Guss:

Papanek, B, Riley, LA, Olson, DG, Lynd, LR, Westpheling, J, Guss, AM "Overcoming restriction for improved transformation efficiency in *Clostridium thermocellum* DSM 1313."

PacBio sequencing was completed by Expression Analysis, REBASE analysis and gene targets were identified by Beth Papanek, *in vivo* expression of Methyltransferase genes was done by Lauren Riley, restriction assays were designed by Beth Papanek and performed by Beth Papanek and Lauren Riley. Restriction Enzyme deletion strains were constructed and transformation efficiency was tested by Beth Papanek, manuscript preparation was done primarily by Beth Papanek and Adam Guss.

Abstract

Clostridium thermocellum is a promising organism for production of advanced biofuels due to its ability to convert cellulosic biomass directly to ethanol, but one of the challenges in performing genetic engineering in this organism is low transformation efficiency. Restriction-Modification (RM) systems are often a barrier to high transformation efficiency by cleaving newly introduced DNA; therefore, we performed methylome analysis to determine if DNA restriction could be hindering transformation in *C. thermocellum*. In addition to the methylation of G(methyl-A)TC, which was previously known, we identified a novel methylated site, CG(methyl-A)CC. Two candidate RM systems (Clo1313_0478-0480 and Clo1313_2366-2367) were identified in the *C. thermocellum* genome that could target this site. Heterologous co-expression of Clo1313_0480, an annotated methyltransferase, and the adjacent Clo1313_0479 resulted in approximately 50% DNA methylation of an *E. coli*/*C. thermocellum* shuttle vector at CGACC *in vivo* in *E. coli*, suggesting a novel 2-subunit structure for this Type III methyltransferase. Using this methylatyed shuttle vector, a 2-8-fold improvement in transformation efficiency was seen. Further deletion of the Clo1313_0478 restriction enzyme from the *C. thermocellum* chromosome resulted in a 20-fold improvement in transformation efficiency relative to the wild type. These strains will facilitate simpler and more rapid genetic modification of *C. thermocellum* and provide a roadmap for improving the transformation efficiency in other microbes.

Introduction

Clostridium thermocellum is a candidate organism for advanced biofuel production because of its ability to convert cellulosic biomass directly to ethanol in a process known as Consolidated Bioprocessing (CBP) (Olson et al., 2012). In addition to ethanol, *C. thermocellum* produces a range of other fermentation products including unusual products such as excreted amino acids (Ellis et al.,

2012a; Holwerda et al., 2014; Ng et al., 1977). These side products divert carbon and electron flux, thereby limiting the yields and titers of ethanol that can be achieved. To improve *C. thermocellum* performance for the industrial production of biofuels, ethanol production must be optimized through genetic engineering of the metabolic pathways in *C. thermocellum.*

While there is a functioning set of genetic tools used in *C. thermocellum* DSM 1313 that has successfully been used to construct a range of engineered strains (Argyros et al., 2011; Biswas et al., 2015b; Deng et al., 2013; Mohr et al., 2013; Papanek et al., 2015; Rydzak et al., 2015b; Tripathi et al., 2010), these systems remain inefficient and time-consuming. The inefficiency limits the types and scope of genetic manipulations that can be made; while gene deletions have been constructed using unstable replicating plasmids, there has been no success in using non-replicating vectors or introducing point mutations. If the genetic tools were expanded and efficiency improved, a greater variety of metabolic engineering of *C. thermocellum* could be done, allowing for the possibility of improved ethanol production from renewable feedstocks.

One limitation to transformation efficiency and genetic tool improvement is the presence of Restriction-Modification (RM) systems in *C. thermocellum* DSM1313. RM systems function as bacterial immune systems by cutting heterologous DNA that is differently methylated than the host organism (Arber and Linn, 1969). RM systems are classified into four groups. Type I, II, and III systems typically recognize and cleave specific sequences 4-8 nucleotides in length (Pingoud and Jeltsch, 1997; Tock and Dryden, 2005). Type I systems recognize non-palindromic, non-contiguous sequences, such as AAC(N_6)GTGC for EcoKI (Murray, 2000; Pingoud and Jeltsch, 1997). Type II systems are the well-known tools of molecular biology, and they typically target palindromic, contiguous sequences such as GAATTC for EcoRI (Pingoud and Jeltsch, 1997). Type III systems typically target contiguous, non-palindromic sequences, such as AGACC for EcoPI. The organism protects its own DNA by methylating those same sites, which prevents cleavage by the respective restriction enzyme. Thus, when foreign plasmid DNA is introduced into a host organism, these restriction systems cleave DNA not methylated in the same patterns the host organism would methylate naturally (Raleigh and Wilson, 1986). Type IV restriction systems, on the other hand, cleave methylated DNA when the sites are methylated in patterns foreign to the host organism (Liu et al., 2010; Roberts et al., 2003). For restriction to be overcome, all of the active RM systems must be avoided. This knowledge has led to numerous studies on proper methylation in order to improve transformation of foreign DNA into host organisms (Chung et al., 2013; Purdy et al., 2002; Yasui et al., 2009). This concept has also been applied directly in *C. thermoclleum* previously. Work to improve transformation in *C. thermocellum* DSM 1313 has shown that by methylating plasmid DNA with the

E. coli Dam methylase, and eliminating Dcm *E. coli* methylation, transformation efficiency could be improved by up to 500-fold (Guss et al., 2012).

Because this previous study only addressed methylation patterns available natively in laboratory strains of *E. coli*, it was hypothesized that the low transformation efficiencies were still being caused by additional restriction-methylation (RM) systems present in *C. thermocellum* DSM 1313. We therefore systematically investigated each putative RM in *C. thermocellum* DSM 1313 using a combination of a methylome analysis, heterologous expression in *E. coli*, and gene deletion in *C. thermocellum* in an attempt to improve transformation in C. thermocellum.

