Improving biological detoxification of furfural and acetate in lignocellulosic hydrolysates using metabolic engineering

By

Jacob Peyton Crigler

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy in Molecular Biosciences

Middle Tennessee State University

May 2018

Thesis Committee:

Dr. Elliot Altman, Chair

Dr. Mary Farone

Dr. Paul Kline

Dr. Brian Robertson

Dr. Mohamed Salem

ACKNOWLEDGEMENTS

Middle Tennessee State University was a great place to learn how to do scientific research. I would like to thank Elliot Altman, first for taking me in to his lab and second for being a wonderful and patient advisor. I would like to thank my graduate committee for taking the time to be on my committee and offering advice. I also would like to thank my fiancé Kim and my parents for being as supportive as anyone could be.

ABSTRACT

Cellulosic ethanol biofuel, made from plant waste products or perennial energy crops like switchgrass, offers many advantages over corn starch-derived ethanol, including less land competition and a lower carbon footprint. However, the efficiency of conversion is currently lower and the cost higher due to the recalcitrance of lignocellulosic biomass. A chemical and/or physical pretreatment step is required to overcome recalcitrance, and common pretreatment methods (e.g., acid, steam, and/or depressurization) release microbial inhibitors, including furfural and acetate, into the hydrolysate, lowering the yield of ethanol via fermentation. Furfural is generated via sugar dehydration, and acetate derives from acetylated xylan in hemicellulose. Utilizing a detoxifying strain is one strategy to overcome the inhibitor dilemma. Biological detoxification potentially allows lower process costs compared to chemical or enzymatic detoxification or alternative pretreatment methods aimed at reducing inhibitor generation. One drawback is time. Thus increasing the rate of furfural and acetate detoxification is desirable. While most microbial species can catabolize acetate, most do not possess the furfural catabolic pathway. A novel *Pseudomonas putida* isolate ALS1267, with a growth rate of 0.25/h in 10 mM minimal furfural medium, was characterized. The genome was sequenced and the furfural pathway cloned into wild-type *P. putida* KT2400, which cannot metabolize furfural, creating a novel strain with an improved growth rate of 0.34/h in 10 mM minimal furfural medium. Genomic library screening was used to find targets for engineering increased acetate consumption in *Escherichia coli.* Sixteen plasmid clones were generated, with growth rate increases of 42.6 to 76.9 percent in 10 g/l minimal

acetate medium, which is a highly inhibitory concentration for this strain. Clones included an uncharacterized oxidoreductase, transporters and other outer membrane proteins, carbon scavengers, and stress defense mechanisms. Genomic mutants with improved acetate consumption were also generated during selection, with mutations in the gluconeogenesis gene *pck* promoter, the '5 UTR of the poorly characterized cold-shock gene *ynaE*, and the global regulator of secondary carbon sources CRP. The second major drawback to biological detoxification is consumption of the sugars to be used to produce ethanol by the detoxifying strain. Elimination of glucose metabolism in *E. coli* was studied by characterizing fast-growing revertants in strains engineered to be glucose minus. All of the revertants either altered or overproduced the N-acetylglucosamine phosphotransferase system. Deletion of the N-acetylglucosamine transporter stabilized the glucose minus phenotype and prevented the occurrence of fast-growing revertants.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Cellulosic Ethanol Biofuel

Cellulosic ethanol biofuel (CE), an alternative to corn starch-derived ethanol biofuel, is produced from forestry, municipal, agricultural, and paper industry waste products or from perennial energy crops like switchgrass, elephant grass, and giant cane. Greenhouse gas (GHG) emissions from CE are 2- to 7-fold lower than corn starch-derived ethanol, due to high energy requirements for corn cultivation, turnover of forests and grasslands for land suitable for corn cultivation, and use of the CE lignin waste stream to provide steam to the CE plant (Cherubini *et al.*, 2009; Searchinger *et al*., 2012; Kumar and Murthy, 2011). Different assessments found the GHG emissions per energy output of corn starch-derived ethanol are similar to or worse than gasoline (de Vries *et al*., 2010; Plevin *et al*., 2010; Jenkins and Alles, 2011). Other environmental disadvantages of using corn for biofuel include increased soil fertility depletion, erosion, and fertilizer usage leading to leaching and water pollution (de Vries *et al*., 2010; Pimentel, 2003). Since the year 2000 corn production in the US has increased 960%, with 31% of the increase due to ethanol production (Figures 1.1 and 1.2) (USDA ERS, 2018).

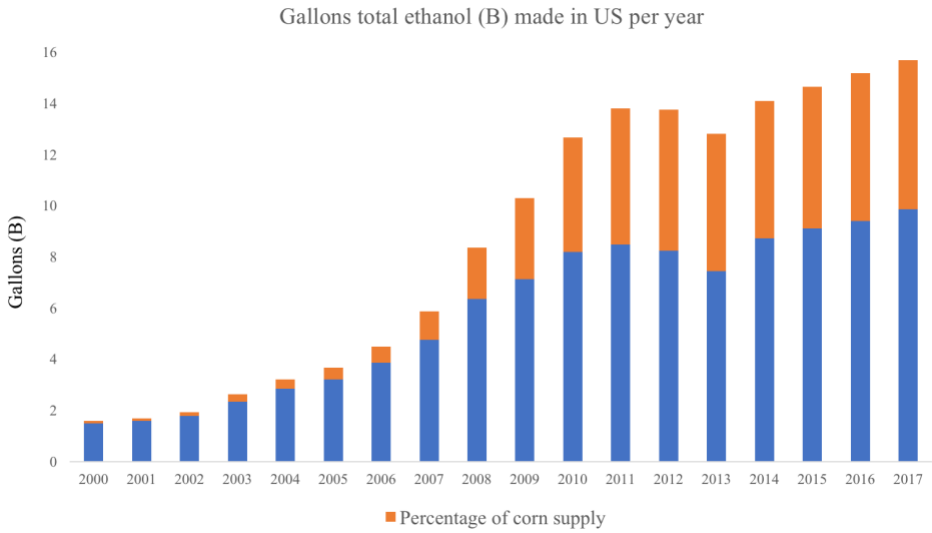

Figure 1.1. Yearly US ethanol biofuel production and percentage of corn harvest
Scale in billion (B) gallons. Data are from USDA Economic Research Service: https://www.ers.usda.gov/data-products/us-bioenergy-statistics/.

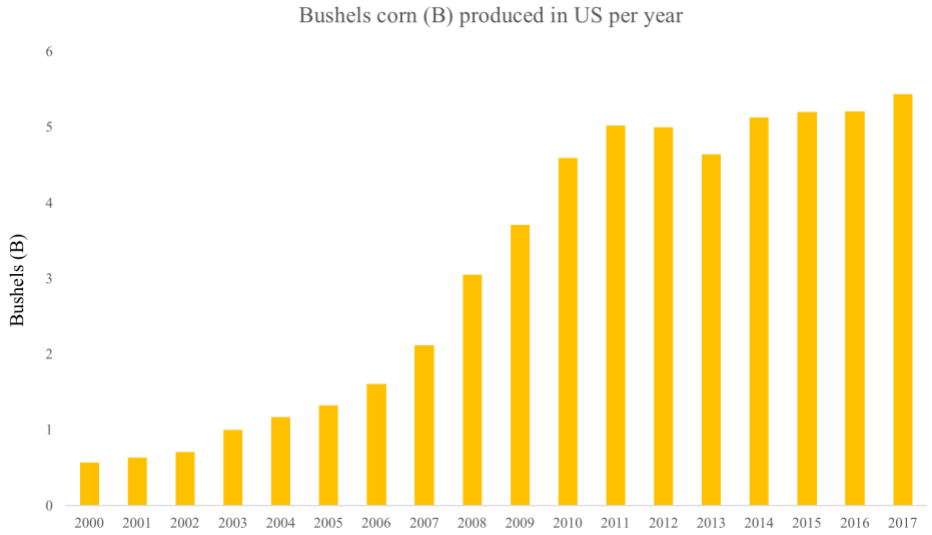

Figure 1.2. Yearly US corn production
Scale in billion (B) bushels. Data are from USDA Economic Research Service: https://www.ers.usda.gov/data-products/us-bioenergy-statistics/.

Approximately 23% of the increase in corn production from 2000 to 2017 (Figure 1.2) is due to improved yield per acre (Nielsen, 2017). Thus the remaining increase in production required increased land use. CE offers the advantage of using industry waste products and thus does not compete for agricultural land. Land competition is lower with switchgrass and other perennial energy crops, as they were selected for the ability to grow on marginal lands (Gelfand *et al*., 2013). Additionally, energy crops require fewer resource inputs, including 7- to 10-fold fewer pesticides, compared to cultivation of corn (McLaughlin and Kszos, 2005; Tilman *et al*., 2006; Hamelinck and Faaij, 2006). CE also offers additional economic benefits to rural, agricultural areas (Zhang *et al*., 2016; Jackson *et al*., 2018). CE cannot replace gasoline due to constraints on available biomass, but estimates are that it could supplement gasoline usage by 20 - 30% (Ohlrogge *et al*., 2009; Schubert, 2006).

The US Renewable Fuels Standard 2 (RSF2) of the Energy Independence and Security Act of 2007 put fourth an ambitious mandate to produce 36 billion (B) gallons of total renewable fuels by 2022. This includes corn starch-derived ethanol, CE, biodiesel, and advanced biofuels. The current production of corn starch-derived ethanol in the US is just over 15 B gallons per year (USDA ERS, 2018), which is the projected maximum production capability (Tyner, 2008). Accordingly, the 2018 - 2022 RSF2 targets for corn starch-derived ethanol production plateau at 15 B gallons per year. The 2017 RSF2 target for CE was 310 million (M) gallons; 60 M gallons were produced (USDA ERS, 2018).

The 2022 target for CE production in 2007, when the regulations were first written, was 16 B gallons per year (EPA, 2013). The mandates are revised yearly; the 2018 target for CE is 288 M gallons (EPA, 2018).

CE is still in limited production because it is not yet cost-competitive with corn starch-derived ethanol. The difficulties hindering CE production lie in the highly recalcitrant nature of lignocellulose, the composition of plant biomass which includes cellulose, hemicellulose, and lignin fractions. After an initial grinding/milling step, a chemical and/or physical pretreatment step is required to deconstruct the lignin barrier and liberate the polysaccharides of cellulose and hemicellulose. Pretreatment methods and their feedstocks that have been employed at the commercial scale include dilute acid (POET-DSM/corn stover and residuals and Abengoa/wheat straw), ammonia and steam (DuPont/corn stover and residuals), and steam explosion (Beta Renewables/giant cane and wheat straw) (Valdivia *et al*., 2016; Somerville, 2015). There are numerous reviews of pretreatment processes and their respective benefits and drawbacks, including Zheng *et al*. (2009), Jönsson and Martín (2016), and Lynd *et al*. (2017). Pretreatment is followed by a chemical or enzymatic polysaccharide hydrolysis step to make sugar monomers available for fermentation to ethanol. Both process steps, along with the cost of enzymes in enzymatic hydrolysis, represent nearly 40% of the total cost of CE production (Yang and Wyman, 2008).

1.2 Inhibitor Generation in Lignocellulosic Biomass Pretreatment Processes

An additional problem is that pretreatment can release inhibitors of microbial growth, including acetic acid and the furan aldehydes furfural and 5-hydroxymethylfurfural (HMF), into the hydrolysate. Acetic acid derives from hydrolysis of acetylated xylan in the hemicellulose fraction, and the furan aldehydes are formed from sugar degradation (Klinke *et al*., 2004; Taherzadeh and Karimi, 2008). The mechanisms of growth inhibition of the furans and acetic acid have been reviewed (Palmqvist and Hahn-Hägerdal, 2000; Mills *et al*., 2009; Piotrowski *et al*., 2014; Almeida *et al*., 2007). Furfural and HMF induce reactive oxygen species, decrease concentrations of reduced redox cofactors, mutagenize DNA, and inhibit enzymes of glycolysis and ethanol fermentation and the pyruvate dehydrogenase complex, as well as act on undetermined hydrophobic targets (Allen *et al*., 2010; Ask *et al*., 2013; Banerjee *et al*., 1981; Modig *et al*., 2002). Besides the acid stress of acetic acid, the anion acetate itself is inhibitory via increasing cytoplasmic osmotic pressure and inhibiting enzymes of methionine and protein synthesis, as well as inhibiting other unknown enzymes (Roe *et al*., 1998; Roe et al., 2002). Concentrations of inhibitors in lignocellulosic hydrolysates highly depend on pretreatment method and feedstock. Steam pretreated corn stover contained 11 g/l furfural, 0.06 g/l HMF, and 1.6 g/l acetic acid (Öhgren *et al*., 2006). Acetic acid levels have been found as high as 14.8 g/l in acid pretreated corn stover (Joiner, 2005). Similarly, acetic acid levels were 7 times higher in dilute acid pretreated corn stover than after ammonia fiber expansion pretreatment (Chundawat *et al*., 2010). Liquid hot water

pretreatment of corn stover produced 0.7 g/l furfural, 0.08 g/l HMF, and 2.2 g/l acetic acid (Cao *et al*., 2013). In sulfuric acid-catalyzed steam pretreated switchgrass, furfural was found from 0.8 - 1.4 g/100g and HMF at 2.1 g/100g (Ewanick and Bura, 2011). 1.8 - 2.4 g/100g acetate was present in hydrolysate from dilute acid pretreated switchgrass (Dien *et al*., 2006). Mild steam pretreatment of switchgrass was found to produce less than 0.04 g/l each of furfural, HMF, and acetic acid (Kumagai *et al*., 2015). Hydrothermolysis pretreatment of switchgrass produced less than 1 g/l of furfural and HMF (Suryawati *et al*., 2009). Pretreatment temperature affects inhibitor formation. Steam explosion of wheat straw at 190ºC resulted in the lowest inhibitor concentrations in one study, at 0.09 g/l furfural, 0.07 g/l HMF, and 0.04 g/l acetic acid, but lowered sugar yields by 26% compared to the optimal treatment for sugar yield at 200ºC, which yielded 0.70 g/l furfural, 0.26 g/l HMF, and 1.0 g/l acetic acid (Linde *et al*., 2008). Various dilute acid pretreatments of corn stover at 190ºC produced 3.0 - 11.5 g/l furfural and 0.2 - 2.7 g/l HMF, while at 170ºC produced 0.1 - 3.8 g/l furfural and 0.1 - 0.7 g/l HMF (Qin *et al*., 2012). Acetic acid levels at 190ºC ranged from 4.2 - 7.6 g/l and 1.2 - 3.0 g/l at 170ºC (Qin *et al*., 2012).

Hardwood and softwood waste products represent approximately 30% of the biomass available for biofuel production but are an as-of-yet commercially untapped sugar source due to a higher recalcitrance than other feedstocks (Perlack *et al*., 2005). The softwoods pine and spruce ranged from 0.6 - 2.3 g/l furfural, 0.4 - 5.9 g/l HMF, and 2.4 - 4.2 g/l

acetate after acid-catalyzed steam explosion pretreatment (Taherzadeh *et al*., 1997; Larsson *et al*., 1999). Dilute acid pretreated spruce contained from 0.5 - 1.0 g/l furfural, 2.0 - 5.9 g/l HMF, and 2.4 g/l acetic acid (Larsson *et al*., 1999; Nilvebrant *et al*., 2003). The hardwoods alder, aspen, and birch had, after acid-catalyzed steam explosion pretreatment, 0.3 - 3.5 g/l furfural, 0.2 - 2.6 g/l HMF, and 2.0 - 10.7 g/l acetate (Taherzadeh *et al*., 1997). Yat *et al*. (2008) found furfural formation in dilute acid treatment of basswood was strongly dependent on incubation time, beginning after maximum xylose accumulation, showing that careful monitoring of the reaction conditions can minimize its formation. The same has been shown in liquid hot water pretreatment of switchgrass, with 0.2 g/l furfural formed after 5 min. but 2.8 g/l formed after 10 min. (Garlock *et al*., 2011). A newer method, sulfite pretreatment to overcome recalcitrance of lignocellulose (SPORL), resulted in 0.4 - 4.2 g/l furfural and 0.2 - 2.4 g/l HMF from lodgepole pine (Zhu *et al*., 2010). SPORL of spruce reduced the acetic acid concentration to 2.7 g/l compared to the dilute acid treated sample at 5.3 g/l (Shuai *et al*., 2010).

Reports on the degree of inhibition of common industrial fermentative species to lignocellulosic hydrolysate inhibitors vary considerably depending of course on the strain used. To give some frame of reference to the degree of inhibition, ethanol yields of *Saccharomyces cerevisiae* were reduced by 43, 80, and 89% at 0.5, 1, and 2 g/l furfural and 71, 83, and 95% for 1, 3, and 5 g/l HMF (Delgenes *et al*., 1996). Ethanol yield

reductions for *Zymomonas mobilis* for the same concentrations of furfural were 4, 18, and 44% and 15, 13, and 53% for HMF (Delgenese *et al*., 1996). *S. cerevisiae* strain ATCC 211239 had no reduction in growth rate in up to 12 g/l furfural and 15 g/l HMF but had growth lag time increases of 24 and 8 h in 3 g/l and 1 g/l furfural and 45 and 4 hours in 7.5 and 1 g/l HMF (Liu *et al*., 2004; Almeida *et al*., 2007). Acetate (at pH 5.6) reduced ethanol yields of *S. cerevisiae* 1, 17, and 38% for 5, 10, and 15 g/l and 10, -2, and 17% for *Z. mobilis* (Delgenes *et al*., 1996). Acetate inhibition depends on pH, with the undissociated form acetic acid being more toxic. For example, ethanol fermentation by *Pichia stipitis* was inhibited 98% in the presence of 8 g/l acetic acid at pH 5.1 but only 25% at pH 6.5 (Pampulha and Loureiro, 1989; van Zyl *et al*., 1991).

Furans and acetic acid are not the only inhibitory compounds present in lignocellulosic hydrolysates. Phenolic compounds from lignin degradation are the third major category and include ferulic acid, coumaric acid, vanillin, syringaldehyde, and *p*-hydroxybenzaldehyde (Klinke *et al*., 2004). Besides acetic acid, other organic acids present are formic and levulinic acid (Almeida *et al*., 2007). Over 60 potential inhibitors have been identified (Olsson and Hahn-Hägerdal, 1996).

When combined the inhibitors show synergistic toxicity (Palmqvist *et al*., 1999; Liu *et al*., 2004; Piotrowski *et al*., 2014). Furthermore, the need for water recirculation results in inhibitor accumulation (Galbe and Zacchi, 1989; Olsson and Hahn-Hägerdal, 1996). Thus

numerous detoxification strategies have been developed (discussed below). As pretreatment methods must be aimed at reducing cost and thus the optimal pretreatment may be too expensive, so to must the ethanol fermentation yield improvement after detoxification overcome the added cost of an additional process step (Taherzadeh and Karimi, 2008; Pienkos and Zhang, 2009).

1.3 Furfural, 5-Hydroxymethylfurfural (HMF), and Acetate Detoxification Methods

Pretreatment techniques aimed at reducing inhibitor concentrations have been thoroughly researched at the laboratory- and pilot plant-scale. Chemical treatments include use of the reducing agents dithionite, dithiothreitol, or hydrogen sulfite (Alriksson *et al*., 2011; Soudham *et al*., 2011) and alkaline treatment (overliming) with calcium hydroxide, sodium hydroxide, or ammonium hydroxide (Alriksson *et al*., 2005; Mohagheghi *et al*., 2006). Alriksson *et al*. (2011) showed reducing agent treatment with a small amount of dithionite improved ethanol yields from dilute sulfuric acid with steam pretreated spruce wood chips and sugarcane bagasse of 76% and 61%, respectively, although the concentrations of furfural, HMF, and acetic acid were not significantly decreased. The beneficial effect was due to improvement in hydrolysis efficiency (Alriksson *et al*., 2011). Millati *et al*. (2002) showed overliming significantly reduced furfural and HMF concentrations but acetic acid levels remained unchanged. Mild alkaline deactylation has also been studied, with a 70% reduction in acetic acid from dilute acid pretreated corn

stover (Shekiro III *et al.*, 2016). Liquid-liquid extraction methods include use of ethyl acetate or trialkylamine (Cantarella *et al*., 2004; Zhu *et al*., 2011). Ethyl acetate extraction almost completely removes acetic acid but is likely a cost-prohibitive option (Aghazadeh *et al*., 2016; Kim, 2018). Zhu *et al*. (2011) used a mixture of trialkylamine, *n*-octanol, and kerosene to remove 100% of furfural, 45% of HMF, and 73% of acetic acid from sulfuric acid pretreated corn stover. Activated carbon and ion exchange are two liquid-solid extraction methods (Berson *et al*., 2005; Nilvebrant *et al.*, 2001). Activated carbon treatment of steam pretreated hardwoods was effective at removal of phenolic inhibitors but only moderately effective at furfural (5.5 to 3.8 g/l) and acetic acid (9.5 to 7.0 g/l) removal (Kim *et al*., 2013). Anion exchange resins decreased furfural up to 38%, HMF 29%, and acetic acid 72% in dilute acid pretreated spruce (Horváth *et al*., 2004). Organic membrane extraction with 15% alamine 336 in octanol removed 47% furfural, 94% HMF, and 97% acetic acid in dilute acid pretreated corn stover (Grzenia *et al*., 2012).

The major drawback to chemical detoxification is the additional processing cost; additional waste streams can also be generated (Pienkos and Zhang, 2009). Biological detoxification, in which a detoxifying microbe catabolizes inhibitors to remove them from the hydrolysate, has received attention as an alternative method, offering the advantages of mild process conditions and potentially lower cost (Lopez *et al*., 2004; Cao *et al*., 2013). The two major disadvantages are slow process time and sugar consumption

by the detoxifying strain (Jönsson and Martín, 2016). While acetate can be catabolized by most microbial species, most species do not possess furfural or HMF catabolic pathways. Larsson *et al*. (1999) found the fungus *Trichoderma reesei* could consume the furans as well as many phenolic inhibitors. Detoxification with *T. reesei* was, along with calcium hydroxide overliming, laccase treatment, and anion exchange, one of the most effective treatments in removing the furans and phenolic compounds, as well as acetic acid, from dilute acid spruce hydrolysate (Larsson *et al*., 1999). Nichols *et al*. (2005) identified a strain of the fungus *Coniochaeta ligniaria* that can catabolize and completely remove furfural and HMF from pretreated corn stover hydrolysate. Various *Coniochaeta* and *Lecythophora* species could also catabolize the lignocellulosic hydrolysate inhibitors levulinic acid and *p*-hydroxybenzaldehyde (Nichols *et al*., 2005). Sugar consumption by *C. ligniaria* is one problem that would need to be addressed for its utilization (Nichols *et al*., 2005; Cao *et al*., 2013). Schneider (1996) engineered a strain of *S. cerevisiae* with gene knockouts in hexokinase I, hexokinase II, and glucokinase I in order to eliminate glucose and mannose consumption and to focus catabolism to acetic acid by reducing glucose catabolite repression of secondary carbon sources induced by hexokinase II. Xia *et al*. (2012) engineered a *ptsG manZ glk crr xylA araA Escherichia coli* strain to eliminate consumption of glucose, xylose, and arabinose and increase catabolism of acetate. In both Schneider (1996) and Xia *et al*. (2012) sugar consumption was not completely eliminated, showing further study and modifications are required. Another strategy is to use microorganisms that natively cannot consume the primary sugars of

cellulose (glucose) and hemicellulose (primarily glucose, xylose, galactose, arabinose, and mannose). Wierckx *et al*. (2010) identified the bacterium *Cupriavidus basilensis* HMF14, which cannot consume glucose, xylose, arabinose, or mannose but can catabolize furfural and HMF, making it an attractive candidate detoxifying strain.

Engineering tolerance in the fermentative strain is another strategy to deal with the inhibitors. Engineering tolerance to the furans usually involves improving or overexpressing reductases that reduce furfural and HMF to their less toxic respective alcohols or dehydrogenases that oxidize them to their less toxic respective acids, as well as redox state engineering to deal with the subsequent redox imbalance (covered in Chapter 2; Heux *et al*., 2006; Heer *et al*., 2009; Ask *et al*., 2013). Due to the complexity of acetate metabolism, many options for engineering acetate tolerance have been explored (covered in Chapter 3). Similar to laboratory directed evolution to improve tolerance, adaptation by re-inoculation of the fermentative strain to successive real-world hydrolysates is another employed strategy (Amartey and Jeffries, 1996; Silva and Roberto, 2001; Sene *et al*., 2001).

