EXPLORING METABOLISM IN THE EXTREMELY THERMOPHILIC CELLULOLYTIC BACTERIUM *CALDICELLULOSIRUPTOR BESCII* FOR ENGINEERED BIOPRODUCT FORMATION

by

GABRIEL MICHAEL RUBINSTEIN

(Under the Direction of Michael W. W. Adams)

ABSTRACT

Petroleum is the primary source of liquid transportation fuels and industrial chemicals, such as solvents and plastic precursors. Due to the finite nature of oil and the need to mitigate environmental damage from the combustion of hydrocarbons, renewable methods for producing these chemicals are necessary. One approach is to generate liquid fuels and commodity chemicals from the microbial conversion of lignocellulosic biomass, an abundant and renewable feedstock. Before lignocellulose can be utilized by most microorganisms, extensive pretreatment is necessary to release the monosaccharides for fermentation. The use of biomass degrading organisms for biological conversion can decrease costs associated with pretreatment. *Caldicellulosiruptor bescii* is an extremely thermophilic anaerobic Gram-positive bacterium with the highest optimum growth temperature of any known cellulolytic organism (78 °C). Due to the availability of genetic tools and the unusual capacity to grow on completely untreated lignocellulosic biomass as a sole carbon substrate, *C. bescii* has gained attention as a potential platform organism to produce biofuels and other bioproducts via consolidated bioprocessing. The work described herein examines metabolism in wild type *C. bescii*, strains containing targeted gene deletions, and strains

expressing engineered metabolic pathways. Study of native metabolism revealed the presence of an alternative enzyme for the oxidation of glyceraldehyde-3-phosphate in glycolysis. This enzyme, a new family of tungsten-containing glyceraldehyde-3-phosphate oxidoreductase, was purified and a knockout strain was characterized. Bioinformatic analysis of the genome revealed a multipurpose ATPase (*msmK*) involved in sugar uptake. A deletion strain showed that *msmK* helps transport di- and oligosaccharides but not C_5 and C_6 substrates. A two-gene engineered metabolic pathway for the reduction of carboxylic acids to alcohols was expressed in *C. bescii*, the first instance of this pathway in a cellulolytic organism. Consisting of an aldehyde oxidoreductase and a primary alcohol dehydrogenase, it enabled the reduction of exogenously added organic acids to the corresponding alcohol, with concomitant ethanol generation from acetate. Finally, heterologous expression of a pyruvate oxidoreductase resulted in increased pyruvate-oxidizing activity in cytosolic extracts. Together, this work describes advances in the understanding of redox cofactor balancing, carbohydrate utilization, carbon metabolism, and heterologous gene expression in this biotechnologically-relevant bacterium.

INDEX WORDS: *Caldicellulosiruptor bescii*, biofuels, bioproducts, anaerobe, bacteria, biotechnology, metabolic engineering, thermophile

EXPLORING METABOLISM IN THE EXTREMELY THERMOPHILIC CELLULOLYTIC BACTERIUM *CALDICELLULOSIRUPTOR BESCII* FOR ENGINEERED BIOPRODUCT FORMATION

by

GABRIEL MICHAEL RUBINSTEIN

B.S., Colorado School of Mines, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

EXPLORING METABOLISM IN THE EXTREMELY THERMOPHILIC CELLULOLYTIC BACTERIUM *CALDICELLULOSIRUPTOR BESCII* FOR ENGINEERED BIOPRODUCT FORMATION

by

GABRIEL MICHAEL RUBINSTEIN

Major Professor:	Michael W. W. Adams
Committee:	Jorge C. Escalante-Semerena
	William N. Lanzilotta
	Yajun Yan

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2020

DEDICATIONS

For my grandfather, Leon Israel Rubinstein (1910-1985), whom I never knew, but in whose footsteps I follow. Born in Poland, he fled religious persecution as a child, emigrating to Palestine in 1920. In 1928, at 18 years old, he arrived in the United States. He spent the rest of his life in New York City. Before World War II, Leon was a kosher butcher. After serving during the war, he entered the textile business. While he lacked formal higher education, never attending college, he was considered a scholar by those who knew him. Remarkably well read, he was a polyglot, speaking at least five languages. Leon was a loving father and a doting husband. I am honored to have inherited his pursuits of both knowledge and justice. May his memory be a blessing.

For Dr. Paul D. Ogg (1969-2016), a dedicated father and husband, a passionate brewer, and a gifted professor in the Bioengineering and Life Sciences program at Colorado School of Mines. During my studies at Mines, I took more courses from Paul than from any other faculty member. P-DOgg, as he was affectionately known, always found ways to connect with students; I was no exception. Dr. Ogg was the first teacher who ever told me, “no one knows,” when I pushed him for answers on a signaling pathway in class. “If you want answers,” he said, “go join a research lab and find them yourself.” And so, I did. It wasn’t until after I arrived in graduate school that I could look back to appreciate what I gained from those countless hours of lectures with Dr. Ogg. In one such class, my interest in thermophiles was first piqued, planting the seed that led me to the Adams Lab. I am saddened that P-DOgg did not get the opportunity to read my first publication, or to be here for my defense. I miss his guidance, and I am thankful for all he taught me.

ACKNOWLEDGEMENTS

My sincerest thanks go first to my parents; without their unending and immeasurable support, this work would not have been possible. I owe a great debt to Dr. Matthew Posewitz, who accepted me as an undergraduate researcher in his lab when I approached him after a biochemistry lecture in the fall of my junior year of college. He has supported me from the application process through to the completion of this work. Thank you to my brother Ari, his wife Rachel, and my nephew Miles, who always managed to put a smile on my face, even when research felt overwhelming. My time in Athens was unforgettable, and countless people made this era of my life special. Without the music, I would never have made it. Thank you to my friends and coworkers in the Adams Lab – especially my cubical mate Dr. Amanda Williams-Rhaesa and our lab manager, Farris Poole. Thank you to my committee members, Dr. William Lanzillota, Dr. Yajun Yan, and Dr. Jorge Escalante-Semerena for their guidance as my projects progressed.

Finally, I must thank my advisor, Dr. Michael W.W. Adams for molding my approach to problem solving and providing me with the scientific principles on which to base my future. It is hard to imagine completing my graduate work under anyone else's direction.

Every time I have stood at my lab bench for several years now, I have stared at the following quote, printed on a white piece of paper. I taped it there one day as motivation – a reminder to myself of the scientists who came before me, and hopefully, of those who will follow.

"If we use fuel to get our power, we are living on our capital and exhausting it rapidly. This method is barbarous and wantonly wasteful and will have to be stopped in the interest of coming generations." ~Nikola Tesla, September 9th, 1915

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

The Need for Alternative Energy Sources

Human reliance on fossil fuels for grid power, transportation fuel, and industrial chemicals has resulted in an untenable and unsustainable status quo; society sits at a fulcrum. The necessity of identifying and deploying oil, coal, and natural gas alternatives is primarily two-pronged: fossil fuels are a finite resource and greenhouse gas emissions will cause irreparable harm to both the planet and its inhabitants.

Over the last century, global fossil fuel consumption has increased exponentially (**Figure 1.1**) (1-4). Technological improvements have allowed continuous access to coal, oil, and natural gas in previously inaccessible locations (5, 6). Fossil fuel consumption is vastly outpacing the geological timescales (millions of years) over which these hydrocarbons form (7, 8); these resources are finite and will eventually run out or become prohibitively expensive to extract (9). Peak oil is the theoretical time when global oil production (or demand in some formulations) reaches a maximum, and after which oil production permanently decreases. By some predictions, including two potential scenarios laid out in a 2020 report by British Petroleum, we may have already reached peak oil, while by other estimations we are still several decades away (1, 10). Either way, as a result of this undetermined limit on the supply of fossil fuels, many developed nations have made notable expenditures to push toward creating and deploying alternative energy solutions. Renewable energy sources (herein: renewables) is a blanket term usually encompassing energy generated from photovoltaics (solar energy), hydroelectric plants (sometimes held as a

separate category, but counted as a renewable in this work), biofuels/biomass, geothermal plants, and wind turbines. In France, 72% of electricity is generated from nuclear power (which, while not a fossil fuel, is also a finite resource) and 18% from renewables, while only 9% is generated from fossil fuels (11). In Spain, 35% of electricity is currently generated from renewables, over 50% of which is wind power. Spain generates 22% of their electricity from nuclear reactors, but fossil fuels still account for 40% of the country's electricity (11). Despite these noteworthy efforts, in 2018 64% of electricity globally was generated by fossil fuel combustion, with only 26% of electricity generated from renewables (11)**.** In 2017, China spent $127 billion USD on investments in renewable energy, representing 45% of the global investments in renewables for the year (12). Similarly, in 2015, China became the largest producer of solar power worldwide (1). Despite these investments, China is still burning more coal than any other nation, representing a stunning 52% of global coal consumption in 2019 (1). Coal burning accounted for 66% of Chinese electricity generation in 2016, as opposed to 26% of electricity generated from renewables (primarily hydroelectric) (13). In the United States, fossil fuels accounted for 80% of total energy consumption and 62% of electricity generation in 2019, the majority of which came from petroleum and natural gas, respectively (**Figure 1.2**). As of 2017, government investment in renewables by the United States is roughly 1/3 the levels of Chinese spending ($40 billion vs. $127 billion) (12). Comparing total electricity generation from renewables, China and the United States, generated 1.84 million GWh and 0.74 million GWh, respectively (1). Thus, China is currently investing roughly three times the amount of money in renewables as the United States and producing roughly three times the amount of renewable energy. Generally, electricity generated from renewable sources is trending upwards globally – at 26% of global electricity in 2018, compared with 13% in 2000 (1). Importantly, money spent on renewables currently represents

most new investments in the energy sector. According to the International Renewable Energy Agency (IRENA), close to 2/3 of new electricity generation capacity in 2018 was from renewables (14).

Independent of the finite nature of fossil fuels, the greenhouse gas emissions from the extraction, refinement, and burning of these hydrocarbons are having demonstrably adverse effects (15-17). Methane and carbon dioxide in the atmosphere trap radiation from the sun that has reflected from the Earth's surface. Trapped ultraviolet radiation results in increasing global temperatures, known as the greenhouse effect. Combustion of hydrocarbons, à la the burning of fossil fuels, releases water and carbon dioxide, and is informally thought of as the primary human contribution to climate change. In the last decade, coal production has decreased globally, except in Asia Pacific nations (e.g. China, Indonesia, Bangladesh, New Zealand, Vietnam), while natural gas (primarily composed of methane) production has continued to increase in all regions (1). Although burning methane releases roughly 50% less carbon dioxide than burning coal, the methane emissions and unintentional release that result from extraction are difficult to quantify (18-20). Methane is at least 20 (and potentially up to 90) times more potent as a greenhouse gas than carbon dioxide, and thus the difference in carbon dioxide release by combustion between natural gas and coal may be largely neutralized by the inadvertent release of methane during extraction (21-24).

Ice cores collected from the arctic can be used to determine historical atmospheric concentrations of methane and carbon dioxide (25, 26). While these levels are indeed cyclical, often fluctuating in association with Milankovitch cycles (regular changes in obliquity, axial procession, and eccentricity of the Earth's orbit), the increases observed since the industrial revolution are significantly outside the scale of previously measured variability (27-29). The broad

consensus among scientists is that greenhouse gas emissions resulting from combustion, production, and refinement of coal, natural gas, and petroleum have led to anthropogenic climate change (15).

Anthropogenic climate change has resulted in increasing global temperatures and shifting weather patterns, with nine of the ten warmest years on record occurring since 2005 (30). Increased global temperatures can result directly in melting icecaps, which facilitates sea level rise and puts coastal communities at risk worldwide (15). Increased concentrations of carbon dioxide in the atmosphere combined with higher global temperatures (favoring gas solubility) have resulted in the acidification of oceans (31). This decrease in oceanic pH has led to a measured decrease in oceanic diversity (31-33). Likewise, seasonal weather-driven events such as wildfires and tropical cyclones (hurricanes and typhoons) are increasing in both number and severity in correlation with rising temperatures (34-38). The impacts of climate change can operate in positive feedback loops, e.g. increasing temperatures result in thawing of the permafrost, which releases large trapped methane pockets, further increasing the temperature (39). The "clathrate gun" hypothesis is an explanation for the periods of rapid temperature increases observed in the Earth's past. According to this hypothesis, rapid increases in temperature, and thus pH of the ocean, associated with large scale extinction events (e.g. the Permian—Triassic extinction event and the Paleocene—Eocene Thermal Maximum) were the result of massive releases of methane from natural gas clathrates under the permafrost and along coastal shelves (40, 41). The thawing of these clathrates acted to rapidly increase temperature by the positive feedback cycle described above. In yet another potential risk, it has been hypothesized that thawing permafrost opens the possibility for dormant pathogens to reenter circulation (42, 43).

Taken together, the results of anthropogenic climate change provide an additional rationalization for developing non-hydrocarbon energy infrastructure; even if coal, petroleum, and natural gas were not finite resources, their unrestricted use is having marked negative impacts that will lead to at minimum a decreased quality of life for inhabitants of the planet, and at worst pose an existential threat to the species. Identifying and deploying alternative energy solutions is a prerequisite for human progress in the coming decades.

The Need for Alternative Transportation Fuels

Although renewable energy resources such as solar, hydroelectric, and wind power discussed above show remarkable promise, they primarily address grid energy demands (electricity). Barring the development of superior battery or fuel cell technologies, these energy sources leave the need for transportation fuels largely unresolved. In the United States, roughly 28% of total energy use is from the transportation sector, closely matching the 25% of total global energy consumption by transportation (44, 45). As oil production slows, or even declines, and energy needs continue to rise, alternative fuel sources that can help fill the looming gap between supply and demand will become vital. Despite improvements in photovoltaics, wind, and other renewables for grid energy, developing liquid transportation fuel alternatives remains a necessity.

The most common transportation fuels are gasoline, diesel, and jet fuel (45). While differing in both production and chemical composition, each of these liquid fuels is derived from crude oil. Each of these fuels can vary in chemical composition depending on the source of crude oil, the methods of refinement, and the downstream processing (addition of stabilizing agents, etc.) (46). Still, even with minor variation in composition, the properties of these mixtures are roughly consistent, especially in terms of energy density (47). A comparison of the properties of different liquid fuels is presented in **Table 1.1**. Gasoline and jet fuel have similar energy densities (MJ/L),

while diesel fuel is more energy dense. The specific energy (MJ/kg) of gasoline, diesel, and jet fuel are all similar. Although combustion of diesel fuel releases more carbon dioxide than combustion of gasoline, diesel engines are more efficient than gasoline burning engines, and so the overall emissions from diesel automobiles tend to be lower than their gasoline counterparts (48).

Biofuels are liquid fuels produced from biological systems, rather than from crude oil. The most notable biofuels in the global energy market are ethanol and biodiesel. In 2019, 29 billion gallons of ethanol were produced globally, primarily by the United States and Brazil. Roughly 11 billion gallons of biodiesel were produced in 2019, with Indonesia leading even the United States and Brazil (1). Ethanol has a significantly lower energy density than the fuels discussed above; roughly 1.3 gallons of ethanol is required to obtain the same energy as 1 gallon of gasoline (47). The properties of biodiesel are ostensibly the same as the properties of diesel derived from crude oil, albeit with slightly higher average carbon chain length and slightly lower energy density and specific energy (**Table 1.1**). This means that biodiesel can be blended with traditional crude oil derived diesel at any ratio, or even used as an unblended drop-in replacement. Unlike the relationship between biodiesel and diesel, ethanol is not a direct drop-in replacement for gasoline because traditional gasoline engines cannot function on pure ethanol. Additionally, ethanol is corrosive to some materials on its own, but also hygroscopic, and thus the water absorbed is corrosive to gasoline infrastructure (49). As a result, most ethanol is blended into gasoline as an additive at well-defined ratios. In 2005, to push the biofuels industry forward, the United States enacted the Renewable Fuel Standard, requiring that transportation fuel contain a prescribed minimum volume of biofuel (50). As a result, almost all gasoline in the United States is now E10, meaning that it contains 10% ethanol by volume (51). E15 ethanol (15% ethanol by volume) has

recently become more widely available in the United States, although its use is limited to more modern vehicles. In the past decade, there has been a sharp rise in Flexible Fuel Vehicles (FFVs) globally, which are designed to operate on any blend ranging from 100% ethanol to 100% gasoline (52). In the United States, FFVs are optimized to operate with E85 (51-83% ethanol), but can readily operate with the national standard E10 (53). As of 2018, major car manufacturers (Ford, GM, Volkswagen, Volvo, etc.) have put more than 60 million FFVs into operation globally, most of which are in Brazil and the United States – the primary producers of ethanol (53-55).

Like other industrial chemical syntheses, the generation of alternative liquid fuels for transportation requires a significant feedstock of some raw material that can then be converted to the desired product by physical, chemical, or biological processes. From the standpoint of carbon emissions, biofuels offer the potential for "carbon neutral" transportation fuels (56, 57). This is because organisms such as plants and algae utilized as feedstocks for biofuels grow by fixing carbon dioxide from the atmosphere to generate biomass. Thus, when biofuels are burned the carbon dioxide that was fixed by these plants is released back into the atmosphere; when a fossil fuel is burned, the carbon dioxide that is released had been previously sequestered for millions of years (8).

The primary feedstocks for biodiesel are vegetable oils, such as soybean oil (58). Soybeans are processed by separating the oil from the meal, which is used as both a food additive and animal feed. The oil, consisting primarily of palmitic acid (16:0), oleic acid (18:1), stearic acid (18:0), linolenic acid (18:3) and linoleic acid (18:2), is converted to a mixture of fatty acid methyl esters (FAMEs) by transesterification, to yield diesel fuel that can be used as-is or blended with traditional oil-derived diesel (59, 60). In addition to vegetable oils, other feedstocks for biodiesel include inedible tallow (rendered animal fat – a byproduct of the meatpacking industry), waste

cooking oils (a byproduct from many processed food plants), and algae (58). Of these, algae hold the strongest prospects as a biodiesel feedstock (58, 61).

Photosynthetic algae from a variety of phylogenetic lineages, such as *Chlamydomonas* and *Nanochloropsis* store excess energy as lipids in the form of triacyl glycerides (TAGs), which are readily converted to biodiesel with high efficiency (62, 63). Algae are a tempting feedstock because they can be grown to high density in outdoor ponds, especially on marginal land, and thus do not compete with land or resources necessary for traditional agriculture (61). This growth requires little investment for energy or nutrient input, as photosynthetic growth relies primarily on input from the sun and fixing atmospheric carbon dioxide. Genetic modifications of these algae have resulted in even higher lipid, and thus FAME yields. For example, the starchless mutant strains of *Chlamydomonas reinhardtii*, *sta6* and *sta7-10*, produce lipids at almost two-fold the levels of the parent strain, as measured by lipid conversion to FAMEs (64). Unfortunately, growth of genetically modified algae in outdoor ponds poses several problems, while indoor farming of these strains hinders the energy advantages of relying on natural sunlight (65). There is large reluctance to farm genetically modified algae in open ponds for the same reasons there is reluctance to utilize many genetically modified organisms on an industrial scale: without the capacity for sterilization, there is a possibility for escape of the modifications into the environment (66). Ironically, this risk is mitigated by the more likely factor inhibiting the use of modified algae in outdoor ponds: modified algae are often hampered by fitness defects that leave large scale cultures open to costly and problematic contaminations or loss of the desired strain in a mixed culture (67, 68). Indeed, risk of contamination during growth in outdoor ponds is currently the most significant hindrance to employing largescale algal biofuel product (67).

Several different feedstocks are utilized for sugar-based ethanol production, including sugar beets, corn, and sugarcane (69). The use of these feedstocks varies regionally, depending on the climate and thus price and availability of the crop in question. In Brazil and Thailand, ethanol is produced from sugarcane, while in the United States corn is the most readily available source of inexpensive starch. Bioethanol production from the fermentation of sugar-crops is a well-developed and mature global industry. In a generic sugar-ethanol plant, the basic workflow is standard (**Figure 1.3A**) (70). Briefly, high-sugar content feedstocks are milled into a flour and mixed with water to form a slurry. Enzymes like amylases are added to the slurry to break down polysaccharides such as starch into monosaccharides. Ammonia is added (which will provide a nitrogen source for yeast during fermentation) and the pH is maintained around 5.5 at 85 °C for several hours to increase saccharification (hydrolysis of polysaccharides to monosaccharides) and decrease the risk of bacterial contamination. Once this mash is cooled to 34 °C, the pH is adjusted even further down to 5 to inhibit bacterial growth during fermentation, additional saccharolytic enzymes, antibiotics, and urea are added, and finally *Saccharomyces cerevisiae* is added to initiate fermentation. Carbon dioxide produced by the yeast during fermentation is also captured for further industrial uses, such as beverage carbonation or dry ice production (71). The ethanol produced from fermentation is purified downstream by distillation and dehydration, and the remnants of the mash are processed into animal feed. Targets for economically viable ethanol fermentation processes are titers over 42 g/L (912 mM) at a rate of greater than 1 g/L/h (72). Below this titer, the energy input required to recover the ethanol by distillation is too high, as the energy costs required for distillation of ethanol dramatically drop once the mixture contains more than 4% w/w ethanol (73).

Improvements over the last 20 years have left the global bioethanol industry in better position in terms of efficiency and thus profitability. Specifically, the industry has switched almost entirely to the dry-milling process described above, compared with less efficient wet-milling methods used previously (74). The industry has also benefitted from decreases in recovery costs of ethanol, as molecular sieves have become the preferred way to dehydrate the final product of fermentation (75). These modest improvements can be seen by the average gallons of ethanol produced per bushel of corn; this metric increased by roughly 7%, from 2.64 in 2001 to 2.83 in 2016 (76, 77). Despite these improvements and the size of the sugar-based bioethanol industry, there are notable problems with using food-based feedstocks. Producing ethanol from edible sources competes with the use of the necessary farming land and the supply of the crop in question. Increased demand combined with this competition is theorized to drive prices of these crops up in the absence of government intervention, but this is a source of significant debate (78-82). Regardless of the overall impact of ethanol manufacturing on food crop prices and arable land, an ideal feedstock for liquid biofuel generation should not compete with food production (83-85). Excluding algal biodiesel, biofuels described thus far (sugar-based bioethanol and plant oil-based biodiesel) are collectively known as first generation biofuels. Second generation biofuels are those that rely on non-food biomass feedstocks. Plausible feedstocks for second generation biofuels include non-food crops, municipal solid waste, and food production waste streams (86). For liquid fuel (e.g. ethanol) production, non-food crops show the most promise, although downstream processing and biological conversion of municipal solid waste has shown significant potential for the production of methane, organic acid mixtures, and hydrogen (87-89).

Lignocellulose as an Energy Feedstock

Inedible plant biomass (herein: lignocellulose) presents an appealing feedstock because it is inexpensive, abundant, and renewable. Cellulose is an unbranched polysaccharide of linear β-1,4 linked D-glucose and the primary component of plant biomass. Cellulose is the most abundant polymer on the planet, and thus its repurposing as a material for the generation of fuels makes intuitive sense (90). Furthermore, plant feedstocks can be specialized for the location where the plant is being grown; native plant species could be utilized in different climates to optimize crop yields (91, 92). Lignocellulosic biomass could be generated as either a dedicated crop, such as perennial grasses (e.g. switchgrass, miscanthus) or fast-growing trees (e.g. poplar, eucalyptus), or from current waste streams that go unutilized (93, 94). Examples of currently unutilized waste lignocellulose are the stover and husks that remain in the field after corn and rice harvest, or the bagasse produced when sugar-rich pulp is extracted from sugarcane for ethanol production. Currently, bagasse is usually burned to produce electricity (95). Unlike sugar-ethanol fermentation, the production of lignocellulosic ethanol is challenging and requires numerous steps (96-98). Of the 15.8 billion gallons of ethanol produced in the United States in 2019, 418 million gallons of biofuel were mandated to be produced from lignocellulosic feedstocks, with the remainder coming from corn (99). However, the 418 million gallon mandate includes lignocellulosic "biogas" generated from anaerobic digesters or gasification (100); realistically, closer to a paltry 10 million gallons of cellulosic ethanol were produced in the US 2017 (99). Lignocellulosic ethanol production is broken into multiple steps, which are generally defined as pretreatment, saccharification, and fermentation. Fermentation can be further subdivided into hexose and pentose fermentation, which are often conducted separately (**Figure 1.3B**).

While cellulose is the primary component of lignocellulosic biomass, a variety of other polymers and decorations are essential to the structure. The other most abundant components of biomass are hemicellulose and lignin (93). Hemicellulose encompasses a broad range of β-1,4 linked polysaccharides, such as xylan, glucuronoxylan, arabinoxylan, and glucomannan (101). While hemicelluloses are structurally diverse, with significant branching, covalent modifications, and sugar substitutions, they commonly consist of pentose sugars such as xylose and arabinose (101). Lignin is a polymer composed of crosslinked phenolic monomers termed "monolignols" (102). Like hemicelluloses, lignin structures are also amorphous due to the ability to form cross-linked non-linear web-like structures conferred by the multiple free -OH groups present on monolignols (103). Together, cellulose, hemicellulose, and lignin generally make up greater than 70% of plant cell walls – the primary component of lignocellulosic biomass. The ratio of these three biopolymers can vary greatly depending on the type of plant, with dicots for example having lower lignin content and softwoods having higher lignin content (93).

Understanding the relationships between cellulose, hemicellulose, and lignin in the plant cell wall is essential to understanding how to utilize lignocellulose as a chemical feedstock. While the content and structure of plant cell walls are extremely diverse, the general structure can be visualized as microfibrils of cellulose scaffolds interwoven with hemicellulose, with lignin filling in the open spaces (93, 104-106). Microfibrils are crystalline arrangements of approximately 24 cellulose chains held in place by a combination of hydrogen bonding and Van der Waals forces (107). The hemicellulose chains also interact with the microfibrils by hydrogen bonding (108). Both lignin and hemicelluloses are amorphous when compared to the crystalline nature of cellulose microfibrils.

The expenses associated with lignocellulosic ethanol production result primarily from biomass recalcitrance – i.e., the difficulty associated with breaking down the comingled chains of complex biopolymers into readily fermentable mono- or disaccharides (109). This is in stark contrast to the readily digestible starches in sugar-ethanol feedstocks. Decades of research have been spent developing methods for overcoming biomass recalcitrance, which is largely viewed as the single greatest hindrance to the realization of widescale lignocellulosic biofuel production (110). Recalcitrance is overcome through a combination of methods, relying on intensive treatment of the biomass before it is fermented (111, 112). Lignocellulose is significantly hardier than the starch that makes up the majority of sugarcane pulp and corn flour, and thus more stringent treatment methods are required to achieve saccharification. These methods generally consist of some mixture of chemical, thermochemical, and physical treatment (113, 114). It is necessary to first pretreat lignocellulose before enzymatic hydrolysis of the polysaccharides is cost-effective. A primary goal of pretreatment is to physically break apart and fractionate the cellulose, hemicellulose, and lignin. Before other pretreatment, biomass is often physically broken down by mechanical treatments, such as milling, to increase surface area (115, 116).

The most common forms of pretreatment are dilute acid, alkaline, organic solvent ("organosolv"), hot water, steam explosion, and ionic liquid (114, 117). Often, several of these methods are combined. For instance, most organosolv treatments also make simultaneous use of either high or low pH and high temperature to improve the biomass breakdown. Importantly, different pretreatment methods are associated with the solubilization or degradation of different fractions of the biomass, and thus hold potential appeal for different downstream processing of treated material (114, 118). The yeast commonly used for fermentation to produce lignocellulosic ethanol are unable to ferment pentoses, and lignin is not a suitable growth substrate for any well-

established anaerobic fermentative microorganism. Therefore, it is often desirable to utilize pretreatment that will separate hemicellulose and lignin from the cellulose fraction. While separating hemicellulose is often desirable, such as by solubilizing this fraction while the cellulose fraction remains solid, these pentoses can be fermented separately downstream by different organisms, and potentially to non-ethanol products. For downstream fermentation, care must be taken during pretreatment to avoid the formation of furans such as furfural and 5-hydroxymethylfurfural (5-HMF) formed by dehydration, as these compounds are well known inhibitors of microbial growth (114, 119). In other cases, furfural and 5-HMF may be targeted as final products of the hemicellulose fraction, as they are both desirable chemical feedstocks for numerous industrial processes (120, 121). Delignification is a desirable quality in most pretreatments for several reasons: degradation products of lignin can be inhibitory to fermentative organisms downstream and the presence of lignin decreases the efficacy of saccharolytic enzymes both by adsorbing the enzymes and by sterically decreasing access to polysaccharide chains (118, 122). Although lignin has been extensively studied for potential utilization as a material for downstream production of value-added chemicals, which could someday play a role in a more robust bio-refinery platform, most of the lignin separated by pretreatment is currently burned to generate electricity (123-125).

High temperature organosolv pretreatments that utilize either high or low pH are some of the most common approaches for fractionation (126, 127). In one recent study, alkali-catalyzed organosolv pretreatment (ACOS) of cornstalks at 110 °C in 60% aqueous ethanol with 4% w/w sodium hydroxide catalyst resulted in removal of >80% of the total lignin, while recovering 85% of the cellulose and 82% of the hemicellulose as a joint fraction (128). In co-solvent-enhanced lignocellulose fractionation (CELF), tetrahydrofuran (THF) is used as a co-solvent 1:1 with water

at high temperature (150 °C) (129). Like ACOS, CELF has demonstrated high levels of delignification of solids and final recovery of up to 95% sugars (xylans, glucans, arabinans) after saccharification with low enzyme loading (2 mg/g glucan) (130, 131). While originally developed for use in conjunction with dilute acid pretreatment (~0.5% w/w sulfuric acid), more recent work has also demonstrated that CELF is similarly efficient without pH adjustment (132). In addition to these thermochemical methods, some pretreatments make use of physical disruption of biomass. In Ammonia Fiber Expansion (AFEX™) treatment, biomass is first treated with liquid ammonia under high pressure (20 atm), and the release of pressure causes expansion and thus separation between the macromolecular fiber structures (133). AFEX™ pretreatment is usually conducted at temperatures from 100-120 °C, and the ammonia is readily recovered and reused after treatment. High monosaccharide yields are recovered from lignocellulose after AFEX™; treatment of corn stover followed by enzymatic hydrolysis can recover >85% of glucose and xylose (134). Ionic liquids (ILs) are another potential class of solvents for biomass pretreatment, which hold promise due to their ability to completely dissolve biomass (135). Afterwards, cellulose can be precipitated from the solvent by use of an anti-solvent, but recovery of the IL, necessary due to the high price of producing these compounds, is often involved and difficult (114). Because ILs are a large class of compounds, their properties are tunable and dozens of ILs have been explored as pretreatment options, with some capable of solubilizing up to 10% w/v cellulose; recovered cellulose exhibits decreased crystallinity and increased susceptibility to enzymatic hydrolysis (114, 136, 137). These examples demonstrate that multiple excellent pretreatment options exist, and the focus moving forward is to tune these methods to minimize costs without losing efficacy.

Once biomass has been fractionated and the desired sugars, e.g. glucans, have been delignified, saccharification is accomplished by the addition of exogenous enzymes, such as fungal

cellulases (138). The addition of exogenous enzymes is, like pretreatment, a source of exorbitant costs in the production of lignocellulosic bioproducts (139-141). Note that the costs of saccharification are inversely correlated with the efficacy of pretreatment, where efficient pretreatment decreases cost by lowering the necessary enzyme loadings (142, 143). Usually, enzyme loadings are either measured by mg protein/g glucan or Filter Paper Units (FPU)/g glucan, where FPU is a measure of activity rather than protein content. Common enzyme loadings range from 2-30 mg/g substrate (glucan/cellulose/xylan/solids), but these measures can be difficult to compare, because different commercial or homebrewed preparations of enzyme cocktail contain different ratios of the relevant proteins – demonstrating the value of utilizing FPU rather than mg protein when comparing loadings (143, 144). Saccharification is most often accomplished by the addition of commercial cellulase cocktails containing mixtures of cellulases, hemicellulases (xylanases), and β-glucosidases (138).

The primary source of cellulases and hemicellulases for industry is from the fungus *Trichoderma reesei*. Random mutagenesis and directed evolution have yielded hyper-cellulolytic strains of *T. reesei* that secrete the desired enzymes in high titers (>100 g/L) (145). One current limit to cellulase production is the need for growth on cellulase-inducing sugars, such as lactose (146). Despite extensive research, the exact mechanisms of *T. reesei* cellulase-induction on different substrates are still not wholly understood (147). Even though strain engineering has allowed the generation of strains that can produce cellulases without the canonically necessary inducing-sugars, growth on cellulase-inducing biomass is still the preferred option. In an ideal biorefinary setting, enzyme production could occur onsite by growth of *T. reesei* on feedstock derived from the primary raw biomass source (147). Indeed, co-location of enzyme production is a key assumption is many techno-economic lifecycle analyses of lignocellulosic ethanol plants,

with speculated cost reductions of 5-25% compared to off-site enzyme production (147). To date, strains have been engineered that can specifically produce high titers of biomass degrading enzymes at low cost from feedstocks of soybean husks, corn stover, coconut mesocarp, and sugarcane molasses, among others (148-151).

When enzymatic saccharification and fermentation are conducted as sequential steps, the process is termed separate hydrolysis and fermentation (SHF), and this separation allows each step to be conducted under optimized conditions (152). Alternatively, these processes can be physically and temporally combined in a simultaneous saccharification and fermentation (SSF) process, which has gained ground as an alternative (153). Often, the more steps and vessels are involved in an industrial process, the more expensive the process will be. In an SSF process, substrate loading of at least 20% w/w is necessary to achieve the essential benchmark of >4% w/w ethanol from fermentation (152). Numerous experimental studies have demonstrated that when similar or identical pretreatment methods are employed, SSF processes were able to generate more ethanol than SHF, more quickly, and with lower enzyme loadings, implying that SHF processes are losing relevance (152-154). As alternative organisms to *S. cerevisiae* are being explored for bioproduct fermentation, it should be noted that some SSF processes now utilize the pentose fraction of the biomass in combination with the cellulose-enriched solids to increase total sugar utilization and thus overall biomass conversion; these processes are known as simultaneous saccharification and co-fermentation (SSCF) (155, 156). When SS(C)F is not employed, by the time lignocellulose has undergone pretreatment and saccharification, the fermentation, generally occurs as it would for sugar-based ethanol production: fermentation by *S. cerevisiae* produces ethanol from solubilized glucans. SHF, SSF, and SSCF processes are presented in **Figure 1.3B**.

Consolidated Bioprocessing

As discussed, most costs (beyond raw feedstocks) associated with lignocellulosic biofuel production can be attributed to pretreatment and enzymatic saccharification of biomass (113). Specifically, the production costs for the exogenously added enzymes is extreme (113). Consolidated Bioprocessing (CBP) could potentially cut costs in cellulosic biofuel production by reducing or eliminating reliance on expensive exogenous enzymes (157). In CBP, enzyme production, enzymatic saccharification, and fermentation are all combined into a single simultaneous process (**Figure 1.3C**) (158). In some variations of CBP, minimal pretreatment is utilized, further increasing potential savings.

Cellulolytic organisms can break down crystalline cellulose to utilize this polysaccharide as a sole carbon source. To date, organisms that are natively both highly cellulolytic and able to ferment sugars to desirable compounds, e.g. ethanol, at necessary yields are not known (159). As a result, there are two approaches to obtaining a microbe suitable for CBP (157). In the first strategy, the recombinant approach, an organism that can ferment a desirable product at high yields is engineered to express biomass degrading enzymes. In the second strategy, the native approach, a microbe that is already sufficiently cellulolytic is engineered to produce desirable end-products at sufficient titers. While the properties necessary for a successful CBP microbe vary depending on the target-product and feedstock, ideal CBP microbes should be able to grow to high density on high substrate loadings, utilize both hexose and pentose sugars, be resistant to known inhibitory compounds produced during pretreatment or biomass degradation in culture, and be resistant to end-product inhibition of fermentation and growth (157-161). To date, neither of the two CBP strategies has yielded an industrially relevant success, although significant improvements have been achieved by both approaches.

For the recombinant approach, the most tempting targets for producing a cellulolytic organism through heterologous expression are organisms where high yields of target chemicals have already been produced from simple sugar feedstocks and genetic systems are available (157). Examples of such targets include *S. cerevisiae*, which has been demonstrated to produce up to 160 g/L ethanol, the gram-negative bacterium *Zymomonas mobilis*, which can produce titers up to 130 g/L ethanol, and *E. coli*, which has been demonstrated to produce high titers of non-ethanol products, such as isopropanol (140 g/L) and butanol (30 g/L) (160, 162, 163). Biomass degradation is a difficult phenotype to transfer, as the exoproteomes of biomass degrading organisms are complicated, and no necessary minimum group of genes has been identified.

For the breakdown of crystalline cellulose, three different classes of enzyme are usually required: endoglucanases that hydrolyze β-1,4 linkages in internal amorphous regions of cellulose, exoglucanases that hydrolyze β-1,4 linkages to release glucose monosaccharides or cellobiose disaccharides processively from the ends of cellulose chains, and β-glucosidases that hydrolyze glucose from cellobiose or cellodextrins released by the exo- and endoglucanases (164). Cellulolytic fungi secrete multiple free enzymes with each of these activities, which act synergistically with both each other and a suite of other glycoside hydrolases (GHs) (125). Like cellulases, hemicellulases (xylanases) with both exo- and endo- activities, and xylosidases are well characterized and are necessary to solubilize lignocellulosic biomass rather than just crystalline cellulose (165).

Expression of enzymes from biomass degrading organisms in *S. cerevisiae* and *Z. mobilis* has been moderately successful, although efficiency of biomass breakdown does not approach that of native cellulolytic organisms or use of exogenous enzymes (166, 167). In one example, a strain of *Z. mobilis* that natively expresses an endoglucanase (zmo1086 CelA) was discovered, and this

gene was over-expressed in combination with expression of an endoxylanase from an uncultured bacterium attained from metagenome data. The resulting strain of *Z. mobilis* was able to secrete both the cellulase and xylanase, hydrolyze xylan and carboxymethyl cellulose, and produce 3 g/L ethanol from pretreated rice straw (168). In *S. cerevisiae*, strains expressing GHs have also been used to produce ethanol from pretreated biomass with either reduced or no enzyme loading, albeit at titers that usually do not approach the prerequisite 40 g/L ethanol (159). In one strain expressing endoglucanase, β-glucosidase, and cellobiohydrolase (Cel7A, exoglucanase) from *T. reesei*, growth on hot-water pretreated rice straw enabled the production of 7.5 g/L ethanol (169). In the best effort to date, an engineered strain of *S. cerevisiae* expressing cell-surface bound endoglucanase, cellobiohydrolase, and β-glucosidase fermented high loads (200 g/L) of IL-pretreated rice straw to a titer of 43 g/L ethanol, although the addition of 10 FPU/g biomass exogenous cellulase was still necessary (170). Similarly, a strain of *S. cerevisiae* expressing endoglucanase and β-glucosidase produced 35 g/L ethanol from pretreated corn cob, but exogenous commercial cellulase was again necessary to achieve this titer (171). Despite these advancements, no "true" industrial CBP process exists utilizing recombinant yeast, but recent work has shown that heterologous enzyme expression can significantly decrease enzyme loading (161, 172). A major difficulty for an industrial CBP strain (either bacteria or yeast) is achieving high levels of enzyme production without sacrificing ethanol production yield, titer, and rate (173).

S. cerevisiae and *Z. mobilis* do not naturally metabolize C_5 sugars, and *E. coli* selectively consumes C_6 sugars before C_5 sugars due to transcriptional regulation termed carbon catabolite repression (CCR) (174-176). Through metabolic engineering, *S. cerevisiae* and *Z. mobilis* have both been engineered to metabolize pentose sugars (175, 177). Cofactor imbalances inherent to the pathway are observed in *S. cerevisiae* (although ethanol titers >30 g/L have been observed from

growth on mixed xylose/glucose substrates), but *Z. mobilis*, where a balanced pathway has been expressed, shows promise on this front (175, 178, 179). Strains of *Z. mobilis* heterologously expressing operons for both xylose assimilation and the pentose phosphate pathway from *E. coli* can grow on xylose as a sole carbon source (177). Growth of one strain on a mixture of xylose and glucose resulted in 55 g/L ethanol, compared with 25 g/L on xylose alone and 30 g/L on glucose alone (180). Further engineering has allowed the metabolism of arabinose in addition to xylose, utilizing expression of a 5 gene operon from *E. coli* (181). Likewise, *E. coli* has been engineered to overcome CCR, and co-utilization of C_6 and C_5 sugars has been demonstrated (182). Pentose metabolism has been effectively engineered in organisms which cannot natively metabolize C_5 sugars, marking an important step for efficient fermentation of lignocellulose. Importantly, even in organisms that do not naturally utilize C_5 sugars, effects of CCR can be seen when pentose utilization pathways are introduced – posing further potential issues with true co-utilization (183).

While engineering non-cellulolytic organisms to degrade biomass is challenging, engineering cellulolytic organisms to produce the desired end-products at viable titers also poses difficulties (157). Among these issues, the most notable are the establishment of sufficient systems for genetic modification, producing relevantly high titers of non-native products in engineered strains, overcoming potential product toxicity, and achieving elevated sugar consumption for biological conversion (158, 159). Cellulolytic organisms mostly consist of fungi and bacteria, although some cellulolytic protozoans have been identified in rumen environments, and both fungi and bacteria have been studied as options for the native strategy (159, 184). Cellulolytic bacteria can be further subdivided into those that utilize a free enzyme strategy similar to fungi and those that rely on cell-surface associated cellulosome complexes (**Figure 1.4**) (185, 186).

Most work on cellulolytic fungi for the native strategy has employed *T. reesei* and *Fusarium oxysporum* (160, 187). As discussed at length, *T. reesei* is already used as a key source for most industrial cellulase cocktails, and thus employing this organism for direct conversion to product also makes sense (139). In addition to a strong biomass degradation phenotype, *T. reesei* is amenable to genetic manipulation and capable of utilizing a variety of lignocellulosic sugar substrates without genetic modification (both C_5 and C_6), despite the fact that native ethanol production in strains is low or non-existent (188, 189). Significant improvements in these ethanol titers have been achieved, with adaptation and strain selection yielding an improvement of 0.4 g/L to 5 g/L in one *Trichoderma* strain, and production up to 10 g/L observed from simple sugars in engineered strains (190, 191). Further efforts to improve ethanol yields in *T. reesei* will likely include traditional steps of knocking out genes responsible for byproducts and either overexpressing native alcohol production genes or heterologously expressing additional sets of genes from *S. cerevisiae,* which have largely been ignored so far (192). *F. oxysporum* is not as well developed as an industrial organism as *T. reesei,* but this fungi is also highly cellulolytic and genetic manipulation has allowed for ethanol titers up to 17.5 g/L on glucose (193). Importantly, when this titer of ethanol was reached, acetate titers of 12.8 g/L were also observed, indicating the opportunity for rerouting additional carbon and reductant toward ethanol production with future engineering efforts. Indeed, based on this work, overexpression of phosphoglucomutase and transaldolase – previously identified as metabolic bottlenecks – increased ethanol production on glucose to over 20 g/L and lowered acetate production to roughly 2 g/L (194). These titers are still too low for meaningful industrial utilization. For both biomass degrading fungi in discussion here, cellulase production takes place under aerobic conditions, while ethanologenic growth occurs under anaerobic conditions. Thus, in a CBP process utilizing these organisms, a first aerobic step

for enzyme production would precede the anaerobic fermentation stage, although the enzymatic degradation would continue into the second stage due to the insensitivity of these enzymes to the presence of oxygen (160).

The best studied native CBP bacterium is *Clostridium thermocellum,* although other *Clostridium* spp. and additional genre from the phylum Firmicutes have been explored as well (159). *C. thermocellum* is cellulolytic and natively produces ethanol as a primary carbon end-product of fermentation, albeit in combination with acetate (195); lactate, formate, valine, and alanine are also produced when metabolism is perturbed (196). Although *C. thermocellum* grows well on pretreated plant biomass, it lacks C_5 sugar metabolism, and to date, there are only a few reports of strains where pentose metabolism has been engineered (197). Nevertheless, there has been extensive manipulation of metabolism in *C. thermocellum*, assisted by ever-increasing genetic tools – including the very recent development of both endogenous and exogenous CRISPR/Cas systems that have drastically reduced the time necessary for strain generation (198).

While wild type ethanol titer from controlled fermentation on high loadings (100 g/L) of crystalline cellulose is 15 g/L, significant improvements have been achieved by sequential gene knockouts in combination with directed evolution (199). The highest ethanol titer achieved to date is 27 g/L from growth on crystalline cellulose, although several other strains have also been generated capable of producing more than 20 g/L ethanol (72, 200, 201). To achieve the highest titer strain, genes to non-ethanol reduced carbon products were deleted, including deletions of lactate dehydrogenase (Δ*ldh),* phosphotransacetylase (Δ*pta),* acetate kinase (Δ*ack*), pyruvate formate lyase (Δ*pfl*), and two hydrogenases *(*Δ*hydG*). These deletions resulted in production of 3.4 g/L ethanol, but the strain had a significant slow growth phenotype (202). Two rounds of adaptive evolution experiments increased the ethanol titer 8-fold by first focusing on improving growth rate

by serially transferring cultures daily on low carbon loading, for 150 rounds and then improving ethanol generation by transferring cultures weekly on higher carbon loading (72). It was concluded that the limit to further ethanol production in the evolved strain is likely ethanol toxicity, and future improvements might result from additional adaptive evolution experiments to increase ethanol tolerance (72). Sequencing of the evolved strains revealed a variety of changes to the genome, with some mutations in expected locations (e.g. *gapdh*, a key glycolytic enzyme and *adhE*, the primary ethanol-producing enzyme) but other mutations were observed that appear to be associated with transcription and sporulation (72). *C. thermocellum* can solubilize and ferment biomass, but pretreatment is necessary for efficient conversion and titers decrease significantly with high loading or substrates where efficient delignification and separation of hemicelluloses has not taken place (203, 204). Other biofuel products besides ethanol have been demonstrated by engineered *C. thermocellum*, most notably butanol (0.4 g/L) and isobutanol (5.4 g/L) from growth on crystalline cellulose (205, 206).

Unlike the free enzymes secreted by cellulolytic fungi, *C. thermocellum* degrades biomass with large multi-enzyme complexes attached to the cell surface dubbed 'cellulosomes' (**Figure 1.4C**) (207). Fungal cellulases are usually single proteins composed of a carbohydrate binding module (CBM) that anchors the enzyme to the substrate, which is connected to the catalytic glycosyl hydrolase (GH) domain by a flexible linker region, sometimes called the 'hinge' (**Figure 1.4B**) (208). In contrast, cellulosomes are often made up of dozens of enzymes, which are attached to the cell surface by large base proteins called scaffoldins (185). These structural scaffoldins contain numerous cohesin domains, which are complementary to dockerin domains on secreted glycoside hydrolases, or even to other scaffoldins, to help form more complicated super-structures. Scaffoldins also contain CBMs, like fungal cellulases, to help anchor the complex, and thus

potentially indirectly the bacterial cell, to substrates (209). In some cases, scaffoldins may not be anchored to the parent cell at all and exist as foundations for cell-free cellulosomes (210). Cellulosome diversity is broad, with some relatively simple structures, and some, like *C. thermocellum* revealing more complex mega-structures. The *C. thermocellum* cellulosome contains up to 63 enzymes, and even this is not the largest known cellulosome (185). Recently, designer cellulosomes have been engineered, containing specific GHs in attempts to optimize biomass degradation through improved synergies and for potential use in enzyme cocktails (211). Like secretion of free enzymes, expression of both simple and complex cellulosomes has been accomplished in *S. cerevisiae*, with some reports of higher cellulose degradation and ethanol production than from strains expressing fungal cellulases (212, 213).

Beyond the native strategy, some researchers have focused on developing co-cultures for CBP, relying on specific properties of more than one organism, rather than trying to create a single ideal CBP microbe (159). Examples of CBP co-cultures include combining *C. thermocellum* with *Thermoanaerobacterium thermosaccharolyticum* or combining *Aspergillus oryzae* with *S. cerevisiae*. In both of these cases, a cellulolytic organism is being paired with a non-cellulolytic organism. In the first case, *C. thermocellum* is incapable of utilizing pentose sugars, while *T. thermosaccharolyticum* can efficiently convert C_5 substrates to ethanol but lacks cellulolytic activity. Still, optimization is necessary as ethanol titers above 2.4 g/L have not been achieved (214). In the second case, *A. oryzae* is a moderately cellulolytic organism, but does not produce ethanol; the combination of *A. oryzae* with *S. cerevisiae* is used industrially for the production of sake in Japan and has now been explored for CBP potential showing production of 37 g/L ethanol from growth on brewers' spent grain (215). Though co-cultures and consortia have shown promise,

there are inherent risks associated with such fermentations on large industrial scale due to the potential culture instability and loss of the desired strains or ratios.

While most of the focus around CBP and, more broadly, cellulosic bioproducts in general, is in relation to biofuels production, it should be noted that fuels make up about 85% of a barrel of crude oil (216). The other 15% is essential for the production of different chemical feedstocks (216). Establishment of bioproducts beyond biofuels, such as plastic precursors, is important for decreasing societal oil dependence. Beyond ethanol, butanol, and isopropanol, lab-scale experiments have explored CBP approaches to the production of compounds such as pinene, polyhydroxyalkanoates, and fatty acid ethyl esters (217-219).

Thermophiles as Industrially Useful Microbes

Thermophiles are organisms that grow optimally above 45 °C. These can be further subdivided into moderate thermophiles, which grow optimally from 45-70 °C, extreme thermophiles which grow optimally between 70-80 °C, and hyperthermophiles, which grow optimally above 80 °C (220). While thermophilic fungi are known, all currently identified extreme thermophiles and hyperthermophiles are prokaryotes. Extreme and hyperthermophiles may be either bacteria or archaea, and significant diversity exists (221); thermophiles may be aerobes or anaerobes and grow in either marine or freshwater environments. To thrive above ambient temperatures, thermophiles exhibit properties that vary from their mesophilic counterparts, such as cell membrane composition and the thermostability of proteins (222, 223). Thermostable enzymes are desirable for many biotechnological processes, both at lab-scale and in industrial applications (224). The most obvious and impactful use of a thermophilic enzyme was the application of DNA polymerase from *Thermus aquaticus* (T_{opt} 70 °C) to improve PCR by eliminating the need to add fresh polymerase after each denaturation step (225); this improvement

to PCR revolutionized the life sciences. Now, thermostable enzymes derived from thermophilic organisms are key parts of a variety of processes in industries such as textiles, pulp, pharmaceuticals, detergent production, and food processing (226). Beyond thermostability, many of these enzymes have other desirable properties, such as high tolerance to solvents, detergents, and other typically denaturing conditions (227). Despite significant efforts to understand and apply the specific mechanisms that lead to thermotolerance in microbes or thermostability of individual proteins, these properties still remain largely outside our ability to engineer, and thus many native thermophilic enzymes are either employed directly or used as a starting point for experimental modification (223, 228-231).

In addition to the use of thermophilic enzymes for catalysis, there is great interest in utilizing thermophilic organisms directly as whole-cell biocatalysts (232). There are several potential advantages to exploiting thermophilic microbes for industrial processes such as CBP. Utilizing thermophilic organisms decreases the risk of contamination from mesophilic competition, which lowers costs associated with potential loss of batches of product, as well as sterilization costs. Additionally, thermophiles are less susceptible to phages, which are a significant source of economic costs associated with microbial food processes like yogurt production (220, 233). For production of desirable products, such as biofuels, employing extreme thermophiles can ease the separation of fermentation products from the growth medium by enabling the potential for directly and continuously stripping products from reaction vessels as they are produced ("temperature dependent product separation") (234). In a large-scale industrial setting, cooling and heating processes are expensive; mesophilic fermentations on industrial scales require significant costs in cooling due to the generation of high levels of metabolic heat (235). Maintaining a constant elevated temperature for a process could decrease overall costs, especially

as only initial input costs required to reach temperature are necessary, with metabolic heat potentially maintaining necessary elevated temperatures for thermophilic fermentations on large scales (236). Finally, many chemical processes are inherently more favorable at high temperature, with biomass degradation among these. Beyond physical implications for biomass degradation, saccharolytic enzymes generally have higher activity above ambient temperatures (157). To this end, anaerobic thermophilic cellulolytic bacteria, all known examples of which belong to the phylum Firmicutes, are of particular interest (237, 238). *C. thermocellum*, discussed in detail above for CBP applications, is one such thermophile, with an optimum growth temperature of 60 °C (239).

To optimize organisms such as *C. thermocellum* for industrial processes, genetic modifications are usually necessary. Systems for the genetic modification of thermophiles, especially extreme and hyperthermophiles, are rare (240). The paucity of these systems arises primarily from a two-fold issue with antibiotic use in these organisms: most antibiotics are not thermostable, degrading at temperatures above 60 °C, and almost all known antibiotic resistance markers are thermolabile (241). As a result, most genetic systems developed in these microbes to date rely on nutritional markers, such as uracil auxotrophy (240). More recently, some antibiotic markers have been established for use up to 80 °C, most notably a thermostable kanamycin resistance marker (242); kanamycin itself is also thermostable, maintaining 90% of activity after autoclaving at 120 °C (243). Modern genetic manipulation systems utilized in mesophilic organisms, such as CRISPR/Cas9, are not viable in thermophiles, due to a lack of appropriately thermostable homologs with the relevant activities. Very recently, some advances have been made in utilizing the Cas9 enzyme (GeoCas9) from *Geobacillus stearothermophilus* (T_{opt} 55 °C) for genetic manipulation of moderate thermophiles, although most use of CRISPR/Cas systems in

thermophiles to date has relied on taking advantage of endogenous systems (198, 244, 245). In extreme and hyperthermophiles, these options are still undeveloped. Despite advancements in genetic modification of non-model thermophilic organisms, the available tools still tend to be relatively blunt, with only minimal availability of more subtle and tunable tools like inducible promoters (220, 240, 246).

The Genus *Caldicellulosiruptor*

In freshwater hot springs with temperatures in the range of 70-80 °C and roughly neutral pH, bacteria from the genus *Caldicellulosiruptor* tend to dominate consortia, as confirmed by metagenomic sequencing (247, 248). These bacteria are Gram-positive and strict anaerobes. The type species for the genus, *C. saccharolyticus,* was first isolated from a hot spring in New Zealand in 1987 as *Caldocellum saccharolyticum,* before being reclassified several years later (249, 250). This organism was interesting at the time because in addition to growing optimally at 75 °C, it produces almost theoretical hydrogen yields from the fermentation of glucose (4 H_2 / glucose) and is weakly cellulolytic (250, 251). In the intervening decades, over a dozen other species of *Caldicellulosiruptor* have been isolated from hot springs on four continents, with optimal growth temperatures ranging from 65-78 °C (**Table 1.2**) (246). Significantly, genome sequences are published for each of the currently verified members of the genus. While optimal growth conditions, carbon substrates, and fermentation end-products vary among *Caldicellulosiruptor* spp., all grow on a wide range of lignocellulose derived sugar substrates (246).

Next to *C. saccharolyticum, Caldicellulosiruptor bescii* is the most studied member of the genus. First isolated in 1990 as *Anaerocellum thermophilum* and not reclassified until nearly two decades later, this species simultaneously co-ferments C_5 and C_6 sugars to acetate, lactate, hydrogen gas and carbon dioxide (252-254). *C. bescii* is a promising CBP host microbe because it

has the highest temperature optimum of any known cellulolytic organism (T_{opt} 78 °C), expresses the most active naturally occurring cellulase yet identified, and genetic tools are available (255, 256). Furthermore, this organism can grow on high substrate loadings of non-pretreated plant biomass as a sole carbon source (257). *C. bescii* utilizes an enormous range of sugar substrates, including the five most common monosaccharides found in lignocellulose: glucose, xylose, arabinose, galactose, and mannose (258).

The cellulolytic and xylanolytic properties of different *Caldicellulosiruptor* species vary significantly, and correlate well with the content of the genome at each of two major loci, termed the Glucan Degradation Locus (GDL) and Xylan Degradation Locus (XDL) (247, 259, 260). Cellulases, mannanases, and xylanses produced by *Caldicellulosiruptor* spp., expressed at these loci along with a variety of sugar transporters and regulators, are different from the fungal free cellulases and cellulosomes discussed so far. These large multidomain enzymes are individual polypeptides akin to their fungal counterparts, but composed of multiple CBMs and GH domains separated by several linker regions (**Figure 1.4A**) (261). Exemplified by *C. bescii* athe_1867 (*celA*), they have modular structures and can approach 2000 amino acids in length (259). *C. bescii* CelA, which out-performs commercial cellulases in crystalline cellulose degradation, is comprised of a GH9 domain and a GH48 domain, separated by multiple linker regions and three CBM domains (**Figure 1.4A**) (186). The five other cellulases encoded in the GDL of *C. bescii* also consist of two GH domains (though these vary) separated by multiple CBM domains. The substrate profiles and hydrolysis characteristics of each of these GHs have been studied and evaluated for relative contribution by relation back to the total *C. bescii* secretome activity (261). As has been observed with other organisms (e.g. *C. thermocellum*), without biological conversion, the process

of biomass degradation is slowed, likely a by a mass action effect due to the lack of removal of degraded glucans (261).

In addition to the secretion of free enzymes, *Caldicellulosiruptor* attach themselves to lignocellulosic carbohydrates by surface-layer homology domains (SLH) that exist in some *Caldicellulosiruptor* GHs and help form an indirect connection, anchoring one region of the enzyme to a cell and another region to the lignocellulosic substrate (262). This genus also produces small novel non-catalytic carbohydrate binding proteins known as tāpirins that are encoded in the GDL (263). Tāpirins from different *Caldicellulosiruptor* have been recombinantly expressed, and their relative binding capabilities have been compared (264). Likewise, deletion of tāpirins in *C. bescii* resulted in a strain that did not bind to crystalline cellulose (265). The direct contribution of adhesion of cells to biomass degradation has not yet been assessed, although it is presumed to be significant.

Fermentative Metabolism in *C. bescii*

Primary carbon metabolism in *Caldicellulosiruptor* spp. utilizes the Embden-Meyerhof Parnas (EMP) glycolytic pathway (**Figure 1.5**) (266). No evidence of CCR has been observed in any species of *Caldicellulosiruptor*, and similar growth is supported by a variety of both pentose and hexose sugars (251, 267). Roughly, pentoses are metabolized by the sequential activities of xylose isomerase and xylulokinase, and then likely enter primary metabolism through glycolysis from the pentose phosphate pathway (PPP) as glyceraldehyde-3-phosphate (GAP) or fructose-6-phosphate (266). The main carbon fermentation products of *C. bescii* are carbon dioxide, lactate, and acetate (268). While none of these native products are desirable as liquid fuels, understanding fermentative metabolism is necessary for any informed efforts to engineer the production of alternative reduced carbon end-products. Although some *Caldicellulosiruptor* spp. produce

ethanol from growth on glucose, *C. bescii* reportedly only produces trace amounts (268, 269). Carbon dioxide is produced stoichiometrically with acetate by pyruvate ferredoxin oxidoreductase (POR), which oxidatively decarboxylates pyruvate to acetyl-CoA by the reduction of ferredoxin. Acetyl-CoA is further converted to acetate with the concomitant production of ATP by phosphotransacetylase (PTA) and acetate kinase (ACK).

C. bescii contains both a cytosolic bifurcating [FeFe] hydrogenase and a membrane-bound [NiFe] hydrogenase. The bifurcating [FeFe] hydrogenase is believed to be the major redox-cofactor recycling enzyme in *C. bescii*, simultaneously oxidizing both ferredoxin and NADH to produce H_2; this enzyme is constitutively expressed at high levels (270). Deletion of the maturation genes necessary for assembly of the membrane-bound [NiFe] hydrogenase has confirmed that the majority of H_2 production during fermentative growth of *C. bescii* results from the activity of the bifurcating hydrogenase (271). When the organism is growing rapidly, or as H_2 partial pressure in a culture increases, NADH is increasingly recycled with reduction of pyruvate by lactate dehydrogenase (LDH). Under such circumstances, reduced ferredoxin is likely recycled by the membrane bound hydrogenase, as has been reported in *C. saccharolyticus* (251).

C. bescii does not express a traditional transhydrogenase for interconversion of NADH and NADPH, but rather expresses a bifurcating NADH-dependent ferredoxin-$NADP^+$ oxidoreductase (Nfn). The role of this enzyme in maintaining redox homeostasis in *C. bescii* is not currently understood, but it is hypothesized to be involved in simultaneously utilizing the NADH and reduced ferredoxin pools required for catabolic processes to generate the NADPH necessary for anabolic processes (**Figure 1.5**). This idea is supported by the fact that *C. bescii*, like other characterized *Caldicellulosiruptor* spp., appears to lack the oxidative branch of the pentose phosphate pathway, a canonical system for generating NADPH (266). A similar role for Nfn has

been demonstrated in hyperthermophilic archaea such as *Pyrococcus furiosus*, which similarly lacks the oxidative PPP (130).

The identification of genes involved in utilizing the metal tungsten in the genomes of all *Caldicellulosiruptor* species (272) recently led to the discovery of a new class of glyceraldehyde-3-phosphate-oxidizing enzyme (GOR) that contains tungsten at its catalytic site (273). A detailed description of the characterization of this novel enzyme composes Chapter 2 of this work. GOR uses ferredoxin rather than NAD^+ as an electron carrier, analogous to the enzyme GAPOR found in many species of archaea. While the canonical GAPDH oxidizes GAP to produce NADH and 1,3-bisphosphoglycerate, which is used by phosphoglycerate kinase (PGK) to generate 3-phosphoglycerate (3PG), GOR generates 3PG directly from GAP, thereby circumventing the energy-conserving step. Thus, *C. bescii* contains two parallel pathways from GAP to 3PG: the ATP-generating GAPDH-PGK pathway, dependent on NAD^+, and the non-energy conserving GOR pathway, which reduces ferredoxin. It has been speculated that these dual glycolytic pathways in *C. bescii* are employed by the cell to manage carbon and electron flux under growth conditions in which electron carriers cannot be efficiently recycled; balance of GAP flux between these two pathways is likely impacted by environmental H_2 conditions, as is the case with pyruvate reduction by LDH. GOR expression is regulated by the Rex repressor, which de-represses transcription when the $[NADH]/[NAD^+]$ ratio in the cell is high (274). GAPDH and PGK expression are likely regulated by other transcription factors, in response to fructose-1,6-bisphosphate. Increased understanding of the interplay between these two pathways will be essential for efficiently directing reductant flow in future metabolic engineering efforts.

Establishment and Utilization of a Genetic System in *C. bescii*

The genetic engineering pipeline for these non-model organisms began with the identification of a necessary methylation patterns within *C. bescii.* Identifying the restriction enzyme, CbeI, and its cut site was crucial for determining how to protect foreign DNA during transformation. Screening and selection methods thus far have been accomplished mostly with nutritional selection (256). In *C. bescii*, uracil auxotrophy was first established via spontaneous mutations with 5-FOA, resulting in the partial deletion of genes (Δ*pyrBCF*) involved in the uracil biosynthetic pathway (256). Once the corresponding methyltransferase to CbeI was identified and a replicating shuttle vector was designed and tested, the basis of the genetic engineering system was established for *C. bescii* (256, 275).

Since this initial parent strain was developed, multiple lineages of modified *C. bescii* strains have been generated. Initially, the focus was on creating new parent strains, including a smaller, yet still useful, spontaneous deletion of uracil biosynthesis genes (Δ*pyrFA*) and of CbeI (Δ*pyrFA*, Δ*cbeI*) (275, 276). This removed the need for methylation of inserted DNA and demonstrated the first targeted knockout of genes in *C. bescii*; similar methods were also shown to work in the weakly cellulolytic *C. hydrothermalis* (276, 277). To date, *C. bescii* and *C. hydrothermalis* are the only genetically tractable species of *Caldicellulosiruptor*. The most recent advance in parent strain development has been a move away from nutritional requirements for transformant selection with *C. bescii*. A high temperature kanamycin resistance gene (*htk*), originally optimized for growth at 80 °C in *Thermus thermophilus*, was utilized in *C. bescii* to design both replicating and non-replicating vectors (242). With *htk* expression, *C. bescii* grew well in concentrations of kanamycin up to 200 and 800 μg/ml in the replicating and genome-integration strains, respectively; parent and wild type strains were normally inhibited by only 10 μg/ml of kanamycin. Not only was *htk* shown

to be a valuable selective marker, but its use also led to the generation of new genomic parent strains (MACB lineage), starting with the targeted, markerless, clean deletion of *pyrE*, and subsequent individual deletions of *cbeI* and *ldh*.

Additionally, early during the pipeline development, a transposon, IS*Cbe4*, was identified in the *ldh* gene of a *C. bescii* genetic parent strain, leading to the identification of 47 autonomous insertion elements across eight sequenced *Caldicellulosiruptor* species (278). This later led to an investigation of the genomic stability of the available parent strains, which showed that the potential of transposon insertions was more severe than originally thought (279). Sequencing both old (JWCB) and new genetic linages indicated that, while seven types were identified, the only active insertion element in *C. bescii* appears to be IS*Cbe4*. The number of instances of IS*Cbe4* increased greatly from 7 in wild type *C. bescii* to 23 in the most modified JWCB parent strain, JWCB018. The MACB lineage on the other hand only gained 1-3 replicates of IS*Cbe4*. Along with transposons, the study also identified several mutations present in the JWCB strains, ranging from single nucleotide polymorphisms to whole gene deletions or genome arrangements. This also highlights the importance of resequencing the complete genome of recombinant strains, especially as a mis-assembly of a highly studied area (which includes *celA*) in the original reference wild type genome of *C. bescii* was recently discovered (279, 280). Overall, the MACB genetic strains have been shown to be more genomically stable and generally more similar to the wild type, making them the better parent strains to continue genetic manipulations in *C. bescii*.

Using genetic engineering, both metabolic and cellulolytic pathways have been targeted within *C. bescii*. A variety of strains have been made to explore the role of specific genes for lignocellulose degradation, including knockouts of pectin lyases, tāpirins, and glycoside hydrolases, and knock-ins of enzymes from species native to or outside of the *Caldicellulosiruptor*

genus (259, 262, 265, 281-288). The knockout of a single glycosyltransferase gene (athe_1864) resulted in the inability to grow on crystalline cellulose, although another report suggested that glycosylation of key cellulases (athe_1867, athe_1865 etc.) located in the Glucan Degradation Locus (GDL) did not affect biomass degradation (261, 289, 290). Other gene deletions have been used to study redox metabolism in *C. bescii,* namely the deletion of the transcriptional repressor Rex and the deletion of the newly discovered alternative glycolytic enzyme, GOR, discussed in Chapter 2 of this work (272-274).