Materials and Methods

Strains, growth media, and culture conditions

Escherichia coli strains were grown on LB supplemented with 12 μg mL^{-1} chloramphenicol as needed. *E. coli* was grown aerobically at 37, except when otherwise noted. Strains of *C. thermocellum* were grown in CTFUD medium (Olson and Lynd, 2012b) at either 50 or 55°C as indicated inside a Coy anaerobic chamber (Coy Laboratory Products, Grass Lake MI). CTFUD medium was supplemented with 10 μg mL^{-1} thiamphenicol (Tm; Sigma-Aldrich St. Louis, MO), 50 μg/ml 5-fluoro-2'-deoxyuridine (FUdR; Sigma-Aldrich St. Louis, MO), or 500 μg/ml 8-azahypoxanthine (AZH; Tokyo Chemical Industry, Co., Tokyo, Japan) as needed.

C. thermocellum plasmid and strain construction

Plasmid pDG065 for deletion of the potential Restriction Enzyme Clo1313_0478 and pDG070 for deletion of the potential Restriction Enzyme Clo1313_2366 were constructed by PCR amplifying the relevant regions of homology from *C. thermocellum* DSM1313 gDNA and cloning the regions into pDGO68 (GenBank KU682275).. Standard methods were used to construct *C. thermocellum* deletions (Olson and Lynd, 2012b). Briefly, plasmids were isolated separately from a *dcm- E. coli* strain (Guss et al., 2012) and electroporated into *C. thermocellum* DSM1313 Δ*hpt*. Electroporated cells were immediately plated on CTFUD agar supplemented with Tm. Individual colonies were picked into liquid CTFUD supplemented with Tm, followed by plating dilutions in CTFUD supplemented with Tm and FUdR. Individual colonies were streaked on the same medium for purification. Purified colonies were then picked into liquid CTFUD in the absence of Tm, and then plated in CTFUD agar supplemented with 8AZH. Colonies were single colony purified, picked into liquid CTFUD, and deletions were confirmed by PCR.

C. thermocellum methylome analysis with PacBio sequencing

Genomic DNA from *C. thermocellum* DSM 1313 was isolated and sent to Expression Analysis (Durham, NC) for PacBio genome sequencing using four SMRT cells. DNA methylation was determined using the SMRT Analysis software.

C. thermocellum methylome analysis with PacBio sequencing

Methyl Transferase genes Clo1313_2367, Clo1313_ 0480 and Clo1313_0479-0480 were cloned into pAH55 and joined to the P_{tac} promoter using Gibson Assembly (New England Biolabs, Ipswich, MA) and integrated into the lambda *attB* site of *E. coli* strain Top 10 *Δdcm::frt* using the CRIM System (Haldimann and Wanner, 2001), resulting in strains AG1955 (Top 10 *Δdcm::frt* . λ::Clo1313_2367), AG1956 (Top 10 *Δdcm::frt* . λ::Clo1313_0480), and AG1957 (Top 10 *Δdcm::frt* . λ::Clo1313_0479-0480). Plasmid pNJ020 was transformed into each of these strains, and the resulting strains were grown in 50 mL LB cultures induced with 0.1 mM IPTG. Optimization of Clo1313_0479-0480 expression and activity was attempted by growth at room temperature, 37°C, and 45°C, as well as growth at 45°C supplemented with 0.1 mM betaine. Plasmid pNJ020 was digested with TaqαI (New England Biolabs, Ipswich, MA) according to the manufacturer's protocol.

Testing transformation efficiency

Electrocompetent cells were prepared fresh for each transformation efficiency experiment. CTFUD medium was inoculated in the afternoon and grown anaerobically overnight at 55°C to early stationary phase (O.D. ~1.0). Cultures were transferred to 500 mL centrifuge bottles and spun at 5000 rpm for 15 minutes. Supernatant was removed and replaced with cold electroporation buffer (10% glycerol, 40 mM sucrose). Cells were spun again and the wash was repeated. Cell pellets were resuspended in ~500 µL electroporation buffer and transferred to microcentrifuge tubes. For each electroporation, between 70 and 100 ng of DNA and 20 µL of electrocompetent cells were added to a 1 mm electroporation cuvette. The cuvette was then pulsed using the square wave protocol on a Bio-Rad electroporator with 1200 V, and a 1.5 second pulse length. Cells were immediately resuspended in 1 mL CTFUD and plated. Liquid CTFUD agar with Tm was poured for each plate. Plates were then incubated at 50°C for 3-4 days until the number of colonies could be accurately counted.

Results

Methylome analysis of C. thermocellum DSM1313

Pacific BioSciences (PacBio) Single-Molecule Real-Time (SMRT) sequencing technology was used to detect patterns of DNA bases modified by methylation in

the N^6-methyladenine (m^6A), N^4-methylcytosine (m^4C), and 5-Methylcytosine (m^5C) positions (Rhoads and Au, 2015). Two different patterns were methylated in *C. thermocellum* DSM1313 (Table 3-1). Of these, the GATC motif is naturally methylated by the *E. coli* Dam methyltransferase gene (Messer and Noyer-Weidner, 1988), while the other, CGACC, appears to be unique. This result implies that there is one active RM system present in *C. thermocellum* DSM1313 that is not being circumvented by the typical lab strains of *E. coli* being used for plasmid cloning and replication. We therefore needed to identify the RM system responsible for the CGACC methylation pattern.

Identification of putative RM system genes in C. thermocellum DSM1313

In order to identify possible RM system genes responsible for methylating and presumably cleaving CGACC motifs, the Restriction Enzyme Database (REBASE) was accessed (Roberts et al., 2015) (Table 3-2). Clo1313_2273-2275 are predicted to target GATC, and thus are likely responsible for the methylation seen above and the restriction described previously (Guss et al., 2012). Clo1313_2372 is predicted to be a Type IV enzyme, and thus likely explains adverse effect of Dcm methylation previously described (Guss et al., 2012). Clo1313_2405-2406 are not predicted to have an associated restriction enzyme, suggesting that either Clo1313_2366-2367 or Clo1313_0478-0480 is likely responsible for CGACC methylation and restriction.