1.4 Dissertation Synopsis

To engineer furfural and HMF degradation, the genome of a novel *Pseudomonas putida* isolate ALS1267, recently characterized for its robust growth on furfural and HMF (Lee *et al*., 2016), was sequenced and the pathway cloned. Additionally, to study engineering

furfural and HMF tolerance, the putative furfural/HMF dehydrogenase and furfuryl alcohol/HMF alcohol dehydrogenase genes from *P. putida* ALS1267 were cloned and characterized. To find engineering targets for increasing catabolism and tolerance to acetate, screening *E. coli* genomic libraries for improved growth on highly inhibitory acetate concentrations was employed. To further the field of engineering a detoxifying strain that cannot consume glucose, glucose minus revertants from one such strain (a *ptsG manZ glk E. coli* strain) were identified, characterized, and an additional gene knockout made to improve the strain.

REFERENCES

Almeida JR, Modig T, Petersson A, Hähn-Hägerdal B, Lidén G, Gorwa-Grauslund MF. 2007. Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. J Chem Technol Biot. 82:340-9.

Alriksson B, Cavka A, Jönsson LJ. 2011. Improving the fermentability of enzymatic hydrolysates of lignocellulose through chemical in-situ detoxification with reducing agents. Bioresour Technol. 102:1254-63.

Alriksson B, Horváth IS, Sjöde A, Nilvebrant NO, Jönsson LJ. 2005. Ammonium hydroxide detoxification of spruce acid hydrolysates. In: Twenty-Sixth Symposium on Biotechnology for Fuels and Chemicals. Humana Press. 911-22.

Amartey S, Jeffries T. 1996. An improvement in *Pichia stipitis* fermentation of acid-hydrolyzed hemicellulose achieved by overliming (calcium hydroxide treatment) and strain adaptation. World J Microbiol Biotechnol. 12:281-3.

Ask M, Bettiga M, Mapelli V, Olsson L. 2013. The influence of HMF and furfural on redox-balance and energy-state of xylose-utilizing *Saccharomyces cerevisiae*. Biotechnol Biofuels. 6:22.

Banerjee N, Bhatnagar R, Viswanathan L. 1981. Inhibition of glycolysis by furfural in *Saccharomyces cerevisiae*. Eur J Appl Microbiol Biotechnol. 11:226-8.

Berson RE, Young JS, Kamer SN, Hanley TR. 2005. Detoxification of actual pretreated corn stover hydrolysate using activated carbon powder. Appl Biochem Biotechnol. 124:923-34.

Cantarella M, Cantarella L, Gallifuoco A, Spera A, Alfani F. 2004. Comparison of different detoxification methods for steam-exploded poplar wood as a substrate for the bioproduction of ethanol in SHF and SSF. Process Biochem. 39:1533-42.

Cao G, Ximenes E, Nichols NN, Zhang L, Ladisch M. 2013. Biological abatement of cellulase inhibitors. Bioresour Technol. 146:604-10.

Cherubini F, Bird ND, Cowie A, Jungmeier G, Schlamadinger B, Woess-Gallasch S. 2009. Energy-and greenhouse gas-based LCA of biofuel and bioenergy systems: Key issues, ranges and recommendations. Resour Conserv Recy. 53:434-47.

Chundawat SP, Vismeh R, Sharma LN, Humpula JF, da Costa Sousa L, Chambliss CK, Jones AD, Balan V, Dale BE. 2010. Multifaceted characterization of cell wall decomposition products formed during ammonia fiber expansion (AFEX) and dilute acid based pretreatments. Bioresour Technol. 101:8429-38.

de Vries SC, van de Ven GW, van Ittersum MK, Giller KE. 2010. Resource use efficiency and environmental performance of nine major biofuel crops, processed by first-generation conversion techniques. Biomass Bioenergy. 34:588-601.

Delgenes JP, Moletta R, Navarro JM. 1996. Effects of lignocellulose degradation products on ethanol fermentations of glucose and xylose by *Saccharomyces cerevisiae*, *Zymomonas mobilis*, *Pichia stipitis*, and *Candida shehatae*. Enzyme Microb Technol. 19:220-5.

Dien BS, Jung HJ, Vogel KP, Casler MD, Lamb JF, Iten L, Mitchell RB, Sarath G. 2006. Chemical composition and response to dilute-acid pretreatment and enzymatic saccharification of alfalfa, reed canarygrass, and switchgrass. Biomass Bioenergy. 30:880-91.

Ewanick S, Bura R. 2011. The effect of biomass moisture content on bioethanol yields from steam pretreated switchgrass and sugarcane bagasse. Bioresour Technol. 102:2651-8.

Galbe M, Zacchi G. 1993. Simulation of processes for conversion of lignocellulosics. Biotechnology in Agriculture. 291.

Garlock RJ, Balan V, Dale BE, Pallapolu VR, Lee YY, Kim Y, Mosier NS, Ladisch MR, Holtzapple MT, Falls M, Sierra-Ramirez R. 2011. Comparative material balances around pretreatment technologies for the conversion of switchgrass to soluble sugars. Bioresour Technol. 102:11063-71.

Gelfand I, Sahajpal R, Zhang X, Izaurralde RC, Gross KL, Robertson GP. 2013. Sustainable bioenergy production from marginal lands in the US Midwest. Nature. 493:514.

Grzenia DL, Wickramasinghe SR, Schell DJ. 2012. Fermentation of reactive membrane-extracted and ammonium hydroxide-conditioned dilute-acid-pretreated corn stover. Appl Biochem Biotechnol. 166:470-8.

Hamelinck CN, Faaij AP. 2006. Outlook for advanced biofuels. Energy Policy. 34:3268-83.

Heer D, Heine D, Sauer U. 2009. Resistance of *Saccharomyces cerevisiae* to high concentrations of furfural is based on NADPH-dependent reduction by at least two oxidoreductases. Appl Environ Microbiol. 75:7631-8.

Heux S, Cachon R, Dequin S. 2006. Cofactor engineering in *Saccharomyces cerevisiae*: expression of a H2O-forming NADH oxidase and impact on redox metabolism. Metab Eng. 8:303-14.

Horváth IS, Sjöde A, Nilvebrant NO, Zagorodni A, Jönsson LJ. 2004. Selection of anion exchangers for detoxification of dilute-acid hydrolysates from spruce. Appl Biochem Biotechnol. 114:525-38.

EPA. 2013. Renewable Fuels: Regulations & Standards. http://www.epa.gov/otaq/fuels/renewablefuels/regulations.htm.

EPA. 2018. Proposed Volume Standards for 2018, and the Biomass-Based Diesel Volume for 2019. https://www.epa.gov/renewable-fuel-standard-program/proposed-volume-standards-2018-and-biomass-based-diesel-volume-2019.

Jackson RW, Neto AB, Erfanian E. 2018. Woody biomass processing: Potential economic impacts on rural regions. Energy Policy. 115:66-77.

Jenkins R, Alles C. 2011. Field to fuel: developing sustainable biorefineries. Ecol Appl. 21:1096-104.

Joiner D. 2005. Cellulose digestibility, ethanol yield, and lignin recovery from corn stover fractionated by a two-stage dilute-acid and dilute-alkaline process (Doctoral dissertation).

Jönsson LJ, Martín C. 2016. Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects. Bioresour Technol. 199:103-12.

Kim Y, Kreke T, Hendrickson R, Parenti J, Ladisch MR. 2013. Fractionation of cellulase and fermentation inhibitors from steam pretreated mixed hardwood. Bioresour Technol. 135:30-8.

Klinke HB, Thomsen AB, Ahring BK. 2004. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Appl Microbiol Biotechnol. 66:10-26.

Kumagai A, Wu L, Iwamoto S, Lee SH, Endo T, Rodriguez Jr M, Mielenz JR. 2015. Improvement of enzymatic saccharification of *Populus* and switchgrass by combined pretreatment with steam and wet disk milling. Renew Energy. 76:782-9.

Kumar D, Murthy GS. 2011. Impact of pretreatment and downstream processing technologies on economics and energy in cellulosic ethanol production. Biotechnol Biofuels. 4:27.

Larsson S, Reimann A, Nilvebrant NO, Jönsson LJ. 1999. Comparison of different methods for the detoxification of lignocellulose hydrolyzates of spruce. Appl Biochem Biotechnol. 77:91-103.

Lee SA, Wrona LJ, Cahoon AB, Crigler J, Eiteman MA, Altman E. 2016 Isolation and characterization of bacteria that use furans as the sole carbon source. Appl Biochem Biotechnol. 178:76-90.

Linde M, Jakobsson EL, Galbe M, Zacchi G. 2008. Steam pretreatment of dilute H_2SO_4-impregnated wheat straw and SSF with low yeast and enzyme loadings for bioethanol production. Biomass Bioenergy. 32:326-32.

Liu ZL, Slininger PJ, Dien BS, Berhow MA, Kurtzman CP and Gorsich SW. 2004. Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2,5-bis-hydroxymethylfuran. J Ind Microbiol Biotechnol 31:345-52.

López MJ, Nichols NN, Dien BS, Moreno J, Bothast RJ. 2004. Isolation of microorganisms for biological detoxification of lignocellulosic hydrolysates. Appl Microbiol Biotechnol. 64:125-31.

Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. Curr Opin Biotechnol. 45:202-11.

McLaughlin SB, Kszos LA. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass Bioenergy. 28:515-35.

Millati R, Niklasson C, Taherzadeh MJ. 2002. Effect of pH, time and temperature of overliming on detoxification of dilute-acid hydrolyzates for fermentation by *Saccharomyces cerevisiae*. Process Biochem. 38:515-22.

Mills TY, Sandoval NR, Gill RT. 2009. Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. Biotechnol Biofuels. 2:26.

Modig T, Liden G, Taherzadeh MJ. 2002. Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. Biochem J. 363:769-76.

Nielsen RL, 2017. Historical Corn Grain Yields for the U.S. https://www.agry.purdue.edu/ext/corn/news/timeless/yieldtrends.html.

Nichols NN, Dien BS, Guisado GM, López MJ. 2005. Bioabatement to remove inhibitors from biomass-derived sugar hydrolysates. Appl Biochem Biotechnol. 121:379-90.

Nilvebrant N-O, Reimann A, Larsson S, Jönsson LJ. 2001. Detoxification of lignocellulose hydrolysates with ion-exchange resins. Appl Biochem Biotechnol. 91:35-49.

Nilvebrant NO, Persson P, Reimann A, De Sousa F, Gorton L and Jönsson LJ. 2003. Limits for alkaline detoxification of dilute-acid lignocellulose hydrolysates. Appl Biochem Biotechnol. 107:615-28.

Ohlrogge J, Allen D, Berguson B, DellaPenna D, Shachar-Hill Y, Stymne S. 2009. Driving on biomass. Science. 324:1019-20.

Okuda N, Soneura M, Ninomiya K, Katakura Y, Shioya S. 2008. Biological detoxification of waste house wood hydrolysate using *Ureibacillus thermosphaericus* for bioethanol production. J Biosci Bioeng. 106:128-33.

Olsson L, Hahn-Hägerdal B. 1996. Fermentation of lignocellulosic hydrolysates for ethanol production. Enzyme Microb Technol. 18:312-31.

Palmqvist E, Grage H, Meinander NQ, Hahn-Haegerdal B. 1999. Main and interaction effects of acetic acid, furfural, and *p*-hydroxybenzoic acid on growth and ethanol productivity of yeasts. Biotechnol Bioeng. 63:46-55.

Palmqvist E, Hahn-Hägerdal B. 2000. Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. Bioresour Technol. 74:25-33.

Pampulha ME, Loureiro-Dias MC. 1989. Combined effect of acetic acid, pH and ethanol on intracellular pH of fermenting yeast. Appl Microbiol Biotechnol. 31:547-50.

Perlack RD, Wright LL, Turhollow AF, Graham RL, Stokes BJ, Erbach DC. 2005. Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply. ORNL.

Pienkos PT, Zhang M. 2009. Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates. Cellulose. 16:743-62.

Pimentel D. 2003. Ethanol fuels: energy balance, economics, and environmental impacts are negative. Nat Resour Research. 12:127-34.

Piotrowski JS, Zhang Y, Sato T, Ong I, Keating D, Bates D, Landick R. 2014. Death by a thousand cuts: the challenges and diverse landscape of lignocellulosic hydrolysate inhibitors. Front Microbiol. 14:90.

Plevin RJ, Jones AD, Torn MS, Gibbs HK. 2010. Greenhouse gas emissions from biofuels' indirect land use change are uncertain but may be much greater than previously estimated. Environ Sci Technol. 44:8015-21.

Qin L, Liu ZH, Li BZ, Dale BE, Yuan YJ. 2012. Mass balance and transformation of corn stover by pretreatment with different dilute organic acids. Bioresour Technol. 112:319-26.

Roe AJ, McLaggan D, Davidson I, O'Byrne C, Booth IR. 1998. Perturbation of anion balance during inhibition of growth of *Escherichia coli* by weak acids. J Bacteriol. 180:767-72.

Roe AJ, O'Byrne C, McLaggan D, Booth IR. 2002. Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology. 148:2215-22.

Schneider H. 1996. Selective removal of acetic acid from hardwood-spent sulfite liquor using a mutant yeast. Enzyme Microb Technol. 19:94-8.

Schubert C. 2006. Can biofuels finally take center stage?. Nature Biotechnol. 24:777.

Searchinger TD. 2010. Biofuels and the need for additional carbon. Environ Res Lett. 5:024007.

Sene L, Converti A, Zilli M, Felipe M, Silva S. 2001. Metabolic study of the adaptation of the yeast *Candida guilliermondii* to sugarcane bagasse hydrolysate. Appl Microbiol Biotechnol. 57:738-43.

Shekiro III J, Chen X, Smith H, Tucker MP. 2016. Development and characterization of a high-solids deacetylation process. Sustainable Chemical Processes. 4:6.

Shuai L, Yang Q, Zhu JY, Lu FC, Weimer PJ, Ralph J, Pan XJ. 2010. Comparative study of SPORL and dilute-acid pretreatments of spruce for cellulosic ethanol production. Bioresour Technol. 101:3106-14.

Silva CJ, Roberto IC. 2001. Improvement of xylitol production by *Candida guilliermondii* FTI 20037 previously adapted to rice straw hemicellulosic hydrolysate. Lett Appl Microbiol. 32:248-52.

Somerville C. 2008. Development of cellulosic biofuels. J Agricultur. Hal. 21-7.

Soudham VP, Alriksson B, Jönsson LJ. 2011. Reducing agents improve enzymatic hydrolysis of cellulosic substrates in the presence of pretreatment liquid. J Biotechnol. 155:244-50.

Suryawati L, Wilkins MR, Bellmer DD, Huhnke RL, Maness NO, Banat IM. 2009. Effect of hydrothermolysis process conditions on pretreated switchgrass composition and ethanol yield by SSF with *Kluyveromyces marxianus* IMB4. Proc Biochem. 44:540-5.

Taherzadeh MJ, Eklund R, Gustafsson L, Niklasson C, Lidén G. 1997. Characterization and fermentation of dilute-acid hydrolyzates from wood. Ind Eng Chem Res. 36:4659-65.

Taherzadeh MJ, Karimi K. 2008. Pretreatment of lignocellulosic wastes to improve ethanol and biogas production: a review. Int J Mol Sci. 9:1621-51.

Tilman D, Hill J, Lehman C. 2006. Carbon-negative biofuels from low-input high-diversity grassland biomass. Science. 314:1598-600.

Tyner WE. 2008. The US ethanol and biofuels boom: Its origins, current status, and future prospects. AIBS Bulletin. 58:646-53.

USDA ERS. 2018. U.S. BioEnergy Statistics. https://www.ers.usda.gov/data-products/us-bioenergy-statistics/.

Valdivia M, Galan JL, Laffarga J, Ramos JL. 2016. Biofuels 2020: Biorefineries based on lignocellulosic materials. Microb Biotechnol. 9:585-94.

van Zyl C, Prior BA, du Preez JC. 1991. Acetic acid inhibition of D-xylose fermentation by *Pichia stipitis*. Enzy Microb Technol. 13:82-6.

Wierckx N, Koopman F, Bandounas L, De Winde JH, Ruijssenaars HJ. 2010. Isolation and characterization of *Cupriavidus basilensis* HMF14 for biological removal of inhibitors from lignocellulosic hydrolysate. Microb Biotechnol. 3:336-43.

Xia T, Eiteman MA, Altman E. 2012. Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of *Escherichia coli* strains. Microb Cell Fact. 11:77.

Yang B, Wyman CE. 2008. Pretreatment: the key to unlocking low-cost cellulosic ethanol. Biofuels, Bioprod Biorefin. 2:26-40.

Yat SC, Berger A, Shonnard DR. 2008. Kinetic characterization for dilute sulfuric acid hydrolysis of timber varieties and switchgrass. Bioresour Technol. 99:3855-63.

Zhang Y, Goldberg M, Tan E, Meyer PA. 2016. Estimation of economic impacts of cellulosic biofuel production: a comparative analysis of three biofuel pathways. Biofuels, Bioprod Biorefin. 10:281-98.

Zheng Y, Pan Z, Zhang R. 2009. Overview of biomass pretreatment for cellulosic ethanol production. Int J Agric Biol Eng. 2:51-68.

Zhu J, Yong Q, Xu Y, Yu S. 2011. Detoxification of corn stover prehydrolyzate by trialkylamine extraction to improve the ethanol production with *Pichia stipitis* CBS 5776. Bioresour Technol. 102:1663-8.

Zhu JY, Zhu W, OBryan P, Dien BS, Tian S, Gleisner R, Pan XJ. 2010. Ethanol production from SPORL-pretreated lodgepole pine: preliminary evaluation of mass balance and process energy efficiency. Appl Microbiol Biotechnol. 86:1355-65.

CHAPTER 2. SEQUENCING, CLONING, AND CHARACTERIZATION OF THE FURFURAL CATABOLIC PATHWAY OF NOVEL ISOLATE *PSEUDOMONAS PUTIDA* ALS1267

Parts of this chapter have been published in Applied Biochemistry and Biotechnology:

Lee SA, Wrona LJ, Cahoon AB, Crigler J, Eiteman MA, Altman E. 2016. Isolation and characterization of bacteria that use furans as the sole carbon source. Appl Biochem Biotechnol. 178:76-90.

Manuscript in preparation:

Crigler J, Eiteman MA, Altman E. Sequencing and cloning the furfural catabolic pathway from the novel isolate *Pseudomonas putida* ALS1267.

2.1 Abstract

Gene assignments for the aerobic bacterial catabolism of furfural and 5-hydroxymethylfurfural (HMF) have been made in two species, *Pseudomonas putida* Fu1 and *Cupriavidus basilensis* HMF14. A third furfural- and HMF-metabolizing strain, *P. putida* ALS1267, was recently characterized. In this work, the genome of *P. putida* ALS1267 was sequenced. In one contiguous 18 kb sequence, *P. putida* ALS1267 contained homologs of each gene in each furfural and HMF cluster of *P. putida* Fu1 and *C. basilensis* HMF14 excluding HMF acid oxidase (*hmfH*) and an unessential gene *hmfH'*. The 18 kb pathway was cloned into *P. putida* KT2440, which cannot metabolize the furans, enabling growth on furfural as the sole carbon source at 0.34/h. The clone was not able to metabolize HMF, most likely due to the lack of *hmfH*. No putative HmfH ortholog in *P. putida* ALS1267 was identified but candidate genes for this function were found. Additionally, the putative furfuryl alcohol/HMF alcohol dehydrogenase (PsfG) and furfural/HMF dehydrogenase (PsfA) of *P. putida* ALS1267, previously identified in *P. putida* Fu1, were cloned to test for conferrance of furfural and HMF tolerance. Neither a *psfA*, *psfG*, nor *psfGA* clone improved tolerance to furfural, furfuryl alcohol. The 18 kb furfural cluster clone improved tolerance to furfuryl alcohol but not furfural.

2.2 Introduction

Furfural and 5-hydroxymethylfurfural (HMF) occur in soils, generated abiotically via sugar dehydration (Huber *et al*., 2010; Cheshire *et al*., 1992). They also are human-made,

from biomass treatments that include burning or acid (Zeitsch, 2000; Christian *et al*., 2003). The list of species known to catabolize furfural and other furans is small (Wierckx *et al*., 2011), but this is a reflection on the amount of study it has been given, as these organisms are expected to be found ubiquitously in decaying plant matter. The interest in the study of microbial degradation of furans arises from the presence of furfural and HMF in pretreated lignocellulosic biomass for green fuel production, which if over a certain concentration can become inhibitory to the fermentative microorganism, resulting in lower yields of ethanol or other fuels. Additionally, an HMF intermediate, 2,5-furan-dicarboxylic acid (FDCA), can be used as a platform chemical in polyester production, and research into producing it biologically as a green alternative is underway (Koopman *et al*., 2010; Hossain *et al*., 2017; Yang and Huang, 2017; Karich *et al*., 2018; Yuan *et al*., 2018).

Aerobic bacterial furfural and HMF degradation converge at the 2-furoic acid pathway, which proceeds via CoA intermediates (Figure 2.1) (Koenig and Andreesen, 1990; Koopman *et al*., 2010; Trudgill, 1969). 2-furoic acid is first converted to 2-furoyl-CoA by furoyl-CoA synthetase (HmfD) and then is hydroxylated to 5-hydroxy-2-furoyl-CoA by furoyl-CoA dehydrogenase, formed from the subunits HmfABC (Koenig and Andreesen, 1989; Koopman *et al*., 2010). After tautomerization of 5-hydroxy-2-furoyl-CoA to the keto form, the lactone is hydrolyzed to 2-oxoglutaroyl-CoA either spontaneously or by an unidentified enzyme (Koopman *et al*., 2010). A tautomerization to the keto form of 2-oxoglutaroyl-CoA occurs, and it is then hydrolyzed by HmfE to produce the TCA cycle intermediate 2-oxoglutarate (alpha-ketoglutarate) (Koopman *et al*., 2010).