The first modification made to explore native metabolism was the deletion of lactate dehydrogenase (Δ*ldh*), which successfully eliminated lactate production and resulted in increased formation of acetate and H_2 (291). More importantly, the Δ*ldh* strain was the parent for engineering ethanol production by *C. bescii*, either by inserting the gene (*adhE*) encoding the bifunctional alcohol/aldehyde dehydrogenase from *Clostridium thermocellum* (T_{opt} 60 °C) or by inserting *adhB* or *adhE* from *Thermoanaerobacter pseudoethanolicus* (T_{opt} 65 °C) (292, 293). These strains produced ethanol at maximum concentrations of approximately 15 mM at 60 °C and 2 mM at 75 °C, respectively (292, 293). Expression of an NADPH-dependent butanol dehydrogenase (*bdhA*) from *Thermoanaerobacter pseudoethanolicus* 39E did not result in ethanol production, but improved furan aldehyde tolerance in an engineered strain (294). More recently, the six-gene operon encoding the energy-conserving reduced ferredoxin NAD^+ oxidoreductase (Rnf) complex from *Thermoanaerobacter* sp. X514 (T_{opt} 60 °C) was expressed in a strain expressing *adhE* from *C. thermocellum*. Rnf provides a route for using reduced ferredoxin from sugar oxidation to generate NADH for ethanol production. In the new genetic lineage of *C. bescii,* co-expression of *adhE* from *C. thermocellum* and *rnfCDGEAB* from *Thermoanaerobacter* sp. X514 increased ethanol production at 60 °C five-fold over expression of *adhE* alone, to 75 mM (3.5 g/L) (270).

This represents a major step towards the goal of efficient CBP in *C. bescii* (295). To explore potential for improvements in bioproduct formation by *C. bescii*, ethanol makes a good model because its production is simple and has been demonstrated already. Beyond the canonical AdhE method, new approaches to ethanol production should aim to balance redox cofactor use and energy (i.e. ATP) generation in primary metabolism.

The AOR-Adh Pathway for Bioalcohol Conversion

In 2014, a non-canonical metabolic pathway for alcohol production was engineered into the hyperthermophilic archaeon *Pyrococcus furiosus* (T_{opt} 100 °C). Insertion of a bacterial primary alcohol dehydrogenase gene (*adhA*) into the *P. furiosus* genome allowed the conversion of glucose to ethanol in a pathway that proceeds through acetate, rather than acetyl-coenzyme A (acetyl-CoA) (296). Acetate is reduced to acetaldehyde, catalyzed by the native *P. furiosus* enzyme aldehyde ferredoxin oxidoreductase (AOR). Acetate reduction is driven by the low redox potential of the electron carrier, a 4Fe-ferredoxin. The heterologously-expressed AdhA from *Thermoanaerobacter* sp. X514 catalyzes acetaldehyde reduction to ethanol using NADPH as the electron donor (**Figure 1.6**) (296). Interestingly, due to the broad substrate specificities of both AOR and AdhA, a variety of exogenously added organic acids were also able to be converted to their corresponding alcohols with the concurrent production of ethanol (296).

While a recent resurgence of interest in the AOR-Adh pathway ostensibly resulted from heterologous expression in *P. furiosus,* several acetogenic bacteria naturally utilize this pathway. Indeed, the natural existence of this pathway was hypothesized three decades ago, but direct evidence in native systems was only recently exhibited (297). Recent work in the mesophilic clostridia *C. ragsdalei*, *C. autoethanogenum,* and *C. ljungdahlii* has confirmed the existence of the AOR-Adh pathway in these organisms, which contain both native AORs and native Adhs (298-

300). Acetate produced by the Wood-Ljungdahl pathway can enter the AOR-Adh pathway to produce ethanol as an end-product; thus, this has also been described as the Acetate Re-assimilation pathway (301). When these species grow on carbon dioxide and hydrogen, ethanol formation by the AOR-Adh pathway is thermodynamically feasible and is believed to act as a valve for avoiding culture acidification due to acetate buildup (299). In addition to several *Clostridium* the native pathway has now been demonstrated in the closely related genus *Thermoanaerobacter* as well as the phylogenetically distinct gram negative acetogen *Sporomusa ovata* (301-304). While alcohol titers from these organisms are robust (up to 16.1 g/L ethanol produced by engineered *C. ragsdalei*), none of these microbes are cellulolytic, and no cellulolytic *Clostridium* spp. has been reported to naturally contain this pathway (305). Furthermore, while many of these organisms can utilize both C_5 and C_6 sugars (namely xylose and glucose), they are likely unable to utilize both substrates simultaneously, as the closely related *Clostridium acetobutylicum* is affected by CCR (306). All three of the above-mentioned clostridial species grow optimally at 37 °C, and grow poorly above 40 °C, implying that the AOR and Adh homologs encoded by these organisms are likely unsuitable for recombinant expression in thermophilic microbes (307-309).

While a large variety of alcohol dehydrogenases are well characterized, less is known about variable characteristics of AOR family enzymes. Originally identified as “carboxylic acid reductase” (CAR) in 1989 from *Clostridium thermoaceticum* (a thermophile) and *Clostridium formicoaceticum* (a mesophile), this family of enzymes were soon renamed (310-312). AORs are tungsten-containing enzymes that rely on a complex tungstobispterin cofactor, analogous to the molybdopterin cofactors found in most molybdenum-containing enzymes. These enzymes are found in both archaea and bacteria and are most well characterized from thermophilic hosts (313).

More recently an unrelated class of CARs have gained attention in the literature. This alternative class of CARs were first observed and characterized in the fungus *Neurospora crassa* in the 1960s and 1970s, but no bacterial homologs were identified until 1997 in *Nocardia iowensis* (314-316). The *Nocardia* CAR was then cloned, sequenced, and fully characterized in 2004 (317); since 2013, several other homologs have now been identified in bacteria such as *Mycobacterium* and *Tsukamurella*, as well as the fungi *Aspergillus* and *Trametes* (318). As opposed to AORs, which rely on the low redox potential of ferredoxin, CARs utilize both NADPH and ATP. Indeed, unlike the strictly oxygen sensitive AORs, CARs are oxygen tolerant, derived from aerobic organisms. Compared with the already broad substrate specificities of AORs, CARs are capable of an even wider range of activities. CARs can reduce both straight and branched chain acids up to C_{18} and are particularly active on aromatics, such as substituted benzoic acids, while still retaining activity on smaller acids like butyrate (318). While CARs have been used in some pathway engineering efforts in organisms like *E. coli* to make unusual products like 3-hydroxytyrosol, efforts are often hampered by the steep cofactor requirements and a trade-off with product selectivity (319); often, aldehyde products are reactive, and thus 'neutralized' or removed by background activities in the cell. CARs in thermophilic hosts have not been identified, and these enzymes are not encountered as part of native ethanol production pathways.

The AOR-Adh pathway is a strong candidate for engineering into *C. bescii* because it allows for the formation of ethanol as well as other short and medium chain alcohols from their corresponding acids. Short and medium chain fatty acids have received attention as a chemical feedstock that could be generated in high concentration from the organic fraction of municipal solid waste, and then catalytically upgraded (either biologically or non-biologically) to produce chemicals with superior properties, such as by the AOR-Adh pathway, to generate higher alcohols.

Alcohols like butanol and hexanol are high-energy drop-in fuels, as compatible with gasoline infrastructure as biodiesel is with petroleum infrastructure. In the AOR-Adh pathway to upgrade acids, *C. bescii* should utilize both reduced ferredoxin and reduced nicotinamide nucleotides generated from glucose oxidation. Chapter 3 of this work is comprised of the characterization of *C. bescii* strains that heterologously express the AOR-Adh pathway. *C. bescii* does not naturally contain either AOR or Adh and so the genes encoding these enzymes are from the same thermophilic sources as comprised the *P. furiosus* pathway. This work provides proof-of-principle that the AOR-Adh pathway can be harnessed for ethanol production and organic acid upgrading in a highly cellulolytic organism.

Research Objectives

There were two overarching goals of this research. 1) Improving understanding of the native *C. bescii* metabolism and 2) Applying this understanding to assist in the informed modification of metabolism to produce bioproducts. To the first aim, work in Chapter 2 identified a key new route through primary metabolism in *C. bescii*. This work increased our understanding of the function and importance of several major enzymes, including broader implications for the utilization and balancing of different redox cofactors in the cell. To the second aim, work in Chapter 3 demonstrated the heterologous expression of a non-native metabolic pathway. This work explored the possibility of using *C. bescii* to produce both ethanol and next-generation drop-in fuels, through the upgrading of carboxylic acids to alcohols. Chapter 4 tied these ideas together by exploring ways to improve product formation on two fronts: investigating a pathway for sugar uptake on the front end of metabolism and attempting to improve ethanol production in the current *C. bescii* benchmark strain by improving redox balance with further genetic modification.

TABLES AND FIGURES

Table 1.1 Properties of common liquid fuels

Jet fuel, diesel, gasoline, and biodiesel are mixtures of different compounds, with variable properties. Jet fuel, diesel, and gasoline are derived from crude oil, and thus exact properties vary depending on the source of crude oil and the distillation procedures utilized.

Fuel	Energy Density (MJ/L)	Specific Energy (MJ/kg)	Density (kg/L)[1]	Boiling Point (°C)	Carbon Number
Diesel	38.6	42-46	0.78-0.84	180-340	C_8-C_{25}
Biodiesel	33.3-35.7	39	0.85	315-350	C_{12}-C_{22}
Gasoline	34.8	43-46	0.74	26.7-225	C_4-C_{12}
Ethanol	21.2	26.8	0.79	78	C_2
Butanol	29.2	33.1	0.81	118	C_4
Jet Fuel	35.1	43.1	0.79-0.83	150-255	C_5-C_{14}

[1] Average density at 15 °C

Table 1.2 Characteristics of known *Caldicellulosiruptor* species

Species	Isolation Location	Isolation Temperature (°C)	Isolation pH	T_{opt} (°C)	Cellulose Degradation[1]	# of GH Domains[2]
C. acetigenus	Hveragerði, Iceland	68	7	65-68	No	66
C. bescii	Kamchatka, Russia	60-98	NR	78	High	52
C. changbaiensis	Changbai Mountains, China	83	7	75	High	66
C. danielii	Waimangu, New Zealand	NR	NR	ND	High	69
C. sp. F32	Qingdao, China	NR	NR	75	NR	45
C. hydrothermalis	Kamchatka, Russia	55-65	4.5-8	65	No	62
C. kristjanssonii	Hveragerði or Fludir, Iceland	78	8.7	78	No	37
C. kronotskyensis	Kamchatka, Russia	55-65	4.5-8	70	High	77
C. lactoaceticus	Hveragerði, Iceland	74	8	68	Moderate	44
C. morganii	Rotorua, New Zealand	63	8.8	ND	High	49
C. naganoensis	Nagano Prefecture, Japan	75-85	9	75	High	44
C. obsidiansis	Wyoming, USA	66	5	78	High	47
C. owensensis	California, USA	32	9	75	No	51
C. saccharolyticus	Taupo, New Zealand	48	NR	70	High	59

NR: Not Reported

ND: Not Determined

[1] Cellulose degradation phenotypes: No (little or no growth observed), Moderate (suboptimal growth compared with glucose), High (growth similar to glucose).

[2] Number of Glycosyl hydrolase (GH) domains represents number of ORFs containing at least one GH domain, not total number of GH domains in the genome. Number of GH containing ORFs reproduced from (247), with the exception of *C. changbaiensis*, which was obtained from http://www.cazy.org.

Figure 1.1 Global energy consumption 1900-2019

Total global energy consumption by source since 1900. Data adapted from (320) and (1), and compiled by http://www.ourworldindata.org (321).

Figure 1.1

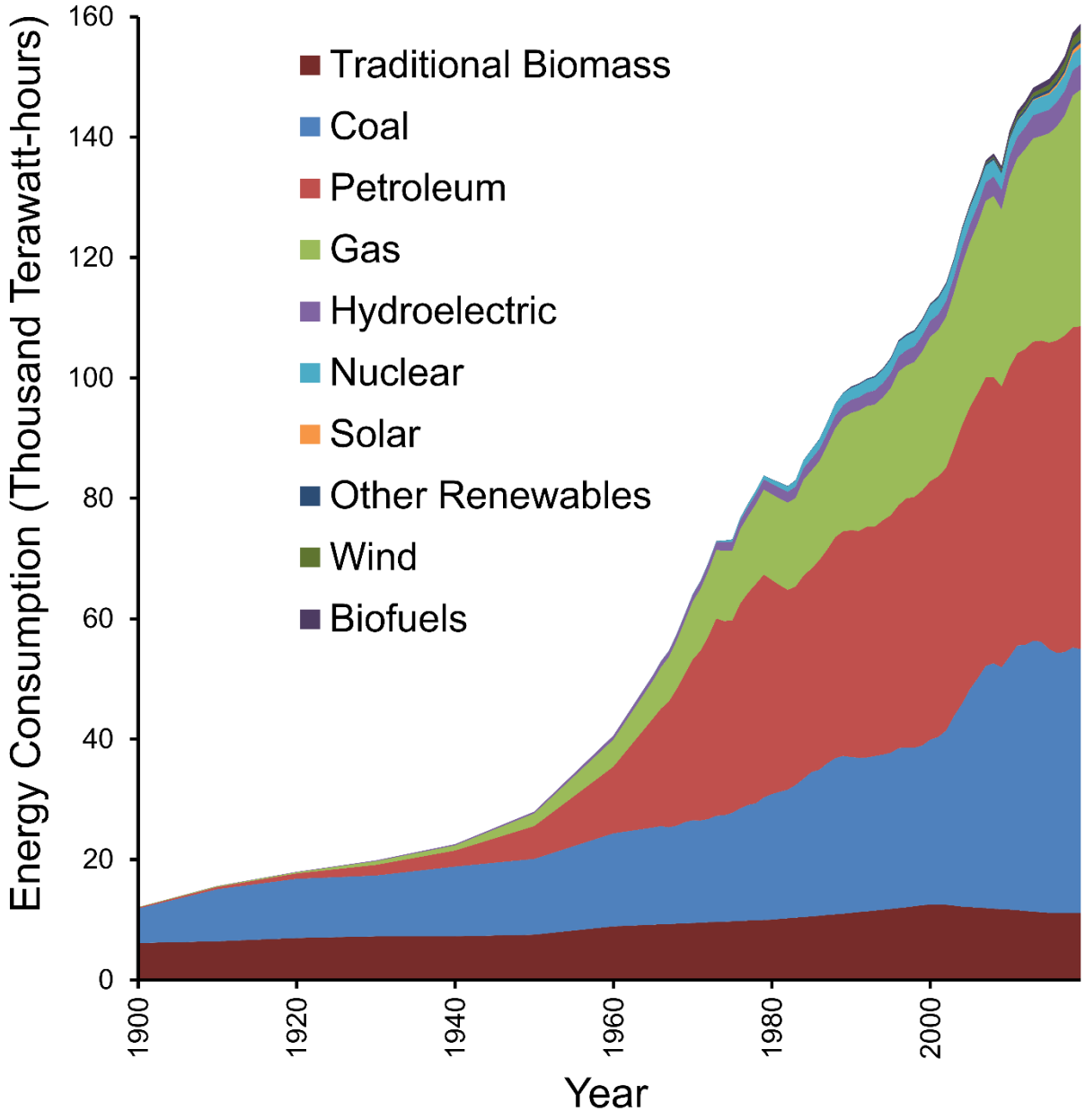

Energy Consumption (Thousand Terawatt-hours)
160
140
120
100
80
60
40
20
0
1900
1920
1940
1960
1980
2000
Year
Traditional Biomass
Coal
Petroleum
Gas
Hydroelectric
Nuclear
Solar
Other Renewables
Wind
Biofuels

Figure 1.2 Sources of energy in the United States

Sources of energy in the United States, with further subdivision of renewable energy. (A) Total energy consumption in the United States in 2019 by energy source, including all five end-use sectors: transportation, industrial, residential, commercial, and electric power. Renewable energy is further subdivided (inset). (B) Total electricity generation in the United States in 2019 by energy source. The composition of renewable electricity production by source is further subdivided (inset). Data obtained from the United States Department of Energy's Energy Information Administration (http://www.eia.gov).

Figure 1.2

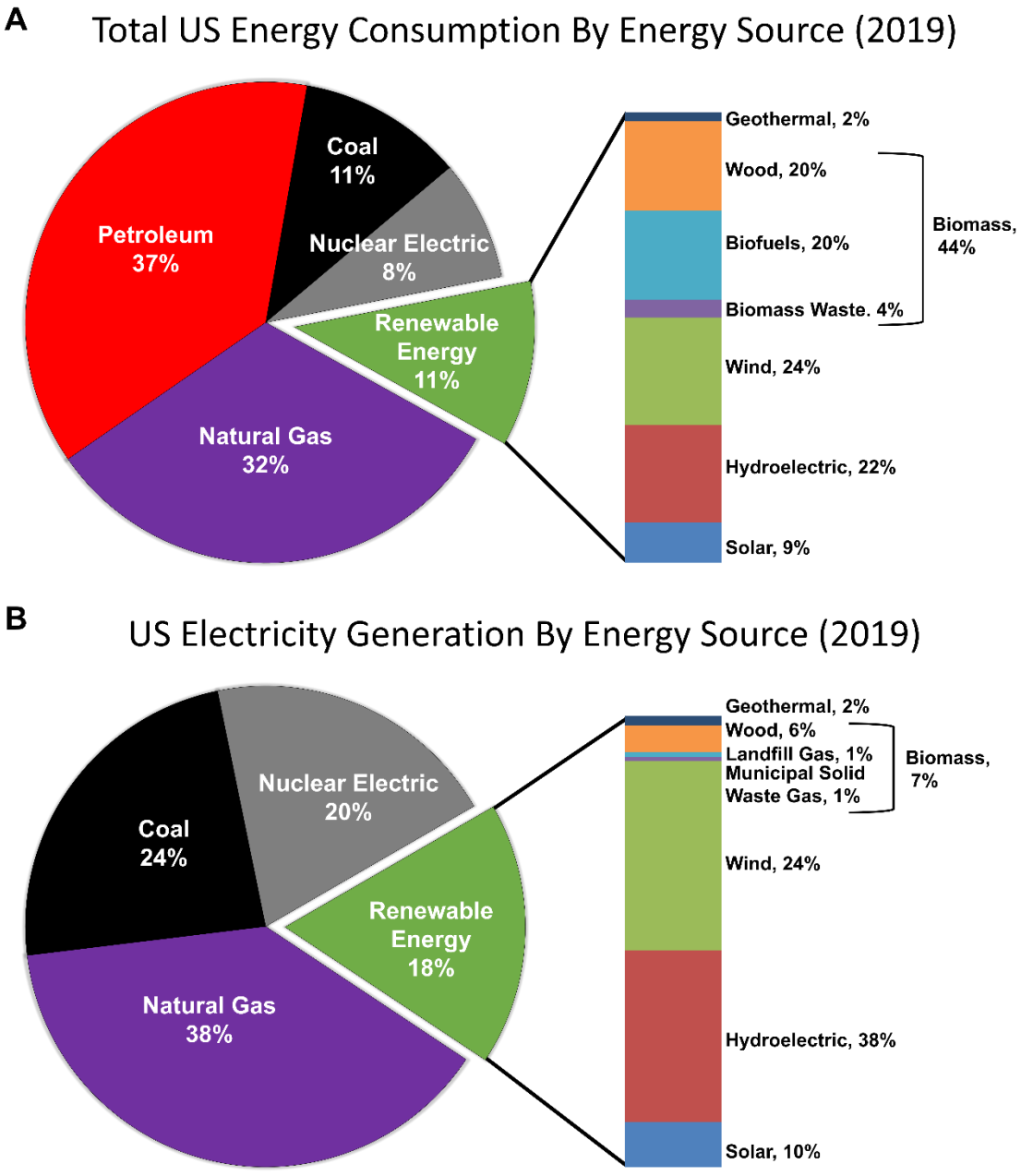

A
Total US Energy Consumption By Energy Source (2019)
Petroleum 37%
Coal 11%
Nuclear Electric 8%
Renewable Energy 11%
Natural Gas 32%
Geothermal, 2%
Wood, 20%
Biofuels, 20%
Biomass Waste. 4%
Biomass, 44%
Wind, 24%
Hydroelectric, 22%
Solar, 9%
B
US Electricity Generation By Energy Source (2019)
Coal 24%
Nuclear Electric 20%
Renewable Energy 18%
Natural Gas 38%
Geothermal, 2%
Wood, 6%
Landfill Gas, 1%
Municipal Solid Waste Gas, 1%
Biomass, 7%
Wind, 24%
Hydroelectric, 38%
Solar, 10%

Figure 1.3 Strategies for bioethanol production

Schematic strategies for three different approaches to bioethanol production. (A) The Starch-based ethanol strategy is utilized for industrial bioethanol production from corn in the United States. Hexose sugars derived from starch are fermented by yeast. (B) Three different approaches for cellulosic ethanol are presented. Soluble pentose sugar streams are shown in orange. Hexose sugar streams are shown in cyan. The Separate Hydrolysis and Fermentation (SHF) strategy is indicated with red accents. The Simultaneous Saccharification and Fermentation (SSF) strategy is indicated by blue accents. A Simultaneous Saccharification and Co-Fermentation (SSCF) strategy is shown with green accents. (C) The Consolidated Bioprocessing strategy (CBP) combines enzyme production, enzymatic saccharification, and fermentation into a single step. In some variations of CBP, both hexose and pentose sugars are fermented.

Figure 1.3

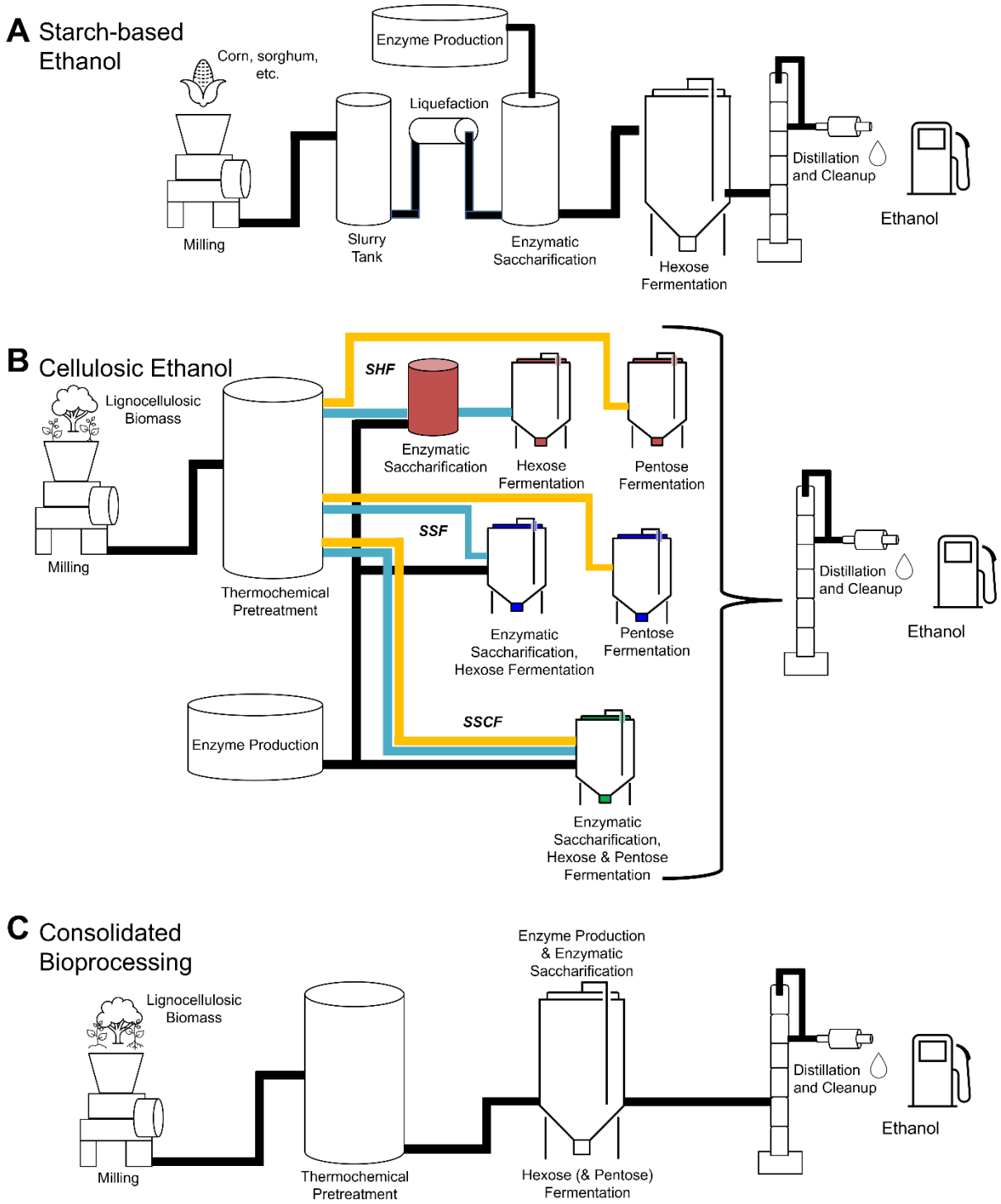

A Starch-based Ethanol
Corn, sorghum, etc.
Enzyme Production
Liquefaction
Milling
Slurry Tank
Enzymatic Saccharification
Hexose Fermentation
Distillation and Cleanup
Ethanol
B Cellulosic Ethanol
Lignocellulosic Biomass
Milling
Thermochemical Pretreatment
SHF
Enzymatic Saccharification
Hexose Fermentation
Pentose Fermentation
SSF
Enzymatic Saccharification, Hexose Fermentation
Pentose Fermentation
Enzyme Production
SSCF
Enzymatic Saccharification, Hexose & Pentose Fermentation
Distillation and Cleanup
Ethanol
C Consolidated Bioprocessing
Lignocellulosic Biomass
Milling
Thermochemical Pretreatment
Enzyme Production & Enzymatic Saccharification
Hexose (& Pentose) Fermentation
Distillation and Cleanup
Ethanol

Figure 1.4 Domain arrangement of cellulose-degrading enzymes

Three categories of cellulose-degrading enzymes are exemplified by the multidomain CelA from *Caldicellulosiruptor bescii* (A), the cellobiohydrolase Cel7A from *Trichoderma reesei* (B), and the cellulosome complex from *Clostridium thermocellum* (C). CelA and Cel7A are presented on the same scale. Glycosyl hydrolase domains, carbohydrate binding module domains, linker regions, and signal peptides are indicated. CelA and Cel7A each consist of a single polypeptide, while the cellulosome consists of multiple polypeptides linked by a scaffoldin backbone. The cellulosome is shown at a smaller scale to account for its larger size. Due to the variable structure of the cellulosome, specific glycoside hydrolase domains are not indicated; variability of glycosyl hydrolase domains in the cellulosome is indicated with colors representing different families.

Figure 1.4

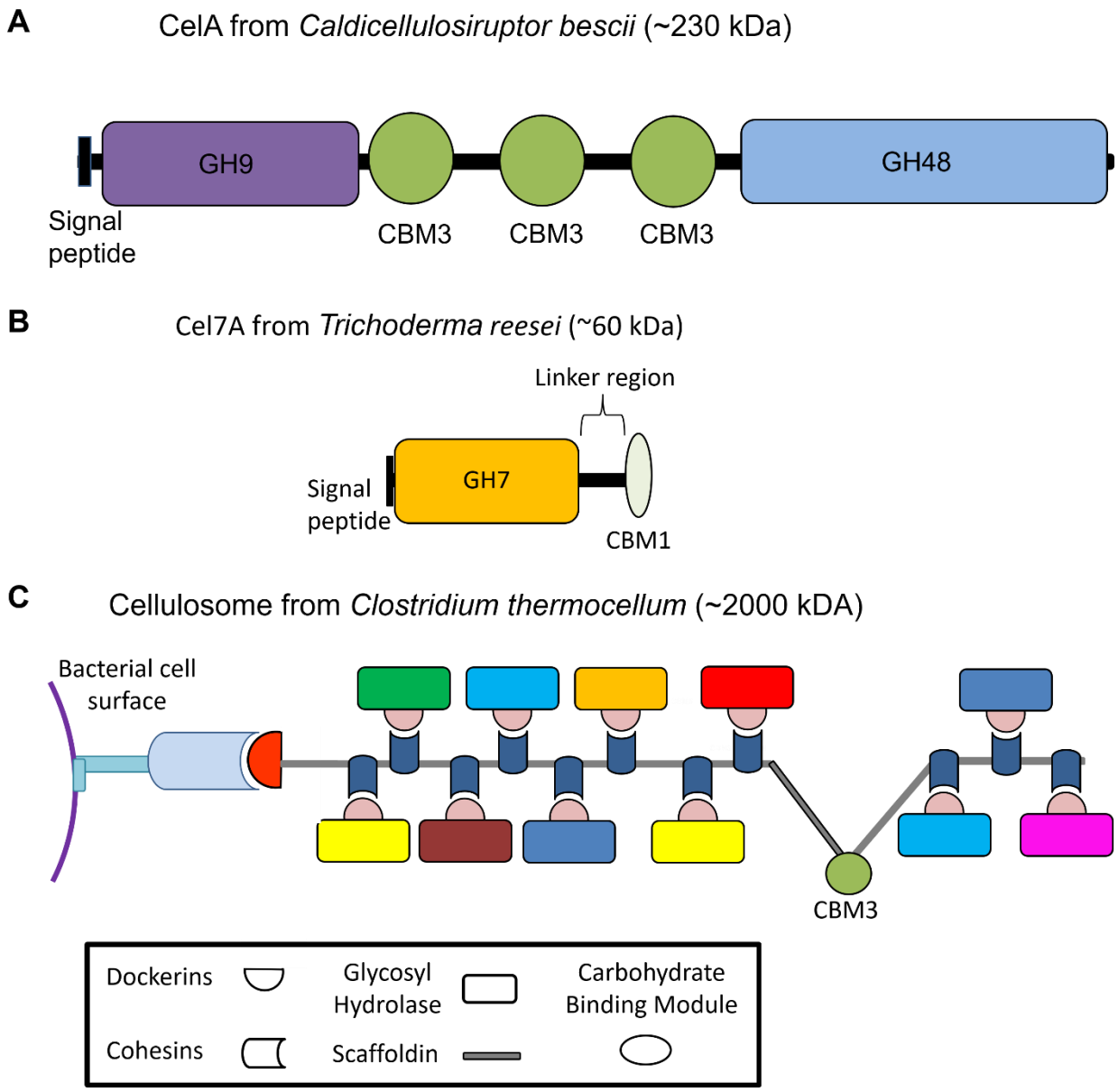

A
CelA from Caldicellulosiruptor bescii (~230 kDa)
GH9
GH48
Signal peptide
CBM3
CBM3
CBM3
B
Cel7A from Trichoderma reesei (~60 kDa)
Linker region
Signal peptide
GH7
CBM1
C
Cellulosome from Clostridium thermocellum (~2000 kDA)
Bacterial cell surface
CBM3
Dockerins
Glycosyl Hydrolase
Carbohydrate Binding Module
Cohesins
Scaffoldin

Figure 1.5 Primary metabolism in *C. bescii*

Caldicellulosiruptor species can utilize both C_5 and C_6 sugars, which enter the EMP glycolytic pathway. Central metabolism includes two parallel pathways for oxidizing glyceraldehyde-3-phosphate to 3-phosphoglycerate. Primary redox cofactors ferredoxin (red) and NAD(H) (green) can be used to generate NADP(H) (blue), as no oxidative pentose phosphate pathway is present. Abbreviations: hexokinase (HK), phosphoglucoisomerase (PGI), triosephosephate isomerase (TPI), fructose-bisphosphate aldolase (FPA) xylose isomerase (XI), xylulokinase (XK) transketolase (TK), phosphofructokinase (PFK), enolase (ENO), phosphoglycerate mutase (PGM), bifurcating hydrogenase (BF-H_2ase), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), glyceraldehyde 3-phosphate oxidoreductase (GOR), phosphoglycerate kinase (PGK), pyruvate kinase (PK), lactate dehydrogenase (LDH), pyruvate oxidoreductase (POR), phosphotransacetylase (PTA), acetate kinase (ACK), membrane bound hydrogenase (MBH), bifurcating NADH-dependent reduced ferredoxin $NADP^+$ oxidoreductase (BF-Nfn),

Figure 1.5

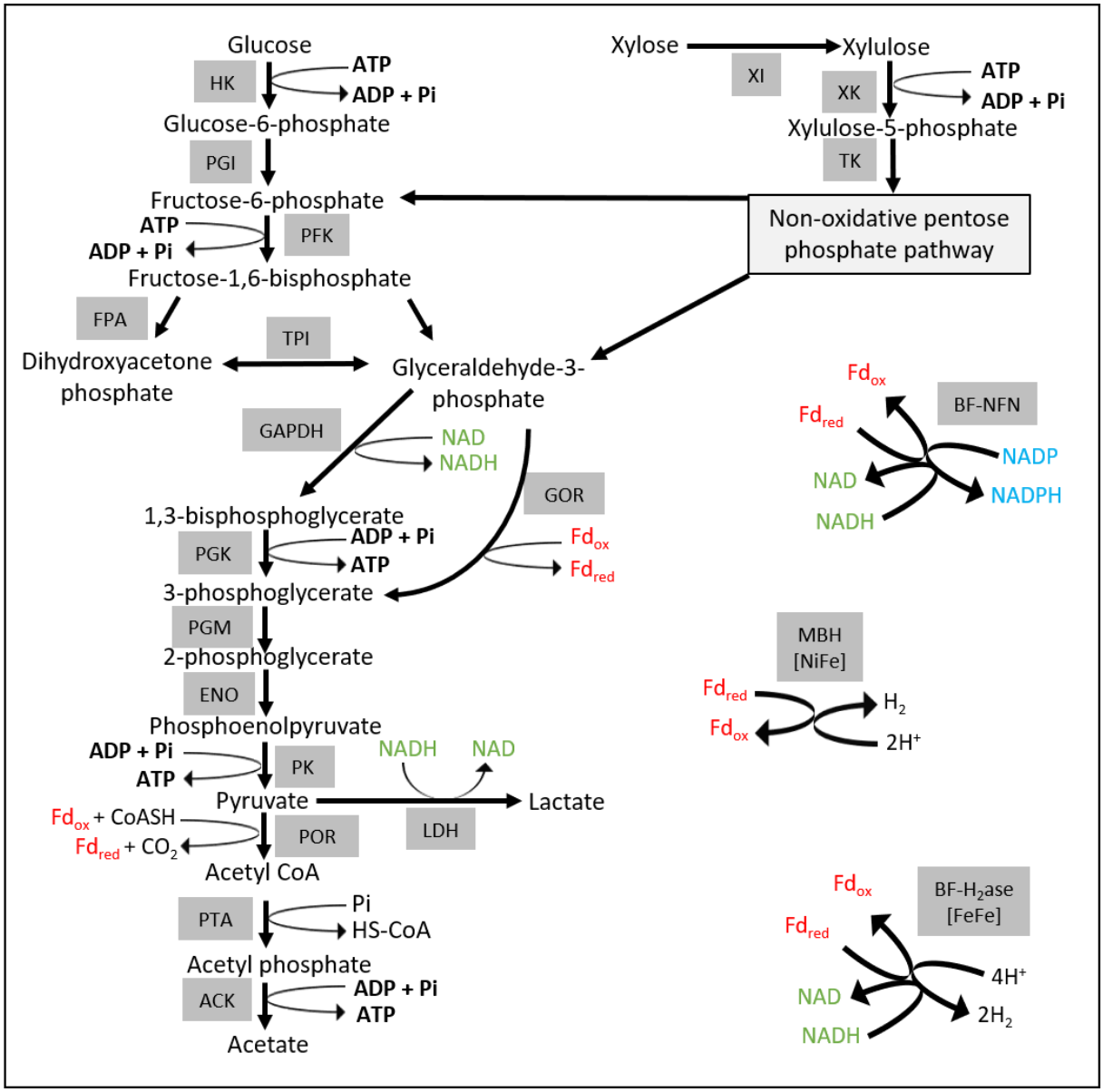
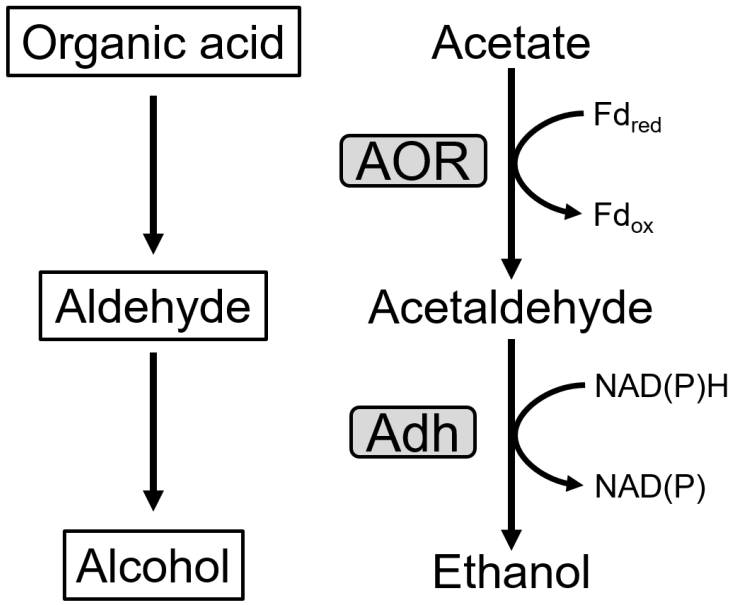

Glucose
HK
ATP
ADP + Pi
Glucose-6-phosphate
PGI
Fructose-6-phosphate
ATP
ADP + Pi
PFK
Fructose-1,6-bisphosphate
FPA
TPI
Dihydroxyacetone phosphate
Glyceraldehyde-3-phosphate
GAPDH
NAD
NADH
GOR
1,3-bisphosphoglycerate
PGK
ADP + Pi
ATP
3-phosphoglycerate
PGM
2-phosphoglycerate
ENO
Phosphoenolpyruvate
ADP + Pi
ATP
PK
NADH
NAD
Pyruvate
Lactate
LDH
Fdox + CoASH
Fdred + CO2
POR
Acetyl CoA
PTA
Pi
HS-CoA
Acetyl phosphate
ACK
ADP + Pi
ATP
Acetate
Xylose
XI
Xylulose
XK
ATP
ADP + Pi
Xylulose-5-phosphate
TK
Non-oxidative pentose phosphate pathway
Fdox
Fdred
BF-NFN
NADP
NAD
NADPH
NADH
MBH [NiFe]
Fdred
Fdox
H2
2H+
Fdox
Fdred
BF-H2ase [FeFe]
4H+
NAD
2H2
NADH

Figure 1.6 The AOR-Adh pathway for alcohol production

The two-step AOR-Adh pathway converts acetate to the ethanol, through acetaldehyde as an intermediate. The source of acetate is variable, depending on the organism. The pathway utilizes two reducing equivalents: reduced ferredoxin for the first step and either NADH or NADPH for the second step, depending on the specificity of the Adh homolog. Due to the broad substrate specificity of these enzymes, the pathway can be generalized to convert organic acids to the corresponding alcohol through the relevant aldehyde intermediate. Abbreviations: reduced ferredoxin (Fd_{red}), oxidized ferredoxin (Fd_{ox}), aldehyde oxidoreductase (AOR), alcohol dehydrogenase (Adh).

Figure 1.6


Organic acid
Aldehyde
Alcohol
Acetate
AOR
Fd$_{red}$
Fd$_{ox}$
Acetaldehyde
Adh
NAD(P)H
NAD(P)
Ethanol

CHAPTER 2

THE THERMOPHILIC BIOMASS-DEGRADING BACTERIUM *CALDICELLULOSIRUPTOR BESCII* UTILIZES TWO ENZYMES TO OXIDIZE GLYCERALDEHYDE-3-PHOSPHATE DURING GLYCOLYSIS[1]

---

[1] **Rubinstein, G.M.** †, Scott, I.S. †, Poole, F.L., Lipscomb, G.L., Schut, G.J., Williams-Rhaesa, A.M., Stevenson, D.M., Amador-Noguez, D., Kelly, R.M., Adams, M.W.W. 2019. *Journal of Biological Chemistry.* 294(25):9995-10005

†Authors contributed equally to this work

ABSTRACT

Caldicellulosiruptor bescii is an extremely thermophilic, cellulolytic bacterium with an optimal growth temperature of 78 °C and is the most thermophilic cellulose degrader known. It is an attractive target for biotechnological applications, but metabolic engineering will require an in-depth understanding of its primary pathways. An analysis of its genome uncovered evidence that *C. bescii* may have a completely uncharacterized aspect to its redox metabolism involving a tungsten-containing oxidoreductase of unknown function. Herein we describe the purification and characterization of this new member of the aldehyde ferredoxin oxidoreductase (AOR) family of tungstoenzymes and show that it is a novel heterodimeric glyceraldehyde-3-phosphate (GAP) ferredoxin oxidoreductase (GOR) present not only in all known species of *Caldicellulosiruptor* but also in 44 mostly bacterial, anaerobic genera. This enzyme is phylogenetically distinct from the monomeric GAP-oxidizing enzyme found previously in several Archaea. Its large subunit (GOR-L) contains a single tungstopterin site and one [4Fe-4S] cluster while the small subunit (GOR-S) contains four [4Fe-4S] clusters. The enzyme uses ferredoxin as an electron acceptor. Deletion of either subunit results in a distinct growth phenotype on both C_5- and C_6-sugars with an increased lag phase but higher cell densities. Using metabolomics and kinetic analyses we show that GOR functions in parallel with the conventional GAP dehydrogenase providing an alternative ferredoxin-dependent glycolytic pathway. These two pathways likely facilitate the recycling of reduced redox carriers (NADH and ferredoxin) in response to environmental H_2 concentrations. This metabolic flexibility has distinct implications for the future engineering of this and related species.

Introduction

Caldicellulosiruptor bescii is an extremely thermophilic bacterium with an optimal growth temperature of 78 °C, making it the most thermophilic cellulose degrader identified to date (268, 322, 323). This organism is an attractive target for use in biotechnological applications due to its ability to ferment high concentrations of unpretreated plant biomass, thereby potentially mitigating the costly thermochemical pretreatment steps that are typically required for lignocellulose conversion (159, 255, 324). *C. bescii* and other species of the *Caldicellulosiruptor* genus do not contain the well-characterized cellulosomes described for other cellulolytic bacteria, such as *Clostridium thermocellum,* but rather secrete a suite of multidomain non-cellulosomal glycosyl hydrolases that work synergistically to convert plant biomass into fermentable sugars (186, 324). *C. bescii* ferments C_6 and C_5 sugars derived from cellulose and hemicellulose primarily to hydrogen gas, lactate, acetate and carbon dioxide (257, 325). The development of a genetic system for this organism has created the potential to metabolically-engineer it for consolidated biomass processing to produce liquid fuels and other bio-based products from renewable biomass. Engineered strains of *C. bescii* that can produce moderate amounts of ethanol directly from the conversion of plant biomass have already been achieved (270, 292). The recent developments of a high temperature kanamycin resistance gene and of an inducible promoter have further increased the potential use of this organism as a platform for biotechnological applications (242, 267, 274, 326). The further development of efficient metabolic engineering strategies will require an in-depth understanding of its primary metabolism.

We recently reported that *C. bescii* has a completely uncharacterized aspect to its primary redox metabolism involving the metal tungsten (272). This was very surprising; tungsten is an element that is rarely utilized in biological systems, as virtually all forms of life utilize the

chemically-analogous element molybdenum (327). Molybdoenzymes are ubiquitous, serving roles in carbon, nitrogen and sulfur metabolism (328). In contrast, the number of microorganisms known to require tungsten is relatively few (327). Tungsten is chemically very similar to molybdenum with comparable coordination chemistries and physical properties. They are both coordinated to enzymes via pyranopterin cofactors and they both stabilize three redox states (IV, V and VI) in the biological range, although for tungsten these are at much lower reduction potentials than those of molybdenum (329). In spite of this, there is a large bias towards molybdoenzymes in all three domains of life, whereas tungstoenzymes appear to be more abundant in thermophilic organisms (330).

Tungsten metabolism is best characterized in hyperthermophilic archaea, represented by *Pyrococcus furiosus*, which grows optimally at 100 °C. It contains five members of the aldehyde oxidoreductase (AOR) family of tungstoenzymes, which are phylogenetically unrelated to any of the three families of molybdoenzymes (331). All five have been characterized and they are referred to as aldehyde ferredoxin oxidoreductase (AOR) (311), glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) (332) and formaldehyde ferredoxin oxidoreductase (FOR) (333), together with two tungstoenzymes of unknown function, WOR4 (334) and WOR5 (335). All of these enzymes oxidize aldehydes of various types. The prototypical AOR has a broad substrate specificity and is thought to be involved in peptide catabolism wherein it oxidizes amino acid-derived aldehydes (327). However, of all of these archaeal tungstoenzymes only the function of GAPOR is definitively understood (332). GAPOR replaces the conventional NAD-dependent glyceraldehyde-3-phosphate (GAP) dehydrogenase (GAPDH) and is part of a modified glycolytic pathway in *P. furiosus* and several archaeal species (332, 336, 337). In addition to using ferredoxin rather than NAD as an electron acceptor, the GAPOR reaction oxidizes GAP directly to 3-phosphoglycerate rather than to the 1,3-bisphosphoglycerate generated by GAPDH (332). Hence, this unusual GAPOR-containing

glycolytic pathway found in these archaea does not contain phosphoglycerate kinase and consequently GAP oxidation does not result in net ATP synthesis by substrate level phosphorylation.

The notion of *C. bescii* utilizing tungsten came from the identification of a 12-gene cluster of highly expressed genes (Athe_0820-Athe_0831) (**Figure S2.1**) that is completely conserved across all fourteen members of the *Caldicellulosiruptor* genus that have been characterized to date (247). These genes encode proteins involved in pyranopterin synthesis and tungstate transport, and a putative tungstopterin-containing enzyme (Athe_0821) belonging to the AOR family of tungstoenzymes that we termed XOR to signify its uncharacterized status. To demonstrate that *C. bescii* could utilize tungsten, the known tungstoenzyme AOR of *P. furiosus* was heterologously-produced in an active form (272). Additionally, we showed that during growth on simple and complex C_6 and C_5 sugars, Athe_0821 is constitutively expressed and its protein product represents ~ 2% of the cytoplasmic protein in *C. bescii.* XOR is phylogenetically distinct from the characterized AOR enzymes and was proposed to represent a new member of this family of tungstoenzymes (272). XOR is also unusual in that the gene encoding (Athe_0821) is adjacent to another highly-expressed gene (Athe_0820) encoding a polyferredoxin-like protein predicted to contain four [4Fe-4S] clusters. It was not known if this functioned as an electron carrier for XOR or was an additional subunit.

It was therefore of importance to elucidate the function of this tungstoenzyme in the *Caldicellulsiruptor* genus, both in terms of understanding its primary metabolism and in utilizing this information to engineer these organisms for biotechnological applications. In contrast to the archaeon *P. furiosus, C. bescii* does not grow on peptides and its genome encodes the standard Embden-Myerhoff (EM) pathway, including a classical phosphorylating GAPDH (Athe_1406). Moreover, cell-free extracts of *C. bescii* did not catalyze the oxidation of any of the aldehydes, including GAP, used by the AOR-type enzymes of *P. furiosus* (272). Herein we show that XOR is

indeed a new member of the AOR family of tungstoenzymes and is a novel bacterial-type of glyceraldehyde-3-phosphate ferredoxin oxidoreductase. We will refer to it as GOR to differentiate it from the phylogenetically-distinct GAPOR found primarily in Archaea. Affinity-tagged GOR was purified and kinetically characterized. Utilizing a deletion mutant strain lacking GOR, metabolomics was used to show that GOR provides *C. bescii* with an alternative glycolytic pathway, generating reduced ferredoxin rather than NADH. This is the first report of a–functional "GAPOR" in a bacterium.

Results

Purification of affinity-tagged GOR

The genes encoding XOR (Athe_0821) and the putative polyferredoxin (PFD, Athe_0820) were modified by the inclusion of a polyhistidine tag at the N-terminus of XOR and the expression of both genes was placed under the control of the high level promoter of the S-layer protein (*slp*) to give the overexpression strain (OE-XOR/PFD, **Figure S2.1**). The *slp* gene is expressed at a level that is ~10-fold higher than that of Athe_0821 and of Athe_0820 in wild type *C. bescii* (272). XOR was purified from the cytoplasmic extract of the OE-XOR/PFD strain by ion exchange, affinity and size exclusion chromatography and its properties are described below. The first priority was to determine a catalytic activity for the purified enzyme. XOR was assayed for its ability to oxidize a variety of aldehydes and reduce the artificial electron acceptor benzyl viologen (BV). These included the following compounds: 2-ethylhexanal, 2-methylbutyraldehyde, 2-methylpentanal, 2-methoxybenzaldehyde, 2-naphthaldehyde, 2-phenylpropionaldehyde, 3-phenylbutyraldehyde, 3,4-dihydroxybenzaldehyde, acetaldehyde, acetoin, arabinose, butyraldehyde, cinnamaldehyde, coniferaldehyde, , dihydroxyacetone, galactose, glucose, glucose-6-phosphate, glyceraldehyde, glycoaldehyde, glycolic acid, glyoxal, glyoxylic acid, hexanal, isobutyraldehyde, mannose, methyl

glyoxal, phenylpropionaldehyde, ribose, sinapaldehyde, syringaldehyde, vanillin, and xylose. However, none of them were utilized as a substrate by XOR. We had previously shown that a cytoplasmic extract of wild type *C. bescii* was unable to oxidize several other aldehydes that are known substrates of the AOR family members in the archaea (272). These included formaldehyde, propionaldehyde, crotonaldehyde, glutaraldehyde, isovaleraldehyde, benzaldehyde and glyceraldehyde-3-phosphate (GAP). Similarly, purified XOR was unable to use these substrates, with the notable exception of GAP. Surprisingly, GAP (1 mM) was readily oxidized by purified XOR with a specific activity (at 70 °C) of 48.3 units/mg and the activity increased with increasing concentrations of XOR in the assay (**Table S2.1**). However, contrary to expectation, only a very low level of GAP oxidation activity was detected in the cytoplasmic extract of the OE-XOR/PFD strain (0.01 units/mg). As previously reported, the same was also observed with extract from the wild type strain (< 0.01 unit/mg). Because the purified XOR enzyme clearly utilizes GAP as a substrate, we will now refer to XOR as a GAP oxidoreductase or GOR.

The lack of significant benzyl viologen (BV)-linked GAP-oxidation activity in the cytoplasmic extract of the wild type strain appeared to be due to the presence of an inhibitor of GOR activity. For example, as shown in **Table 2.1**, the total GAP-oxidizing units increased ~4-fold following the affinity purification of GOR. This was accompanied by a 77-fold purification yielding 2.1 mg of purified protein from 50 g (wet weight) of cells. A similar response to the partial purification from the cytoplasmic fraction was observed with GAPDH, where a greater than 100-fold increase in the NAD-dependent oxidation of GAP was observed after an anion exchange chromatography step (**Figure S2.2**). The addition of the cytoplasmic extract to purified GOR resulted in a ~90% reduction in its GAP-oxidizing activity (**Figure 2.1**) indicating that the cytosol contains an inhibitory factor. To investigate the nature of this inhibitory factor, the cytoplasmic

extract from the OE-XOR/PFD strain was fractionated using a 30-kDa filter (and the protein fraction < 30 kDa was washed with buffer). The washed extract (nominally > 30 kDa) and the resulting filtrate (< 30 kDa) were each tested for their capacity to inhibit the GAP oxidizing activity of purified GOR (**Figure 2.1**). The inhibitor appeared only in the high molecular weight fraction (> 30 kDa) as the low molecular weight fraction (< 30 kDa) did not have any inhibitory effect on GOR (**Figure 2.1**). Moreover, heat treatment (95 °C for 60 min) of the washed cytoplasmic extract eliminated the inhibitory effect, indicating that the inhibitor is likely to be a protein.

Based on its ability to inhibit purified GOR activity, the inhibiting factor was purified from the flow-through fraction of the GOR affinity purification step by hydrophobic interaction chromatography. Analysis of the purified inhibitor fraction by liquid chromatography-tandem mass spectrometry (LC-MS) (Proteomics and Mass Spectrometry Facility, University of Georgia) identified a 70 kDa bifunctional phosphoglycerate kinase/triosephosphate isomerase (PGK/TPI) fusion protein (Athe_1405). Addition of the purified PGK/TPI protein to the GOR assay virtually abolished GAP-oxidizing activity (**Figure 2.1**).

TPI catalyzes the interconversion of GAP and dihydroxyacetone phosphate (DHAP) with a strong bias in favor of DHAP (~95:5) (338). Hence, TPI activity is responsible for the inhibition of GOR and GAPDH in cytoplasmic extracts by a "substrate-stealing" mechanism and thus its separation from these enzymes during their purification results in a dramatic increase in their GAP-dependent activities. The presence of the PGK/TPI protein in the standard GOR assay enabled reduction of BV to be measured in the presence of DHAP (which generates GAP) in the absence of added GAP (**Table S2.2**). Likewise, the presence of PGK/TPI in the standard GAPDH assay enabled the reduction of NAD if DHAP (rather than GAP) was added (**Table S2.2**). Interestingly, DHAP itself partially inhibits GOR activity in the absence of PGK/TPI and, conversely, increases GAPDH

activity in the absence of PGK/TPI (**Table S2.2)**. GOR activity is reduced by 70% in the presence of 10 mM DHAP in the absence of TPI, but only decreases by 22% in the presence of 1 mM DHAP. In contrast, GAPDH activity increased 4-fold in the presence of 10 mM DHAP. Notably, TPI inhibition of both can be overcome by sufficiently high concentrations of GAP (5 and 15 mM) as measured by the initial activities (**Table S2.2)**. However, in all cases where PGK/TPI was present, activity diminished within 15 seconds, compared to >30 seconds in the standard assays.

A kinetic analysis of purified GOR with varying concentrations of GAP revealed a linear increase in activity, with no saturation observed up to 20 mM GAP with BV (1 mM) as the electron acceptor. While this data precludes the determination of K_m and k_{cat} values, k_{cat}/K_m was determined as the slope of the plot of activity versus substrate concentration to be 4.7×10^4 sec^{-1} M^{-1} at 70 °C (**Figure S2.3A**). However, with *C. bescii* ferredoxin as the electron carrier, saturation kinetics were observed, with an estimated k_{cat} value of 5.57 s^{-1} and a K_m value for oxidized Fd of 38.9 ± 10.7 µM at 70 °C (using 5 mM GAP) (**Figure S2.4**). Like GOR, partially purified GAPDH also showed a linear increase in activity with increasing GAP concentrations utilizing NAD (1 mM) as the electron acceptor, with no saturation observed at GAP concentrations up to 20 mM (**Figure S2.3B**).

Size exclusion chromatography (SEC) of affinity purified GOR showed that GOR eluted as two distinct symmetrical peaks that had the same specific activities with approximate molecular weights of 315 and 73 kDa (**Figure S2.5**). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) revealed identical profiles for the peaks with two major bands near 60 and 15 kDa. These corresponded to the expected gene products of Athe_0821 (65,916 Da) and Athe_0820 (14,618 Da), which will now be referred to as the large (GOR-L) and small (GOR-S) subunits of GOR. The identity of the two subunits was confirmed by peptide analysis using LC-MS (**Figure S2.6**). We propose that the two peaks from the SEC column represent the holoenzyme

(L_4S_4), which is a tetramer of heterodimers, and the dissociated heterodimer (LS, **Figure S2.5**). GOR-L has 29% sequence identity (47% similarity) to *P. furiosus* GAPOR (which is 67 aa larger) and is 31% identical (48% similar) to the structurally-characterized AOR of *P. furiosus*, which contains one tungstopterin site and one [4Fe-4S] cluster. The cluster coordinating cysteine residues and the tungstopterin-binding residues of AOR are conserved in GOR-L and in GAPOR (although they lack the residues that coordinate the monomeric iron atom in AOR, see **Figure S2.7**). GOR-S, annotated as a putative polyferredoxin, contains 16 cysteinyl residues in four motifs and these are predicted to coordinate four [4Fe-4S] clusters (**Figure S2.7**) (339). Metal analysis of purified GOR yielded an iron to tungsten ratio (Fe/W) of 20.4:1, consistent with one and four [4Fe-4S] clusters in GOR-L and GOR-S, respectively, together with a monomeric tungstopterin site in GOR-L.

Deletion of the genes encoding GOR

To assess the functional importance of GOR, mutants were generated by individually disrupting *gorL* (Athe-0821) and *gorS* (Athe_0820) by replacing each with the gene encoding kanamycin resistance (Cb*htk*) (**Figure S2.1**) to generate the mutant strains *ΔgorL* and *ΔgorS* (**Table S2.3**). As shown in **Figure 2.2**, compared to the parent, growth on xylose (C_5) of both mutant strains exhibited a longer lag phase. Although the exponential growth rates were comparable (**Figures 2.2 and S2.8**), both mutant strains reached a significantly higher cell density in stationary phase – up to 50% higher than that of the parent. Similar results were obtained during growth on other sugars such as: glucose, fructose, mannose, galactose, rhamnose, arabinose, and lactose (**Figure S2.8**).

Metabolomic analyses of the Δgor strain

To provide further insight into the metabolic role of GOR, MS-based metabolomics were employed to evaluate changes in the concentrations of key metabolites that result from the deletion of g*orL*. A total of 19 intermediary metabolites were found to have significant changes in

concentration in the Δ*gorL* strain relative to the parent strain (**Table S2.4**), the most notable of which were intermediates of the glycolytic and pentose phosphate pathways. As shown in **Figure 2.3**, concentrations of glucose-6-phosphate, fructose-6-phosphate and fructose-1,6-bisphosphate were higher in the in the Δ*gor* strain while those of 3-phosphoglycerate and phosphoenolpyruvate were lower, consistent with a bottleneck in the glycolytic pathway at the GAP oxidation step.

Phylogenetic analyses of GOR

A phylogenetic tree of the AOR family of enzymes was constructed using 4,063 homologs of the GOR-L sequence based on the InterPro domains (IPR013983 and IPR001203). Although the five previously characterized types of tungsten-containing oxidoreductases, AOR, FOR, GAPOR, WOR4 and WOR5, are all from Archaea such as *P. furiosus*, the AOR family is also well represented in the bacterial domain with 64% of the sequences from Bacteria and 30% from Archaea, with the remaining (6%) from uncharacterized metagenomes or environmental samples (**Figure S2.9**). The five types of previously characterized members of the AOR family from *P. furiosus* form distinct clades that are well separated from each other and *C. bescii* GOR is clearly more closely related to GAPOR than it is to any of the other enzymes, sharing a more recent common ancestor. Nevertheless, the GOR and GAPOR branches also sub-divide into bacterial and archaeal sub-clades (**Figure S2.9B**). Hence, the phylogeny supports distinctive terms for GOR and GAPOR, especially as they are found predominantly in Bacteria and Archaea, respectively.

Discussion

Herein we describe the first glyceraldehyde-3-phosphate ferredoxin oxidoreductase from a bacterium and designate it GOR to show that is it is distinct from the previously characterized GAPOR from archaea (272). We also demonstrate that GOR is directly involved in the glycolytic pathway of *C. bescii*. This result was unexpected since the genome of *C. bescii* encodes the standard

pathway that includes the classical phosphorylating GAPDH. This catalyzes the conversion of GAP to 1,3-bisphosphoglycerate, which is then used to conserve energy in the form of ATP by PGK generating 3-phosphoglycerate. Note that GOR oxidizes GAP directly to 3-phoshoglycerate and so ATP is not synthesized in the GOR-dependent glycolytic pathway. Both GAPDH and PGK/TPI enzymes were partially purified in this study indicating that they are present at substantial concentrations in the cell. However, the results of our genetic and metabolomics-based studies clearly indicate an important glycolytic role for GOR under the usual growth conditions. Moreover, deletion of either of the genes encoding GOR affects the efficient oxidation of GAP, as evident from the accumulation of metabolic intermediates upstream of this step in the glycolytic pathway and depletion of intermediates downstream (**Figure 2.3**).

Lack of saturation kinetics for both GOR and GAPDH is a highly unusual feature for enzymes of this type. While there are other examples of non-saturable enzymes (e.g. AhpC from *Helicobacter pylori*) this phenomenon is rare (340). In the case of GOR, the order of k_{cat}/K_m ($4.7x10^4$ sec^{-1} M^{-1}) indicates that the lack of saturation is not because the enzyme is diffusion limited. Instead, it likely has an extremely low affinity for GAP but with a high turnover rate, such that formation of the [ES] complex is slow, but turnover occurs rapidly upon complex formation. Non-saturation kinetics is even more surprising for GAPDH. This enzyme has been extremely well studied from many organisms with K_m values for GAP ranging from 20 μM to 1.4 mM (341, 342). However, in an organism like *C. bescii* with two competing primary metabolic pathways for GAP oxidation, it is not unreasonable that the two GAP-oxidizing enzymes have low affinities for the substrate as a means of regulating substrate flux. It is also possible that another component is necessary in the assay for these enzymes to be saturable with GAP, as in the case with human ADH5 (Class III

alcohol dehydrogenase), which cannot be saturated with ethanol (up to 3 M) in the absence of activation by fatty acids (343).

We show here that both GAPDH and GOR activities are masked by the presence of a highly active TPI, which is biased towards DHAP formation thereby maintaining only a relatively low concentration of GAP that is available to both enzymes. The high TPI activity might play a unique thermoprotective role by minimizing the pool of the thermally-labile glycolytic intermediate GAP by reversibly converting it to the more thermostable DHAP, thus regulating the glycolytic flux through GOR and GAPDH (344). This in contrast to the situation in some archaea where GAPOR is the sole GAP-oxidizing enzyme in a modified EM pathway generating 3-phosphoglycerate in a single irreversible step, where GAPDH serves solely a gluconeogenic role (332, 345, 346). The concurrent presence of GOR and GAPDH activity suggest that in *C. bescii* both enzymes operate in parallel. The operation of parallel pathways in glucose degradation is unusual, although not without precedent in other thermophilic microorganisms. For example, the archaeon *Thermoproteus tenax* has been demonstrated to utilize modified versions of both EMP and ED pathways (347-350).

The inhibition of the activity of GOR by DHAP, while GAPDH appears to be activated by DHAP (**Table S2**) was unanticipated. However, it does explain our inability to measure significant GAP-dependent oxidoreductase activity (BV-dependent) in cytoplasmic extracts previously That DHAP is a regulator of both GOR and GAPDH cannot be ruled out as it could potentially activate GAPDH while inhibiting GOR, but further study will be required to clarify this, especially since high and likely non-physiological concentrations of DHAP (10 mM) were required to observe significant effects.

If the inhibition of GAPDH activity in cytosolic extracts results from PGK/TPI activity, it would be expected that this effect should have been observed in the purification of GAPDH from

other organisms where TPI is present. That this situation in *C. bescii* is extremely unusual is shown by comparison with a related anaerobic fermentative thermophile, *Thermotoga maritima* (T_{opt} 80 °C). Like *C. bescii*, *T. maritima* also contains a PGK/TPI fusion protein but its GAPDH behaves like a conventional enzyme. It shows no inhibition in cytoplasmic extracts and has been purified by conventional chromatography with no significant increases in activity (351). Notably, *T. maritima* lacks GOR. Thus, we hypothesize that the PGK/TPI effect is related to the presence of GOR in *C. bescii.* Similarly, this effect may be related to the lack of saturation kinetics discussed above as a means of regulating the dynamic pathways for GOR oxidation present in this organism.

We speculate that the role of GOR in *C. bescii* and other *Caldicellulosiruptor* species is to serve as an alternative glycolytic pathway (**Figure 2.4**) that is used by the cell to regulate carbon and electron flux during growth under conditions in which *C. bescii* is unable to efficiently recycle its electron carriers. This might be particularly important when *C. bescii* is exposed to excess carbon loading, such as during growth on unprocessed biomass (257). For example, the genome of *C. bescii* encodes two hydrogenases, one using only reduced ferredoxin as the electron donor while the other simultaneously uses both reduced ferredoxin and NADH. The latter is a cytoplasmic ferredoxin/NADH-dependent bifurcating [FeFe] hydrogenase that is the primary source of H_2 production during growth on glucose-based substrates (271). The former is a membrane bound ferredoxin-dependent [NiFe] hydrogenase (Ech) that can generate a proton gradient across the cell membrane (352). The standard EM pathway utilizing the classic phosphorylating GAPDH is likely the primary glycolytic route used by *C. bescii* as this pathway produces ATP. While ferredoxin-dependent conversion of GAP to 3-phosphoglycerate by GOR yields less ATP than the standard EM pathway, it allows glycolysis to proceed with only ferredoxin as the redox carrier. This can be recycled via H_2 production by Ech, avoiding the potential generation of excess NADH. This

hypothesis is supported by the finding that genes necessary for the biosynthesis of GOR are upregulated in the closely-related *C. saccharolyticus* in the presence of high H_2 concentrations (353), along with the recent discovery that GOR (XOR) is regulated by Rex, a redox-sensing transcription factor that responds to the intracellular NADH/NAD ratio (354). These findings suggest that high concentrations of NADH or H_2, either of which could interfere with the regeneration of oxidized cofactors via the bifurcating hydrogenase (274, 353), steer carbon flux towards GOR to utilize the more thermodynamically favorable Ech pathway thereby maintaining a high rate of H_2 production (274, 353). However, it is not possible at present to reconcile such an analysis with the observed growth phenotype of the Δ*gor* mutant strains, which exhibit a longer lag phase but higher cell densities compared to the parent strain (**Figures 2.2 and S2.8**).

Phylogenetic analyses of the AOR family of enzymes (**Figure S2.9**) shows that each of the five characterized members of the AOR family of enzymes from *P. furiosus* can be assigned to distinct clusters, which is in accordance with their function in terms of both their substrate specificities and oligomeric structures. Hence AOR from *P. furiosus* and also from *Pyrobaculum aerophilum* are homodimeric enzymes and catalyze the oxidation of a wide range of both aliphatic and aromatic aldehydes (339, 355). FOR from *Thermococcus litoralis* and *P. furiosus* are homotetrameric enzymes but with a more limited substrate range than AOR, oxidizing short chain aliphatic aldehydes (333, 356). WOR4 and WOR5, both from *P. furiosus,* are also homodimeric enzymes. WOR5, like AOR, can oxidize a wide range of aldehydes but the substrate for WOR4 is not known and this enzyme was isolated purely based on its tungsten content (335, 357). GAPORs from *Methanococcus maripaludis*, *P. furiosus*, and *P. aerophilum* are monomeric enzymes and are absolutely specific for glyceraldhyde-3-phosphate as they are not known to oxidize any other aldehyde (332, 336, 337). We show here that *C. bescii* GOR-L is part of a distinct clade, separate

from other members of the AOR family, including GAPOR (**Figure S2.9**), its closest characterized homolog. The GOR-L clade contains all of the *Caldicellulosiruptor* species plus 44 other anaerobic genera, representing 9 taxonomic classes: Archaeoglobi, Bacilli, Clostridia, Deltaproteobacteria, Fusobacteriia, Methanomicrobia, Nitrospira, Thermodesulfobacteria, and Thermoprotei. While many of the genera are thermophilic or hyperthermophilic some are mesophilic; however, almost all species with GOR-L also have a co-located GOR-S homolog (SynTax web server) (358). Moreover, this analysis shows that, unlike GAPOR, which is found primarily in Archaea, GOR is found primarily in Bacteria. For example, the GOR-L clade could be divided into three sub-clades representing a bacterial-only clade, making 54% of all GOR-L clade sequences, an archaeal-only making up 26% and a mixed clade (split 50/50) making up 20% (**Figure S2.9**).