In vivo methylation of plasmid DNA with expression of Clo1313_0479-0480

In order to experimentally determine the methyltransferase responsible for methylating CGACC, the methyltransferase genes Clo1313_0480, Clo1313_2367, and Clo1313_0479-0480 from *C. thermocellum* were cloned and inserted into the chromosome of a dam^+ dcm^- strain of *E. coli.* To test the functionality of the expressed methyltransferase genes, a restriction enzyme-based assay was devised. Because a fraction of the CGACC sites overlap with the restriction enzyme TaqαI restriction pattern (TCGA), the locations on a plasmid where this overlap occurs will be blocked from TaqαI restriction by CGACC methylation, resulting in a different restriction pattern. Therefore, comparing an unmethylated plasmid digest to a digest of the same plasmid extracted from a methylating strain of *E. coli* should show a distinct and predictable change in the banding pattern. Plasmid pNJ020 has 7 CGACC motifs, including 1 that overlaps with TaqαI, resulting in a distinctive banding pattern to compare methylated and unmethylated plasmid (loss of 1409 bp and 441 bp bands, generation of 1850 bp band; Figure 3-1) and was used for all of the assays.

Initially, only strains expressing Clo1313_0480 and Clo1313_2366 from the *E. coli* chromosome were constructed. Plasmid pNJ020 was transformed into each

Table 3-1 Methylome analysis of *C. thermocellum*

Motif	Position Methylated	Percent Modification Detected
GATC	2	65.1%
CGACC	3	55.4%

Table 3-2 REBASE prediction of restriction systems present in *C. thermocellum* DSM1313

Type of System	Gene Type	Locus Tag	Predicted Sequence
II	DNA Methylase	Clo1313_2273	None
II	Restriction Enzyme	Clo1313_2274	GATC
II	DNA Methylase	Clo1313_2275	GATC
II	Helicase	Clo1313_2366	None
II	DNA Methylase	Clo1313_2367	None
II	DNA Methylase	Clo1313_2405	None
II	DNA Methylase	Clo1313_2406	None
III	Restriction Enzyme	Clo1313_0478	None
III	DNA Methylase	Clo1313_0480	None
IV	HNH Endonuclease	Clo1313_2372	N/A

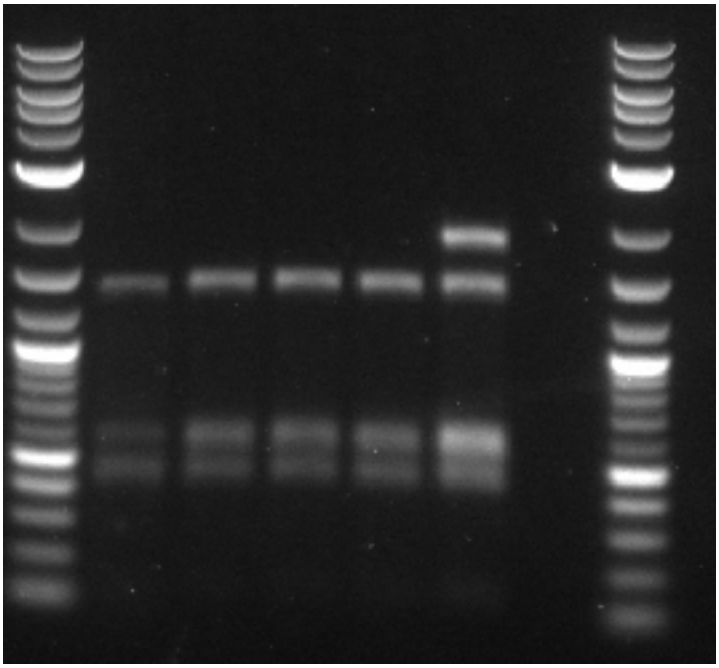

Figure 3-1 Gel image of pNJ020 digested after different methylation conditions

Lane 1: 2-log DNA ladder; Lane 2: pNJ20 from Top 10 dcm-; Lane 3: pNJ020 from Top 10 dam- dcm-; Lane 4: pNJ020 from λ::Clo1313_2367; Lane 5: pNJ020 from λ::Clo1313_0480; Lane 6: pNJ020 from λ::Clo1313_0479-0480 at 45°C; Lane 7: Blank Digest; Lane 8: 2-log DNA ladder

Table 3-3 Results of three separate transformation efficiency experiments

	Experiment	Methylation State	CFU/µg DNA	Fold Improvement over Wild Type
Wild type	1	-	1570	
	2	-	1807	
	3	-	120	
Wild type	1	+	4106	2.6
	2	+	3765	2.1
	3	+	1054	8.8
ΔClo1313_2366	1	-	1811	
	2	-	2711	
	3	-	331	
ΔClo1313_0478	1	-	18,357	11.7
	2	-	15,060	8.3
	3	-	2560	21.3

strain, methyltransferase gene expression was induced, and functionality was tested by TaqαI restriction digest. No plasmid methylation was observed in either strain (Figure 3-1, Lanes 4 and 5). Closer examination of gene Clo1313_0479, the gene located between the Clo1313_0480 methyltransferase and Clo1313_0478 restriction enzyme and annotated as a protein of unknown function, revealed Clo1313_0479 to contain an EVE domain (formerly DUF55). The EVE domain was recently discovered to be a member of the PUA superfamily and is predicted to be involved in RNA binding (Bertonati et al., 2009). We considered the possibility that Clo1313_0479 is instead a DNA-binding domain that confers sequence specificity to the adjacent methyltransferase subunit. Therefore, we co-expressed Clo1313_0479 and Clo1313_0480 from the *E. coli* chromosome and tested methylation of pNJ020. This time, a banding pattern consistent with partial methylation of CGACC was observed (Figure 3-1, Lane 6). Though the expected banding pattern was observable, the banding pattern of the unmethylated plasmid was also present, and initially more prevalent. To improve the ratio of methylated to unmethylated plasmid, the *E. coli* strain was grown at different temperatures in an attempt to achieve better protein folding. When the strain of *E. coli* expressing Clo1313_0479-0480 was grown and induced at 45°C, the resulting banding pattern showed approximately equal quantities of methylated and unmethylated plasmid DNA (Figure 3-1, Lane 6). Other variations, including addition of 0.1 M of the thermoprotectant glycine betaine, lower temperatures, and various induction schemes, did not result in improved methylation.