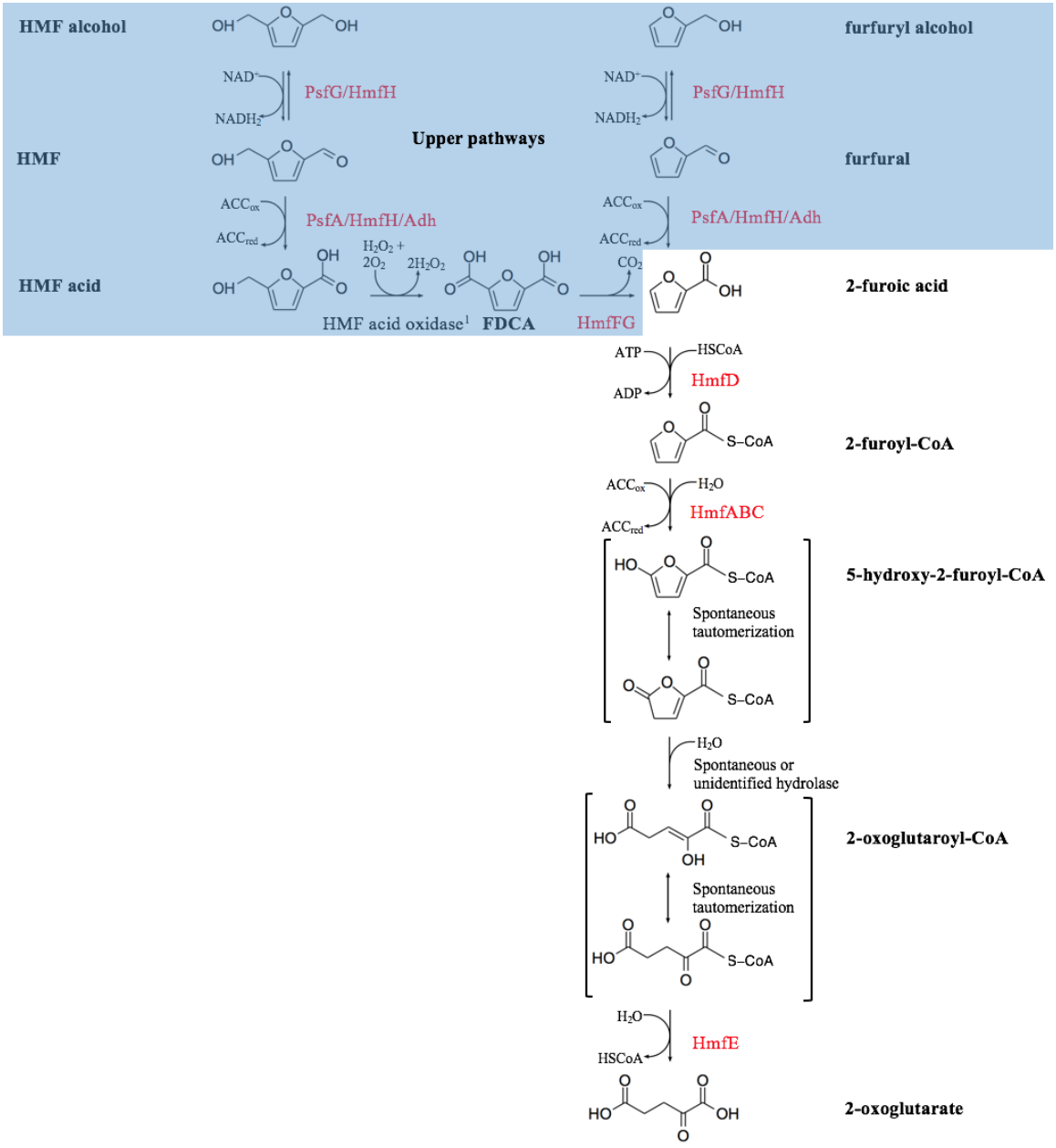

Figure 2.1. Furfural and HMF degradation pathway
Adapted from Koopman *et al*. (2010), Nichols *et al*. (2008), and Wiercke *et al*. (2011). The "upper pathway" reactions, the pathways leading to 2-furoic acid, are highlighted in blue.
[1] HmfH performs this step in *C. basilensis* HMF14 in a reaction requiring O_2 (Koopman *et al*., 2010).

In the "upper pathway" reactions, furfuryl alcohol and HMF alcohol, also called furan dimethanol or 2,5-bis(hydroxymethyl)furan, are oxidized to furfural and HMF, respectively. Furfural is oxidized directly to 2-furoic acid, and HMF is converted to 2-furoic acid (hereafter called furoic acid) in a three step pathway, first via oxidation to HMF acid by either HmfH in a reaction requiring oxygen or by a nonspecific aldehyde dehydrogenase, then oxidation to 2,5-furan-dicarboxylic acid (FDCA), catalyzed by HmfH, and then decarboxylation to furoic acid by HmfFG (Koopman *et al*., 2010). In *Methylovorus* sp. strain MP688, HMF conversion to FDCA proceeds via 5-formyl-2-furancarboxylic acid (FFA), also performed by an oxidase, HMFO (Dijkman and Fraaije, 2014). HmfH can perform each of the four upper pathway oxidations (Koopman *et al*., 2010). Similarly, HMFO has a wide substrate range, including HMF but not furfural (Dijkman and Fraaije, 2014). Additionally, the host into which the pathway has been previously cloned (*P. putida* S12), which cannot metabolize the furans, possesses native furfuryl alcohol, furfural, HMF alcohol, and HMF oxidation ability (Koopman *et al*., 2010). In *Pseudomonas putida* Fu1, an alcohol dehydrogenase (PsfG) and an aldehyde dehydrogenase (PsfA) were found in the furoic acid sequence cluster (Nichols *et al*., 2008). They were shown to be induced by furoic acid but their roles in the upper pathway reactions were not investigated.

Recently, a novel strain, *P. putida* ALS1267, was isolated from a wastewater treatment facility, selected and characterized for its robust growth on furfural as a sole carbon source (Table 2.1 and Figure 2.2) (Lee *et al*., 2016). When compared to two of the

Table 2.1. Furfural dehydrogenase activity measured from whole cell extracts

Strain	Furfural dehydrogenase activity (mIU/mg)
E. coli MG1655	0.0 ± 0.4
P. putida KT2440	72 ± 12
B. phytofirmans PsJN	26 ± 12
ALS1131	0.8 ± 0.2
ALS1172	1.0 ± 04
ALS1267	170 ± 59
ALS1279	2.1 ± 1.6
ALS1280	2.3 ± 1.4

From Lee *et al*. (2016). ALS1131, 1172, 1267, 1279, and 1280 were isolated from wastewater treatment facilities and selected for growth on furfural as the sole carbon source. ALS1131 was identified as *Pseudomonas mendocina*, ALS1172 *Pigmentiphaga* spp., ALS1267 *Pseudomonas putida*, ALS1279 *Pseudomonas* spp. BWDY9, and ALS1280 *Cupriavidus pinatubonensis*.

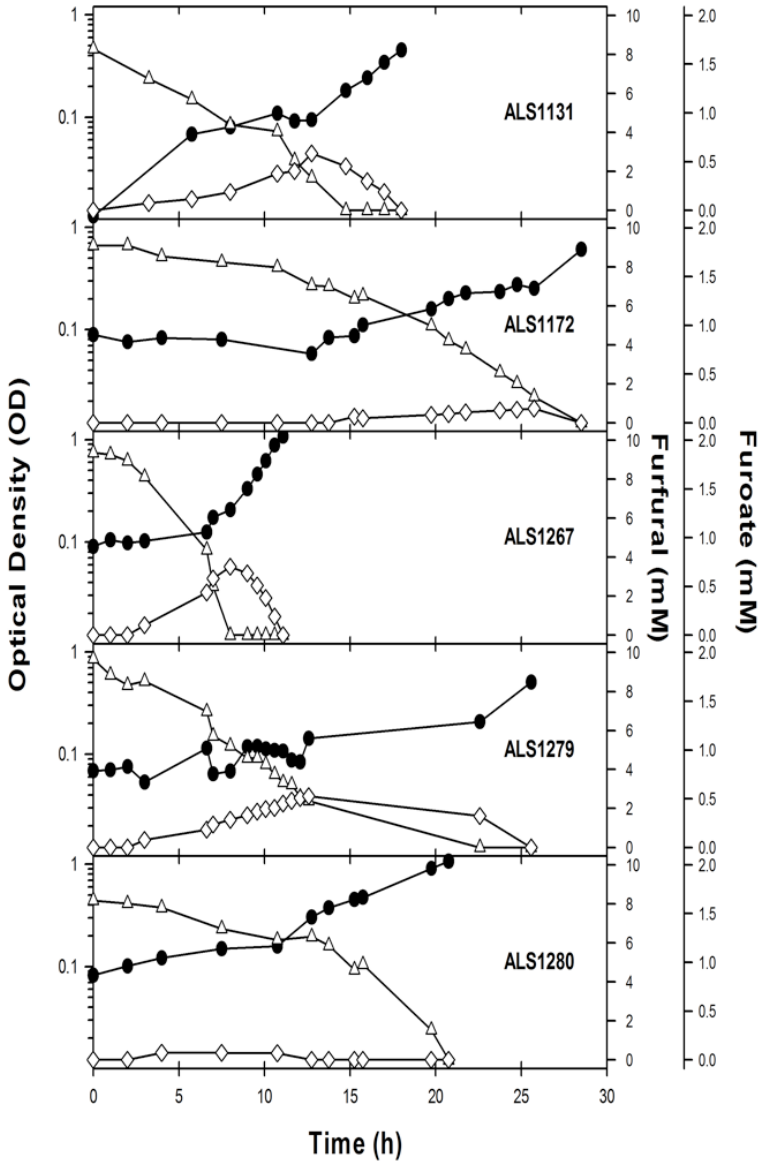

Figure 2.2. Growth of *P. putida* ALS1267 compared to four other isolates that can metabolize furfural as the sole carbon source

From Lee *et al*. (2016). Bioreactor experiments were performed in 9 mM minimal furfural medium as described in Lee *et al*. (2016). ● is optical density, △ is furfural concentration, and ◇ is furoic acid (furoate) concentration.

available, characterized aerobic furfural-metabolizing species, *P. putida* Fu1 and *Burkholderia phytofirmans* PsJN, *P. putida* ALS1267 had a 2-fold and 3-fold higher tolerance and a 170-fold and 6-fold higher furfural oxidation rate, respectively. It also metabolized HMF at a high rate. In this work, the genome of *P. putida* ALS1267 was sequenced and the pathway cloned into the heterologous host *P. putida* KT2440.

Furfural and HMF are toxic compounds at a high enough concentration, the amount depending of course on species and strain. Furfuryl alcohol and HMF alcohol are less toxic than furfural and HMF (Table 2.2). Additionally, furoic acid is less inhibitory than furfuryl alcohol. Many organisms that cannot metabolize furfural or HMF, including *Escherichia coli*, *Corynebacterium glutamicum*, and *Saccharomyces cerevisiae*, have nonspecific dehydrogenases that reduce furfural and HMF to their respective alcohols and/or oxidize them to their respective acids to detoxify them (Boopathy *et al*., 1993; Gutiérrez *et al*., 2002; Taherzadeh *et al*., 1999; Tsuge *et al*., 2014). Thus it was investigated if cloning either the putative furfural/HMF dehydrogenase or furfuryl alcohol/HMF alcohol dehydrogenase of *P. putida* ALS1267 would confer tolerance to furfural and HMF.

Table 2.2. Inhibition of furfural, furoic acid, and furfuryl alcohol

Strain	mM Furfural	mM Furoic acid	mM Furfuryl alcohol
P. putida KT2440	22	> 100	50
P. aeruginosa PAO1	18	> 100	46
E. coli MG1655	44	> 100	64

Numbers listed are mM furfural, furoic acid, or furfuryl alcohol in LB medium that lead to an OD_{600} of less than 0.05 after 24 h of growth. Inoculation was done at a 1/200 dilution from LB overnight cultures. Furfural, furoic acid, and furfuryl alcohol stocks were at pH 7.0. Measurements done in 2 mM increments.

2.3 Materials and Methods

2.3.1 Bacterial strains and growth conditions

Bacterial strains used in this study are listed in Table 2.3. Lysogeny broth (LB) was used for rich culture and MM minimal medium used was M9 medium (Miller, 1972) supplemented with (in mg/L): 35 $(NH_4)_6Mo_7O_{24}\cdot 4H_2O$, 247 H_3BO_3, 71 $CoCl_2$, 24 $CuSO_4$, 158 $MnCl_2$, 28 $ZnSO_4$, and 13 $FeSO_4\cdot 7H_2O$. All *Pseudomonas* spp. and *B. phytofirmans* PsJN were grown at 30ºC. *E. coli* MG1655 and *S. enterica* LT2 were grown at 37ºC. When cloned with the furfural degradation pathway from *P. putida* ALS1267, *P. putida* KT2440 and *P. aeruginosa* PAO1, *E. coli* MG1655, and *S. enterica* LT2 were tested at both 30 and 37ºC. Furoic acid, furfural, and HMF stocks were at pH 7.0 and added to MM medium at 10 mM as indicated. 40 µg/ml kanamycin was used in rich media for selection of all bacterial strains used with kanamycin-resistant plasmids. In MM minimal media, 40 µg/ml kanamycin was used for *Pseudomonas* spp. and 20 µg/ml for *E. coli* and *S. enterica*. Tetracycline was used at 40 µg/ml in rich and minimal medium for all *Pseudomonas* spp. containing tetracycline-resistant plasmids.

2.3.2 Genome sequencing and identification of the furfural degradation pathway of *P. putida* ALS1267

A Nextera XT paired-end library of *P. putida* ALS1267 gDNA was sequenced on a MiSeq system using the MiSeq v2 500 Cycle Reagent Kit. CLC Genomics Workbench version 10.1.1 was used for *de novo* assembly. The three previously identified operons involved in furoic acid, furfural, and/or HMF metabolism, EU290170, GU556182, and

Table 2.3. Bacterial strains used in this study

Strain	Reference
P. putida ALS1267	Lee *et al*., 2016
E. coli MC1061 (DSM-7140)	Laboratory stock
P. putida KT2440 (DSM-6125)	Trevisan, 1889
P. putida KT2440-pBBR1MCS(Kan)'fur	This study
P. aeruginosa PAO1 (DSM-22644)	Schroeter, 1872
P. aeruginosa PAO1-pMP220'fur	This study
E. coli MG1655 (DSM-18039)	Guyer *et al*., 1981
E. coli MG1655-pBBR1MCS(Kan)'fur	This study
S. enterica LT2 (DSM-17058)	Le Minor and Popoff, 1987
S. enterica LT2-pBBR1MCS(Kan)'fur	This study
P. putida Fu1 (DSM-6384)	Koenig and Andreesen, 1989
B. phytofirmans PsJN (DSM-17436)	Sessitsch *et al*., 2005
P. putida KT2440-pBBR1MCS(Kan)'*psfGAB*	This study
P. putida KT2440-pBBR1MCS(Kan)'*psfA*-P	This study
P. putida KT2440-pBBR1MCS(Kan)'*psfA*	This study
P. putida KT2440-pBBR1MCS(Kan)'*psfG*	This study
P. putida KT2440-pBBR1MCS(Kan)'*psfGA*	This study

'fur contains *hmfABCDEIFT-psfGAB-hmfG*-hydrolase-*psfD*-araC from *P. putida* ALS1267. '*psfGAB* contains *hmfIF*T-*psfGAB*. '*psfA*-P contains *psfA* with the native *hmfA* promoter. Control strains of each wild-type parent were also made containing the respective empty cloning vector.

GU556183, were used for bla stn searches of *P. putida* ALS1267 contigs. Contigs were annotated using both Prokka v1.12 and RAST v1.0.1, using the KBase platform.

2.3.3 Cloning the furfural degradation pathway and putative furfural/HMF dehydrogenase and furfuryl alcohol/HMF alcohol dehydrogenase of *P. putida* ALS1267

E. coli MC1061 (DSM-7140), a high transformation efficiency strain, was used for all intermediate cloning transformations. A clone in pBBR1MCS-3 and a kanamycin-resistant derivative of pBBR1MCS-3, designated pBBR1MCS(Kan), was constructed containing *P. putida* ALS1267 genes *hmfABCDEIFT-psfGAB-hmfG*-hydrolase-*psfD*-araC, along with 142 bp upstream of *hmfA* and 38 bp downstream from araC, abbreviated 'fur. A BglII BsrGI AscI SpeI restriction enzyme linker sequence was first added to pBBR1MCS-3 using the primers (forward primer abbreviated "F", reverse primer abbreviated "R", and, if applicable, the restriction enzyme(s) used in parentheses and restriction site(s) underlined in the primer sequence) F(XbaI BglII BsrGI AscI SpeI Kpn1): 5'-AAGAAGATTCTAGAAGATCTTTGCTAGCTGTACAAACTTAAGGGCGCGCCAAATGCATACTAGTGGTACCAAGAAGAA-3' and R(KpnI SpeI AscI BsrGI BglII XbaI): 5'-TTCTTCTTGGTACCACTAGTATGCATTTGGCGCGCCCTTAAGTTTGTACAGCTAGCAAAGATCTTCTAGAATCTTCTT-3' by boiling and then cooling to anneal the two primers and then cloning the restriction enzyme linker into pBBR1MCS-3 with XbaI

and Kpn1.The kanamycin resistance gene *kanR* from Tn5 was cloned into the NheI site of pBBR1MCS-3 using F(NheI): 5'-AATCAACTGCTAGCCAAGCGCAAAGAGAAAGCAG-3' and R(NheI): 5'-TCAACTAAGCTAGCGCTCAGAAGAACTCGTCAAG-3'. The 'fur clone was constructed in a three step PCR cloning using native BsrGI and AscI restriction enzyme sites in the insert, using the primers: 'furF1(BglII): 5'-TTCTTCTTAGATCTCTGTCCTGAAAATCAGACGC-3', 'furR1: 5'-TTTCTTCTTCTTTCAAAGGTTTTGCGCTACGTGG-3', 'furF2: 5'-CGAAGACTTTTCACTGCTGC-3', 'furR2: 5'-TGTTGTTGCGCACGATATCC-3', 'furF3: TTTCTTCTTCTTTCAGATCGTCAACATTGCCTCG-3', and 'furR3 (SpeI): 5'-TTCTTCTTACTAGTTGGGATAGTTGCCCAATACC-3'. 'fur was subcloned into pMP220 for replication in *P. aeruginosa* PAO1, with the BglII site of pMP220 first removed using Klenow and then self-ligating the vector.

hmfABCDE appears expressed as a single unit, driven by the promoter of *hmfA*. There are intergenic sequences with possible promoters upstream of *hmfI*, *hmfF*, and *hmfT*. To clone *psfG* and *psfA* without *hmfABCDE* in a construct that could potentially be expressed by a native promoter as well as by the lac promoter of pBBR1MCS(Kan), a construct containing 158 bp upstream of the *hmfI* up until the start of *psfC* was constructed, containing *hmfIFT-psfGAB*, abbreviated '*psfGAB*, using F(KpnI): 5'-AACAACAAGGTACCAACGGTGTGCGCCCCTA-3' and R(XbaI): 5'-TTCCTTCCTCTAGATGCGGCCGGCTATTACC-3'.

psfA was cloned with the *hmfA* promoter added to pBBR1MCS(Kan), abbreviated '*psfA*-P. The primers to add the *hmfA* promoter were F(KpnI): 5'-ATATATATATATATATATATATTGGTACCCTGTCCTGAAAATCAGACGC-3' and R(SpeI): 5'-TTAATTAATCTAGAAAATTTACTAGTATACGCGGCTCTTGTTGTTATAC-3'. *psfA* was added to that construct with F(SpeI): 5'-ATATATATATACTAGTGATTCAAGGAGGACTCATGC-3' and R(XbaI): 5'-TTAATTAATTTCTAGAAATACTTCAGATGGCGCAGC-3'. *psfA* was cloned with its native ribosomal binding site (RBS) by including 23 bp upstream of the start codon, with expression driven by the lac promoter of the pBBR1MCS(Kan), designated '*psfA*, using F(KpnI): 5'-TACTATGGTACCCCCATTGATTCAAGGAGGAC-3' and R(XbaI): 5'-AGAGAGTCTAGACTCTTCATAGAGGCAGGTGA-3'.

psfG and *psfGA* were cloned to include the native RBS of *psfG,* with 44 bp upstream of *psfG*, designated '*psfG* and '*psfGA*, respectively. '*psfG* was cloned with F(KpnI): 5'-TAATATGGTACCTGATTCCGGAAACCCAAGGC-3' and R(XbaI): 5'-TAGAGATCTAGAAGCGGTCCATTGTCCATCGA-3'. '*psfGA* was cloned with the same forward primer as '*psfG* and the reverse primer (XbaI): 5'-TAGAGATCTAGACCTCTTCATAGAGGCAGGTG-3'.

Samples of *P. putida* KT2440 transformed with '*psfGAB*, '*psfA*-P, '*psfA*, '*psf*G, and '*psfGA* for SDS-PAGE were grown in LB kanamycin medium with 1 mM IPTG, as well as in the

same medium with 10 mM furoic acid added, which has been shown to induce the furoic acid pathway as well as *psfG* and *psfA* (Nichols *et al*., 2008).

2.3.4 Growth rate experiments

Growth rate experiments were performed by washing rich overnight cultures twice in MM with no carbon source, normalizing the inoculation volume by OD_{600} readings, and diluting 1/200 in 5 ml of the indicated growth medium. Growth rates of the 'fur clone in *P. putida* KT2440, *P. aeruginosa* PAO1, *E. coli* MG1655, and *S. enterica* LT2 were performed in MM medium with the carbon source listed at 10 mM, with the appropriate antibiotic. 'fur clones were also tested in the same medium without antibiotic selection. Cultures were grown with continuous shaking. Growth rate was calculated as the slope of the plot of the time in hours versus the natural logarithm of the OD_{600} reading.

2.3.5 Furfural, furoic acid, and HMF tolerance experiments

'*psfGAB*, '*psfA*-P, '*psfA*, '*psf*G, and '*psfGA* clones were grown in LB medium with, in 1 mM increments, 12 - 26 mM furfural, 44 - 60 mM furfuryl alcohol, and 14 - 30 mM HMF with kanamycin and 1, 0.5, or 0.05 mM IPTG. OD_{600} readings were taken after 24 hours of growth, and inhibition was defined as the lowest concentration that yielded an OD_{600} of 0.050 or lower.

2.4 Results

2.4.1 Furfural pathway identification

The furfural catabolic pathway of *P. putida* ALS1267 was found in one contiguous 18,057 bp sequence (to be added to the Genbank database after a patent submission) (Table 2.4 and Figure 2.3). Two coverage gaps of approximately 700 and 300 bp were filled using PCR and conventional sequencing (data not shown). The pathway contained each gene in the Nichols *et al*. (2008) sequence cluster EU290170 from *P. putida* Fu1 in the same order and orientation with an average protein identity of 96% over the sequence lengths stated in Table 2.4. These genes were: *psfGABC*, hydrolase, *psfD*, and araC. It also contained each gene in the Koopman *et al*. (2010) furoic acid pathway sequence cluster GU556182 from *Cupriavidus basilensis* HMF14: *hmfR2*, *hmfABCDE*, and *hmfT2*, with *hmfABCDE* in the same order. The name *msf1* in GU556182 was changed to *hmfT2* by the research group in Wierkcz *et al*. (2016). The average protein identity was 61% for all seven genes and 65% for *hmfABCDE*. The second *C. basilensis* HMF14 sequence cluster (GU556183) from Koopman *et al*. (2010) contains the HMF catabolic genes *hmfFGH*, the LysR regulator *hmfR1* and putative major facilitator superfamily (msf) transporter *hmfT1* that are highly homologous to *hmfR2* and *hmfT2* in the furoic acid sequence, respectively, and the uncharacterized gene *hmfH'*. As there are two msf transporters in the *C. basilensis* HMF14 clusters and only one in the *P. putida* ALS1267 cluster, the msf transporter was designated *hmfT*, although it had a closer identity to *hmfT2* than *hmfT1*. The *P. putida* ALS1267 cluster contained *hmfF* and *hmfG* at 72% and 57% protein identity, respectively, but not *hmfH* or *hmfH'*. In this cluster *P. putida*

Table 2.4. *P. putida* ALS1267 furfural cluster: protein functions and protein identity to other furfural- and HMF-degrading species

Protein	Function	Identity[1]
HmfA	Furoyl-CoA dehydrogenase, large subunit	63% (632/1000)
HmfB	Furoyl-CoA dehydrogenase, FAD-binding subunit	57% (154/270)
HmfC	Furoyl-CoA dehydrogenase, 2Fe-2S iron sulfur subunit	69% (115/166)
HmfD	Furoyl-CoA synthetase	60% (301/501)
HmfE	2-oxoglutaroyl-CoA hydrolase	77% (201/260)
HmfI	Putative transporter	none
HmfF	2,5-furan-dicarboxylic acid decarboxylase 1	72% (351/486)
HmfT	Msf transporter	70% (291/413)
PsfG	Putative furfuryl alcohol/HMF alcohol dehydrogenase	98% (249/252)
PsfA	Putative furfural/HMF dehydrogenase	96% (463/480)
PsfB (HmfR1/ HmfR2)	LysR positive regulator	96% (306/317)
PsfC (HmfG)	2,5-furan-dicarboxylic acid decarboxylase 2	98% (202/206)
Hydrolase	Putative 2-oxoglutaroyl-CoA hydrolase[2]	93% (243/259)
PsfD	Putative furoyl-CoA dehydrogenase accessory factor	96% (312/323)
AraC	AraC-type transcriptional regulator	97% (278/284)

Gene assignments from Koopman *et al.* (2010) and Wierkcz *et al*. (2016) and adapted from Nichols *et al*. (2008). *hmfT* is called *msf1* in Koopman *et al*. (2010) operon GU556183 but was renamed *hmfT1* by Wierckx *et al.* (2016). HmfT also has 58% identity to HmfT2, which was originally also named *msf1* in Koopman *et al*. (2010) cluster GU556182. HmfR1 and HmfR2 from GU556183 and GU556182, respectively, are 61% (182/298) and 48% (82/171) identical to PsfB of *P. putida* Fu1 (cluster EU290170). HmfG from GU556183 is 56% (108/192) identical to PsfC of *P. putida* Fu1 and is 57% (105/184) identical to it in *P. putida* ALS1267.