We also show here that GOR is part of a small minority of AOR family members that are heteromeric. These include a homolog of AOR (termed Aa AorB) that is a heterotrimer containing the equivalent of the GOR-S subunit with a string of iron-sulfur clusters (359). In addition, a second heteromeric member of the AOR family described recently is part of an enzyme complex that reduces benzoyl CoA (360). Termed BamB, the AOR homolog also has a partner subunit (BamC) with multiple iron-sulfur clusters. Furthermore, of the characterized AOR family members from hyperthermophilic archaea, the gene encoding WOR5 has an adjacent gene homologous to that encoding GOR-S, suggesting that WOR5, like GOR, has an additional iron-sulfur containing subunit that was likely not detected during standard SDS analysis. Finally, the heteromeric AOR family group also includes an AOR from *E. coli* termed Ec YdhV that contains a polyferredoxin-like subunit similar to GOR-S. Interestingly, YdhV contains molybdenum rather than tungsten as *E. coli* does not metabolize tungsten (361). It is not clear why GOR and these other AOR family members require an additional subunit with four [4Fe-4S] clusters, which in GOR are proposed to transfer electrons

between the catalytic subunit (GOR-L) and ferredoxin. GOR has a high affinity for *C. bescii* ferredoxin (K_m ~ 40 μM), consistent with this being the physiological electron acceptor, although it is possible that GOR could provide reductant to the Ech hydrogenase directly while ferredoxin is required to interact with the bifurcating hydrogenase.

In any event, the results presented here lay the foundation for a more comprehensive understanding of the two glycolytic pathways that are present in the cellulolytic genus *Caldicellulosiruptor* and for more effective metabolic engineering projects in the future. Moreover, this work suggests that alternative glycolytic pathways using either GOR or GAPOR may be more widespread than previously thought in both the bacterial and archaeal worlds.

Experimental procedures

Growth of C. bescii

C. bescii strains used or constructed in the present study are listed in **Table S2.3**. Strains were transformed as previously described (272, 275). For growth of uracil auxotrophic strains, uracil was added at a concentration of 20 μM. When required, kanamycin and 5-fluoroorotic acid (5-FOA) were used at concentrations of 50 μg/mL and 4 mM, respectively unless otherwise noted. Growth experiments were performed with 5 mL cultures in screw cap Hungate tubes sealed with butyl rubber stoppers in modified DSMZ 516 medium with the following composition per liter: 1× salt solution, 1× vitamin solution, 1× trace element solution, 0.16 μM sodium tungstate, 0.25 mg resazurin, 5 g xylose (unless otherwise noted), 1 g cysteine hydrochloride, 1 g sodium bicarbonate, and 1 mM potassium phosphate buffer (pH 7.2). The 50× stock salt solution contained the following per liter: 16.5 g NH_4Cl, 16.5 g KCl, 16.5 g $MgCl_2 \cdot 6H_2O$ and 7 g $CaCl_2 \cdot 2H_2O$. Stock solutions of 200× vitamins and 1,000× trace elements were prepared as previously described (268). Cultures were also

grown in a 15 L fermenter at 75 °C with 20 g/L xylose as the growth substrate. The culture was sparged at 2 L/h with 80% (v/v) nitrogen and 20% (v/v) carbon dioxide and stirred at 150 rpm with flat blade impeller. The pH was maintained at 7.0 (25 °C) by the addition of 10% (w/v) sodium bicarbonate. Cells were harvested in the late-exponential phase by centrifugation at 6,000 *g* for 10 min (Beckman Avanti J-30I JLA 10.500 rotor) and the resulting cell pellets were flash frozen in liquid nitrogen and stored at 80 °C. The cell yield from the fermenter was approximately 12 g (wet weight).

Vector construction

Plasmids were constructed from PCR products using NEBuilder® HiFi DNA Assembly Master Mix (New England BioLabs). PCR was performed using Takara PrimeSTAR Max DNA polymerase (Takara R045A). Genomic DNA, plasmid DNA, and PCR products were purified with kits from Zymo Research and Stratagene. The templates for PCR were either previously sequenced plasmids or genomic DNA from *C. bescii*. All primers used in this study can be found in **Table S2.5** and plasmids constructed for this study can be found in **Figure S2.10**. Plasmids were cloned into NEB 10-beta competent *E. coli* (New England BioLabs C3019I), followed by Sanger sequencing through GeneWiz.

Metabolomics

For the MS-based metabolomics, parent and *ΔgorL* strains of *C. bescii* were grown in duplicate in 16 mm sealed Hungate tubes in modified DSMZ 516 medium as described above with glucose as the carbon source. Cultures were grown to an OD_{600} of ~0.15 and each culture was deposited by vacuum filtration onto a 0.2 μm nylon membrane (47 mm diameter). The membrane was then placed (cells down) into 1.5 mL cold (−20 °C) on dry ice extraction solvent (20:20:10 v/v/v acetonitrile:methanol:water) in a small petri dish and swirled. After a few moments the filter was

inverted (cells up) and solvent was passed over the surface of the membrane several times to maximize extraction. Finally, the cell extract was stored at −80 °C. Cell extracts from *E. coli* (K-12 substrain MG1655 rph+ ilvG+) grown in ^{13}C-glucose were used as internal standards for quantitation of metabolites in *C. bescii* strains. *E. coli* cells were grown aerobically to an OD_{600} of ~0.45 in M9 minimal medium containing 0.4% universally labeled ^{13}C-glucose as sole carbon source. To assure complete labeling of metabolites in *E. coli*, inoculation was performed using 1/50 dilution from an overnight culture that was also grown on ^{13}C-labeled glucose. *E. coli* metabolites were extracted in a similar manner to the procedure used for *C. bescii*.

C. bescii and ^{13}C-labeled *E. coli* extracts were thawed, centrifuged at 14,000 rpm for 10 min to remove cell debris and the *C. bescii* supernatants mixed with ^{13}C-labeled *E. coli* supernatant in ratios of 1:1, 5:1 and 1:0 (no ^{13}C labeled internal standards). These mixed samples were then fully dried under a stream of nitrogen and subsequently re-suspended with HPLC running solvent (97:3 water: methanol containing 10 mM tributylamine (TBA) and 9.8 mM acetate, pH 8.2) in half the initial volume. Samples were then centrifuged to remove particulates, transferred to HPLC vials, and analyzed by HPLC/MS for quantification as previously described (362).

Enzyme assays

All assays were carried out in serum-stoppered cuvettes under anaerobic conditions at 70 °C, unless noted otherwise. To remove trace amounts of O_2, sodium dithionite was added to the assay mixture to give an A_{600} of ˜0.2. The extract was added and after a 1-min incubation period the reaction was initiated by addition of the substrate. Aldehyde-oxidizing activity of cytoplasmic extracts of *C. bescii* strains and purified GOR were determined by following the reduction of ferredoxin or benzyl viologen (BV, 1 mM) in 50 mM EPPS buffer, pH 8.0, at 425 or 600 nm, respectively, using various aldehydes (1 mM) as the substrate. With GAP as the substrate the assay temperature was 70 °C.

GAPDH activity was measured by following GAP-dependent reduction of NAD (1 mM) at 320 nm. Extinction coefficients of 8.2 $mM^{-1}cm^{-1}$, 6.2 $mM^{-1}cm^{-1}$, and 8.1 $mM^{-1}cm^{-1}$ were used for reduced benzyl viologen, NADH and reduced ferredoxin, respectively. Specific enzyme activities are expressed as units per mg of protein, where one unit represents 1 µmol of substrate oxidized per min.

Purification of affinity-tagged GOR

All purification steps were performed under strictly anaerobic conditions. Frozen cells (50 g, wet weight) of *C. bescii* strain OE XOR/PFD were thawed and suspended in 25 mM Tris/HCl, pH 8.0, containing 1 mg mL^{-1} lysozyme (Sigma-Aldrich) in a ratio of 3 mL per g of cells. The suspended cells were incubated at room temperature for 15 min followed by four 15-s intervals of sonication (amplitude 40; Qsonica Q55) interspaced by at least 30 sec. Cell lysates were clarified by ultracentrifugation at 100,000 *g* for 1 h (Beckman L90K ultracentrifuge 70.1Ti rotor) yielding clarified cell lysate (235 mL, 14.3 mg protein mL^{-1}). This was loaded onto a 250 mL QHP column equilibrated with buffer A (25 mM Tris, pH 8.0, containing 2 mM cysteine), at a rate of 10 mL min^{-1} diluted 1:5 with buffer A. Protein was eluted with a linear gradient (2500 mL) from 0-500 mM NaCl at 10 mL min^{-1} and collected as 75 mL fractions. Fractions were assayed for GOR and GAPDH activity. Fractions containing GAPDH activity were reserved for further purification while fractions containing GOR activity were pooled and applied to a 5 mL HisTrap Excel column (GE Healthcare) equilibrated in buffer B (50 mM phosphate, pH 7.2, containing 300 mM NaCl). The column was washed with five volumes of buffer B, and GOR was eluted with a gradient of 0 to 500 mM imidazole in buffer C (buffer B containing 500 mM imidazole) over 20 column volumes. Fractions were collected in sealed serum bottles. Fractions containing GOR activity were pooled and concentrated (15 kDa Amicon Ultra centrifugal filter) to yield 2.9 ml of GOR (1.2 mg/ml). This was loaded onto a HiLoad 16/60 Superdex 200 column (GE Healthcare) equilibrated with 50 mM Tris/HCl, pH 8.2,

containing 200 mM NaCl at a rate of 1.25 mL $min^{-1}$and collected in 4 ml fractions in sealed serum bottles. Protein containing fractions were analyzed for purity using denaturing SDS-PAGE gradient electrophoresis using precast 4 - 12% Bis-Tris gels; Novex NuPAGE gels (Invitrogen) and stained with Imperial protein stain (Novex). Protein bands were identified via a Mascot MS/MS Ions Search of a custom *C. bescii* database (version 2.6) using liquid chromatography-tandem mass spectrometry (LC-MS/MS) data (Proteomics and Mass Spectrometry Facility, University of Georgia).

Partial Purification of GAPDH

In order to further purify GAPDH, the fraction containing the highest GAPDH activity from the QHP column was concentrated (10 kDa Amicon Ultra centrifugal filter) to yield 4.1 ml (14.2 mg/ml) and loaded onto a HiLoad 16/60 Superdex 200 and eluted in the same manner as above. Fractions containing GAPDH activity were pooled and applied to a 1 ml HiTrap Blue HP column equilibrated in 20 mM EPPS, pH 8. The column was washed with five column volumes of the equilibration buffer. Proteins were eluted with 2 mL of the equilibration buffer containing 20 mM NAD. The eluate was concentrated and buffer exchanged (10 kDa Amicon Ultra centrifugal filter) into 50 mM EPPS, pH 8 to remove NAD, yielding 0.4 mL of partially purified GAPDH (1.6 mg/mL).

Estimation of molecular mass and metal analysis

For size determination of the purified GOR, the elution volumes (V_e) were determined by measuring the absorption of the peaks at 280 nm and the total void volume (V_o) was determined by measuring the elution of blue dextran. The known molecular weight standards thyroglobulin (669 kDa), apoferritin (443 kDa), amylase (200 kDa), alcohol dehydrogenase (150 kDa), albumin (66 kDa) and carbonic anhydrase (29 kDa) were used as a reference. The molecular mass of the purified protein in each peak was estimated from the plot of V_e/V_o against the logarithm of the molecular mass (kDa). Metal concentrations of the final chromatography fractions were measured

using an Agilent 7900 inductively coupled plasma mass spectrometer (ICP-MS) fitted with MicroMist nebulizer, UHMI-spray chamber, Pt cones and an Octopole Reaction System (ORS) collision cell (Agilent Technologies, Santa Clara, CA) as described previously (363).

Purification of PGK/TPI

In order to isolate the GOR inhibiting factor, the flow through from the affinity purification of GOR described above was applied to a 60 mL Phenyl Sepharose column equilibrated with buffer A containing 1M $(NH_4)_2SO_4$. Protein was eluted with a linear gradient (800 mL) from 1.0 to 0 M $(NH_4)_2SO_4$ at 7.5 mL min^{-1}. The resulting fractions were buffer exchanged with 50 mM EPPS, pH 8.0, using ultrafiltration with a 30 kDa filter and tested for the ability to inhibit GAP (1 mM) oxidation by purified GOR. Fractions exhibiting the ability to inhibit GOR activity were analyzed by SDS-PAGE using precast 4-12% Bis-Tris Novex NuPAGE gels (Invitrogen) and protein bands were identified via MS/MS as described above.

Phylogenetic analyses

In InterPro (version 63.0), GOR-L (Athe_0821; UniProt ID: B9MQI2_CALBD) is annotated to have one aldehyde ferredoxin oxidoreductase N-terminal (IPR013983) domain and one aldehyde ferredoxin oxidoreductase C-terminal (IPR001203) domain (364). One of each domain was entered into the InterPro Domain Architecture (IDA) search tool, yielding 4,107 GOR-L sequence homologs. The 4,107 sequences were aligned using Clustal Omega, version 1.2.1, with the default parameters (365). TrimAl was used to remove multiple alignment positions with greater than 99.7% gaps (i.e. -gt 0.003) (366), and the resulting alignment was used to construct a maximum likelihood phylogenetic tree using IQ-Tree (version 1.5.5) (367). The IQ-Tree standard model selection test was used to automatically determine the best-fit model (i.e. LG+F+R10) (368). This model tree was refined using ultrafast bootstrap approximation (UFBoot) and Shimodaira-Hasegawa-like

approximate likelihood ratio tests with 3,000 bootstrap replicates for each. Only sequences that were not marked deleted by Uniprot were used to display the tree, leaving 4063 sequences. iTOL (version 4.2.1) was used for analysis and display of the phylogenetic tree (369). For visualization the tree was re-rooted and branches with ultrafast bootstrap confidence values less than 94% were removed. Uniprot Retrieve was used to lookup the kingdom of the organism for each of the 4,063 sequences and label them in iTOL (370).

Acknowledgments

We thank Jeffrey Zurawski, Piyum Khatibi, and Laura Lee for many helpful discussions and Daehwan Chung and Janet Westpheling for providing strains and plasmids of *C. bescii*. We also acknowledge the University of Georgia Proteomics and Mass Spectrometry Facility for performing mass spectrometry analysis and computational support. The phylogenetic analysis was made possible by the resources and technical expertise of the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-SC0019391), by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center at Oak Ridge National Laboratory, and by the US National Science Foundation (CBET1264053).

Tables and Figures

Table 2.1 Affinity purification of His-tagged GOR

Step	Protein (mg)	Volume (mL)	Units (U)	Specific Activity (U/mg)	Purification (Fold)	Yield (%)
S100	3361.9	235	33.6	0.01	-	-
Q-HP	518.1	370	339	0.7	1	100
Ni-NTA	2.9	2.4	144	50.1	76	43
Sephedex 200	2.1	2.1	101	48.3	74	30

Table S2.1 Protein concentration dependent activity of GOR

Standard assays were conducted with 1 mM BV in 2 mL of 50 mM EPPS buffer, pH 8.0, in sealed anaerobic cuvettes. Reactions were carried out with varying concentrations of purified His-tagged GOR and reactions were initiated by the addition of GAP. Increasing concentrations of enzyme increased the total BV reduction activity, but specific activity remained constant. Without the addition of GOR, no GAP dependent reduction of BV was observed, demonstrating that GOR is responsible for the measured activity.

Enzyme Assay	[GAP] (mM)	GOR (mg)	Total Activity (U)	Specific Activity (U/mg)
GOR	1	0	0	0.0
GOR	1	0.005	0.31	61.3
GOR	1	0.010	0.62	61.7
GOR	1	0.015	0.92	61.6

Table S2.2 Effect of TPI on standard GAPDH and GOR activities

Standard GOR and GAPDH assays were performed in sealed anaerobic cuvettes in 2 mL of EPPS buffer, pH 8.0. GOR assay included 1 mM BV and 0.005 mg of purified His-tagged GOR. GAPDH assays included 20 mM dipotassium phosphate, 1 mM NAD and 20 µL of partially purified GAPDH (1.6 mg/mL). Reactions marked "+" contained 30 uL of partially purified TPI (10 mg/mL) while those designated "-" did not include TPI (although minor TPI contamination of GAPDH is indicated by background activity when DHAP but no GAP was added). Reactions containing GAP were initiated by the addition of GAP to the assay mixture. Reactions containing TPI but no GAP were initiated by the addition of TPI. Specific activity is expressed per mg of GOR or GAPDH.

Enzyme Assay	TPI	[GAP] (mM)	[DHAP] (mM)	Predicted Activity	Specific Activity (U/mg)
GOR	-	0	10	No	0.0
GOR	-	1	0	Yes	61.3
GOR	-	1	1	Yes	47.6
GOR	-	1	10	Yes	18.1
GOR	+	0	10	Yes	16.7
GOR	+	1	0	No	1.5
GOR	+	5	0	Yes	190.1
GOR	+	15	0	Yes	645.5
GAPDH	-	1	0	Yes	0.9
GAPDH	-	0	10	No	0.1
GAPDH	-	1	10	Yes	4.4
GAPDH	+	0	10	Yes	2.6
GAPDH	+	1	0	No	0.0
GAPDH	+	5	0	Yes	2.2
GAPDH	+	15	0	Yes	23

Table S2.3 Strains used and constructed in this study

Strain	Parent	Genotype	Alias	Reference
MACB1032		*ΔpyrE ΔcbeI*		(242)
JWCB018		*ΔpyrFA ΔcbeI*		(276)
MACB1010	JWCB018	*ΔpyrFA ΔcbeI:*P_{slp}*xor-pfd*	OE-XOR/PFD	This study
MACB1050	MACB1032	*ΔpyrE ΔcbeIΔgorL*	*ΔgorL*	This study
MACB1074	MACB1032	*ΔpyrE ΔcbeIΔgorS*	*ΔgorS*	This study

Table S2.4 MS-based metabolomics of Δ*gorL* and parent strains

Metabolite	Log_2 Fold change[a]	
Homocysteine	3.555 ±	0.363
Ribose 5-phosphate	1.675 ±	0.234
Sedoheptulose 7-phosphate	1.614 ±	0.179
4-hydroxybenzoate	1.478 ±	0.276
Glucose-6-phosphate	1.341 ±	0.097
Fructose 6-phosphate	1.318 ±	0.082
Malate	0.916 ±	0.026
Ribulose 5-phosphate	0.892 ±	0.267
Glycerol-3-phosphate	0.525 ±	0.016
Fructose 1_6-bisphosphate	0.505 ±	0.06
Dihydroxy acetone phosphate	0.485 ±	0.013
Threonine	0.477 ±	0.069
Valine	0.278 ±	0.04
Phenylalanine	0.253 ±	0.015
Myoinositol	0.238 ±	0.005
Glutamate	0.164 ±	0.006
Serine	0.164 ±	0.044
Leucine	0.044 ±	0.001
Succinate	0.002 ±	0
Pyruvate	-0.019 ±	0.001
UDP-N-acetylglucosamine	-0.127 ±	0.005
Phenylpyruvate	-0.177 ±	0.005
Methionine	-0.354 ±	0.029
Aspartic acid	-0.445 ±	0.022
2-ketoglutarate	-0.454 ±	0.031
Tyrosine	-0.772 ±	0.030
Dihydroorotate	-0.820 ±	0.046
Isoleucine	-1.071 ±	0.012
Fumarate	-1.326 ±	0.135
NAD+	-1.407 ±	0.145
Citrate	-1.769 ±	0.022
Quinolinate	-2.060 ±	0.824
3-phosphoglycerate	-4.122 ±	0.142
Phosphoenolpyruvate	-4.525 ±	0.178

[a]Changes in concentration are given relative to the parent strain.

Table S2.5 Primers used in this study

Primer	Target locus	Sequence (5'-3')
ISAORKO3'F	3' flank *gor*-L	AAGTTAGGCTGGTGGGTACCTTCCCAAATC TGCACCCTATC
JAB0004	3' flank *gor*-L	CAAAAAAACTATTTACCTCTACTCCAATCT TCATTTAAAATGGAATTCTTTTTGAAAC
JAB0003	pAR003	TTAAAATGGAATTCTTTTTGAAACATCAAC
JAB002	pAR003	CTGCATAGCCTCTTTTAAATCCTGTCTTACT CACTCACCTCTTCCATTG
JAB001	5' flank *gor*-L	CTTACTCACTCACCTCTTCCATTG
ISAORKO5'R	5' flank *gor*-L	GGACTATGAAGGAGAGCTGAATTCTCTGAC GCTCAGT
ISAORKOvectorF	pGL104 (242)	TGAATTCTCTGACGCTCAG
ISCB003	pGL104	CTTTCTACATAGAAAGGATGGTCTCTAGAA TGAATAAAGATGCTTACATTCAAATGTTC
GLCB096	pGL104	ATGAATAAAGATGCTTACATTCAAATGTTC
ISAORKOvectorR	pGL104	GTTAGGCTGGTGGGTACC
ISCB004	pIMS006	CATAGCCTCTTTTAAATCCTGT
JAB0003	pIMS006	TTAAAATGGAATTCTTTTTGAAACATCAAC
ISCB005	5' flank *gor*-S	CATAGCCTCTTTTAAATCCTGTTTAAATCAC CACCTGTAAAATC
ISCB006	5' flank *gor*-S	GGAAAGCTTATTAATTATGTAAAATAAGAA TTCTCTGACGCTCAG
ISCB007	pIMS006	GAATTCTCTGACGCTCAG
ISCB008	pIMS006	GTTAGGCTGGTGGGTACC
ISCB009	3' flank *gor*-S	CTTATAAAGTCTTTGCCTATCATTTAAAAT GGAATTCTTTTTGAAAC
ISCB010	3' flank *gor*-S	GTTAGGCTGGTGGGTACCCTGTTGTCTTTA TCCCCAAAAG
IS AORKOvectorR2	pIMSPFAOR	CCAATGATCGAAGTTAGGCTGGTGGTACCT TATTTTACATAATTAATAAG
IS AORKOvectorF2	pIMSPFAOR	GCAGATAGAAGGAATTCTCTGACGCTCAGT GGAAC
IS tagAOR5'F2	5' flank *gor*-S	CGAAGTTAGGCTGGTGGTACCTTATTTTAC ATAATTAATAAGCTTTCC
IS tagAOR5'R	5' flank *gor*-S	GGTTCTTCGAGCTGATAACACAGGATTTA

Figure 2.1 Activity of purified GOR in the presence and absence of inhibitors

The activity of purified GOR (0.006 mg/mL) was measured in the presence of cytoplasmic extract (S100, 0.174 mg/mL), heat-treated (95 °C for 60 min) cytoplasmic extract (HT-S100, 0.036 mg/mL), cytoplasmic extract washed using a 30 kDa membrane (> 30 kDa-S100, 0.325 mg/mL), the flow-through material that passed through the 30 kDa membrane (< 30 kDa-S100, protein was not measured), or partially purified PGK/TPI (*C. bescii* PGK/TPI, 0.002 mg/mL). The concentrations given in parentheses are the final concentration of each component in the assay cuvette.

Figure 2.1

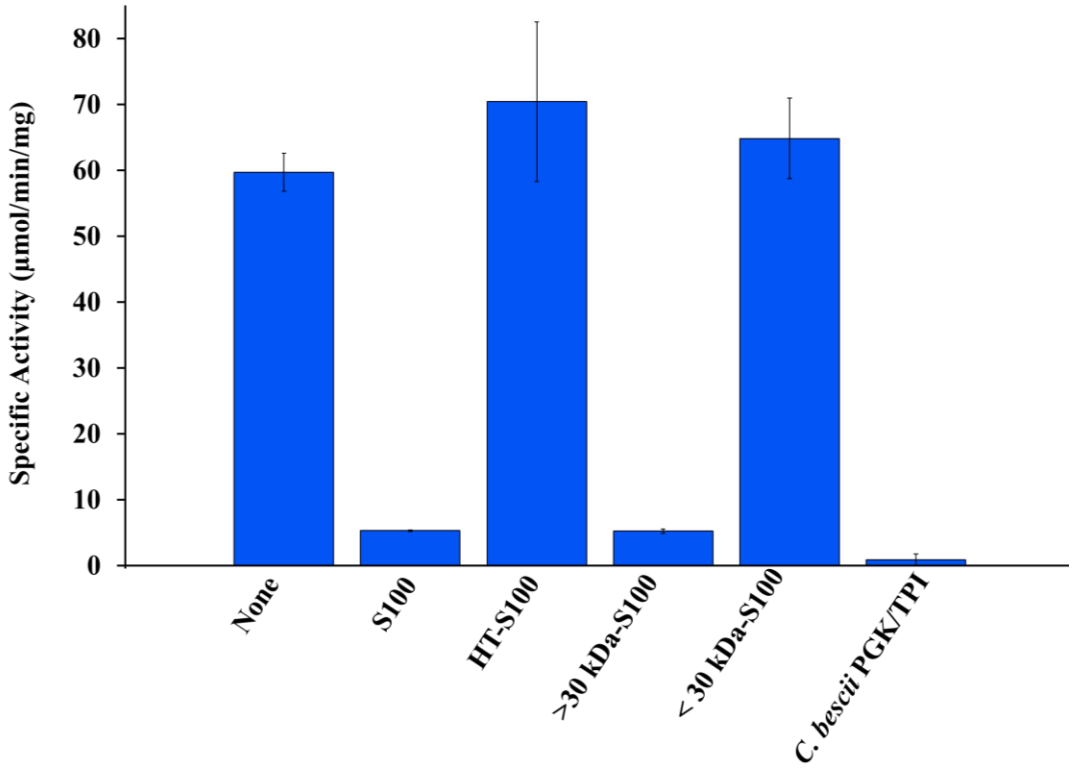

Specific Activity (µmol/min/mg)
80
70
60
50
40
30
20
10
0
None
S100
HT-S100
>30 kDa-S100
<30 kDa-S100
C. bescii PGK/TPI

Figure 2.2 Growth of Δ*gor* strains on xylose

Growth of parent strain (red) and of the Δ*gorL* (green) and Δ*gorS* (blue) strains using xylose as the carbon source at 75 °C (n=3).

Figure 2.2

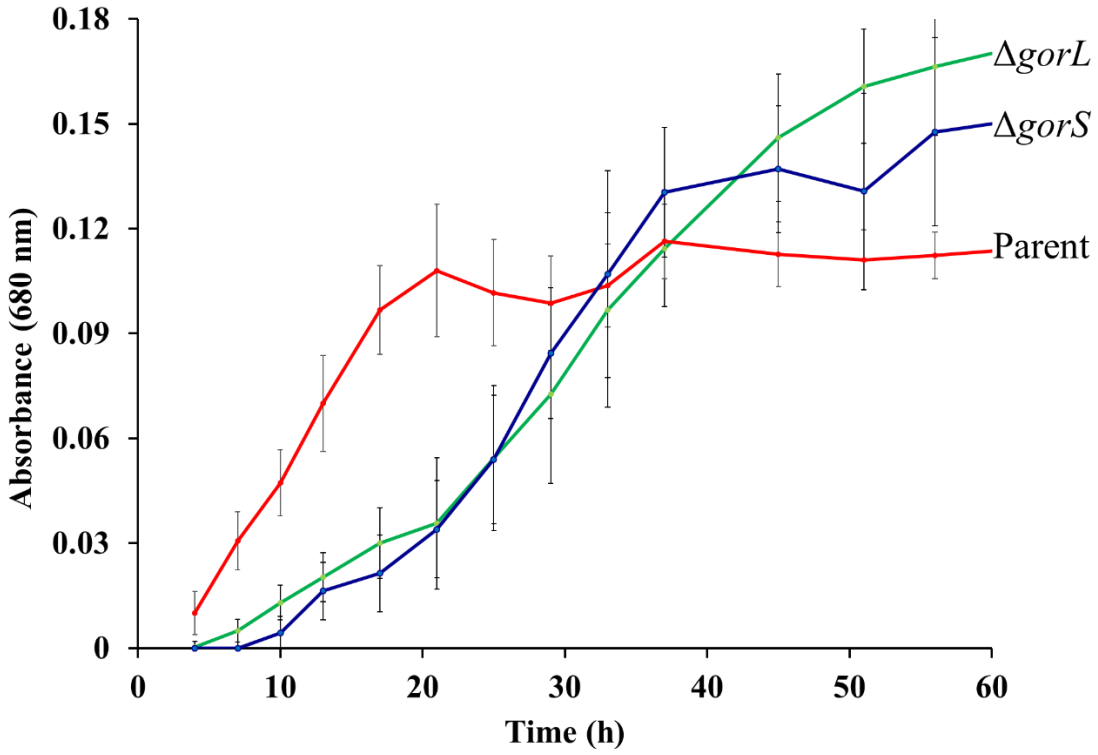

0.18
0.15
0.12
0.09
0.06
0.03
0
Absorbance (680 nm)
0
10
20
30
40
50
60
Time (h)
ΔgorL
ΔgorS
Parent

Figure 2.3 Relative abundance of glycolytic metabolites in Δ*gorL* and parent strains

A). Relative concentrations of intracellular metabolites in the Δ*gorL* strain relative to the parent strain as indicated by the $\log_2$-fold change. Error bars represent the standard deviation s (n=3 replicates). B). Glycolytic and pentose phosphate pathways in *C. bescii* with relative abundance of key metabolites. Those with higher abundance in the mutant strain are shown in red while those with lower abundance are shown in green. The further metabolism of intermediates of the pathways to amino acids and nucleotide precursors are indicated in orange. Abbreviations: PEP, phosphoenolpyruvate; 3PG, 3-phosphoglycerate; DHAP, dihydroxyacetone phosphate; FBR, fructose-1,6-phosphate; RU5P, ribulose-5-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; S7P, sedoheptulose-7-phosphate; R5P, ribose-5-phosphate; PPRP, phosphoribosyl pyrophosphate.

Figure 2.3

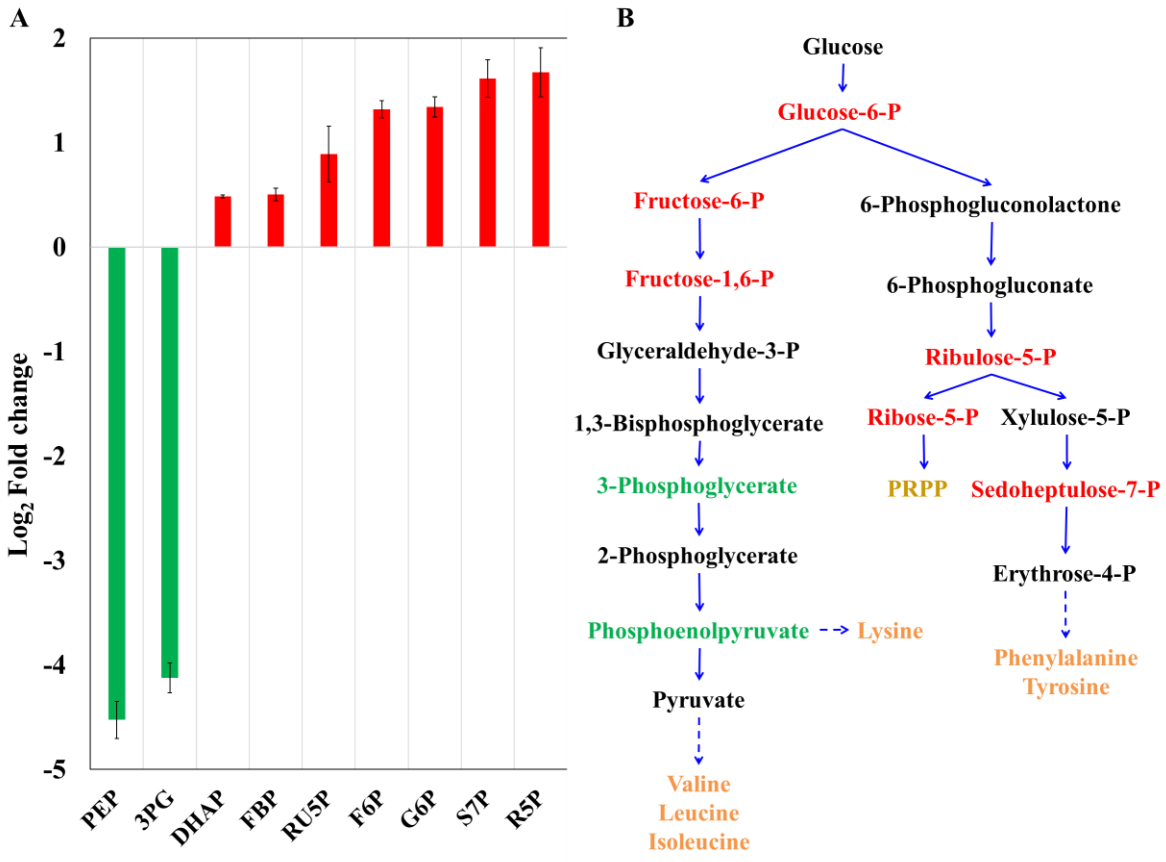

A
2
1
0
-1
-2
-3
-4
-5
Log2 Fold change
PEP
3PG
DHAP
FBP
RU5P
F6P
G6P
S7P
R5P
B
Glucose
Glucose-6-P
Fructose-6-P
6-Phosphogluconolactone
Fructose-1,6-P
6-Phosphogluconate
Glyceraldehyde-3-P
Ribulose-5-P
1,3-Bisphosphoglycerate
Ribose-5-P
Xylulose-5-P
3-Phosphoglycerate
PRPP
Sedoheptulose-7-P
2-Phosphoglycerate
Erythrose-4-P
Phosphoenolpyruvate
Lysine
Phenylalanine
Tyrosine
Pyruvate
Valine
Leucine
Isoleucine

Figure 2.4 Proposed role for GOR in primary metabolism

Proposed dual glycolytic pathways in *C. bescii*. The conventional pathway is shown in green, where oxidation of NADH and reduced ferredoxin leads to H_2 production via the bifurcating hydrogenase. The alternative GOR-dependent pathway and oxidation of the resulting reduced ferredoxin by the Ech hydrogenase and the associated ATP synthase is shown in dark yellow. Abbreviations: HK, hexokinase; PGI, phosphoglucose isomerase; PFK, phosphofructokinase; FBA, fructose-bisphosphate aldolase; TPI, triosephosphate isomerase BFH_2ase, bifurcating [FeFe] hydrogenase; GAPDH, glyceraldehyde-3-phosphate; PGK, phosphoglycerate kinase; GOR, glyceraldehyde-3-phosphate ferredoxin oxidoreductase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; ENO, enolase; PK. Pyruvate kinase Fd, ferredoxin; Ech, [NiFe] hydrogenase; POR, pyruvate ferredoxin oxidoreductase; LDH, lactate dehydrogenase; PTA, phosphotransacetylase; ACK, acetate kinase.

Figure 2.4

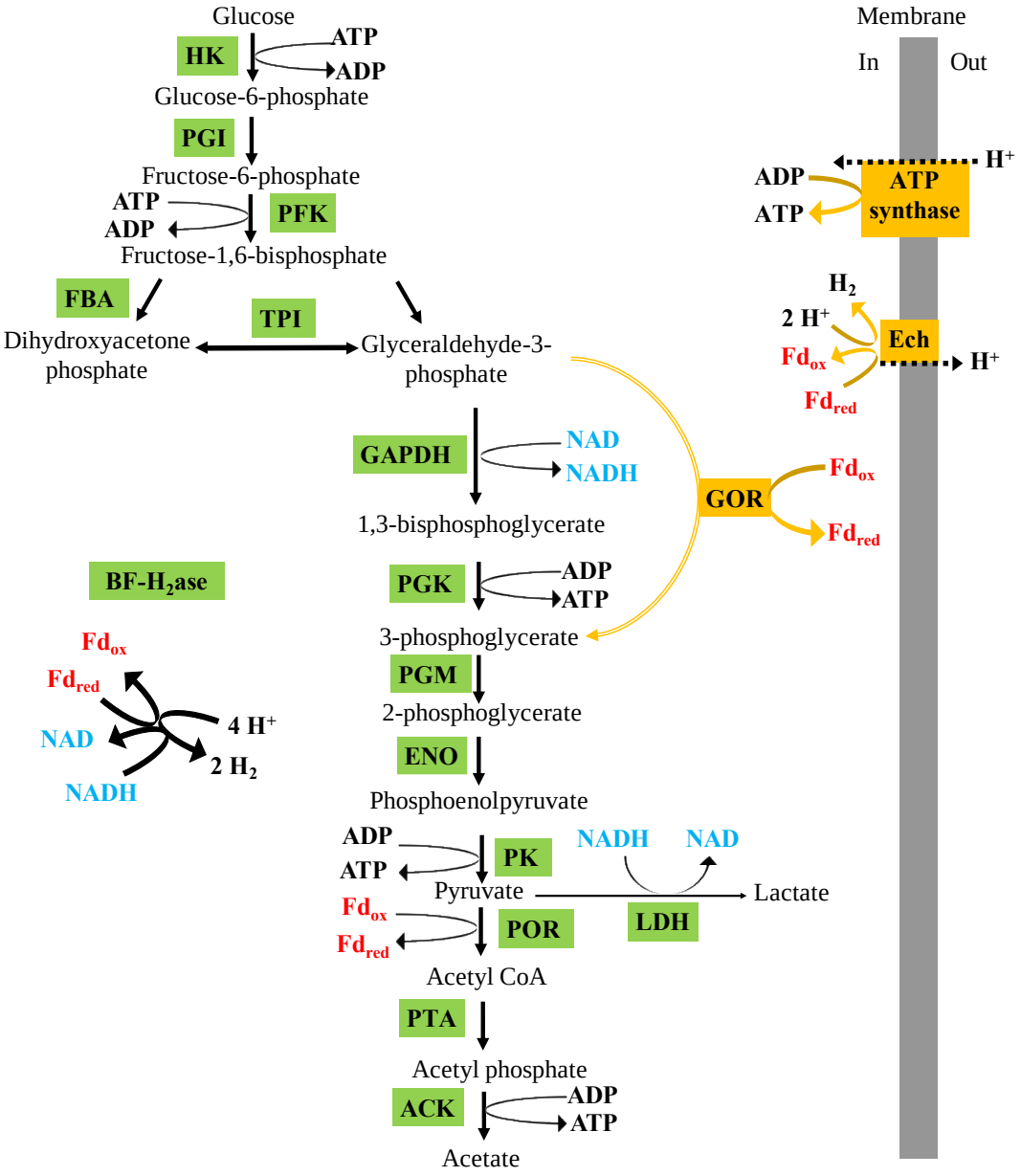

Glucose
HK
ATP
ADP
Glucose-6-phosphate
PGI
Fructose-6-phosphate
ATP
ADP
PFK
Fructose-1,6-bisphosphate
FBA
TPI
Dihydroxyacetone phosphate
Glyceraldehyde-3-phosphate
GAPDH
NAD
NADH
1,3-bisphosphoglycerate
PGK
ADP
ATP
3-phosphoglycerate
PGM
2-phosphoglycerate
ENO
Phosphoenolpyruvate
ADP
ATP
PK
Pyruvate
NADH
NAD
Lactate
LDH
Fdox
Fdred
POR
Acetyl CoA
PTA
Acetyl phosphate
ACK
ADP
ATP
Acetate
BF-H2ase
Fdox
Fdred
NAD
NADH
4 H+
2 H2
GOR
Fdox
Fdred
Membrane
In
Out
ADP
ATP
ATP synthase
H+
H2
2 H+
Fdox
Fdred
Ech
H+

Figure S2.1 Genomic context of *gorSL* in wild type and mutant strains

A). *C. bescii* gene cluster (Athe_0820 to Athe_0831) encoding GOR-L and GOR-S (orange) and the proteins necessary for pyranopterin biosynthesis (green) and tungstate transport (blue). B). Gene organization of recombinant strains of *C. bescii* Δ*gorL* (MACB1050), Δ*gorS* (MACB1074) and OE-XOR/PFD (MACB1010). Abbreviations: slp, S-layer protein; Cb*htk*, thermostable kanamycin resistance gene; pfd, polyferredoxin, gorS.

Figure S2.1

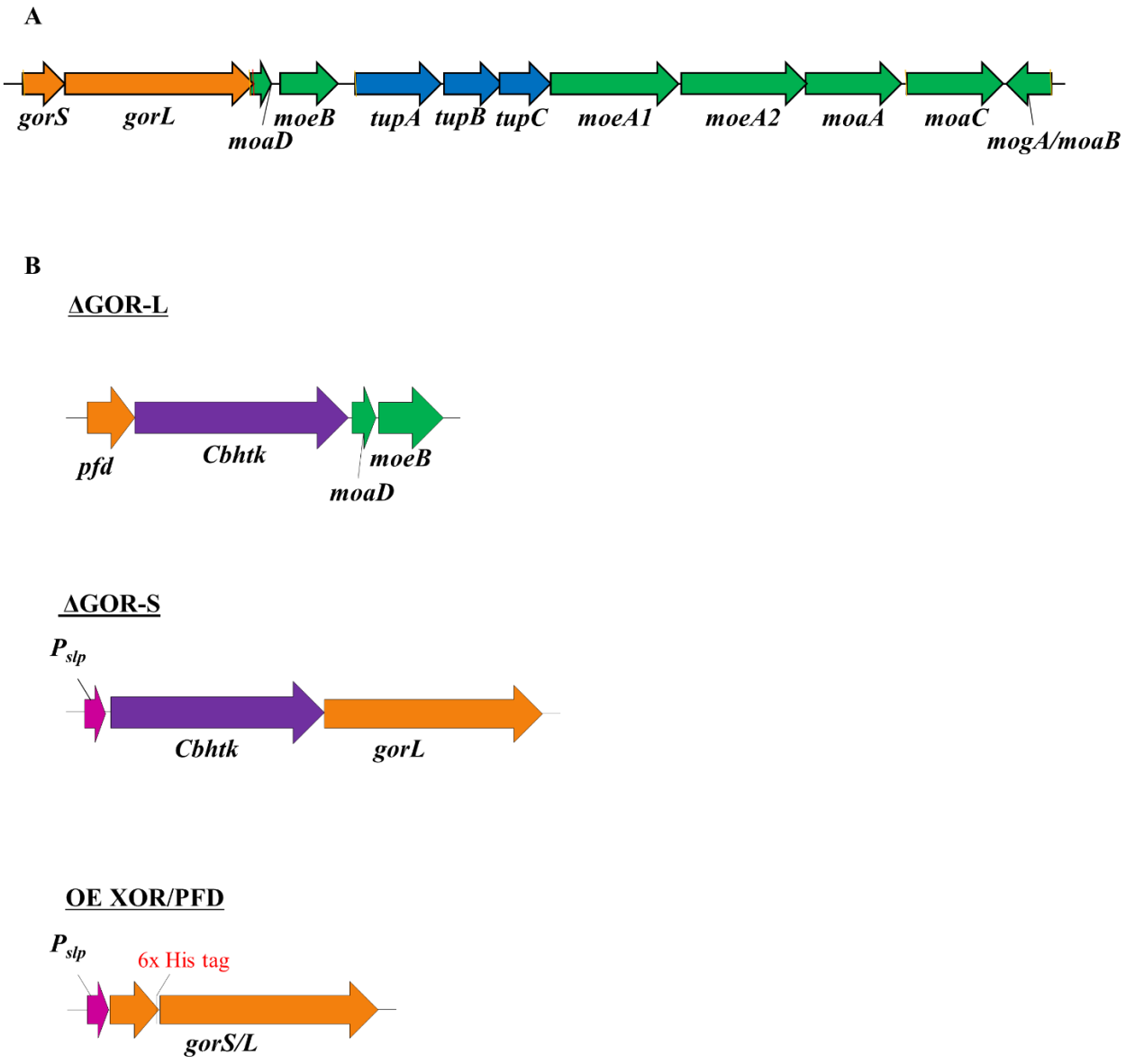

A
gorS
gorL
moaD
moeB
tupA
tupB
tupC
moeA1
moeA2
moaA
moaC
mogA/moaB
B
ΔGOR-L
pfd
Cbhtk
moaD
moeB
ΔGOR-S
P slp
Cbhtk
gorL
OE XOR/PFD
P slp
6x His tag
gorS/L

Figure S2.2 Activities of GOR and GAPDH following separation by anion exchange chromatography

The activity of GAPDH (red) and GOR (blue) were determined by the GAP-dependent reduction of NAD and benzyl viologen (BV), respectively. The initial substrate concentrations were GAP (5 mM), NAD (1 mM), and BV (1 mM).

Figure S2.2

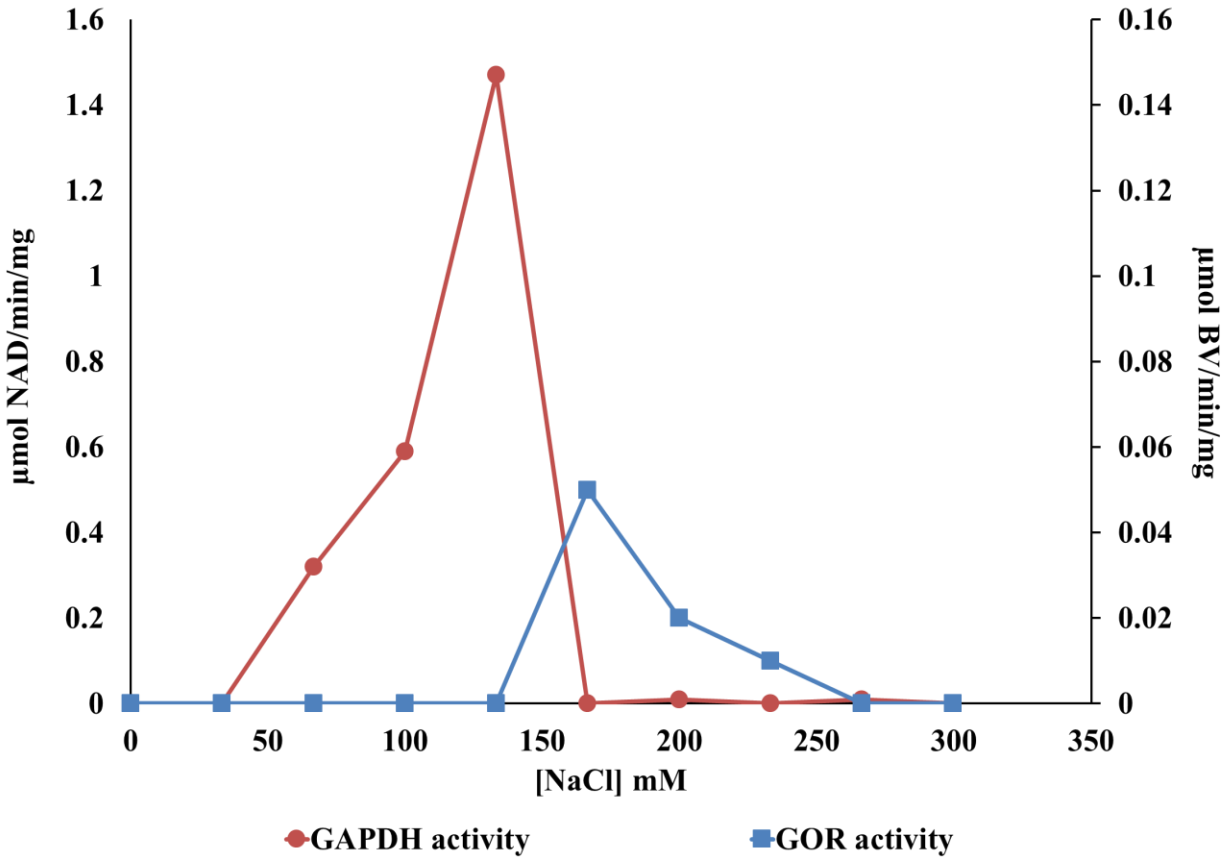

1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
µmol NAD/min/mg
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
µmol BV/min/mg
0
50
100
150
200
250
300
350
[NaCl] mM
GAPDH activity
GOR activity

Figure S2.3 Kinetic properties of purified GOR and partially purified GAPDH

Purified GOR (A) and partially purified GAPDH (B) activities were measured by reduction of BV and NAD, respectively, each at a concentration 1 mM.

Figure S2.3

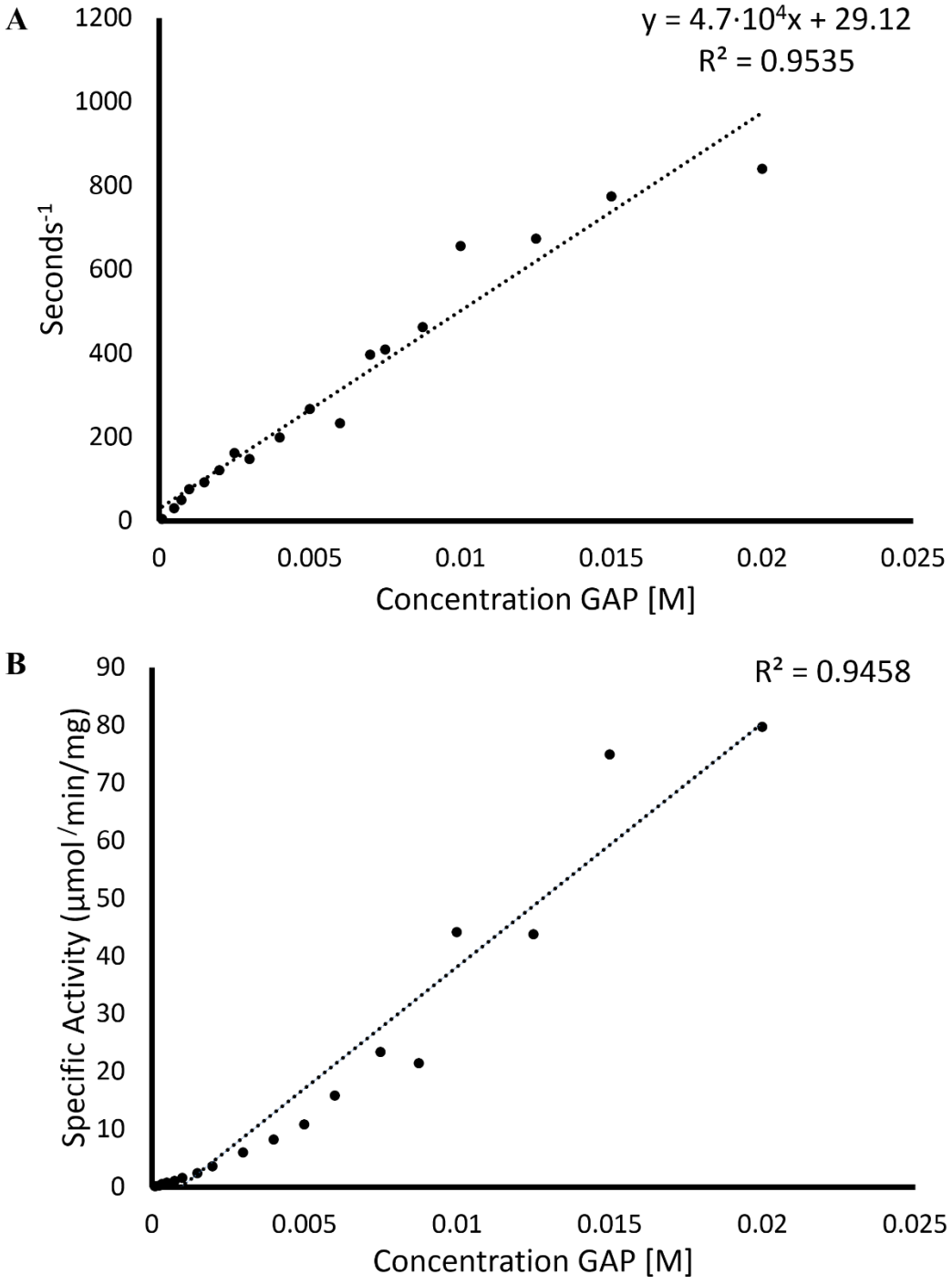

A
y = 4.7·10⁴x + 29.12
R² = 0.9535
Seconds⁻¹
1200
1000
800
600
400
200
0
0
0.005
0.01
0.015
0.02
0.025
Concentration GAP [M]
B
R² = 0.9458
Specific Activity (µmol/min/mg)
90
80
70
60
50
40
30
20
10
0
0
0.005
0.01
0.015
0.02
0.025
Concentration GAP [M]

Figure S2.4 Kinetic properties of purified GOR with a physiological electron acceptor

Kinetic properties of purified GOR using *C. bescii* ferredoxin as the electron acceptor. The initial concentration of GAP was 5 mM.

Figure S2.4

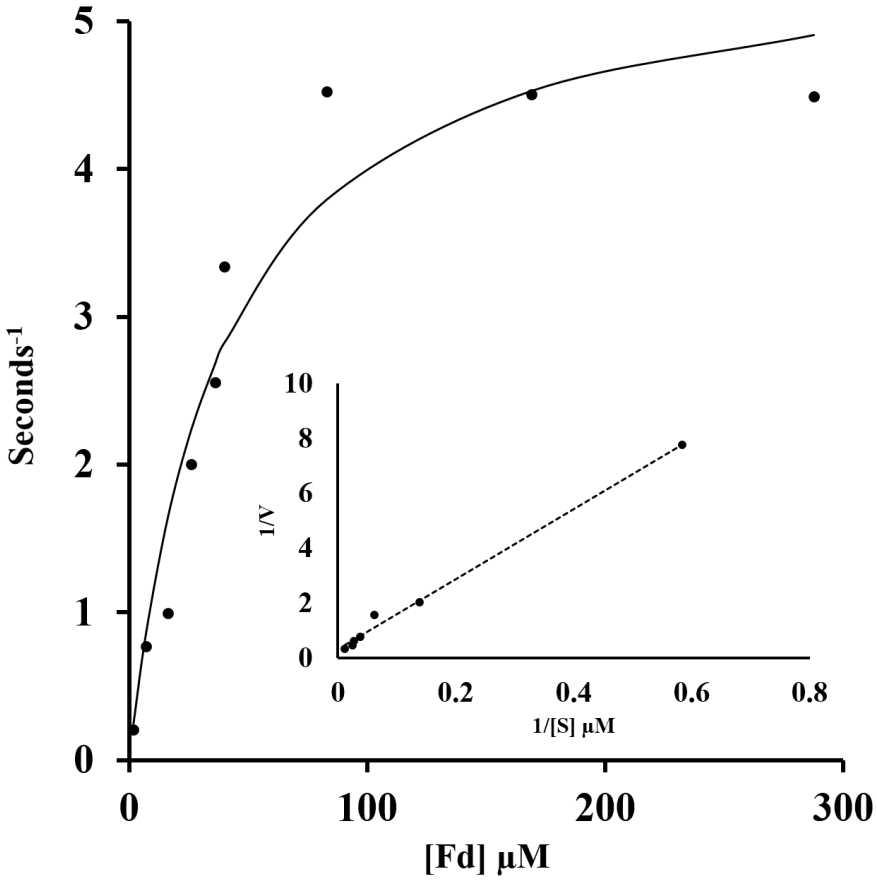

5
4
3
2
1
0
Seconds-1
0
100
200
300
[Fd] µM
10
8
6
4
2
0
1/V
0
0.2
0.4
0.6
0.8
1/[S] µM

Figure S2.5 Size exclusion chromatography of purified GOR

A). Estimated molecular mass of GOR holoenzyme (Peak B) and GOR heterodimer (Peak A; open circles/arrows). The standard proteins (closed circles) are (a) thyroglobulin (669 kDa), (b) apoferritin (443 kDa), (c) amylase (200 kDa), (d) alcohol dehydrogenase (150 kDa), (e) albumin (66 kDa) and (f) carbonic anhydrase (29 kDa). B). Size exclusion chromatography elution profile of affinity purified GOR showing GOR holoenzyme (Peak B) and GOR heterodimer (Peak A). Fractions were monitored in the UV (280 nm for protein) and visible (390 nm for iron-sulfur clusters) regions.

Figure S2.5

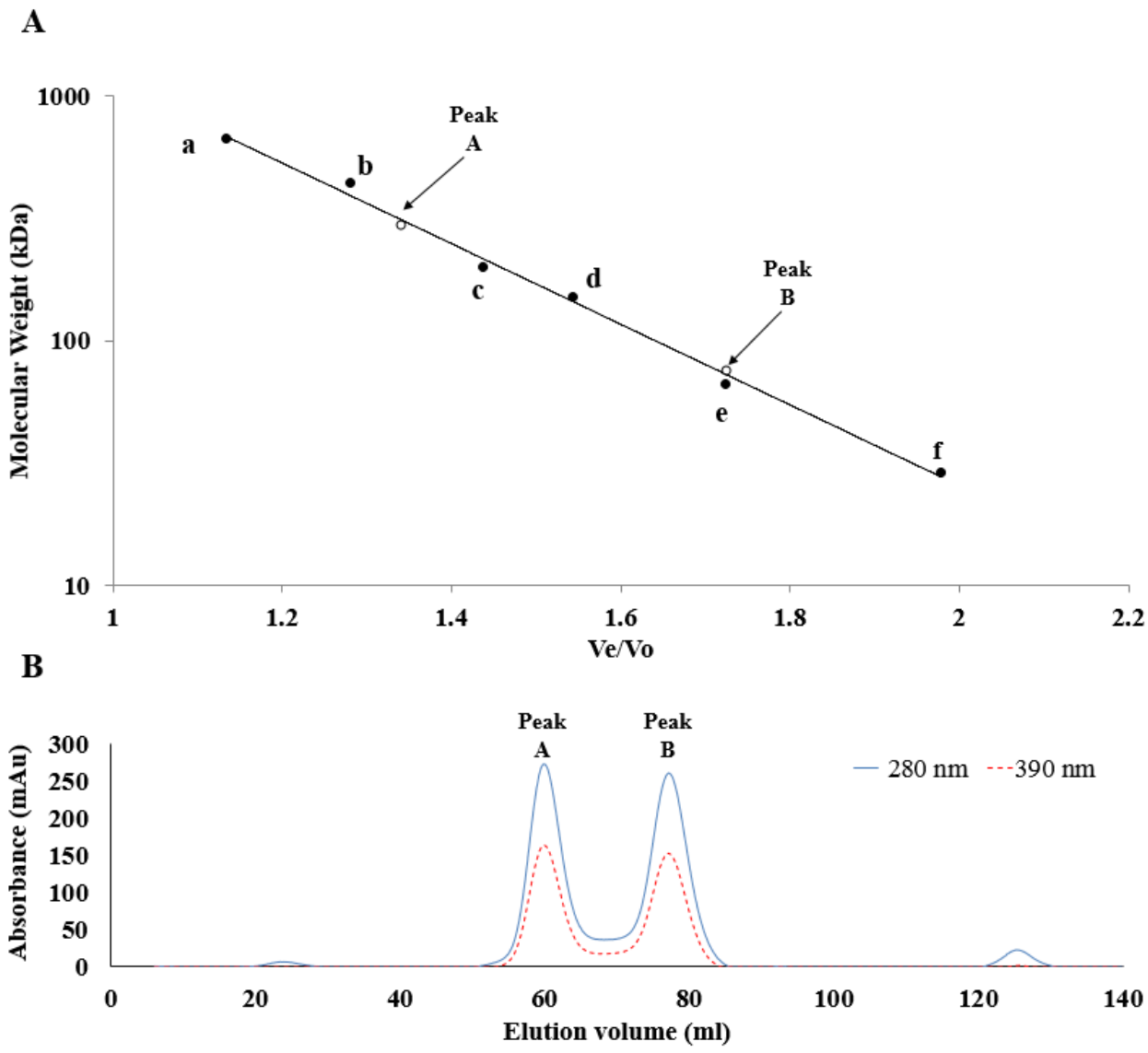

A
1000
100
10
Molecular Weight (kDa)
a
b
c
d
e
f
Peak A
Peak B
1
1.2
1.4
1.6
1.8
2
2.2
Ve/Vo
B
Absorbance (mAu)
300
250
200
150
100
50
0
Peak A
Peak B
280 nm
390 nm
0
20
40
60
80
100
120
140
Elution volume (ml)

Figure S2.6 Identification of GOR-S and GOR-L by SDS-PAGE and MS/MS analyses

A). SDS-PAGE analysis of GOR following ion exchange, affinity and size exclusion chromatography steps from cell extract of the *C. bescii* strain OE-XOR/PFD. Lane 1 affinity purified GOR lanes 2 and 3, peaks A and B respectively from size exclusion chromatography (refer to Figure S5B). Gel was stained with Imperial Protein Stain B). MS/MS analysis of the indicated bands. Peptides that matched the sequence of GOR-L and GOR-S are shown in red. Red boxes indicate bands excised for MS analysis.

Figure S2.6

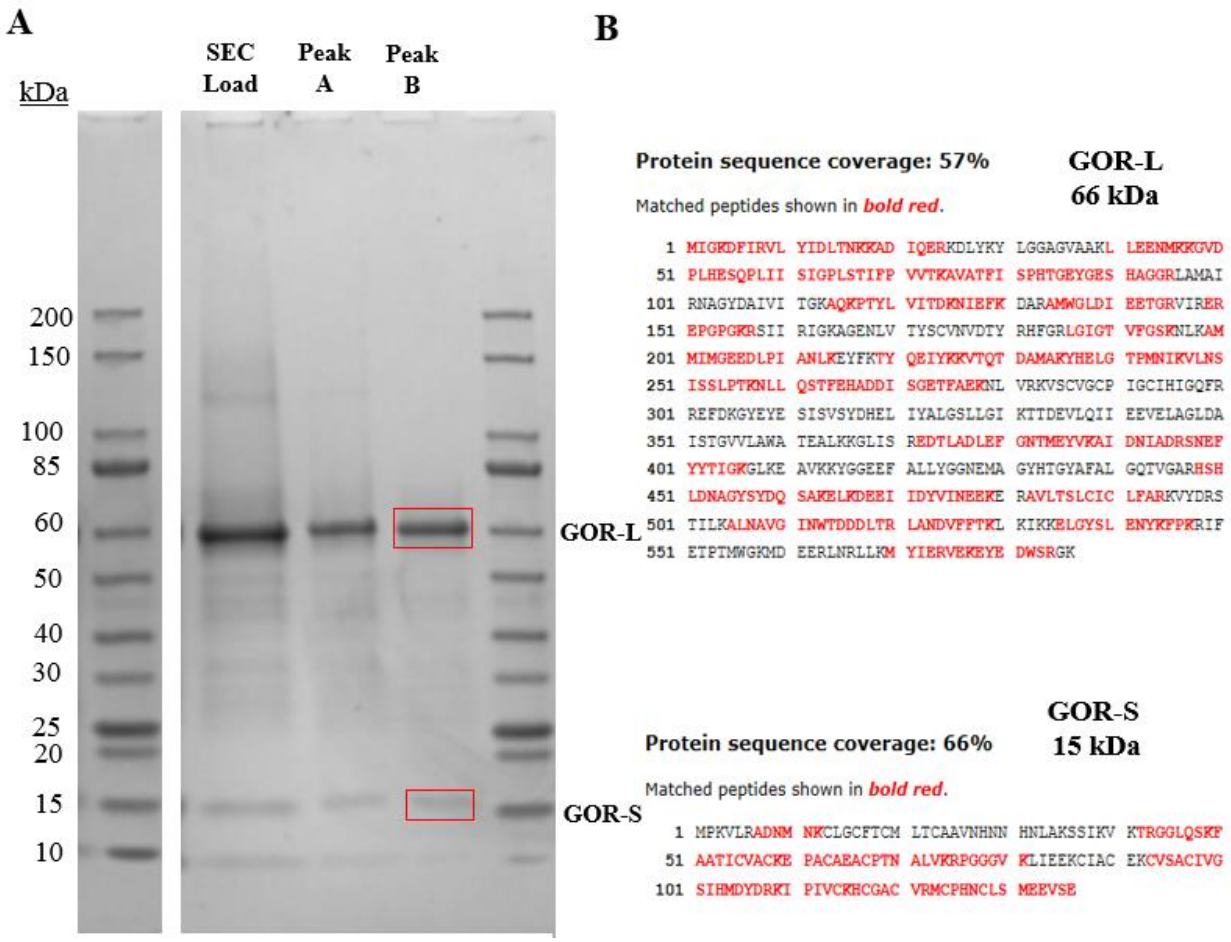

A
kDa
SEC Load
Peak A
Peak B
200
150
100
85
60
50
40
30
25
20
15
10
GOR-L
GOR-S
B
Protein sequence coverage: 57%
GOR-L
66 kDa
Matched peptides shown in bold red.
1 MIGKDFIRVL YIDLTNKKAD IQERKDLYKY LGGAGVAAKL LEENMKKGVD
51 PLHESQPLII SIGPLSTIFP VVTKAVATFI SPHTGEYGES HAGGRLAMAI
101 RNAGYDAIVI TGKAQKPTYL VITDKNIEFK DARAMWGLDI EETGRVIRER
151 EPGPGKRSII RIGKAGENLV TYSCVNVDTY RHFGRLGIGT VFGSKNLKAM
201 MIMGEEDLPI ANLKEYFKTY QEIYKKVTQT DAMAKYHELG TPMNIKVLNS
251 ISSLPTKNLL QSTFEHADDI SGETFAERNL VRKVSCVGCP IGCIHIGQFR
301 REFDKGYEYE SISVSYDHEL IYALGSLLGI KTTDEVLQII EEVELAGLDA
351 ISTGVVLAWA TEALKKGLIS REDTLADLEF GNTMEYVKAI DNIADRSNEF
401 YYTIGKGLKE AVKKYGGEEF ALLYGGNEMA GYHTGYAFAL GQTVGARHSH
451 LDNAGYSYDQ SAKELKDEEI IDYVINEEKE RAVLTSLCIC LFARKVYDRS
501 TILKALNAVG INWTDDDLTR LANDVFFTKL KIKKELGYSL ENYKFPKRIF
551 ETPTMWGKMD EERLNRLLKM YIERVEKEYE DWSRGK
GOR-S
15 kDa
Protein sequence coverage: 66%
Matched peptides shown in bold red.
1 MPKVLRADNM NKCLGCFTCM LTCAAVNHNN HNLAKSSIKV KTRGGLQSKF
51 AATICVACKE PACAEACPTN ALVKRPGGGV KLIEEKCIAC EKCVSACIVG
101 SIHMDYDRKI PIVCKHCGAC VRMCPHNCLS MEEVSE

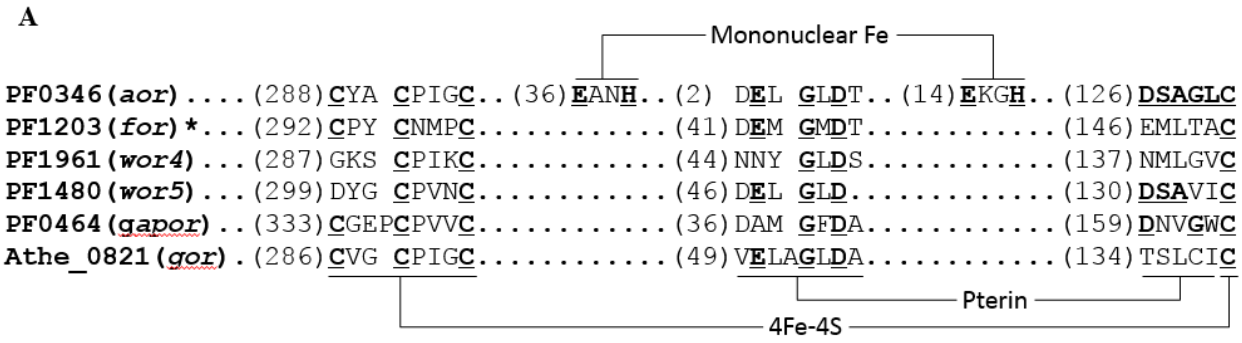
Figure S2.7 Cofactor binding motifs in WOR family enzymes

A). Pterin and [4Fe-4S]-binding motifs in the WOR family enzymes. B). Sequence of GOR-S showing the four cysteine motifs.