Methylated plasmid improves transformation efficiency in C. thermocellum

Methylation of CGACC should block restriction by Clo1313_0478 in *C. thermocellum*. Therefore, transformation efficiency was tested using the plasmid DNA partially methylated by Clo1313_0479-0480 from above. Plasmid methylation resulted in a transformation efficiency improvement of 2.1 to 8.8-fold in the wild type strain relative to the control that was not methylated by Clo1313_0479-0480 (Table 3-3).

Deletion of Clo1313_0478 further improves transformation efficiency

Plasmid DNA was only partially methylated by Clo1313_0479-0480, which could explain the relatively small impact on transformation efficiency. To completely eliminate restriction as a barrier to transformation in *C. thermocellum*, we deleted putative restriction enzymes Clo1313_0478 and Clo1313_2366, resulting in strains AG929 and AG930, respectively. Because electrocompetent cells can vary widely in terms of concentration, competency, and ability to recover, the three experiments done to test for efficiencies were done with freshly made electrocompetent cells grown in three separate batches of medium. While *C. thermocellum* ΔClo1313_2366 showed no statistical improvement (p value,

0.6314) in transformation efficiencies over the parent strain, the *C. thermocellum* ΔClo1313_0478 exhibited a 8.3 to 21.3-fold increase in transformation efficiency relative to wild type (Table 3-3).

Discussion

Identifying, and overcoming RM systems is a crucial step in improving the speed and complexity of metabolic engineering strategies, such as in order to increase ethanol yields and titers for industrial application. Here, we have demonstrated that genes Clo1313_0478-Clo1313_0480 encode a Restriction modification system that targets the motif CGACC and has a direct impact on the transformation efficiency of *C. thermocellum.* By increasing transformation efficiency, genetic engineering will become more reliable in *C. thermocellum*. Plasmids will be easier to transform, and as new tools are developed, they will be easier to deploy in strain AG929 (*C. thermocellum ΔClo1313_0478*).

A key step in understanding RM systems both in *C. thermocellum* and other organisms is to characterize the native methylome to identify potential restrictions systems that may be decreasing efficiency. Utilizing methylation data provided by PacBio sequencing has resulted in the improvement of transformation efficiencies in organisms several non-model organisms (Kolek et al., 2016; Monk et al., 2015; Pyne et al., 2014). In *C. thermocellum* DSM1313 specifically, PacBio analysis revealed two active RM systems, one targeting GATC and another targeting CGACC. Because *E. coli* natively methylates GATC, it is not a barrier to transformation; however, CGACC is both common on plasmids and unprotected by *E. coli*. We have achieved higher transformation efficiencies in two ways, each of which were informed by a prior knowledge of the methylome of *C. thermocellum* DSM 1313. First, the necessity of Clo1313_0479 and the functional expression of Clo1313_0479-0480 were enabled by a simple restriction assay of functionality that required knowledge of the methylated DNA sequence. Further, the specific deletion of Clo1313_0478 was informed by the knowledge that one remaining restriction system needed to be eliminated, resulting in a successful increase in transformation efficiency by approximately 10 – 20 fold.

Interestingly, the poorly characterized gene in between Clo1313_0478 and Clo1313_0480 contains an EVE pfam domain (formerly DUF55), which is part of the PUA superfamily and thought to bind RNA (Bertonati et al., 2009). Because expression of Clo1313_0480 alone was not sufficient to methylate DNA in *E. coli*, we considered the possibility that Clo1313_0479 may instead bind DNA and could be part of a DNA methyltransferase complex consisting of Clo1313_0480 and Clo1313_0479. Only by co-expressing Clo1313_0479 and Clo1313_0480 *E. coli* chromosome was *in vivo* methylation of the CGACC motif achieved. While it is common for Type I RM systems to have multiple genes associated with the methylation function, to or knowledge no Type III system has previously been

reported as requiring a two-gene complex. However, expressing only the annotated methytransferase Clo1313_0480 did not exhibit any *in vivo* methylation of plasmid DNA, but expressing both genes on the *E. coli* chromosome resulted in approximately 50% methylation. This two-gene system implies a novel architecture of Type III DNA methyltransferases and allows for a better understanding of RM systems broadly.

Though methylating the plasmid completely would have been ideal for optimal transformation efficiency, attempts at optimization only increased methylation to approximately 50%. Likely, each particular CGACC site on a given plasmid molecule has a 50% likelihood of being methylated, meaning that amongst the population of plasmids present in a plasmid preparation, there is a Poisson distribution of the pNJ020 plasmids containing between zero and seven methylated CGACC sites, with an average of half the sites. Interestingly, the level of methylation increased with the temperature of the *E. coli* culture, suggesting that the major barrier for in vivo methylation for this enzyme is insufficient activity of this thermophilic enzyme at a presumably sub-optimal temperature. Alternate approaches such as in vitro methylation with purified protein or in vivo methylation in another thermophilic host could potentially circumvent this problem in the future.

Conclusion

Using a combination of methylome analysis, heterologous gene expression in *E. coli*, and gene deletions in *C. thermocellum*, we have demonstrated increased transformation efficiency of *C. thermocellum* by avoiding DNA restriction associated with the sequence CGACC. *E. coli* in vivo methylation of CGACC required the presence of both Clo1313_0479 and Clo1313_0480, suggesting a novel architecture for Type III DNA methyltransferases. Restriction was then completely avoided by deleting the restriction enzyme gene Clo1313_0478 from the *C. thermocellum* chromosome, increasing the transformation efficiency by up to 20 fold. This makes strain AG929 the most transformable strain of *C. thermocellum* to date, which will facilitate more rapid and reliable genetic transformation of *C. thermocellum*.

References

Arber, W., Linn, S., 1969. DNA modification and restriction. Annu Rev Biochem. 38**,** 467-500.

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., Caiazza, N. C., 2011. High ethanol titers from cellulose using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microb. 77**,** 8288-8294.

Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W., Kuzin, A. P., Seetharaman, J., Abashidze, M., Ramelot, T. A., Kennedy, M. A., Cort, J. R., Belachew, A., Hunt, J. F., Tong, L., Montelione, G. T., Rost, B., 2009. Structural genomics reveals EVE as a new ASCH/PUA-related domain. Proteins. 75**,** 760-73.