[1] Protein identity of HmfABCDEFT to *C. basilensis* HMF14 (Koopman *et al*., 2010) and PsfGAB-HmfG-hydrolase-PsfD-araC to *P. putida* Fu1 (Nichols *et al.*, 2008).[2] No experimental evidence.

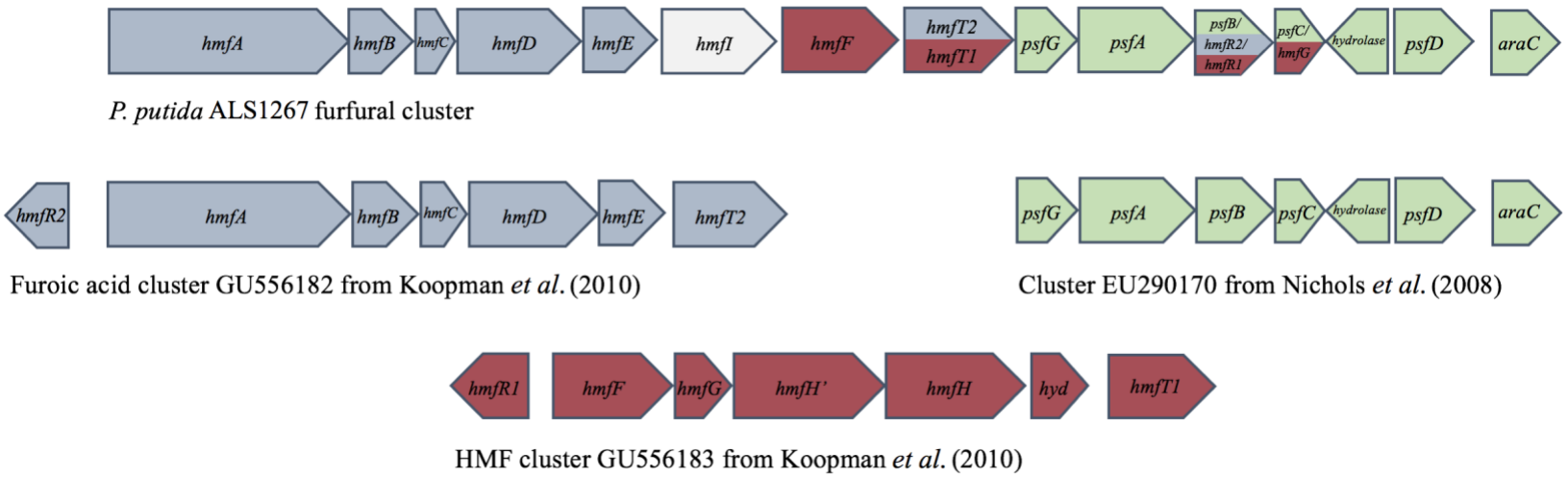

Figure 2.3. Organization of the furfural cluster of *P. putida* ALS1267 and comparison to clusters in *C. basilensis* HMF14 and *P. putida* Fu1

ALS1267 had one gene that encodes a putative sodium:solute symporter that has little sequence identity to *hmfT1*, *hmfT2*, or *hmfH'*. It is not present in *P. putida* KT2440 [closest protein identity match 27% (101/375)]. As neither the genome of *C. basilensis* HMF14 nor *P. putida* Fu1 is available, it is not known if it is present in those species. It was designated *hmfI*.

The furoic acid, furfural, and HMF clusters previously sequenced in *C. basilensis* HMF14 and *P. putida* Fu1 contain two genes in common, both found in the *P. putida* ALS1267 furfural cluster. The LysR regulator *psfB* in *P. putida* Fu1 has 61% and 48% protein identity to *hmfR2* and *hmfR1* of *C. basilensis* HMF14, respectively. In this paper it is called *psfB*. The *P. putida* Fu1 sequence also contains *psfC*, one of the two FDCA decarboxylases of *C. basilensis* HMF14, *hmfG*. In this paper it is called *hmfG*.

2.4.2 Cloning the furfural degradation pathway of *P. putida* ALS1267

The entire 18,057 bp pathway, genes *hmfA* through araC, along with the promoter of *hmfA*, was cloned into *P. putida* KT2440 and growth rate experiments were performed in minimal furfural, furoic acid, and HMF medium. It was found that the control strains *P. putida* KT2440 and *P. aeruginosa* PAO1 with the empty cloning vectors pBBR1MCS-3 and pMP220, respectively, could grow in MM-furoic acid, -furfural, and -HMF (with tetracycline) medium (data not shown). Subsequent experiments showed *P. putida* KT2440-pBBR1MCS-3 and *P. aeruginosa* PAO1-pMP220 grew in MM medium with tetracycline as the sole carbon source, but not furoic acid, furfural, or HMF as sole carbon source without tetracycline (data not shown). Wild-type *P. putida* KT2440 and *P.*

aeruginosa PAO1 without the vectors could not grow in MM with tetracycline as the sole carbon source. Because of this, the kanamycin resistance gene from Tn5 was cloned into the middle of the tetracycline resistance gene using the NheI restriction site in the pBBR1MCS-3 vector, called pBBR1MCS(Kan), and in the 'fur clone in pBBR1MCS-3. The pBBR1MCS(Kan) empty vector did not confer tetracycline resistance, nor could *P. putida* KT2440 nor *P. aeruginosa* PAO1 transformed with it then grow on tetracycline as the sole carbon source (data not shown). pMP220'fur in *P. aeruginosa* PAO1 was grown without tetracycline selection for the growth rates listed in table 2.5. pBBR1MCS(Kan)'fur clones were grown in MM-furoic acid, -furfural, and -HMF with and without kanamycin selection, and their respective growth rates were not significantly different. The growth rates in Table 2.5 are in the pBBR1MCS(Kan)'fur clones grown with kanamycin. *P. putida* KT2440-pBBRMCS-3'fur, *P. putida* KT2440-pBBR1MCS(Kan)'fur and *P. aeruginosa* PAO1-pMP220'fur each grew in MM-furfural without antibiotic selection. Samples were taken out and streaked on LB plates containing the respective antibiotic. Plasmid mini preps were made, digested with SpeI, and run on a gel with the respective digested 'fur vector controls, showing the respective intact 'fur construct in each was retained under furfural selection. The strains were grown in LB medium with no antibiotic for one day, transferred to fresh medium for a second day, and LB-streaked colonies of 5 of each screened for loss of the 'fur clone construct. After 2 days of curing, the 'fur plasmid was lost in each 5 of the three strains.

The 'fur clone in *P. putida* KT2440 enabled it to grow on MM-furoic acid with a growth rate of 0.566/h and MM-furfural at 0.34/h (Table 2.5). 'fur did not enable *P. putida*

Table 2.5. Growth rates (in /h) of furfural pathway clone ('fur) in MM-furoic acid, -furfural, and -HMF

Strain	Furoic Acid	Furfural	HMF
P. putida ALS1267	0.684	0.250	0.311
P. putida KT2440	NG	NG	NG
P. putida KT2440'fur	0.566	0.340	NG
P. aeruginosa PAO1	NG	NG	NG
P. aeruginosa PAO1'fur	0.523	0.312	NG
E. coli MG1655	NG	NG	NG
E. coli MG1655'fur	NG	NG	NG
S. enterica LT2	NG	NG	NG
S. enterica LT2'fur	NG	NG	NG
P. putida Fu1	0.139	0.0850	NG
B. phytofirmans PsJN	0.157	0.0597	0.00998

'fur clones all in pBBR1MCS(Kan) vector, except for *P. aeruginosa* PAO1, which is in pMP220. Growth rates were performed with 10 mM of furoic acid (pH 7.0), furfural (pH 7.0), or HMF (pH 7.0) in MM. NG = no growth.

KT2440 or *P. aeruginosa* PAO1 to grow on MM-HMF. 'fur in *P. aeruginosa* PAO1 enabled it to grow on MM-furoic acid at a growth rate of 0.523/h and MM-furfural at 0.312/h. Both constructs had higher growth rates at 30°C compared to 37°C (data not shown for 37°C). 'fur did not confer growth in any of the three media for *E. coli* MG1655 or *S. enterica* LT2, at 30 or 37°C. *Shewanella putrefaciens* NRRL B-951 cloned with 'fur was able to grow on MM-furfural and -FA, but not MM-HMF (data not shown).

It was found that transformation of the *Pseudomonas* spp. used in this study with the large (23 - 28 kb) 'fur plasmid constructs using standard electrocompetent cell preparation protocols always resulted in degraded plasmids. Thus it is worthwhile to note which preparation worked for both strains. For *P. putida* KT2440, cells were grown to late stationary phase, to an OD_{600} of > 2.0, and washed twice in room temperature magnesium electroporation buffer (Dennis and Sokol, 1995), containing 1 mM HEPES and 1 mM $MgCl_2$ at pH 7.0. Cell were pelleted in a microcentrifuge for 30 s at 13,000 rpm. The same protocol was used for *P. aeruginosa* PAO1, except 300 mM sucrose, pH 7.0, was used. Electrotransformation was done with the preset *Pseudomonas aeruginosa* 2.5 kV protocol on a BioRad Gene Pulser Xcell system. SOC medium and cuvettes were at room temperature. Outgrowth was done for 2 hours at 30°C.

2.4.3 Homology to other species

From a blastn search, two species contained the entire 18 kb furfural cluster of *P. putida* ALS1267, from *hmfA* to the *araC* regulator in the same order and orientation. *Pseudomonas umsongensis* strain BS3657 (LT629767) had 95% DNA identity

(16987/17932) and *Pseudomonas* sp. A3(2016) (CP014870) 94% DNA identity (16913/17938). *Pseudomonas stutzeri* strain 28a24 (CP007441) contained every gene in the cluster except *hmfI* and *hmfT*. It is not known if *P. putida* Fu1 contains the other genes in the cluster, as its genome is not available. Three species, all outside the *Pseudomonas* genus, contained every gene in the cluster except *hmfI*, *hmfT*, and the hydrolase: *Alcanivorax dieselolei* B5 (CP003466) from 3148727 - 3180798, a span that also contained other genes including a series of TRAP-T family transporters, *Bosea* sp. RAC05 (CP016464) from 1136359 - 1148911, and one in a plasmid in *Dinoroseobacter shibae* DFL 12 (plasmid pDSHI01, CP000831 from 117527 - 131284).

2.4.4 Furfural, furfuryl alcohol, and HMF tolerance experiments

None of the *psfA*, *psfG*, or *psfGA* clones ('*psfGAB*, '*psfA*-P, '*psfA*, '*psfG*, and '*psfGA*) in *P. putida* KT2440 showed an improved tolerance to furfural, furfuryl alcohol, or HMF when grown with 1 mM, 0.5 mM, or 0.05 mM IPTG (data not shown). '*psfG* and '*psfGA* clones both showed a decreased tolerance to furfuryl alcohol (data not shown). As this was not seen when grown in LB-furfural or -HMF (OD_{600} readings were not significantly different than the control strain), this was not due to the energy burden of recombinant protein overexpression.

SDS-PAGE of '*psfG* and '*psfGA* showed a protein band corresponding to the 26 kDa size of PsfG (Figure 2.4). PsfA is a 51 kDa protein, and no overexpresson was observed for that size in any of the clones. It appears that *psfG* and *psfA* are expressed in a single unit, so it is believed that PsfA was expressed at least in the '*psfGA* construct.

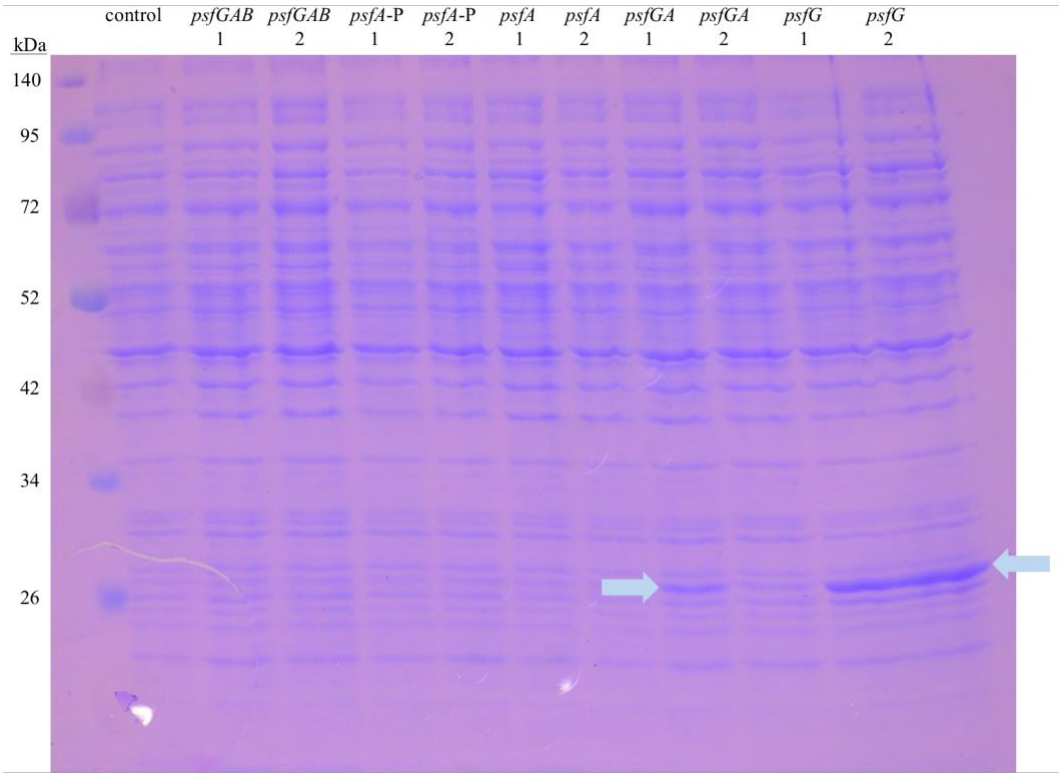

Figure 2.4. SDS-PAGE of *P. putida* ALS1267 *psfGAB*, *psfA*-P, *psfA*, *psfGA*, and *psfG* cloning constructs in *P. putida* KT2440

The *psfA*-P clone contains the promoter of *hmfA* inserted upstream of *psfA*, as described in *Methods*. Samples were grown in LB medium with kanamycin and induced with IPTG to an OD_{600} of 0.5 - 0.6 and normalized by the OD_{600} reading.

2.5 Discussion

Two papers previously identified the gene assignments for the bacterial aerobic catabolism of furfural, HMF, and furoic acid in the species *P. putida* Fu1 and *C. basilensis* HMF14, which are in three distinct sequence clusters (Figure 2.3) (Nichols *et al*., 2008; Koopman *et al*., 2010). This study shows a third strain, *P. putida* ALS1267, contains every gene except two in each of the three clusters in one contiguous 18 kb sequence. When this 18 kb pathway, containing *hmfABCDEIFT-psfGAB-hmfG*-hydrolase-*psfD*-araC, designated 'fur, was cloned into *P. putida* KT2440, which cannot metabolize any of the furans, it enabled fast growth on minimal furoic acid and minimal furfural media (Table 2.5). The clone was not able to metabolize HMF. The 18 kb cluster did not contain *hmfH*, encoding HMF acid oxidase (also called furfural/HMF oxidoreductase). The clone contained the two other HMF-specific genes, *hmfF* and *hmfG*, which decarboxylate FDCA to furoic acid. The other gene not present, *hmfH'*, which has some sequence identity to a probable extracytoplasmic solute receptor of *Ralstonia eutropha* H16, was shown nonessential in metabolism of the furan compounds (Koopman *et al*., 2010).

The step of conversion of HMF acid to FDCA by HmfH is of special consideration, as FDCA is one of the top 12 most potentially valuable biologically produced platform chemicals, which can be used as a replacement for terephthalic acid in polyester industries (Werpy and Peterson, 2004). Several research groups have advanced biological production of FDCA (Koopman *et al*., 2010; Hossain *et al*., 2017; Yang and Huang,

2017; Karich *et al*., 2018; Yuan *et al*., 2018). Due to this interest, and to complete the gene assignments of *P. putida* ALS1267 for HMF catabolism, it was attempted to find the gene encoding this activity in the genome assembly annotation. HmfH is a glucose-methanol-choline (GMC) oxidoreductase family protein. Along with HmfH', HmfH is the least conserved protein in both of the *C. basilensis* HMF14 clusters, in terms of homology in other furoic acid-degrading species (Koopman *et al*., 2010). In species that contain the more highly conserved furoic acid pathway, *hmfABCDE*, as well as *hmfFG*, HmfH protein identity was found from 68 to 43% (Koopman *et al*., 2010). *Bradyrhizobium japonicum* USDA110 could catabolize HMF and had 45% identity to HmfH of *C. basilensis* HMH14, but *Burkholderia xenovorans* LB400 had 44% identity and could catabolize furfural but not HMF (Koopman *et al*., 2010).

There are only two matches to HmfH of 20% or greater protein identity over a 300 amino acid span or greater in *P. putida* ALS1267: a predicted glucose-methanol-choline oxidoreductase, at 32% (187/581) identity and a predicted choline dehydrogenase at 30% (185/606). Both of these proteins are found in *P. putida* KT2440, choline dehydrogenase (NP_742226; 99% (541/548) identity) and oxygen-dependent choline dehydrogenase [WP_010955611; 99% (558/565) identity], respectively. Thus, with the hypothesis that the lack of HmfH in 'fur is the reason why *P. putida* KT2440'fur cannot metabolize HMF, neither could be the HmfH ortholog. The three next closest matches to HmfH in *P. putida* ALS1267 are only at 21% (59/277), 26% (33/127), and 31% (21/68). The former two are found in *P. putida* KT2440 (GMC family oxidoreductase NP_747223 and dehydrogenase NP_743818, respectively). The latter, a predicted glucose-methanol-choline

oxidoreductase:NAD binding site, has no ortholog in *P. putida* KT2440. When searched through species possessing the furoic acid pathway there was no correlation, and it was not found in the three *Pseudomonas* spp. that contained all (*P. umsongensis* strain BS3657 and *Pseudomonas* sp. A3(2016)) or most (*P. stutzeri* strain 28a24) of the furfural cluster of *P. putida* ALS1267. When HmfH from *C. basilensis* HMF14 is searched through all known proteins in the genus *Pseudomonas*, there are over 35 species at 44% or higher protein identity over a 500 amino acid span, none of which possess the furoic acid pathway. The closest two matches were 47% (266/562) identity for glucose dehydrogenase (WP_082056588; *Pseudomonas* sp. 10-1B) and 46% (254/558) identity for sorbosone dehydrogenase (WP_032628900; *Pseudomonas syringae*). The closest matches of these two *Pseudomonas* dehydrogenases in *P. putida* ALS1267 where to the aforementioned predicted GMC oxidoreductases. The likelihood that HmfH is in a gap in coverage in the *P. putida* ALS1267 genome assembly is low, as neither *P. umsongensis* strain BS3657, *Pseudomonas* sp. A3(2016) nor *P. stutzeri* strain 28a24 have HmfH orthologs over 34% protein identity. The same is true for the three other species that possess all genes of the *P. putida* ALS1267 furfural cluster excluding *hmfI*, *hmfT*, and the hydrolase (*A. dieselolei* B5, *Bosea* sp. RAC05, and *D. shibae* DFL 12 plasmid pDSHI01, although the latter species has been reported to metabolize furfural but not HMF and thus was excluded) (Wierckx *et al*., 2011). It appears that there is no HmfH ortholog in *P. putida* ALS1267. Similarly, candidate orthologs of HMF acid oxidase, HMFO, from *Methylovorus* sp. strain MP688, which converts HMF to FDCA via an FFA intermediate instead of an HMF acid intermediate, were identified (Dijkman and Fraaije, 2014). The closest matches were to the aforementioned predicted GMC oxidoreductases.

To attempt to find a candidate gene for HMF acid oxidation to FDCA and to further study the difference between the *P. putida* ALS1267 and *C. basilensis* HMF14 HMF pathway, a list of species was made that possess the furoic acid pathway, HmfABCDE, as well as two of the HMF-specific enzymes, HmfFG, and the presence of putative HmfH orthologs was noted (presence of a putative ortholog was defined as having a 44% or greater protein identity over 80% or greater of the length of the protein sequence) (Table 2.6). HmfH can also carry out each of the four upper pathway oxidations (Koopman *et al*., 2010). Neither PsfA (the putative furfural/HMF dehydrogenase) nor PsfG (the putative furfuryl alcohol/HMF alcohol dehydrogenase) have any significant sequence identity to HmfH. There is no HmfH ortholog in *P. putida* ALS1267. Wiercke *et al*. (2016) identified an enzyme in *C. basilensis* HMF14, termed "aldehyde dehydrogenase" (Adh), which increased the rate of FDCA formation from HMF when added to a *hmfH-hmfT1* clone. Adh was also shown to increase yield from 5-formylfuran-2-carboxylic acid (FFA), which can also be used as a feedstock for FDCA production. *C. basilensis* HMF14 Adh has a 62% (294/477) protein sequence identity to PsfA, with the next closest protein match in *P. putida* ALS1267 at 36% (162/454) and the closest match in *P. putida* KT2440 at 34% (161/476), giving good support to the prediction of PsfA as a HMF (and furfural) dehydrogenase. Putative orthologs for HmfH, PsfA, and PsfG were identified in the species in Table 2.6. 7 species possessed PsfA and PsfG but not HmfH (all in the genera *Acidovorax*, *Bosea*, *Pseudomonas*, and *Alcanivorax*, as well as one in a plasmid in *Dinoroseobacteri*). 8 species possessed PsfA but not HmfH or PsfG (all in the genera *Azoarcus*, *Bordetella*, and *Rhodopseudomonas*, *Pelagibaca*, as well as one in a plasmid in

Table 2.6. HmfH, PsfG, and PsfA orthologs in species with HmfABCDE and HmfFG

Class[1]	Order	Species	Accession Number
List 1. Have HmfH ortholog (No PsfG or PsfA orthologs)			
Beta	Burkholderiales	*Paraburkholderia phytofirmans* PsJN	CP001052.1
Beta	Burkholderiales	*Paraburkholderia* sp. BN5 plasmid pBN1	CP022991.1
Beta	Burkholderiales	*Paraburkholderia caribensis* MBA4 plasmid	CP012748.1
Beta	Burkholderiales	*Burkholderia* sp. CCGE1001	CP002520.1
Beta	Burkholderiales	*Burkholderia* sp. CCGE1003	CP002217.1
Beta	Burkholderiales	*Paraburkholderia* fungorum strain ATCC BAA-463	CP010026.1
Beta	Burkholderiales	*Paraburkholderia hospita* strain DSM 17164	CP026107.1
Beta	Burkholderiales	*Paraburkholderia phenoliruptrix* BR3459a	CP003864.1
Beta	Burkholderiales	*Paraburkholderia terrae* strain DSM 17804	CP026113.1
Beta	Burkholderiales	*Paraburkholderia phymatum* STM815 plasmid pBPHY01	CP001045.1
Beta	Burkholderiales	*Burkholderia* sp. PAMC 26561	CP014307.1
Beta	Burkholderiales	*Burkholderia* sp. PAMC 28687 strain PAMC28687	CP014507.1
Alpha	Rhizobiales	*Methylobacterium nodulans* ORS 2060	CP001349.1
Alpha	Rhizobiales	*Methylobacterium* sp. 4-46	CP000943.1
Alpha	Rhizobiales	*Bradyrhizobium japonicum* strain J5	CP017637.1
Alpha	Rhizobiales	*Bradyrhizobium diazoefficiens* USDA 110	CP011360.1