*The NCBI RefSeq protein start position shown here for FOR (PF1203) differs from that of the previously purified protein, which contains an additional eight residues at the N-terminus (333).

Figure S2.7

A



B

MPKVLRADNMNKCLGCFTCMLTCAAVNHNNHNLAKSSIKVKTRGGLQSKFAATICVACKEPACAEACP
TNALVKRPGGGVKLIEEKCIACEKCVSACIVGSIHMDYDRKIPIVCKHCGACVRMCPHNCLSMEEVSE

Figure S2.8 Growth of the Δ*gorL* strain on common carbohydrate substrates

Growth of the parent (blue) and Δ*gorL* (red) strains on the indicated carbon sources at 75 °C (n=3).

Figure S2.8

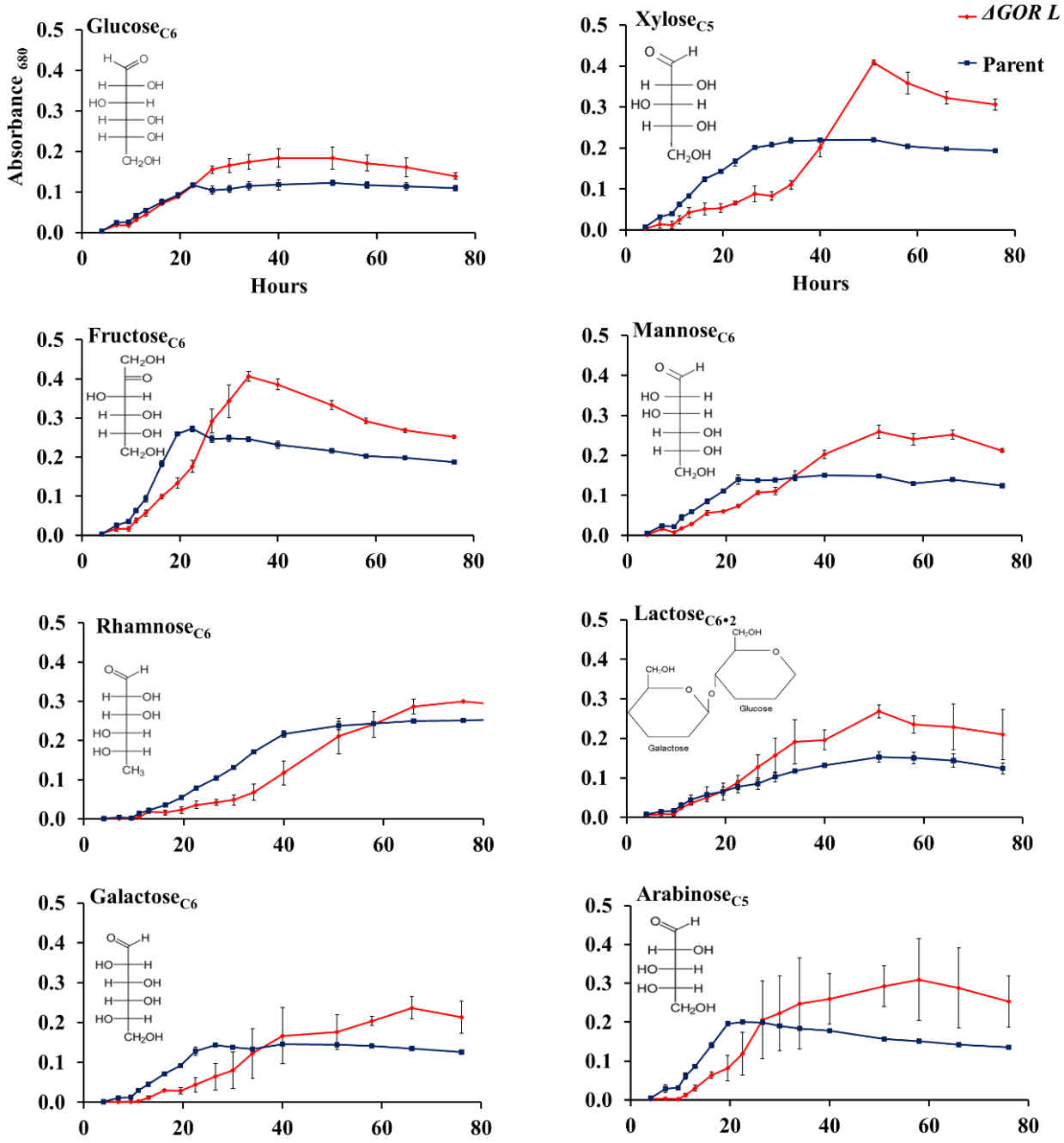

Glucose C6
Xylose C5
ΔGOR L
Parent
Fructose C6
Mannose C6
Rhamnose C6
Lactose C6•2
Glucose
Galactose
Galactose C6
Arabinose C5
Absorbance 680
Hours
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80

Figure S2.9 Phylogenetic tree of WOR family enzymes

A). Phylogenetic tree of the WOR family of enzymes 4,063 *C. bescii* GOR-L sequence homologs based on InterPro domains (IPR013983 and IPR001203). Only branches with ultrafast bootstrap (UFBoot) values greater or equal to 94% confidence are displayed. represent GOR-L (red) and five previously identified *P. furious* tungsten-containing oxidoreductase enzymes, AOR (dark green), WOR4 (purple), WOR5 (pink), FOR (gold), GAPOR (blue), and three others, Aa AorB (lime green), Ec YdhV (brown) and Gm BamB 1 & 2 (cyan). The outer ring indicates the source of the sequences where bacteria are gray, archaea are yellow and unknown environmental sequences are red. B). An enlarged region of a phylogenetic tree featuring *C. bescii* GOR (Cb GOR) and *P. furious* GAPOR (Pf GAPOR).

Figure S29

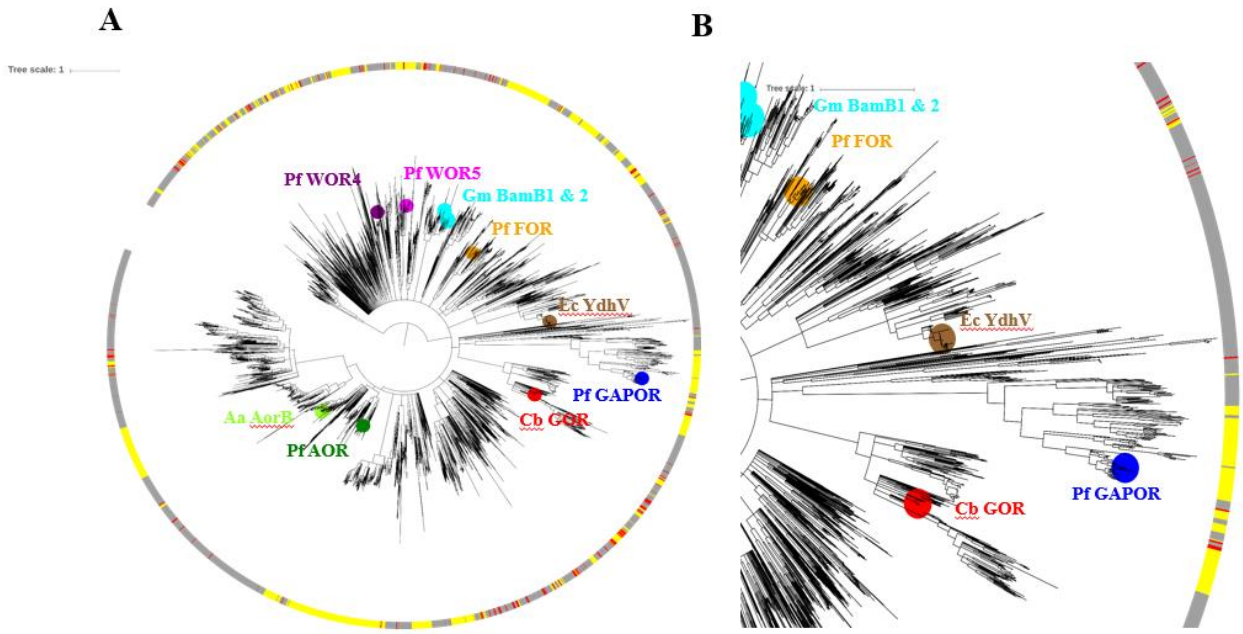
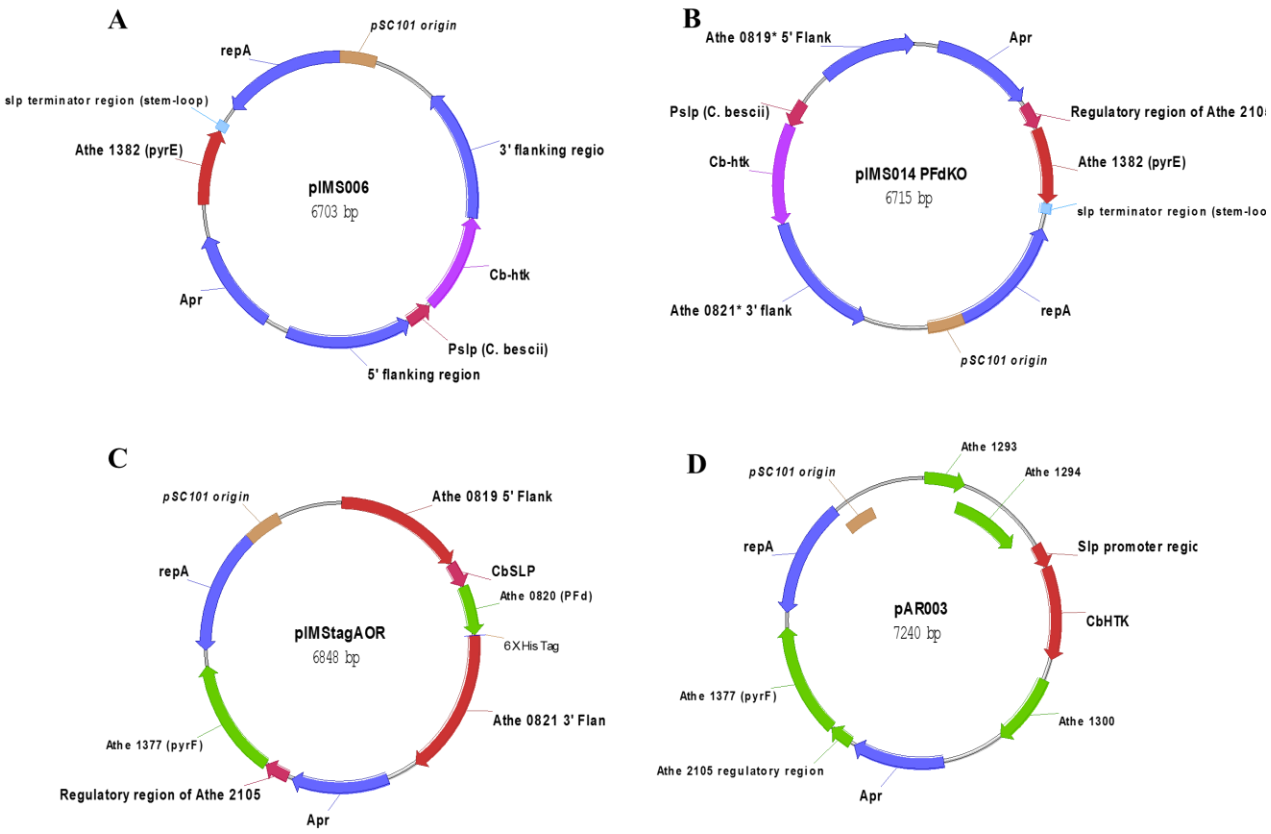

A
Tree scale: 1
Pf WOR4
Pf WOR5
Gm BamB1 & 2
Pf FOR
Ec YdhV
Pf GAPOR
Cb GOR
Aa AorB
Pf AOR
B
Tree scale: 1
Gm BamB1 & 2
Pf FOR
Ec YdhV
Pf GAPOR
Cb GOR

Figure S2.10 Plasmids utilized in this work

Vectors used in this study for A). deletion of *gorL* (pIMS006), B). deletion of *gorS* (Pfd; pIMS014) and C). overexpression of His-tagged GOR (pIMStagAOR). CbHTK under control of the *slp* promoter was cloned from D). (pAP003).

Figure S2.10


A
pSC101 origin
repA
slp terminator region (stem-loop)
Athe 1382 (pyrE)
3' flanking regio
pIMS006
6703 bp
Cb-htk
Apr
Pslp (C. bescii)
5' flanking region
B
Athe 0819* 5' Flank
Apr
Pslp (C. bescii)
Regulatory region of Athe 210
Cb-htk
pIMS014 PFdKO
6715 bp
Athe 1382 (pyrE)
slp terminator region (stem-loo
Athe 0821* 3' flank
repA
pSC101 origin
C
pSC101 origin
Athe 0819 5' Flank
repA
CbSLP
Athe 0820 (PFd)
pIMStagAOR
6848 bp
6XHis Tag
Athe 0821 3' Flan
Athe 1377 (pyrF)
Regulatory region of Athe 2105
Apr
D
Athe 1293
pSC101 origin
Athe 1294
repA
Slp promoter regic
pAR003
7240 bp
CbHTK
Athe 1377 (pyrF)
Athe 1300
Athe 2105 regulatory region
Apr

CHAPTER 3

ENGINEERING THE CELLULOLYTIC EXTREME THERMOPHILE *CALDICELLULOSIRUPTOR BESCII* TO REDUCE CARBOXYLIC ACIDS TO ALCOHOLS USING PLANT BIOMASS AS THE ENERGY SOURCE[1]

---

[1] **Rubinstein, G.M.**, Lipscomb, G.L., Williams-Rhaesa, A.M., Schut, G.J., Kelly, R.M., Adams, M.W.W. 2020. *Journal of Industrial Microbiology & Biotechnology.* 47:585-597

Abstract

Caldicellulosiruptor bescii is the most thermophilic cellulolytic organism yet identified (T_{opt} 78 °C). It grows on untreated plant biomass and has an established genetic system thereby making it a promising microbial platform for lignocellulose conversion to bio-products. Here, we investigated the ability of engineered *C. bescii* to generate alcohols from carboxylic acids. Expression of aldehyde ferredoxin oxidoreductase (*aor* from *Pyrococcus furiosus*) and alcohol dehydrogenase (*adhA* from *Thermoanaerobacter* sp. X514) enabled *C. bescii* to generate ethanol from crystalline cellulose and from biomass by reducing the acetate produced by fermentation. Deletion of lactate dehydrogenase in a strain expressing the AOR-Adh pathway increased ethanol production. Engineered strains also converted exogenously-supplied organic acids (isobutyrate and *n*-caproate) to the corresponding alcohol (isobutanol and hexanol) using both crystalline cellulose and switchgrass as sources of reductant for alcohol production. This is the first instance of an acid to alcohol conversion pathway in a cellulolytic microbe.

Introduction

As global energy demands and concerns about anthropogenic climate change continue to mount, alternatives to fossil fuels are becoming increasingly desirable. One potential source is lignocellulosic biomass (371-373). Traditionally, biomass must be pre-treated by expensive thermochemical methods and then enzymatically hydrolyzed before microorganisms (e.g., yeast) can ferment the individual sugars to a desired product (e.g., ethanol) (111, 374). Over the last decade, the idea of consolidated bioprocessing (CBP) has received attention as a potential cost-saving alternative to this first-generation cellulosic biofuel production scheme (157-159). Exploiting thermophilic microbes for CBP has several potential advantages. Notably, saccharolytic enzymes typically have higher activity at elevated temperatures (323) and utilizing thermophilic microbes can decrease the risk of contamination, lower the costs of cooling associated with large-scale fermentation processes, and facilitate separation of fermentation products from the growth medium via temperature-dependent product separation (159, 295, 375).

Caldicellulosiruptor bescii is an extremely thermophilic, cellulolytic, strictly anaerobic bacterium (257, 268), capable of simultaneously fermenting pentose and hexose sugars to acetate, lactate, hydrogen gas, and carbon dioxide (252, 253). *C. bescii* is a promising CBP host microbe because it has the highest temperature optimum of any known cellulolytic organism (T_{opt} 78 °C) and a genetic system is available (242, 255, 256, 275). Previous metabolic engineering of *C. bescii* has demonstrated the fermentation of untreated switchgrass directly to ethanol (**Figure 3.1**). Specifically, in a strain in which the lactate dehydrogenase gene (*ldh*) was inactivated, heterologous expression of the bifunctional aldehyde/alcohol dehydrogenase (*adhE*) from *Clostridium thermocellum* resulted in the production of 10 mM ethanol from growth on 5 g • L^{-1} switchgrass at 65 °C (below the optimum growth temperature for *C. bescii*) (292). Alternatively,

expression of an *adhE* homolog from *Thermoanaerobacter pseudoethanolicus* allowed *C. bescii* to ferment biomass to ethanol at 75 °C, albeit at lower titer, i.e. 2.5 mM ethanol (293).

Recently, a novel synthetic metabolic pathway for alcohol production was engineered into the hyperthermophilic archaeon *Pyrococcus furiosus,* which also grows by C_6 sugar fermentation (although it does not utilize cellulose). Insertion of a bacterial alcohol dehydrogenase gene (*adhA*) into the *P. furiosus* genome enabled the conversion of glucose to ethanol in a pathway that proceeds through acetate, rather than acetyl-coenzyme A (acetyl-CoA) (296). Acetate is reduced to acetaldehyde, catalyzed by the native *P. furiosus* enzyme aldehyde ferredoxin oxidoreductase (AOR). Acetate reduction is driven by the low redox potential of the electron carrier, a 4Fe-ferredoxin. The heterologously-expressed alcohol dehydrogenase (AdhA) from *Thermoanaerobacter* sp. X514 catalyzes acetaldehyde reduction to ethanol using NADPH as the electron donor (296). Interestingly, due to the broad substrate specificities of both AOR and AdhA, a variety of exogenously added organic acids were also able to be converted to their corresponding alcohols with the concurrent production of ethanol (296).

While the AOR-Adh pathway has been expressed heterologously in *P. furiosus,* several bacteria naturally utilize this pathway (376). Recent work in the mesophilic clostridia, *C. ragsdalei*, *C. autoethanogenum*, *C. ljungdahlii*, and *C. carboxidivorans,* has confirmed the existence of the AOR-Adh pathway in these bacteria (298-300, 377), which contain both native AORs and Adhs. While alcohol titers from these species are robust (up to 350 mM ethanol produced by engineered *C. ragsdalei*), none of these microbes are cellulolytic, and no cellulolytic *Clostridium* spp. has been reported to contain this pathway (305). A variety of *Thermoanaerobacter* species have also been shown to utilize this pathway, but likewise lack cellulolytic activity (302, 303).

The AOR-Adh pathway is an attractive candidate for engineering into *C. bescii* because it allows for the concomitant formation of ethanol as well as other short- and medium-chain alcohols from their corresponding acid. In doing so, the bacterium utilizes both reduced ferredoxin and reduced nicotinamide nucleotides generated from glucose oxidation (**Figure 3.1**). Herein, we report the characterization of *C. bescii* strains that heterologously express the AOR-Adh pathway. *C. bescii* does not naturally contain either AOR or AdhA, and so the genes encoding these enzymes are from other thermophilic sources. In these engineered *C. bescii* strains, we demonstrate production of up to 4 mM ethanol, as well as the reduction of exogenously-supplied organic acids into their corresponding alcohols, specifically isobutanol and hexanol, which are drop-in fuels compatible with the current petroleum-based fuel infrastructure. This work demonstrates potential for the AOR-Adh pathway to be harnessed for both the production of ethanol and the conversion of other organic acids to alcohols by the fermentation of disaccharides, crystalline cellulose and, most importantly, untreated plant biomass.

Materials and methods

Growth of C. bescii

C. bescii strains used and generated in this study are listed in **Table 3.1**. DSMZ 6725 (wild type *C. bescii*) was obtained from the German Collection of Microorganisms and Cell Cultures. Strain JWCB018 (control strain) was obtained from J. Westpheling (University of Georgia). Strain MACB1013 (AOR-Adh strain, Δ*pyrFA*, Δ*cbeI*, $P_{S\text{-}layer}$ *aor adhA* [PF0346, Teth514_0564]) was generated as described previously (279). Strain MACB1038 (AOR-Adh/Δ*ldh* strain, Δ*pyrFA*, Δ*cbeI*, Δ*ldh*, $P_{S\text{-}layer}$ *aor adhA* [PF0346, Teth514_0564]) was generated in this study. Recombinant strains of *C. bescii* were generated using allelic exchange, as described previously (256). For preparation of competent cells, cultures were grown at 70 °C in Low Osmolarity Defined (LOD)

medium supplemented with amino acids, as described elsewhere (256, 276). Unless otherwise noted, strains were grown in 100 mL serum bottles in 50 mL of modified DSM516 growth medium, at 75 °C with shaking at 150 rpm, as described previously (242). Carbon substrates used were: maltose (Sigma, St. Louis, MO), cellobiose (MP Biomedicals, Santa Ana, CA), crystalline cellulose (Avicel PH-101, Sigma, St. Louis, MO), and switchgrass (sieved 20/80-mesh fraction; provided by Dr. Brian Davison, Oak Ridge National Laboratory, Oak Ridge, TN). The only treatment that was carried out on plant biomass samples involved washing in distilled water for 18 h at 78 °C to remove soluble free sugars. All strains were revived from glycerol stocks and transferred to fresh medium before each experiment. When working with uracil auxotrophic strains, all growth medium was supplemented with 20 µM uracil. Growth medium containing exogenous organic acids (20 mM isobutyrate or 10 mM *n*-caproate) were adjusted to pH 7.2 after the addition of the acids. Isobutyric acid and *n*-caproic acid were both obtained from Sigma (St. Louis, MO). For controlled fermentation experiments, cultures were grown in 15 L of medium in a custom-built 20 L fermentor containing 20 g • L^{-1} crystalline cellulose. The fermentor was inoculated with 1 L of late-log phase culture. Medium was supplemented with ammonium chloride (J.T. Baker, Phillipsburg, NJ) to a final concentration of 20 mM to prevent nitrogen limitation during growth. Agitation by a dual Rushton stack impeller was constant at 150 rpm, and 80/20 nitrogen/carbon dioxide was sparged through the vessel at 1.5 L • min^{-1}. The culture pH was allowed to acidify from 7.0 to 6.2 (measured at 75°C), at which point the automated addition of 10% w/v sodium bicarbonate solution maintained the pH for the remainder of growth.

Cell counts and protein estimations

Cell counts were performed using a Petroff-Hausser counting chamber. Cultures were counted undiluted or diluted 10-fold in *C. bescii* base salts depending on cell density. Protein

estimations were performed using the Bradford Assay with bovine serum albumin (BSA) standards (Biorad).

C. bescii transformation and strain construction

Primers utilized for strain construction and confirmation are presented in **Table 3.2**. Transformation of *C. bescii* was performed as described previously (242, 256). Strain MACB1038 was generated by transforming MACB1013 with pDCW121, acquired from J. Westpheling (University of Georgia). Competent cells of MACB1013 were prepared as described elsewhere (242, 256, 279).

Preparation of whole cell extracts and protein purification

C. bescii strains DSMZ 6725 and MACB1038 were each grown in 1 L of modified DSM516 medium with glucose until the late-exponential phase. Cells were harvested by centrifugation at 6,000 x g for 10 minutes. Cell pellets were flash frozen in liquid nitrogen and stored at -80 °C. The remaining steps of extract preparation were conducted under strictly anaerobic conditions. Cell pellets were thawed and then suspended in 3 mL of 25 mM Tris/HCl buffer containing 1 mg • mL^{-1} lysozyme (Sigma, St. Louis, MO), and washed with excess buffer over a regenerated cellulose centrifugation filter (10k, Amicon, EMD Millipore, Burlington, MA). After washing, extracts were concentrated to 3 mL buffer per g of wet weight cells, aliquoted into sealed vials, and stored for up to one week at 4 °C until protein content and enzyme activities were measured. AOR and ferredoxin were purified from *P. furiosus* under strictly anaerobic conditions, as described previously (311, 378). *C. bescii* ferredoxin was purified by a procedure similar to that used for *P. furiosus* ferredoxin (data not shown).

Enzyme assays

AOR and Adh activities were measured spectrophotometrically in *C. bescii* cell extracts using an Agilent Technologies Cary 100 UV-Vis Spectrophotometer equipped with an Agilent Cary Series Temperature Controller. Unless otherwise specified, assays were performed in 3 mL silicone-stoppered quartz cuvettes under an argon atmosphere. AOR activity was measured by the acetaldehyde (1 mM)-dependent reduction of benzyl viologen (1 mM) at 600 nm ($\varepsilon = 8.23\ M^{-1} \bullet cm^{-1}$) in 50 mM EPPS buffer (pH 8.0) containing sodium dithionite (1 µM) at 75 °C, as described elsewhere (311). Activity is expressed as AOR units • mg^{-1} of protein, where 1 unit equals 1 µmol of acetaldehyde oxidized per minute. Adh-dependent reduction of acetaldehyde (1 mM) was measured in 50 mM MOPS buffer (pH 7.5) by the oxidation of NADPH at 340 nm ($\varepsilon = 6.22\ M^{-1} \bullet cm^{-1}$), as described elsewhere (296). NADPH was added to each cuvette to a final absorbance of ~0.2 AU and allowed to stabilize before the addition of substrate. Assays using purified AOR and ferredoxin were performed in 700 µL silicone-stoppered quartz cuvettes. The oxidation of crotonaldehyde to crotonate was measured by the reduction of ferredoxin, calculated by the decrease in absorbance at 425 nm ($\varepsilon = 14.60\ M^{-1} \bullet cm^{-1}$). Assays were performed at 75 °C in 50 mM EPPS buffer (pH 8.0) with 200 µM crotonaldehyde and 0.01 mg • mL^{-1} AOR.

Chemical analysis

Acetate, lactate, and ethanol were measured on a Waters 2690 high performance liquid chromatograph (HPLC) over a Biorad Aminex Fast Acid Analysis Column with a photodiode array (PDA) detector. Cellobiose and glucose were measured by the same instrument and column with a refractive index (RI) detector. Hydrogen gas was measured by sampling the gas phase of the serum bottle using a Shimadzu GC-8A Gas Chromatograph (GC) with a thermal conductivity detector (TCD) (Restek Molesieve 5A packed column, argon carrier gas). Carbon dioxide

production was not measured, and was assumed to be equivalent to the sum of acetate and ethanol production, based on reaction stoichiometry. Conversion of isobutyrate to isobutanol and *n*-caproate to hexanol were measured by both HPLC-PDA and GC with a flame ionization detector (FID). The Agilent 7890A GC-FID (Restek Stabilwax 30m column, argon carrier gas) has higher sensitivity than the HPLC-PDA, and measurement by this method is preferable. Acetate and ethanol concentrations, measured by HPLC, were verified by GC.

Carbon balances

Carbon balances (substrate utilized versus products accounted for) were calculated as previously described (257). The amount of carbon consumed was reported as glucose equivalents, calculated from the dry mass of residual crystalline cellulose after fermentation. Products were determined as the sum of measured acetate, ethanol, pyruvate, glucose, cellobiose, calculated carbon dioxide, and estimated biomass. Production of carbon dioxide was assumed to equal one molecule per molecule of acetate or ethanol formed. Biomass was estimated by protein content.

Results

AOR-Adh strain produces ethanol and lactate

To explore the feasibility of utilizing the AOR-Adh pathway for ethanol production in *C. bescii* (**Figure 3.1**), a strain was engineered to heterologously express the genes encoding *P. furiosus* AOR and *T.* sp. X514 AdhA, as described previously (279). However, ethanol production by this strain was not characterized in the previous study. Both genes were expressed from a synthetic operon under the control of the constitutively-expressed high-level promoter of the gene encoding the *C. bescii* S-layer protein (*slp*), using a ribosomal binding site from the same *slp* promoter between the genes. The resulting strain (MACB1013; Δ*pyrFA*, Δ*cbeI*, $P_{S\text{-}layer}$ *aor adhA* [PF0346, Teth514_0564]) is referred to herein as the ‘AOR-Adh strain’. The AOR-Adh strain

parent (JWCB018; (291)) is an example of where the IS*Cbe4* transposon disrupts *ldh* (278), and so this strain does not produce lactate and also exhibits increased hydrogen production relative to the wild type (**Figure 3.2A**). Surprisingly, the AOR-Adh strain produced a significant amount of lactate (4.1 mM) when cultured in medium containing cellobiose as a carbon source (**Figure S3.1**) (279). PCR screening of the *ldh* region in the AOR-Adh strain suggested the presence of an intact *ldh* gene and sequencing confirmed this to be the case (**Figure 3.2B**) (279). Hence, it appeared that during construction of the AOR-Adh strain, the transposon migrated, leaving an intact and functional gene encoding LDH. This is discussed further below.

To investigate ethanol production by the AOR-Adh strain, it was cultivated in medium containing cellobiose as a carbon source, alongside the AOR-Adh parent (JWCB018) control strain. The AOR-Adh strain was found to produce 1.2 mM ethanol after 30 h of growth (**Figure S3.1**). Since a functional LDH was restored in this strain, we hypothesized that deletion of *ldh* from the AOR-Adh strain would lead to improved ethanol production because reducing power previously utilized by LDH to produce lactate would be directed toward ethanol as an end-product of fermentation (**Figure 3.1**) (264, 291, 292).

Deletion of ldh in the AOR-Adh strain increases alcohol production

With the goal of increasing the ethanol production of the AOR-Adh pathway in *C. bescii*, a targeted deletion of the restored *ldh* locus in the AOR-Adh strain was made, yielding the 'AOR-Adh/Δ*ldh*' strain (MACB1038; Δ*pyrFA*, Δ*cbeI*, Δ*ldh*, $P_{S\text{-}layer}$ *aor adhA* [PF0346, Teth514_0564]). A schematic of the relevant strain genotypes as well as PCR confirmation of the *ldh* deletion in the AOR-Adh strain background is detailed in **Figure 3.2**. To examine ethanol production by the new strain, it was cultured alongside the control strain (JWCB018) in medium containing cellobiose (**Figure S3.2**). After 30 h of growth at 75 °C, the AOR-Adh/Δ*ldh* strain produced 2.4

mM ethanol, while the control strain produced less than 0.5 mM (**Figure 3.3A**). A slightly lower concentration of ethanol was produced when the AOR-Adh/Δ*ldh* strain was grown with crystalline cellulose as the growth substrate, and with switchgrass, less than 0.5 mM was produced (**Figure 3.3A**). As expected, neither strain produced detectable lactate.

Activities of AOR and AdhA in the C. bescii AOR-Adh/Δldh strain

To demonstrate that the heterologously produced enzymes AOR and AdhA were active in *C. bescii,* and to verify that no competing activities existed, the activities of AOR and AdhA (measured by acetaldehyde oxidation and reduction, respectively) were determined in cell-free extracts of the wild type and AOR-Adh/Δ*ldh* strains. While wild type *C. bescii* did not display detectable activity for either enzyme, the extracts of the AOR-Adh/Δ*ldh* strain exhibited both AOR and AdhA activities (**Table 3.3**).

Activities of purified AOR with C. bescii and P. furiosus ferredoxins

To compare the activity of *P. furiosus* AOR with the 4Fe-ferredoxin from the same archaeon versus activity with the 8Fe-ferredoxin from *C. bescii,* activities were measured using the purified redox proteins at concentrations from 1 to 100 µM (**Figure S3.3**). For the 4Fe-ferredoxin from *P. furiosus*, the K_M value was 14.9 µM with a V_{max} of 10.1 units • mg^{-1}. For the 8Fe-ferredoxin from *C. bescii,* the K_M value was 149.0 µM with a V_{max} of 18.9 units • mg^{-1}. K_M and V_{max} values were calculated using a non-linear regression to fit the kinetic data to the Michaelis-Menten model (**Figure S3.3**) (379).

Reduction of organic acids to alcohols by the AOR-Adh/Δldh strain

To further test the functionality of the AOR-Adh pathway in *C. bescii*, the recombinant strain was grown in the presence of isobutyrate to determine if it was reduced to isobutanol. Isobutyrate was chosen as the model organic acid because a recombinant strain of *P. furiosus*

expressing the AOR-Adh pathway reduced isobutyrate to isobutanol at a higher titer than any of the other six organic acids that were tested (296). With 20 mM exogenous isobutyrate in cellobiose-containing growth medium, the AOR-Adh/Δ*ldh* strain produced 3.5 mM isobutanol, along with 1.2 mM ethanol, after 30 h of growth (**Figure 3.3B**). A similar concentration of isobutanol was produced using crystalline cellulose as the growth substrate, but less than 1 mM was produced on switchgrass. No isobutanol was produced by the control strain (JWCB018) with any of the three growth substrates (**Figure 3.3B**).

As a rule, alcohols with a longer carbon chain length have better fuel properties (380). *n*-Caproate (or hexanoate, C_6) is the longest straight-chain organic acid that is suitably water soluble, making it a desirable target for reduction by the AOR-Adh pathway. The addition of 10 mM *n*-caproate to the growth medium (containing cellobiose) did not inhibit *C. bescii* growth (**Figure S3.2**) and resulted in the production of 0.2 mM hexanol along with 1.0 mM ethanol (**Figure 3.3C**). Growth with crystalline cellulose as the substrate improved hexanol production to 0.4 mM, and growth on switchgrass yielded the same concentration of hexanol as that produced during growth on cellobiose.

The highest concentrations of isobutanol and hexanol were achieved from growth on crystalline cellulose (4.6 mM and 0.4 mM, respectively). In all experiments, alcohol production from growth on switchgrass was lower, correlating with less robust growth on this substrate. The final ethanol, isobutanol, and hexanol concentrations during growth on the untreated biomass were 0.4 mM, 0.7 mM, and 0.2 mM, respectively (**Figure 3.3, Figure S3.2**).

Controlled fermentation conditions improve alcohol production of the AOR-Adh/Δldh strain

To examine the effect of pH control and gas flow control on the fermentation product profile of the AOR-Adh/Δ*ldh* strain, growth in a 15 L fermentor was followed under controlled

conditions with and without 20 mM isobutyrate (**Figure 3.4**). Cell growth using crystalline cellulose as the carbon source in the fermentor was higher than in closed serum bottles, reaching cell densities from 6-8 x 10^8 cells • mL^{-1}, compared to 4 x 10^8 cells • mL^{-1} in bottles. After 40 h of growth, the AOR-Adh/Δ*ldh* strain produced 3.9 mM ethanol, together with 30.8 mM acetate and 0.9 mM pyruvate. In medium supplemented with 20 mM exogenous isobutyrate, the AOR-Adh/Δ*ldh* strain produced up to 2.3 mM isobutanol, along with 1.2 mM ethanol, 24.0 mM acetate, and 0.4 mM pyruvate after 40 h. The decrease in ethanol and isobutanol concentrations observed near the end of growth was potentially due to evaporation during the continuous sparging of the vessel (**Figure 3.4**).

Products formed do not account for carbon utilized

Under controlled fermentation conditions, closure of the carbon balance (cellulose utilized versus products measured) for the AOR-Adh/Δ*ldh* strain was only 78% after 122 hours (**Figure 3.5**), indicating the formation of one or more unidentified carbon end-products. For controlled fermentation with the addition of isobutyrate, the carbon balance closure was 91% after 155 hours (**Figure 3.5**). Under comparable fermentation conditions, wild type *C. bescii* carbon balances were approximately 100% closed (257).

Discussion

Herein we have established that the expression of the AOR-Adh pathway in *C. bescii* supports the production of alcohols, demonstrating the first example of this pathway in a cellulolytic microbe. While ethanol titers from this pathway are lower than those reported in previously engineered strains of *C. bescii*, the AOR-Adh pathway has the significant benefit of reducing a variety of exogenous organic acids to the corresponding alcohol (270, 292, 293). Expression of the AOR-Adh pathway in a cellulolytic bacterium enables the production of bio-

alcohols, such as isobutanol and hexanol – both of which are high-energy drop-in biofuels – to be powered by the oxidation of sugars from lignocellulosic biomass, an abundant and renewable feedstock.

The energy density (energy per unit volume) of isobutanol (30.4 MJ • L^{-1}) is significantly higher than that of ethanol (21 MJ • L^{-1}) and approaches that of gasoline (31 MJ • L^{-1}) (381, 382). The energy density of hexanol (31.7 MJ • L^{-1}) is even higher than that of isobutanol, and is roughly equivalent to that of gasoline (383). Furthermore, unlike ethanol, neither isobutanol nor hexanol is hygroscopic, and thus, their use minimizes corrosion problems in end-use applications (384). The combination of these properties has made both of these medium-chain alcohols desirable biofuel targets. Isobutanol production from engineered *E. coli* has been achieved with high yield (68% of the theoretical maximum) and titer (50 g • L^{-1}) from growth on glucose (384, 385). *De novo* hexanol production has been demonstrated in prokaryotes through three separate engineered pathways: the keto-acid pathway and two different CoA dependent pathways (reverse β-oxidation and a modified extension of the clostridial butanol pathway) (386). In this work, hexanol is produced by the reduction of exogenously supplemented *n*-caproate. The addition of *n*-caproate is a potentially viable means of achieving hexanol production by engineered *C. bescii* because recent work has shown the high titer (23.4 g • L^{-1}) production of *n*-caproate from the oxidation of lactic acid by a *Clostridium* dominated microbial consortium (387, 388). Alternatively, anaerobic digestion technology has been demonstrated to convert municipal solid waste to *n*-caproate (and a variety of other medium chain fatty acids) in a mixed-culture reactor (89). Anticipating improved AOR-Adh pathway yields in the future, the biological production of *n*-caproate from lactic acid or municipal solid waste may provide an inexpensive feedstock for hexanol production, driven by the oxidation of sugars derived from lignocellulosic biomass.

The reduction of millimolar concentrations of isobutyrate to isobutanol in the *C. bescii* strains containing the AOR-Adh pathway provides secondary confirmation that the pathway is functional, despite low ethanol yields. The general increase in end-product concentrations (hydrogen, acetate, and presumably carbon dioxide) observed in serum bottle growth with isobutyrate for all strains can be attributed to the increased buffering capacity in the physiologically relevant pH range for the isobutyrate supplemented medium (**Figure S3.4**). This conclusion is further supported by the similar concentrations of total end-products generated during growth with and without exogenous acids under pH-controlled fermentation conditions, where initial buffering capacity of the medium was less important (**Figure 3.4**).

In the engineering of microbes for biofuel or bio-commodity production, lactate is an undesirable reduced product. Reducing power in the form of NADH could be better directed toward the desired compound (i.e., ethanol), rather than funneling electrons into lactate formation. Therefore, metabolic engineering efforts toward biofuel production in *C. bescii* have relied on strains lacking *ldh* (291-293). The restoration of *ldh* by movement of IS*Cbe4* is thus unfavorable for biofuel production efforts. This is one of several instances of unanticipated transposon movements in this parent strain background explored in a recent examination of genome stability in *C. bescii* lineages (279). Deleting the restored *ldh* gene from the AOR-Adh strain increased ethanol production by two-fold, likely because reducing power (in the form of NADH), previously utilized by LDH, can instead be leveraged by AdhA to reduce acetaldehyde to ethanol.

Ethanol production by the engineered AOR-Adh pathway reached a maximum of 4 mM ethanol in *C. bescii*. Significantly, the strains expressing the AOR-Adh pathway produced up to 3 mM pyruvate during growth, compared to <0.2 mM for the control strain (data not shown). The production of pyruvate by these strains, rather than its further metabolism to generate ATP and

either acetate or ethanol, implies an imbalance in the redox chemistry of the pathway and presumably the bottleneck is at pyruvate ferredoxin oxidoreductase (POR). The oxidative decarboxylation of pyruvate by POR yields carbon dioxide, acetyl CoA, and, importantly, reduced ferredoxin. In the parent strain, this ferredoxin is oxidized by the bifurcating hydrogenase in combination with NADH to reduce protons to hydrogen gas (352). The AOR-Adh pathway in the recombinant strains may be unable to compete with the hydrogenase for reduced ferredoxin and NADH, even though AOR uses reduced ferredoxin and AdhA has a preference for NADPH over NADH (389). Indeed, the bottleneck at the level of pyruvate might occur by the increase in reduced ferredoxin concentration due to the lack of AOR activity, which is limited by the ability of AdhA to reduce acetaldehyde in a NADPH-dependent reaction. A bottleneck at both POR and AOR would also explain both the lower than expected ethanol yield in these strains, and also the secretion of pyruvate into the growth medium.

Carbon balances were 91% closed during fermentation with the addition of isobutyrate, but only 78% closed during fermentation in its absence in the standard medium. One plausible explanation for this disparity is a technical one stemming from the difference in fermentation times (155 hours compared with 122 hours) and the method used to measure residual sugars in the supernatant. While the assay for total reducing sugar measures soluble polysaccharides resulting from the degradation of cellulose, the HPLC measurement only detects mono- and disaccharides. Hence solubilized polysaccharides would be included as substrates consumed (loss of insoluble cellulose), but not accounted for as products measured in the supernatant. Thus, at 155 hours these solubilized polysaccharides would be degraded beyond the sugars that were measured. This hypothesis aligns with the observation that the largest difference in carbon balances between these experiments is a shift in the amount of cellobiose and glucose present in the supernatant at the end

of growth (**Figure 3.5**). Generally, glucose and cellobiose begin to accumulate in the supernatant of *C. bescii* fermentation on cellulose after 40 hours, and continue until at least 150 hours (257). For insoluble growth substrates, such as cellulose, carbon balances must be calculated at the end of growth, when residual cellulose can be dried and weighed. It is also possible that the presence of an exogenously provided electron acceptor (isobutyrate) disentangles carbon and electron flow by providing a way to dispose of electrons that, unlike acetate, is not directly derived from sugar oxidation. Further evidence for this is provided by the decrease in pyruvate production with the addition of isobutyrate (2.8 mM vs. 0.9 mM). As noted above, pyruvate secretion has been proposed to indicate unbalanced redox metabolism in engineered strains of *C. bescii* (270).

Notably, unknown products have been previously implicated in other engineered strains of *C. bescii* that lack functional *ldh*. In one case, the introduction of a non-redox balanced ethanol production pathway in a Δ*ldh* background increased the closure of carbon balances at 60 °C, significantly lowering the relative amount of carbon going to unknown products (270). The addition of another enzyme to make the pathway redox balanced completely eliminated formation of unknown products, allowing approximately 100% closure of the carbon balance at 60 °C (270). Taken together with the previous results, it seems likely that the unknown products are generated as a means of overcoming metabolic perturbations in strains lacking redox balanced pathways.

An understanding of how the NADH and NADPH pools are maintained in *C. bescii* is incomplete. For the direct interconversion between NADH and NADPH, *C. bescii* does not encode a canonical transhydrogenase in its genome (246) but does contain a bifurcating NADH-dependent reduced ferredoxin $NADP^+$ oxidoreductase (Nfn, Athe_0644-0645), which is capable of oxidizing both NADH and reduced ferredoxin to generate two NADPH, but the role of this enzyme in regulating or maintaining these three redox cofactor pools is unclear (**Figure 3.1**) (130, 239, 390).

Understanding redox cofactor recycling mechanisms in *Caldicellulosiruptor* spp. is essential for improving the yields of heterologously expressed pathways in future metabolic engineering efforts.

Low ethanol yields may also result from a mismatched ferredoxin cofactor for *P. furiosus* AOR in *C. bescii*. Production of ethanol in the AOR-Adh pathway strains indicates that the 8Fe-ferredoxin from *C. bescii* is, indeed, capable of transferring electrons to the AOR from *P. furiosus*. However, the ten-fold higher K_M for the *C. bescii* 8Fe-ferredoxin compared to the *P. furiosus* 4Fe-ferredoxin (14.9 μM vs. 149.0 μM) indicates that a high intracellular concentration of reduced ferredoxin might be necessary for pathway functionality in engineered strains; saturation kinetics were not observed with purified AOR at the measured concentrations of *C. bescii* ferredoxin (**Figure S3.3**). The low affinity of *C. bescii* ferredoxin for the heterologously-expressed *P. furiosus* AOR could help to explain low ethanol yields in strains containing the AOR-Adh pathway. Rather than interacting with AOR, reduced ferredoxin at low intracellular concentrations might be more prone to interact with native *C. bescii* enzymes, such as the bifurcating hydrogenase, for which it likely has a higher affinity.

The high specific activities for heterologously produced AOR and AdhA enzymes in cell-free extracts of *C. bescii* indicate that the alcohol yields are not due to lack of the relevant enzymes, but rather an imbalance in pathway stoichiometry of redox cofactors. The specific activity of AOR in *P. furiosus* cell free extract (S100) has been measured as 4.0 ± 0.1 units • mg^{-1} at 75 °C, which is comparable to the 3.3 ± 0.5 units • mg^{-1} measured in whole cell extract of the recombinant AOR-Adh/*Δldh* strain of *C. bescii* (272). Likewise, cell-free extracts of recombinant *P. furiosus* expressing the *T.* sp. X514 AdhA exhibited 0.9 units • mg^{-1} NADPH linked Adh activity at 65 °C, compared to 0.8 ± 0.1 units • mg^{-1} in the AOR-Adh/*Δldh* strain of *C. bescii* at 75 °C (389). The

cell-free extract of *T.* sp. X514 reportedly contains 3.4 units • mg^{-1} of NADPH linked Adh activity at 65 °C, but much of this activity likely arises from a combination of AdhA homologs (e.g., BdhA) and the Adh domain of the bifunctional AdhE present in this species (389). Nevertheless, the AOR-Adh pathway in these engineered strains is clearly inefficient and it appears to be acting as an overflow valve rather than a primary metabolic channel. Hence, the challenges of pathway bottlenecks and cofactor mismatches must be overcome if we are to significantly improve yields. Notwithstanding, the demonstrated functionality of the AOR-Adh pathway in *C. bescii* presents a promising avenue for future metabolic engineering efforts toward bio-alcohol production in this cellulolytic extreme thermophile.

Conclusions

This study is the first demonstration of using the AOR-Adh pathway for bio-alcohol production in a cellulolytic microbe, enabling the conversion of untreated biomass to ethanol by a bacterium that does not natively produce alcohols. Furthermore, the addition of exogenous organic acids to the growth medium allowed engineered strains of *C. bescii* to reduce these acids to the corresponding alcohol – namely, the next generation biofuels isobutanol and hexanol. Future metabolic engineering efforts should improve redox balance in engineered strains, and thereby increase product yields. The capability of this pathway to generate a range of alcohols with varying chain lengths (C_1-C_6) makes it an attractive target for microbial biofuel generation from plant biomass.

Acknowledgements

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DE-SC0019391) and by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center at Oak Ridge National Laboratory. We

thank Israel Scott, Laura Lee, Piyum Khatibi, Jeff Zurawski, Jonathan Conway, Diep Nguyen, Farris Poole, Mirko Basen, and Matthew Keller for helpful discussions, and Daewhan Chung and Janet Westpheling for providing *C. bescii* strain JWCB018 and plasmids pDCW121 and pDCW88.

Tables and Figures

Table 3.1 *C. bescii* strains used and generated in this study

Strain	Alias	Parent	Transforming Plasmid	Description	Reference
DSMZ 6725	Wild type	---	---	Wild type	(268)
JWCB018	Control	JWCB005	pDCW88	Δ*pyrFA*, Δ*cbeI*, *ldh*::IS*Cbe4*	(276)
MACB1013	AOR-Adh	JWCB018	pGR002	Δ*pyrFA*, Δ*cbeI*, $P_{S\text{-}layer}$ *aor adhA* (PF0346, Teth514_0564)	(279)
MACB1038	AOR-Adh/Δ*ldh*	MACB1013	pDCW121	Δ*pyrFA*, Δ*cbeI*, Δ*ldh*, $P_{S\text{-}layer}$ *aor adhA* (PF0346, Teth514_0564)	This study

Table 3.2 Primers used in this study

Primer	Target Locus[a]	Sequence (5' to 3')
GR012	*cbeI / slp*	CAATTTGCCTTCCAACATGATAGGATCCCGGGACAGGATTTAAAAGA
GR013	*cbeI / slp*	TCTTTTAAATCCTGTCCCGGGATCCTATCATGTTGGAAGGCAAATTG
GR014	*cbeI / T.* sp. X514 *adhA*	GGGTTGCATGCGGCCGCAGATTAGGTAGGTGGTTTTAACA
GR015	*cbeI / T.* sp. X514 *adhA*	TGTTAAAACCACCTACCTAATCTGCGGCCGCATGCAACCC
LDH_conf R	*ldh*	CAGCCCTCAAACAAATAGCATTTT
LDH_conf F	*ldh*	CAAGCGAATAGTCCCTGTTAGAT
GR_adhA_F	*T.* sp. X514 *adhA*	GGAGGTTTGGTGAGTAGTTATGTGGGAAACAAAAATAAAT
GR122gibson2 F	*T.* sp. X514 *adhA*	AATTTAATTTAAAAATTTCACAGCAAG
GR_s-pro_R	*slp / T.* sp. X514 *adhA*	ATTTATTTTTGTTTCCCACATAACTACTCACCAAACCTCC
(MB)adhA qF	*T.* sp. X514 *adhA*	GGCCTGGGATGTGGTAAAACC
(MB)adhA qR	*T.* sp. X514 *adhA*	GCATTTTTATCAGTGTATTCCAGCAG
GR122fusionAOR R	*P. furiosus aor / T.* sp. X514 *adhA*	TGTGAAATTTTTAAATTAAATTTCAATAGAACTCAGCGATTC
GR122fusionAOR F	*slp / T.* sp. X514 *adhA*	TTTATGAAGAATCTTTCTAAAGGAGGTTTGGTGAGTAG
(MB)Cb001	*slp*	ATCCCGGGACAGGATTTAAAAGAGGCTATG
(MB)Cb009	*cbeI*	GTTGGAAGAAGCTGTTGTTCACTATG
(MB)Cb010	*cbeI*	CAAGTCTCCATTTTTTCCATACAGGC
GL313	*P. furiosus aor*	CATCAAAGACGAGCACATTGAG
GL314	*P. furiosus aor*	GCGAACTTGACCAAATTCTCAC
GLCB003	*ldh*	CTGCAATTAAAGGACCCAAATACG
GLCB004	*ldh*	GTTTGTGCAACCTTCTATGCC
GLCB005	*cbeI*	CTGGTGTATCTCCTGCCAAG
GLCB006	*cbeI*	GGCGAGATTCAGTGGTACTG
GLCB007	*pyrF*	TTGTTGAACTTCTCTGATAGGC
GLCB008	*pyrF*	TTAAGAGATTGCTGCGTTGATAC
GLCB010	pDCW88	CAACCCGTTCCATGTGCTC
GLCB012	*E. coli aac*	ACGATCAACGGCACTGTTG
GLCB044	*cbeI*	TTCAAAACTCCTTCTGTCGAAGG

[a]All primers were designed to target the *C. bescii* genome unless otherwise noted

Table 3.3 AOR and AdhA specific activities in *C. bescii* crude extracts

Strain	Enzyme	Substrate	Electron Carrier	Specific Activity[a]
AOR-Adh/Δ*ldh*	AOR	Acetaldehyde	Benzyl viologen	3.3 ± 0.5
Wild type	AOR	Acetaldehyde	Benzyl viologen	<0.01
AOR-Adh/Δ*ldh*	ADH	Acetaldehyde	NADPH	0.8 ± 0.2
Wild type	ADH	Acetaldehyde	NADPH	<0.01

[a]μmol activity min^{-1} mg^{-1} protein ± standard deviation (σ)

Figure 3.1 Primary metabolism and engineered ethanol pathways in *C. bescii*

Primary redox cofactors are NAD/H (gray italics), an 8Fe-ferredoxin (bold) and NADP/H (underlined). Native enzymes are shown in gray boxes. Recombinantly-expressed enzymes are outlined in black boxes, and two different non-native pathways for ethanol production are indicated by dotted lines and gray arrows. Abbreviations: bifurcating hydrogenase (BF-H_2ase), glyceraldehyde 3-phosphate dehydrogenase (GAPDH), glyceraldehyde 3-phosphate oxidoreductase (GOR), phosphoglycerate kinase (PGK), pyruvate kinase (PK), lactate dehydrogenase (LDH), pyruvate oxidoreductase (POR), phosphotransacetylase (PTA), acetate kinase (ACK), membrane bound hydrogenase (MBH), bifurcating NADH-dependent reduced ferredoxin $NADP^+$ oxidoreductase (BF-Nfn), aldehyde ferredoxin oxidoreductase (AOR), bifunctional aldehyde/alcohol dehydrogenase (AdhE), primary alcohol dehydrogenase (AdhA).

Figure 3.1

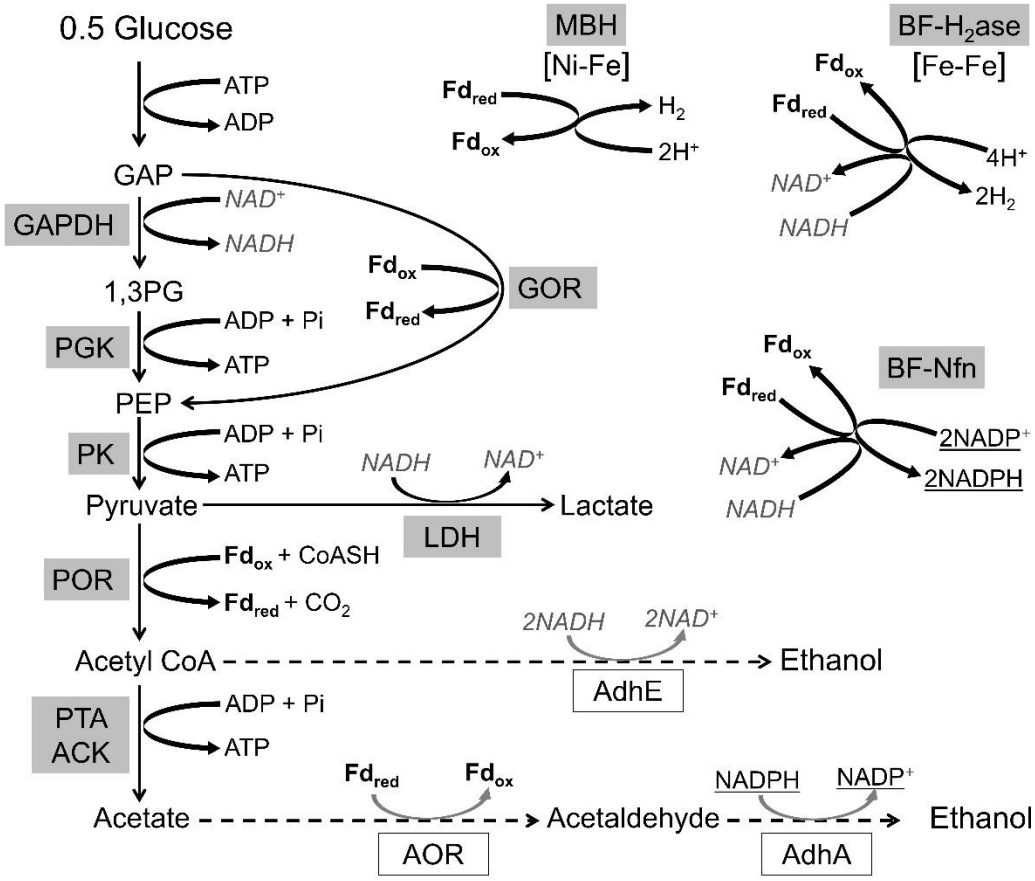

0.5 Glucose
ATP
ADP
GAP
GAPDH
NAD+
NADH
1,3PG
PGK
ADP + Pi
ATP
PEP
PK
ADP + Pi
ATP
Pyruvate
NADH
NAD+
Lactate
LDH
POR
Fdox + CoASH
Fdred + CO2
Acetyl CoA
2NADH
2NAD+
Ethanol
AdhE
PTA
ACK
ADP + Pi
ATP
Acetate
Fdred
Fdox
AOR
Acetaldehyde
NADPH
NADP+
Ethanol
AdhA
GOR
Fdox
Fdred
MBH
[Ni-Fe]
Fdred
H2
Fdox
2H+
BF-H2ase
[Fe-Fe]
Fdox
Fdred
4H+
NAD+
2H2
NADH
BF-Nfn
Fdox
Fdred
2NADP+
NAD+
2NADPH
NADH

Figure 3.2 Genotypes of engineered *C. bescii* strains at the *cbeI* and *ldh* loci

A). Schematic showing the structure of the genome at the *ldh* (left) and *cbeI* (right) loci in the wild type, control, AOR-Adh, and AOR-Adh/Δ*ldh* strains. Native *ldh* and *cbeI* are shown in black while a native insertion sequence, *ISCbe4*, is shown in hatch, and recombinantly expressed genes are show in white. P_{slp} is the 200 bp promoter region from the *C. bescii* S-layer protein gene. *aor* is the gene encoding AOR in *P. furiosus* and *adhA* is the gene encoding AdhA in *Thermoanaerobacter* sp. X514. The black lines labelled a-c and d-f correspond to the PCR products amplified from each strain using primers targeting either the *ldh* or *cbeI* locus, as appropriate. PCR product sizes: a, 2427 bp; b, 4053 bp; c, 1474 bp; d, 2205 bp; e, 1167 bp; f, 4668 bp. B). DNA agarose gel showing the PCR products amplified from each strain at either locus, *ldh* (left) or *cbeI* (right). Bands labeled a-f correspond to the lines and product sizes indicated in panel A.

Figure 3.2

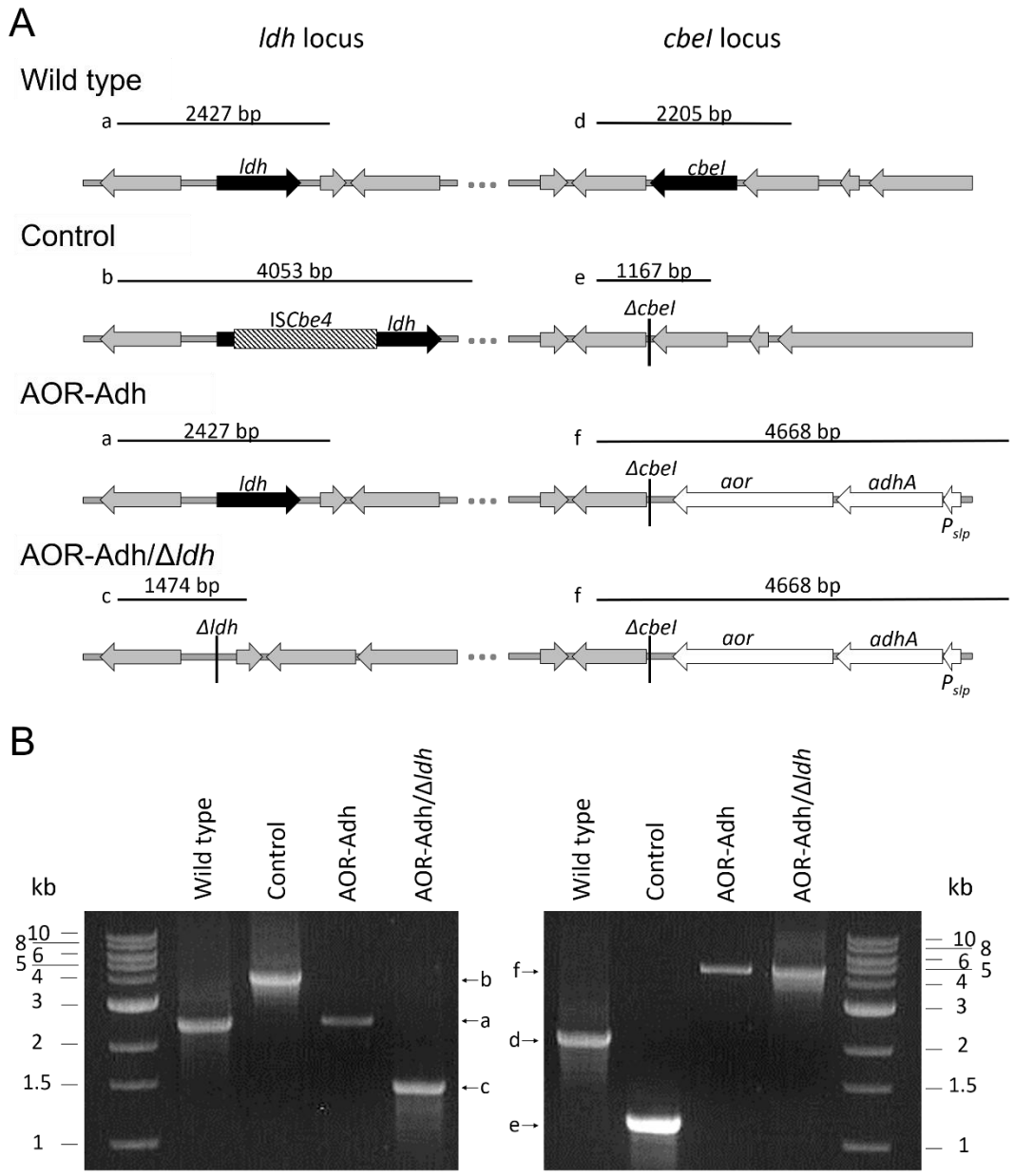

A
ldh locus
cbeI locus
Wild type
a
2427 bp
ldh
d
2205 bp
cbeI
Control
b
4053 bp
ISCbe4
ldh
e
1167 bp
ΔcbeI
AOR-Adh
a
2427 bp
ldh
f
4668 bp
ΔcbeI
aor
adhA
P slp
AOR-Adh/Δldh
c
1474 bp
Δldh
f
4668 bp
ΔcbeI
aor
adhA
P slp
B
kb
Wild type
Control
AOR-Adh
AOR-Adh/Δldh
10
8
6
5
4
3
2
1.5
1
b
a
c
f
d
e
Wild type
Control
AOR-Adh
AOR-Adh/Δldh
kb
10
8
6
5
4
3
2
1.5
1

Figure 3.3 Alcohol production by strains expressing the AOR-Adh pathway

Alcohol production by *C. bescii* strains expressing the AOR-Adh pathway during growth on different substrates. Strains were grown on 5 g • L^{-1} cellobiose (white) or crystalline cellulose (gray) for 30 h, or untreated switchgrass (black) for 128 h, at 75 °C with shaking at 150 rpm. A). Production of ethanol by the AOR-Adh/Δ*ldh* strain compared to the control strain. B). Production of isobutanol by the AOR-Adh/Δ*ldh* strain compared to the control strain. Strains were grown with the addition of 20 mM isobutyrate to the growth medium. C). Production of hexanol by the AOR-Adh/Δ*ldh* strain compared to the control strain. Strains were grown with the addition of 10 mM *n*-caproate to the growth medium. Section signs (§) represent alcohol concentrations below the instrument's limit of detection. All experiments were performed in biological triplicate (n=3) with error bars representing standard deviation (σ).

Figure 3.3

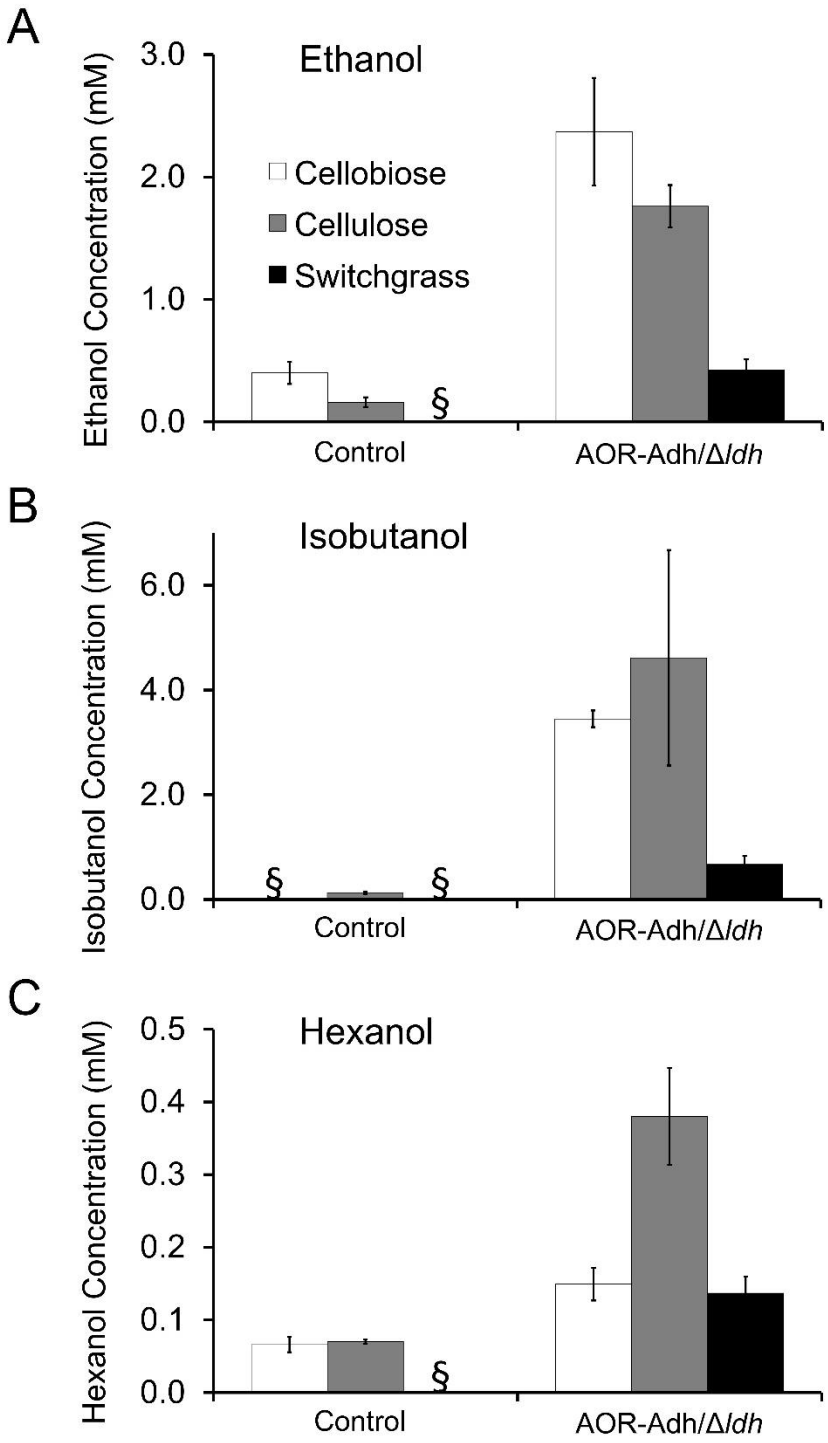

A
Ethanol
Ethanol Concentration (mM)
3.0
2.0
1.0
0.0
Cellobiose
Cellulose
Switchgrass
§
Control
AOR-Adh/Δldh
B
Isobutanol
Isobutanol Concentration (mM)
6.0
4.0
2.0
0.0
§
§
Control
AOR-Adh/Δldh
C
Hexanol
Hexanol Concentration (mM)
0.5
0.4
0.3
0.2
0.1
0.0
§
Control
AOR-Adh/Δldh

Figure 3.4 Growth of the AOR-Adh/Δ*ldh* strain under controlled fermentation conditions

A 15 L culture of the AOR-Adh/Δ*ldh* strain was grown on 20 g • L^{-1} crystalline cellulose in a 20 L fermentor with pH (6.2), temperature (75 °C), impeller agitation (150 rpm), and gas-flow control (1.5 L • min^{-1}) in standard medium without (A) and with (B) the addition of 20 mM isobutyrate. Growth was monitored by cell density (diamonds, dotted line). Concentrations in the medium of ethanol (triangles, dashed line), pyruvate (circles), free sugars measured as cellobiose and glucose combined (crosses), and isobutanol (squares, dashed line) over the course of fermentation are shown.