Biswas, R., Zheng, T., Olson, D. G., Lynd, L. R., Guss, A. M., 2015. Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in Clostridium thermocellum. Biotechnology for biofuels. 8**,** 20.

Chung, D., Farkas, J., Westpheling, J., 2013. Overcoming restriction as a barrier to DNA transformation in Caldicellulosiruptor species results in efficient marker replacement. Biotechnology for biofuels. 6**,** 82.

Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Joe Shaw, A., Lynd, L. R., 2013. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. Metabolic engineering. 15**,** 151-8.

Ellis, L. D., Holwerda, E. K., Hogsett, D., Rogers, S., Shao, X., Tschaplinski, T., Thorne, P., Lynd, L. R., 2012. Closing the carbon balance for fermentation by *Clostridium thermocellum* (ATCC 27405). Bioresource technology. 103**,** 293-9.

Guss, A. M., Olson, D. G., Caiazza, N. C., Lynd, L. R., 2012. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum*. Biotechnology for biofuels. 5**,** 30.

Haldimann, A., Wanner, B. L., 2001. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. J Bacteriol. 183**,** 6384-93.

Holwerda, E. K., Thorne, P. G., Olson, D. G., Amador-Noguez, D., Engle, N. L., Tschaplinski, T. J., van Dijken, J. P., Lynd, L. R., 2014. The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading. Biotechnology for biofuels. 7**,** 155.

Kolek, J., Sedlar, K., Provaznik, I., Patakova, P., 2016. Dam and Dcm methylations prevent gene transfer into Clostridium pasteurianum NRRL B-598: development of methods for electrotransformation, conjugation, and sonoporation. Biotechnology for biofuels. 9**,** 1-14.

Liu, G., Ou, H. Y., Wang, T., Li, L., Tan, H., Zhou, X., Rajakumar, K., Deng, Z., He, X., 2010. Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. PLoS Genet. 6**,** e1001253.
Messer, W., Noyer-Weidner, M., 1988. Timing and targeting: the biological functions of Dam methylation in E. coli. Cell. 54**,** 735-7.
Mohr, G., Hong, W., Zhang, J., Cui, G. Z., Yang, Y., Cui, Q., Liu, Y. J., Lambowitz, A. M., 2013. A targetron system for gene targeting in thermophiles and its application in Clostridium thermocellum. PloS one. 8**,** e69032.
Monk, I. R., Tree, J. J., Howden, B. P., Stinear, T. P., Foster, T. J., 2015. Complete Bypass of Restriction Systems for Major Staphylococcus aureus Lineages. MBio. 6**,** e00308-15.
Murray, N. E., 2000. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). Microbiol Mol Biol Rev. 64**,** 412-34.
Ng, T. K., Weimer, T. K., Zeikus, J. G., 1977. Cellulolytic and physiological properties of Clostridium thermocellum. Arch Microbiol. 114**,** 1-7.
Olson, D. G., Lynd, L. R., 2012. Transformation of *Clostridium thermocellum* by electroporation. Method Enzymol. 510**,** 317-30.
Olson, D. G., McBride, J. E., Shaw, A. J., Lynd, L. R., 2012. Recent progress in consolidated bioprocessing. Curr Opin Biotechnol. 23**,** 396-405.
Papanek, B., Biswas, R., Rydzak, T., Guss, A. M., 2015. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in Clostridium thermocellum. Metab Eng. 32**,** 49-54.
Pingoud, A., Jeltsch, A., 1997. Recognition and cleavage of DNA by type-II restriction endonucleases. Eur J Biochem. 246**,** 1-22.
Purdy, D., O'Keeffe, T. A., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., Minton, N. P., 2002. Conjugative transfer of clostridial shuttle vectors from Escherichia coli to Clostridium difficile through circumvention of the restriction barrier. Mol Microbiol. 46**,** 439-52.
Pyne, M. E., Moo-Young, M., Chung, D. A., Chou, C. P., 2014. Expansion of the genetic toolkit for metabolic engineering of Clostridium pasteurianum: chromosomal gene disruption of the endogenous CpaAI restriction enzyme. Biotechnology for biofuels. 7**,** 163.
Raleigh, E. A., Wilson, G., 1986. Escherichia coli K-12 restricts DNA containing 5-methylcytosine. Proc Natl Acad Sci U S A. 83**,** 9070-4.
Rhoads, A., Au, K. F., 2015. PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics. 13**,** 278-289.
Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., Blumenthal, R. M., Degtyarev, S., Dryden, D. T., Dybvig, K., Firman, K., Gromova, E. S., Gumport, R. I., Halford, S. E., Hattman, S., Heitman, J., Hornby, D. P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T. R., Kobayashi, I., Kong, H., Kruger, D. H., Lacks, S.,

Marinus, M. G., Miyahara, M., Morgan, R. D., Murray, N. E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D. N., Reich, N., Repin, V. E., Selker, E. U., Shaw, P. C., Stein, D. C., Stoddard, B. L., Szybalski, W., Trautner, T. A., Van Etten, J. L., Vitor, J. M., Wilson, G. G., Xu, S. Y., 2003. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. 31**,** 1805-12.

Roberts, R. J., Vincze, T., Posfai, J., Macelis, D., 2015. REBASE-a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. 43**,** D298-D299.

Rydzak, T., Lynd, L. R., Guss, A. M., 2015. Elimination of formate production in Clostridium thermocellum. J Ind Microbiol Biotechnol. 42**,** 1263-72.

Tock, M. R., Dryden, D. T., 2005. The biology of restriction and anti-restriction. Curr Opin Microbiol. 8**,** 466-72.

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., 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 Microb. 76**,** 6591-9.

Yasui, K., Kano, Y., Tanaka, K., Watanabe, K., Shimizu-Kadota, M., Yoshikawa, H., Suzuki, T., 2009. Improvement of bacterial transformation efficiency using plasmid artificial modification. Nucleic Acids Res. 37**,** e3.