Table 2.6 continued

Alpha	Rhizobiales	*Bradyrhizobium diazoefficiens* USDA 110	BA000040.2
Alpha	Rhizobiales	*Bradyrhizobium diazoefficiens* strain USDA 122	CP013127.1
Alpha	Rhizobiales	*Bradyrhizobium diazoefficiens* DNA strain: NK6	AP014685.1
Alpha	Rhizobiales	*Bradyrhizobium japonicum* USDA 6 DNA	AP012206.1
Alpha	Rhizobiales	*Bradyrhizobium japonicum* SEMIA 5079	CP007569.1
Alpha	Rhizobiales	*Bradyrhizobium japonicum* strain E109	CP010313.1
Alpha	Rhizobiales	*Methylobacterium phyllosphaerae* strain CBMB27	CP015367.1
Alpha	Rhizobiales	*Methylobacterium oryzae* CBMB20	CP003811.1
Alpha	Rhizobiales	*Methylobacterium radiotolerans* JCM 2831	CP001001.1
Alpha	Rhizobiales	*Bradyrhizobium* sp. G22	LN907826.1
List 2. Have neither HmfH, PsfG, nor PsfA orthologs			
Beta	Burkholderiales	*Paraburkholderia xenovoran*s LB400	CP008760.1
Beta	Burkholderiales	*Paraburkholderia xenovorans* LB400	CP000270.1
Beta	Burkholderiales	*Burkholderia* sp. OLGA172	CP014578.1
Alpha	Rhizobiales	*Pseudorhodoplanes sinuspersici* strain RIPI110	CP021112.1
Alpha	Rhizobiales	*Bradyrhizobium erythrophlei* strain GAS401	LT670849.1
List 3. Have HmfH and PsfA orthologs (no PsfG ortholog)			
Beta	Burkholderiales	*Cupriavidus* sp. NH9 plasmid pENH92	CP017759.1
Beta	Burkholderiales	*Burkholderia* sp. HB1	CP012193.1

Table 2.6 continued

Beta	Burkholderiales	*Burkholderia* sp. CCGE1002	CP002015.1
Beta	Burkholderiales	*Paraburkholderia caribensis* strain Bcrs1W	CP013349.1
Beta	Burkholderiales	*Paraburkholderia caribensis* strain DSM 13236	CP026103.1
Beta	Burkholderiales	*Paraburkholderia caribensis* strain MWAP64 plasmid 1	CP013104.1
Beta	Burkholderiales	*Alcaligenaceae* bacterium LMG 29303 Orrdi1	LT907988.1
Beta	Burkholderiales	*Bordetella flabilis* strain AU10664	CP016172.1
Alpha	Rhizobiales	*Methylobacterium aquaticum* DNA, strain MA-22AA	AP014704.1
List 4. Have PsfA and PsfG orthologs (No HmfH ortholog)			
Beta	Burkholderiales	*Acidovorax* sp. KKS102	CP003872.1
Alpha	Rhizobiales	*Bosea* sp. RAC05	CP016464.1
Gamma	Pseudomonadales	*Pseudomonas stutzeri* strain 28a24	CP007441.1
Gamma	Pseudomonadales	*Pseudomonas umsongensis* strain BS3657	LT629767.1
Gamma	Pseudomonadales	*Pseudomonas* sp. A3(2016)	CP014870.1
Gamma	Oceanospirllales	*Alcanivorax dieselolei* B5	CP003466.1
Alpha	Rhodobacterales	*Dinoroseobacter shibae* DFL 12 plasmid pDSHI01	CP000831.1
List 5. Have PsfA ortholog (No HmfH or PsfG orthologs)			
Beta	Burkholderiales	*Bordetella bronchialis* strain AU3182	CP016170.1
Beta	Burkholderiales	*Bordetella bronchialis* strain AU17976	CP016171.1

Table 2.6 continued

Beta	Burkholderiales	*Bordetella* sp. H567	CP012334.1
Beta	Rhodocyclales	*Azoarcus* sp. SY39	CP025682.1
Alpha	Rhodospirillales	*Azospirillum* sp. B510 plasmid pAB510a	AP010947.1
Beta	Burkholderiales	*Bordetella genomosp*. 8 strain AU19157	CP021108.1
Alpha	Rhizobiales	*Rhodopseudomonas palustris* BisB18	CP000301.1
Alpha	Rhodobacterales	*Pelagibaca abyssi* strain JLT2014	CP015093.1

[1]All species found were in the phylum Proteobacteria; listed is the class of Proteobacteria, Alphaproteobacteria, Betaproteobacteria, or Gammaproteobacteria. Matches were based on blastn searches for homology to *C. basilensis* HMF14. Species in each category are listed in descending order of homology by DNA identity to *hmfABCDE* of *C. basilensis* HMF14.

Azospirillum). 27 species possessed HmfH but not PsfA nor PsfG (all in the genera *Paraburkholderia*, *Burkholderia*, *Methylobacterium*, and *Bradyrhizobium*). 9 species had both PsfA and HmfH, but no species was found to have both PsfG and HmfH.
Four species from Table 2.6 that have been determined to metabolize both furfural and HMF are all in List 1, having HmfH but not PsfG or PsfA (*Burkholderia phytofirmans* PsJN 2010, *Burkholderia phymatum* STM815 2010, *Bradyrhizobium japonicum* USDA110 2010, and *Methylobacterium radiotolerans* 2010 JCM2831). The three species in Table 2.6 that have been determined to metabolize furfural but not HMF are *Paraburkholderia xenovorans* LB400 (List 2: No HmfH, PsfG, or PsfA orthologs), *D. shibae* DFL 12 plasmid pDSHI01 (List 4: have PsfA and PsfG but no HmfH), and *Rhodopseudomonas palustris* BisB18 2010 (List 5: have PsfA but no HmfH or PsfG). *P. putida* ALS1267 is the first species reported that has HmfABCDEFG but not HmfH and can metabolize HMF. Thus in the Proteobacteria it appears there are at least two evolutionary routes for furfural/HMF/furfuryl alcohol/HMF alcohol oxidations and the oxidation of HMF acid to FDCA.

6010 proteins were annotated from the *P. putida* ALS1267 genome assembly; 5555 have orthologs in *P. putida* KT2440. A comparative genomics study of the additional 455 genes and encoded proteins was conducted. There are no prior experimentally validated Proteobacterial species that metabolize both furfural and HMF that do not possess an HmfH ortholog, with *P. putida* ALS1267 being the first. Thus the comparative study could only rely on the assumption that at least one of the two species that possess the entire 18 kb furfural cluster of *P. putida* ALS1267, *P. umsongensis* strain BS3657 and

Pseudomonas sp. A3(2016) (both of which do not possess an HmfH ortholog), can metabolize HMF and thus possesses an HMF acid dehydrogenase. Likewise, correlation was assessed for *P. stutzeri* strain 28a24, *A. dieselolei* B5, and *Bosea* sp. RAC05. First, genomic sequences of *P. putida* ALS1267 that do not align to the *P. putida* KT2440 genome were aligned separately to *P. umsongensis* strain BS3657 and *Pseudomonas* sp. A3(2016). Matches were found (from 88 - 95% DNA identity) in a cluster in *P. umsongensis* strain BS3657 at (LT629767) 1391324 - 1397942, 1404847 - 1406475 and 1458400 - 1465342. These contained outer membrane porin, OprD family (SDS74007), zinc-binding alcohol dehydrogenase family protein (SDS74079), quinol monooxygenase YgiN (SDS74123), 2,4-dienoyl-CoA reductase (SDS74181), DNA-binding transcriptional regulator, ArsR family (SDS74241), DNA-binding transcriptional regulator, LysR family (SDS74288), NADP-dependent 3-hydroxy acid dehydrogenase YdfG (SDS74329), zinc-binding alcohol dehydrogenase family protein (SDS74720), quinol monooxygenase YgiN (SDS74763), luciferase-type oxidoreductase, BA3436 family (SDS74786), predicted Fe^{2+}/Mn^{2+} transporter, VIT1/CCC1 family (SDS77080), thiamine biosynthesis lipoprotein (SDS77129), sulfite reductase (NADPH) flavoprotein alpha-component (SDS77186), predicted protein (SDS77230), hypothetical protein (SDS77277), two-component system, OmpR family, response regulator (SDS77338), and two-component system, OmpR family, sensor kinase (SDS77393). The same matches were found in *Pseudomonas* sp. A3(2016), excluding the DNA-binding transcriptional regulator, LysR family, YgiN (SDS74763), and YdfG. *Bosea* sp. RAC05 only possessed orthologs of 2,4-dienoyl-CoA reductase and sulfite reductase (NADPH) flavoprotein alpha-component. *A. dieselolei* B5 only possessed orthologs of the two zinc-binding

alcohol dehydrogenase family proteins. Using more stringent homology criteria for *P. stutzeri* strain 28a24, it did not possess any of the candidate proteins over 67% identity. All other species in Table 2.6 were then analyzed for the presence of orthologs of possible candidates for HMF acid oxidase: zinc-binding alcohol dehydrogenase family SDS74079 (here called zinc-binding alcohol dehydrogenase 1), quinol monooxygenase SDS74123 (called YgiN1), NADP-dependent 3-hydroxy acid dehydrogenase YdfG, zinc-binding alcohol dehydrogenase family protein SDS74720 (called zinc-binding alcohol dehydrogenase 2), quinol monooxygenase SDS74763 (called YgiN2), luciferase-type oxidoreductase, BA3436 family (called luciferase-type oxidoreductase), and sulfite reductase (NADPH) flavoprotein alpha-component (called sulfite reductase alpha-component). First, the three species known to metabolize furfural but not HMF were used to exclude candidates (*P. xenovorans* LB400, *D. shibae* DFL 12 plasmid pDSHI01 (the host chromosome was included), and *R. palustris* BisB18 2010). *P. xenovorans* LB400 and *R. palustris* BisB18 2010 had an ortholog of both Zn-binding alcohol dehydrogenases and *D. shibae* DFL 12 had an ortholog of 2,4-dienoyl-CoA reductase. *R. palustris* BisB18 2010 had a 47% (113/238) match for the NADP-dependent 3-hydroxy acid dehydrogenase YdfG. As it is the most likely candidate for an HMF oxidase, the relatively low identity match should not cause it to be excluded. There was no correlation with YdfG in the four species in List 1 known to metabolize HMF [*B. phytofirmans* PsJN 2010 and *M. radiotolerans* 2010 JCM2831 did not possess orthologs of YdfG but *B. phymatum* STM815 2010 and *B. japonicum* USDA110 2010 did at 68% (167/244) and 55% (135/244)]. Neither the three species from Table 2.6 known to metabolize furfural but not HMF nor the four known to metabolize both furfural and HMF possessed

orthologs of the luciferase-type oxidoreductase or sulfite reductase alpha-component. Of all species in Table 2.6, the luciferase-type oxidoreductase was found only in *P. umsongensis* strain BS3657, *Pseudomonas* sp. A3(2016), *P. caribensis* strain Bcrs1W, *P. caribensis* strain DSM 13236, *P. caribensis* strain MWAP64 plasmid 1, and *Azoarcus* sp. SY39, and the sulfite reductase alpha-component was only found in *Paraburkholderia* sp. BN5 plasmid pBN1, *M. nodulans* ORS 2060, *Methylobacterium* sp. 4-46, *Bosea* sp. RAC05, *P. umsongensis* strain BS3657, and *Pseudomonas* sp. A3(2016). Of all species in Table 2.6, only *P. umsongensis* strain BS3657 possessed the quinol monooxygenases YgiN1 and YgiN2; *Pseudomonas* sp. A3(2016) possessed only YgiN1. Analysis of the ability of the other species in Table 2.6 for the ability to metabolize HMF and the presence or absence of orthologs of the aforementioned proteins would help further narrow the list of possible candidates for HMF acid oxidase.

The *C. basilensis* HMF14 furoic acid pathway (*hmfABCDE*) cloned into *P. putida* S12 was reported as having poor growth on furfural and furoic acid, and adaptation on minimal furfural medium improved the growth rate to 0.3/h on furfural (initial growth rate not reported) (Koopman *et al*., 2010), showing that (1) there are innate factors in *P. putida* S12 that can contribute to furfural and furoic acid metabolism, including a nonspecific furfural/HMF dehydrogenase but also including at least one more, and (2) adding additional gene(s) would improve the clone. The poor initial growth of the *hmfABCDE* clone, as well as the fact that adaptation on minimal furfural medium improved the growth rate on furfural, demonstrates innate factors of the cloning host that must be considered when analyzing the gene assignments for furfural and furoic acid

metabolism. As the cloning host *P. putida* S12 has innate furfural oxidation ability (Koopman *et al*., 2010), all that strain needs added for growth on furfural or furoic acid is the furoic acid pathway. The enzyme functions of HmfABCDE in furoic acid metabolism have been experimentally validated (Koenig and Andreesen, 1989; Koenig and Andreesen, 1990; Koopman *et al*., 2010). There is a hydrolysis reaction of 5-hydroxy-2-furoyl-CoA to 2-oxoglutaroyl-CoA in the furoic acid catabolic pathway (Figure 2.1) that was presumed to be either spontaneous or catalyzed by an unidentified, nonspecific hydrolase (Koopman *et al*., 2010). If the latter is the case, then this ability must occur in the cloning host *P. putida* S12. In that scenario, it is possible that suboptimal flux through this nonspecific hydrolase is the or a contributing cause to the poor growth on furoic acid of the clone. A hydrolase in the *P. putida* Fu1 cluster was also found in the *P. putida* ALS1267 cluster, although its role in carrying out this reaction was not investigated. It was included in the *P. putida* ALS1267 'fur clone. A putative hydroxylase was found in the HMF cluster of *C. basilensis* HMF14 (that was not included in the clone from *C. basilensis* HMF14), although it has little sequence identity to the hydrolase of *P. putida* ALS1267, and has not been experimentally tested for a role in furoic acid metabolism.

As the *C. basilensis* HMF14 *hmfABCDE* clone in *P. putida* S12 had poor initial growth on furfural as the sole carbon source, including additional genes from *C. basilensis* HMF14 would improve the clone. Candidates are *hmfT2* and *hmfT1*, the msf transporters found in the two *C. basilensis* HMF14 clusters. *hmfT1*, when added to a clone containing *hmfH*, improved FDCA formation over a clone containing *hmfH* alone (Wierckx *et al*., 2016). The *P. putida* ALS1267 'fur clone made in this study contained the ortholog *hmfT*.

Another gene not included in the *C. basilensis* HMF clone was *hmfR2*, the lysR positive regulator, which has an ortholog in *P. putida* ALS1267, *psfB*. Nichols *et al*. (2008) showed that deleting *psfB* increased the doubling time of *P. putida* Fu1 on furoic acid from 121 minutes to 11 - 15 hours. Thus it appeared that a mutation occurred somewhere in the *P. putida* Fu1 genome to compensate for the loss of PsfB (Nichols *et al*., 2008). The *C. basilensis* HMF14 *hmfABCDE* clone contained 13 bp upstream of *hmfA*, to include the native RBS, but expression was driven by the promoter of the cloning vector (Koopman *et al*., 2010). In this study, the *P. putida* ALS1267 'fur clone included 142 bp upstream to include the native promoter of *hmfA*, thus likely requiring *psfB* for robust growth. Nichols *et al*. (2008) showed that PsfB positively regulates expression of *psfC* and *psfD*, but not *psfG* or *psfA*. As *psfG* and *psfA* are located between *hmfABCDE* and *psfC* and *psfD*, there are multiple promoters in the *P. putida* ALS1267 furfural cluster. Thus a furoic acid-degrading clone from *P. putida* ALS1267 in the native orientation would require *psfB* for robust growth on furoic acid, regardless of whether *hmfABCDE* expression is driven from the vector promoter.

Innate host ability also complicates comparing, based on the literature, the genes responsible for the upper pathways (the furfuryl alcohol, furfural, HMF alcohol, and HMF oxidations). *P. putida* S12 can perform each of the four upper pathway oxidations, as well as the furfural and HMF reductions to their respective alcohols (Koopman *et al*., 2010). Likewise, it has been shown that *P. putida* KT2440 has native furfural oxidation ability (Lee *et al*., 2017). A *hmfABCDEFGH* clone from *C. basilensis* HMF14 enabled *P. putida* S12 to metabolize HMF (Koopman *et al*., 2010). HmfFGH were added to the *C.*

basilensis HMF14 *hmfABCDE* clone after adaptation on minimal furfural medium to improve the growth rate, and the rate of HmfH oxidation of furfural and HMF was not stated (Koopman *et al*., 2010). Thus it is not possible to determine the extent of HmfH oxidation of furfural and HMF based on the literature. However, addition of Adh of *C. basilensis* HMF14 to a *hmfH* clone decreased the lag time of conversion of HMF to FDCA (Wierckx *et al*., 2016). As this clone was also made in *P. putida* S12, this demonstrates that Adh of *C. basilensis* HMF14 has a more robust HMF oxidation activity than the nonspecific HMF dehydrogenase possessed by the host, at least when overexpressed on a plasmid. Thus including *C. basilensis* HMF14 *adh* (or, based on homology, *P. putida* ALS1267 *psfA*) would improve furfural catabolism of the *hmfABCDE* cloning construct. This is worthwhile in mentioning because another lab group has replicated the same construct (*hmfABCDEFGH* from *B. phytofirmans* PsJN), for the aim of engineering a furfural and HMF-degrading industrial strain, in *P. putida* KT2440, and the construct was also aided by adaptation on minimal furfural medium (Guarnieri *et al*., 2017). It is also possible that other, unknown genes of *C. basilensis* HMF14 would improve that clone, although the genome is not available for analysis.

Although the genome of *C. basilensis* HMF14 is not available for comparison, the genome of *B. phytofirmans* PsJN, also in the order Burkholderiales and which can metabolize furfural and HMF, is. Similar to the difference in the presence of HmfH orthologs, *B. phytofirmans* PsJN, as well as the other species that possess *hmfABCDEFGH* and have been validated on growing on furfural and HMF (*B. diazoefficiens* USDA 110, *B. phymatum* STM815 2010, and *M. radiotolerans* JCM

2831), do not possess orthologs of HmfI, PsfGA, the hydrolase, PsfD, or the araC regulator. The cloning hosts, *P. putida* S12 and *P. putida* KT2440, also do not possess orthologs of those proteins (PsfD is found at 62% and 61% identity to xanthine dehydrogenase accessory factor, but in *Pseudomonas* sp. A3(2016) it is found at 95% identity). The xanthine dehydrogenase accessory factor-like protein PsfD has been shown to be involved in furoic acid metabolism, likely as an accessory factor for HmfABC (Nichols *et al*., 2008; Koenig and Andreesen, 1989). Deleting *psfD* in *P. putida* Fu1 caused a 5-fold decrease in growth rate in growth on furoic acid (Nichols *et al*., 2008) but did not eliminate growth on furoic acid. As there is no PsfD ortholog in *C. basilensis* HMF14, it is likely not required for its HmfABC, but it is required for robust growth on furoic acid in a clone from the *P. putida* ALS1267 cluster. As with the hydrolase, the roles of HmfI and the araC regulator of the *P. putida* ALS1267 furfural cluster were not experimentally investigated, but the genes were included in the 'fur clone.

As furfural and HMF toxicity is the major reason for the study of the pathway due to their presence in a variety of pretreated lignocellulosic hydrolysates, characterization of the enzymes responsible for furfural and HMF reduction and oxidation was performed. To date, the only characterization of PsfA and PsfG (of *P. putida* Fu1) was in response to growth on furoic acid (Nichols *et al*., 2008). Both were induced in growth on furoic acid, and neither was regulated by PsfB, although PsfC and PsfD were upregulated by PsfB (Nichols *et al*., 2008). The paper stated that a *psfA* knockout in *P. putida* Fu1 lost all ability to grow on furoic acid as the sole carbon source; growth on furfural was not tested. In the assumption that PsfA is the furfural/HMF dehydrogenase, this result was

puzzling, as PsfA would act in the upper pathway oxidation of furfural to furoic acid and thus should have no effect on catabolism of furoic acid. An author of that paper kindly sent the Δ*psfA* strain and its parent for further characterization. It was found that although *P. putida* Fu1 Δ*psfA* had a decreased growth rate in furoic acid as the sole carbon source, it could grow (Table 2.7). The *psfA* knockout was not an in-frame deletion and thus the decreased growth rate in furoic acid could have been due to a polar effect. The growth rate decrease of the Δ*psfA* strain in furoic acid was 2.0-fold relative to wild-type *P. putida* Fu1, and the difference in furfural was 4.1-fold. Thus, although a decrease in the growth rate on furoic acid would affect growth on furfural, the *psfA* knockout decreases furfural catabolism further, giving supporting, correlative evidence to the role of PsfA in furfural metabolism. As all *P. putida* species that have been tested possess native furfural oxidation ability, deleting *psfA*, if it is indeed the furfural/HMF dehydrogenase, would not eliminate that ability.

Species that cannot metabolize furfural or HMF reduce them to their respective alcohols and/or oxidize them to their respective acids to detoxify them, as all are less toxic than the aldehydes (Table 2.2) (Boopathy *et al*., 1993; Gutiérrez *et al*., 2002; Taherzadeh *et al*., 1999; Tsuge *et al*., 2014). In this study, none of the cloning constructs of *psfA* or *psfG* (or *psfA* and *psfG* together) conferred tolerance to furfural or HMF (data not shown). Due to that result, the 'fur clone in *P. putida* KT2440 was also tested for conferring tolerance to furfural and furfuryl alcohol. It was found that 'fur increased tolerance to furfuryl alcohol but not furfural (Tables 2.8 and 2.9).

Table 2.7. Growth rate of *P. putida* Fu1 Δ*psfA* in minimal furoic acid and furfural medium

Strain	Growth rate (/h) in furoic acid	Growth rate (/h) in furfural
P. putida Fu1	0.136	0.0822
P. putida Fu1 Δ*psfA*	0.0690	0.0201

Growth rates were done wit *P. putida* KT2440 as a control, which did not grow in either medium.

Table 2.8. Tolerance of *P. putida* ALS1267 furfural pathway clone ('fur) in LB-furfuryl alcohol

Strain	48 mM	50 mM	52 mM	54 mM	56 mM
P. putida KT2440'fur	> 2.000	> 2.000	> 2.000	1.100	0.410
P. putida KT2440-pBBR1MCS(Kan)	0.142	0.058	0.043	0.024	0.040

Units listed are OD_{600} readings taken after 24 hours of growth.

Table 2.9. Tolerance of *P. putida* ALS1267 furfural pathway clone ('fur) in LB-furfural

Strain	16 mM	18 mM	20 mM	22 mM	24 mM
P. putida KT2440'fur	1.720	0.490	0.240	0.130	0.084
P. putida KT2440-pBBR1MCS(Kan)	1.680	0.384	0.229	0.104	0.072

Units listed are OD_{600} readings taken after 24 hours of growth.

This result demonstrates that the upper pathway is active in the 'fur clone, but it is unclear why furfural tolerance was not improved. These results and the *psfA*, *psfG*, and *psfGA* cloning experiments show that, despite the fast furfural catabolism of *P. putida* ALS1267, using these upper pathway enzymes is not an effective route for furfural detoxification.

In enzyme assays of whole cell extracts, *P. putida* ALS1267 has a 2.3-fold higher furfural oxidation activity than *P. putida* KT2440 (Table 2.1) (Lee *et al*., 2017). *P. putida* ALS1267 has a 2.5-fold higher furfural tolerance and 170-fold higher furfural oxidation activity compared to *P. putida* Fu1 and a 3.6-fold higher tolerance and 6.5-fold higher furfural oxidation activity than *B. phytofirmans* PsJN (Lee *et al*., 2017). Given the sequence similarity, it is unknown why *P. putida* ALS1267 has such a higher growth rate on furfural than *P. putida* Fu1 since the genome of the latter is not available for comparison. *P. putida* Fu1 has been previously reported to metabolize HMF (Lee *et al*., 2017), but in this study, in the minimal medium used, it could not. Also of note, *P. putida* ALS1267 has a lower furfural tolerance than *P. putida* KT2440 (Lee *et al*., 2017), as well as a lower growth rate in LB medium (data not shown), and thus *P. putida* ALS1267 is not itself a candidate strain for industrial use.