Figure 3.4

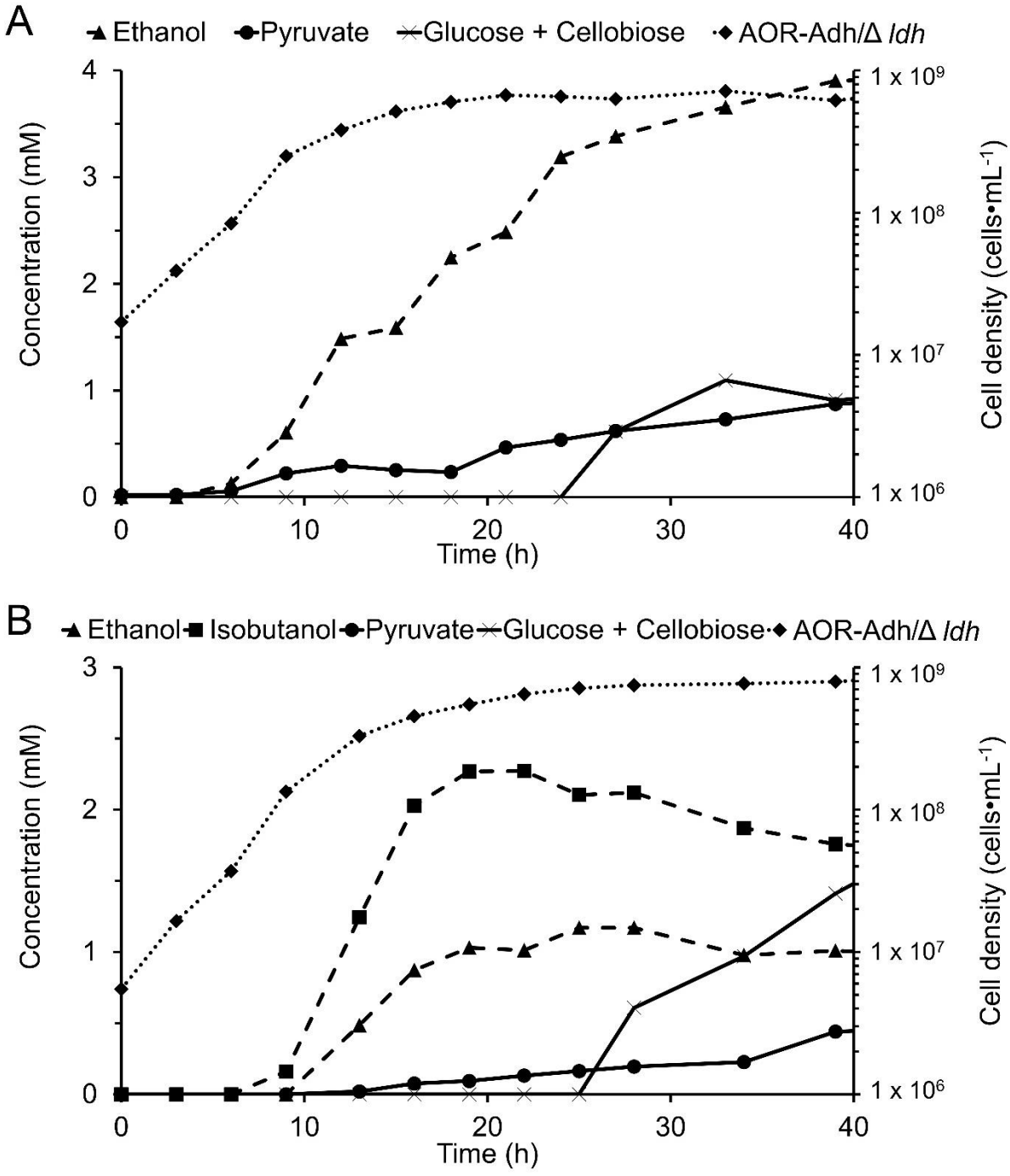

A
Ethanol
Pyruvate
Glucose + Cellobiose
AOR-Adh/Δ *ldh*
Concentration (mM)
Cell density (cells•mL^{-1})
Time (h)
B
Ethanol
Isobutanol
Pyruvate
Glucose + Cellobiose
AOR-Adh/Δ *ldh*
Concentration (mM)
Cell density (cells•mL^{-1})
Time (h)

Figure 3.5 Carbon balances for the AOR-Adh/Δ*ldh* strain under controlled fermentation conditions

Substrate utilized versus products quantified for the AOR-Adh/Δ*ldh* strain under controlled fermentation conditions. The AOR-Adh/Δ*ldh* strain was grown on 20 g • L^{-1} crystalline cellulose without (left) and with (right) the addition of 20 mM exogenous isobutyrate. Products in the carbon balance include: biomass formation estimated from total protein (checker), ethanol (black), acetate (cross hatch), pyruvate (gray), carbon dioxide (stripe), glucose (white) and cellobiose (hatch).

Figure 3.5

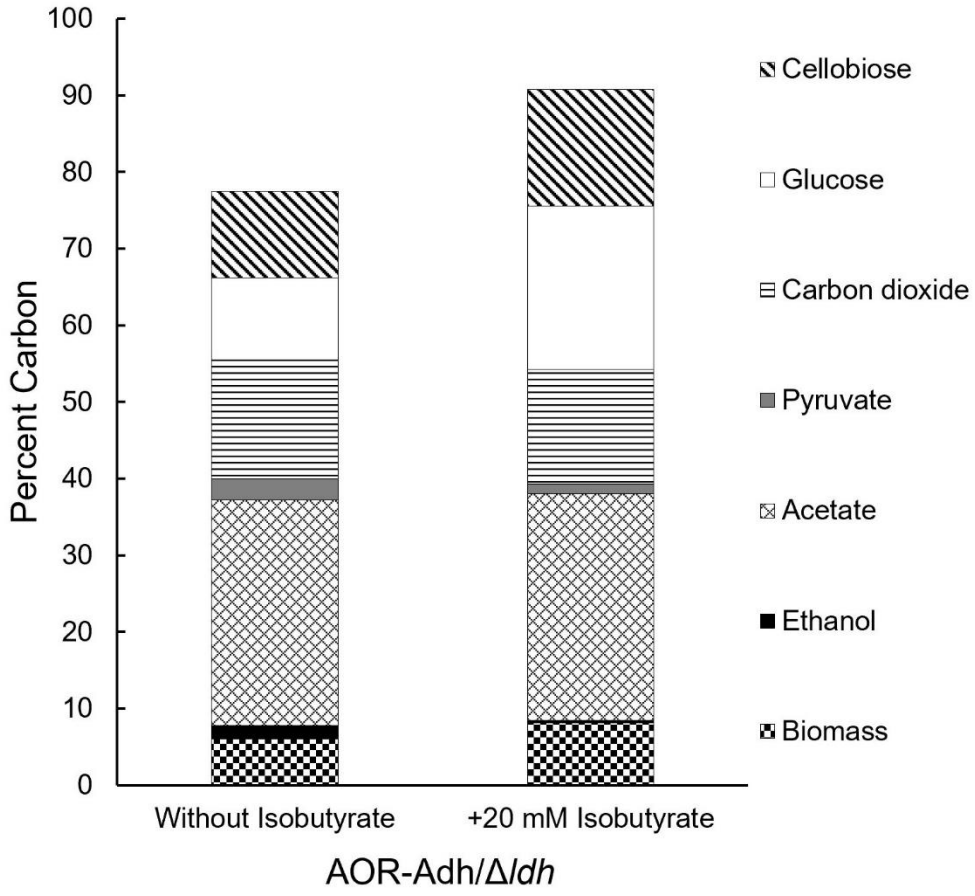

100
90
80
70
60
50
40
30
20
10
0
Percent Carbon
Cellobiose
Glucose
Carbon dioxide
Pyruvate
Acetate
Ethanol
Biomass
Without Isobutyrate
+20 mM Isobutyrate
AOR-Adh/Δ*ldh*

Figure S3.1 Growth and fermentation products of the AOR-Adh strain over time in sealed Hungate tubes

The control strain (A) and the AOR-Adh strain (B) were grown side by side with or without the addition of 20 mM isobutyrate to the growth medium (C, D) for 30 hours. Cultures were grown in 30 mL crimp sealed Hungate tubes containing 15 mL of culture volume. During growth, culture and headspace samples were taken to monitor the concentrations of hydrogen (■), acetate (▲), ethanol (●), isobutanol (+), and isobutyrate (×) in the growth medium. Cell growth was measured by OD_{680} (♦).

Figure S3.1

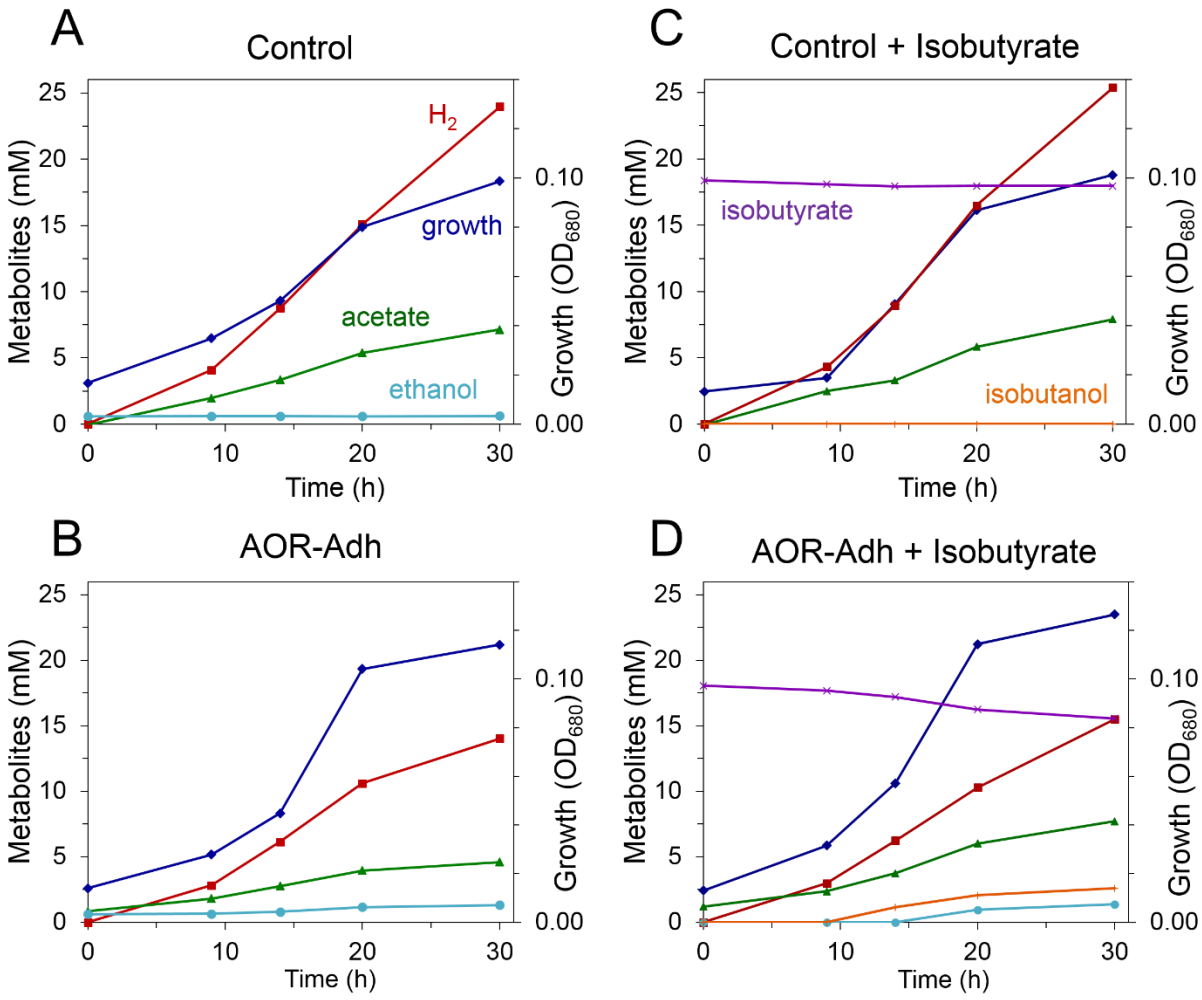

A
Control
C
Control + Isobutyrate
B
AOR-Adh
D
AOR-Adh + Isobutyrate
Metabolites (mM)
Growth (OD680)
Time (h)
H2
growth
acetate
ethanol
isobutyrate
isobutanol

Figure S3.2 Effects of carbon substrate and exogenous acid addition on growth

The control strain (♦) and the AOR-Adh/Δ*ldh* strain (▲) were grown on 5 g/L cellobiose, crystalline cellulose, or switchgrass in triplicate with no supplemented organic acid (A-C), 20 mM exogenous isobutyrate (isb) (D-F), or 10 mM exogenous n-caproate (cap) (G-I). The AOR-Adh strain (■) was grown on 5 g/L cellobiose in triplicate with no supplemented organic acid (A) or with 20 mM exogenous isobutyrate (D). All three strains grew at similar rates and to a similar final density (1-4 x 10^8 cells mL^{-1}), with *n*-caproate supplemented medium and growth on switchgrass yielding the lowest cell densities.

Figure S3.2

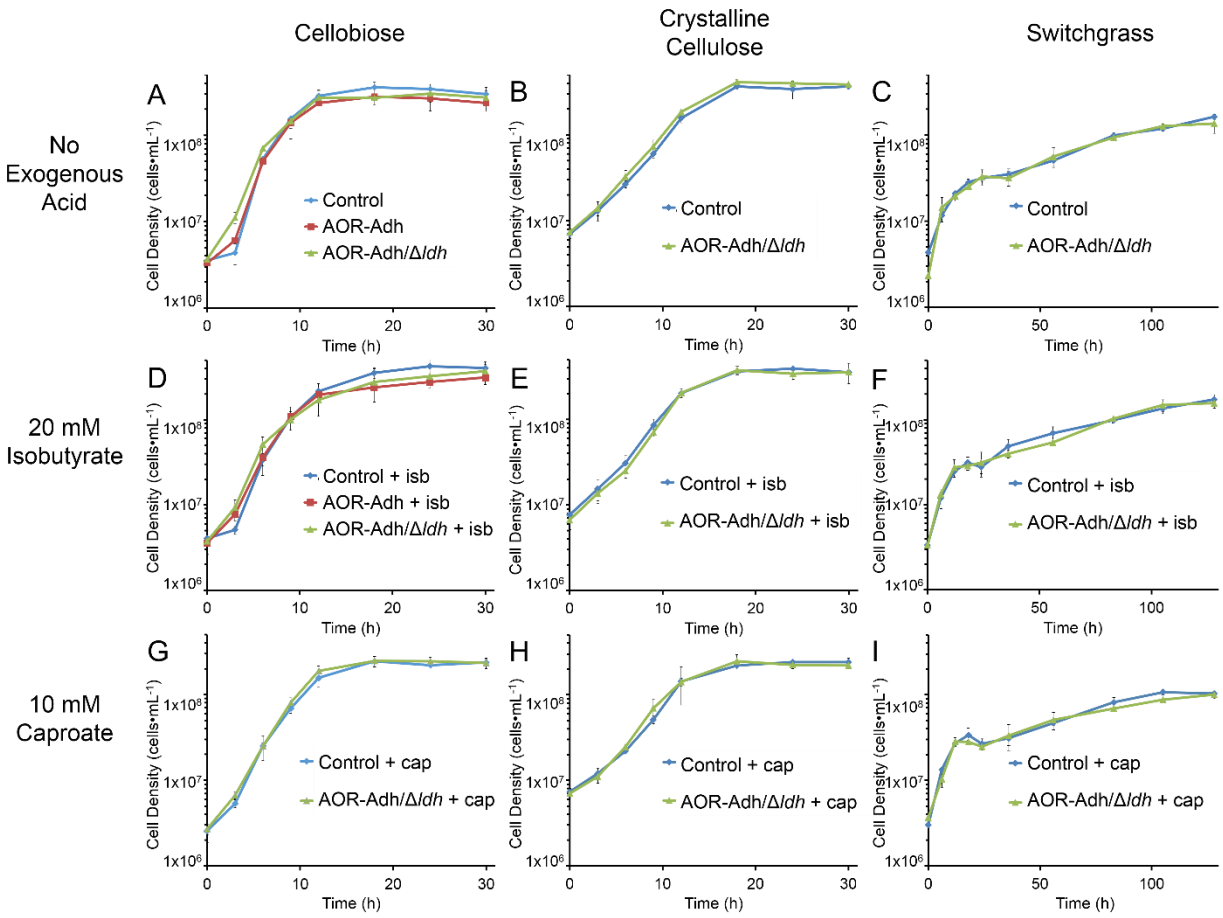

Cellobiose
Crystalline Cellulose
Switchgrass
No Exogenous Acid
20 mM Isobutyrate
10 mM Caproate
A
B
C
D
E
F
G
H
I
Cell Density (cells•mL⁻¹)
Time (h)
Control
AOR-Adh
AOR-Adh/Δldh
Control + isb
AOR-Adh + isb
AOR-Adh/Δldh + isb
Control + cap
AOR-Adh/Δldh + cap

Figure S3.3 Kinetic properties of *P. furiosus* AOR with *P. furiosus* and *C. bescii* ferredoxins

AOR activity was measured at 75 °C by the crotonaldehyde (200 µM) dependent reduction of ferredoxin at 425 nm in 50 mM EPPS buffer (pH 8.0). *P. furiosus* AOR was less active with *C. bescii* ferredoxin (■) than with *P. furiosus* ferredoxin (●). Both trendlines were calculated using a non-linear regression to fit the Michaelis-Menten equation.

Figure S3.3

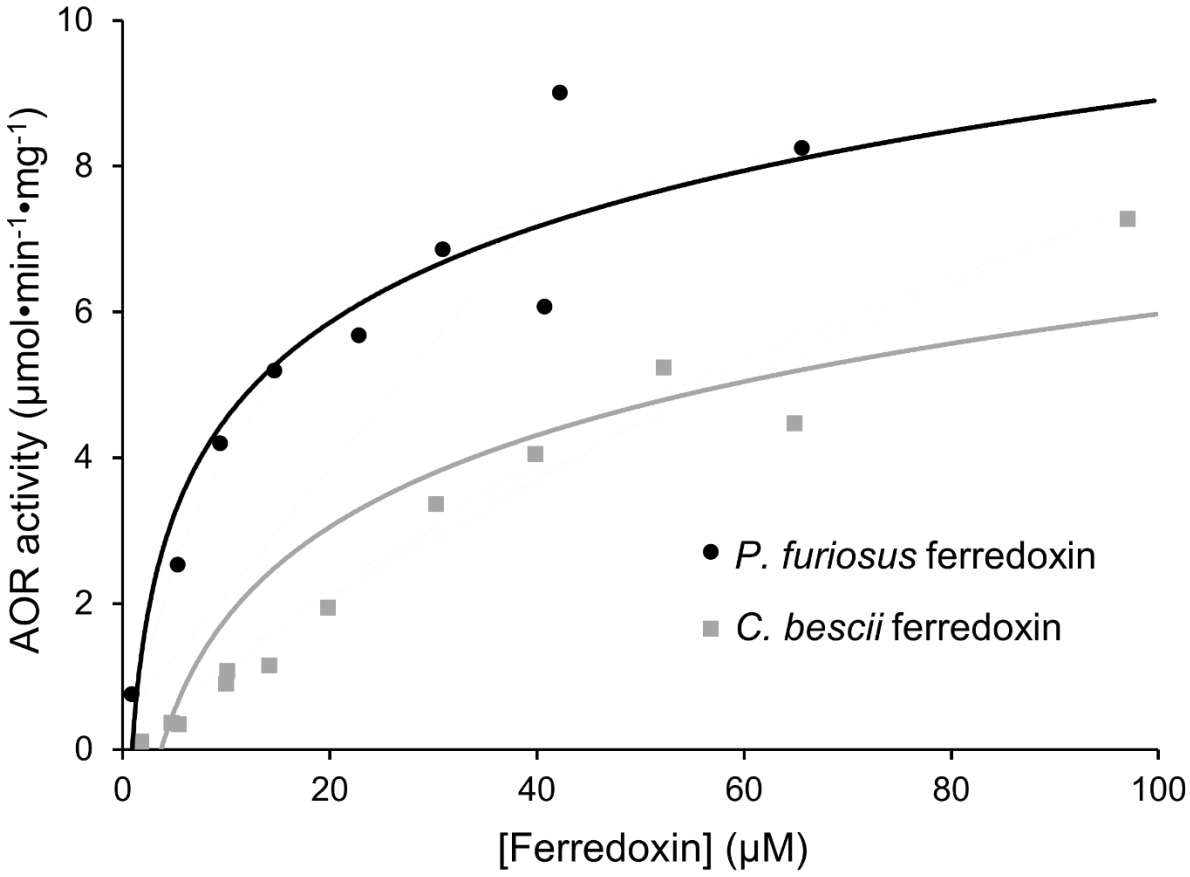
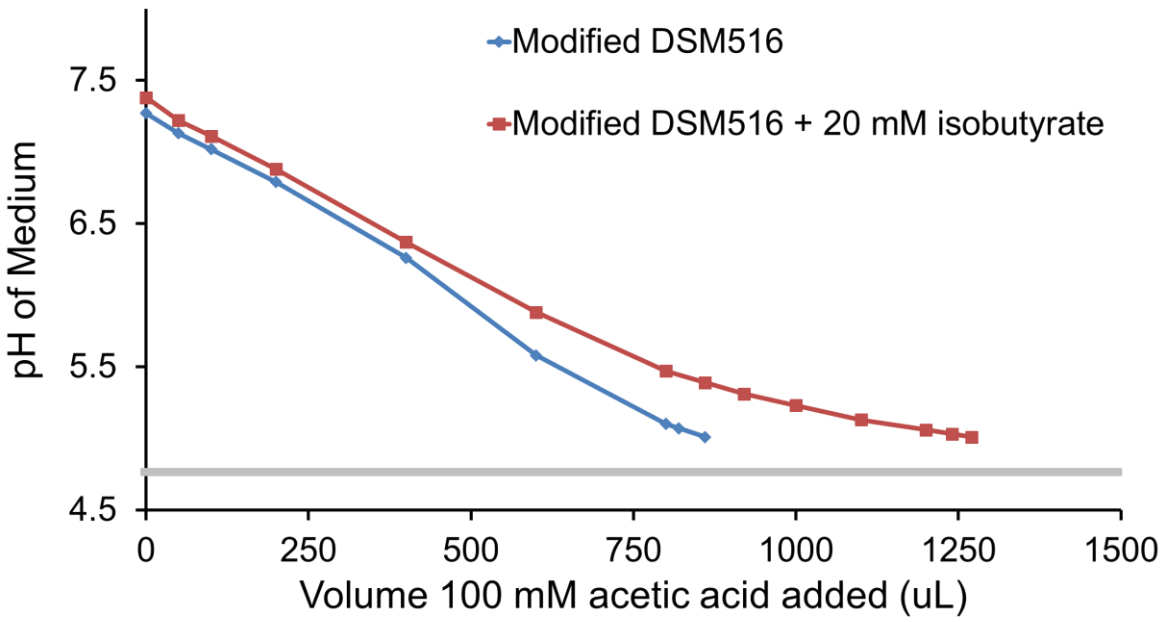

AOR activity (µmol•min-1•mg-1)
[Ferredoxin] (µM)
P. furiosus ferredoxin
C. bescii ferredoxin
0
2
4
6
8
10
0
20
40
60
80
100

Figure S3.4 Impact of isobutyrate on buffering capacity of modified DSM516 medium

Modified DSM516 medium (♦) and modified DSM516 medium supplemented with 20 mM isobutyrate (■) were prepared as described. With continuous pH monitoring, 100 mM acetic acid was slowly added to 5 mL of media until a final pH of 5.01 was achieved. The pKa of isobutyrate (4.8) and the pK_a of acetate (4.75) are represented by the gray line. The addition of 20 mM isobutyrate increased the medium's capacity for acetate by almost 1.5-fold.

Figure S3.4


Modified DSM516
Modified DSM516 + 20 mM isobutyrate
pH of Medium
7.5
6.5
5.5
4.5
0
250
500
750
1000
1250
1500
Volume 100 mM acetic acid added (uL)

CHAPTER 4

UNDERSTANDING SUGAR UPTAKE AND ELIMINATING UNDESIRABLE PRODUCTS AS STRATEGIES TO IMPROVE BIOPRODUCT FORMATION BY *CALDICELLULOSIRUPTOR BESCII*

Summary

Caldicellulosiruptor bescii is the most thermophilic cellulolytic microbe identified to date (T_{opt} 78 °C). Due to its rare ability to grow on non-pretreated plant biomass as a sole carbon source and because a genetic system is available, this organism is a potential platform for consolidated bioprocessing to generate bioproducts. *C. bescii* has been engineered to produce desirable non-native end-products, such as ethanol and acetone, but product titers are too low for industrial relevance. To examine the feasibility of improving product formation in *C. bescii*, this work took two approaches: examining sugar uptake to better understand carbohydrate utilization and redirecting undesirable side-products to the target compound by alleviating metabolic bottlenecks. For the first approach, the examination of carbohydrate uptake, we identified a promiscuous ATPase responsible for energizing numerous CUT2 family ABC transporters for sugar import. Deletion of this ATPase, *msmK*, resulted in diminished growth phenotypes on all tested di- and oligosaccharides, while growth on monosaccharides was unimpaired. Understanding sugar uptake is essential to improving carbohydrate utilization in this organism. The second approach to increasing bioproduct formation was to decrease an undesirable side-product, pyruvate, in a previously engineered ethanologenic strain of *C. bescii.* A pyruvate oxidoreductase (*pfor3*) from *Clostridium thermocellum* (T_{opt} 60 °C) was heterologously expressed in the ethanol producing

strain, yielding the low temperature POR strain (LT-POR). While POR activity measured *in vitro* increased in the LT-POR strain, genome rearrangement during strain construction disrupted the heterologously expressed AdhE, eliminating ethanol production.

Introduction

Transportation energy needs continue to rise, even as perils from increasing fossil fuel consumption mount (2, 16). To decrease greenhouse gas emissions while meeting global energy demand, non-hydrocarbon alternative fuels must be developed and implemented. To this end, ethanol is a stop-gap biofuel additive that is blended with gasoline. Almost all global ethanol supply is generated by fermentation of corn, sugarcane, or sugar beets—a dated and wasteful process (1, 69). To decrease both cost and greenhouse gas emissions, efforts are being made to develop methods for ethanol production from an abundant and renewable feedstock: lignocellulosic biomass (94).

Several approaches to produce ethanol from lignocellulosic biomass are under investigation, including consolidated bioprocessing (CBP). The largest advantage of CBP over other methods of lignocellulosic ethanol production is decreasing or eliminating the need for the addition of expensive exogenously produced sugar-hydrolyzing enzymes (157). Another prospective benefit of CBP is the potential for decreasing the necessity of pretreating biomass, specifically by utilizing cellulolytic extreme thermophiles that can directly ferment lignocellulose to ethanol at high temperature in a one-step process (159). Advantages of utilizing thermophiles include decreased risk of contamination, enhanced lignocellulose breakdown, and potential for separating volatile products like ethanol directly from the fermentation vessel (159, 239). Currently, no known cellulolytic thermophile can natively ferment non-pretreated biomass directly to ethanol at above trace titers.

Caldicellulosiruptor bescii is an anaerobic thermophilic bacterium, with the highest optimum growth temperature of any known cellulolytic organism (T_{opt} 78 °C) (255). With the ability to ferment both C_6 and C_5 based carbohydrates in lignocellulosic biomass to acetate, lactate, carbon dioxide, and hydrogen without any pretreatment and because a genetic system exists, *C. bescii* is a promising CBP platform microbe (256, 257). To this end, several strains have been engineered to produce ethanol, using different recombinant expression strategies, at temperatures ranging from 60-75 °C (270, 292, 293, 391). The greatest titer of ethanol currently achieved for an engineered strain of *C. bescii* is 75 mM (3.5 g/L) from the fermentation of crystalline cellulose under controlled fermentation conditions (gas sparging and pH control) (270). Under non-controlled fermentation conditions in sealed serum bottles, the same strain produced 33 mM (1.5 g/L) ethanol. This strain expresses the bifunctional alcohol/aldehyde dehydrogenase, *adhE*, from *Clostridium thermocellum* (T_{opt} 60 °C) and the membrane-bound Rnf complex, *rnfCDGEAB*, from *Thermoanaerobacter* sp. X514 (T_{opt} 60 °C) (**Figure 4.1**). These genes were expressed under the control of native high-level constitutive *C. bescii* promoters, and in accordance with the sources of the heterologously expressed genes, ethanol production was maximal when growth was conducted at 60 °C.

AdhE is the canonical enzyme for ethanol production in thermophilic bacteria, reducing acetyl-CoA produced by pyruvate oxidoreductase (POR) in primary metabolism to acetaldehyde and then ethanol, with each step oxidizing NADH (239). The heterologously expressed Rnf complex links ferredoxin and NADH pools, the two primary redox cofactors in *C. bescii*. To this end, reduced ferredoxin is oxidized while NAD^+ is reduced, and an ion is pumped across the membrane. Debate is ongoing as to the nature of the ion—hydrogen or sodium—pumped across by the *Thermoanaerobacter* sp. X514 Rnf, as forms of Rnf that utilize each have been identified

and this specificity cannot be bioinformatically resolved (392, 393). Expression of the Rnf complex should increase the pool of NADH available for ethanol production by AdhE. Interestingly, ethanol production is identical in strains expressing only AdhE or AdhE and Rnf during growth in sealed bottles, but maximum ethanol production increases by 25% with Rnf expression under pH-controlled and gas-sparging fermentation conditions. Also of note, expression of the same AdhE under the same promoter at the same genomic locus in two different background strains (JWCB018: Δ*pyrFA* Δ*cbeI* IS*Cbe4*::*ldh*, MACB1034: Δ*pyrE* Δ*ldh*) resulted in the production of dramatically different amounts of ethanol. In the older parent strain (JWCB018), ethanol production was roughly 15 mM, as previously reported (292). However, in the newer genomic background (MACB1034) the ethanol production under identical conditions doubled (30 mM). The differences between these genomic backgrounds were explored in a recent publication, wherein JWCB018 was found to contain significant genome rearrangements and high instances of the insertion sequence IS*Cbe4* relative to both the wild type and MACB1034 (279).

While *C. bescii* grows to similar cell densities and produces the same amount of fermentation products at 60 °C and 78 °C, the growth rate is approximately cut in half at the lower temperature (270, 292). In addition to slower growth, certain carbon end-products are produced at 60 °C that are not produced at higher growth temperatures – specifically pyruvate and acetoin are released into the growth medium during low temperature growth (270). While these products are also observed in the wild type during growth at 60 °C, when metabolism is altered, the ratios of these side products to sugar-consumed increases. During growth at 75 °C, little or no pyruvate or acetoin production occurs in the wild type, however pyruvate is produced in Δ*ldh* strains (391). This makes intuitive sense, as these strains are deprived of one potential reduced fate for pyruvate (lactate). Under controlled fermentation conditions on 20 g/L crystalline cellulose, the AdhE-Rnf

strain produced 75 mM ethanol, 52 mM acetate, and 11 mM pyruvate, while only trace acetoin was detected, unlike during growth in sealed bottles with no pH control or gas exchange (3 mM) (270).

Two limiting factors of bioproduct formation for *C. bescii* are product inhibition and sugar uptake. Early studies of *C. bescii* concluded that growth was inhibited by osmolarity of 250 mOsmol, but a subsequent study implied growth in medium with >550 mOsmol is feasible, albeit with a significant (>50h) lag phase (257, 394). This subsequent work also indicated that the more significant limit on growth is organic acid formation. In sealed bottles with no pH control or gas exchange, growth is inhibited by acid formation quite quickly, as media acidifies below pH 5.0 and growth ceases with acid concentrations of roughly 5-20 mM acetate and 5-20 mM lactate, and total acid formation of 10-30 mM when extra buffering capacity (such as MOPS) is not added to the standard media. Under pH-controlled growth conditions where media is not allowed to acidify, organic acid concentrations can reach 165 mM, consisting of roughly 140 mM acetate and 25 mM lactate. In the presence of 200 mM sodium acetate, *C. bescii* did not grow, and a lag phase (30 h) was observed with an initial concentration of 150 mM sodium acetate, similar to the maximum concentration observed to be physiologically produced (257). The authors proposed that high organic acid concentrations in the medium act to uncouple membrane function by interfering with the ion gradient, a known phenomenon in acetogenic *Clostridium* spp. (395). Thus, one plausible approach to increasing product formation is the generation of uncharged products of neutral pH – e.g. ethanol. Indeed, the ethanol-producing AdhE-Rnf strain of *C. bescii* consumed more substrate and produced more carbon end-products than the wild type or parent strains during growth in sealed bottles. Under pH-controlled fermentation conditions, the total soluble carbon end-products

were roughly equivalent, despite the production of markedly less organic acids (137 mM total products, 62 mM organic acids) (257, 270).

In addition to producing more neutral products, elimination of acidic or undesirable reduced end-products also holds promise. Indeed, while some level of acetate production may be necessary to maintain ATP levels for biomass formation, lactate production has already been eliminated (291). The pyruvate secretion that results from lactate dehydrogenase deletion is more prominent at lower temperature, and redirection of pyruvate shows promise both directly as a further carbon source for ethanol and to potentially increase total products by removing another acidic end-product (270). The factors defining the upper bounds of growth for *C. bescii* remain unidentified, although for growth on biomass (pretreated and unpretreated) inhibition by a lignocellulose derived inhibitor (e.g. lignin-derived phenolics) has been suggested (257, 325, 396). Quorum sensing autoinducer mechanisms are also plausible (397).

Concerning the second limiting factor, less study has been dedicated to the uptake of sugars by *C. bescii*. While all species of *Caldicellulosiruptor* can utilize a large range of both C_5 and C_6 sugar substrates, this varies species to species (**Table 4.1**). No carbon catabolite repression (CCR) has been observed in any of the 14 species currently known (246). Lack of CCR is a distinctive trait among bacteria, highly desirable for CBP organisms – especially combined with the ability to utilize both the primary and minor decorative sugars found in the most studied sources of lignocellulose feedstocks (e.g., poplar, switchgrass). In *C. bescii,* carbohydrate uptake is accomplished by either a variety of ABC transporters or by a single phosphotransferase system (PTS) in the case of fructose. The ABC transport systems can be divided into CUT1 and CUT2 families, which are hypothesized to uptake di/oligo- and monosaccharides, respectively. While both CUT2 family transporters contain a chromosomally linked ATPase, none of the 14 predicted

CUT1 family transporters are associated with a dedicated ATPase. Bioinformatic identification of these transporters and the related regulatory motifs and transcription factors is the focus of an upcoming publication (398). A complete reconstruction of the primary carbohydrate transport and utilization pathways and the associated regulons is presented in **Figure 4.2**.

Eliminating side product formation and improving sugar consumption are both viable paths to increase product formation in future engineering efforts, specifically to increase ethanol production by *C. bescii*. To increase sugar uptake and consumption, it necessary to understand how carbohydrate uptake is achieved and regulated. To this end, we bioinformatically identified a promiscuous ATPase, *msmK* (athe_1803), hypothesized to energize all 12 of the CUT1 family transport systems in *C. bescii,* each of which lacks a dedicated ATPase (398). Herein, I report that the deletion of *msmK* resulted in impaired growth of *C. bescii* on the examined di- and oligosaccharides relative to the parent strain, while growth on the tested monosaccharides was mostly unimpaired. To decrease undesirable side product formation, this work aimed to lower pyruvate secretion in the AdhE-Rnf strain. Secretion of pyruvate was previously hypothesized to result from a bottleneck in metabolism at POR at 60 °C, due to the presumptive lower activity of the enzyme at this sub-optimal growth temperature. Herein, expression of a POR from *Clostridium thermocellum* (T_{opt} 60 °C) in the AdhE-Rnf strain background demonstrated that decreased POR activity is not the cause of pyruvate secretion at sub-optimal growth temperatures. Unfortunately, genome rearrangement of the AdhE-Rnf expression locus during strain construction precluded direct examination of the impact of expressing a low temperature (LT) POR on ethanol production.

Materials and Methods

Culture conditions and growth of C. bescii

DSMZ 6725 (wild type *C. bescii*) was obtained from the German Collection of Microorganisms and Cell Cultures. Unless otherwise specified, *C. bescii* was cultured under anaerobic conditions (80% (v/v) nitrogen and 20% (v/v) carbon dioxide headspace) at 75 °C in modified DSM 516 medium with plate shaking at 150 rpm, as described previously (391). To overcome uracil auxotorophy in modified strains, all media contained 20 µM uracil. Media contained soluble substrates at a concentration of 25 mM or insoluble substrate loadings of 5 g/L. Xylose, crystalline cellulose (Avicel PH-101), pectin from apple, and beechwood xylan were obtained from Sigma (St. Louis, MO, USA). Glucose and fructose were obtained from Amresco (Solon, Ohio, USA). Prior to media preparation, pectin and xylan were washed to remove contaminating soluble sugars that could interfere with growth experiments. Pectin and xylan were suspended in cold 70% ethanol, soaked for 30 minutes, and pelleted by centrifugation at 10,000 x *g* for 10 min. The wash procedure was repeated three times before samples were dried overnight under airflow in a fume hood. Solid growth medium and cultures were prepared using modified DSM 516 medium with 3% agar, as described (242). Solid medium cultures were grown at 70 °C for 48 h under argon atmosphere.

Cultures for end-product analysis and growth studies were conducted in 100 mL serum vials with 50 mL of culture sealed with butyl rubber stoppers for up to 40 hours. For ethanol production experiments, strains were cultured at both 60 °C and 75°C. *C. bescii* competent cells were grown at 70 °C in 1 L media bottles containing 500 mL of Low Osmolarity Defined medium with amino acids mix (LOD-AA) under and sealed with butyl rubber stoppers to a final OD_{680} of 0.06-0.07, as described previously. Cultures for preparing cell extracts were grown until mid-late

log phase at 60 °C in 1 L media bottles containing 500 mL of modified DSM516 medium with glucose.

Cell counts and protein estimation

Cells were counted on a Petroff-Hausser Counting Chamber (Hausser Scientific, Horsham, PA, USA). Cultures were counted un-diluted or diluted either fivefold or tenfold in 1x *C. bescii* base salts, depending on culture density. For protein estimations, the Bradford method was utilized (Bio-Rad #5000006, Hercules, CA, USA) in 96 well plate format with bovine serum albumin (BSA) standards (Bio-Rad #5000007, Hercules, CA, USA).

Plasmid and strain construction

Strains used and generated in this study are listed in **Table 4.2**. Generation of *C. bescii* strains was accomplished with allelic exchange, as described previously (256). Plasmids for strain construction were generated from PCR products with NEBuilder HiFi DNA Assembly (New England Biolabs, Ipswich, MA, USA). PCR products for construct assembly were generated with PrimeSTAR HS DNA polymerase (TaKaRa Bio, Kusatsu, Shiga, Japan). PCR products for screening were generated with SpeedSTAR HS DNA polymerase (TaKaRa Bio, Kusatsu, Shiga, Japan). Plasmids were transformed into NEB 10-beta Competent *E. coli* by heat-shock at 42 °C, per manufacturer's directions (NEB #C3019H). PCR products, plasmid DNA, and genomic DNA were purified with kits from Stratagene and Zymo Research. Prior to transformation into *C. bescii*, plasmids were methylated with M.CbeI, as described; methylation was confirmed by lack of digestion in the presence of *HaeIII* restriction endonuclease (New England Biolabs, Ipswich, MA, USA). Plasmid sequences were verified by Sanger sequencing (GeneWiz, South Plainfield, NJ, USA).

C. bescii competent cells were prepared as described previously and transformed by electroporation with a Bio-Rad Gene Pulser (242). Electroporation was performed using 500-1000 ng of methylated plasmid DNA with voltages from 1.8-2.2 kV, resistances of 200-400 Ω, and capacitance of 25 μF. After electroporation, cells were promptly transferred to recovery media, preheated to 70 °C. At 0 min, 45 min, 90 min, and 180 min post electroporation, 5 mL samples were transferred to 50 mL of selective medium containing 50 μg / mL kanamycin and incubated at 75 °C with shaking at 150 rpm until growth was observed (24-72 h). Cultures with growth were screened for plasmid integration by PCR, and colony purified by 2-4 rounds of plating and selection to eliminate merodiploidy before counter-selection for second crossover (plasmid loss). Counter selection was accomplished on solid medium using modified DSM 516 medium with 4 mM 5-floroorotic acid (5-FOA) and 20 μM uracil. After strain construction, modified genomic loci were sequence-verified by Sanger sequencing (GeneWiz, South Plainfield, NJ, USA).

End-product analysis

Ethanol and acetate production were measured by gas chromatography with a flame ionization detector (GC-FID) on an Agilent 7890 equipped with a Restek Stabilwax 30 m column and argon carrier gas.

Preparation of cell extracts

For each assayed strain, cells were harvested from 500 mL of mid-late log phase culture by centrifugation at 6000 x *g* for 10 min (Beckman Avanti J-30I JLA 10.500 rotor), immediately flash frozen in liquid nitrogen, and stored at -80 °C. All subsequent steps of extract preparation were conducted anaerobically in a Coy chamber or sealed centrifugation tubes. Cell pellets were thawed and resuspended in 25 mM MOPS buffer pH 7.0 containing 1 mg / mL lysozyme at a ratio of 3 mL / g cells. Suspended cells were incubated at room temperature for 30 minutes, followed

by 3 x 10 s rounds of sonication (Qsonica Q55, amplitude 40). Lysates were clarified by ultracentrifugation at 100,000 x *g* for 1 h (Beckman L90K ultracentrifuge 70.1Ti rotor) and then clarified lysates (S100) were washed with excess buffer over a regenerated cellulose centrifugation filter (10 kDa Amicon EMD Millipore) to remove small molecules and redox cofactors. Extracts were sealed anaerobically and either stored at -80 °C or immediately assayed for activities.

Enzymatic activity assays

Enzyme activities in *C. bescii* and *C. thermocellum* S100 cytosolic extracts were measured spectrophotometrically using an Agilent Technologies Cary 100 UV-Vis Spectrophotometer equipped with an Agilent Cary Series Temperature Controller. All assays were conducted in 3 mL silicone-stoppered glass cuvettes under anoxic conditions (argon atmosphere) with 2 mL reaction volume. POR activity was measured by the pyruvate-dependent reduction of methyl viologen (MV) at temperatures ranging from 25-85 °C. Reduction of methyl viologen was followed at 600 nm and activities were calculated using an extinction coefficient of 13.7 mM^{-1} cm^{-1}. Standard POR reactions were conducted in 50 mM EPPS buffer, pH 8.4. with 2 mM $MgCl_2$, 0.4 mM thiamine pyrophosphate (TPP), 0.1 mM CoA-SH, 1 mM MV and S100 cytosolic extract. Anaerobic conditions were maintained by the addition of trace sodium dithionite (DT, > 0.2 mM). After the addition of cytosolic extract, absorbance was allowed to stabilize to account for background activity before the reaction was begun by the addition of 5 mM pyruvate. Both the aldehyde dehydrogenase (ALDH) and alcohol dehydrogenase (ADH) activities of AdhE were measured at 60 °C. Reaction mixtures contained 50 mM MOPS, pH 7.0, 5 µM ferrous sulfate, 0.5 mM NADH or NADPH, and reactions were initiated by the addition of either 0.35 mM acetyl-CoA (ALDH activity) or 10 mM acetaldehyde (ADH activity). Acetaldehyde or acetyl-CoA dependent oxidation of NAD(P)H was measured at 340 nm ($\varepsilon = 6.22$ mM^{-1} cm^{-1}). For all assays,

activities were calculated as the difference in rate before and after the addition of substrate. Activity is expressed as Units (U) / mg of protein where 1 unit equals the reduction of 1 µmol of MV / min or the oxidation of 1 µmol NAD(P)H / min.

Quantitative reverse transcriptase polymerase chain reaction (RT-qPCR)

Cells were harvested from 25 mL of mid-late log phase culture grown at 60 °C by centrifuging at 6000 x *g* for 10 minutes (Beckman Avanti J-30I JLA 10.500 rotor). Cells were flash frozen in liquid nitrogen and stored at -80 °C until RNA extraction (1 week). RNA was harvested by phenol:chloroform extraction and isopropanol precipitation. Genomic DNA was removed by digestion with Ambion Turbo DNase (Invitrogen #AM2238) followed by an additional round of phenol:chloroform extraction. RNA was quantified with a Thermoscientific NANODROP 2000c spectrophotometer and cDNA was synthesized with AffinityScript QPCR cDNA Synthesis Kit (Agilent #60059). RT-qPCR was performed on an Agilent MX3000p qPCR instrument with Brilliant III Ultra-Fast SYBR Green QRT-PCR Master Mix (Agilent #60082). Samples were prepared in biological duplicate and reactions were run in technical duplicate. Data was processed and analyzed using MxPro software (Agilent).

Results

Deletion of msmK impairs growth on di- and oligosaccharides

To verify the putative function of athe_1803 (*msmK*) and further explore its role in sugar uptake, a deletion strain was constructed. Clean deletion of the target gene was verified by PCR followed by Sanger sequencing (**Figure 4.3**). The markerless deletion of *msmK* in a Δ*pyrE* (MACB1018) background resulted in strain MACB1080 (Δ*pyrE*, Δ*msmK*).

Under standard growth conditions on the monosaccharide substrates xylose and glucose, the *msmK* KO strain grew to the same final density as the parent strain (**Figure 4.4**). On both

substrates, while the same final cell density was reached, a minor growth phenotype was observed whereby the growth rate was lower during log phase than the parent strain. In the parent strain the maximum growth rate on xylose was 0.74 h^{-1}, while the *msmK* KO strain was 0.53 h^{-1}. On glucose, the maximum growth rate of the parent strain was 0.84 h^{-1}, while the *msmK* KO strain was 0.64 h^{-1}. During growth on both glucose and xylose, maximum cell density was reached at 16 h, while the deletion strain did not reach maximum density until 20 h.

Growth of the Δ*msmK* strain was severely impaired relative to the parent strain on all tested di- and oligosaccharide substrates (**Figure 4.4**). Tested substrates with diminished growth capacity were cellobiose, crystalline cellulose, xylan, and pectin. The growth phenotype was least severe on cellobiose, where the final cell density was roughly 8-fold lower than the parent strain. During growth on cellulose and pectin, the final cell densities were over an order of magnitude lower than the parent strain.

Construction and analysis of the LT-POR strain

The low-temperature POR or LT-POR strain (MACB1078) was constructed by the heterologous expression of a single gene, cthe3120, from *C. thermocellum* in an MACB1062 (Δ*pyrE,* Δ*ldh,* $P_{S\text{-}layer}$ *adhE* $P_{BF\text{-}H2ase}$ *rnfCDGEAB*) background. Cthe3120, coding for the single subunit enzyme PFOR3, was expressed at the intergenic region between athe_0962-0963 under the control of the highly expressed constitutive promoter for athe_1708 – a putative 2-ketoacid oxidoreductase of unknown specificity.

Growth of the LT-POR strain at 60 °C on glucose revealed no ethanol production compared to roughly 30 mM in the parent strain. To investigate this decreased ethanol production relative to the parent strain, diagnostic PCR was performed on the athe_0949 locus – the site of heterologous *rnf* and *adhE* in the parent (**Figure 4.5**). Disruption of the locus clearly took place, as even with

significant elongation time, no PCR product was observed with primers that should bridge the insertion site (**Figure 4.5C**). Internal diagnostic reactions indicate that all subunits of *rnf* are present in the genome, while potentially only part of *adhE* is present (**Figure 4.5D**). Interestingly, the appropriate product was observed bridging the 3' portion of *adhE* with the downstream athe_0949 flanking region used for recombination during MACB1062 strain construction (data not shown). Due to the discrepancies in diagnostic reactions, the exact genotype of the LT-POR strain cannot be verified, although the presence of *pfor3* (cthe3120) and the absence of *ldh* have both been confirmed.

AdhE activity in the LT-POR strain

As a result of lower than expected ethanol titers in the LT-POR strain, activity for both the ADH and ALDH domains of AdhE were assayed. For both reactions, activity was not detected above the levels in wild type *C. bescii* extracts (~0.1 U/mg). In each case, the viability of assay conditions were verified by the spike-in of *C. thermocellum* S100 after the initial reaction with LT-POR strain extracts failed to yield a reaction; both ALDH and ADH activities were verified in *C. thermocellum* cytosolic extract.

Heterologous gene expression in the LT-POR strain

Expression of *pfor3* (cthe3120), *adhE*, *rnf*, *por* (athe_0874), and *gapdh* were measured by qPCR in the LT-POR strain. Expression of other genes was correlated to *gapdh* expression (**Figure 4.6**). Primers utilized for *adhE, gapdh, por,* and *rnf* have been previously verified experimentally (270). Lack of expression of *adhE* and *rnf* in the LT-POR strain (no cycle threshold observed) correlated with the lack of ALDH and ADH activity in the assays discussed above. Expression of the native *por* was approximately 1.6 times the expression of *gapdh*. The heterologously expressed *pfor3* was present at 0.5 times the expression of *gapdh.*

POR activity in the LT-POR strain

POR activity assays were conducted in both the LT-POR and wild type strains every 10 °C from 25 °C to 85 °C (**Figure 4.7**). For both strains, the maximum activity was observed at 65 °C. At all temperatures, the LT-POR strain had a higher specific activity than that of the wild type. By comparing the activity differences between these two strains, it was possible to identify a maximum difference at 55 °C, which should indicate the optimal temperature for activity of the heterologously expressed POR. Subtracting the activity over the range of temperatures for the LT-POR strain from the wild type, a temperature dependent activity profile was established for the *C. thermocellum* PFOR3 (**Figure 4.7**).

Discussion

This work describes two different experimental focuses, each addressing one of two methods to help improve the viability of *C. bescii* as a microbial bioproduct platform. In the first approach, the deletion of a promiscuous ATPase helped to elucidate the role of this protein in regulating uptake for multiple different carbohydrate substrates present in lignocellulosic biomass. In the second approach, the heterologous expression of a second POR aimed to relieve a metabolic bottleneck to eliminate side-product formation and thus improve flux to the desired end-product (ethanol). While unanticipated genomic rearrangement prevented an increase in ethanol production, examination of the LT-POR strain revealed that the metabolic bottleneck did not result from the native *C. bescii* POR functioning poorly at low temperature, as originally hypothesized.

C. bescii can utilize a broad range of carbohydrates, which contribute to this organism's ability to degrade and ferment unprocessed plant biomasses such as hardwoods (poplar), low lignin grasses (Napier and Bermuda), and high lignin grasses (switchgrass) (255). Growth has been observed on arabinose, carboxymethyl cellulose, cellobiose, crystalline cellulose, β-cyclodextrin,

filter paper (Whatman), fructose, galactose, glucose, glycogen, lactose, maltose, mannose, melibiose, pectin, pullulan, rhamnose, fucose, starch, sucrose, trehalose, xylan (beechwood, birchwood, and oat spelts) and xylose. Weak growth was also observed on mannitol, polygalacturonate, raffinose, sorbitol and yeast extract. Currently, no evidence is available for CCR, meaning that *C. bescii* does not preferentially uptake or metabolize different sugars when provided with a complex mixture. Indeed, to the contrary, it has been demonstrated that xylose (C_5) and glucose (C_6) are metabolized simultaneously in a 50:50 (w/w) mixture during fermentation (267).

The specific mechanisms of sugar uptake in the genus *Caldicellulosiruptor* remain largely unstudied. In a forthcoming publication, we analyze the transcriptional regulation of carbohydrate degradation, utilization, and transport in this genus (398). During this regulatory analysis, an effort was made to provide accurate functional assignment of carbohydrate utilization genes such as these transporters. Accurate functional assignments can be difficult due to functionally divergent paralogs that only vary slightly in specificity. To overcome some of these difficulties, our collaborators employed a subsystems-based comparative genomics approach to improve the accuracy of genomic annotations, predict functions of uncharacterized carbohydrate utilization genes, and reconstruct regulatory and metabolic networks in this genome. In the course of this analysis, it was noted that while the CUT2 family ABC transporters (putatively responsible for C_5 and C_6 monosaccharide transport) are each chromosomally associated with a dedicated ATPase subunit, the CUT1 family transporters each lack an associated ATPase. A similar scenario has been previously observed in *Streptococcus pneumoniae*, where *msmK* was characterized as a shared ATPase for six different CUT1 family ATP transporters that each lacked their own ATPase (399). Likewise, in *Bacillus subtilis*, a homolog (*msmX*) was demonstrated to be a multipurpose

ATPase involved in the uptake of several pectin related oligosaccharides (400). Mutagenesis studies of the transmembrane domain of a *msmX* associated transporters showed two specific residues that are necessary for interaction with *msmX*. Markedly, both residues identified in *B. subtilis* AraPQ, E208 and D213, are conserved in each of the 12 putative *msmK*-associated transporters in *C. bescii* (**Figure 4.8**) (398).

Deletion of *msmK* resulted in the predicted phenotypes on all tested carbohydrate substrates, indicating that this ATPase is indeed functionally promiscuous (**Figure 4.4**). Growth on cellobiose was not diminished by the same level as growth on the tested oligosaccharides. We hypothesize that the difference observed in growth of the *msmK* KO strain on cellobiose versus cellulose is because *msmK*-independent CUT2 family transporters are contributing to cellobiose uptake. Indeed, RNAseq data included in the aforementioned forthcoming publication indicates that the CUT2 family GxgABC system is weakly induced during growth on cellobiose, which could contribute to the observed growth. Understanding the role of different sugar transporters and their regulation can enable improved understanding of the complex biomass degradation process.

Pyruvate is an undesirable product of the AdhE-Rnf strain of *C. bescii*, representing 5-10% of the total carbon in end-products. Previously, we hypothesized that the production of pyruvate resulted from the inability of POR to keep up with metabolism at sub-optimal growth temperatures in *C. bescii*, such as the temperature required for optimal ethanol production in this recombinant strain (270). This hypothesis was driven by several factors, including the extreme thermostability of enzymes from this organism. In the hyperthermophilic archaeon *Pyrococcus furiosus* (T_{opt} 100 °C), the homologous POR enzyme is most active close to the optimum growth temperature of the organism, with activity increasing up to 90 °C, beyond which assays were technically limited (401). In *C. thermocellum*, previous engineering efforts showed that expression of an additional

enzyme to metabolize pyruvate could result in increased ethanol production. The mesophilic enzyme pyruvate decarboxylase (PDC) directly decarboxylates pyruvate to acetaldehyde, without an acetyl-CoA intermediate. Heterologous expression of a variety of PDC homologs from different mesophilic organisms in *C. thermocellum* increased ethanol production up to 20% during growth at 55 °C (200). Of the eight tested PDC homologs, the candidate from *Zymobacter palmae* was the most effective, which correlates well with the enzymes optimum activity at 55 °C, and the reasonable thermostability of the enzyme ($T_{1/2}$ 10 h at 50 °C). Interestingly, there are conflicting reports over the ability of the PDC from *Zymomonas mobilis* to increase ethanol production in *C. thermocellum*, despite this enzyme's optimum activity at 60 °C and $T_{1/2}$ of 24 h at 50 °C. Kannuchamy et al. reported that expression of *pdc* from *Z. mobilis* in *C. thermocellum* resulted in a two-fold increase in both titer and yield of ethanol, while in the screen of eight PDC homologs described above, Tian et al. found that expression of this PDC did not improve ethanol titers (200, 402). No explanation for this discrepancy is apparent.

Several factors contributed to the decision to utilize PFOR3 from *C. thermocellum* as the heterologous POR with the aim of improving pyruvate flux in the AdhE-Rnf strain of *C. bescii* during growth at 60 °C. Enzymes from *C. thermocellum*, such as *adhE*, have already been successfully utilized in *C. bescii* – specifically in this target strain; in *C. thermocellum*, both enzymes, POR and AdhE, should form the native ethanol production pathway. The *C. thermocellum* genome encodes five annotated 2-ketoacid oxidoreductases (all designated as pyruvate ferredoxin oxidoreductases) (403). It is likely that several of these enzymes represent other 2-ketoacid oxidoreductases (KORs) besides POR, such as 2-ketoglutarate oxidoreductase (KGOR) and 2-ketoisovalerate oxidoreductase (VOR). In *C. bescii* for instance, the genome encodes three KORs, at least one of which I recently showed has presumptive VOR activity

(unpublished data). It has been demonstrated that in the proteome of *C. thermocellum*, PFOR3 was more highly represented than POR1 (404). Further, PFOR3 has higher homology to tsac_0046, the demonstrated primary POR in *T. saccharolyticum* (405). Shortly after the LT-POR strain was constructed, a publication examining the relative contributions of all five KORs in *C. thermocellum* to POR activity showed that PFOR1 is significantly more active than PFOR3 (403). However, the activity of the latter clearly makes a major contribution to the total POR activity of the LT-POR strain (**Figure 4.7**).

While comparisons of temperature curves between the LT-POR strain and the wild type revealed that the optimum temperature for PFOR3 is roughly 55 °C, the more surprising result was that the T_{opt} for *C. bescii* POR is 65 °C. This optimum temperature, well below the optimum temperature for growth, implies that even in non-ethanol producing strains the build-up of pyruvate is the result of a bottleneck elsewhere in metabolism. Other sources for this backup may be related to redox cofactor build up (e.g. reduced ferredoxin) or downstream carbon flux issues (e.g. acetate production or acetyl-CoA dependent pathways). Although *in vitro* activity assays do not strictly correlate to *in vivo* activities due to a variety of potential conditional differences and complicating factors, such as allosteric regulation, the lack of any POR activity observed at 85 °C and above could be a contributing factor in determining the upper limits of *C. bescii* growth. It has been reported that *C. bescii* grows, albeit weakly, up to 90 °C. Realistically, above 84 °C, growth as measured by both final cell density and protein falls off precipitously (268). POR is believed to be an essential enzyme in the primary metabolism of *C. bescii*, as it is the key source of acetyl-CoA. Hence, it is surprising to find that the optimum temperature differs by over 10 °C from the optimum temperature of the bacterium itself. Still, there are numerous mechanisms beyond specific activity

that could be used to tune POR activity *in vivo*, including expression, post translational modification, and allostery.

The original goal of constructing the LT-POR strain was to increase ethanol production in the AdhE-Rnf strain. This goal was not achieved due to presumed genome rearrangement during strain construction that resulted in inactivation of the ethanol producing enzyme AdhE, which was previously heterologously expressed in its active form in the parent strain. PCR with primers designed to bridge the athe_0949 locus, the site of AdhE and Rnf expression, produced no product in the LT-POR strain, despite annealing time allowing for products up to 15 kB (**Figure 4.5**). This result implies that the orientation of these genome regions is not the same as it was in the parent, and thus genome rearrangement likely occurred. Genome rearrangements, such as by inversions of large portions of the chromosome, have been reported in *C. bescii,* and are believed to result from activity of the mobile element IS*Cbe4* (279). Based on previous reporting, the more rounds of genetic manipulation (transformation by electroporation, selection, and counterselection) a strain of *C. bescii* endures, the more instances of IS*Cbe4* occur in the genome, increasing the likelihood of genome instability (279). In the case of the LT-POR strain, this was the fifth round of genetic manipulation: deletion of *pyrE,* followed by deletion of *ldh,* followed by insertion of *adhE,* then insertion of *rnfCDGEAB,* and finally insertion of *pfor3.* Strains created by similar numbers of rounds of genetic manipulation (3-4) have shown genome instability before, and it may be necessary to design future strategies for strain construction that minimize the number of transformations required to create a target strain (279).

No ethanol production was observed in the LT-POR strain, which lacked ADH/ALDH activities and expression of both *adhE* and *rnf* as measured by qPCR. Wild type *C. bescii* has never been shown to produce more than trace amounts of ethanol (<0.4 mM), although some

Caldicellulosiruptor species can natively produce the alcohol (269). The only alcohol dehydrogenase encoded in the *C. bescii* genome is a truncated monofunctional enzyme, missing the C-terminal domain; no native ADH activity has been measured in cytosolic extracts (unpublished data). PCR revealed that at least part of the *adhE* gene is still present in the genome, but its location is unknown, and the qPCR results imply that at least the section of the gene being amplified is not expressed. While PCR revealed that all subunits of Rnf appear to be present, although likely not at the athe_0949 locus where they were originally expressed), qPCR data shows that, at least for the target region in *rnfD,* no transcript is present. The development of a metabolic model for *C. bescii*, focusing specifically on applications for the AdhE-Rnf strain, could shed further light on what metabolic engineering steps might improve ethanol production. In the future, computational approaches could elucidate unexpected methodologies for improving product formation or increasing overall metabolic flux (carbohydrates → fermentation products) by revealing unanticipated connections in metabolism that are not readily discerned by human observation and intuition.

Tables and Figures

Table 4.1 Experimentally determined sugar utilization in *Caldicellulosiruptor* spp.

Caldicellulosiruptor spp.	Arabinose	Xylose	Fructose	Rhamnose	Fucose	Ribose	Maltose	Cellobiose	Lactose	Pectin	Sucrose	Trehalose	Glucose	Galactose	Mannitol	Sorbitol	Arabinogalactan	Crystalline cellulose
C. bescii	+	+	+	+	+	nd	+	+	+	+	+	+	+	+	+	-	+	+
C. kronotskyensis	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	nd	nd
C. hydrothermalis	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	nd	nd
C. owensensis	+	+	+	+	+	+	+	+	+	+	+	-	+	+	+	-	-	-
C. obsidiansis	+	+	+	nd	nd	-	+	+	+	+	+	nd	nd	nd	-	nd	nd	nd
C. acetigenus	+	+	+	+	+	nd	+	+	+	-	+	+	+	+	nd	-	nd	-
C. kristjanssonii	-	+	+	+	+	nd	+	+	+	+	+	+	+	+	nd	-	nd	+
C. lactoaceticus	-	+	-	+	+	-	+	+	+	+	-	-	-	-	-	-	nd	nd
C. danielli (Wai35.B1)	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
C. saccharolyticus	+	+	+	+	+	-	+	+	+	+	+	+	+	+	-	-	+	+
C. sp. F32	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
C. changbaiensis (CBS-Z)	+	nd	+	+	+	nd	nd	nd	nd	+	+	+	+	+	nd	+	+	+
C. morganii (Rt8.B8)	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd
C. naganoensis	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd

+	growth observed
-	no growth observed
nd	not determined

Table 4.2 Strains used and generated in this study

Strain	Alias	Parent	Transforming Plasmid	Description	Reference
DSMZ 6725	Wild type	---	---	Wild type	(268)
MACB1018	Δ*pyrE*	DSMZ 6725	pGL090	Δ*pyrE*	(242)
MACB1062	AdhE-Rnf	MACB1058	pAR008	Δ*pyrE*, Δ*ldh*, $P_{S\text{-}layer}$ *Cthe-adhE* $P_{BF\text{-}H2ase}$ *rnfCDGEAB*	(270)
MACB1078	LT-POR	MACB1062	pMA001	Δ*pyrE*, Δ*ldh*, P_{Athe_1709} *Cthe-pfor3*	This study
MACB1080	Δ*msmK*	MACB1018	pGR027	Δ*pyrE*, Δ*msmK*	This study

Figure 4.1 Primary metabolism in the recombinant ethanologenic AdhE-Rnf strain

Primary metabolic pathways and ethanol production in the recombinant AdhE-Rnf strain, with implications for the LT-POR strain. Native pathways and enzymes are shown in white boxes. Recombinantly-expressed enzymes are shown in green boxes. The target product, ethanol, is indicated in yellow. The undesirable product pyruvate, indicating a potential pathway bottleneck, is emphasized with a red outline. POR, a target enzyme for decreasing pyruvate and increasing ethanol by heterologous expression of an additional homolog, is emphasized by a green outline. Primary redox cofactors, NAD(H) and ferredoxin ($Fd_{red/ox}$), are emphasized in bold. Abbreviations: Embden-Meyerhof-Parnas glycolytic pathway (Upper EMP), bifurcating hydrogenase (BF-H_2ase), membrane bound hydrogenase (MBH), reduced ferredoxin NAD^+ oxidoreductase from *Thermoanaerobacter* sp. X514 (Rnf), pyruvate oxidoreductase (POR), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), glyceraldehyde-3-phosphate oxidoreductase (GOR), phosphoglycerate kinase (PGK), bifunctional alcohol/aldehyde dehydrogenase from *Clostridium thermocellum* (AdhE).