CONCLUSION

Research Summary

The work completed for this thesis covers many aspects of working with a non-model microorganism for the production of a valuable chemical. My work within the BioEnergy Sciences Center has focused on the organism *Clostridium thermocellum*, but understanding that one system has ranged from metabolic engineering, to characterization, to understanding and optimization of the genetic systems. Metabolic engineering of *C. thermocellum* resulted in the highest ethanol yielding strain to date, extensive characterization of several different strains provided deeper insight into the physiological responses in the cell, and the development of better genetic tools not only improved transformation efficiency in *C. thermocellum*, but will also enable further advances in *C. thermocellum* as a platform Consolidated Bioprocessing strain. Many of the mutant strains of *C. thermocellum* resulting from the work in this thesis are currently some of the most promising strains available, both in terms of ethanol yields and as platform strains for additional engineering efforts.

AG553 is the highest ethanol yielding strain of C. thermocellum on model substrates and pretreated biomass

C. thermocellum AG553, as described in Chapter 1, directed almost all of central metabolism to the production of ethanol. Lactate, formate, acetate, and hydrogen were eliminated or greatly reduced in strain while keeping the cell viable. High ethanol yields were maintained on both the model substrates cellobiose and Avicel as well as the pretreated real world switchgrass and poplar samples. As second generation biofuels are increasingly applied in an industrial setting, biomass as the carbon source is the only economical option. It is crucial in constructing an industrially relevant microorganism that it exhibits strong growth and product yield on biomass, and not just expensive refined laboratory sugars. The ethanol yield reached by AG553 on both switchgrass and poplar makes it an important CBP platform organism.

Aside from its ethanol yields, AG553 showed a dramatically inhibited growth rate. While this was not altogether surprising after genetically eliminating most of its central metabolic pathways, it did limit the rate of ethanol production and therefore limit the industrial relevance of AG553. Improving and understanding the growth phenotype was an important component of the work with *C thermocellum.*

Laboratory evolution of AG553 improved growth rate

Further rational engineering of AG553 proved unsuccessful, likely due to the slow growth and substantially disrupted metabolism of the strain. Therefore,

other approaches were employed for specifically improving the growth rate of *C. thermocellum* AG553. The laboratory evolution of AG553 resulted in numerous strains of *C. thermocellum* that showed improved growth rates while maintaining or increasing final ethanol yields. Genomic resequencing of the evolved strains showed fewer acquired genomic mutations than we anticipated, and none were clear indicators of a mechanism for improved growth. Strain AG601 was selected as a promising strain because of its higher ethanol yield, unique mutation in *spo0A*, as well as a mutation in the *rho* termination factor. We hypothesized that these unique mutations were the most likely cause of improved growth rate as they likely shut down the sporulation process and allowed the cells to grow at a faster rate.

Medium supplementation also improved growth rate

Previous work with *C. thermocellum* showed that formate supplementation improved the growth in *Δpfl* (Rydzak et al., 2015b). With *pfl* deleted in AG553 as well as AG601, it was extrapolated that formate supplementation would also improve the growth rate for these mutant strains. The comparison of wild type, AG553, and AG601 with and without the addition of formate was used to elucidate the mechanisms for improved growth in *C. thermocellum*. Stress response genes were the most highly upregulated in the strains with genetic mutations. Because most of central metabolism was genetically deleted, this stress response was unsurprising. However, we anticipated this stress response was primarily the result of limited biosynthetic pathways. Formate specifically acts as a donor group for tetrahydrofolate, which in turn is a precursor for purine, formylmethionine, methionine, and S-adenosyl-methionine. It was expected that the elimination of formate inhibited these pathways enough to cause the stress response. Interestingly, transcriptomic results indicated that despite formate playing a critical role in C1 metabolism, the addition of formate did not mitigate the stress response in the cell. This implies that the stress response seen in AG553 and AG601 is due to an overall metabolic redox imbalance rather than the absence of necessary precursor metabolites for growth.

Formate medium supplementation provided important insights to future work with *C. thermocellum*. Reduced growth rate caused by formate elimination is likely due to the inhibition of biosynthetic pathways. Keeping biosynthesis balanced in future metabolic engineering attempts will be important for developing industrial strains with sufficient growth rates. Though redox imbalances were not addressed in this work, the stress response seen in the RNA sequencing data also implies that there is more to be understood about redox in *C. thermocellum* and that it too will be important in developing robust industrial strains.

Genetic tool improvement will accelerate progress in C. thermocellum

After the characterization work outlined in chapters 1 and 2 was completed, we hoped to further investigate specific genetic mutations and differentially expressed genes. Genetically manipulating the altered genes in a wild type background, would have allowed for better correlation between observed phenotypes and specific mutations. Unfortunately, the genetic tools for *C. thermocellum* are still too inefficient for the introduction of point mutations or gene complementation. This genetic bottleneck forced us to reassess and improve the available genetic tools for *C. thermocellum*.

Several approaches were explored to improve genetic tools in *C. thermocellum*. The presence of an active Restriction-Methylation (RM) system was the hypothesized cause of the low efficiency. To overcome this active RM system, two separate annotated Restriction Enzymes were deleted from *C. thermocellum's* genome. One of these two strains, ΔClo1313_0478 showed a 20-fold improvement in transformation efficiency supporting the argument that RM systems play a large role in the transformation of foreign DNA. As corroborating evidence for the Clo1313_0478-Clo1313_0480 gene cluster as the active RM system, the methyltransferase gene was cloned onto the *E. coli* chromosome to perform *in vivo* plasmid methylation. Through this process, it was discovered that expressing the annotated methyltransferase (Clo1313_0480) alone did not replicate the expected methylation pattern. When the EVE domain-containing protein Clo1313_0479 was co-expressed with Clo1313_0480, the expected *in vivo* methylation was observed. This system is the first reported Type III methyltransferase that requires the activity of two separate genes for functionality and expands the predicted function of EVE domain proteins to include not RNA binding, but likely DNA binding as well.

The work completed for genetic tools in *C. thermocellum* substantially improved transformation efficiencies. Moving forward, *C. thermocellum ΔClo1313_0478* can act as a platform strain for more efficient strain engineering. Additionally, the concepts and research methods that were applied to *C. thermocellum* can be applied to other microorganisms.