REFERENCES

Cheshire MV, Russell JD, Fraser AR, Bracewell JM, Robertsons GW, Benzig-Purdie LM, Ratcliffe CI, Ripmeester JA, Goodman BA. 1992. Nature of soil carbohydrate and its association with soil humic substances. Eur J Soil Sci. 43:359-73.

Christian TJ, Kleiss B, Yokelson RJ, Holzinger R, Crutzen PJ, Hao WM, Saharjo BH, Ward DE. 2003. Comprehensive laboratory measurements of biomass-burning emissions: 1. Emissions from Indonesian, African, and other fuels. J Geophys Res Atmos. 108:D23.

Dennis JJ, Sokol PA. 1995. Electrotransformation of *Pseudomonas*. In: Electroporation Protocols for Microorganisms. Humana Press. 125-33.

Dijkman WP, Groothuis DE, Fraaije MW. 2014. Enzyme-Catalyzed Oxidation of 5-Hydroxymethylfurfural to Furan-2, 5-dicarboxylic Acid. Angewandte Chemie International Edition. 53:6515-8.

Guyer MS, Reed RR, Steitz JA, Low KB. 1981. Identification of a sex-factor-affinity site in *E. coli* as γδ. In: Cold Spring Harbor symposia on quantitative biology. Cold Spring Harbor Laboratory Press. 45:135-140.

Hossain GS, Yuan H, Li J, Shin HD, Wang M, Du G, Chen J, Liu L. 2017. Metabolic engineering of *Raoultella ornithinolytica* BF60 for production of 2, 5-furandicarboxylic acid from 5-hydroxymethylfurfural. Appl Environ Microbiol. 283:e02312-16.

Huber SG, Wunderlich S, Schöler HF, Williams J. 2010. Natural abiotic formation of furans in soil. Environ Sci Technol. 44:5799-804.

Karich A, Kleeberg SB, Ullrich R, Hofrichter M. 2018. Enzymatic preparation of 2, 5-furandicarboxylic acid (FDCA)—a substitute of terephthalic acid—by the joined action of three fungal enzymes. Microorganisms. 6:5.

Koenig K, Andreesen JR. 1989. Molybdenum involvement in aerobic degradation of 2-furoic acid by *Pseudomonas putida* Fu1. Appl Environ Microbiol. 55:1829-34.

Koenig K, Andreesen JR. 1990. Xanthine dehydrogenase and 2-furoyl-coenzyme A dehydrogenase from *Pseudomonas putida* Fu1: two molybdenum-containing dehydrogenases of novel structural composition. J Bacteriol. 172:5999-6009.

Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ. 2010. Efficient whole-cell biotransformation of 5-(hydroxymethyl) furfural into FDCA, 2, 5-furandicarboxylic acid. Bioresource Technol. 101:6291-6.

Koopman F, Wierckx N, de Winde JH, Ruijssenaars HJ. 2010. Identification and characterization of the furfural and 5-(hydroxymethyl) furfural degradation pathways of *Cupriavidus basilensis* HMF14. Proc Natl Acad Sci USA. 107:4919-24.

Le Minor L, Popoff MY. 1987. Designation of *Salmonella enterica* sp. nov., nom. rev., as the Type and Only Species of the Genus *Salmonella*: Request for an Opinion. Int J Syst Evol Microbiol. 37:465-8.

Lee SA, Wrona LJ, Cahoon AB, Crigler J, Eiteman MA, Altman E. 2016. Isolation and characterization of bacteria that use furans as the sole carbon source. Appl Biochem Biotechnol. 178:76-90.

Nichols NN, Mertens JA. 2008. Identification and transcriptional profiling of *Pseudomonas putida* genes involved in furoic acid metabolism. FEMS Microbiol Lett. 284:52-7.

Schroeter J. 1872. Ueber einige durch Bacterien gebildete Pigmente. Beitr Biol. Pflanz. 1:109-26.

Sessitsch A, Coenye T, Sturz AV, Vandamme P, Barka EA, Salles JF, Van Elsas JD, Faure D, Reiter B, Glick BR, Wang-Pruski G. 2005. *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. Int J Syst Evol Microbiol. 55:1187-92.

Trevisan V. 1889. I generi e le specie delle Batteriacee. Zanaboni Gabuzzi. 1-35.

Trudgill PW. 1969. The metabolism of 2-furoic acid by *Pseudomonas* F2. Biochem J. 113:577.

Wierckx N, Koopman F, Ruijssenaars HJ, de Winde JH. 2011. Microbial degradation of furanic compounds: biochemistry, genetics, and impact. Appl Microbiol Biotechnol. 92:1095-105.

Wierckx NJ, Schuurman TD, Kuijper SM, Ruijssenaars HJ, inventors; Purac Biochem BV, assignee. 2016. Genetically modified cell and process for use of said cell. United States patent US 9,309,546.

Yang CF, Huang CR. 2017. Isolation of 5-hydroxymethylfurfural biotransforming bacteria to produce 2, 5-furan dicarboxylic acid in algal acid hydrolysate. J Biosci Bioeng. S1389-172330737-5.

Yuan H, Li J, Shin HD, Du G, Chen J, Shi Z, Liu L. 2018. Improved production of 2, 5-furandicarboxylic acid by overexpression of 5-hydroxymethylfurfural oxidase and 5-hydroxymethylfurfural/furfural oxidoreductase in *Raoultella ornithinolytica* BF60. Bioresource Technol. 247:1184-8.

Zeitsch KJ. 2000. The chemistry and technology of furfural and its many by-products. Elsevier.

CHAPTER 3. CLONES AND MUTANTS THAT IMPROVE ACETATE CONSUMPTION IN *ESCHERICHIA COLI* AT HIGH ACETATE CONCENTRATIONS

Parts of this chapter have been published in Applied Biochemistry and Biotechnology:

Rajaraman E, Agarwal A, Crigler J, Seipelt-Thiemann R, Altman E, Eiteman MA. 2016. Transcriptional analysis and adaptive evolution of *Escherichia coli* strains growing on acetate. Appl Microbiol Biotechnol. 100:7777-85.

Manuscript in preparation:

Crigler J, Hani FH, Altman E, Eiteman MA. Clones and mutants that improve acetate consumption in *Escherichia coli* at high acetate concentrations.

3.1 Abstract

Acetate growth inhibition is problematic in the production of fuels from lignocellulose, due to the presence of acetylated xylans in the hemicellulose fraction. A pretreatment step is required to hydrolyze the highly recalcitrant biomass, and the acetyl groups are hydrolyzed during this process and thus are present in the hydrolysate. Acetate inhibition is also problematic in industrial recombinant protein fermentations due to acetate excretion by the recombinant host. The effect of the acetate anion on cellular inhibition in *Escherichia coli* was investigated by genomic library screening for clones with improved growth with acetate as the sole carbon source. The focus of this work was improving acetate consumption at a highly inhibitory concentration of acetate. 16 clones were generated with improvement of up to a 3.4-fold increase in growth on acetate at 10 g/l. Most clones fell into five categories: outer membrane, transporter, alternative carbon source catabolism, stress response, and DNA repair. Additionally, mutant strains were generated with improved growth on acetate by up to 4.6-fold. The three fastest growing mutants were in regulatory sequences of the gluconeogenic gene *pck* and the cold-shock gene *ynaE*, as well as an E12A mutation in CRP, the global regulator of secondary carbon sources.

3.2 Introduction

Acetate can be both an inhibitor of growth and an energy source. Acetate metabolism in bacteria is studied for reducing "overflow metabolism" causing acetate dissimilation in

recombinant protein fermentations and for engineering increased tolerance and/or assimilation of inhibitory levels of acetate in pretreated lignocellulosic biomass for the production of green fuels. Additionally, acetate can be used as an inexpensive carbon source for biological production of platform chemicals like polyhydroxyalkanoates and itaconic acid (Lemos *et al.,* 1998; Dai *et al*., 2007; Noh *et al*., 2018).

Acetic acid is inhibitory to *E. coli* at concentrations as low as 0.5 g/l, due to decreasing intracellular pH, uncoupling the transmembrane pH gradient, and the resulting energy demands of responding to the acid stress to maintain homeostasis (Roe *et al*., 1998; Salmond *et al*., 1984; Axe and Bailey, 1995). Internal acidification also causes DNA depurination and double stranded lesions (Choi *et al*., 2000). At neutral pH used in most bioprocessing facilities, where the dissociated form acetate predominates, inhibition can occur from 5 g/l in rich media and 1.2 g/l in minimal media (Luli and Strohl, 1990; Diez-Gonzalez and Russell, 1997). Acetic acid freely permeates the cell, and the anion enters via ion-coupled transport at nearly the same rate as the acid (Gimenez *et al*., 2003; Jolkver *et al*., 2009; Axe and Bailey, 1995). Inhibition of the anion is due to disturbing cytoplasmic osmolarity via increasing internal osmotic pressure and inhibiting enzymes involved in methionine biosynthesis and protein synthesis, as well as inhibiting other unknown enzymes (Roe *et al*., 1998; Roe *et al*., 2002).

Many researchers have addressed engineering acetate tolerance in bacteria (Joachimsthal *et al*., 1998; Treves *et al*., 1998; Yang *et al*., 2010; Chen *et al*., 2011; Zhang *et al*., 2011; Fernandez-Sandoval *et al*., 2012; Luan *et al*., 2013; Mordukhova and Pan, 2013; Steiner

and Sauer, 2013; Lennen and Herrgård, 2014; Rau *et al*., 2016; Noh *et al*., 2018) or altering growth conditions to improve tolerance (Han *et al*., 1993; Xue *et al*., 2010; Sandoval *et al*., 2011; Wang *et al*., 2014). Fewer papers have examined the effects on acetate metabolism directly, using acetate as the sole carbon source. Genomic library screening at 1.75 g/l minimal acetate, pH 7.0, medium by Sandoval *et al*. (2011) found the acetate assimilating acetyl-CoA synthetase, *acs*, and the TCA cycle gene *fumB* overexpressed. In addition to genes involved directly in acetate metabolism, the authors found overexpressing genes of peptidoglycan biosynthesis, dipeptide transport, arginine biosynthesis, multidrug efflux transport, and tetrahydrofolate biosynthesis improved acetate consumption. Chong *et al*. (2013) found that a D138Y mutation in the global regulator CRP, altering expression levels of over 400 genes, led to a five-fold increase in growth rate on acetate as the sole carbon source at 11 g/l. Directed evolution in acetate minimal medium at 5.1 g/l caused a single S266P mutation in the RNA polymerase subunit RpoA, increasing the growth rate by 25 percent (Rajaraman *et al*., 2016).

Acetate assimilation begins with AMP-forming acetyl-CoA synthetase. Growth on two-carbon compounds such as acetate requires the glyoxylate bypass of the decarboxylation steps of the TCA cycle to replenish the TCA cycle intermediates oxaloacetate and alpha-ketoglutarate for anabolic reactions. Growth also requires gluconeogenesis to generate sugar phosphates as well as the intermediates 3-phosphoglycerate and phosphoenolpyruvate for anabolic reactions. Thus acetyl-CoA sits at the intersection of both central catabolism and anabolism. Additionally, the intracellular acetyl-phosphate pool, which is highly transient and depends on the energy state of the cell, is a global

signal, functioning by regulating two-component signal transduction pathways and possibly by other unknown mechanisms (Feng *et al*., 1992; Bouche *et al*., 1998). Acetyl-phosphate is generated in the primarily acetate dissimilating phosphotransacetylase-acetate kinase pathway.

Acetate provokes changes in gene expression of 1/10 to 1/5 of the genome (Yang *et al*., 2014; Treitz *et al*., 2016). There are many possible strategies for engineering increased acetate tolerance and/or consumption, and most require engineering multiple related and unrelated genes and pathways. Genomic libraries provide the benefit of genome-wide screening and simultaneous production of functional clones with the desired phenotype. A cloning selection scheme was devised based on the differential acetate tolerance of *E. coli* MG1655 and *E. coli* MC1061, a high transformation efficiency strain. In minimal medium with acetate, pH 7.0, as the sole carbon source MC1061 cannot grow above 8 g/l acetate, and MG1655 cannot grow above 14 g/l. A genomic library of MG1655 in the high copy number plasmid pTrc99a was transformed into MC1061 and plated on 10 g/l acetate minimal medium to select for clones with improved acetate consumption. Some transformants with improved growth on acetate did not contain a plasmid insert. The genomes of the four with the greatest increase in growth were sequenced for variant determination.

3.3 Materials and Methods

3.3.1 Bacterial strains and growth conditions

Parental bacterial strains used in this study are listed in Table 3.1. Lysogeny broth (LB) was used for rich medium and M9 (Miller, 1972) for minimal medium. Acetate, pH 7.0, was added at 10 g/l, from Sigma-Aldrich sodium acetate trihydrate stock. Culture of all strains was at 37ºC. Ampicillin, used at 100 μg/ml in LB and 50 μg/ml in M9, and 1 mM IPTG were added when indicated. Thiamine (2.5 mg/ml) and leucine (344 mg/ml) were added to M9 media for culture of MC1061. Tetracycline for transduction selection was used at 20 μg/ml.

3.3.2 Construction and screening of MG1655 genomic library

Sau3AI fragments of 3 - 8 kb of MG1655 gDNA were ligated into pTrc99a digested with BamHI and dephosphorylated with shrimp alkaline phosphatase. Genomic libraries were transformed into MC1061 and serial dilutions selected on LB ampicillin plates to verify greater than 10^5 transformants per transformation. Genomic libraries were transformed into MC1061, grown for 1 hour in SOC medium, washed in M9 medium with no carbon source, and plated on M9 10 g/l acetate, pH 7.0, medium supplemented with ampicillin, IPTG, thiamine, and leucine. The earliest transformant colonies were re-streaked on the selection plates. Re-streaked colonies that grew faster than MC1061 with an empty pTrc99a plasmid were screened for plasmid insertions. Transformants with plasmid insertions were cured of the plasmid to verify the increased acetate growth phenotype was lost. Plasmid inserts were sequenced using conventional sequencing methods.

Table 3.1. Parental bacterial strains used in this study

Strain	Genotype	Reference
MC1061	F- *araD139* Δ(*ara-leu*)*7696* Δ(*lac*)X74 *galU galK rpsL hsr- hsm*+	Laboratory stock
MG1655	F- l - (wild-type)	Guyer *et al*., 1981
CAG18456	F-, λ - *cysG yhfT-3084*::Tn*10*	Singer *et al*., 1989
CAG12081	F-, λ - *recT-3061*::Tn*10*	Singer *et al*., 1989

3.3.3 Growth rate experiments

Clone plasmids were transformed into MG1655 for growth rate experiments. Overnight cultures were washed twice in M9 medium with no carbon source and diluted 1/200 in M9 10 g/l acetate, pH 7.0, medium with 50 μg/ml ampicillin and 1 mM IPTG. The growth rate was calculated as the slope of the plot of the time in hours versus the natural logarithm of the OD_{600} reading. Growth rates for the mutants were performed with the mutation transduced into wild-type MG1655 and grown in the same medium as above, excluding ampicillin and IPTG.

3.3.4 Genome sequencing and variant detection

Paired-end Nextera XT libraries of the gDNA of the cured mutant strains and the parent strain MC1061 were run on a MiSeq system using a v2 500-cycle MiSeq Reagent Kit. CLC Genomics Workbench version 10.1.1 tools Fixed Ploidy Variant Detection and Indels and Structural Variants were used for variant detection. Variants were verified with conventional Sanger sequencing.

3.3.5 Variant transductions

P1 lysates of CAG18456 (34% linked to *pck* and 51% linked to *crp*) and CAG12081 (45% linked to *ynaE*) were transduced into the mutant strains KD1137, KD1167, and KD1151, respectively. 8 transductants were streaked on M9 10 g/l acetate (pH 7.0) thiamine and leucine plates to verify which retained the improved acetate growth phenotype. Lysates were made on one of each and transduced into wild-type MC1061 and MG1655. Transductants in MC1061 were streaked on the same plates to verify gain

of function. Transductants in MG1655 were screened in growth rate cultures in 10 g/l M9 acetate medium to verify gain of function.

3.4 Results

3.4.1 Genomic library clones with improved acetate consumption

750 transformants were streaked on the acetate selective plates and 136 with increased growth were screened for inserts. Of these, 26 contained inserts. When cured of the plasmid, all lost the increased acetate growth phenotype. The clone plasmids were transformed into wild-type MG1655 for growth rate experiments. Table 3.2 shows the gene(s) contained in the top 16 clones and their respective growth rates on acetate. Genes in the clones fall into five general categories: outer/inner membrane (*yehDC* fimbrial adhesins, colanic acid biosynthetic gene *ypdI*, and cold temperature LPS biosynthetic gene *lpxP*), catabolic (*p*-aminobenzoyl-glutamate catabolic genes *abgAB*, and the fucose and arabinose catabolic gene *fucK*), transporter (galactose/methyl-galactoside ABC transporter subunit *mglB*, putrescine transporter subunits *potFG*, and the arsenite/antimonite:H^+ antiporter *arsB*), stress response (the toxin-antitoxin system *yfjZ-ypjF*, putative stress response chaperone *yfdX*, arsenate stress response gene *arsC*, and the oxidative stress response SoxR reducing system components *rsxDGE*), and DNA repair (resolvasome genes *ruvC* and *yebC* and putative DNA repair gene *yfjY*). Additionally, one clone contained the viral defense genes *casABCD* and another contained the folate biosynthetic gene *nudB*. One clone contained exclusively genes unknown function,

Table 3.2. Clones with improved growth on acetate

Gene(s) contained in insert	Insert flipped	Function(s)	Growth rate (/h)	Location, u00096.3 chromosome
amiD' *ybjS*'	✓	peptidoglycan catabolism/ predicted NAD(P)H-binding oxidoreductase	0.109	905176..906599
casE' *casD casC casB casA cas3*'	✓	CRISPR antiviral defense complex	0.107	2879918..2886219
yagF'	✓	xylonate catabolism	0.104	284182..284913
metI' *metN*'		methionine ABC transporter subunits	0.103	221275..221808
mglA' *mglB galS*		galactose ABC transporter subunits/ repressor of galactose assimilation	0.0765	2238438..2242715
abgT'*abgB abgA abgR*'	✓	*p*-aminobenzoyl-glutamate assimilation	0.0650	1401337..1405297
pabC' *yceG*'	✓	folate biosynthesis/ peptidoglycan polymerization terminase	0.0622	1153875..1154297
ruvA' *yobI yebB ruvC yebC nudB*		resolvasome component/ uncharacterized protein/ predicted amidase/resolvasome component/ regulator of *ruvABC*/folate biosynthesis	0.0600	1945714..1949375

Table 3.2 continued

yfjW yfjX yfjY ypjJ yfjZ ypjF		uncharacterized protein/predicted antirestriction protein/ predicted DNA repair protein/ uncharacterized protein/toxin-antitoxin system	0.0590	2772654..2778759
yhfW' yhfX yhfY yhfZ		predicted mutase/predicted DNA repair protein/ uncharacterized protein/predicted transcriptional regulator	0.0566	3508803..3513001
fucI' fucK fucU'	✓	fucose and arabinose catabolism	0.0556	2935885..2939130
yehB' yehC yehD yehE		cryptic fimbrial adhesins/ uncharacterized protein	0.0530	2189165..2193522
frc' yfdX ypdI yfdY lpxP'		formyl-CoA:oxalate-CoAtransferase/ putative chaperone/ colanic acid biosynthesis/ uncharacterized protein/ cold-induced LPS biosynthesis	0.0493	2492998..2496480
ybjN' potF potG potH'		pleiotropic suppressor/ putrescine ABC transporter subunits	0.0493	893122..896549
arsR arsB arsC'	✓	arsenate and antimonite metals resistance	0.0491	3648669..3651998
rsxC rsxD rsxG rsxE' nth'	✓	SoxR reducing system/ DNA repair	0.0439	1708090..1711840
E. coli MG1655 with pTrc99a		n/a (control strain)	0.0303	n/a

Table 3.2 continued

' indicates incomplete gene contained in plasmid insert. Genes listed in orientation on plus strand of u00096.3 chromosome. Insert flipped indication is relative to orientation of the trc promoter of pTrc99a. Growth rates performed with genomic library clone plasmids from *E. coli* MC1061 transformed into *E. coli* MG1655. Growth rates for clones in M9 10 g/l acetate, pH 7.0, medium supplemented with 50 μg/ml ampicillin and 1 mM IPTG. Growth rate is taken from $time_0$ to an OD_{600} of 0.5, with the OD_{600} at t_0 set to 0.005.

yhfXW and *yhfZY*. YhfX may be a DNA repair protein (Kumar *et al*., 2016). Other genes of unknown function were *yobI*, *yebB*, *yfjW*, *yfjX*, *ypjJ*, *yehE*, and *yfdY*.

In addition, four clones did not contain a complete gene in the insert. The clone that had the highest growth rate contained the incomplete genes *ybjS* and *amiD*, with *ybjS* on the plus strand relative to expression driven by the trc promoter of pTrc99a but lacking the first 154 bp and *amiD* on the minus strand, lacking the first 263 bp. YbjS is a predicted outer membrane oxidoreductase and AmiD is involved in peptidoglycan turnover. This construct also had the highest growth rate of the clones when tested in 5 and 15 g/l acetate (data not shown). Three other clone inserts also do not contain a complete gene, containing *yagF* (xylonate catabolism), *metI*/*metN* (methionine ABC transporter subunits), and *pabC*/*yceG* (folate biosynthesis/ peptidoglycan polymerization terminase) and having the 3rd, 4th, and 7th highest growth rates on 10 g/l acetate, respectively. *yagF* contains only bp 982 to 1713 of the middle of the 1,968 bp *yagF* gene, in the reverse orientation of the trc promoter. The first 347 bp of *metI* and last 195 bp of *metN* are located on the minus strand relative to the trc promoter, as are base pairs 1 - 182 of *yceG* and 581 - 810 of *pabC* on the forth construct. The *amiD*/*ybjS, metN*/*metI*, and *pabC*/*yceG* clones all lose the increased acetate growth phenotype when the trc promoter is removed, whereas the *yagF* clone retains the phenotype (data not shown). pTrc99a has an ATG site between the trc promoter and the BamHI restriction site that was used for plasmid library construction that can that can transcribe the plasmid insert into an mRNA capable of being translated into a protein. If the gene fragment inserted at the BamHI site is in-frame with this ATG, then a truncated version of the protein can be made. The *ybjS* gene was

not in-frame and so this construct can not produce any protein or truncated protein product.

SDS-PAGE was performed, and in only one clone was increased protein expression detected: the clone containing *mglB* and *galS*. (Figure 3.1; data not shown for all clones). MglB and GalS are 35.7 and 37.4 kDa, respectively. Also in the construct is the incomplete gene *mglA*, which would produce a 32.0 kDa protein product. All three genes are in the reverse orientation relative to the trc promoter of pTrc99a, but expression of each is possible with the native promoters in the insert. The band for this clone was estimated at 32.3 kDa.

3.4.2 Growth rates of clones compared to reduction in growth lag time

The growth rates listed in Table 3.2 are taken from the starting inoculation time (all set at an OD_{600} of 0.005) to growth to an OD_{600} of 0.5. When taken in the exponential growth phase, from an OD_{600} of 0.1 to 0.5, the growth rates of the clones range from 59.9% lower to 39.7% higher than the control (Table 3.3).