Figure 4.1

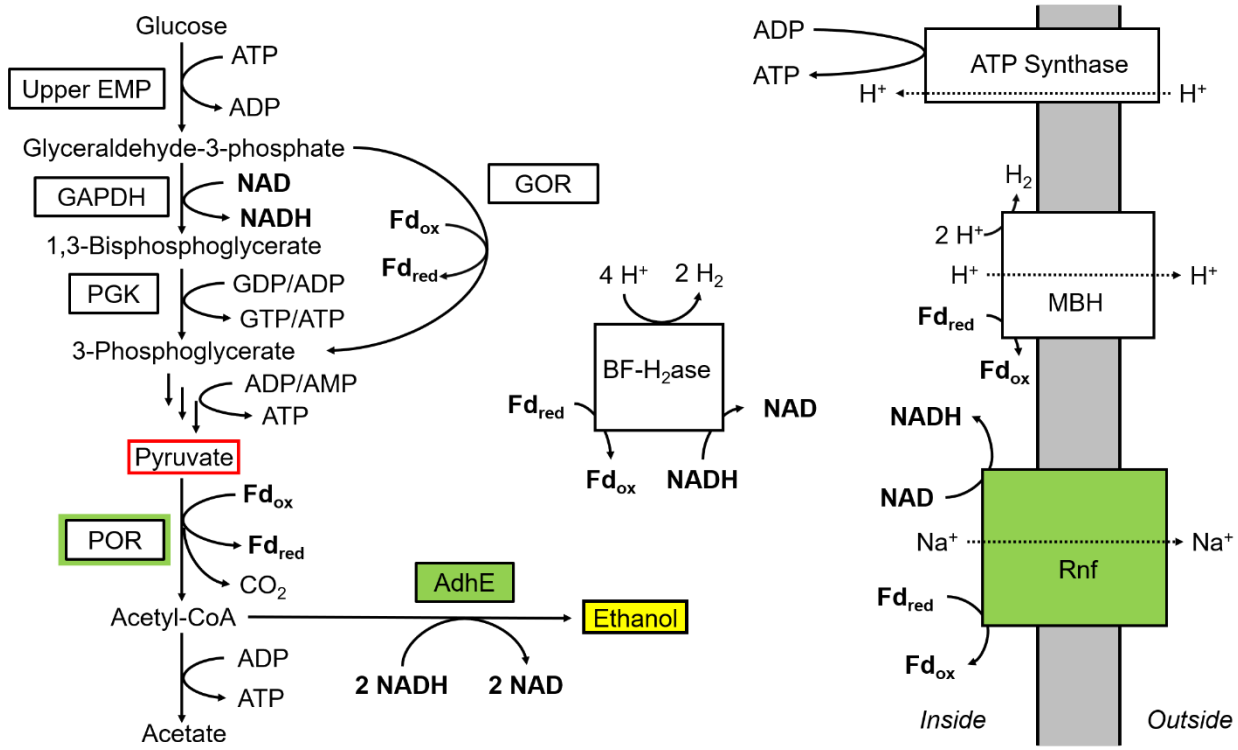

Glucose
ATP
Upper EMP
ADP
Glyceraldehyde-3-phosphate
GAPDH
NAD
NADH
1,3-Bisphosphoglycerate
PGK
GDP/ADP
GTP/ATP
3-Phosphoglycerate
ADP/AMP
ATP
Pyruvate
POR
Fdox
Fdred
CO2
Acetyl-CoA
ADP
ATP
Acetate
GOR
Fdox
Fdred
AdhE
Ethanol
2 NADH
2 NAD
4 H+
2 H2
BF-H2ase
Fdred
NAD
Fdox
NADH
ADP
ATP
ATP Synthase
H+
H+
H2
2 H+
H+
H+
MBH
Fdred
Fdox
NADH
NAD
Na+
Na+
Rnf
Fdred
Fdox
Inside
Outside

Figure 4.2 Reconstruction of metabolic pathways and regulons involved in plant polysaccharide degradation and carbohydrate utilization in *C. bescii*

Reproduced from Rodionov, et al. 2020 (398). Polysaccharide substrates are in rounded red rectangles. Extracellular enzymes, transporters, and cytoplasmic enzymes are shown by rectangles with red, blue, and black text, respectively. CAZymes and other classes of enzymes are shown by large and small arrows, respectively. Novel enzymes predicted by this work are highlighted in yellow. Novel carbohydrate-specific ABC transporters from the CUT1 and CUT2 families are in pink boxes with red and green outlines, respectively. The PTS transporter for fructose is in an orange box. Genes regulated by the same transcription factor are shown by matching colored symbols, described in the lower inset. The central carbohydrate metabolism, fermentation, and hydrogen production pathways are shown with light yellow background.

Figure 4.2

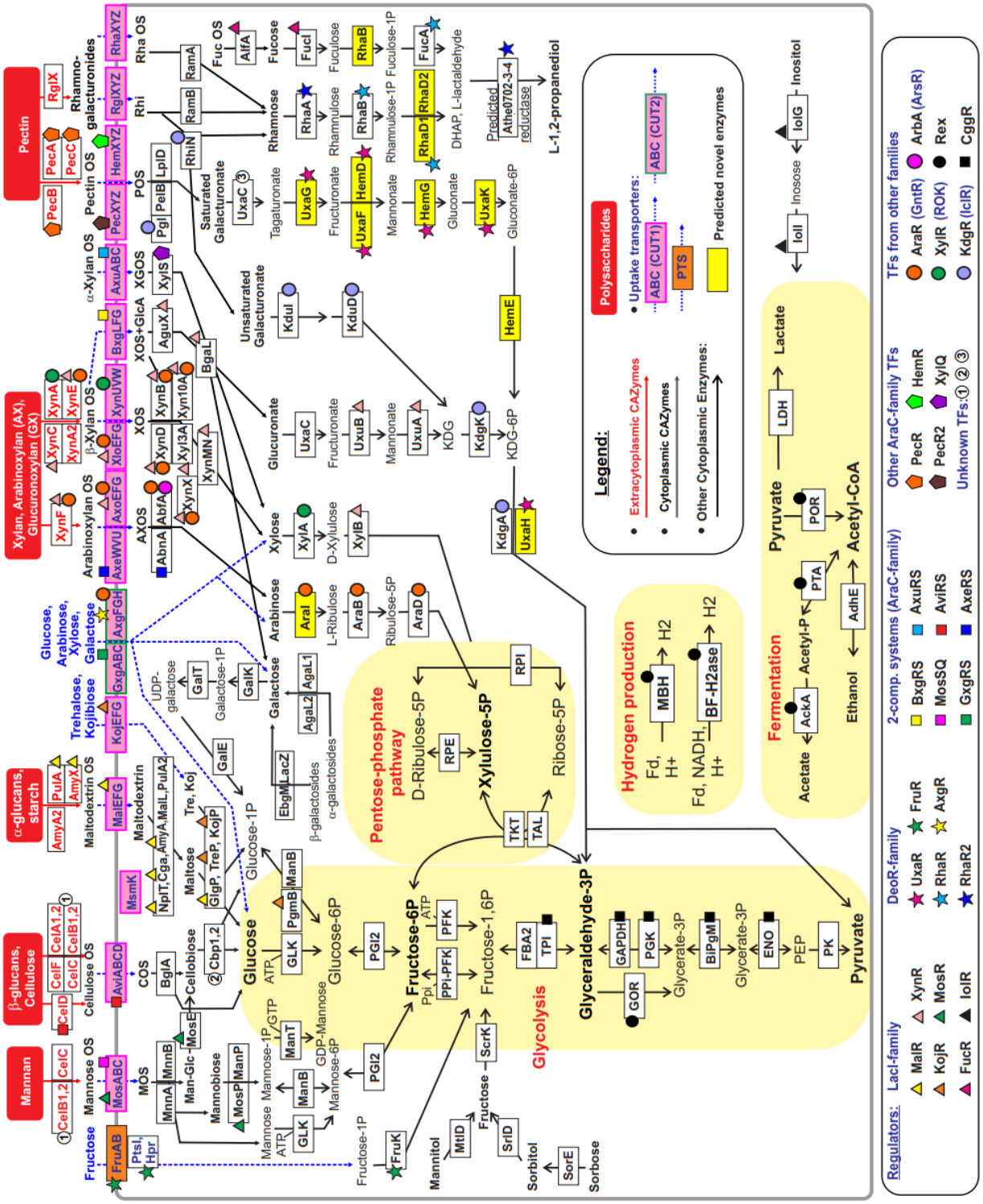

Glycolysis
Pentose-phosphate pathway
Hydrogen production
Fermentation
Glucose
Fructose-6P
Fructose-1,6P
Glyceraldehyde-3P
Pyruvate
Xylulose-5P
Ribose-5P
D-Ribulose-5P
Acetyl-CoA
Ethanol
Acetate
Lactate
Legend:
Polysaccharides
Extracytoplasmic CAZymes
Cytoplasmic CAZymes
Other Cytoplasmic Enzymes:
Uptake transporters:
Predicted novel enzymes
Regulators:
LacI-family
DeoR-family
2-comp. systems (AraC-family)
Other AraC-family TFs
TFs from other families

Figure 4.3 Verification of *msmK* deletion

Deletion of *msmK* was verified by PCR using primers outside the flanking regions used for strain construction (GR0126 and GR0125) and internal to the deletion target (GR0128 and GR0127). PCR products were visualized by gel electrophoresis (A) to compare products between the parent strain (B) and Δ*msmK* strain (C). All products matched expected sizes, as indicated in panels B and C (thin black lines). Flanking regions utilized for homologous recombination during strain construction are shown in purple. Surrounding genes in the athe_1803 locus are shown in blue. The deletion target, *msmK* (Athe_1803), is shown in green. Primers are indicated with black arrows.

Figure 4.3

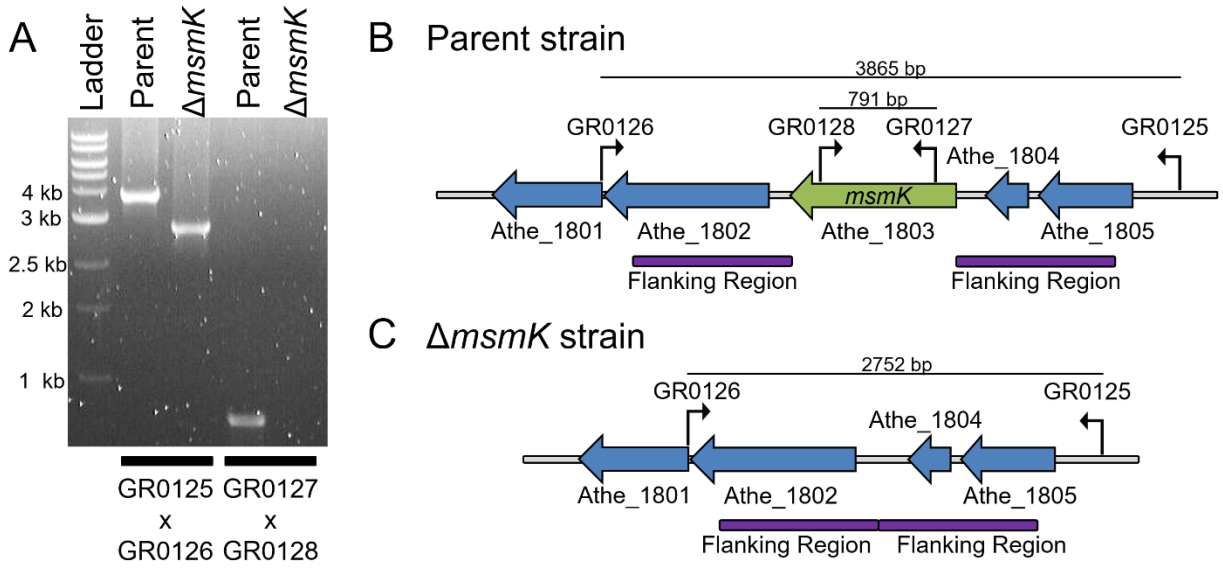

A
Ladder
Parent
ΔmsmK
Parent
ΔmsmK
4 kb
3 kb
2.5 kb
2 kb
1 kb
GR0125 x GR0126
GR0127 x GR0128
B Parent strain
3865 bp
791 bp
GR0126
GR0128
GR0127
GR0125
Athe_1804
msmK
Athe_1801
Athe_1802
Athe_1803
Athe_1805
Flanking Region
Flanking Region
C ΔmsmK strain
2752 bp
GR0126
GR0125
Athe_1804
Athe_1801
Athe_1802
Athe_1805
Flanking Region
Flanking Region

Figure 4.4 Comparison of Δ*msmK* strain and parent strain on varied carbon substrates

The Δ*msmK* strain and the parent strain were grown in biological triplicate (n=3) at 75 °C on a variety of soluble and insoluble carbohydrate substrates. Growth was compared on glucose (A), crystalline cellulose (B), xylan (C), cellobiose (D), pectin (E), and xylose (F). Growth on xylan was reported as end-point (28 h) protein concentration. For all other substrates, growth was reported as cell density (cells / mL) over the course of growth (24-28 h).

Figure 4.4

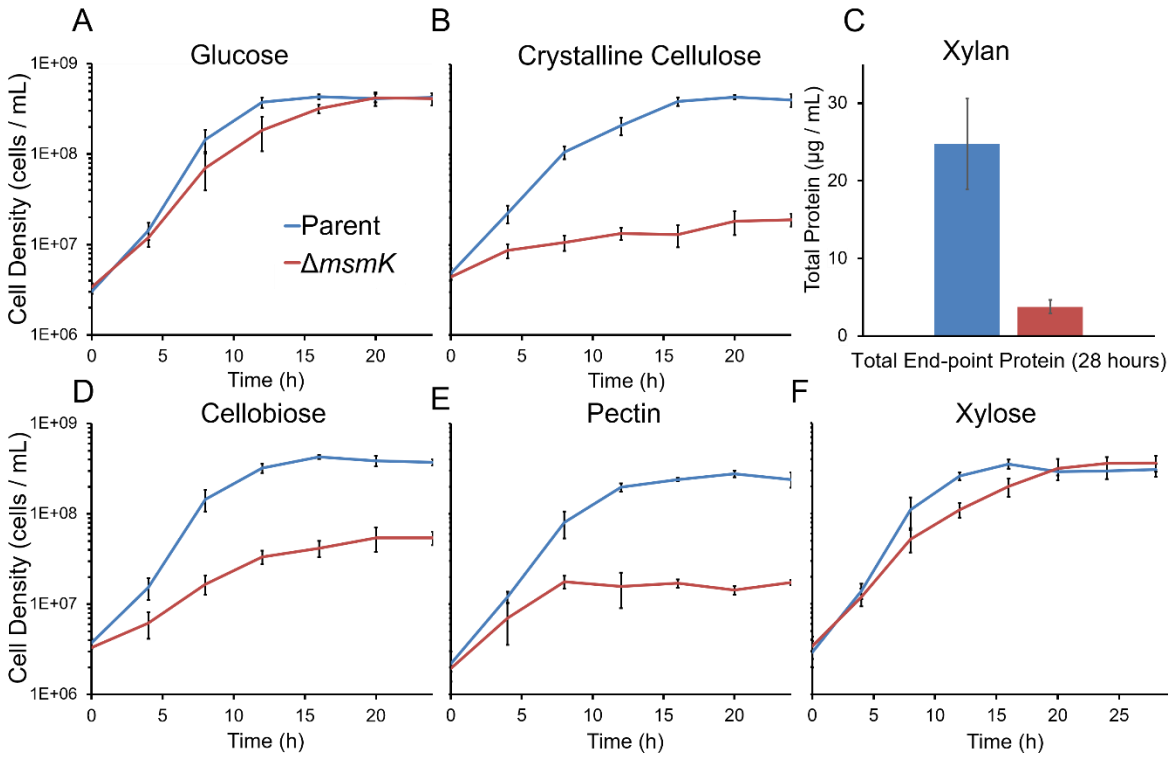

A
Glucose
B
Crystalline Cellulose
C
Xylan
D
Cellobiose
E
Pectin
F
Xylose
Cell Density (cells / mL)
1E+09
1E+08
1E+07
1E+06
Time (h)
0
5
10
15
20
25
Parent
ΔmsmK
Total Protein (µg / mL)
30
20
10
0
Total End-point Protein (28 hours)

Figure 4.5 Diagnostic PCR in the LT-POR strain

To assess the genotype of the LT-POR strain, diagnostic PCRs were conducted comparing the wild type, Δ*ldh*, AdhE-Rnf, and LT-POR strains. Amplification of the *ldh* locus (A) showed the appropriate band, indicating the absence of *ldh* in all strains except the wild type. Amplification of the athe_963 locus (B) shows the appropriate 3500 bp increase in the LT-POR strain, corresponding to integration of *pfor3* at this locus. Amplification of the athe_0949 locus (C) yielded no product for the LT-POR strain, while the desired 10.7 kb product corresponding to *rnfCDGEAB* and *adhE* integration was present in the AdhE-Rnf strain. PCR targeting products internal to *rnfCDGEAB* and *adhE* (D) imply that these genes are still present in the genome, if not at the expected (athe_0949) locus. Native *C. bescii* genes are indicated in blue, with heterologously expressed genes in gray. Primers are indicated by black arrows. All reactions in panel D utilized LT-POR strain gDNA as template.

Figure 4.5

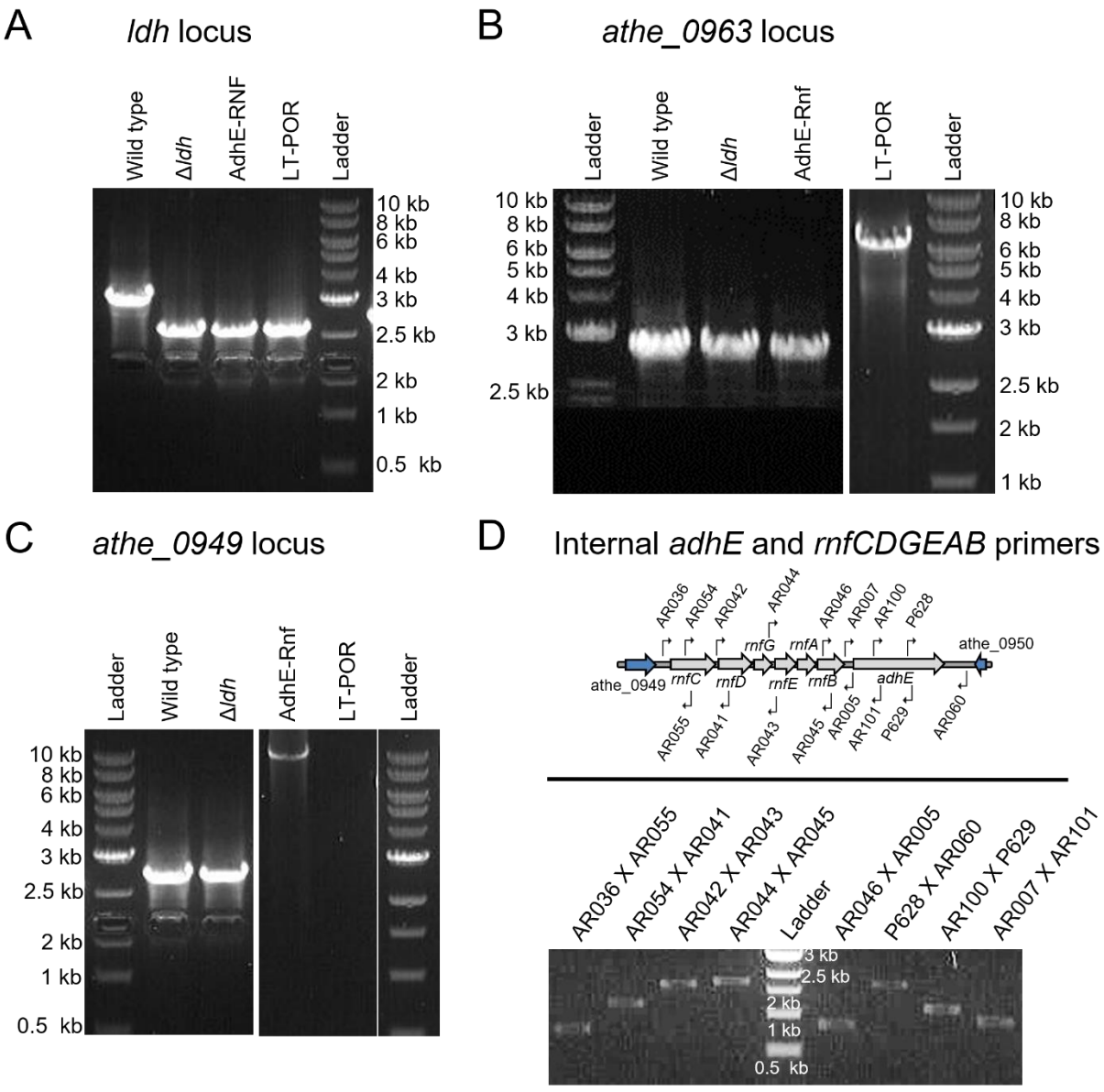

A
ldh locus
Wild type
Δldh
AdhE-RNF
LT-POR
Ladder
10 kb
8 kb
6 kb
4 kb
3 kb
2.5 kb
2 kb
1 kb
0.5 kb
B
athe_0963 locus
Ladder
Wild type
Δldh
AdhE-Rnf
LT-POR
Ladder
10 kb
8 kb
6 kb
5 kb
4 kb
3 kb
2.5 kb
2 kb
1 kb
C
athe_0949 locus
Ladder
Wild type
Δldh
AdhE-Rnf
LT-POR
Ladder
10 kb
8 kb
6 kb
4 kb
3 kb
2.5 kb
2 kb
1 kb
0.5 kb
D
Internal adhE and rnfCDGEAB primers
AR036
AR054
AR042
AR044
AR046
AR007
AR100
P628
athe_0950
rnfG
rnfA
athe_0949
rnfC
rnfD
rnfE
rnfB
adhE
AR055
AR041
AR043
AR045
AR005
AR101
P629
AR060
AR036 X AR055
AR054 X AR041
AR042 X AR043
AR044 X AR045
Ladder
AR046 X AR005
P628 X AR060
AR100 X P629
AR007 X AR101
3 kb
2.5 kb
2 kb
1 kb
0.5 kb

Figure 4.6 Expression of *C. bescii por* and *C. thermocellum pfor3* in the LT-POR strain

Expression of the native *C. bescii por* (athe_0874) and the heterologously expressed *pfor3* from *Clostridium thermocellum* were measured by RT-qPCR. Expression is presented relative to the glycolytic enzyme *gapdh* (athe_1406).

Figure 4.6

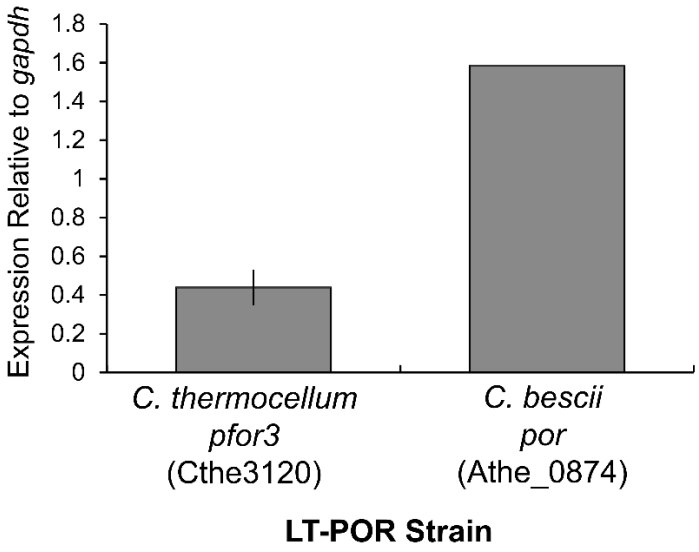

Expression Relative to gapdh
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
C. thermocellum pfor3 (Cthe3120)
C. bescii por (Athe_0874)
LT-POR Strain

Figure 4.7 Temperature dependence of POR activity in cytosolic extract

POR activity was measured at temperatures ranging from 25 °C to 85 °C in cytosolic extracts (S100) of the wild type and LT-POR strains of *C. bescii.* At all temperatures, activity was higher in the LT-POR strain (red triangles) than in the wild type (blue circles). For both strains, enzyme activity was highest at 65 °C. The difference between activities (wild type activity subtracted from LT-POR activity) in the two strains at each temperature (green squares) represents the contribution of the heterologously expressed *C. thermocellum pfor3*. The largest difference in activity between the strains, representing the T_{opt} for *C. thermocellum* PFOR3, was observed at 55 °C. Error bars represent standard deviations (σ) of at least three assay replicates (n=3).

Figure 4.7

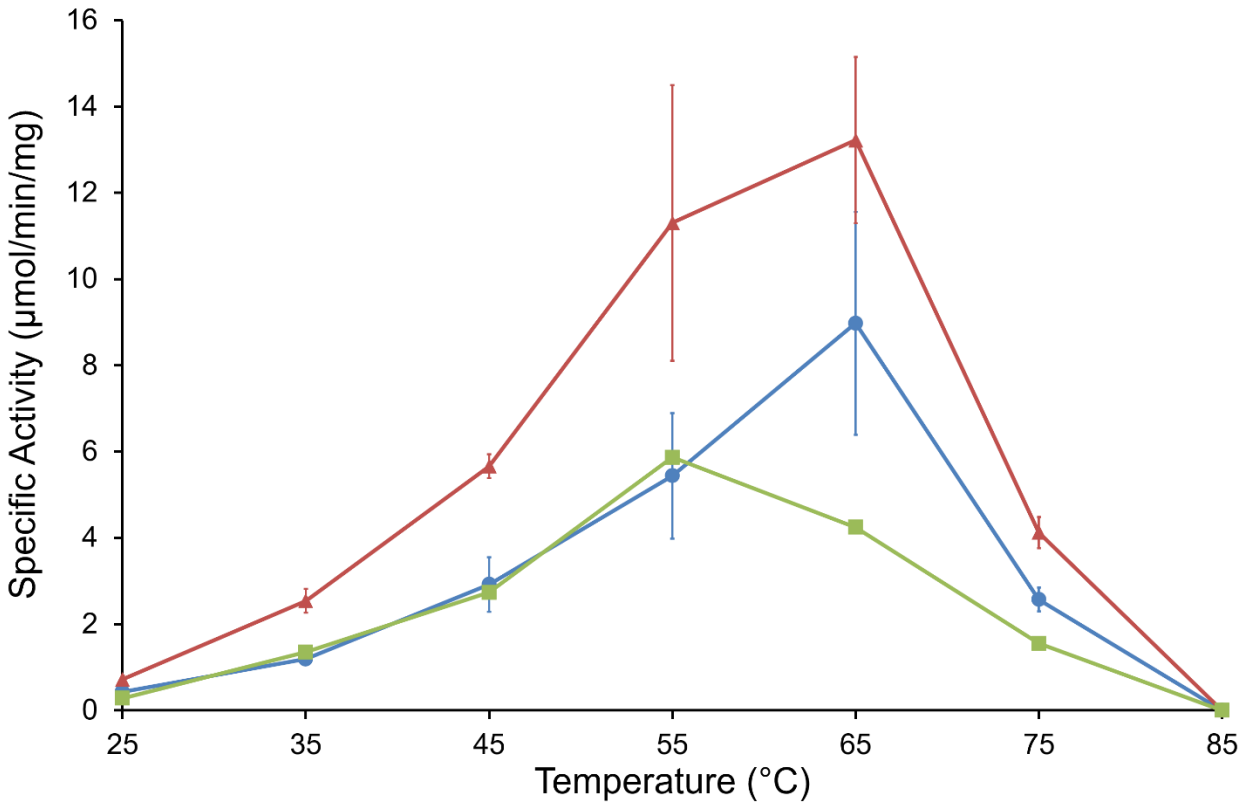
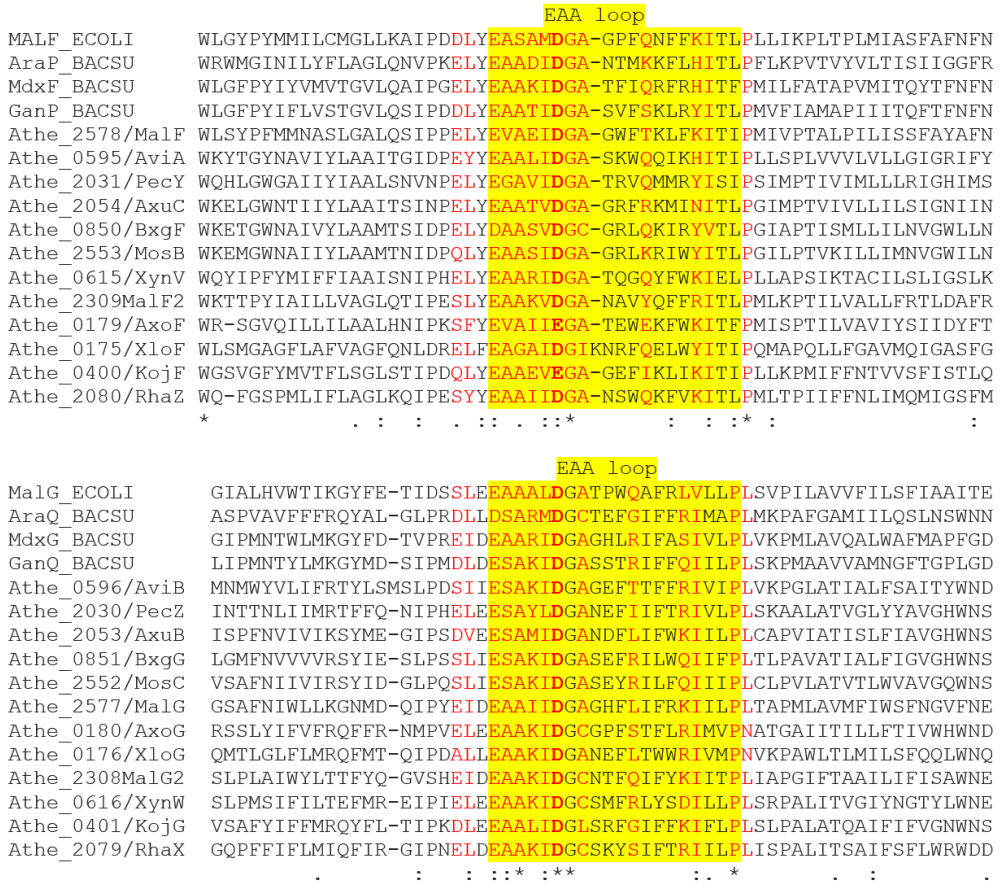

Specific Activity (µmol/min/mg)
16
14
12
10
8
6
4
2
0
25
35
45
55
65
75
85
Temperature (°C)

Figure 4.8 Conserved residues in CUT1 ABC transporter permeases for ATPase contacts

Multipurpose ATPases that energize multiple CUT1 family ABC transporters have been observed in several bacteria (e.g. *msmX* in *Bacillus subtilis*, *msmK* in *Streptococcus pneumonia*). Mutagenesis of AraPQ transporter permeases in *B. subtilis* revealed residues essential for interaction with *msmX*. These residues (red, bold – E208 and D213) are conserved in other CUT1 transporter permeases known to interact with MsmX in *B. subtilis* (MdxFG, GanPQ), and all 12 CUT1 transporter permeases identified in the *C. bescii* genome. Residues know from 3D structure analysis to establish contacts with the ATPase MalK in the *E. coli* CUT1 family MalEFG-MalK maltose transport system are show in red. Residues fall within or adjacent to the EAA loop region, where the permeases dimerize and interact with the ATPase.

Figure 4.8


EAA loop
MALF_ECOLI WLGYPYMMILCMGLLKAIPDDLYEASAMDGA-GPFQNFFKITLPLLIKPLTPLMIASFAFNFN
AraP_BACSU WRWMGINILYFLAGLQNVPKELYEAADIDGA-NTMKKFLHITLPFLKPVTVYVLTISIIGGFR
MdxF_BACSU WLGFPYIYVMVTGVLQAIPGELYEAAKIDGA-TFIQRFRHITFPMILFATAPVMITQYTFNFN
GanP_BACSU WLGFPYIFLVSTGVLQSIPDDLYEAATIDGA-SVFSKLRYITLPMVFIAMAPIIITQFTFNFN
Athe_2578/MalF WLSYPFMMNASLGALQSIPPELYEVAEIDGA-GWFTKLFKITIPMIVPTALPILISSFAYAFN
Athe_0595/AviA WKYTGYNAVIYLAAITGIDPEYYEAALIDGA-SKWQQIKHITIPLLSPLVVVLVLLGIGRIFY
Athe_2031/PecY WQHLGWGAIIYIAALSNVNPELYEGAVIDGA-TRVQMMRYISIPSIMPTIVIMLLLRIGHIMS
Athe_2054/AxuC WKELGWNTIIYLAAITSINPELYEAATVDGA-GRFRKMINITLPGIMPTVIVLLILSIGNIIN
Athe_0850/BxgF WKETGWNAIVYLAAMTSIDPELYDAASVDGC-GRLQKIRYVTLPGIAPTISMLLILNVGWLLN
Athe_2553/MosB WKEMGWNAIIYLAAMTNIDPQLYEAASIDGA-GRLKRIWYITLPGILPTVKILLIMNVGWILN
Athe_0615/XynV WQYIPFYMIFFIAAISNIPHELYEAARIDGA-TQGQYFWKIELPLLAPSIKTACILSLIGSLK
Athe_2309MalF2 WKTTPYIAILLVAGLQTIPESLYEAAKVDGA-NAVYQFFRITLPMLKPTILVALLFRTLDAFR
Athe_0179/AxoF WR-SGVQILLILAALHNIPKSFYEVAIIEGA-TEWEKFWKITFPMISPTILVAVIYSIIDYFT
Athe_0175/XloF WLSMGAGFLAFVAGFQNLDRELFEAGAIDGIKNRFQELWYITIPQMAPQLLFGAVMQIGASFG
Athe_0400/KojF WGSVGFYMVTFLSGLSTIPDQLYEAAEVEGA-GEFIKLIKITIPLLKPMIFFNTVVSFISTLQ
Athe_2080/RhaZ WQ-FGSPMLIFLAGLKQIPESYYEAAIIDGA-NSWQKFVKITLPMLTPIIFFNLIMQMIGSFM
EAA loop
MalG_ECOLI GIALHVWTIKGYFE-TIDSSLEEAAALDGATPWQAFRLVLLPLSVPILAVVFILSFIAAITE
AraQ_BACSU ASPVAVFFFRQYAL-GLPRDLLDSARMDGCTEFGIFFRIMAPLMKPAFGAMIILQSLNSWNN
MdxG_BACSU GIPMNTWLMKGYFD-TVPREIDEAARIDGAGHLRIFASIVLPLVKPMLAVQALWAFMAPFGD
GanQ_BACSU LIPMNTYLMKGYMD-SIPMDLDESAKIDGASSTRIFFQIILPLSKPMAAVVAMNGFTGPLGD
Athe_0596/AviB MNMWYVLIFRTYLSMSLPDSIIESAKIDGAGEFTTFFRIVIPLVKPGLATIALFSAITYWND
Athe_2030/PecZ INTTNLIIMRTFFQ-NIPHELEESAYLDGANEFIIFTRIVLPLSKAALATVGLYYAVGHWNS
Athe_2053/AxuB ISPFNVIVIKSYME-GIPSDVEESAMIDGANDFLIFWKIILPLCAPVIATISLFIAVGHWNS
Athe_0851/BxgG LGMFNVVVVRSYIE-SLPSSLIESAKIDGASEFRILWQIIFPLTLPAVATIALFIGVGHWNS
Athe_2552/MosC VSAFNIIVIRSYID-GLPQSLIESAKIDGASEYRILFQIIIPLCLPVLATVTLWVAVGQWNS
Athe_2577/MalG GSAFNIWLLKGNMD-QIPYEIDEAAIIDGAGHFLIFRKIILPLTAPMLAVMFIWSFNGVFNE
Athe_0180/AxoG RSSLYIFVFRQFFR-NMPVELEEAAKIDGCGPFSTFLRIMVPNATGAIITILLFTIVWHWND
Athe_0176/XloG QMTLGLFLMRQFMT-QIPDALLEAAKIDGANEFLTWWRIVMPNVKPAWLTLMILSFQQLWNQ
Athe_2308MalG2 SLPLAIWYLTTFYQ-GVSHEIDEAAKIDGCNTFQIFYKIITPLIAPGIFTAAILIFISAWNE
Athe_0616/XynW SLPMSIFILTEFMR-EIPIELEEAAKIDGCSMFRLYSDILLPLSRPALITVGIYNGTYLWNE
Athe_0401/KojG VSAFYIFFMRQYFL-TIPKDLEEAALIDGLSRFGIFFKIFLPLSLPALATQAIFIFVGNWNS
Athe_2079/RhaX GQPFFIFLMIQFIR-GIPNELDEAAKIDGCSKYSIFTRIILPLISPALITSAIFSFLWRWDD

CHAPTER 5

DISCUSSION AND CONCLUSIONS

Over the last decade, the genus *Caldicellulosiruptor* has matured as a microbial platform for both native and engineered bioproduct formation (246). Among the 14 isolated and sequenced members of this genus, *C. bescii* is the subject of the second-most publications, has the highest optimum growth temperature (78 °C), and is among the most cellulolytic (247, 268). The work with *C. bescii* accomplished here explores a previously unidentified feature of primary metabolism (Chapter 2), heterologous expression of a pathway to produce non-native reduced carbon products (Chapter 3), heterologous expression of an enzyme to explore bottlenecks in engineered metabolism (Chapter 4), and deletion of a native multipurpose enzyme responsible for the uptake of a wide variety of carbohydrates (Chapter 4). Together, this work improves the prospects for using *C. bescii* as a CBP platform organism, although significant improvements in product titers, strain stability, and overall metabolic flux are still necessary before this bacterium is industrially viable.

Bioproduct Formation by *Caldicellulosiruptor* spp.

The best studied member of this genus is *C. saccharolyticus* (T_{opt} 70 °C), which has been examined extensively as a platform for H_2 production from numerous feedstocks (406). While no genetically tractable lineage of *C. saccharolyticus* has been developed, studies have explored the production of H_2 in both batch and continuous culture on substrates such as sorghum bagasse, carrot pulp, agarose, alginic acid, paper sludge hydrolysate, dilute-acid pretreated barley straw, wheat straw hydrolysate, the organic fraction of municipal solid waste, sugarcane bagasse, maize

leaves, sweet sorghum juice, sugar beet molasses, potato steam peels, and hemicellulose, as well as a variety of monosaccharides (266, 407-418). Compared to *C. bescii*, *C. saccharolyticus* is not as cellulolytic, and thus many of the forms of biomass studied as substrates required pretreatment prior to efficient fermentation (260). While hydrogen production in *C. bescii* has been improved over wild type levels by the deletion of *ldh*, batch fermentation of C. *saccharolyticus* in pH-controlled bioreactors has resulted in hydrogen production at 95% of the theoretical yield (3.8 mol H_2 / mol glucose vs 4.0 mol H_2 / mol glucose) (291, 417). Indeed, lignocellulosic H_2 production by *C. saccharolyticus*, which has been fine-tuned by studies on the impacts of pH and gas sparging, among other characteristics, is the closest a strain of *Caldicellulosiruptor* has come to being industrially relevant (353, 419, 420). Even so, recent advances in photovoltaic electrolysis mean that scale-up of fermentative H_2 production may not remain viable for long; recent technoeconomic assessments of H_2 production by photovoltaic electrolysis and fermentation of food waste showed that fermentation was at least twice as expensive (421, 422).

Besides H_2, other desirable products from *Caldicellulosiruptor* spp. have all been produced by *C. bescii.* These products include 1,2-propanediol (propylene glycol), acetone, ethanol, and the variety of non-*de novo* alcohols produced from the relevant acids by the AOR-Adh pathway described in Chapter 3 (e.g., hexanol, isobutyrate) (270, 292, 293, 423). While acetone, ethanol, and other alcohols were produced by metabolic engineering, 1,2-propanediol (1,2-PD) is a native product of some *Caldicellulosiruptor* spp. during growth on the methylpentose sugars fucose and rhamnose (269, 424); no 1,2-PD is produced during growth on more traditional substrates, e.g. glucose and xylose. During growth on fucose and rhamnose, *C. bescii* produced 3.3 mM and 0.73 mM 1,2-PD, respectively (269). Based on the gene content of *C. saccharolyticus*, rhamnose and fucose were both previously hypothesized to enter primary carbon metabolism as

dihydroxyacetone phosphate (DHAP), with 1,2-PD formation resulting as a side-product of methylpentose catabolism (251). Both of these sugars have been observed as decorations on plant cell surfaces, and thus, their presence in the biomass encountered by *C. bescii* in natural hot-spring environments is not improbable. Methylpentose metabolism in *Caldicellulosiruptor* spp. has not been studied and warrants further investigation.

Short-chain diols such as 1,2-PD are desirable industrial compounds for the synthesis of polymers (425). 1,2-propanediol is industrially produced from propylene oxide, which is generated by steam cracking during oil processing. Thus, identifying alternative microbial sources for this compound is desirable for all the same reasons as biofuel production. Unfortunately, while some organisms, such as *Thermoanaerobacter thermosaccharolyticum*, can ferment 1,2-PD from common monosaccharide like glucose and xylose at high titer (up to 9 g/L), production by *C. bescii* is simply a side product of growth on the deoxy sugars discussed above (426). Thus, 1,2-PD production by *C. bescii* is likely directly limited (1:1) by how much fucose or rhamnose is consumed, and because both of these substrates are more expensive than this product, such a pathway has little biotechnological value.

Acetone is the most recent product of engineered *C. bescii* metabolism (423) . To produce acetone, three genes were heterologously expressed: those encoding thiolase (*thl* from *Caldanaerobacter subterraneus* subsp. *Tengcongensis*), CoA transferase (*ctfAB* from *Thermosipho melanesiensis*), and acetoacetate decarboxylase (*adc* from *Clostridium acetobutylicum)* (**Figure 5.1**). The acetone pathway (from acetyl-CoA) does not require reductant and thus regeneration of redox cofactors utilized during glycolysis in the generation of acetyl CoA is simply accomplished by H_2 generation. This four gene pathway for acetone production was expressed in a strain that expresses the bifunctional AdhE from *Clostridium thermocellum*, but

without the expression of the *Thermoanaerobacter* sp. X514 Rnf (present in the most ethanologenic engineered strain of *C. bescii*). In this AdhE strain background (MACB1058, Δ*pyrE,* Δ*ldh,* P_{slp}Ct*adhE*), expression of the acetone pathway yielded a strain that produced 9.1 mM (0.53 g/L) acetone, as well as 3.3 mM (0.15 g/L) ethanol. Due to the high K_m of Ctf for acetate, exogenous acetate had to be added to the media before significant acetone titers were observed. Addition of up to 50 mM acetate increased acetone production, but at higher acetate concentrations, acetone decreased again. Interestingly, net acetate production was negative, with the final acetate concentration decreasing by 2.3 mM from the initial concentration of 50 mM in the growth medium (423). This pathway is attractive as a target in *C. bescii* because no reducing equivalents are required, and so if some portion of acetyl CoA is used to generate acetone rather than acetate, no disruption of redox balance should occur. While some theoretical loss in ATP yield might result, this has not proven to be a problem in strains like the ethanol producing AdhE-Rnf strain (examined in Chapter 4) which also utilizes acetyl CoA (270). Like the AdhE-Rnf ethanol pathway, production of acetone was highest below the optimal growth temperature of *C. bescii,* at 65 °C. Moving forward, the most readily apparent tactic for improving acetate production is to decrease the K_m of Ctf for acetate, and/or to improve the overall activity (V_{max}) or turnover number (K_{cat}) of this enzyme through rational design approaches.

The production of ethanol and other alcohols in engineered *C. bescii* strains has been extensively discussed in this work (Chapters 1, 3, and 4). Notably, while multiple routes to ethanol production have been demonstrated, none of these strains approach the necessary titers for industrial viability – the 40 g/L ethanol threshold. The AOR-Adh pathway converted 20 mM isobutyrate to as much as 3.5 mM isobutanol – a yield of 17.5%. While this product was not produced *de novo,* isobutyrate and similar short or medium chain fatty acid (SCFA/MCFA)

feedstocks can be inexpensively produced from fermentation of municipal solid waste by a variety of consortia (89, 427). If the flux or redox balance in the pathway were improved, such as by the use of homologous variants of AOR and Adh with higher activities or different cofactor specificities, organic acid upgrading to alcohols could be improved. Future work on such strains might examine the utilization of exogenous addition of mixtures of SCFAs and MCFAs rather than just single pure compounds, to potentially produce mixed alcohol product streams. Variants of the AOR-Adh pathway designed to overcome the redox imbalances described in Chapter 3 are included as Appendix A. Despite expression on a replicating shuttle vector, which has been demonstrated to yield lower transcript levels than chromosomal expression under the same promoter, one of the pathway variants examined revealed a 60% increase in ethanol production in sealed serum bottles (Appendix A) (423). Expression of this pathway variant from the chromosome in the future should yield even more robust improvements.

Approaches for Enhancing Bioproduct Formation

Improving pathway titers and yields of engineered products in recombinant strains of *C. bescii* is essential to utilizing this microbe as a CBP platform; making such improvements is likely to require strategies beyond the traditional methods employed to date. So far, approaches to engineering this bacterium have relied on assessing what is known about native metabolism (including in related organisms) and following through on informed speculation for strain construction. The next step is to take an integrated systems biology approach for strain engineering. Systems biology will enable quantitative improvements beyond what we have currently achieved by considering connections in metabolism that human observation might not discern.

Genome scale models (GEMs) of metabolic networks attempt to derive information about an organism by linking chemical reactions with the set of enzymes encoded in the genome (428). An ideal GEM can represent the totality of metabolic reactions in relation to proteins in the genome, enabling prediction of substrate consumption, product formation, biomass formation, and relative flux through each individual reaction. Improvements in computational power in combination with the genome-sequencing revolution have enabled the reconstruction of GEMs across a broad range of organisms for a dynamic set of applications, including identification of drug targets in pathogens, identification of unknown gene functions, cancer modelling, microbial community interactions, and design/optimization of metabolic engineering strategies (429-431).

The development of a genome scale metabolic model for *C. bescii*, which includes 610 genes, 718 metabolites, 714 metabolic reactions, and 41 exchange reactions, is the subject of an impending publication from a collaborative effort that includes our laboratory (432). While several metabolic models have been reconstructed for related organisms, such as *C. thermocellum* and *T. saccharolyticum*, this is the first example of a GEM in the genus *Caldicellulosiruptor* (433-436). When compared with the most recent *C. thermocellum* GEM, our *C. bescii* GEM (GEM-iCbes) contains a higher percentage of protein coding genes. Also, when compared to models in other biomass-degrading organisms, GEM-iCbes includes significantly more genes involved in carbohydrate utilization and transport, enabled by a companion publication about the transcriptional regulation of the carbohydrate regulatory network discussed in Chapter 4 (398). GEM-iCbes was extensively curated and validated by manual approaches, including verification of predicted enzyme activities in cell extracts. For example, I conducted assays to validate phosphoglycerate kinase activity in *C. bescii*, which revealed that this activity can be either GTP or ATP dependent, rather than just the ATP dependent reaction (432). The model was tuned and

validated by comparing simulations with experimentally measured biomass and product yields in the wild type, Δ*ldh* strain, and AdhE-Rnf strain during growth on both glucose and fructose, as well as crystalline cellulose.

In its current state, GEM-iCbes is consistent with the experimental results in the aforementioned strains, capable of accurately predicting a variety of phenotypes, such as carbon end-product yields in the wild type and mutant strains and differences in biomass flux observed during growth on different carbohydrates. For example, when the lactate dehydrogenase reaction was removed from the model, the simulations accurately predicted the redirection of carbon and electrons to acetate and hydrogen, respectively, as experimentally observed in the Δ*ldh* strain (291). Similarly, the model accurately predicted the experimentally observed higher biomass in the Δ*ldh* strain during growth on fructose than during growth on glucose.

In addition to establishment of GEM-iCbes, the manuscript focuses on using this model to predict ways to maximize ethanol production in the AdhE-Rnf strain. While the model helped to describe what enzymes might carry the most flux, improving our understanding of metabolism, few physiologically relevant targets were identified that might readily improve ethanol production. In the simulations, ethanol production was maximized when the two hydrogenases were coupled, such that the membrane bound hydrogenase (MBH, ferredoxin dependent) produced H_2 while the bifurcating hydrogenase (BF-H_2ase, ferredoxin and NAD(H) dependent) operated in the direction of H_2 uptake. Whether or not these enzymes can be coupled in this manner under physiological conditions is unknown. Importantly, the BF-H_2ase is the primary H_2 producing enzyme in the wild type strain; deletion of MBH by knocking out processing genes did not decrease H_2 production relative to the wild type (271).

One notable engineering suggestion for the improvement of ethanol production in the AdhE-Rnf strain derived from GEM-iCbes was the expression of a sodium/proton antiporter (Mrp); no antiporter with this activity is predicted in the *C. bescii* genome. As discussed in Chapter 4, both proton and sodium pumping homologs of Rnf are known (392, 393). While the strategy for optimal ethanol production utilized in *C. bescii* would require a proton dependent Rnf, it is possible that the *Thermoanaerobacter* sp. X514 Rnf might be sodium dependent. Therefore, lack of a sodium gradient leaves the enzyme non-functional under physiological conditions, even if activity can be measured by *in vitro* assays. This hypothesis could help explain why no difference in ethanol production between the AdhE and AdhE-Rnf strains is observed during growth in sealed serum bottles (although a significant difference has been observed under controlled fermentation conditions) (270). The addition of an Mrp antiporter to the model in the AdhE-Rnf strain led to Rnf carrying electron flux where it previously had not, increasing predicted ethanol production. Still, as noted in Chapter 4, further genetic modifications to the AdhE-Rnf strain are fraught with risk, as genome rearrangement and loss of the ethanol production phenotype have been observed.

In another approach implied and tested by the model, Rnf could be replaced by soluble hydrogenase-1, SH1, from *Pyrococcus furiosus*. In this strategy, additional NADH for ethanol production would be generated through a series of reactions rather than the single Rnf reaction. Heterologous expression of SH1 would allow H_2 generated by either of the native *C. bescii* hydrogenases, MBH or BF-H_2ase, to be oxidized to generate NADPH. Indeed, SH1 is believed to function in NADPH generation from H_2-reuptake physiologically in native *P. furiosus* metabolism (437). In a strain of *C. bescii* expressing AdhE and SH1, this additional NADPH could be oxidized by the BF-Nfn, yielding NADH and Fd_{red}. In such a strain, GEM-iCbes again predicts increased ethanol production relative to the AdhE-Rnf strain, due to the expanded NADH pool.

Experimentally, the AdhE-Rnf strain generated 3.5 g/L (75 mM) ethanol from the consumption of 11.8 g/L crystalline cellulose. When GEM-iCbes was constrained to consume this same amount of crystalline cellulose and produce the same experimentally determined biomass yield, the maximum predicted ethanol titer was 4 g/L (88 mM). This difference of under 20% between the experimentally determined and simulated ethanol titers is impressive, but still reveals a sub-optimal yield of ethanol per glucose. Based on pathway stoichiometry, AdhE in *C. bescii* should enable the production of 2 moles of ethanol per mole of glucose consumed. From 11.8 g/L crystalline cellulose, a maximum of 6.7 g/L (146 mM) ethanol can be produced, revealing yields of 52% and 60% of the theoretical maximum for the experimentally determined and simulated titers, respectively. With both approaches for improving ethanol production proposed from GEM-iCbes, the Mrp strategy and the SH1 strategy, the model predicted a maximum ethanol titer of 6.4 g/L (140 mM), representing 96% of the theoretical maximum. In each scenario, no lactate and only minimal acetate (1.2 mM) was produced, leaving ethanol as the single reduced carbon end-product of fermentation. Notably, these were not the only strategies which resulted in this ethanol titer, and several other approaches have been proposed because of the model. However, at least while the biomass production was confined to the experimentally observed level, ethanol yields above 96% of the theoretical maximum were not observed *in silico*.

To further improve GEM-iCbes, both transcriptomics data and thermodynamic considerations are currently being added to the model. Application of these kinds of data sets to metabolic models is an emerging field, but developing these considerations should help fortify the reliability of the model while allowing the possibility for exploring growth temperature variations and alternative engineering strategies (e.g. adjustment of expression levels for target enzymes) (438, 439). As GEM-iCbes develops, predictions for improving product formation (e.g. ethanol,

acetone), developing and testing pathways for alternative metabolic engineering products, and even improving total flux from carbohydrate utilization to product formation will also improve, expanding the utility of *C. bescii* as a platform organism for bioproduct formation.

Compared with systems biology approaches like the utilization of a genome scale metabolic model, a less sophisticated methodology to improve metabolic engineering efforts is the use of adaptive evolution. Such methods have yielded significant success in a variety of applications, most notably by improving tolerances to toxic compounds, growth rates of impaired strains, or product yields (440); many of the proprietary strains of bacteria and fungi used in a range of industrial processes were developed with this strategy. In the production of cellulases for lignocellulosic biomass pretreatment discussed in Chapter 1, a strain of *Aspergillus niger* that produces 5-fold more cellulase than the wild type has been developed through adaptive evolution (441). In *Saccharomyces cerevisiae,* adaptive evolution has yielded strains capable of producing high ethanol titers from waste corncob residues without pretreatment, a substrate where the presence of phenolic compounds is usually toxic and inhibitory (442).

Often in adaptive evolution experiments, some form of selective pressure is exerted to help develop the desired phenotype, but the strength of this pressure is variable. For example, to increase tolerance to an end-product, an organism might be grown for many generations in the presence of the compound at just below the inhibitory concentrations. Over the course of dozens, or even hundreds, of culture transfers, mutations which improve fitness under the growth conditions are likely to accumulate, resulting in a strain with the desired traits. This process can be used iteratively by increasing the concentration of inhibitor as the tolerance increases, or successively, to select for a secondary trait after the first has been developed to a desired degree. In some scenarios, especially if a selective pressure for the desired trait cannot be readily exerted,

these experiments can be conducted in many replicates with the inclusion of a mutagen, to help encourage genetic diversity. Under such conditions, the range of developed strains will require extensive screening to identify a lineage where the target phenotype has been achieved.

An engineered strain of *C. thermocellum* (AG553) was developed in which the primary routes to other carbon fermentation end-products besides ethanol were eliminated by gene deletions (202). While this strain produced 3.8 g/L ethanol from growth on 50 g/L cellobiose, it had a severely impaired growth phenotype; despite high yield of ethanol, the growth rate was low and very little substrate was consumed. For further study, this strain was subjected to two successive adaptive evolution experiments (72). In the first experiment, the strain was serially transferred daily in rich medium with 5 g/L cellobiose. Each transfer represented roughly 10 generations, and 150 transfers were conducted. After these 1500 generations, the growth rate roughly doubled. In the second adaptive evolution experiment, substrate concentration was increased to 50 g/L cellobiose and the medium was adjusted from rich to defined. Instead of being transferred daily, strains were transferred weekly, with ethanol measured every three transfers. After 13 weeks (approximately 1000 generations) ethanol production had increased from 9.5 g/L to 22.1 g/L. At this point, the culture was plated on solid medium and a single colony was isolated, characterized, and sequenced.

To date, no adaptive evolution experiments have been conducted in *C. bescii.* Before the adaptive evolution experiments, the *C. thermocellum* AG553 strain produced a remarkably similar amount of ethanol to the gold standard *C. bescii* AdhE-Rnf strain (3.4 g/L vs 3.5 g/L) from growth on 20 g/L crystalline cellulose. Considering the success in *C. thermocellum*, repeating these experiments with the *C. bescii* strain might elicit similarly impressive increases in ethanol generation. Notably, no negative growth phenotype is associated with the AdhE-Rnf strain, and so

it is also plausible that any increase in ethanol titer will not yield the nearly 10-fold improvement seen in *C. thermocellum*. Other potential directed evolution experiments that might yield positive results in *C. bescii* include mutagenic studies of the AOR-Adh strain described in Chapter 3, and transfers of ethanologenic or other genetically tractable strains of *C. bescii* in high ethanol concentrations to improve ethanol tolerance. *C. bescii* growth is slightly inhibited by 430 mM ethanol (20 g/L, 2% w/v), which is significantly above the yield achieved to date, but too low for industrial use (minimum 912 mM ethanol titers required); growth of wild type *C. bescii* is completely inhibited by 650 mM ethanol (30 g/L, 3% w/v) (**Figure 5.2**). Exposing *C. bescii* to mutagenic compounds (e.g. diethyl sulfate) runs the risk of increasing instances of IS*Cbe4* transpositions as a stress response, such as was observed previously in a strain developed by growth on the toxic uracil analog 5-floroorotic acid (279). While the influence of transposable elements in adaptive evolution is well studied, and often considered a boon, the risk of genome instability associated with IS*Cbe4* in *C. bescii* must be considered (443).

An important last step in long-term evolution experiments is sequencing the genome of any strains that yield the desired properties. In an ideal scenario, parallel experiments can be conducted, such that sequencing of multiple strains can reveal differences or similarities correlated with phenotypes. For instance, in strains with increased alcohol tolerance, one might observe similar mutations in lipid biosynthesis pathways. In an evolved strain that produces higher ethanol titers, perhaps mutations in AdhE improved enzyme turnover, or changed redox cofactor specificity. Often, as with the evolved strains of *C. thermocellum*, a variety of mutations of unknown function are observed, which may not seem to reveal useful insights (72). Here, another potential advantage of employing a metabolic model can be noted, whereby if gene deletions or disruptions are observed, exploring the changes with a GEM could be informative. Still, mutations resulting in

changes in specificity are less likely to be informative in such circumstances compared, for instance, with a loss of function mutation that might be more readily probed by a model.

Assessing the current commercialization potential of *C. bescii*

Commercialization of ethanol production requires minimum titers of 42 g/L (912 mM). To achieve such a titer, 100% conversion of 73.9 g/L crystalline cellulose is necessary, based on the theoretical yield of 2 moles ethanol per mole of glucose. While *C. bescii* has been shown to grow on high loads of untreated plant biomass (switchgrass) and crystalline cellulose, each at substrate loadings up to 200 g/L, conversion of such high loads has not been observed (257). To date, complete conversion of more than 12 g/L crystalline cellulose has not been demonstrated with *C. bescii.* Recent work has shown that in continuous culture, solubilization of up to 27% of 50 g/L loadings of untreated switchgrass can be attained, but this does not approach the levels necessary for industrial viability in terms of sugar conversion even if product formation for ethanol could reach theoretical yields (444). For conversion of plant biomass to reach the 42 g/L ethanol benchmark, over 100 g/L switchgrass would need to be both solubilized and fully converted to product. While *C. bescii* certainly has an advantage over *C. thermocellum* in sugar conversion because it can utilize C_5 sugar substrates, the evolved strain of *C. thermocellum* discussed in detail above generated 22.4 g/L ethanol from 60 g/L crystalline cellulose. This result, 75% of the theoretical yield of ethanol, is significantly better than any titer or yield observed with *C. bescii*. Still, this titer is only slightly more than half of the minimum requirement (42 g/L) for economically viable ethanol production.

To date, no CBP organism utilizing the “native strategy” (whereby a naturally cellulolytic microbe is engineered for product formation) has demonstrated sugar conversion and product titers approaching commercial viability for ethanol production. To put these titers in broader

perspective, commercial yeast strains utilized for starch-based ethanol production, e.g. corn ethanol, can reach titers over 100 g/L (2.2 M). For further comparison, lignocellulosic ethanol production using waste streams of rice milling with a traditional non-CBP approach has demonstrated the generation of over 50 g/L ethanol using *S. cerevisiae* (445).

Potential for Development of Improved Genetic Tools

Despite the moderate successes achieved with genetic engineering of *C. bescii* to date, the genetic system is still crude and extremely time-consuming compared with the precise modern techniques available in model organisms such as *E. coli* and *S. cerevisiae.* The arrival of CRISPR/Cas9 genome editing tools has allowed remarkable improvements in the genetic modification of even non-model mesophiles. Unfortunately, until extremely recently, no Type II CRISPR systems, the source of Cas9, had been identified in thermophiles (244, 446, 447). Thus, the thermolability of these enzymes precluded their use in thermophilic prokaryotes. Examination of prokaryotic genomes reveals the presence of CRISPR loci in roughly 40% of the sequenced bacterial genomes and 80% of the archaeal genomes (448). As an alternative to utilizing exogenous Type II systems, there have been successes in "hijacking" the endogenous Type I and Type III systems in a range of species for targeted genome alterations (Type I and Type III) or RNA targeting (Type III) (449, 450).

In the hyperthermophilic archaeon *Sulfolobus islandicus*, both the native Type I-A and Type III-B CRISPR/Cas systems have been utilized to accomplish gene deletions and site directed mutagenesis (449). The Type III system was even used to demonstrate *in situ* gene tagging, by incorporating a 6xHis tag into a target gene. The Type I system was used to make four specific point mutation in a target gene in a single transformation experiment, although in some transformants, only one, two, or three of the mutations were obtained. The use of this system in *S.*

islandicus is of note compared to similar work with a variety of Type I systems in *Clostridium tyrobutyricum, Clostridium difficile, Clostridium pasteurianum, Heliobacterium modesticaldum, Zymomonas mobilis, Lactobacillus crispatus,* and *Haloarcula hispanica* because this was the first instance of utilizing these systems in a thermophile (451-456). Very recently, *C. thermocellum* was engineered using both its native Type I-B system, and the newly discovered thermophilic Cas9 variants from *Geobacillus stearothermophilus*, *Acidothermus cellulolyticus,* and *Geobacillus thermodentrificans* T12 (198). Of these Cas9 variants, *G. stearothermophilus* was the most effective. Due to the low recombination efficiency of *C. thermocellum*, expression of recombineering machinery (*beta/exo*) from the thermophile *Acidithiobacillus caldus* on a replicating vector prior to the CRISPR-based transformations was necessary for satisfactory transformation efficiency to be achieved. In the end, the strain generation time was halved utilizing these CRISPR systems compared with traditional homologous recombination and auxotrophy or antibiotic selection approaches. The *G. stearothermophilus* Cas9 has also been utilized in *Thermoanaerobacter ethanolicus* JW200 to delete thymidine kinase, AdhE, and a homolog of the redox-sensing transcription factor Rex, termed redox sensing protein (RSP); this work represents the highest temperature at which Cas9 has been used for gene editing (65 °C) (457).

The *C. bescii* genome contains seven CRISPR loci, and encodes at least one Type I-B system, with an additional putative I-B system, as well as putative III-A and III-D systems. While neither type III system present in *C. bescii* has been utilized as an endogenous gene editing system before, the Type I-B system closely resembles the system utilized in *C. thermocellum*. Development of this system in the same manner as the work in *C. thermocellum* described above could significantly decrease the time required for *C. bescii* strain construction and allow the achievement of previously difficult to obtain gene deletions. Efficient transformation would likely

require a two-plasmid, two-step process in *C. bescii* as well, where the recombineering machinery is introduced before the CRISPR/Cas system can be effectively utilized. The *G. stearothermophilus* Cas9 variant may also function in *C. bescii*, as growth at 65 °C (demonstrated in *T. ethanolicus* JW200) is feasible, exemplified by the ethanologenic strains.

The advantages of CRISPR/Cas systems over traditional genetic approaches include the ability to make multiple mutations at once, high efficiency, high specificity, decreased off-target effects, and low incidence of false positives (458). Of these, the ability to make multiple mutations at once is one of the most alluring for use in *C. bescii.* Successive rounds of transformation in *C. bescii* increase the instances of the mobile genetic element IS*Cbe4* in the genome, likely because of the high stress endured during competent cell preparation, electroporation, selection, and counter-selection (279). Increased instances of IS*Cbe4* can interfere with individual gene functions, leading to undesirable phenotype, or can help facilitate global genome rearrangements via large scale inversions. Indeed, genome rearrangements observed in *C. bescii* have always been flanked by instances of IS*Cbe4*. Thus, the ability to generate multiple changes to the genome at once, e.g. to simultaneously delete *ldh* and insert an AdhE-Rnf operon elsewhere in the genome could increase genome stability in downstream strains. Even more attractive is the opportunity to use such a system to simultaneously delete, or at least disrupt, all instances of IS*Cbe4* within the genome. Such a strain would hypothetically have improved genome stability, thereby avoiding some of the problems of unanticipated changes in engineered strains that have been previously observed (Chapters 3 and 4). It is unknown what complications could arise from such efforts. It has been previously observed that removal of a restriction modification system (CbeI) in *C. bescii* increased the transformation efficiency, but increased IS*Cbe4* activity (276, 279). It is possible that decreased activity of IS*Cbe4* might increase the activity of other recognized insertion

sequences in *C. bescii,* which have seemed immobile to this point. Despite the advantage of stricter selection and multiplexing mutations, modifications to the genome with a CRISPR/Cas system still rely on recombination events, and so using such a system is not a panacea, especially for the use of Type I-B systems discussed above; if the probability of observing multiple unlikely events is low enough, no strength of selectivity can overcome this. The expression of recombineering machinery from a replicating plasmid during the initial work, as in *C. thermocellum*, should help (198).

Beyond multiplexing insertions/deletions and the creation of a Δ*IS*Cbe4 strain, utilization of a CRISPR/Cas system could enable the deletion of important (bordering on essential) metabolic genes, such as the BF-H_2ase and GAPDH. Currently, researchers in our laboratory have been unable to knock out the genes that encode either of them. The most highly expressed gene knocked out to date is that encoding GOR, which was achieved with the additional selective pressure of marker replacement, as described in Chapter 2. Deletion of *gapdh* is necessary for increasing our understanding of the relative roles of GOR and GAPDH as parallel pathways in primary metabolism. According to the *C. bescii* GEM described above, deletion of GAPDH should be viable; comparison of the Δ*gorSL* strain and a future Δ*gapdh* strain could help further validate the metabolic model while providing insight into why the organism needs two highly expressed parallel mechanisms for GAP oxidation.

A final issue of note in the discussion of strain construction is the precise control of gene expression. In the most sophisticated metabolic engineering efforts, pathway efficiencies are often tuned stoichiometrically by adjusting the expression of individual genes using varied promoters (459). In *C. bescii,* the promoters utilized to date are quite limited. While a single inducible promoter ($P_{\text{xylose isomerase}}$) has been deployed, the majority of heterologous expression has been

conducted with high level constitutively expressed promoters, such as those controlling expression of the *C. bescii* s-layer protein, bifurcating hydrogenase, 30S ribosomal protein, and pyruvate oxidoreductase (246, 267). While expressing pathway genes at high levels might be desirable in some cases, utilizing multiple high-level promoters in conjunction might unduly increase the metabolic burden on the cell and should be approached with caution. It should also be noted that functional use of Cas9 systems (such as the recent work in *C. thermocellum*) may require an inducible promoter, due to the toxicity of the Cas9 in the presence of functional guide RNA (198). In all cases, because promoters themselves have not been well characterized, heterologous expression has involved simply using the 200 bp upstream of a given transcription start site as a promoter region. In our forthcoming publication examining sugar utilization and regulation in *C. bescii*, we also explore the entire regulome of *C. bescii* (398). When utilizing native promoters for genetic work, it is important to consider what regulatory binding sites might be unintentionally included in the 200 bp region, such as the Rex binding sites on the BF-H_2ase and POR promoters utilized previously.

Factors beyond the promoter strength and transcription factor binding can regulate gene expression in *C. bescii* in ways that we do not yet understand. The most obvious example of this point is that expression of the same gene under the same promoter varies depending on where on the chromosome the expression cassette is integrated (270); expression from a replicating shuttle vector is also notably lower than chromosomal integration driven by the same promoter (423). The mechanisms limiting expression based on genomic location are currently unknown but underline the necessity of identifying preferred sites for reliable and consistent recombinant gene expression. The most likely explanation is that these differences result from three-dimensional folding patterns and spatial organization of the chromosome (460, 461). While the architecture of bacterial

chromosomes is an ascendent field, it seems likely that this organization is even more complex in extreme thermophiles. For now, establishing baseline loci for consistent expression of engineered pathways should be sufficient.