Future application of genetic tools

Microscopic, single-celled organisms have been employed by humanity for thousands of years. While beer, wine and bread are the most recognizable results of this microbiological phenomenon, they are certainly not the only examples. For the last half a century, the beneficial applications of microbial systems have expanded to include the production of antibiotics, herbicides and insecticides, anticancer pharmaceuticals, and commodity chemicals (Rondon et al., 1999). Transitioning to a biology-based economy is the result of increasing pressure to decouple economic growth from environmental destruction and to

improve the sustainability of many fossil fuel manufacturing processes (Sheppard, 2011). The commodities currently produced by microbial systems are a tiny fraction of the beneficial chemicals that microorganisms are capable of producing, but one of the largest hurdles to making the transition more encompassing is humanity's limited knowledge of the microbial universe. Our ability to both culture and manipulate a wide range of microbial hosts for a wide range of metabolic products remains rudimentary.

The insights from *C. thermocellum* are being applied to other non-model microorganisms for the broad expansion of genetic tools and genetic engineering. Consolidated bioprocessing is just one example of a important microbial metabolism for potential industrial use. Utilizing other interesting metabolisms in industrial applications will require advanced genetic tool strategies in many non-model microbial hosts. For the first time, the problem of rudimentary genetic tools in non-model organisms is beginning to be able to be addressed systematically, on a large scale, with a variety of organisms. The goal is not to improve one microorganism for the production of one chemical, but to understand the fundamental barriers to foreign DNA transformation and successful genetic engineering. Understanding these barriers is essential to the ubiquity of genetic engineering of non-model organisms ubiquitous.

We hypothesize that several barriers to successful engineering exist in non-model organisms including *C. thermocellum.* These include: 1. Delivering foreign DNA into the host cell, 2. Circumventing the native immune system of the host organism, which is a network of Restriction Enzymes (REs) designed to cleave foreign DNA, 3. Maintaining the foreign DNA stably once present in the host organism, and 4. Selecting only the cells that maintain and replicate the foreign DNA. While all of these components must be addressed in genetic tool development, we believe that the primary focus of research efforts should be circumventing the native immune systems through *in vivo* methylation and RE deletion and finding ways to maintain foreign DNA in the host organisms. Without these two components, targeted metabolic engineering will not be possible. All of these barriers have been overcome to improve *C. thermocellum* transformation efficiency and genetic engineering techniques.

The initial step to addressing transformation in new organisms is to understand the native RM systems present and active in the organism of interest. This allows for a more systematic and thorough approach than has been previously utilized for microorganisms. In this approach, a combination of genome annotation and advanced sequencing technology allows for the identification of annotated restriction enzymes and methyltransferase genes through the curated Restriction Enzyme Database (REBASE, New England Biolabs, Ipswich, MA). REBASE data is continually expanding because of the ever-increasing affordability of genome sequencing technology. This growing

database provides an excellent starting point for understanding the RM systems present in thousands of bacterial genomes. REBASE will allow us to compile the annotated Restriction Enzyme and Methyltransferase genes for the organisms of interest. Often, however, REBASE does not connect specific DNA recognition to individual genes. REBASE allows us to identify the RM systems that are present, but not the sequences that those RM systems are targeting.

Methyltransferase genes target DNA sequences with three different methylation modifications: 6-methyl Adenine (6mA), 4-methyl Cytosine (4mC) and 5-methyl Cytosine (5mC) (Nelson and McClelland, 1991). Single-molecule, real-time sequencing (SMRT) technology has detected 4-methyl Cytosine and 6-methyl Adenine methylation motifs for several years (Davis et al., 2013; Flusberg et al., 2010). However, 5mC motifs are also frequently found in prokaryotic cells, and have previously been largely overlooked in methylome analyses (Yu et al., 2015). In order to expand understanding of active RM systems, DNA sequences methylated in each organism must be identified. To lay the groundwork for future work, I have already completed methylome analysis for a number of different non-model organisms (Appendix 1). Utilizing both SMRT sequencing and bisulfite treatment, the complete methylomes of the target organisms were identified for the first time allowing for all of the active RM systems to be identified.

Some of the target organisms, such as *C. acetobutylicum* and *M. thermoacetica* have genetic tools and have been previously engineered. However, these organisms still have low transformation efficiencies and inefficient engineering protocols. Similar to *C. thermocellum,* the RE genes can be targets for gene knockouts. For organisms that have not been previously transformed such as *Fibrobacter succinogenes* or *Clostridium clariflavum*, *in vivo* plasmid methylation in *E. coli* will likely be required. By expressing active methyltransferase genes on the *E. coli* chromosome, plasmids can be methylated to mimic the host organism and circumvent the host's REs (Suzuki and Yasui, 2011). This allows for improved transformation using well-established protocols such as electroporation and conjugation (Chen and Dubnau, 2004).

Once a plasmid is introduced to a host organism, plasmid maintenance must be considered (Meacock and Cohen, 1980). Plasmid maintenance requires a functioning origin of replication in each new organism. A collection of different origins of replication will be collected in order to quickly test a variety of parameters. These origins will be sourced from different genuses and species; they will have different mechanisms for replication as well as a variety of optimal temperatures. Isolating cells that contain the plasmid also requires selective markers such as antibiotic sensitivity and a corresponding gene that confers antibiotic resistance. Organisms that have not been engineered previously do not have literature available that identifies antibiotic sensitivity. Therefore, each organism must be tested with multiple antibiotics at multiple concentrations to

identify the Minimum Inhibitory Concentrations (Table A-8). Combining new origins of replication and antibiotic resistance genes will be the basis of an extensive plasmid library. This library will consist of a variety of origins of replication cloned with many antibiotic resistance genes. These numerous plasmids will expedite testing and identification of a functioning plasmid for a new organism in the future.

Multiple aspects of transformation require consideration in order to improve genetic engineering in non-model microorganisms, as discussed in Chapter 3. This hypothesis has been tested and supported by our work in *C. thermocellum*, and it is being applied in numerous other organisms. Moving forward, methylomes from non-model organisms must be reproduced in *E. coli* strains and plasmids must be constructed and identified to contain all of the necessary components for replication. Expanding genetic engineering to more microbial metabolisms will allow for the industrial production of many more chemicals and commodities. Utilizing biological systems and allowing human consumption to transition away from fossil fuel dependence will result in a more equitable and sustainable economy.