The OD_{600} 0.5 cutoff value was chosen because growth was nonlinear on a semi-log plot after this point. The general trend is that the growth rate taken in the exponential phase increases as time until growth increases. However, the time until growth relative to the control ranges from 108 to 45 hours improvement. As the growth rate in exponential phase in most clones is not significantly different from the control but the time until growth is significantly lower, most clones' effects appear to be due to decreasing the

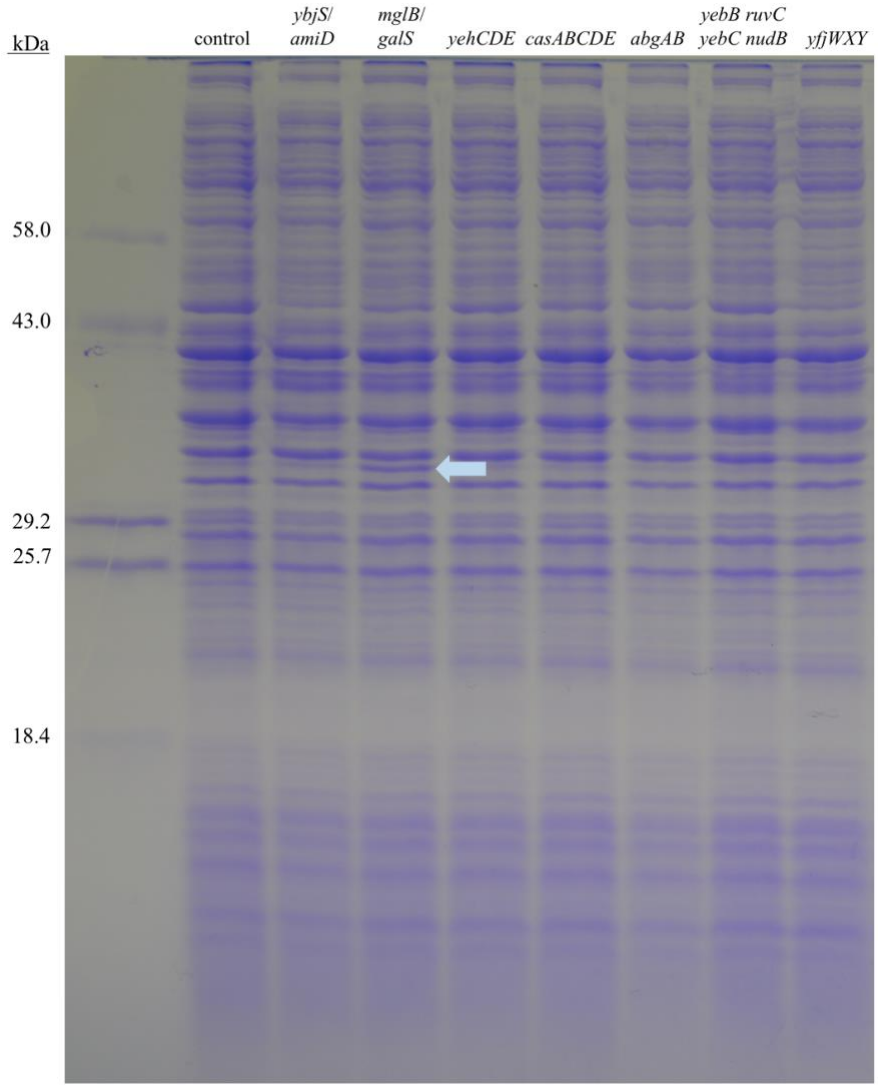

Figure 3.1. SDS-PAGE of select genomic library clones with improved acetate consumption

Samples were grown in LB medium supplemented with ampicillin and IPTG and grown to an $O.D._{600}$ of 0.45 - 0.5. SDS-PAGE of the 9 other genomic library clones was also performed and no protein overexpression was detected.

Table 3.3. Comparison of the clones' growth rates in exponential phase and reduction in growth lag times

Gene(s) contained in insert	Growth rate (/h) when measured from t_0 to OD_{600} of 0.5	Growth rate (/h) during exponential phase, OD_{600} of 0.1 - 0.5	Hours until growth
amiD' ybjS'	0.109	0.0816	44
casE' casD casC casB casA cas3'	0.107	0.101	45
yagF'	0.104	0.0851	45
metI' metN'	0.103	0.0828	46
mglA' mglB galS	0.0765	0.0510	61
abgT' abgB abgA abgR'	0.0650	0.0803	72
pabC' yceG'	0.0622	0.0847	75
ruvA' yobI yebB ruvC yebC nudB	0.0600	0.0653	79
yfjW yfjX yfjY ypjJ yfjZ ypjF	0.0590	0.109	82
yhfW' yhfX yhfY	0.0566	0.119	83
yhfZ			
fucI' fucK fucU'	0.0556	0.0522	87
yehB' yehC yehD yehE	0.0533	0.0642	94
frc' yfdX ypdI yfdY lpxP'	0.0493	0.0843	95
ybjN' potF potG potH'	0.0493	0.118	95

Table 3.3 continued

arsR arsB arsC'	0.0491	0.118	106
rsxC rsxD rsxG rsxE nth'	0.0439	0.109	107
E. coli MG1655 with pTrc99a	0.0303	0.0852	152

' indicates incomplete gene contained in plasmid insert. Growth rates performed with genomic library clone plasmids from *E. coli* MC1061 transformed into *E. coli* MG1655. Growth rates in M9 10 g/l acetate, pH 7.0, medium supplemented with 50 μg/ml ampicillin and 1 mM IPTG. Hours until growth is defined as hours until reaching an OD_{600} of 0.1.

growth lag caused by acetate inhibition, as opposed to increasing acetate catabolism. No clone contained genes involved directly in acetate catabolism.

3.4.3 Mutants with improved acetate consumption

When streaked on the selection plates, the control strain, MC1061 with an empty pTrc99a plasmid, grew colonies after 7 days. In all of the 136 transformants screened for inserts, re-streaked colonies appeared after 3 - 4 days. Re-streaked colonies on all clones containing inserts appeared at 4 days. 48 transformants had colonies that appeared in 3 days, none of which contained plasmid inserts. Four of these were cured of the plasmid to verify the phenotype was retained, and the genome was sequenced. One strain contained a T to G mutation 136 bp upstream of *pck* and another contained a C to T mutation 304 bp upstream of *ynaE*. Two strains had the same mutation, an E12A mutation in CRP. Only one (KD1167) was further characterized. The mutations were transduced into wild-type MC1061 and MG1655 backgrounds, which all recreated the increased acetate growth phenotype. The variants and growth rates of the variants transduced into MG1655 are shown in Table 3.4.

Table 3.4. Mutants with improved growth on acetate

Strain	Mutant	Function	Growth rate (/h)	Variant, u00096.3 chromosome
KD1137	136 bp upstream of *pck*	gluconeogenesis	0.194	3532682 T to G
KD1151	304 bp upstream of *ynaE*	cold shock protein; putative transcriptional regulator	0.143	1434528 C to T
KD1167, KD1142	CRP E12A	global regulator	0.201	3486157 A to C
E. coli MG1655		n/a (control strain)	0.0415	n/a

Growth rates in M9 10 g/l acetate, pH 7.0, medium. Growth rate is taken from $time_0$ to an OD_{600} of 0.5. Variants were transduced into wild-type MG1655.

3.5 Discussion

To produce functional clones with improved acetate consumption at a highly inhibitory acetate concentration, *E. coli* MG1655 genomic libraries were screened on minimal plates with acetate as the sole carbon source at 10 g/l. (The growth rate of MG1655 grown in M9 10 g/l acetate, pH 7.0, medium is 0.0415/h and in M9 0.2% glucose medium is 0.446/h.) A pH of 7.0 was used to find engineering targets related solely to the effect of the acetate anion on inhibition and metabolism. Four inserts contained no complete gene, and they are among the fastest growing clones. There is little information on *ybjS*, contained in the best-performing clone, which encodes an uncharacterized, putative NAD(P)H-dependent oxidoreductase and predicted outer membrane protein. It has been correlated with UV light protection and tolerance to 1-napthol and is positively regulated by CpxR, an envelope stress response regulator (Rooney *et al*., 2011; Gall *et al*., 2008; Bury-Moné *et al*., 2009). Interestingly, in a proteomics study it was found to be expressed less in acetate-grown cultures (at 2.7 g/l acetate) compared to glucose-grown cultures (Table 3.5) (Treitz *et al*., 2016). *amiD*, involved in peptidoglycan turnover, while it cannot be expressed in this construct, was also found down-regulated in growth on acetate. The ends of the divergently expressed genes overlap by 3 bp, and there is no putative regulatory site in the insert. *ybjS*, when cloned via PCR with its native promoter and Shine-Dalgarno sequence as well as with the trc promoter and consensus Shine-Dalgarno sequences, showed no improvement in growth on acetate (data not shown). The function of this construct is unknown; however, due to its pronounced effect on increasing growth on acetate at high and low concentrations, it warrants further study.

Table 3.5. Genes in clones previously found to be up- or down-regulated in minimal acetate medium

Down-regulated	Expression ratio	Reference
abgR	0.97	Treitz *et al*., 2016
amiD	0.58	Treitz *et al*., 2016
arsC	0.70	Treitz *et al*., 2016
frc	0.34	Treitz *et al*., 2016
metN	0.84	Treitz *et al*., 2016
ybjS	0.72	Treitz *et al*., 2016
yceG	0.85	Treitz *et al*., 2016
yfdX	0.98	Treitz *et al*., 2016
yhfZ	0.67	Treitz *et al*., 2016
ypjF	0.72	Treitz *et al*., 2016
Up-regulated	**Expression ratio**	**Reference**
casA	1.60	Rajaraman *et al*., 2016
casB	1.12	Rajaraman *et al*., 2016
casC	1.16	Rajaraman *et al*., 2016
casD	1.12	Rajaraman *et al*., 2016
casE	1.06	Rajaraman *et al*., 2016
fucI	1.42	Treitz *et al*., 2016
fucU	1.09	Treitz *et al*., 2016
mglA	2.26	Treitz *et al*., 2016

Table 3.5 continued

mglB	4.16	Treitz *et al*., 2016
nudB	1.08	Treitz *et al*., 2016
potF	1.42	Treitz *et al*., 2016
ruvA	1.09	Treitz *et al*., 2016
yebC	1.18	Treitz *et al*., 2016
ypjA	1.66	Treitz *et al*., 2016

In Treitz *et al*. (2016), *E. coli* MG1655 was grown in M9 2.7 g/l acetate minimal medium relative to M9 glucose minimal medium. In Rajaraman *et al*. (2016) strains with high (*E. coli* ATCC8739) and low (*E. coli* SMS-3-5) acetate tolerance were compared in M9 5.1 g/l acetate minimal medium.

metN, encoding a methionine ABC transporter subunit, and *yceG*, encoding a peptidoglycan polymerization terminase were in two of the other three clones with no complete gene in the insert and were also found to be down-regulated in acetate minimal medium relative to glucose minimal medium (Table 3.5) (Treitz *et al*., 2016). Additionally, *metJ*, the repressor of the methionine ABC transporter operon, was found to be up-regulated. Similar to the *amiD*/*ybjS* clone, the function of these constructs points to an unknown regulatory effect.

The role of the clone containing the antiviral defense genes *casABCD*, which was the clone with the second highest growth rate, is also unclear. However, this operon has previously been seen to be overexpressed in growth on minimal acetate in an acetate-tolerant *E. coli* strain (Table 3.5) (Rajaraman *et al*., 2016).

The three clones containing catabolic genes (*mglB*, *fucK*, and *abgAB*) have roles in carbon scavenging. The galactose/methyl-galactoside ABC transporter *mgl* operon and the fucose and arabinose catabolic *fuc* operon have been shown to be highly induced in growth on minimal acetate medium, as well as in a starvation state (Table 3.5) (Treitz *et al*., 2016; Liu *et al*., 2005; Oh *et al*., 2002; Death and Ferenci, 1994). In minimal acetate medium *E. coli* will induce expression of many alternative carbon utilization genes, including *mgl*, *fuc*, and *ara* genes, even if those carbon sources are not in the growth medium (Liu *et al*., 2005). The same happens in carbon-limited transition to stationary phase (Hua *et al*., 2004; Death and Ferenci, 1994). The mgl ABC transporter is a high-affinity transporter for both galactose and glucose, induced for scavenging micromolar or

submicromolar concentrations of the sugars (Death and Ferenci, 1994). The clone also contains the negative regulator of the gal regulon, *galS*. *galS* has been shown to be slightly repressed in these growth conditions and, given the overexpression of its regulon, its overexpression in this construct is not likely the cause for the increased acetate growth phenotype. The growth lag times in the high concentration of acetate used in this experiment were 61 and 87 hours for the *mglB*/*galS* and *fucK* clones, respectively (152 hours for the control) (Table 3.3). Even though the carbon source acetate is abundant, the cells must first overcome its inhibition. During this time cell growth is arrested. Although induction of gene expression of alternative carbon sources not present in the growth medium in response to growth on minimal acetate medium has been observed (Oh *et al.*, 2002; Liu *et al*., 2005), one would expect it would be more of a response to starvation than to acetate. It is unclear what the role of *mglB* and *fucK* is in these clones. The role of the *abgAB* clone, which is involved in folate catabolism, seems more straightforward. Starving cells will degrade and turnover endogenous proteins during starvation and in the process may catabolize the coenzyme (Reeve *et al*., 1984).

Overexpressing the folate biosynthetic gene *folM* has been shown to confer tolerance during growth on acetate (Sandoval *et al*., 2011). One clone contained *nudB*, which encodes the committed enzymatic step in folate biosynthesis. Folate is essential for nucleic acid and amino acid biosynthesis, both of which are repressed during growth on acetate (Oh *et al*., 2002). Some of the inhibition of the acetate anion has been attributed to interfering with methionine biosynthesis, which also increases pools of its toxic precursor homocysteine (Roe *et al*., 2002). Tetrahydrofolate derivatives are cofactors of

the two enzymes that catalyze this last step of the pathway. This clone also contained the DNA repair resolvasome component *ruvC* and its regulator *yebC*. RuvA and YebC are also overexpressed in acetate-grown cultures (Table 3.5) (Treitz *et al*., 2016).

A common finding was genes involved in outer membrane structure and cell adhesion. The colanic acid biosynthetic gene *ypdI* was found in one clone. Colanic acid plays a role in survival outside of the host and has a low basal level of expression in growth over 30°C, which can be offset by mutations in several regulators (Gottesman, 1995; Markovitz, 1977; Gottesman *et al*., 1985). It contributes to the arrangement of cells in biofilms and is induced during osmotic shock (Prigent-Combaret *et al*., 2000; Sledjeski and Gottesman, 1996). Also in this clone were *yfdX* and *frc*, genes in an oxalic acid resistance operon. Frc has been shown to be heavily repressed and YfdX mildly repressed during growth on minimal acetate (at neutral pH) and thus is not likely responsible for the increased acetate growth phenotype in this clone (Trietz *et al*., 2016). Also present is the uncharacterized gene *yfdY*. The only information on this gene in the literature is that it is induced in biofilms (Ren *et al*., 2004). Another clone contains the fimbrial adhesion genes *yehC* and *yehD*. *yehABCD* is a cryptic operon but is functional and promotes biofilm formation when overexpressed on a plasmid (Korea *et al*., 2010). *yehE*, contained in the insert with *yehC* and *yehD* but which is not in the same operon, has no known function, but is downregulated when the stringent starvation protein SspA is deleted (Hansen *et al*., 2005).

Clones also contained stress response functions. SoxR senses superoxide and nitric oxide and activates SoxS, a transcriptional regulator which activates superoxide resistance genes, including superoxide dismutase *sodA*, as well as genes involved in resistance to antibiotic and organic solvent stress (Greenberg *et al*., 1990; Martin and Rosner, 2002; Aono, 1998). One clone contained members of the *rsx* operon, which act to reduce SoxR to maintain the reducing environment during oxidative stress. When cells are in a starvation state, respiration and ROS generation are at a higher rate than DNA and protein biosynthesis (Dukan and Nyström, 1999). As growth on minimal acetate medium causes a similar metabolic state as starvation, this scenario of acetate provoking oxidative stress is possible (Liu *et al*., 2005; Oh *et al*., 2002; Hua *et al*., 2004). Also supported by this is the over-production of superoxide dismutases SodA, SodB, and SodC in growth on minimal acetate (Trietz *et al*., 2016). One clone contained the entire operon *yfjXY-ypjJ-yfjZ-ypjF*. *yfjZ* and *ypjF* encode a Type IV toxin-antitoxin system, which has been shown to respond to acid stress (Wang *et al*., 2010). Deletion of both genes independently increases biofilm formation (Wang *et al*., 2009). There is little information on the first three genes in this operon. *yfjX* is a predicted antirestriction protein and a *yfjY* mutant confers a mutator phenotype when overexpressed (Yang *et al*., 2004). In a clone containing the entire operon, the YfjY mutator phenotype was increased 200-fold when *yfjZ* and *ypjF* were disrupted (Yang *et al*., 2004).

One clone contains the putrescine- (*potF)* and ATP-binding (*potG*) subunits of the putrescine importer PotFGHI but lacks the majority of the membrane channels *potH* and *potI*. In a proteomics study of growth of *E. coli* on minimal acetate medium relative to

minimal glucose medium, the protein with the largest decrease in growth on acetate was the putrescine importer PlaP (Treitz *et al*., 2016). Putrescine biosynthetic enzymes were all heavily repressed, and putrescine catabolic pathways were induced. Putrescine is a polycation with roles in regulating nucleic acid and protein biosynthesis. Putrescine is also exported from the cell during hyperosmotic shock, and internal pools increase in stationary phase (Munro *et al*., 1972; McLaggan *et al*., 1994; Tweeddale *et al*., 1998). When both putrescine catabolic pathways are deleted, *E. coli* is more sensitive to oxidative, temperature, and antibiotic stresses (Schneider *et al*., 2013). There are at least five putrescine importers, each with different cellular roles (Furuchi *et al*., 1991; Pistocchi *et al*., 1993; Kashiwagi *et al*., 1997; Terui *et al*., 2014; Kuihara *et al*., 2011). The role of the putrescine importer PlaP that is highly repressed in growth on acetate is the induction of cell motility in response to extracellular putrescine (Kuihara *et al*., 2011). PotFGHI and PuuP are the major putrescine importers, with PotFGHI maintaining intracellular putrescine levels during growth on glucose (Terui *et al*., 2014). PotFGHI is required for fast growth on glucose, although the mechanisms are not fully understood (Terui *et al*., 2014). PuuP is required for putrescine assimilation during growth without glucose (Terui *et al*., 2014). Thus it seems this clone may increase putrescine import for a regulatory role.

As seen in Tables 3.2 and 3.4, *E. coli* MG1655 has a 1.4-fold increase in growth rate in 10 g/l minimal acetate medium when grown without IPTG. Additionally, PCR-targeted cloning constructs of *abgABT*, *mglBAC*, and *cas3-casABCDE12* all resulted in a decreased growth rate relative to the vector-only control, likely due to the energy burden

of recombinant protein overexpression (data not shown). Furthermore, only 4 of the 16 clones (*casABCD*, *abgAB*, *yfjWXY-ypjJ-yfjZ-ypjF*, and *potFG*) had genes driven by the trc promoter of pTrc99a. Thus the experimental conditions primarily selected for genes driven by the native promoters in the inserts, presumably to relieve excessive overexpression. pTrc99a is a high-copy number plasmid, but gene overexpression from this vector may be too energetically costly for a large number of genes, which would limit true genome coverage.

No clones contained genes involved directly in acetate metabolism, such as acetyl-CoA synthetase, or TCA, glyoxylate bypass, or gluconeogenic genes. Two of the three mutant strains, however, were variants in the gluconeogenic gene *pck* promoter and the global regulator of secondary carbon sources, CRP. The third variant was in the upstream sequence of the poorly characterized cold-shock protein *ynaE*.

KD1137, the first mutant strain with increased acetate consumption, had a T to G mutation in the phosphoenolpyruvate carboxykinase (*pck*) promoter located 137 bp upstream of the start codon. The *pck* promoter has two binding sites for the global regulator cAMP receptor protein (CRP) (Figure 3.2), with CRP box-1 promoting transcription and box-2 repressing transcription, but only in the presence of high cAMP-CRP for the latter (Nakano *et al*., 2014). The consensus CRP binding site is the 22 bp palindrome AAATGTGATCTAGATCACATTT, where the most conserved base pairs are underlined (Nakano *et al*., 2014). The wild-type sequence of box-2 in the *pck*

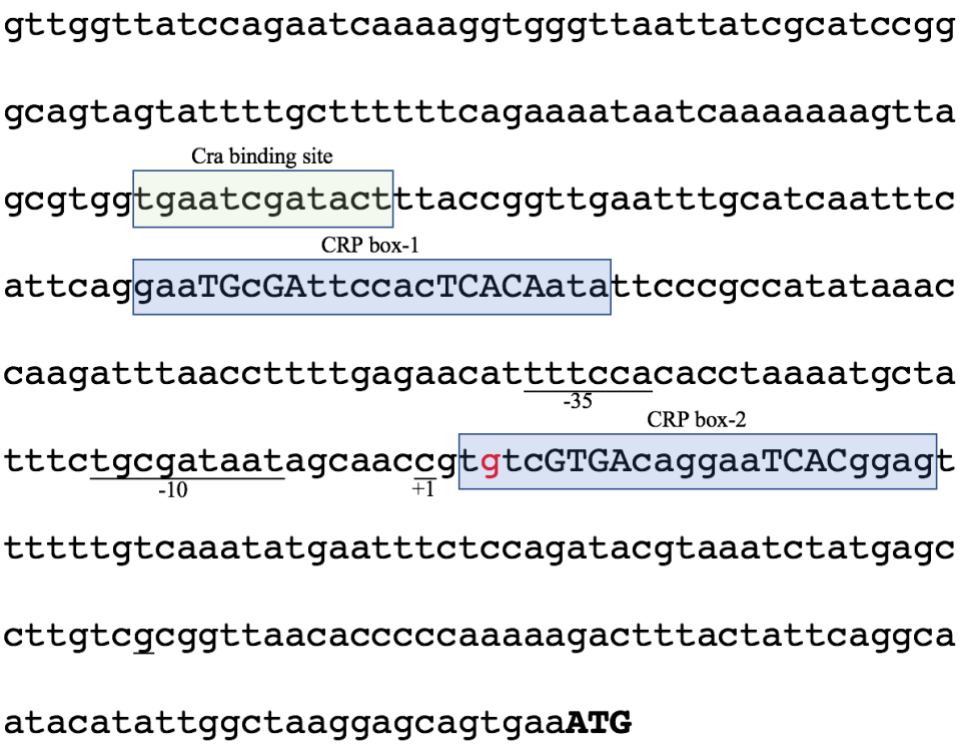

Figure 3.2. KD1137 variant in *pck* promoter and location of regulatory sequences
Upstream sequence from ATG start codon of *pck*. The T to G variant of KD1137 is shown in red. Blue boxes are CRP binding sites in which capital letters represent identical base pairs to the two highly conserved motifs in each CRP box (Nakano *et al*., 2014). The Cra binding site is shown in green (Ramseier *et al*., 1995). -35, -10, and +1 transcription start sites are underlined (Prost and Cozzone, 1999).

promoter is TTTCGTGACAGGAATCACGGAG. The T to G variant of KD1137 is in the second base pair, which is adjacent to the first conserved motif, to give TgTCGTGACAGGAATCACGGAG, where the variant is in lower case.

Growth of *E. coli* on acetate requires gluconeogenesis for the biosynthesis of cell wall polysaccharides as well as various amino acids and phospholipids from the intermediates phosphoenolpyruvate (PEP) and 3-phosphoglycerate. Pck catalyzes the first step of gluconeogenesis, the phosphorylation and decarboxylation of oxaloacetate to PEP, during growth on two-carbon compounds and succinate (PEP synthase, *pps*, is used for PEP generation during growth on three carbon compounds like pyruvate and, in combination with malate dehydrogenase, can also generate PEP from two carbon compounds and succinate) (Goldie *et al*., 1980). Growth on acetate is well known to increase *pck* expression. *pck* is subject to carbon catabolite repression during growth on glucose and is positively regulated by both CRP and Cra, the other major regulator of alternate carbon sources, in the absence of glucose. Interestingly, expression of *pck* in glucose-grown wild-type *E. coli* is similar to that in null mutants of both CRP and Cra independently (Chin *et al*., 1989; Gossett *et al*., 2004). Although during growth on poor carbon sources *pck* is natively induced, *E. coli* has suboptimal flux to Pck (Chao *et al*., 1993). Overexpression of *pck* on a multicopy plasmid increases the specific growth rate and specific oxygen consumption rate on succinate in a *pck pps* strain, peaking at a 10-fold increase over wild-type expression followed by a decrease in growth rate and a slight decrease to a plateau in oxygen consumption rate (Chao *et al*., 1993). Additional upregulation of *pck* by the mutation in KD1137 improves growth on acetate, possibly by

relieving some residual repression of cAMP-CRP bound to CRP box-2 found in the absence of glucose catabolite repression.