Figures

Figure 5.1 Acetone production pathway in engineered *C. bescii*

Primary metabolic pathways and acetone production in recombinant *C. bescii*. Native pathways and enzymes are shown in gray boxes. Recombinantly-expressed enzymes are shown in purple boxes. Native metabolites are shown in blue boxes, while metabolites resulting from engineered metabolism are in green boxes. Abbreviations: Embden-Meyerhof-Parnas glycolytic pathway (Upper EMP), non-oxidative branch of the pentose phosphate pathway (Non-Ox PPP), bifurcating hydrogenase (BF-H_2ase), membrane bound hydrogenase (MBH), bifurcating NADH-dependent reduced ferredoxin $NADP^+$ oxidoreductase (BF-Nfn), thiolase from *Caldanaerobacter subterraneus* subsp. *tengcongensis* (Thl), CoA transferase from *Thermosipha melanesiensis* (Ctf), acetoacetate decarboxylase from *Clostridium acetobutylicum* (Adc). Adapted from Straub, et al (2020) (423).

Figure 5.1

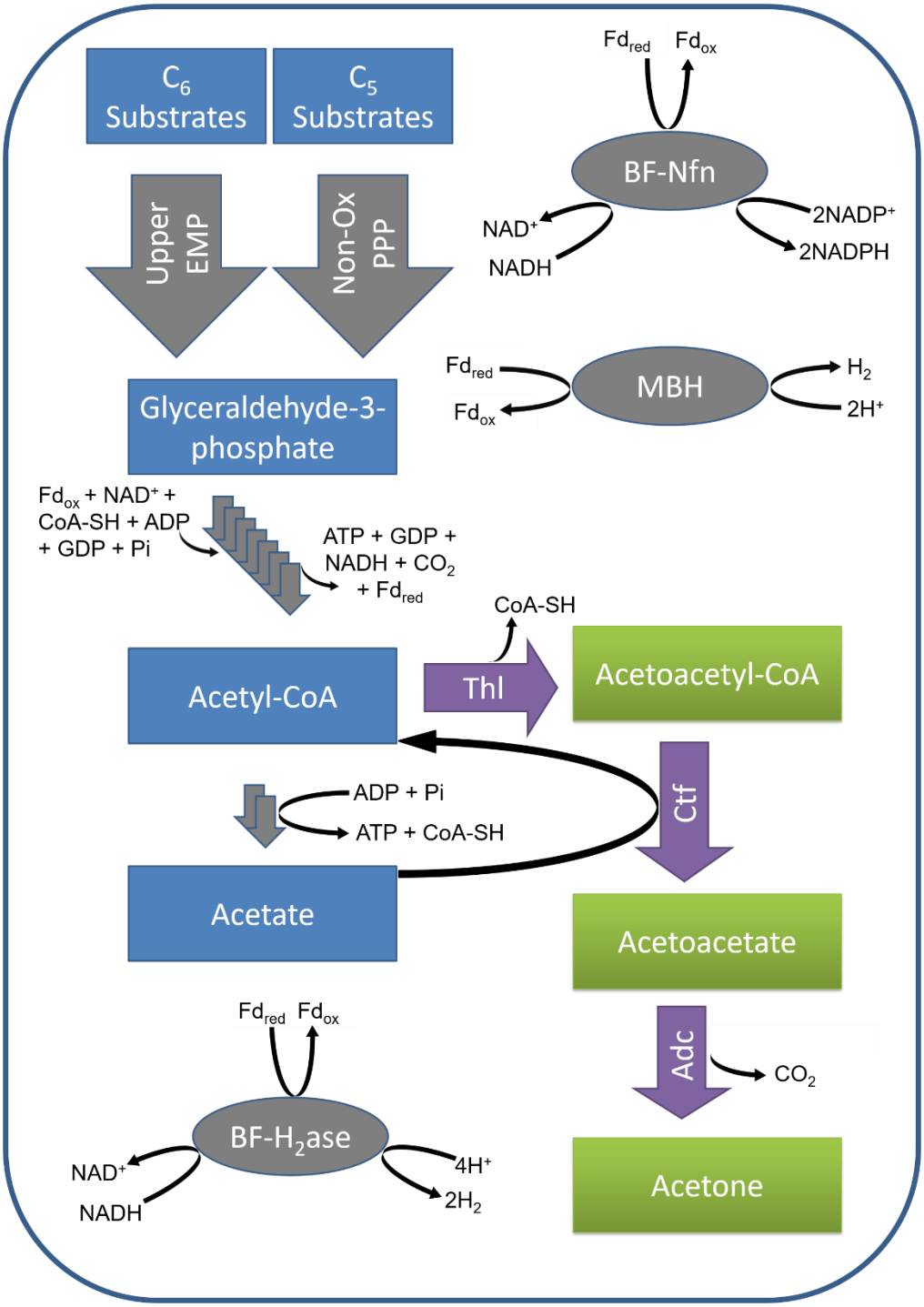
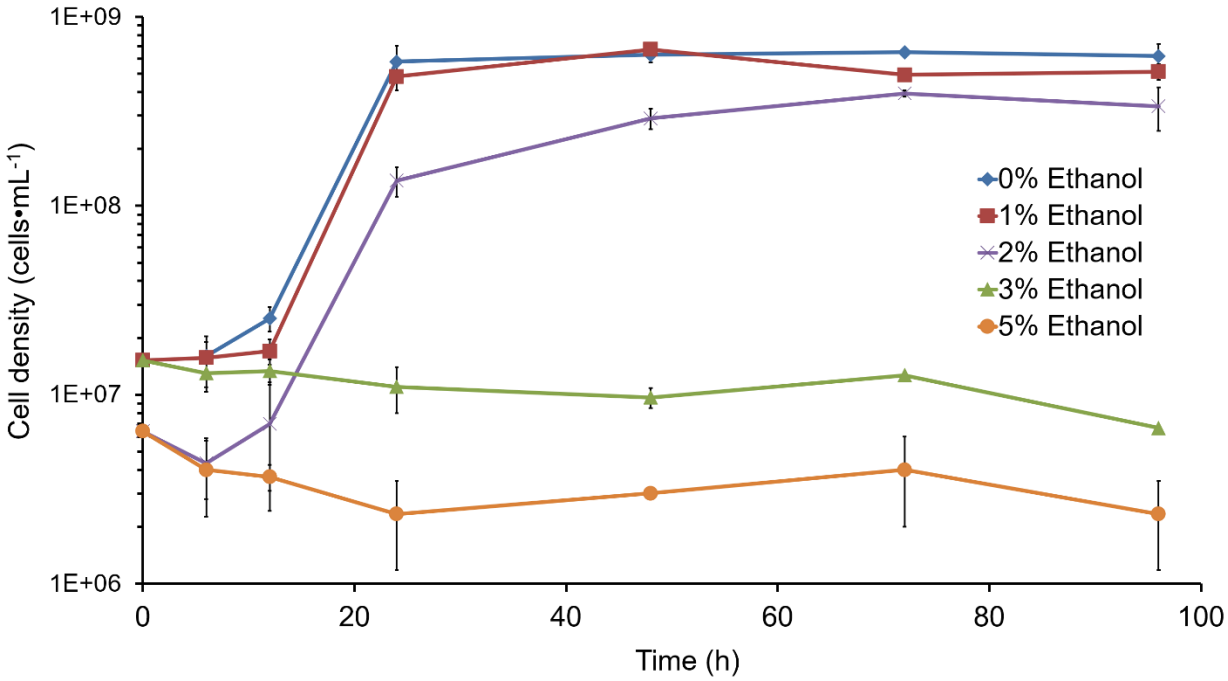

C6 Substrates
C5 Substrates
Upper EMP
Non-Ox PPP
Glyceraldehyde-3-phosphate
Fdred Fdox
BF-Nfn
NAD+
NADH
2NADP+
2NADPH
Fdred
Fdox
MBH
H2
2H+
Fdox + NAD+ + CoA-SH + ADP + GDP + Pi
ATP + GDP + NADH + CO2 + Fdred
CoA-SH
Acetyl-CoA
Thl
Acetoacetyl-CoA
ADP + Pi
ATP + CoA-SH
Ctf
Acetate
Acetoacetate
Adc
CO2
Acetone
Fdred Fdox
BF-H2ase
NAD+
NADH
4H+
2H2

Figure 5.2 Tolerance of *C. bescii* to growth in ethanol

Growth of wild type *C. bescii* (DSMZ 6725) at 78 °C in varying initial concentrations of ethanol (0-5% w/v) with agitation at 150 rpm. Cultures were grown anaerobically in sealed 100 mL serum vials containing 50 mL of modified DSM 516 growth medium with cellobiose as the substrate. Growth was uninhibited by 1% w/v ethanol (217 mM, 10 g/L, red squares), and only minimal impairment of growth was observed in 2% w/v ethanol (435 mM, 20 g/L, purple crosses). No growth was observed in ethanol of concentrations of 3% w/v (652 mM, 30 g/L, green triangles) and 5% w/v (1087 mM, 50 g/L, orange circles).

Figure 5.2


1E+09
1E+08
1E+07
1E+06
Cell density (cells•mL-1)
0
20
40
60
80
100
Time (h)
0% Ethanol
1% Ethanol
2% Ethanol
3% Ethanol
5% Ethanol

REFERENCES

1. BP Statistical Review of World Energy. 2020. 69th edition ed.
2. Smil V. 2016. Energy transitions: global and national perspectives. ABC-CLIO.
3. Etelmad B, Luciani J, Bairoch P, Toutain J. 1991. World Energy Production, 1800-1985. Geneve: Droz.
4. International Energy Statistics: Total Oil Production 1980-2014. 2014. U.S. Energy Information Administration.
5. Miller RG, Sorrell SR. 2013. The future of oil supply. Philosophical transactions Series A, Mathematical, physical, and engineering sciences 372:20130179-20130179.
6. Maugeri L. 2013. The shale oil boom: a US phenomenon. Harvard Kennedy School, Belfer Center for Science and International Affairs.
7. Hunt JM. 1991. Generation of gas and oil from coal and other terrestrial organic matter. Organic Geochemistry 17:673-680.
8. Tissot BP, Welte DH. 2013. Petroleum formation and occurrence. Springer Science & Business Media.
9. Hubbert MK. Nuclear energy and the fossil fuel, p. *In* Institute AP (ed),
10. Hughes L, Rudolph J. 2011. Future world oil production: growth, plateau, or peak? Current Opinion in Environmental Sustainability 3:225-234.
11. World energy balances. 2020. https://www.iea.org.
12. Global trends in renewable energy investment report 2018. 2018. Bloomberg New Energy Finance.

13. Basic electric power statistics. 7/5/20. 2016. https://chinaenergyportal.org/en/2016-detailed-electricity-statistics-updated/. Accessed

14. Renewable capacity statistics 2019. 2019. International Renewable Energy Agency (IRENA).

15. Shukla P, Skea J, Calvo Buendia E, Masson-Delmotte V, Pörtner H, Roberts D, Zhai P, Slade R, Connors S, Van Diemen R. 2019. Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

16. State of the Climate in 2019. 2020. Bulletin of the American Meteorological Society 101:S1-S429.

17. State of the climate: global climate report for annual 2019. NOAA National Centers for Environmental Information, https://www.ncdc.noaa.gov/sotc/global/201913.

18. Böttner C, Haeckel M, Schmidt M, Berndt C, Vielstädte L, Kutsch JA, Karstens J, Weiß T. 2020. Greenhouse gas emissions from marine decommissioned hydrocarbon wells: leakage detection, monitoring and mitigation strategies. International Journal of Greenhouse Gas Control 100:103119.

19. Schout G, Griffioen J, Hassanizadeh SM, Cardon de Lichtbuer G, Hartog N. 2019. Occurrence and fate of methane leakage from cut and buried abandoned gas wells in the Netherlands. Science of the Total Environment 659:773-782.

20. Carbon dioxide emissions coefficients by fuel. 2016. *on* U.S. Energy Information Administration. https://www.eia.gov/environment/emissions/co2_vol_mass.php. Accessed 9/16/2020.

21. Alvarez RA, Zavala-Araiza D, Lyon DR, Allen DT, Barkley ZR, Brandt AR, Davis KJ, Herndon SC, Jacob DJ, Karion A, Kort EA, Lamb BK, Lauvaux T, Maasakkers JD, Marchese AJ, Omara M, Pacala SW, Peischl J, Robinson AL, Shepson PB, Sweeney C, Townsend-Small A, Wofsy SC, Hamburg SP. 2018. Assessment of methane emissions from the U.S. oil and gas supply chain. Science 361:186-188.

22. Shine KP, Fuglestvedt JS, Hailemariam K, Stuber N. 2005. Alternatives to the global warming potential for comparing climate impacts of emissions of greenhouse gases. Climatic Change 68:281-302.

23. Frolkis VA, Karol IL, Kiselev AA. 2002. Global warming potential, global warming commitment and other indexes as characteristics of the effects of greenhouse gases on Earth's climate. Ecological Indicators 2:109-121.

24. Lashof DA, Ahuja DR. 1990. Relative contributions of greenhouse gas emissions to global warming. Nature 344:529-531.

25. Abram NJ, Wolff EW, Curran MAJ. 2013. A review of sea ice proxy information from polar ice cores. Quaternary Science Reviews 79:168-183.

26. Stauffer B, Oeschger H. 1985. Gaseous components in the atmosphere and the historic record revealed by ice cores. Annals of Glaciology 7:54-59.

27. Marsh GE. 2014. Interglacials, Milankovitch cycles, solar activity, and carbon dioxide. Journal of Climatology 2014:345482.

28. Mawad R. 2015. On the correlation between Earth's orbital perturbations and oscillations of sea level and concentration of greenhouse gases. Journal of Modern Trends in Physics Research 15:1-9.

29. Tarling D. Milankovitch cycles in climate change, geology and geophysics, p 1-14. *In* (ed),

30. Arguez A, Hurley S, Inamdar A, Mahoney L, Sanchez-Lugo A, Yang L. 2020. Should we expect each year in the next decade (2019–28) to be ranked among the top 10 warmest years globally? Bulletin of the American Meteorological Society 101:E655-E663.

31. Mostofa KM, Liu CQ, Zhai W, Minella M, Vione DV, Gao K, Minakata D, Arakaki T, Yoshioka T, Hayakawa K. 2016. Reviews and syntheses: ocean acidification and its potential impacts on marine ecosystems. Biogeosciences 13:1767-1786.

32. Hutchins DA, Fu F. 2017. Microorganisms and ocean global change. Nature Microbiology 2:17058.

33. Colossi Brustolin M, Nagelkerken I, Moitinho Ferreira C, Urs Goldenberg S, Ullah H, Fonseca G. 2019. Future ocean climate homogenizes communities across habitats through diversity loss and rise of generalist species. Global Change Biology 25:3539-3548.

34. Aponte C, de Groot WJ, Wotton BM. 2016. Forest fires and climate change: causes, consequences and management options. International Journal of Wildland Fire 25:i-ii.

35. Barbero R, Abatzoglou JT, Larkin NK, Kolden CA, Stocks B. 2015. Climate change presents increased potential for very large fires in the contiguous United States. International Journal of Wildland Fire 24:892-899.

36. Flannigan MD, Stocks BJ, Wotton BM. 2000. Climate change and forest fires. Science of the Total Environment 262:221-229.

37. Moon I-J, Kim S-H, Chan JCL. 2019. Climate change and tropical cyclone trend. Nature 570:E3-E5.

38. Walsh KJE, McBride JL, Klotzbach PJ, Balachandran S, Camargo SJ, Holland G, Knutson TR, Kossin JP, Lee T-c, Sobel A, Sugi M. 2016. Tropical cyclones and climate change. WIREs Climate Change 7:65-89.

39. Walter Anthony KM, Anthony P, Grosse G, Chanton J. 2012. Geologic methane seeps along boundaries of Arctic permafrost thaw and melting glaciers. Nature Geoscience 5:419-426.

40. Maslin M, Owen M, Day S, Long D. 2004. Linking continental-slope failures and climate change: testing the clathrate gun hypothesis. Geology 32:53-56.

41. Dickens GR. 2003. A methane trigger for rapid warming? Science 299:1017-1017.

42. Gross M. 2019. Permafrost thaw releases problems. Current Biology 29:R39-R41.

43. Smith AW, Skilling DE, Castello JD, Rogers SO. 2004. Ice as a reservoir for pathogenic human viruses: specifically, caliciviruses, influenza viruses, and enteroviruses. Medical Hypotheses 63:560-566.

44. Monthly energy review. 2020. *on* U.S. Energy Information Administration. https://www.eia.gov/. Accessed 9/1/2020.

45. Administration EI, Office GP. 2016. International energy outlook 2016, with projections to 2040. Government Printing Office.

46. Ahmed A, Waqas M, Naser N, Singh E, Roberts W, Chung S, Sarathy M. 2016. Compositional effects of gasoline fuels on combustion, performance and emissions in engine. SAE International Journal of Fuels and Lubricants 9:460-468.

47. Fuel properties comparison. https://afdc.energy.gov/fuels/properties. Accessed 4/28/20.

48. Sullivan JL, Baker RE, Boyer BA, Hammerle RH, Kenney TE, Muniz L, Wallington TJ. 2004. CO_2 emission benefit of diesel (versus gasoline) powered vehicles. Environmental Science and Technology 38:3217-3223.

49. Larsen U, Johansen T, Schramm J. 2009. Ethanol as a future fuel for road transportation: main report, DTU Mekanik.

50. Energy Policy Act. 2005. USA.

51. How much ethanol is in gasoline, and how does it affect fuel economy? 6/24/20 2020. https://www.eia.gov/tools/faqs/faq.php?id=27&t=4. Accessed 9/10/20.

52. Policarpo NA, Frutuoso FS, Cassiano DR, Cavalcante FSA, Araújo RS, Bertoncini BV, Oliveira MLM. 2018. Emission estimates for an on-road flex-fuel vehicles operated by ethanol-gasoline blends in an urban region, Brazil. Urban Climate 24:111-120.

53. Flexible fuel vehicles. https://afdc.energy.gov/vehicles/flexible_fuel.html. Accessed 9/10/20.

54. Automotores AaoNdFdVc. 2019. Anuário da indústria automobilística brasileira 2019. ANFAVEA São Paulo.

55. Why hybrid car sales are stalling. p *In* The Toronto Star. www.thestar.com. https://www.thestar.com/business/2015/06/19/why-hybrid-car-sales-are-stalling.html.

56. Lal R. 2007. Biofuels from crop residues. Soil and Tillage Research:237-238.

57. Campbell MN. 2008. Biodiesel: algae as a renewable source for liquid fuel. Guelph Engineering Journal 1:2-7.

58. Maroa S, Inambao F. 2020. Biodiesel feedstocks, p 29-43, Biodiesel, Combustion, Performance and Emissions Characteristics. Springer.

59. Pryde E. 1980. Composition of soybean oil, p 13-31, Handbook of Soy Oil Processing and Utilization.

60. Noureddini H, Zhu D. 1997. Kinetics of transesterification of soybean oil. Journal of the American Oil Chemists' Society 74:1457-1463.

61. Kumar M, Sun Y, Rathour R, Pandey A, Thakur IS, Tsang DCW. 2020. Algae as potential feedstock for the production of biofuels and value-added products: opportunities and challenges. Science of the Total Environment 716:137116.

62. Siaut M, Cuiné S, Cagnon C, Fessler B, Nguyen M, Carrier P, Beyly A, Beisson F, Triantaphylidès C, Li-Beisson Y, Peltier G. 2011. Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves. BMC Biotechnology 11:7.

63. Ma Y, Wang Z, Yu C, Yin Y, Zhou G. 2014. Evaluation of the potential of 9 *Nannochloropsis* strains for biodiesel production. Bioresource Technology 167:503-509.

64. Work VH, Radakovits R, Jinkerson RE, Meuser JE, Elliott LG, Vinyard DJ, Laurens LML, Dismukes GC, Posewitz MC. 2010. Increased lipid accumulation in the *Chlamydomonas reinhardtii* starchless isoamylase mutant and Increased carbohydrate synthesis in complemented strains. Eukaryotic Cell 9:1251-1261.

65. Borowitzka MA. 1999. Commercial production of microalgae: ponds, tanks, and fermenters, p 313-321. *In* Osinga R, Tramper J, Burgess JG, Wijffels RH (ed), Progress in Industrial Microbiology, vol 35. Elsevier.

66. Flynn KJ, Mitra A, Greenwell HC, Sui J. 2013. Monster potential meets potential monster: pros and cons of deploying genetically modified microalgae for biofuels production. Interface Focus 3:20120037.

67. Long-term cultivation of algae in open-raceway ponds: lessons from the field. 2015. Industrial Biotechnology 11:213-220.

68. Szyjka SJ, Mandal S, Schoepp NG, Tyler BM, Yohn CB, Poon YS, Villareal S, Burkart MD, Shurin JB, Mayfield SP. 2017. Evaluation of phenotype stability and ecological risk of a genetically engineered alga in open pond production. Algal Research 24:378-386.

69. Vohra M, Manwar J, Manmode R, Padgilwar S, Patil S. 2014. Bioethanol production: feedstock and current technologies. Journal of Environmental Chemical Engineering 2:573-584.

70. Madson P, Monceaux D. 1995. Fuel ethanol production. The Alcohol Textbook 3.

71. Xu Y, Isom L, Hanna MA. 2010. Adding value to carbon dioxide from ethanol fermentations. Bioresource Technology 101:3311-3319.

72. Tian L, Papanek B, Olson DG, Rydzak T, Holwerda EK, Zheng T, Zhou J, Maloney M, Jiang N, Giannone RJ, Hettich RL, Guss AM, Lynd LR. 2016. Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. Biotechnology for Biofuels 9:116.

73. Zacchi G, Axelsson A. 1989. Economic evaluation of preconcentration in production of ethanol from dilute sugar solutions. Biotechnology and Bioengineering 34:223-233.

74. Ethanol production and distribution. https://afdc.energy.gov/fuels/ethanol_production.html. Accessed 9/1/20.

75. Baeyens J, Kang Q, Appels L, Dewil R, Lv Y, Tan T. 2015. Challenges and opportunities in improving the production of bio-ethanol. Progress in Energy and Combustion Science 47:60-88.

76. Mueller S. 2010. Detailed report: 2008 national dry mill corn ethanol survey. University of Illinois at Chicago, Chicago, IL.

77. Good D. 2016. Prospects for corn use for ethanol production. farmdoc daily 6.

78. Zhang Z, Lohr L, Escalante C, Wetzstein M. 2010. Food versus fuel: what do prices tell us? Energy Policy 38:445-451.

79. Thompson PB. 2012. The agricultural ethics of biofuels: the food vs. fuel debate. Agriculture 2:339-358.

80. Ajanovic A. 2011. Biofuels versus food production: does biofuels production increase food prices? Energy 36:2070-2076.

81. Pimentel D, Marklein A, Toth MA, Karpoff MN, Paul GS, McCormack R, Kyriazis J, Krueger T. 2009. Food versus biofuels: environmental and economic costs. Human Ecology 37:1.

82. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proceedings of the National Academy of Sciences 103:11206-11210.

83. Mitchell R, Schmer M, Anderson W, Jin V, Balkcom K, Kiniry J, Coffin A, White P. 2016. Dedicated energy crops and crop residues for bioenergy feedstocks in the central and eastern USA. Bioenergy Research 9:384-398.

84. Wang M, Han J, Dunn JB, Cai H, Elgowainy A. 2012. Well-to-wheels energy use and greenhouse gas emissions of ethanol from corn, sugarcane and cellulosic biomass for US use. Environmental Research Letters 7:045905.

85. M-F Johnson J. 2019. A “Soil Lorax” perspective on corn stover for advanced biofuels. Agronomy Journal 111:59-62.

86. Callegari A, Bolognesi S, Cecconet D, Capodaglio AG. 2020. Production technologies, current role, and future prospects of biofuels feedstocks: a state-of-the-art review. Critical Reviews in Environmental Science and Technology 50:384-436.

87. Campuzano R, González-Martínez S. 2016. Characteristics of the organic fraction of municipal solid waste and methane production: a review. Waste Management 54:3-12.

88. Zheng X, Ying Z, Wang B, Chen C. 2018. Hydrogen and syngas production from municipal solid waste (MSW) gasification via reusing CO2. Applied Thermal Engineering 144:242-247.

89. Grootscholten TIM, Strik DPBTB, Steinbusch KJJ, Buisman CJN, Hamelers HVM. 2014. Two-stage medium chain fatty acid (MCFA) production from municipal solid waste and ethanol. Applied Energy 116:223-229.

90. Ioelovich M. 2014. Cellulose: nanostructured natural polymer. LAP LAMBERT Academic Publishing.

91. Lin Y, Tanaka S. 2006. Ethanol fermentation from biomass resources: current state and prospects. Applied Microbiology and Biotechnology 69:627-642.

92. Antizar-Ladislao B, Turrion-Gomez JL. 2008. Second-generation biofuels and local bioenergy systems. Biofuels, Bioproducts and Biorefining 2:455-469.

93. Prasad RK, Chatterjee S, Mazumder PB, Gupta SK, Sharma S, Vairale MG, Datta S, Dwivedi SK, Gupta DK. 2019. Bioethanol production from waste lignocelluloses: a review on microbial degradation potential. Chemosphere 231:588-606.

94. Parsons S, McManus MC, Taylor CM. 2018. Chapter 13 - Second-Generation Ethanol from Lignocellulose, p 193-206. *In* Thornley P, Adams P (ed), Greenhouse Gas Balances of Bioenergy Systems doi:https://doi.org/10.1016/B978-0-08-101036-5.00013-6. Academic Press.

95. Hofsetz K, Silva MA. 2012. Brazilian sugarcane bagasse: energy and non-energy consumption. Biomass and Bioenergy 46:564-573.

96. Hoover F-A, Abraham J. 2009. A comparison of corn-based ethanol with cellulosic ethanol as replacements for petroleum-based fuels: a review. International Journal of Sustainable Energy 28:171-182.

97. Somma D, Lobkowicz H, Deason JP. 2010. Growing America's fuel: an analysis of corn and cellulosic ethanol feasibility in the United States. Clean Technologies and Environmental Policy 12:373-380.

98. McAloon A, Taylor F, Yee W, Ibsen K, Wooley R. 2000. Determining the cost of producing ethanol from corn starch and lignocellulosic feedstocks. National Renewable Energy Lab, Golden, CO (US),

99. EPA finalizes Renewable Fuel Standard for 2019, reflecting cellulosic biofuel shortfalls. 2018. *on* U.S. Energy Information Administration. https://www.eia.gov/todayinenergy/detail.php?id=37712#. Accessed 9/15/2020.

100. Renewable natural gas production. *on* U.S. Department of Energy. https://afdc.energy.gov/fuels/natural_gas_renewable.html. Accessed 9/5/20.

101. Smith PJ, Wang H-T, York WS, Peña MJ, Urbanowicz BR. 2017. Designer biomass for next-generation biorefineries: leveraging recent insights into xylan structure and biosynthesis. Biotechnology for Biofuels 10:286.

102. Vanholme R, De Meester B, Ralph J, Boerjan W. 2019. Lignin biosynthesis and its integration into metabolism. Current Opinion in Biotechnology 56:230-239.

103. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. 2010. Lignin biosynthesis and structure. Plant Physiology 153:895-905.

104. Zhao X, Zhang L, Liu D. 2012. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. Biofuels, Bioproducts and Biorefining 6:465-482.

105. Li M, Pu Y, Ragauskas AJ. 2016. Current understanding of the correlation of lignin structure with biomass recalcitrance. Frontiers in Chemistry 4.

106. Lee HV, Hamid SBA, Zain SK. 2014. Conversion of lignocellulosic biomass to nanocellulose: structure and chemical process. The Scientific World Journal 2014:631013.

107. Cosgrove DJ. 2014. Re-constructing our models of cellulose and primary cell wall assembly. Current Opinion in Plant Biology 22:122-131.

108. Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annual Review of Plant Biology 61:263-289.

109. Wyman CE. 2007. What is (and is not) vital to advancing cellulosic ethanol. Trends in Biotechnology 25:153-157.

110. Padella M, O'Connell A, Prussi M. 2019. What is still limiting the deployment of cellulosic ethanol? Analysis of the current status of the sector. Applied Sciences 9:4523.

111. Wyman CE, Dale BE, Elander RT, Holtzapple M, Ladisch MR, Lee YY. 2005. Coordinated development of leading biomass pretreatment technologies. Bioresource Technology 96:1959-1966.

112. Chen H, Liu J, Chang X, Chen D, Xue Y, Liu P, Lin H, Han S. 2017. A review on the pretreatment of lignocellulose for high-value chemicals. Fuel Processing Technology 160:196-206.

113. Yang B, Wyman CE. 2008. Pretreatment: the key to unlocking low-cost cellulosic ethanol. Biofuels, Bioproducts and Biorefining 2:26-40.

114. Karimi K, Shafiei M, Kumar R. 2013. Progress in physical and chemical pretreatment of lignocellulosic biomass, p 53-96, Biofuel Technologies. Springer.

115. Zheng Y, Pan Z, Zhang R. 2009. Overview of biomass pretreatment for cellulosic ethanol production. International Journal of Agricultural and Biological Engineering 2:51-68.

116. Taherzadeh MJ, Karimi K. 2008. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review. International Journal of Molecular Sciences 9:1621-1651.

117. Amiri H, Karimi K, Zilouei H. 2014. Organosolv pretreatment of rice straw for efficient acetone, butanol, and ethanol production. Bioresource Technology 152:450-456.

118. Li C, Knierim B, Manisseri C, Arora R, Scheller HV, Auer M, Vogel KP, Simmons BA, Singh S. 2010. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: biomass recalcitrance, delignification and enzymatic saccharification. Bioresource Technology 101:4900-4906.

119. van der Pol EC, Bakker RR, Baets P, Eggink G. 2014. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Applied Microbiology and Biotechnology 98:9579-9593.

120. Zhang L, Xi G, Zhang J, Yu H, Wang X. 2017. Efficient catalytic system for the direct transformation of lignocellulosic biomass to furfural and 5-hydroxymethylfurfural. Bioresource Technology 224:656-661.

121. Smith MD, Cai CM, Cheng X, Petridis L, Smith JC. 2018. Temperature-dependent phase behaviour of tetrahydrofuran–water alters solubilization of xylan to improve co-production of furfurals from lignocellulosic biomass. Green Chemistry 20:1612-1620.

122. Saritha M, Arora A, Lata. 2012. Biological pretreatment of lignocellulosic substrates for enhanced delignification and enzymatic digestibility. Indian Journal of Microbiology 52:122-130.

123. Ragauskas AJ, Beckham GT, Biddy MJ, Chandra R, Chen F, Davis MF, Davison BH, Dixon RA, Gilna P, Keller M, Langan P, Naskar AK, Saddler JN, Tschaplinski TJ, Tuskan GA, Wyman CE. 2014. Lignin valorization: improving lignin processing in the biorefinery. Science 344:1246843.

124. Cao L, Yu IKM, Liu Y, Ruan X, Tsang DCW, Hunt AJ, Ok YS, Song H, Zhang S. 2018. Lignin valorization for the production of renewable chemicals: state-of-the-art review and future prospects. Bioresource Technology 269:465-475.

125. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. 2016. Opportunities and challenges in biological lignin valorization. Current Opinion in Biotechnology 42:40-53.

126. Wei Kit Chin D, Lim S, Pang YL, Lam MK. 2020. Fundamental review of organosolv pretreatment and its challenges in emerging consolidated bioprocessing. Biofuels, Bioproducts and Biorefining 14:808-829.

127. Zhou Z, Lei F, Li P, Jiang J. 2018. Lignocellulosic biomass to biofuels and biochemicals: a comprehensive review with a focus on ethanol organosolv pretreatment technology. Biotechnology and Bioengineering 115:2683-2702.

128. Tang C, Chen Y, Liu J, Shen T, Cao Z, Shan J, Zhu C, Ying H. 2017. Sustainable biobutanol production using alkali-catalyzed organosolv pretreated cornstalks. Industrial Crops and Products 95:383-392.

129. Patri AS, McAlister L, Cai CM, Kumar R, Wyman CE. 2019. CELF significantly reduces milling requirements and improves soaking effectiveness for maximum sugar recovery of

Alamo switchgrass over dilute sulfuric acid pretreatment. Biotechnology for Biofuels 12:177.

130. Nguyen DMN, Schut GJ, Zadvornyy OA, Tokmina-Lukaszewska M, Poudel S, Lipscomb GL, Adams LA, Dinsmore JT, Nixon WJ, Boyd ES, Bothner B, Peters JW, Adams MWW. 2017. Two functionally distinct NADP+-dependent ferredoxin oxidoreductases maintain the primary redox balance of *Pyrococcus furiosus*. Journal of Biological Chemistry 292:14603-14616.

131. Nguyen TY, Cai CM, Osman O, Kumar R, Wyman CE. 2016. CELF pretreatment of corn stover boosts ethanol titers and yields from high solids SSF with low enzyme loadings. Green Chemistry 18:1581-1589.

132. Smith MD, Mostofian B, Cheng X, Petridis L, Cai CM, Wyman CE, Smith JC. 2016. Cosolvent pretreatment in cellulosic biofuel production: effect of tetrahydrofuran-water on lignin structure and dynamics. Green Chemistry 18:1268-1277.

133. Dale BE. 1986. Method for increasing the reactivity and digestibility of cellulose with ammonia. Google Patents.

134. Lau MW, Dale BE. 2009. Cellulosic ethanol production from AFEX-treated corn stover using *Saccharomyces cerevisiae* 424A(LNH-ST). Proceedings of the National Academy of Sciences 106:1368-1373.

135. Kilpeläinen I, Xie H, King A, Granstrom M, Heikkinen S, Argyropoulos DS. 2007. Dissolution of wood in Ionic Liquids. Journal of Agricultural and Food Chemistry 55:9142-9148.

136. Chen Z, Long J. 2016. Organosolv liquefaction of sugarcane bagasse catalyzed by acidic ionic liquids. Bioresource Technology 214:16-23.

137. Cheng F, Sun J, Wang Z, Zhao X, Hu Y. 2019. Organosolv fractionation and simultaneous conversion of lignocellulosic biomass in aqueous 1,4-butanediol/acidic ionic-liquids solution. Industrial Crops and Products 138:111573.

138. Lopes AM, Ferreira Filho EX, Moreira LRS. 2018. An update on enzymatic cocktails for lignocellulose breakdown. Journal of Applied Microbiology 125:632-645.

139. Sukumaran RK, Singhania RR, Mathew GM, Pandey A. 2009. Cellulase production using biomass feed stock and its application in lignocellulose saccharification for bio-ethanol production. Renewable Energy 34:421-424.

140. Hong Y, Nizami A-S, Pour Bafrani M, Saville BA, MacLean HL. 2013. Impact of cellulase production on environmental and financial metrics for lignocellulosic ethanol. Biofuels, Bioproducts and Biorefining 7:303-313.

141. Olofsson J, Barta Z, Börjesson P, Wallberg O. 2017. Integrating enzyme fermentation in lignocellulosic ethanol production: life-cycle assessment and techno-economic analysis. Biotechnology for Biofuels 10:51.

142. Kabel MA, Bos G, Zeevalking J, Voragen AGJ, Schols HA. 2007. Effect of pretreatment severity on xylan solubility and enzymatic breakdown of the remaining cellulose from wheat straw. Bioresource Technology 98:2034-2042.

143. Wyman CE, Balan V, Dale BE, Elander RT, Falls M, Hames B, Holtzapple MT, Ladisch MR, Lee YY, Mosier N, Pallapolu VR, Shi J, Thomas SR, Warner RE. 2011. Comparative data on effects of leading pretreatments and enzyme loadings and formulations on sugar yields from different switchgrass sources. Bioresource Technology 102:11052-11062.

144. Liu G, Zhang J, Bao J. 2016. Cost evaluation of cellulase enzyme for industrial-scale cellulosic ethanol production based on rigorous Aspen Plus modeling. Bioprocess and Biosystems Engineering 39:133-140.

145. Cherry JR, Fidantsef AL. 2003. Directed evolution of industrial enzymes: an update. Current Opinion in Biotechnology 14:438-443.

146. Dashtban M, Buchkowski R, Qin W. 2011. Effect of different carbon sources on cellulase production by *Hypocrea jecorina* (*Trichoderma reesei*) strains. International Journal of Biochemistry and Molecular Biology 2:274.

147. Novy V, Nielsen F, Seiboth B, Nidetzky B. 2019. The influence of feedstock characteristics on enzyme production in *Trichoderma reesei*: a review on productivity, gene regulation and secretion profiles. Biotechnology for Biofuels 12:238.

148. Dey P, Singh J, Scaria J, Anand AP. 2018. Improved production of cellulase by *Trichoderma reesei* (MTCC 164) from coconut mesocarp-based lignocellulosic wastes under response surface-optimized condition. 3 Biotech 8:402.

149. Ellilä S, Fonseca L, Uchima C, Cota J, Goldman GH, Saloheimo M, Sacon V, Siika-aho M. 2017. Development of a low-cost cellulase production process using *Trichoderma reesei* for Brazilian biorefineries. Biotechnology for Biofuels 10:30.

150. Coffman AM, Li Q, Ju L-K. 2014. Effect of natural and pretreated soybean hulls on enzyme production by *Trichoderma reesei*. Journal of the American Oil Chemists' Society 91:1331-1338.

151. Rana V, Eckard AD, Teller P, Ahring BK. 2014. On-site enzymes produced from *Trichoderma reesei* RUT-C30 and *Aspergillus saccharolyticus* for hydrolysis of wet exploded corn stover and loblolly pine. Bioresource Technology 154:282-289.

152. Alfani F, Gallifuoco A, Saporosi A, Spera A, Cantarella M. 2000. Comparison of SHF and SSF processes for the bioconversion of steam-exploded wheat straw. Journal of Industrial Microbiology and Biotechnology 25:184-192.

153. Wingren A, Galbe M, Zacchi G. 2003. Techno-economic evaluation of producing ethanol from softwood: comparison of SSF and SHF and identification of bottlenecks. Biotechnology Progress 19:1109-1117.

154. Tomás-Pejó E, Oliva JM, Ballesteros M, Olsson L. 2008. Comparison of SHF and SSF processes from steam-exploded wheat straw for ethanol production by xylose-fermenting and robust glucose-fermenting *Saccharomyces cerevisiae* strains. Biotechnology and Bioengineering 100:1122-1131.

155. Bondesson P-M, Galbe M. 2016. Process design of SSCF for ethanol production from steam-pretreated, acetic-acid-impregnated wheat straw. Biotechnology for Biofuels 9:222.

156. Zhang Q, Bao J. 2017. Industrial cellulase performance in the simultaneous saccharification and co-fermentation (SSCF) of corn stover for high-titer ethanol production. Bioresources and Bioprocessing 4:17.

157. Lynd LR, Van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology 16:577-583.

158. Parisutham V, Kim TH, Lee SK. 2014. Feasibilities of consolidated bioprocessing microbes: from pretreatment to biofuel production. Bioresource Technology 161:431-440.

159. Olson DG, McBride JE, Shaw AJ, Lynd LR. 2012. Recent progress in consolidated bioprocessing. Current Opinion in Biotechnology 23:396-405.

160. Xu Q, Singh A, Himmel ME. 2009. Perspectives and new directions for the production of bioethanol using consolidated bioprocessing of lignocellulose. Current Opinion in Biotechnology 20:364-371.

161. Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. 2013. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresource Technology 135:513-522.

162. Shen CR, Lan EI, Dekishima Y, Baez A, Cho KM, Liao JC. 2011. High titer anaerobic 1-butanol synthesis in *Escherichia coli* enabled by driving forces. Applied and Environmental Microbiology doi:10.1128/aem.03034-10.

163. Inokuma K, Liao JC, Okamoto M, Hanai T. 2010. Improvement of isopropanol production by metabolically engineered *Escherichia coli* using gas stripping. Journal of Bioscience and Bioengineering 110:696-701.

164. Béguin P, Aubert J-P. 1994. The biological degradation of cellulose. FEMS Microbiology Reviews 13:25-58.

165. Juturu V, Wu JC. 2012. Microbial xylanases: engineering, production and industrial applications. Biotechnology Advances 30:1219-1227.

166. den Haan R, van Rensburg E, Rose SH, Görgens JF, van Zyl WH. 2015. Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Current Opinion in Biotechnology 33:32-38.

167. Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U. 2014. Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. Biotechnology for Biofuels 7:135.

168. Todhanakasem T, Sowatad A, Kanokratana P, Havanapan P-o, Champreda V. 2019. Expression and extracellular secretion of endo-glucanase and xylanase by *Zymomonas mobilis*. Applied Biochemistry and Biotechnology 187:239-252.

169. Yamada R, Taniguchi N, Tanaka T, Ogino C, Fukuda H, Kondo A. 2011. Direct ethanol production from cellulosic materials using a diploid strain of *Saccharomyces cerevisiae* with optimized cellulase expression. Biotechnology for Biofuels 4:8.

170. Matano Y, Hasunuma T, Kondo A. 2012. Display of cellulases on the cell surface of *Saccharomyces cerevisiae* for high yield ethanol production from high-solid lignocellulosic biomass. Bioresource Technology 108:128-133.

171. Feng C, Zou S, Liu C, Yang H, Zhang K, Ma Y, Hong J, Zhang M. 2016. Ethanol production from acid-and alkali-pretreated corncob by endoglucanase and β-glucosidase co-expressing *Saccharomyces cerevisiae* subject to the expression of heterologous genes and nutrition added. World Journal of Microbiology and Biotechnology 32:86.

172. Liu Z, Ho S-H, Sasaki K, den Haan R, Inokuma K, Ogino C, van Zyl WH, Hasunuma T, Kondo A. 2016. Engineering of a novel cellulose-adherent cellulolytic *Saccharomyces cerevisiae* for cellulosic biofuel production. Scientific Reports 6:24550.

173. Wang D, Hong J. 2018. Expression of Cellulolytic Enzymes in Yeast, p 201-221, Fungal Cellulolytic Enzymes. Springer.

174. Stülke J, Hillen W. 1999. Carbon catabolite repression in bacteria. Current Opinion in Microbiology 2:195-201.

175. Verho R, Londesborough J, Penttilä M, Richard P. 2003. Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*. Applied and Environmental Microbiology 69:5892-5897.

176. Feldmann SD, Sahm H, Sprenger GA. 1992. Pentose metabolism in *Zymomonas mobilis* wild-type and recombinant strains. Applied Microbiology and Biotechnology 38:354-361.

177. Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S. 1995. Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. Science 267:240-243.

178. Xia J, Yang Y, Liu C-G, Yang S, Bai F-W. 2019. Engineering *Zymomonas mobilis* for robust cellulosic ethanol production. Trends in Biotechnology 37:960-972.

179. Nakamura N, Yamada R, Katahira S, Tanaka T, Fukuda H, Kondo A. 2008. Effective xylose/cellobiose co-fermentation and ethanol production by xylose-assimilating *S. cerevisiae* via expression of β-glucosidase on its cell surface. Enzyme and Microbial Technology 43:233-236.

180. Kim IS, Barrow KD, Rogers PL. 2000. Kinetic and nuclear magnetic resonance studies of xylose metabolism by recombinant *Zymomonas mobilis* ZM4(pZB5). Applied and Environmental Microbiology 66:186-193.

181. Deanda K, Zhang M, Eddy C, Picataggio S. 1996. Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering. Applied and Environmental Microbiology 62:4465-4470.

182. Kim J-H, Block DE, Mills DA. 2010. Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass. Applied Microbiology and Biotechnology 88:1077-1085.

183. Bettiga M, Hahn-Hägerdal B, Gorwa-Grauslund MF. 2008. Comparing the xylose reductase/xylitol dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting *Saccharomyces cerevisiae* strains. Biotechnology for Biofuels 1:16.

184. Wilson DB. 2011. Microbial diversity of cellulose hydrolysis. Current Opinion in Microbiology 14:259-263.

185. Artzi L, Bayer EA, Moraïs S. 2017. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nature Reviews Microbiology 15:83-95.

186. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang S-J, Resch MG, Adams MW, Lunin VV. 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 342:1513-1516.

187. Ali SS, Nugent B, Mullins E, Doohan FM. 2016. Fungal-mediated consolidated bioprocessing: the potential of *Fusarium oxysporum* for the lignocellulosic ethanol industry. AMB Express 6:13.

188. Xu Q, Himmel ME, Singh A. 2015. Chapter 11 - Production of Ethanol from Engineered *Trichoderma reesei*, p 197-208. *In* Himmel ME (ed), Direct Microbial Conversion of Biomass to Advanced Biofuels doi:https://doi.org/10.1016/B978-0-444-59592-8.00011-7. Elsevier, Amsterdam.

189. Chou Y-C, Singh A, Xu Q, Himmel ME, Zhang M. 2020. Methods for metabolic engineering of a filamentous *Trichoderma reesei*, p 45-50, Metabolic Pathway Engineering. Springer.

190. Stevenson DM, Weimer PJ. 2002. Isolation and characterization of a *Trichoderma* strain capable of fermenting cellulose to ethanol. Applied Microbiology and Biotechnology 59:721-726.

191. Huang J, Chen D, Wei Y, Wang Q, Li Z, Chen Y, Huang R. 2014. Direct ethanol production from lignocellulosic sugars and sugarcane bagasse by a recombinant *Trichoderma reesei* strain HJ48. The Scientific World Journal 2014:798683.

192. Liu G, Qu Y. 2019. Engineering of filamentous fungi for efficient conversion of lignocellulose: tools, recent advances and prospects. Biotechnology Advances 37:519-529.

193. Panagiotou G, Villas-Bôas SG, Christakopoulos P, Nielsen J, Olsson L. 2005. Intracellular metabolite profiling of *Fusarium oxysporum* converting glucose to ethanol. Journal of Biotechnology 115:425-434.

194. Anasontzis GE, Kourtoglou E, Villas-Boâs SG, Hatzinikolaou DG, Christakopoulos P. 2016. Metabolic engineering of *Fusarium oxysporum* to improve Its ethanol-producing capability. Frontiers in Microbiology 7.

195. Freier D, Mothershed CP, Wiegel J. 1988. Characterization of *Clostridium thermocellum* JW20. Applied and Environmental Microbiology 54:204-211.

196. Rydzak T, Lynd LR, Guss AM. 2015. Elimination of formate production in *Clostridium thermocellum*. Journal of Industrial Microbiology and Biotechnology 42:1263-1272.

197. Xiong W, Reyes LH, Michener WE, Maness P-C, Chou KJ. 2018. Engineering cellulolytic bacterium *Clostridium thermocellum* to co-ferment cellulose- and hemicellulose-derived sugars simultaneously. Biotechnology and Bioengineering 115:1755-1763.

198. Walker JE, Lanahan AA, Zheng T, Toruno C, Lynd LR, Cameron JC, Olson DG, Eckert CA. 2020. Development of both type I–B and type II CRISPR/Cas genome editing systems in the cellulolytic bacterium *Clostridium thermocellum*. Metabolic Engineering Communications 10:e00116.

199. Holwerda EK, Thorne PG, Olson DG, Amador-Noguez D, Engle NL, Tschaplinski TJ, van Dijken JP, Lynd LR. 2014. The exometabolome of *Clostridium thermocellum* reveals overflow metabolism at high cellulose loading. Biotechnology for Biofuels 7:155.

200. Tian L, Perot SJ, Hon S, Zhou J, Liang X, Bouvier JT, Guss AM, Olson DG, Lynd LR. 2017. Enhanced ethanol formation by *Clostridium thermocellum* via pyruvate decarboxylase. Microbial Cell Factories 16:171.

201. Sato K, Tomita M, Yonemura S, Goto S, Sekine K, Okuma E, Takagi Y, Hon-Nami K, Saikit T. 1993. Characterization of and ethanol hyper-production by *Clostridium thermocellum* I-1-B. Bioscience, Biotechnology, and Biochemistry 57:2116-2121.

202. Papanek B, Biswas R, Rydzak T, Guss AM. 2015. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. Metabolic Engineering 32:49-54.

203. Kothari N, Holwerda EK, Cai CM, Kumar R, Wyman CE. 2018. Biomass augmentation through thermochemical pretreatments greatly enhances digestion of switchgrass by *Clostridium thermocellum*. Biotechnology for Biofuels 11:219.

204. Shao X, Murphy SJ, Lynd LR. 2020. Characterization of reduced carbohydrate solubilization during *Clostridium thermocellum* fermentation with high switchgrass concentrations. Biomass and Bioenergy 139:105623.

205. Lin PP, Mi L, Morioka AH, Yoshino KM, Konishi S, Xu SC, Papanek BA, Riley LA, Guss AM, Liao JC. 2015. Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. Metabolic Engineering 31:44-52.

206. Tian L, Conway PM, Cervenka ND, Cui J, Maloney M, Olson DG, Lynd LR. 2019. Metabolic engineering of *Clostridium thermocellum* for *n*-butanol production from cellulose. Biotechnology for Biofuels 12:186.

207. Gold ND, Martin VJJ. 2007. Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. Journal of Bacteriology 189:6787-6795.

208. Suurnäkki A, Tenkanen M, Siika-aho M, Niku-Paavola ML, Viikari L, Buchert J. 2000. *Trichoderma reesei* cellulases and their core domains in the hydrolysis and modification of chemical pulp. Cellulose 7:189-209.

209. Mayer F. 2019. The bacterial cellulosome-mini-review. Laboratory Medicine 1:102-105.

210. Mohand-Oussaid O, Payot S, Guedon E, Gelhaye E, Youyou A, Petitdemange H. 1999. The extracellular xylan degradative system in Clostridium cellulolyticum cultivated on xylan: evidence for cell-free cellulosome production. Journal of Bacteriology 181:4035-4040.

211. Moraïs S, Barak Y, Caspi J, Hadar Y, Lamed R, Shoham Y, Wilson DB, Bayer EA. 2010. Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. mBio 1:e00285-10.

212. Oh EJ, Jin Y-S. 2020. Engineering of *Saccharomyces cerevisiae* for efficient fermentation of cellulose. FEMS Yeast Research 20.

213. Tsai S-L, DaSilva NA, Chen W. 2013. Functional display of complex cellulosomes on the yeast surface via adaptive assembly. ACS Synthetic Biology 2:14-21.

214. Beri D, York WS, Lynd LR, Peña MJ, Herring CD. 2020. Development of a thermophilic coculture for corn fiber conversion to ethanol. Nature Communications 11:1937.

215. Wilkinson S, Smart KA, James S, Cook DJ. 2017. Bioethanol production from brewers spent grains using a fungal consolidated bioprocessing (CBP) approach. Bioenergy Research 10:146-157.

216. Oil and petroleum products explained. https://www.eia.gov/energyexplained/oil-and-petroleum-products/. Accessed 9/5/20.

217. Sawant SS, Salunke BK, Tran TK, Kim BS. 2016. Lignocellulosic and marine biomass as resource for production of polyhydroxyalkanoates. Korean Journal of Chemical Engineering 33:1505-1513.

218. Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, Keasling JD. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463:559-562.

219. Bokinsky G, Peralta-Yahya PP, George A, Holmes BM, Steen EJ, Dietrich J, Soon Lee T, Tullman-Ercek D, Voigt CA, Simmons BA, Keasling JD. 2011. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. Proceedings of the National Academy of Sciences 108:19949-19954.

220. Zeldes BM, Keller MW, Loder AJ, Straub CT, Adams MWW, Kelly RM. 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Frontiers in Microbiology 6.

221. Counts JA, Zeldes BM, Lee LL, Straub CT, Adams MWW, Kelly RM. 2017. Physiological, metabolic and biotechnological features of extremely thermophilic microorganisms. WIREs Systems Biology and Medicine 9:e1377.

222. Koga Y. 2012. Thermal adaptation of the archaeal and bacterial lipid membranes. Archaea 2012.

223. Razvi A, Scholtz JM. 2006. Lessons in stability from thermophilic proteins. Protein Science 15:1569-1578.

224. L. Bergquist P, W. Morgan H, Saul D. 2014. Selected enzymes from extreme thermophiles with applications in biotechnology. Current Biotechnology 3:45-59.

225. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-91.

226. Bruins ME, Janssen AEM, Boom RM. 2001. Thermozymes and their applications. Applied Biochemistry and Biotechnology 90:155.

227. Atalah J, Cáceres-Moreno P, Espina G, Blamey JM. 2019. Thermophiles and the applications of their enzymes as new biocatalysts. Bioresource Technology 280:478-488.

228. Kumar S, Tsai C-J, Nussinov R. 2000. Factors enhancing protein thermostability. Protein Engineering, Design and Selection 13:179-191.

229. Yeoman CJ, Han Y, Dodd D, Schroeder CM, Mackie RI, Cann IKO. 2010. Chapter 1 - Thermostable Enzymes as Biocatalysts in the Biofuel Industry, p 1-55, Advances in Applied Microbiology, vol 70. Academic Press.

230. Van der Linden MG, de Farias ST. 2006. Correlation between codon usage and thermostability. Extremophiles 10:479-481.

231. Mozo-Villiarías A, Querol E. 2006. Theoretical analysis and computational predictions of protein thermostability. Current Bioinformatics 1:25-32.

232. Frock AD, Kelly RM. 2012. Extreme thermophiles: moving beyond single-enzyme biocatalysis. Current Opinion in Chemical Engineering 1:363-372.

233. Marcó MB, Moineau S, Quiberoni A. 2012. Bacteriophages and dairy fermentations. Bacteriophage 2:149-158.

234. Vane LM. 2008. Separation technologies for the recovery and dehydration of alcohols from fermentation broths. Biofuels, Bioproducts and Biorefining 2:553-588.

235. Curran JS, Smith J, Holms W. 1989. Heat-and-power in industrial fermentation processes. Applied Energy 34:9-20.

236. Keller M, Loder A, Basen M, Izquierdo J, Kelly RM, Adams MWW. 2014. Production of lignofuels and electrofuels by extremely thermophilic microbes. Biofuels 5:499-515.

237. Lemos LN, Pereira RV, Quaggio RB, Martins LF, Moura LMS, da Silva AR, Antunes LP, da Silva AM, Setubal JC. 2017. Genome-centric analysis of a thermophilic and cellulolytic bacterial consortium derived from composting. Frontiers in Microbiology 8.

238. Xia Y, Ju F, Fang HH, Zhang T. 2013. Mining of novel thermo-stable cellulolytic genes from a thermophilic cellulose-degrading consortium by metagenomics. PloS ONE 8:e53779.

239. Olson DG, Sparling R, Lynd LR. 2015. Ethanol production by engineered thermophiles. Current Opinion in Biotechnology 33:130-141.

240. Crosby JR, Laemthong T, Lewis AM, Straub CT, Adams MWW, Kelly RM. 2019. Extreme thermophiles as emerging metabolic engineering platforms. Current Opinion in Biotechnology 59:55-64.

241. Peteranderl R, Shotts EB, Wiegel J. 1990. Stability of antibiotics under growth conditions for thermophilic anaerobes. Applied and Environmental Microbiology 56:1981-1983.

242. Lipscomb GL, Conway JM, Blumer-Schuette SE, Kelly RM, Adams MWW. 2016. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. Applied and Environmental Microbiology 82:4421-4428.

243. Finegold SM. 1959. Kanamycin. AMA Archives of Internal Medicine 104:15-28.

244. Harrington LB, Paez-Espino D, Staahl BT, Chen JS, Ma E, Kyrpides NC, Doudna JA. 2017. A thermostable Cas9 with increased lifetime in human plasma. Nature Communications 8:1424.

245. Liu T, Liu Z, Ye Q, Pan S, Wang X, Li Y, Peng W, Liang Y, She Q, Peng N. 2017. Coupling transcriptional activation of CRISPR–Cas system and DNA repair genes by Csa3a in *Sulfolobus islandicus*. Nucleic Acids Research 45:8978-8992.

246. Lee LL, Crosby JR, Rubinstein GM, Laemthong T, Bing RG, Straub CT, Adams MWW, Kelly RM. 2020. The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi World'. Extremophiles 24:1-15.

247. Lee LL, Blumer-Schuette SE, Izquierdo JA, Zurawski JV, Loder AJ, Conway JM, Elkins JG, Podar M, Clum A, Jones PC, Piatek MJ, Weighill DA, Jacobson DA, Adams MWW, Kelly RM. 2018. Genus-wide assessment of lignocellulose utilization in the extremely thermophilic genus *Caldicellulosiruptor* by genomic, pangenomic, and metagenomic analyses. Applied and Environmental Microbiology 84:e02694-17.

248. Vishnivetskaya TA, Hamilton-Brehm SD, Podar M, Mosher JJ, Palumbo AV, Phelps TJ, Keller M, Elkins JG. 2015. Community analysis of plant biomass-degrading microorganisms from Obsidian Pool, Yellowstone National Park. Microbial Ecology 69:333-345.

249. Sissons CH, Sharrock KR, Daniel RM, Morgan HW. 1987. Isolation of cellulolytic anaerobic extreme thermophiles from New Zealand thermal sites. Applied and Environmental Microbiology 53:832-838.

250. Rainey FA, Donnison AM, Janssen PH, Saul D, Rodrigo A, Bergquist PL, Daniel RM, Stackebrandt E, Morgan HW. 1994. Description of *Caldicellulosiruptor saccharolyticus*

gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. FEMS Microbiology Letters 120:263-266.

251. van de Werken HJ, Verhaart MR, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE, van Niel EW, Stams AJ, Ward DE, de Vos WM, van der Oost J, Kelly RM, Kengen SW. 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Applied and Environmental Microbiology 74:6720-9.

252. Dam P, Kataeva I, Yang S-J, Zhou F, Yin Y, Chou W, Poole FL, Westpheling J, Hettich R, Giannone R, Lewis DL, Kelly R, Gilbert HJ, Henrissat B, Xu Y, Adams MWW. 2011. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. Nucleic Acids Research 39:3240-3254.

253. Carere CR, Rydzak T, Verbeke TJ, Cicek N, Levin DB, Sparling R. 2012. Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H_2 and ethanol-producing bacteria. BMC Microbiology 12.

254. Svetlichny V, Svetlichnaya T, Chernykh N, Zavarzin G. 1990. *Anaerocellum thermophilum* gen. nov. sp. nov, an extreme thermophilic celluloselytic eubacterium isolated from hot springs in the Valley of Geysers. Microbiology (Moscow) 59:871-879.

255. Yang S-J, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MW. 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe *Anaerocellum thermophilum* DSM 6725. Applied and Environmental Microbiology 75:4762-9.

256. Chung D, Farkas J, Huddleston JR, Olivar E, Westpheling J. 2012. Methylation by a unique α-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. PLoS ONE 7:e43844.

257. Basen M, Rhaesa AM, Kataeva I, Prybol CJ, Scott IM, Poole FL, Adams MWW. 2014. Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium *Caldicellulosiruptor bescii*. Bioresource Technology 152:384-392.

258. Foyle T, Jennings L, Mulcahy P. 2007. Compositional analysis of lignocellulosic materials: evaluation of methods used for sugar analysis of waste paper and straw. Bioresource Technology 98:3026-3036.

259. Conway J, McKinley B, Seals N, Hernandez D, Khatibi P, Poudel S, Giannone R, Hettich R, Williams-Rhaesa A, Lipscomb G, Adams M, Kelly R. 2017. Functional analysis of the Glucan Degradation Locus (GDL) in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. Applied and Environmental Microbiology doi:10.1128/aem.01828-17.

260. Zurawski JV, Conway JM, Lee LL, Simpson HJ, Izquierdo JA, Blumer-Schuette S, Nookaew I, Adams MWW, Kelly RM. 2015. Comparative analysis of extremely thermophilic *Caldicellulosiruptor* species reveals common and unique cellular strategies for plant biomass utilization. Applied and Environmental Microbiology 81:7159-7170.

261. Conway JM, Crosby JR, McKinley BS, Seals NL, Adams MWW, Kelly RM. 2018. Parsing i*n vivo* and *in vitro* contributions to microcrystalline cellulose hydrolysis by multidomain glycoside hydrolases in the *Caldicellulosiruptor bescii* secretome. Biotechnology and Bioengineering 115:2426-2440.

262. Conway J, Pierce W, Le J, Harper G, Wright J, Tucker A, Zurawski J, Lee L, Blumer-Schuette S, Kelly R. 2016. Multi-Domain, surface layer associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. Journal of Biological Chemistry 291:6732-6747.

263. Blumer-Schuette SE, Alahuhta M, Conway JM, Lee LL, Zurawski JV, Giannone RJ, Hettich RL, Lunin VV, Himmel ME, Kelly RM. 2015. Discrete and structurally unique proteins (tāpirins) mediate attachment of extremely thermophilic *Caldicellulosiruptor* species to cellulose. Journal of Biological Chemistry 290:10645-10656.

264. Biswas R, Prabhu S, Lynd LR, Guss AM. 2014. Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of *Clostridium thermocellum*. *PLoS ONE* 9:e86389.

265. Lee LL, Hart WS, Lunin VV, Alahuhta M, Bomble YJ, Himmel ME, Blumer-Schuette SE, Adams MWW, Kelly RM. 2019. Comparative biochemical and structural analysis of novel cellulose binding proteins (tāpirins) from extremely thermophilic *Caldicellulosiruptor* species. Applied and Environmental Microbiology 85:e01983-18.

266. de Vrije T, Mars AE, Budde MA, Lai MH, Dijkema C, de Waard P, Claassen PA. 2007. Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Applied Microbiology and Biotechnology 74:1358-67.

267. Williams-Rhaesa AM, Awuku NK, Lipscomb GL, Poole FL, Rubinstein GM, Conway JM, Kelly RM, Adams MWW. 2018. Native xylose-inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic bacterium Caldicellulosiruptor bescii. Extremophiles 22:629-638.

268. Yang SJ, Kataeva I, Wiegel J, Yin Y, Dam P, Xu Y, Westpheling J, Adams MW. 2010. Classification of '*Anaerocellum thermophilum*' strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. International Journal of Systematic and Evolutionary Microbiology 60:2011-5.

269. Ingvadottir EM, Scully SM, Orlygsson J. 2017. Evaluation of the genus of *Caldicellulosiruptor* for production of 1,2-propanediol from methylpentoses. Anaerobe 47:86-88.

270. Williams-Rhaesa AM, Rubinstein GM, Scott IM, Lipscomb GL, Poole Ii FL, Kelly RM, Adams MWW. 2018. Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*. Metabolic Engineering Communications 7:e00073.

271. Cha M, Chung D, Westpheling J. 2016. Deletion of a gene cluster for [Ni-Fe] hydrogenase maturation in the anaerobic hyperthermophilic bacterium *Caldicellulosiruptor bescii* identifies its role in hydrogen metabolism. Applied Microbiology and Biotechnology 100:1823-31.

272. Scott I, Rubinstein G, Lipscomb G, Basen M, Schut G, Rhaesa A, Lancaster W, Poole F, 2nd, Kelly R, Adams M. 2015. A new class of tungsten-containing oxidoreductase in *Caldicellulosiruptor*, a genus of plant biomass-degrading thermophilic bacteria. Applied and Environmental Microbiology 81:7339-47.

273. Scott IM, G.M. Rubinstein, F.L. Poole, G.L. Lipscomb, G.J. Schut, A.M. Williams-Rhaesa, D.M. Stevenson, D. Amador-Noguez, R.M Kelly, and M.W. W. Adams. 2019. The thermophilic biomass-degrading bacterium *Caldicellulosiruptor bescii* utilizes two

enzymes to oxidize glyceraldehyde-3-phosphate during glycolysis. Journal of Biological Chemistry jbc-RA118.

274. Sander K, Chung D, Hyatt D, Westpheling J, Klingeman DM, Rodriguez Jr M, Engle NL, Tschaplinski TJ, Davison BH, Brown SD. 2019. Rex in *Caldicellulosiruptor bescii*: novel regulon members and its effect on the production of ethanol and overflow metabolites. MicrobiologyOpen 8:e00639.

275. Chung D, Cha M, Farkas J, Westpheling J. 2013. Construction of a stable replicating shuttle vector for *Caldicellulosiruptor* species: use for extending genetic methodologies to other members of this genus. PLoS ONE 8:e62881.

276. 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 6:1.

277. Groom J, Chung D, Young J, Westpheling J. 2014. Heterologous complementation of a pyrF deletion in *Caldicellulosiruptor hydrothermalis* generates a new host for the analysis of biomass deconstruction. Biotechnology for Biofuels 7:132.

278. Cha M, Wang H, Chung D, Bennetzen J, Westpheling J. 2013. Isolation and bioinformatic analysis of a novel transposable element, IS*Cbe4*, from the hyperthermophilic bacterium, *Caldicellulosiruptor bescii*. Journal of Industrial Microbiology and Biotechnology 40:1443-8.

279. Williams-Rhaesa A, Poole F, Dinsmore J, Lipscomb G, Rubinstein G, Scott I, Conway J, Lee L, Khatibi P, Kelly R, Adams M. 2017. Genome stability in engineered strains of the extremely thermophilic, lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. Applied and Environmental Microbiology doi:10.1128/aem.00444-17.

280. Kataeva I, Yang S, Dam P, Poole F, 2nd, Yin Y, Zhou F, Chou W, Xu Y, Goodwin L, Sims D, Detter J, Hauser L, Westpheling J, Adams M. 2009. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. Journal of Bacteriology 191:3760-1.

281. Chung D, Pattathil S, Biswal A, Hahn M, Mohnen D, Westpheling J. 2014. Deletion of a gene cluster encoding pectin degrading enzymes in *Caldicellulosiruptor bescii* reveals an important role for pectin in plant biomass recalcitrance. Biotechnology for Biofuels 7:147.

282. Young J, Chung D, Bomble Y, Himmel M, Westpheling J. 2014. Deletion of *Caldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. Biotechnology for Biofuels 7:142.

283. Chung D, Young J, Cha M, Brunecky R, Bomble Y, Himmel M, Westpheling J. 2015. Expression of the *Acidothermus cellulolyticus* E1 endoglucanase in *Caldicellulosiruptor bescii* enhances its ability to deconstruct crystalline cellulose. Biotechnology for Biofuels 8:113.

284. Kim S, Chung D, Himmel M, Bomble Y, Westpheling J. 2016. Heterologous expression of family 10 xylanases from *Acidothermus cellulolyticus* enhances the exoproteome of *Caldicellulosiruptor bescii* and growth on xylan substrates. Biotechnology for Biofuels 9:176.

285. Kim S, Chung D, Himmel M, Bomble Y, Westpheling J. 2017. Heterologous expression of a beta-D-glucosidase in *Caldicellulosiruptor bescii* has a surprisingly modest effect on the activity of the exoproteome and growth on crystalline cellulose. Journal of Industrial Microbiology and Biotechnology 44:1643-1651.

286. Kim S, Chung D, Himmel M, Bomble Y, Westpheling J. 2017. Engineering the N-terminal end of CelA results in improved performance and growth of *Caldicellulosiruptor bescii* on crystalline cellulose. Biotechnology and Bioengineering 114:945-950.