References

Chen, I., Dubnau, D., 2004. DNA uptake during bacterial transformation. Nat Rev Micro. 2**,** 241-249.

Davis, B. M., Chao, M. C., Waldor, M. K., 2013. Entering the era of bacterial epigenomics with single molecule real time DNA sequencing. Curr Opin Microbiol. 16**,** 192-8.

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., Turner, S. W., 2010. Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat Methods. 7**,** 461-5.

Meacock, P. A., Cohen, S. N., 1980. Partitioning of bacterial plasmids during cell division: a cis-acting locus that accomplishes stable plasmid inheritance. Cell. 20**,** 529-42.

Nelson, M., McClelland, M., 1991. Site-specific methylation: effect on DNA modification methyltransferases and restriction endonucleases. Nucleic Acids Res. 19 Suppl**,** 2045-71.

Rondon, M. R., Goodman, R. M., Handelsman, J., 1999. The Earth's bounty: assessing and accessing soil microbial diversity. Trends Biotechnol. 17**,** 403-9.

Rydzak, T., Lynd, L. R., Guss, A. M., 2015. Elimination of formate production in Clostridium thermocellum. J Ind Microbiol Biotechnol. 42**,** 1263-72.

Sheppard, A., Gillespie, I, Hirsch M, Begley, C, 2011. Biosecurity and sustainability within the growing global bioeconomy. Current Opinion in Environmental Sustainability. 3**,** 4-10.

Suzuki, T., Yasui, K., 2011. Plasmid artificial modification: a novel method for efficient DNA transfer into bacteria. Methods in molecular biology. 765**,** 309-26.

Yu, M., Ji, L., Neumann, D. A., Chung, D. H., Groom, J., Westpheling, J., He, C., Schmitz, R. J., 2015. Base-resolution detection of N4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite- sequencing. Nucleic Acids Res. 43**,** e148.

APPENDIX

Methylation results for target organims

Table A-1Methylation analysis for *Clostridium beijerinckii* NCIMB 8052

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
CNTAYNNNNNNCTTC	3	m6A	81.9	Cbei_3161

Table A-2 Methylation analysis for *Clostridium acetobutylicum*

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
CAAAAAR	5	m6A	88.2	Cac_2309
CTGAAG	4	m6A	95.9	Cac_1222/ Cac_3535
GASTC	1	m6A	64.4	Cac_3348
DNNHCTGCAGD	8	m6A	49.1	Cac_1222/ Cac_3535
GCNGC	2	m5C	98.8	Cac_1501

Table A-3 Methylation analysis for *Clostridium kluyveri*

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
GATC	2	m6A	74.4	Ckl_2314
CCAAG	4	m6A	71.4	Unknown
GWTAAT	5	m6A	61.5	Unknown
CAAAAAR	6	m6A	56.5	Ckl_3158
CCGG	1	m5C	Unknown	Ckl_2671

Table A-4 Methylation analysis for *Clostridium clariflavum* 4-2A

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
GATC	1	m6A	99.6	Unknown
GACAT	3	m6A	96.4	Unknown
TGAAAG	4	m6A	99.7	Unknown
CAGAAG	4	m6A	99.7	Unknown
GCGATD	3	m6A	96.3	Ccla_2133
CNAYNNNNCTC	2	m6A	98	Unknown
ATGCAT	4	m6A	97.7	Unknown
GAGNNNNNNNRTC	2	m6A	99.6	Ccla_1830-1831

Table A-5 Methylation analysis for *Moorella thermoacetica*

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
SATC	2	m6A	99.8	Moth_1737
AACCA	4	m6A	99.7	Unknown
GAWTC	2	m6A	99.8	Moth_0026
CTCCG	3	m4C	88.3	Unknown
GGGCCC	4	m4C	88.4	Unknown
GAGNNNNNGAT	2	m6A	100	Moth_1672

Table A-6 Methylation analysis for *Fibrobacter succinogenes* S85

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
GCAGG	3	m6A	99.8	Unknown
CACAT	4	m6A	99.6	Unknown
YACAGC	4	m6A	100	Unknown
CTGCAGH	4	m6A	100	Fisuc_2878
GAANNNNNTGC	2	m6A	100	Fisuc_1346/ Fisuc_1722
GACGTC	2	m6A	100	Fisuc_0069
GWCAG	3	m4C	52.3	Unknown
GCTGTAG	5	m6A	93.6	Unknown
GGWCC	4	m5C	100	Fisuc_0930

Table A-7 Methylation analysis for *Glycomyces* MMG10089

Motif	Position Modified	Methylation Type	Percent Modified	Target Gene
CAGCTG	1	m5C		Unknown
TCGANNNNNNNCTC	4 & 13	m6A		Unknown

Table A-8 MIC data for target organisms

	***C. obsidiansis* OB47**	***M. elsdenii* DSM 20460**	***C. clariflavum* 4-2a**	***F. succinigens* S85**	***Cl. kluyveri***
Antibiotic	**MIC (ug/mL)**	**MIC (ug/mL)**	**MIC (ug/mL)**	**MIC (ug/mL)**	**MIC (ug/mL)**
Tet	2	2	2	2	2
Tm	3	4	5	5	2.7
Cm	3	32	3	2	2.7
Erm	16	16	2	2	17.3
Carb	384	32	> 512	48	192
Strep	64	256	> 512	17	42.7
Gent	24	128	48	2	2
Kan	48	128	256	17	2.7

Abbreviations:
Tet: tetracycline
Tm: thiamphenicol
Cm: chloramphenicol
Erm: erythromycin
Carb: carbenicillin
Strep: streptomycin
Gent: gentamicin
Kan: kanamycin

VITA

Beth Papanek is from Woodstock, IL. She completed her Bachelor's of Science in Chemistry and her Professional Science Masters in Bioenergy at the University of Illinois. During her Masters, she worked on the production of butanol as a biofuel alternative using *Clostridium beijerinckii*. She moved to Knoxville, Tennessee in the summer of 2012 when she began her PhD work through the Bredesen Center for Interdisciplinary Research and Graduate Education. After graduation, she will continue to work for the Bredesen Center and Oak Ridge National Laboratory teaching entrepreneurship to young scientists and pursuing a startup company of her own.