287. Kim S, Himmel M, Bomble Y, Westpheling J. 2017. Expression of a cellobiose phosphorylase from *Thermotoga maritima* in *Caldicellulosiruptor bescii* improves the phosphorolytic pathway and results in a dramatic increase in cellulolytic activity. Applied and Environmental Microbiology doi:10.1128/aem.02348-17.

288. Kim SK, Chung D, Himmel ME, Bomble YJ, Westpheling J. 2019. Heterologous co-expression of two beta-glucanases and a cellobiose phosphorylase resulted in a significant increase in the cellulolytic activity of the *Caldicellulosiruptor bescii* exoproteome. Journal of Industrial Microbiology and Biotechnology doi:10.1007/s10295-019-02150-0.

289. Russell J, Kim SK, Duma J, Nothaft H, Himmel ME, Bomble YJ, Szymanski CM, Westpheling J. 2018. Deletion of a single glycosyltransferase in *Caldicellulosiruptor bescii* eliminates protein glycosylation and growth on crystalline cellulose. Biotechnology for Biofuels 11:259.

290. Alahuhta M, Chandrayan P, Kataeva I, Adams M, Himmel M, Lunin V. 2011. A 1.5 A resolution X-ray structure of the catalytic module of *Caldicellulosiruptor bescii* family 3 pectate lyase. Acta crystallographica Section F, Structural biology and crystallization communications 67:1498-500.

291. Cha M, Chung D, Elkins J, Guss A, Westpheling J. 2013. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. Biotechnology for Biofuels 6:85.

292. Chung D, Cha M, Guss A, Westpheling J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. Proceedings of the National Academy of Sciences 111:8931-6.

293. Chung D, Cha M, Snyder E, Elkins J, Guss A, Westpheling J. 2015. Cellulosic ethanol production via consolidated bioprocessing at 75 °C by engineered *Caldicellulosiruptor bescii*. Biotechnology for Biofuels 8:163-163.

294. Chung D, Verbeke TJ, Cross KL, Westpheling J, Elkins JG. 2015. Expression of a heat-stable NADPH-dependent alcohol dehydrogenase in *Caldicellulosiruptor bescii* results in furan aldehyde detoxification. Biotechnology for Biofuels 8:102.

295. Lynd L, Weimer P, van Zyl W, Pretorius I. 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and Molecular Biology Reviews 66:506-577.

296. Basen M, Schut GJ, Nguyen DM, Lipscomb GL, Benn RA, Prybol CJ, Vaccaro BJ, Poole FL, Kelly RM, Adams MW. 2014. Single gene insertion drives bioalcohol production by a thermophilic archaeon. Proceedings of the National Academy of Sciences 111:17618-17623.

297. Simon H, White H, Lebertz H, Thanos I. 1987. Reduction of 2-enoates and alkanoates with carbon monoxide or formate, viologens, and *Clostridium thermoaceticum* to saturated acids and unsaturated and saturated alcohols. Angewandte Chemie (International Edition in English) 26:785-787.

298. Isom CE, Nanny MA, Tanner RS. 2015. Improved conversion efficiencies for *n*-fatty acid reduction to primary alcohols by the solventogenic acetogen *Clostridium ragsdalei*. Journal of Industrial Microbiology and Biotechnology 42:29-38.

299. Mock J, Zheng Y, Mueller AP, Ly S, Tran L, Segovia S, Nagaraju S, Kopke M, Durre P, Thauer RK. 2015. Energy conservation associated with ethanol formation from H_2 and CO_2 in *Clostridium autoethanogenum* involving electron bifurcation. Journal of Bacteriology 197:2965-80.

300. Perez JM, Richter H, Loftus SE, Angenent LT. 2013. Biocatalytic reduction of short-chain carboxylic acids into their corresponding alcohols with syngas fermentation. Biotechnology and Bioengineering 110:1066-1077.

301. Ammam F, Tremblay P-L, Lizak DM, Zhang T. 2016. Effect of tungstate on acetate and ethanol production by the electrosynthetic bacterium *Sporomusa ovata*. Biotechnology for Biofuels 9:163.

302. Scully SM, Orlygsson J. 2020. Branched-chain amino acid catabolism of *Thermoanaerobacter pseudoethanolicus* reveals potential route to branched-chain alcohol formation. Extremophiles 24:121-133.

303. Hitschler L, Kuntz M, Langschied F, Basen M. 2018. *Thermoanaerobacter* species differ in their potential to reduce organic acids to their corresponding alcohols. Applied Microbiology and Biotechnology 102:8465-8476.

304. Scully SM, Orlygsson J. 2020. Biotransformation of carboxylic acids to alcohols: characterization of *Thermoanaerobacter* strain AK152 and 1-propanol production via propionate reduction. Microorganisms 8:945.

305. Sun X, Atiyeh HK, Kumar A, Zhang H. 2018. Enhanced ethanol production by *Clostridium ragsdalei* from syngas by incorporating biochar in the fermentation medium. Bioresource Technology 247:291-301.

306. Mitchell WJ, Albasheri KA, Yazdanian M. 1995. Factors affecting utilization of carbohydrates by clostridia. FEMS Microbiology Reviews 17:317-329.

307. Abrini J, Naveau H, Nyns E-J. 1994. *Clostridium autoethanogenum*, sp. nov., an anaerobic bacterium that produces ethanol from carbon monoxide. Archives of Microbiology 161:345-351.

308. Tanner RS, Miller LM, Yang D. 1993. *Clostridium ljungdahlii* sp. nov., an acetogenic species in clostridial rRNA homology group I. International Journal of Systematic and Evolutionary Microbiology 43:232-236.

309. Huhnke RL, Lewis RS, Tanner RS. 2010. Isolation and characterization of novel clostridial species. Google Patents.

310. White H, Strobl G, Feicht R, Simon H. 1989. Carboxylic acid reductase: a new tungsten enzyme catalyses the reduction of non-activated carboxylic acids to aldehydes. European Journal of Biochemistry 184:89-96.

311. Mukund S, Adams MW. 1991. The novel tungsten-iron-sulfur protein of the hyperthermophilic archaebacterium, *Pyrococcus furiosus*, is an aldehyde ferredoxin oxidoreductase. Evidence for its participation in a unique glycolytic pathway. Journal of Biological Chemistry 266:14208-16.

312. STROBL G, FEICHT R, WHITE H, LOTTSPEICH F, SIMON H. 1992. The tungsten-containing aldehyde oxidoreductase from *Clostridium thermoaceticum* and its complex with a viologen-accepting NADPH oxidoreductase. Biological Chemistry 373:123.

313. Seelmann CS, Willistein M, Heider J, Boll M. 2020. Tungstoenzymes: occurrence, catalytic diversity and cofactor synthesis. Inorganics 8:44.

314. Li T, Rosazza JP. 1997. Purification, characterization, and properties of an aryl aldehyde oxidoreductase from *Nocardia* sp. strain NRRL 5646. Journal of Bacteriology 179:3482-3487.

315. Gross GG. 1969. Evidence for enzyme-substrate intermediates in the aryl-aldehyde: NADP oxidoreductase catalysed reduction of salicylate. FEBS Letters 5:177-179.

316. Gross GG. 1972. Formation and reduction of intermediate acyladenylate by aryl-aldehyde. European Journal of Biochemistry 31:585-592.

317. He A, Li T, Daniels L, Fotheringham I, Rosazza JPN. 2004. *Nocardia* sp. carboxylic acid reductase: cloning, expression, and characterization of a new aldehyde oxidoreductase family. Applied and Environmental Microbiology 70:1874-1881.

318. Winkler M. 2018. Carboxylic acid reductase enzymes (CARs). Current Opinion in Chemical Biology 43:23-29.

319. Napora-Wijata K, Robins K, Osorio-Lozada A, Winkler M. 2014. Whole-cell carboxylate reduction for the synthesis of 3-hydroxytyrosol. ChemCatChem 6:1089-1095.

320. Vaclav S. 2017. Energy transitions: global and national perspectives.

321. Ritchie H. Energy, *on* Our World in Data. https://ourworldindata.org/energy. Accessed 10/20/2020.

322. Straub CT, Zeldes BM, Schut GJ, Adams MWW, Kelly RM. 2017. Extremely thermophilic energy metabolisms: biotechnological prospects. Current Opinion in Biotechnology 45:104-112.

323. Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MWW, Kelly RM. 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiology Reviews 38:393-448.

324. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL, Adams MW. 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. Journal of Bacteriology 194:4015-4028.

325. Kataeva I, Foston MB, Yang SJ, Pattathil S, Biswal A, Poole FL, Basen M, Rhaesa AM, Thomas TP, Azadi P, Olman V, Saffold TD, Mohler KE, Lewis DL, Doeppke C, Zeng Y, Tschaplinski T, York WS, Davis M, Mohnen D, Xu Y, Ragauskas AJ, Ding SY, Kelly RM, Hahn MG, Adams MWW. 2013. Carbohydrate and lignin are simultaneously solubilized from unpretreated switchgrass by microbial action at high temperature. Energy and Environmental Science 6:2186-2195.

326. Groom J, Chung D, Kim S-K, Guss A, Westpheling J. 2018. Deletion of the *Clostridium thermocellum* recA gene reveals that it is required for thermophilic plasmid replication but not plasmid integration at homologous DNA sequences. Journal of Industrial Microbiology and Biotechnology 45:753-763.

327. Kletzin A, Adams MW. 1996. Tungsten in biological systems. FEMS Microbiology Reviews 18:5-63.

328. Iobbi-Nivol C, Leimkuhler S. 2013. Molybdenum enzymes, their maturation and molybdenum cofactor biosynthesis in *Escherichia coli*. Biochimica et Biophysica Acta Bioenergetics 1827:1086-101.

329. Bevers LE, Hagedoorn P-L, Hagen WR. 2009. The bioinorganic chemistry of tungsten. Coordination Chemistry Reviews 253:269-290.

330. Havarushka N, Fischer-Schrader K, Lamkemeyer T, Schwarz G. 2014. Structural basis of thermal stability of the tungsten cofactor synthesis protein MoaB from *Pyrococcus furiosus*. PloS ONE 9:e86030.

331. Hille R, Hall J, Basu P. 2014. The mononuclear molybdenum enzymes. Chemical Reviews 114:3963-4038.

332. Mukund S, Adams MWW. 1995. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. Journal of Biological Chemistry 270:8389-8392.

333. Roy R, Mukund S, Schut GJ, Dunn DM, Weiss R, Adams MW. 1999. Purification and molecular characterization of the tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*: the third of a putative five-member tungstoenzyme family. Journal of Bacteriology 181:1171-1180.

334. Roy R, Adams MW. 2002. Characterization of a fourth tungsten-containing enzyme from the hyperthermophilic archaeon *Pyrococcus furiosus*. Journal of Bacteriology 184:6952-6956.

335. Bevers LE, Bol E, Hagedoorn P-L, Hagen WR. 2005. WOR5, a novel tungsten-containing aldehyde oxidoreductase from *Pyrococcus furiosus* with a broad substrate specificity. Journal of Bacteriology 187:7056-7061.

336. Reher M, Gebhard S, Schönheit P. 2007. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase (GAPOR) and nonphosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN), key enzymes of the respective modified Embden–Meyerhof

pathways in the hyperthermophilic crenarchaeota *Pyrobaculum aerophilum* and *Aeropyrum pernix*. FEMS Microbiology Letters 273:196-205.

337. Park M-O, Mizutani T, Jones PR. 2007. Glyceraldehyde-3-phosphate ferredoxin oxidoreductase from *Methanococcus maripaludis*. Journal of Bacteriology 189:7281-7289.

338. Rozovsky S, McDermott AE. 2007. Substrate product equilibrium on a reversible enzyme, triosephosphate isomerase. Proceedings of the National Academy of Sciences 104:2080-2085.

339. Chan MK, Mukund S, Kletzin A, Adams M, Rees DC. 1995. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. Science 267:1463-1469.

340. Baker LM, Raudonikiene A, Hoffman PS, Poole LB. 2001. Essential thioredoxin-dependent peroxiredoxin system from *Helicobacter pylori*: genetic and kinetic characterization. Journal of Bacteriology 183:1961-1973.

341. Bell RA, Smith JC, Storey KB. 2014. Purification and properties of glyceraldehyde-3-phosphate dehydrogenase from the skeletal muscle of the hibernating ground squirrel, *Ictidomys tridecemlineatus*. PeerJ 2:e634.

342. Seidler NW. 2012. GAPDH: biological properties and diversity, vol 985. Springer Science & Business Media.

343. Engeland K, Höög J, Holmquist B, Estonius M, Jörnvall H, Vallee BL. 1993. Mutation of Arg-115 of human class III alcohol dehydrogenase: a binding site required for formaldehyde dehydrogenase activity and fatty acid activation. Proceedings of the National Academy of Sciences 90:2491-2494.

344. Ahmed H, Tjaden B, Hensel R, Siebers B. 2004. Embden–Meyerhof–Parnas and Entner–Doudoroff pathways in *Thermoproteus tenax*: metabolic parallelism or specific adaptation? Biochemical Society Transactions 32:303-304.

345. van der Oost J, Schut G, Kengen SM, Hagen WR, Thomm M, de Vos WM. 1998. The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the hyperthermophilic archaeon *Pyrococcus furiosus* represents a novel site of glycolytic regulation. Journal of Biological Chemistry 273:28149-28154.

346. Schäfer T, Schönheit P. 1993. Gluconeogenesis from pyruvate in the hyperthermophilic archaeon *Pyrococcus furiosus*: involvement of reactions of the Embden-Meyerhof pathway. Archives of Microbiology 159:354-363.

347. Selig M, Xavier KB, Santos H, Schönheit P. 1997. Comparative analysis of Embden-Meyerhof and Entner-Doudoroff glycolytic pathways in hyperthermophilic archaea and the bacterium *Thermotoga*. Archives of Microbiology 167:217-232.

348. Siebers B, Hensel R. 1993. Glucose catabolism of the hyperthermophilic archaeum *Thermoproteus tenax*. FEMS Microbiology Letters 111:1-7.

349. Siebers B, Tjaden B, Michalke K, Dörr C, Ahmed H, Zaparty M, Gordon P, Sensen CW, Zibat A, Klenk H-P. 2004. Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. Journal of Bacteriology 186:2179-2194.

350. Bräsen C, Esser D, Rauch B, Siebers B. 2014. Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. Microbiology and Molecular Biology Reviews 78:89-175.

351. Wrba A, Schweiger A, Schultes V, Jaenicke R, Zavodszky P. 1990. Extremely thermostable D-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium *Thermotoga maritima*. Biochemistry 29:7584-7592.

352. Schut GJ, Adams MWW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. Journal of Bacteriology 191:4451-4457.

353. Bielen AA, Verhaart MR, VanFossen AL, Blumer-Schuette SE, Stams AJ, van der Oost J, Kelly RM, Kengen SW. 2013. A thermophile under pressure: transcriptional analysis of the response of *Caldicellulosiruptor saccharolyticus* to different H_2 partial pressures. International Journal of Hydrogen Energy 38:1837-1849.

354. McLaughlin KJ, Strain-Damerell CM, Xie K, Brekasis D, Soares AS, Paget MS, Kielkopf CL. 2010. Structural basis for NADH/NAD+ redox sensing by a Rex family repressor. Molecular Cell 38:563-575.

355. Hagedoorn PL, Chen T, Schröder I, Piersma SR, De Vries S, Hagen WR. 2005. Purification and characterization of the tungsten enzyme aldehyde: ferredoxin oxidoreductase from the hyperthermophilic denitrifier *Pyrobaculum aerophilum*. Journal of Biological Inorganic Chemistry 10:259-269.

356. Mukund S, Adams M. 1993. Characterization of a novel tungsten-containing formaldehyde ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Thermococcus litoralis*. A role for tungsten in peptide catabolism. Journal of Biological Chemistry 268:13592-13600.

357. Roy R, Adams MWW. 2002. Tungsten-dependent aldehyde oxidoreductase: a new family of enzymes containing the pterin cofactor. Metal Ions in Biological Systems 39:673-697.

358. Oberto J. 2013. SyntTax: a web server linking synteny to prokaryotic taxonomy. BMC Bioinformatics 14:4.

359. Heider J, Arndt F, Schmitt G, Winiarska A, Saft M, Seubert A, Kahnt J. 2019. Characterisation of an aldehyde oxidoreductase from the mesophilic bacterium *Aromatoleum aromaticum* EbN1, a member of a new subfamily of tungsten-containing enzymes. Frontiers in Microbiology 10:71.

360. Weinert T, Huwiler SG, Kung JW, Weidenweber S, Hellwig P, Stärk H-J, Biskup T, Weber S, Cotelesage JJ, George GN. 2015. Structural basis of enzymatic benzene ring reduction. Nature Chemical Biology 11:586.

361. Reschke S, Duffus BR, Schrapers P, Mebs S, Teutloff C, Dau H, Haumann M, Leimkühler S. 2019. Identification of YdhV as first molybdoenzyme binding a bis-Mo-MPT cofactor in *Escherichia coli*. Biochemistry doi:10.1021/acs.biochem.9b00078.

362. Rydzak T, Garcia D, Stevenson DM, Sladek M, Klingeman DM, Holwerda EK, Amador-Noguez D, Brown SD, Guss AM. 2017. Deletion of type I glutamine synthetase deregulates nitrogen metabolism and increases ethanol production in *Clostridium thermocellum*. Metabolic Engineering 41:182-191.

363. Vaccaro BJ, Menon AL, Lancaster WA, Adams MW. 2012. Metallomics using inductively coupled plasma mass spectrometry. Current Protocols in Chemical Biology 4:249-274.

364. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR. 2016. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research 45:D200-D203.

365. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7:539.

366. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-1973.

367. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32:268-274.

368. Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587.

369. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Research 44:W242-W245.

370. Consortium U. 2016. UniProt: the universal protein knowledgebase. Nucleic Acids Research 45:D158-D169.

371. Greenwell HC, Lloyd-Evans M, Wenner C. 2013. Biofuels, science and society. Interface Focus 3:1-4.

372. Roddy DJ. 2013. Biomass in a petrochemical world. Interface Focus 3:20120038.

373. Dale BE. 2015. A new industry has been launched: the cellulosic biofuels ship (finally) sails. Biofuels, Bioproducts and Biorefining 9:1-3.

374. Agbor VB, Cicek N, Sparling R, Berlin A, Levin DB. 2011. Biomass pretreatment: fundamentals toward application. Biotechnology Advances 29:675-685.

375. Zeldes BM, Straub CT, Otten JK, Adams MWW, Kelly RM. 2018. A synthetic enzymatic pathway for extremely thermophilic acetone production based on the unexpectedly thermostable acetoacetate decarboxylase from *Clostridium acetobutylicum*. Biotechnology and Bioengineering 115:2951-2961.

376. Nissen LS, Basen M. 2019. The emerging role of aldehyde:ferredoxin oxidoreductases in microbially-catalyzed alcohol production. Journal of Biotechnology 306:105-117.

377. Cheng C, Li W, Lin M, Yang S-T. 2019. Metabolic engineering of *Clostridium carboxidivorans* for enhanced ethanol and butanol production from syngas and glucose. Bioresource Technology 284:415-423.

378. Aono S, Bryant FO, Adams MW. 1989. A novel and remarkably thermostable ferredoxin from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Journal of Bacteriology 171:3433-3439.

379. Dias AA, Pinto PA, Fraga I, Bezerra RMF. 2014. Diagnosis of enzyme inhibition using excel solver: a combined dry and wet laboratory exercise. Journal of Chemical Education 91:1017-1021.

380. Zhang F, Rodriguez S, Keasling JD. 2011. Metabolic engineering of microbial pathways for advanced biofuels production. Current Opinion in Biotechnology 22:775-783.

381. Mann J, Yao N, Bocarsly AB. 2006. Characterization and analysis of new catalysts for a direct ethanol fuel cell. *Langmuir* 22:10432-10436.

382. Lu J, Brigham CJ, Gai CS, Sinskey AJ. 2012. Studies on the production of branched-chain alcohols in engineered *Ralstonia eutropha*. Applied Microbiology and Biotechnology 96:283-297.

383. Chockalingam SR, Saravanan G. 2011. Influence of hexanol-diesel blends on constant speed diesel engine. Thermal Science 15:1215-1222.

384. Atsumi S, Hanai T, Liao JC. 2008. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451:86-89.

385. Baez A, Cho K-M, Liao JC. 2011. High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with in situ product removal. Applied Microbiology and Biotechnology 90:1681-1690.

386. Gronenberg LS, Marcheschi RJ, Liao JC. 2013. Next generation biofuel engineering in prokaryotes. Current Opinion in Chemical Biology 17:462-471.

387. Kucek LA, Nguyen M, Angenent LT. 2016. Conversion of l-lactate into *n*-caproate by a continuously fed reactor microbiome. Water Research 93:163-171.

388. Zhu X, Tao Y, Liang C, Li X, Wei N, Zhang W, Zhou Y, Yang Y, Bo T. 2015. The synthesis of *n*-caproate from lactate: a new efficient process for medium-chain carboxylates production. Scientific Reports 5:e14360.

389. Keller MW, Lipscomb GL, Nguyen DM, Crowley AT, Schut GJ, Scott I, Kelly RM, Adams MWW. 2017. Ethanol production by the hyperthermophilic archaeon *Pyrococcus furiosus* by expression of bacterial bifunctional alcohol dehydrogenases. Microbial Biotechnology 10:1535-1545.

390. Wang S, Huang H, Moll J, Thauer RK. 2010. NADP+ reduction with reduced ferredoxin and NADP+ reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*. Journal of Bacteriology 192:5115-5123.

391. Rubinstein GM, Lipscomb GL, Williams-Rhaesa AM, Schut GJ, Kelly RM, Adams MWW. 2020. Engineering the cellulolytic extreme thermophile *Caldicellulosiruptor bescii*

to reduce carboxylic acids to alcohols using plant biomass as the energy source. Journal of Industrial Microbiology and Biotechnology 47:585-597.

392. Biegel E, Schmidt S, González JM, Müller V. 2011. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. Cellular and Molecular Life Sciences 68:613-634.

393. Tremblay P-L, Zhang T, Dar SA, Leang C, Lovley DR. 2013. The Rnf complex of *Clostridium ljungdahlii* is a proton-translocating ferredoxin:NAD+ oxidoreductase essential for autotrophic growth. mBio 4:e00406-12.

394. Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J. 2012. Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii*. Journal of Industrial Microbiology and Biotechnology 40:41-49.

395. Baronofsky JJ, Schreurs WJA, Kashket ER. 1984. Uncoupling by acetic acid limits growth of and acetogenesis by *Clostridium thermoaceticum*. Applied and Environmental Microbiology 48:1134-1139.

396. Yee KL, Rodriguez Jr M, Tschaplinski TJ, Engle NL, Martin MZ, Fu C, Wang Z-Y, Hamilton-Brehm SD, Mielenz JR. 2012. Evaluation of the bioconversion of genetically modified switchgrass using simultaneous saccharification and fermentation and a consolidated bioprocessing approach. Biotechnology for Biofuels 5:1.

397. Kaur A, Capalash N, Sharma P. 2018. Quorum sensing in thermophiles: prevalence of autoinducer-2 system. BMC Microbiology 18:62.

398. Rodionov D, Rodionova I, Rodionov V, Zhang K, Crosby J, Rubinstein G, Nookaew I, Basen M, Brown S, Klingeman D, Poole F, Zhang Y, Kelly R, Adams M. 2020.

Transcriptional regulation of plant biomass degradation and carbohydrate utilization in *Caldicellulosiruptor bescii*. mSystems doi:To Be Submitted.

399. Marion C, Aten AE, Woodiga SA, King SJ. 2011. Identification of an ATPase, MsmK, which energizes multiple carbohydrate ABC transporters in *Streptococcus pneumoniae*. Infection and Immunity 79:4193-4200.

400. Ferreira MJ, Mendes AL, de Sá-Nogueira I. 2017. The MsmX ATPase plays a crucial role in pectin mobilization by *Bacillus subtilis*. PLoS ONE 12:e0189483.

401. Blamey JM, Adams MWW. 1993. Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology 1161:19-27.

402. Kannuchamy S, Mukund N, Saleena LM. 2016. Genetic engineering of *Clostridium thermocellum* DSM1313 for enhanced ethanol production. BMC Biotechnology 16:34.

403. Hon S, Holwerda EK, Worthen RS, Maloney MI, Tian L, Cui J, Lin PP, Lynd LR, Olson DG. 2018. Expressing the *Thermoanaerobacterium saccharolyticum pforA* in engineered *Clostridium thermocellum* improves ethanol production. Biotechnology for Biofuels 11:242.

404. Rydzak T, McQueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, Shamshurin D, Levin DB, Wilkins JA, Sparling R. 2012. Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. BMC Microbiology 12:214.

405. Zhou J, Olson DG, Lanahan AA, Tian L, Murphy SJ-L, Lo J, Lynd LR. 2015. Physiological roles of pyruvate ferredoxin oxidoreductase and pyruvate formate-lyase in

Thermoanaerobacterium saccharolyticum JW/SL-YS485. Biotechnology for Biofuels 8:138.

406. Bielen A, Verhaart M, van der Oost J, Kengen S. 2013. Biohydrogen production by the thermophilic bacterium *Caldicellulosiruptor saccharolyticus*: current status and perspectives. Life 3:52.

407. Özgür E, Mars AE, Peksel B, Louwerse A, Yücel M, Gündüz U, Claassen PAM, Eroğlu İ. 2010. Biohydrogen production from beet molasses by sequential dark and photofermentation. International Journal of Hydrogen Energy 35:511-517.

408. Mars AE, Veuskens T, Budde MAW, van Doeveren PFNM, Lips SJ, Bakker RR, de Vrije T, Claassen PAM. 2010. Biohydrogen production from untreated and hydrolyzed potato steam peels by the extreme thermophiles *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. International Journal of Hydrogen Energy 35:7730-7737.

409. Pawar SS, Nkemka VN, Zeidan AA, Murto M, van Niel EWJ. 2013. Biohydrogen production from wheat straw hydrolysate using *Caldicellulosiruptor saccharolyticus* followed by biogas production in a two-step uncoupled process. International Journal of Hydrogen Energy 38:9121-9130.

410. Panagiotopoulos IA, Bakker RR, de Vrije T, Claassen PAM, Koukios EG. 2012. Dilute-acid pretreatment of barley straw for biological hydrogen production using *Caldicellulosiruptor saccharolyticus*. International Journal of Hydrogen Energy 37:11727-11734.

411. van Niel EWJ, Budde MAW, de Haas GG, van der Wal FJ, Claassen PAM, Stams AJM. 2002. Distinctive properties of high hydrogen producing extreme thermophiles,

Caldicellulosiruptor saccharolyticus and *Thermotoga elfii*. International Journal of Hydrogen Energy 27:1391-1398.

412. de Vrije T, Bakker RR, Budde MAW, Lai MH, Mars AE, Claassen PAM. 2009. Efficient hydrogen production from the lignocellulosic energy crop *Miscanthus* by the extreme thermophilic bacteria *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. Biotechnology for Biofuels 2:12.

413. Willquist K, van Niel EWJ. 2012. Growth and hydrogen production characteristics of *Caldicellulosiruptor saccharolyticus* on chemically defined minimal media. International Journal of Hydrogen Energy 37:4925-4929.

414. Ivanova G, Rákhely G, Kovács KL. 2008. Hydrogen production from biopolymers by *Caldicellulosiruptor saccharolyticus* and stabilization of the system by immobilization. International Journal of Hydrogen Energy 33:6953-6961.

415. de Vrije T, Budde MAW, Lips SJ, Bakker RR, Mars AE, Claassen PAM. 2010. Hydrogen production from carrot pulp by the extreme thermophiles *Caldicellulosiruptor saccharolyticus* and *Thermotoga neapolitana*. International Journal of Hydrogen Energy 35:13206-13213.

416. Panagiotopoulos IA, Bakker RR, de Vrije T, Koukios EG, Claassen PAM. 2010. Pretreatment of sweet sorghum bagasse for hydrogen production by *Caldicellulosiruptor saccharolyticus*. International Journal of Hydrogen Energy 35:7738-7747.

417. Ivanova G, Rákhely G, Kovács KL. 2009. Thermophilic biohydrogen production from energy plants by *Caldicellulosiruptor saccharolyticus* and comparison with related studies. International Journal of Hydrogen Energy 34:3659-3670.

418. Kádár Z, de Vrije T, van Noorden GE, Budde MAW, Szengyel Z, Réczey K, Claassen PAM. 2004. Yields from glucose, xylose, and paper sludge hydrolysate during hydrogen production by the extreme thermophile *Caldicellulosiruptor saccharolyticus*. Applied Biochemistry and Biotechnology 114:497-508.

419. Willquist K, Claassen PAM, van Niel EWJ. 2009. Evaluation of the influence of CO_2 on hydrogen production by *Caldicellulosiruptor saccharolyticus*. International Journal of Hydrogen Energy 34:4718-4726.

420. Ljunggren M, Willquist K, Zacchi G, van Niel EWJ. 2011. A kinetic model for quantitative evaluation of the effect of hydrogen and osmolarity on hydrogen production by *Caldicellulosiruptor saccharolyticus*. Biotechnology for Biofuels 4:31.

421. Han W, Hu YY, Li SY, Li FF, Tang JH. 2016. Biohydrogen production from waste bread in a continuous stirred tank reactor: a techno-economic analysis. Bioresource Technology 221:318-323.

422. Grimm A, de Jong WA, Kramer GJ. 2020. Renewable hydrogen production: a techno-economic comparison of photoelectrochemical cells and photovoltaic-electrolysis. International Journal of Hydrogen Energy 45:22545-22555.

423. Straub CT, Bing RG, Otten JK, Keller LM, Zeldes BM, Adams MWW, Kelly RM. 2020. Metabolically engineered *Caldicellulosiruptor bescii* as a platform for producing acetone and hydrogen from lignocellulose. Biotechnology and Bioengineering n/a.

424. Isern NG, Xue J, Rao JV, Cort JR, Ahring BK. 2013. Novel monosaccharide fermentation products in *Caldicellulosiruptor saccharolyticus* identified using NMR spectroscopy. Biotechnology for Biofuels 6:47.

425. Saxena R, Anand P, Saran S, Isar J, Agarwal L. 2010. Microbial production and applications of 1,2-propanediol. Indian Journal of Microbiology 50:2-11.

426. Cameron DC, Cooney CL. 1986. A novel fermentation: the production of R(–)–1,2–propanediol and acetol by *Clostridium thermosaccharolyticum*. Nature Biotechnology 4:651-654.

427. Aiello-Mazzarri C, Agbogbo FK, Holtzapple MT. 2006. Conversion of municipal solid waste to carboxylic acids using a mixed culture of mesophilic microorganisms. Bioresource Technology 97:47-56.

428. Kim WJ, Kim HU, Lee SY. 2017. Current state and applications of microbial genome-scale metabolic models. Current Opinion in Systems Biology 2:10-18.

429. Nilsson A, Nielsen J. 2017. Genome scale metabolic modeling of cancer. Metabolic Engineering 43:103-112.

430. Raškevičius V, Mikalayeva V, Antanavičiūtė I, Ceslevičienė I, Skeberdis VA, Kairys V, Bordel S. 2018. Genome scale metabolic models as tools for drug design and personalized medicine. PLoS ONE 13:e0190636.

431. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. 2019. Current status and applications of genome-scale metabolic models. Genome Biology 20:121.

432. Zhang K, Zhao W, Rodionov D, Rubinstein G, Nguyen D, Crosby J, Bing R, Kelly R, Adams M, Zhang Y. 2020. Genome-scale metabolic model of *Caldicellulosiruptor bescii* reveals optimal metabolic engineering strategies for ethanol producton. mSystems doi:To Be Submitted.

433. Dash S, Khodayari A, Zhou J, Holwerda EK, Olson DG, Lynd LR, Maranas CD. 2017. Development of a core *Clostridium thermocellum* kinetic metabolic model consistent with multiple genetic perturbations. Biotechnology for Biofuels 10:108.

434. Garcia S, Thompson RA, Giannone RJ, Dash S, Maranas CD, Trinh CT. 2020. Development of a genome-scale metabolic model of *Clostridium thermocellum* and its applications for Integration of multi-omics datasets and computational strain design. Frontiers in Bioengineering and Biotechnology 8.

435. Salimi F, Zhuang K, Mahadevan R. 2010. Genome-scale metabolic modeling of a clostridial co-culture for consolidated bioprocessing. Biotechnology Journal 5:726-738.

436. Currie DH, Raman B, Gowen CM, Tschaplinski TJ, Land ML, Brown SD, Covalla SF, Klingeman DM, Yang ZK, Engle NL, Johnson CM, Rodriguez M, Shaw AJ, Kenealy WR, Lynd LR, Fong SS, Mielenz JR, Davison BH, Hogsett DA, Herring CD. 2015. Genome-scale resources for *Thermoanaerobacterium saccharolyticum*. BMC Systems Biology 9:30.

437. Vansuch GE, Wu C-H, Haja DK, Blair SA, Chica B, Johnson MK, Adams MWW, Dyer RB. 2020. Metal–ligand cooperativity in the soluble hydrogenase-1 from *Pyrococcus furiosus*. Chemical Science 11:8572-8581.

438. Kim MK, Lun DS. 2014. Methods for integration of transcriptomic data in genome-scale metabolic models. Computational and Structural Biotechnology Journal 11:59-65.

439. Krumholz EW, Libourel IGL. 2017. Thermodynamic constraints improve metabolic networks. Biophysical Journal 113:679-689.

440. SuRin L, Pil K. 2020. Current status and applications of adaptive laboratory evolution in industrial microorganisms. Journal of Microbiology and Biotechnology 30:793-803.

441. Patyshakuliyeva A, Arentshorst M, Allijn IE, Ram AFJ, de Vries RP, Gelber IB. 2016. Improving cellulase production by *Aspergillus niger* using adaptive evolution. Biotechnology Letters 38:969-974.

442. Gu H, Zhang J, Bao J. 2014. Inhibitor analysis and adaptive evolution of *Saccharomyces cerevisiae* for simultaneous saccharification and ethanol fermentation from industrial waste corncob residues. Bioresource Technology 157:6-13.

443. Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. Molecular Ecology 28:1537-1549.

444. Straub CT, Khatibi PA, Otten JK, Adams MWW, Kelly RM. 2019. Lignocellulose solubilization and conversion by extremely thermophilic *Caldicellulosiruptor bescii* improves by maintaining metabolic activity. Biotechnology and Bioengineering 116:1901-1908.

445. Favaro L, Cagnin L, Basaglia M, Pizzocchero V, van Zyl WH, Casella S. 2017. Production of bioethanol from multiple waste streams of rice milling. Bioresource Technology 244:151-159.

446. Mougiakos I, Mohanraju P, Bosma EF, Vrouwe V, Finger Bou M, Naduthodi MIS, Gussak A, Brinkman RBL, van Kranenburg R, van der Oost J. 2017. Characterizing a thermostable Cas9 for bacterial genome editing and silencing. Nature Communications 8:1647.

447. Tsui TKM, Hand TH, Duboy EC, Li H. 2017. The impact of DNA topology and guide length on target selection by a cytosine-specific Cas9. ACS Synthetic Biology 6:1103-1113.

448. Stanley SY, Maxwell KL. 2018. Phage-encoded anti-CRISPR defenses. Annual Review of Genetics 52:445-464.

449. Li Y, Pan S, Zhang Y, Ren M, Feng M, Peng N, Chen L, Liang YX, She Q. 2015. Harnessing Type I and Type III CRISPR-Cas systems for genome editing. Nucleic Acids Research 44:e34-e34.

450. Li Y, Peng N. 2019. Endogenous CRISPR-Cas system-based genome editing and antimicrobials: review and prospects. Frontiers in Microbiology 10.

451. Baker PL, Orf GS, Kevershan K, Pyne ME, Bicer T, Redding KE. 2019. Using the endogenous CRISPR-Cas system of *Heliobacterium modesticaldum* to delete the photochemical reaction center core subunit gene. Applied and Environmental Microbiology 85:e01644-19.

452. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. 2016. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in *Clostridium*. Scientific Reports 6:25666.

453. Zhang J, Zong W, Hong W, Zhang Z-T, Wang Y. 2018. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in *Clostridium tyrobutyricum* and engineer the strain for high-level butanol production. Metabolic Engineering 47:49-59.

454. Maikova A, Kreis V, Boutserin A, Severinov K, Soutourina O. 2019. Using an endogenous CRISPR-Cas system for genome editing in the human pathogen *Clostridium difficile*. Applied and Environmental Microbiology 85:e01416-19.

455. Hidalgo-Cantabrana C, Goh YJ, Pan M, Sanozky-Dawes R, Barrangou R. 2019. Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. Proceedings of the National Academy of Sciences 116:15774-15783.

456. Zheng Y, Han J, Wang B, Hu X, Li R, Shen W, Ma X, Ma L, Yi L, Yang S, Peng W. 2019. Characterization and repurposing of the endogenous Type I-F CRISPR–Cas system of *Zymomonas mobilis* for genome engineering. Nucleic Acids Research 47:11461-11475.

457. Le Y, Fu Y, Sun J. 2020. Genome editing of the anaerobic thermophile *Thermoanaerobacter ethanolicus* using thermostable Cas9. Applied and Environmental Microbiology doi:10.1128/aem.01773-20:AEM.01773-20.

458. Donohoue PD, Barrangou R, May AP. 2018. Advances in industrial biotechnology using CRISPR-Cas systems. Trends in Biotechnology 36:134-146.

459. Lu H, Villada JC, Lee PKH. 2019. Modular metabolic engineering for biobased chemical production. Trends in Biotechnology 37:152-166.

460. Umbarger Mark A, Toro E, Wright Matthew A, Porreca Gregory J, Baù D, Hong S-H, Fero Michael J, Zhu Lihua J, Marti-Renom Marc A, McAdams Harley H, Shapiro L, Dekker J, Church George M. 2011. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. Molecular Cell 44:252-264.

461. Dame RT, Tark-Dame M. 2016. Bacterial chromatin: converging views at different scales. Current Opinion in Cell Biology 40:60-65.

462. Wellington M, Rustchenko E. 2005. 5-Fluoro-orotic acid induces chromosome alterations in *Candida albicans*. Yeast 22:57-70.

463. Keller MW, Lipscomb GL, Loder AJ, Schut GJ, Kelly RM, Adams MWW. 2015. A hybrid synthetic pathway for butanol production by a hyperthermophilic microbe. Metabolic Engineering 27:101-106.

APPENDIX A

ALTERNATIVE FORMULATIONS OF THE AOR-ADH PATHWAY

In Chapter 3 of this work, extensive discussion was dedicated to the successful expression of the AOR-Adh pathway in *Caldicellulosiruptor bescii*. The work described here is a continuation of this effort with the aim of improving the titers and yields of both ethanol and exogenous acid conversion in strains which heterologously express the pathway. For this purpose, additional AOR-Adh pathway strains of *C. bescii* were generated using homologs of aldehyde ferredoxin oxidoreductase (AOR) and alcohol dehydrogenase (Adh) with varying substrate specificities in a different genetic background than the strains described in Chapter 3.

The AOR-Adh and AOR-Adh /Δ*ldh* strains (MACB1013 and MACB1038, respectively) were constructed in the JWCB005 genetic background. 5-FOA is a known mutagen; by selecting for a random uracil auxotrophic mutant on 5-FOA additional mutations were made to the wild type *C. bescii* genome (**Figure A.1A**) (275, 462). Although, when provided with exogenous uracil, strain JWCB005 and its daughter strains display the same growth phenotype as the wild type, these strains also display a cell aggregation phenotype. Strains from this lineage, including MACB1013 and MACB1038, form visible aggregates of cells in liquid culture, most notably under static growth conditions, but also during growth with agitation. The source of this phenotype is unknown. Moving forward, it was desirable to have a genetic background as similar to the wild type as possible. MACB1018 is a new wild type derived, genetically tractable lineage of *C. bescii*, containing a clean deletion of *pyrE* (242). Like *pyrF*, deletion of *pyrE* confers uracil auxotrophy and 5-FOA resistance. The MACB1018 strain contains less genome rearrangements, single

nucleotide polymorphisms (SNPs, insertions/deletions (InDels), and instances of the IS*Cbe4* transposable element than the parent strain used to generate the AOR-Adh pathway strains from Chapter 3 (**Figures A.1 and A.2**) (279). In the strain MACB1018 background, a Δ*ldh* strain (MACB1034) was generated to serve as the new parent strain for metabolic engineering (See Chapter 4).

In addition to the established pathway, three other variants are needed to represent all four combinations of two homologs of each pathway gene. Pathway variants are abbreviated 4H, 8H, 4P, and 8P (**Figure A.3**). P and N imply NADPH or NADH specificity of the Adh, respectively, while 4 and 8 refer to the iron content of the ferredoxin preference of the AOR. Ferredoxin preference was inferred from the structure of the primary ferredoxin in the source organism. Source organisms and pathway gene homologs are listed in **Table A.1** and **Table A.2**, respectively.

The AOR-Adh pathway in *Pyrococcus furiosus,* the impetus for the work in *C. bescii*, relied on the insertion of the AdhA from *Thermoanaerobacter* sp. X514. This enzyme utilizes an NADPH cofactor, and only has ~10% NADH linked activity in enzyme assays (296). In *P. furiosus,* insertion of this gene allowed NADPH generated by the oxidation of hydrogen through the soluble hydrogenase (SHI) to reduce acetaldehyde to ethanol. For a balanced AOR-AdhA pathway in *C. bescii,* the reduction of acetaldehyde to ethanol would need to recycle NADH produced by glyceraldehyde-3-phosphate dehydrogenase (GAPDH) during glycolysis. A homolog of AdhA, butanol dehydrogenase (Bdh) can also catalyze the reduction of acetaldehyde to ethanol, but its activity is NADH specific rather than NADPH specific (463). Expressing *bdh* from the same source organism, *T.* sp. X514, in *C. bescii* might create a balanced pathway where reductant flows more efficiently towards the desired reduced end-product: ethanol.

In addition to the well-characterized AOR from *P. furiosus,* I expressed the putative AOR from *T.* sp. X514. Although this homolog is uncharacterized, whole cell extract shows benzyl viologen linked AOR activity comparable to that of *P. furiosus,* and cell suspension assays showed the reduction of exogenously added isobutyrate to isobutanol (303). AOR family enzymes utilize a low potential ferredoxin cofactor. The primary ferredoxin in *P. furiosus* contains only one 4Fe-4S cluster, while the ferredoxin in *T.* sp. X514 contains two 4Fe-4S clusters. This 8Fe ferredoxin is akin to the primary ferredoxin in *C. bescii,* which also contains two 4Fe-4S clusters. It is possible that the native *C. bescii* ferredoxin has higher affinity for the AOR from the closely related bacterium *T.* sp. X514, rather than the AOR from the archaeon *P. furiosus* (see: **Figure S3.3**).

Two different combinations of the four genes were expressed in *C. bescii* using replicating shuttle vectors expressing the thermostable kanamycin resistance gene, Cb*htk*. While not all four variant strains were constructed, the 4H and 8H strains were successfully generated. Comparison of these strains with the original 4P strain in the unstable genetic background (Chapter 3) and deductive reasoning reveal that the 8P strain (not completed) would likely produce the least ethanol. Both new strains (MACB1060 – 8H, and MACB1061 – 4H) express the AOR-Adh pathway using the same operon structure: the genes are expressed under the control of the high level constitutive *C. bescii* S-layer protein promoter, with a ribosomal binding site situated between the two genes, and a putative rho-independent stem loop terminator from *Thermoanaerobacter* sp. X514 downstream of the alcohol dehydrogenase. All strain construction, growth, and product characterization experiments reported here were conducted with the same materials and methods described in Chapter 4.

Side-by-side growth of the new 4H and 8H strains with the MACB1034 (Δ*ldh*) parent strain and the previously examined 4P strain (Chapter 3) revealed no difference in growth phenotype

(**Figure A.4**). Before strain construction, the hypothesis was that the 8H strain should provide the best cofactor match and produce the most ethanol. Experimentally, the 4H strain produced the most ethanol, with the 8H strain producing less ethanol than the 4P strain (**Figure A.4A**). The implication of these results is that *bdh* is a better Adh variant for the pathway than *adhA*, and that the *T.* sp. X514 *aor* does not provide an improvement over the *P. furiosus aor*. Interestingly, in the presence of 20 mM exogenous isobutyrate, the 4H strain did not produce as much isobutanol as the 4P strain. This result suggests that the specificity of Bdh for isobutyrate is lower than that of AdhA; it is possible that this trend extends further, and the substrate specificity of Bdh is generally narrower. The 4H strain, which uses the *P. furiosus* AOR and the NADH dependent Bdh, produced 4 mM ethanol, which is as much ethanol as the 4P strain produced under controlled fermentation conditions in Chapter 3, and 1.6-fold higher than the 4P strain produced in bottles (2.4 mM).

Overall, even with improved ethanol production, the 4H strain did not demonstrate a switch from acetate to ethanol as the primary reduced end-product of *C. bescii* metabolism. It is important to emphasize that the 4P strain was integrated into the genome, while the 4H and 8H strains were expressed from a replicating shuttle vector. As discussed in Chapter 5, expression from shuttle vectors under the same promoter in *C. bescii* produces less transcript than expression from the genome (423). Thus, it is possible that expression of the 4H and 8H pathway variants from the genome would increase ethanol production. The 8H strain might produce more ethanol than the 4P strain under conditions of equal gene expression, and the 4H strain might produce the highest ethanol formation by the AOR-Adh pathway in *C. bescii* seen to date. Further, the difference in ethanol production and gene expression for the AdhE-Rnf pathway in different parent backgrounds (stable vs unstable) revealed a 2-fold change in ethanol production; for a direct comparison, the 4P and 4H strains should be expressed from the genome in the stable genetic lineage.

Tables and Figures

Table A.1 Gene sources for expression of the AOR-Adh pathway

Two exogenous gene donors (*Pyrococcus furiosus*, green, *Thermoanaerobacter* sp. X514, pink) are compared to the optimum growth temperature (T_{opt}) of the expression host, *C. bescii*.

ORF Prefix	Organism	T_{opt} (ºC)
PF	*Pyrococcus furiosus*	100
Teth514	*Thermoanaerobacter* sp. X514	60
Athe	*Caldicellulosiruptor bescii*	78

Table A.2 Genes for expression of the AOR-Adh pathway in *C. bescii*

Homologs of AOR and Adh for alternative formulations of the AOR-Adh pathway in *C. bescii*. Genes from *Thermoanaerobacter* sp. X514 are shaded in pink and the *aor* from *Pyrococcus furiosus* is shaded in green.

Enzyme	Gene Annotation	Cofactor	Comments
AdhA	Primary alcohol dehydrogenase	NADPH	20% activity w/ NADH
Bdh	Butanol dehydrogenase	NADH	No activity w/ NADPH
AOR	Aldehyde ferredoxin oxidoreductase	4Fe-Fd	Prefers PF ferredoxin
AOR	Aldehyde ferredoxin oxidoreductase	8Fe-Fd	Untested

Figure A.1 Stability of genetic lineages in *C. bescii*

Two lineages of *C. bescii* strains differ in their genome stability. (A) Single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) in the PacBio re-sequenced genomes of the wild type, JWCB005, JWCB018, and MACB1018 strains of *C. bescii* are indicated. Numbers in the wild type column are relative to the original published genome sequence of the wild type. Data from (279). (B) AOR-Adh pathway variant strains in *C. bescii* are presented as a tree to demonstrate lineages. The newer stable lineage is presented in blue, while the older unstable lineage is shown in red.

Figure A.1

A

	Wild Type	JWCB005	JWCB018	MACB1018
SNPs	4	50	56	9
InDels	21	73	125	30

B

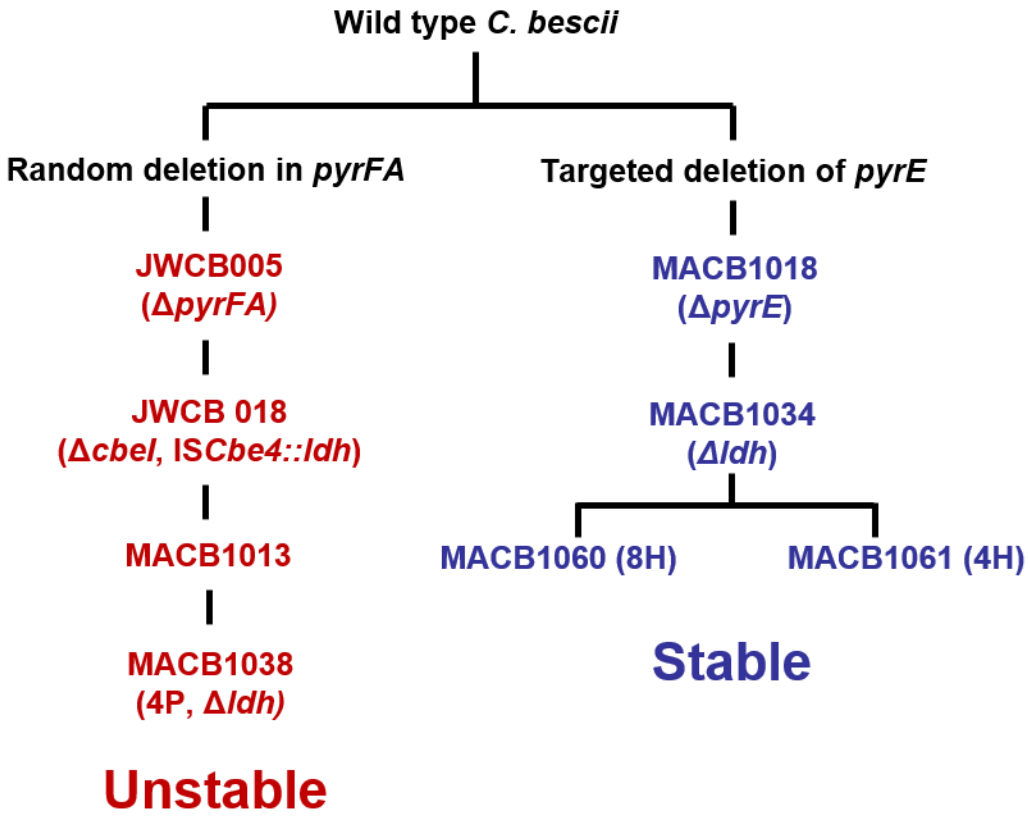


Figure A.2 PacBio sequencing of *C. bescii* strains reveals genome rearrangements

Reprinted from (279). Overall genome arrangements for the wild type, JWCB005, MACB1018, and JWCB018 strains of *C. bescii*. IS*Cbe4* elements are shown in black lines along the outside of the circular genome diagrams. There are no large-scale rearrangements in JWCB005 and MACB1018 compared to wild type. There are two large rearrangements and inversions in JWCB018 (green and red regions) compared to the wild type and parent strain, JWCB005.

Figure A.2

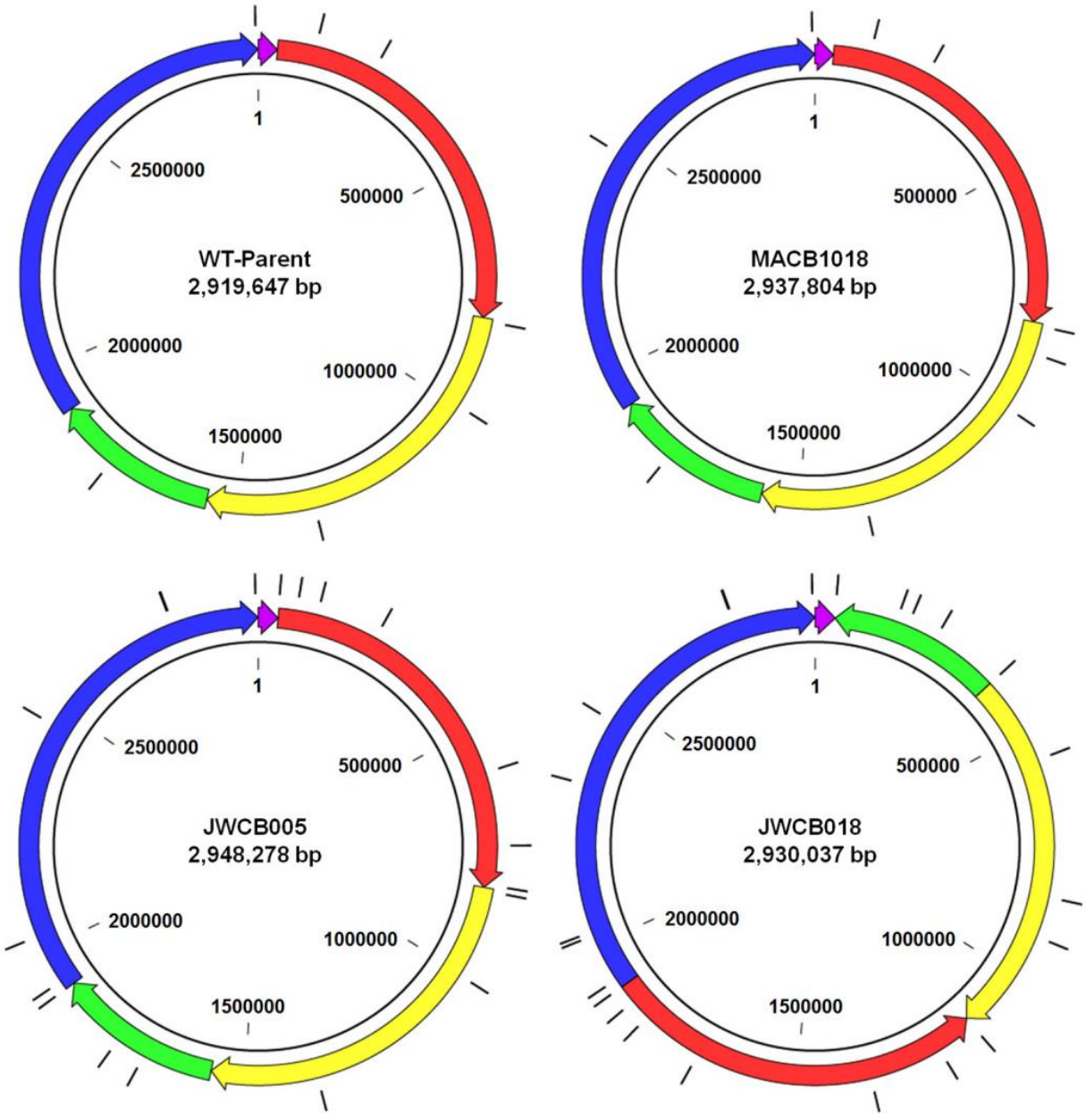

1
500000
1000000
1500000
2000000
2500000
WT-Parent
2,919,647 bp
MACB1018
2,937,804 bp
JWCB005
2,948,278 bp
JWCB018
2,930,037 bp

Figure A.3 Primary metabolism in proposed AOR-Adh pathway variants

C. bescii (purple) redox metabolism in proposed strains, expressing all combinations of *Pyrococcus furiosus* (green) AOR, and Bdh, AdhA, and AOR from *Thermoanaerobacter* sp. X514 (pink).

Figure A.3

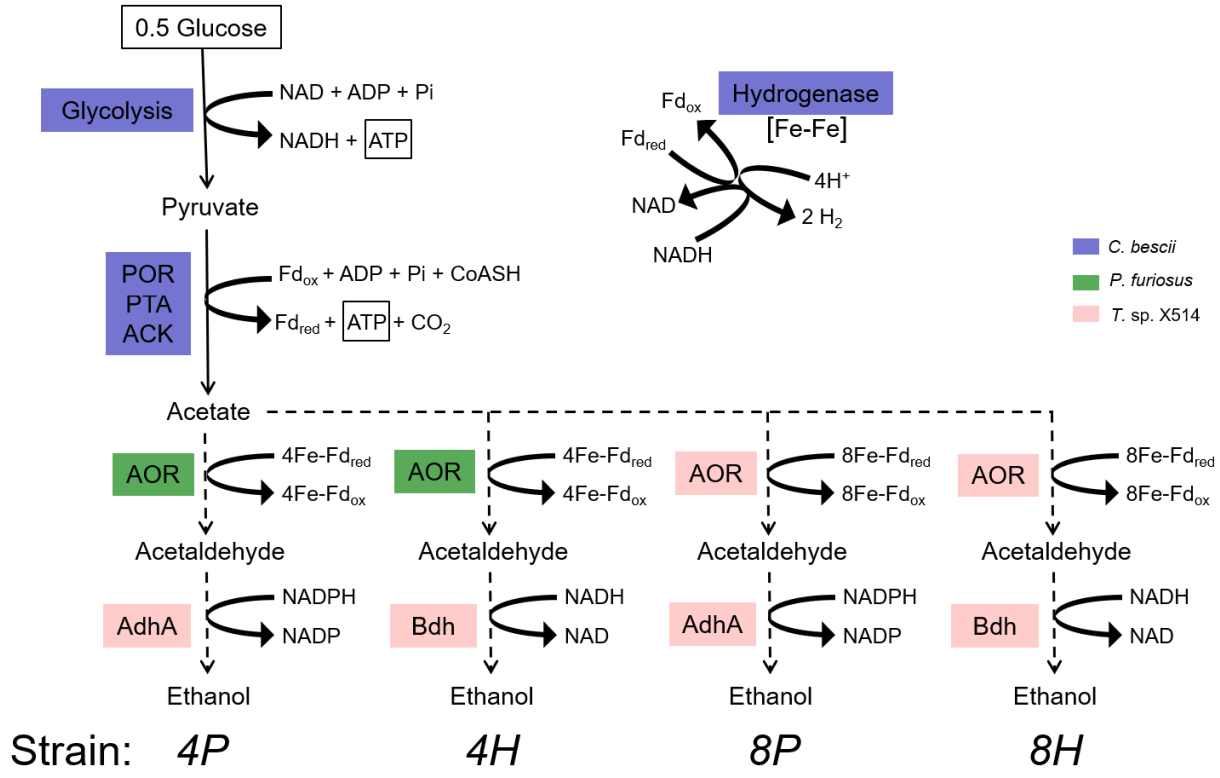
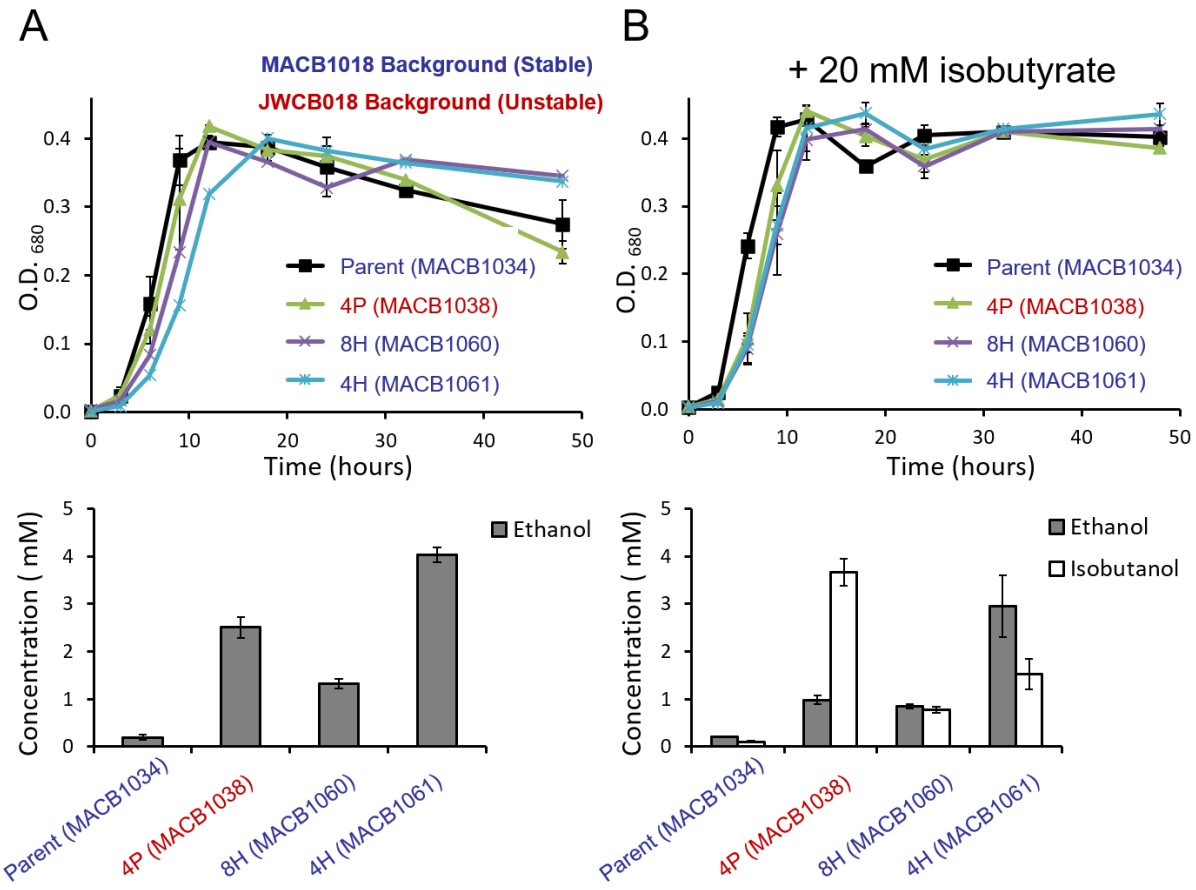

0.5 Glucose
Glycolysis
NAD + ADP + Pi
NADH + ATP
Pyruvate
POR
PTA
ACK
Fdox + ADP + Pi + CoASH
Fdred + ATP + CO2
Acetate
Hydrogenase
[Fe-Fe]
Fdox
Fdred
4H+
NAD
2 H2
NADH
C. bescii
P. furiosus
T. sp. X514
AOR
4Fe-Fdred
4Fe-Fdox
8Fe-Fdred
8Fe-Fdox
Acetaldehyde
AdhA
Bdh
NADPH
NADP
NADH
NAD
Ethanol
Strain: 4P 4H 8P 8H

Figure A.4 Comparing ethanol production and isobutyrate reduction in pathway variants

Three variants of the AOR-Adh pathway were compared with the parent strain (Δ*ldh*) for growth and alcohol production in the absence (A) and presence (B) of 20 mM isobutyrate. Strains constructed in the stable background (MACB1018) are written in blue, while the strain constructed in the unstable background (JWCB018) is shown in red. In the stable background, pathway genes were expressed from a replicating shuttle vector, while in the unstable background strain pathway genes were integrated into the genome. Growth in complex glucose medium (measured by $O.D._{680}$) was followed for 48 hours (top). Endpoint ethanol (gray) and isobutanol (white) were measured by gas chromatography (bottom). Error bars represent biological triplicates (n=3).

Figure A.4


A
MACB1018 Background (Stable)
JWCB018 Background (Unstable)
O.D. 680
Time (hours)
Parent (MACB1034)
4P (MACB1038)
8H (MACB1060)
4H (MACB1061)
B
+ 20 mM isobutyrate
Concentration (mM)
Ethanol
Isobutanol

APPENDIX B

LIST OF PUBLICATIONS

1. †Zhang K, †Zhao W, Rodionov DA, **Rubinstein GM**, Nguyen DN, Crosby JR, Bing RG, Kelly RM, Adams MWW, and Zhang Y. "Genome-Scale Metabolic Model of *Caldicellulosiruptor bescii* Reveals Optimal Metabolic Engineering Strategies for Ethanol Producton." mSystems. To be submitted: November 2020
2. Rodionov DA, Rodionova IA, Rodionov V, Zhang K, Crosby J, **Rubinstein GM**, Nookaew I, Basen M, Brown SD, Klingeman DM, Poole FL, Zhang Y, Kelly RM, and Adams MWW. "Transcriptional regulation of plant biomass degradation and carbohydrate utilization in *Caldicellulosiruptor bescii*." mSystems. To be submitted: November 2020
3. **Rubinstein GM**, Lipscomb GL, Williams-Rhaea AM, Schut GJ, Kelly RM, and Adams MWW. (2020) "Engineering the cellulolytic thermophile *Caldicellulosiruptor bescii* to reduce carboxylic acids to alcohols using plant biomass as the energy source." Journal of Industrial Microbiology & Biotechnology Accepted July 2020.
4. **†Rubinstein GM**, †Lee LL, †Crosby JR, Laemthong T, Bing RG, Straub CT, Adams MWW, and Kelly RM. (2020) "The biology and biotechnology of the genus *Caldicellulosiruptor*: recent developments in 'Caldi World'." Extremophiles 24:1-15. doi:10.1007/s00792-019-01116-5

5. †**Rubinstein GM**, †Scott IM, Poole FL, Lipscomb GL, Schut GJ, Williams-Rhaesa AM, Stevenson DM, Amador-Noguez D, Kelly RM, and Adams MWW. (2019) "The thermophilic biomass-degrading bacterium *Caldicellulosiruptor bescii* utilizes two enzymes to oxidize glyceraldehyde 3-phosphate during glycolysis." Journal of Biological Chemistry 294(25):9995-10005.
6. Williams-Rhaesa AM, **Rubinstein GM**, Scott IM, Lipscomb GL, Poole FL, Kelly RM, and Adams MWW. (2018) "Engineering redox-balanced ethanol production in the cellulolytic and extremely thermophilic bacterium, *Caldicellulosiruptor bescii*." Metabolic Engineering Communications doi: 10.1016/j.mec.2018.e00073.
7. Williams-Rhaesa AM, Awuku NK, Lipscomb GL, Poole FL, **Rubinstein GM**, Conway JM, Kelly RM, and Adams MWW. (2018) "Native xylose-inducible promoter expands the genetic tools for the biomass-degrading, extremely thermophilic *Caldicellulosiruptor bescii*." Extremophiles 22 (4):629-638.
8. Williams-Rhaesa AM, Poole FL, Dinsmore J, Lipscomb GL, **Rubinstein GM**, Scott IM, Conway JM, Lee LL, Khatibi PA, Kelly RM, and Adams MWW. (2017) "Genome stability of engineered strains of the extremely thermophilic, lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*." Applied and Environmental Microbiology doi:10.1128/aem.00444–17.
9. Loder AJ, Zeldes BM, Conway JM, Counts JA, Straub CT, Khatibi PA, Lee LL, Vitko NP, Keller MW, Rhaesa AM, **Rubinstein GM**, Scott IM, Lipscomb GL, Adams MWW, and Kelly RM. (2016) "Extreme Thermophiles as Metabolic Engineering Platforms:

Strategies and Current Perspective." In Industrial Biotechnology (eds C. Wittmann and J.C. Liao). doi:10.1002/9783527807796.ch14

10. Scott IM, **Rubinstein GM**, Lipscomb GL, Basen M, Schut GJ, Rhaesa AM, Lancaster WA, Poole FL, Kelly RM, and Adams MWW. (2015) "A New Class of Tungsten-Containing Oxidoreductase in *Caldicellulosiruptor*, a Genus of Plant Biomass-Degrading Thermophilic Bacteria." Applied and Environmental Microbiology 81:7339-7347. doi:10.1128/aem.01634-15

† These Authors Contributed Equally