The second mutant strain, KD1151, contains a mutation in the *ynaE* 5' UTR. YnaE is a cold shock-induced protein and putative transcription factor which is motif-predicted to control the transcription of *cspB*, *cspG*, *cspH*, *rhsE*, *yncM*, and *ynaD* as a possible stress response regulon. (Polissi *et al*., 2003; Faith *et al*., 2007). It has been found overexpressed in response to the YoeB toxin, the extracellular signal indole, and the antibiotic norfloxacin (Faith *et al*., 2007; Lee *et al*., 2007). *ynaE* has a regulatory RNA 5' leader sequence (Raghavan *et al*., 2011a). Leader sequences form stem loop structures which regulate transcription, mRNA stability, or translation upon ligand binding (riboswitches) or response to temperature (RNA thermometers) and can act in both *cis* and *trans*, although *cis*-acting leaders have been found more often (Serganov and Nudler, 2013; Waters and Storz, 2009). There is no experimental evidence for the regulatory role of the leader sequence of *ynaE*, although the mRNA structure of its 99% identical homolog *ydfK* has been predicted (Raghavan *et al*., 2011b). 544 bp upstream of *ydfK* is 98% identical to that of *ynaE*, and it is also cold shock-induced (Polissi *et al*., 2003; Raghavan *et al*., 2011b). The authors predicted a temperature-sensitive change in mRNA structure causing the ribosomal binding site (RBS) to be bound and blocked by an anti-RBS when grown at 37ºC and relieved when grown at 10ºC. The 5' UTR is predicted to be 317 bp and the variant in KD1151 is 304 bp upstream from the *ynaE* start codon. It is not found within the RBS or anti-RBS but is at the base of a minor stem loop structure that is not predicted to change much upon temperature shock and is outside of the

conserved sequence, -280 to +120 of the translation start site (Raghavan *et al*., 2011b). The roles of this protein in increasing acetate tolerance, as well as the aforementioned upregulation in response to other stressors and any other modes of its regulation, are unclear.

The third mutant strain with improved acetate consumption, KD1167, has an E12A mutation in CRP, the global regulator of secondary carbon sources. E12 does not have a direct role in cAMP binding or in DNA binding (McKay *et al*., 1982; Schultz *et al*., 1991). Despite the vast amount of research performed on CRP structure and function, there is little information on the role of residue E12. This area of the protein (when bound to cAMP) is surface-exposed, away from the DNA and cAMP binding sites, and adjacent to an RNA polymerase binding site (Meibom *et al*., 1999; McKay *et al*., 1982). One paper showed a role of E12 in repressing transcription by binding to the negative regulator CytR, which regulates assimilation of nucleosides and deoxynucleosides (Meibom *et al*., 1999; Hammer-Jespersen, 1973). The CRP-CytR complex represses transcription, with activation requiring the inducer cytidine (Pedersen *et al*., 1991). Residues 12, 13, 17, 105, 108, and 110 were found essential for binding CytR, termed the "repressor region", and interact with one of the three RNA polymerase binding sites of CRP, termed "activating region 2" (essential residues 19, 21, and 101). A W13A mutation had the largest effect on CytR binding. The mutant tested by Meibom *et al*. (1999) was an E12A mutation, so the variant of KD1167 does affect CytR repression. The E12A variant also could potentially affect other regulators or alter the interaction between the repressor region and activating region 2. *cytR* is down-regulated in growth

on acetate, and members of its regulon (*cdd*, *deoCABD*, *nupC*, *nupG*, *ppiA*, and *udp*) are up-regulated (Trietz *et al*., 2016). Likewise, nucleotide biosynthetic genes are down-regulated (Oh *et al*., 2002; Trietz *et al*., 2016). Thus during the growth lag in minimal acetate medium *E. coli* favors catabolism of nucleic acids over cell growth. This likely only occurs during the lag period, as the growth rate is relatively high after the lag.

Little work has been done to find gene targets to engineer an acetate tolerant industrial strain by investigating growth on acetate at neutral pH as the sole carbon source at high acetate concentrations. This study sought to identify potential targets for engineering an industrial strain with increased acetate consumption. The majority of the genomic library clones with improved growth on acetate generated in this study did not have a significantly higher growth rate during the exponential growth phase but rather a significantly decreased growth lag time. No clones were generated that contained genes directly involved in acetate metabolism. Two of the three genomic mutant variants generated with improved growth on acetate, in the *pck* promoter and in CRP, have direct roles in acetate assimilation. The role of the third variant, in the 5' UTR of of the poorly characterized cold shock gene *ynaE*, in increasing growth on acetate is unclear.

REFERENCES

Aono R, Tsukagoshi N, Yamamoto M. 1998. Involvement of outer membrane protein TolC, a possible member of the mar-sox regulon, in maintenance and improvement of organic solvent tolerance of *Escherichia coli* K-12. J Bacteriol. 180:938-44.

Axe DD, Bailey JE. 1995. Transport of lactate and acetate through the energized cytoplasmic membrane of *Escherichia coli*. Biotechnol Bioeng. 47:8-19.

Bury-Moné S, Nomane Y, Reymond N, Barbet R, Jacquet E, Imbeaud S, Jacq A, Bouloc P. 2009. Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. PLoS Genetics. 5:e1000651.

Chao YP, Patnaik RA, Roof WD, Young RF, Liao JC. 1993. Control of gluconeogenic growth by *pps* and *pck* in *Escherichia coli*. J Bacteriol. 175:6939-44.

Chen T, Wang J, Yang R, Li J, Lin M, Lin Z. 2011. Laboratory-evolved mutants of an exogenous global regulator, IrrE from *Deinococcus radiodurans*, enhance stress tolerances of *Escherichia coli*. PLoS One. 6:e16228.

Chin AM, Feldheim DA, Saier MH. 1989. Altered transcriptional patterns affecting several metabolic pathways in strains of *Salmonella typhimurium* which overexpress the fructose regulon. J Bacteriol. 171:2424-34.

Choi SH, Baumler DJ, Kaspar CW. 2000. Contribution of dps to acid stress tolerance and oxidative stress tolerance in *Escherichia coli* O157: H7. Appl Environ Microbiol. 66:3911-6.

Chong H, Yeow J, Wang I, Song H, Jiang R. 2013. Improving acetate tolerance of *Escherichia coli* by rewiring its global regulator cAMP receptor protein (CRP). PloS One. 8:e77422.

Dai Y, Yuan Z, Jack K, Keller J. 2007. Production of targeted poly (3-hydroxyalkanoates) copolymers by glycogen accumulating organisms using acetate as sole carbon source. J. Biotechnol. 129:489-97.

Death A, Ferenci T. 1994. Between feast and famine: endogenous inducer synthesis in the adaptation of *Escherichia coli* to growth with limiting carbohydrates. J Bacteriol. 176:5101-7.

Diez-Gonzalez F, Russell JB. 1997. The ability of Escherichia coli O157: H7 to decrease its intracellular pH and resist the toxicity of acetic acid. Microbiology. 143:1175-80.

Dukan S, Nyström T. 1999. Oxidative stress defense and deterioration of growth-arrested *Escherichia coli* cells. J Biol Chem. 274:26027-32.

Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biology. 5:e8.

Fernández-Sandoval MT, Huerta-Beristain G, Trujillo-Martinez B, Bustos P, González V, Bolivar F, Gosset G, Martinez A. 2012. Laboratory metabolic evolution improves acetate tolerance and growth on acetate of ethanologenic *Escherichia coli* under non-aerated conditions in glucose-mineral medium. Appl Microbiol Biotechnol. 96:1291-300.

Furuchi T, Kashiwagi K, Kobayashi H, Igarashi K. 1991. Characteristics of the gene for a spermidine and putrescine transport system that maps at 15 min on the *Escherichia coli* chromosome. J Biol Chem. 266:20928-33.

Gall S, Lynch MD, Sandoval NR, Gill RT. 2008. Parallel mapping of genotypes to phenotypes contributing to overall biological fitness. Metab Eng. 10:382-93.

Gimenez R, Nuñez MF, Badia J, Aguilar J, Baldoma L. 2003. The gene *yjcG*, cotranscribed with the gene *acs*, encodes an acetate permease in *Escherichia coli*. J Bacteriol. 185:6448-55.

Goldie AH, Sanwal BD. 1980. Genetic and physiological characterization of *Escherichia coli* mutants deficient in phosphoenolpyruvate carboxykinase activity. J Bacteriol. 141:1115-21.

Gosset G, Zhang Z, Nayyar S, Cuevas WA, Saier MH. 2004. Transcriptome analysis of Crp-dependent catabolite control of gene expression in *Escherichia coli*. J Bacteriol. 186:3516-24.

Gottesman S. 1995. Regulation of capsule synthesis: modification of the two-component paradigm by an accessory unstable regulator. In: Two-component signal transduction. American Society of Microbiology. 253-262

Gottesman SU, Trisler PA, Torres-Cabassa AN. 1985. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K-12: characterization of three regulatory genes. J Bacteriol. 162:1111-9.

Greenberg JT, Monach P, Chou JH, Josephy PD, Demple B. 1990. Positive control of a global antioxidant defense regulon activated by superoxide-generating agents in *Escherichia coli*. Proc Natl Acad Sci USA. 87:6181-5.

Guyer MS, Reed RR, Steitz JA, Low KB. 1981. Identification of a sex-factor-affinity site in E. coli as γδ. In: Cold Spring Harbor symposia on quantitative biology. Cold Spring Harbor Laboratory Press. 45:135-140.

Hammer-Jespersen KA. 1983. Nucleoside catabolism. In: Metabolism of nucleotides, nucleosides and nucleobases in microorganisms (ed.) Munch-Petersen A.

Han K, Hong J, Lim HC. 1993. Relieving effects of glycine and methionine from acetic acid inhibition in *Escherichia coli* fermentation. Biotechnol Bioeng. 41:316-24.

Hansen AM, Qiu Y, Yeh N, Blattner FR, Durfee T, Jin DJ. 2005. SspA is required for acid resistance in stationary phase by downregulation of H-NS in *Escherichia coli*. Mol Microbiol. 256:719-34.

Hua Q, Yang C, Oshima T, Mori H, Shimizu K. 2004. Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. Appl Environ Microbiol. 70:2354-66.

Joachimsthal E, Haggett KD, Jang JH, Rogers PL. 1998. A mutant of *Zymomonas mobilis* ZM4 capable of ethanol production from glucose in the presence of high acetate concentrations. Biotechnol Lett. 20:137-42.

Jolkver E, Emer D, Ballan S, Krämer R, Eikmanns BJ, Marin K. 2009. Identification and characterization of a bacterial transport system for the uptake of pyruvate, propionate, and acetate in *Corynebacterium glutamicum*. J Bacteriol. 191:940-8.

Kashiwagi K, Shibuya S, Tomitori H, Kuraishi A, Igarashi K. 1997. Excretion and uptake of putrescine by the PotE protein in *Escherichia coli*. J Biol Chem. 272:6318-23.

Korea CG, Badouraly R, Prevost MC, Ghigo JM, Beloin C. 2010. *Escherichia coli* K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. Environ microbiol. 12:1957-77.

Kumar A, Beloglazova N, Bundalovic-Torma C, Phanse S, Deineko V, Gagarinova A, Musso G, Vlasblom J, Lemak S, Hooshyar M, Minic Z. 2016. Conditional epistatic interaction maps reveal global functional rewiring of genome integrity pathways in *Escherichia coli*. Cell reports. 14:648-61.

Kurihara S, Suzuki H, Oshida M, Benno Y. 2011. A novel putrescine importer required for type 1 pili-driven surface motility induced by extracellular putrescine in *Escherichia coli* K-12. J Biol Chem. 286:10185-92.

Lee J, Jayaraman A, Wood TK. 2007. Indole is an inter-species biofilm signal mediated by SdiA. BMC Microbiol. 7:e42.

Lemos PC, Viana C, Salgueiro EN, Ramos AM, Crespo JP, Reiszcorr MA. 1998. Effect of carbon source on the formation of polyhydroxyalkanoates (PHA) by a phosphate-accumulating mixed culture. Enzyme Microb Technol. 22:662-71.

Lennen RM, Herrgård MJ. 2014. Combinatorial strategies for improving multiple-stress resistance in industrially relevant *Escherichia coli* strains. Appl Environ Microbiol. 80:6223-42.

Liu M, Durfee T, Cabrera JE, Zhao K, Jin DJ, Blattner FR. 2005. Global transcriptional programs reveal a carbon source foraging strategy by *Escherichia coli*. J Biol Chem. 2280:15921-7.

Luan G, Cai Z, Li Y, Ma Y. 2013. Genome replication engineering assisted continuous evolution (GREACE) to improve microbial tolerance for biofuels production. Biotechnol Biofuels. 6:137.

Luli GW, Strohl WR. 1990. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. Appl Environ Microbiol. 56:1004-11.

Markovitz AL. 1977. Genetics and regulation of bacterial ca2lar polysaccharide biosynthesis and radiation sensitivity. In: Surface carbohydrates of the prokaryotic cell. Academic Press. 415-62.

Martin RG, Rosner JL. 2002. Genomics of the *marA*/*soxS*/*rob* regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. Mol Microbiol. 44:1611-24.

McKay DB, Weber IT, Steitz TA. 1982. Structure of catabolite gene activator protein at 2.9-A resolution. Incorporation of amino acid sequence and interactions with cyclic AMP. J Biol Chem. 257:9518-24.

McLaggan D, Naprstek J, Buurman ET, Epstein W. 1994. Interdependence of K^+ and glutamate accumulation during osmotic adaptation of *Escherichia coli*. J Biol Chem. 269:1911-7.

Meibom KL, Søgaard-Andersen L, Mironov AS, Valentin-Hansen P. 1999. Dissection of a surface-exposed portion of the cAMP-CRP complex that mediates transcription activation and repression. Mol Microbiol. 32:497-504.

Miller JH. 1972. Experiments in molecular genetics. Cold Spring Laboratory Press.

Mordukhova EA, Pan JG. 2013. Evolved cobalamin-independent methionine synthase (MetE) improves the acetate and thermal tolerance of *Escherichia coli*. Appl Environ Microbiol. 79:7905-15.

Munro GF, Hercules K, Morgan J, Sauerbier W. 1972. Dependence of the putrescine content of *Escherichia coli* on the osmotic strength of the medium. J Biol Chem. 247:1272-80.

Nakano M, Ogasawara H, Shimada T, Yamamoto K, Ishihama A. 2014. Involvement of cAMP-CRP in transcription activation and repression of the *pck* gene encoding PEP carboxykinase, the key enzyme of gluconeogenesis. FEMS Microbiol Lett. 355:93-9.

Noh MH, Lim HG, Woo SH, Song J, Jung GY. 2018. Production of itaconic acid from acetate by engineering acid-tolerant *Escherichia coli* W. Biotechnol Bioeng. 3:729-38.

Oh MK, Rohlin L, Kao KC, Liao JC. 2002. Global expression profiling of acetate-grown *Escherichia coli*. J Biol Chem. 277:13175-83.

Pedersen H, Søgaard-Andersen L, Holst B, Valentin-Hansen P. 1991. Heterologous cooperativity in *Escherichia coli*. The CytR repressor both contacts DNA and the cAMP receptor protein when binding to the deoP2 promoter. J Biol Chem. 266:17804-8.

Pistocchi R, Kashiwagi K, Miyamoto S, Nukui E, Sadakata Y, Kobayashi H, Igarashi K. 1993. Characteristics of the operon for a putrescine transport system that maps at 19 minutes on the *Escherichia coli* chromosome. J Biol Chem. 268:146-52.

Polissi A, De Laurentis W, Zangrossi S, Briani F, Longhi V, Pesole G, Dehò G. 2003. Changes in *Escherichia coli* transcriptome during acclimatization at low temperature. Res Microbiol. 154:573-80.

Prigent-Combaret C, Prensier G, Le Thi TT, Vidal O, Lejeune P, Dorel C. 2000. Developmental pathway for biofilm formation in curli-producing *Escherichia coli* strains: role of flagella, curli and colanic acid. Environ Microbiol. 2:450-64.

Prost JF, Cozzone AJ. 1999. Detection of an extended -10 element in the promoter region of the *pckA* gene encoding phosphoenolpyruvate carboxykinase in *Escherichia coli*. Biochimie. 81:197-200.

Raghavan R, Groisman EA, Ochman H. 2011a. Genome-wide detection of novel regulatory RNAs in *E. coli*. Genome Res. 21:1487-97.

Raghavan R, Sage A, Ochman H. 2011b. Genome-wide identification of transcription start sites yields a novel thermosensing RNA and new cyclic AMP receptor protein-regulated genes in *Escherichia coli*. J Bacteriol. 193:2871-4.

Rajaraman E, Agarwal A, Crigler J, Seipelt-Thiemann R, Altman E, Eiteman MA. 2016. Transcriptional analysis and adaptive evolution of *Escherichia coli* strains growing on acetate. Appl Microbiol Biotechnol. 100:7777-85.

Ramseier TM, Bledig S, Michotey V, Feghali R, Saier MH. 1995. The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*. Mol Microbiol. 16:1157-69.

Rau MH, Calero P, Lennen RM, Long KS, Nielsen AT. 2016. Genome-wide *Escherichia coli* stress response and improved tolerance towards industrially relevant chemicals. Microb Cell Fact. 15:176.

Reeve CA, Bockman AT, Matin AB. 1984. Role of protein degradation in the survival of carbon-starved *Escherichia coli* and *Salmonella typhimurium*. J Bacteriol. 157:758-63.

Ren D, Bedzyk LA, Thomas SM, Ye RW, Wood TK. 2004. Gene expression in *Escherichia coli* biofilms. Appl Microbiol Biotechnol. 64:515-24.

Roe AJ, McLaggan D, Davidson I, O'Byrne C, Booth IR. 1998. Perturbation of anion balance during inhibition of growth of *Escherichia coli* by weak acids. J Bacteriol. 180:767-72.

Roe AJ, O'Byrne C, McLaggan D, Booth IR. 2002. Inhibition of *Escherichia coli* growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology. 148:2215-22.

Rooney JP, Patil A, Joseph F, Endres L, Begley U, Zappala MR, Cunningham RP, Begley TJ. 2011. Cross-species Functionome analysis identifies proteins associated with DNA repair, translation and aerobic respiration as conserved modulators of UV-toxicity. Genomics. 97:133-47.

Salmond CV, Kroll RG, Booth IR. 1984. The effect of food preservatives on pH homeostasis in *Escherichia coli*. Microbiology. 130:2845-50.

Sandoval NR, Mills TY, Zhang M, Gill RT. 2011. Elucidating acetate tolerance in *E. coli* using a genome-wide approach. Metab Eng. 13:214-24.

Schneider BL, Hernandez VJ, Reitzer L. 2013. Putrescine catabolism is a metabolic response to several stresses in *Escherichia coli*. Mol Microbiol. 88:537-50.

Schultz SC, Shields GC, Steitz TA. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science. 253:1001-7.

Serganov A, Nudler E. 2013. A decade of riboswitches. Cell. 152:17-24.

Singer M, Baker TA, Schnitzler G, Deischel SM, Goel M, Dove W, Jaacks KJ, Grossman AD, Erickson JW, Gross CA. 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. Microbiol Rev. 153:1-24.

Sledjeski DD, Gottesman S. 1996. Osmotic shock induction of capsule synthesis in *Escherichia coli* K-12. J Bacteriol. 178:1204-6.

Steiner P, Sauer U. 2003. Long-term continuous evolution of acetate resistant *Acetobacter aceti*. Biotechnol Bioeng. 84:40-4.

Terui Y, Saroj SD, Sakamoto A, Yoshida T, Higashi K, Kurihara S, Suzuki H, Toida T, Kashiwagi K, Igarashi K. 2014. Properties of putrescine uptake by PotFGHI and PuuP and their physiological significance in *Escherichia coli*. Amino acids. 46:661-70.

Treitz C, Enjalbert B, Portais JC, Letisse F, Tholey A. 2016. Differential quantitative proteome analysis of *Escherichia coli* grown on acetate versus glucose. Proteomics. 16:2742-6.

Treves DS, Manning S, Adams J. 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. Mol Biol Evol. 15:789-97.

Tweeddale H, Notley-McRobb L, Ferenci T. 1998. Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis. J Bacteriol. 180:5109-16.

Wang H, Wang F, Wang W, Yao X, Wei D, Cheng H, Deng Z. 2014. Improving the expression of recombinant proteins in *E. coli* BL21 (DE3) under acetate stress: an alkaline pH shift approach. PloS One. 9:e112777.

Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, Wood TK. 2010. Cryptic prophages help bacteria cope with adverse environments. Nat Commun. 1:147.

Wang X, Kim Y, Wood TK. 2009. Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. ISME J. 3:1164.

Waters LS, Storz G. 2009. Regulatory RNAs in bacteria. Cell. 136:615-28.

Xue W, Fan D, Shang L, Zhu C, Ma X, Zhu X, Yu Y. 2010. Effects of acetic acid and its assimilation in fed-batch cultures of recombinant *Escherichia coli* containing human-like collagen cDNA. J Biosci Bioeng. 109:257-61.

Yang H, Wolff E, Kim M, Diep A, Miller JH. 2004. Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. Mol Microbiol. 53:283-95.

Yang S, Pan C, Hurst GB, Dice L, Davison BH, Brown SD. 2014. Elucidation of *Zymomonas mobilis* physiology and stress responses by quantitative proteomics and transcriptomics. Front Microbiol. 5:246.

Yang S, Pelletier DA, Lu TY, Brown SD. 2010. The *Zymomonas mobilis* regulator *hfq* contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. BMC Microbiol. 10:135.

Zhang X, Zhang Y, Li Z, Xia Y, Ye Q. 2011. Continuous culture and proteomic analysis of *Escherichia coli* DH5α and its acetate-tolerant mutant DA19 under conditions of nitrogen source limitation. Bioprocess Biosyst Eng. 34:179-87.

CHAPTER 4: IMPROVING AN *ESCHERICHIA COLI* *PTSG MANZ GLK* STRAIN ENGINEERED TO BE GLUCOSE MINUS

Published in Microbiology:

Crigler J, Bannerman-Akwei L, Cole AE, Eiteman MA, Altman E. 2018 Glucose can be transported and utilized in *Escherichia coli* by an altered or overproduced N-acetylglucosamine phosphotransferase system (PTS). Microbiology. 164: 163-172.

4.1 Abstract

Biological detoxification of inhibitors in lignocellulosic hydrolysates for cellulosic ethanol production requires a detoxifying strain that does not consume, and thus waste, the sugars available for fermentation to ethanol. In the glucose$^-$ *Escherichia coli* Δ*glk* Δ*manZ* Δ*ptsG* strain, that lacks the glucose phosphotransferase system (PTS) and the mannose PTS as well as glucokinase, it was found that both fast- and slow- growing spontaneous glucose$^+$ revertants could be readily obtained. All of the fast-growing revertants either altered the N-acetylglucosamine PTS or caused its overproduction by inactivating the NagC repressor protein, which regulates the N-acetylglucosamine PTS, and these revertants could utilize either glucose or N-acetylglucosamine as a sole carbon source. When a Δ*nagE* deletion, which abolishes the N-acetylglucosamine PTS, was introduced into the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains, fast-growing revertants could no longer be isolated. Based on these results and other studies, it is clear that the N-acetylglucosamine PTS is the most easily adaptable PTS for transporting and phosphorylating glucose other than the glucose PTS and mannose PTS, which are the primary glucose transport systems. While the slow-growing glucose$^+$ revertants were not characterized, they were likely mutations that other researchers have observed before which affect other PTSs or sugar kinases.

4.2 Introduction

All bacteria require a carbon source for growth, and sugars are by far the most preferred carbon source. The vast majority of sugars are actively transported into the cell by three different transport systems, the phosphoenolpyruvate (PEP) phosphotransferase system (PTS), ATP-binding cassette (ABC) transporters or cation gradient-driven symporters (for a general review see Kaback, 2013). Sugars must be phosphorylated in order to be utilized by the Emden-Meyerhof-Parnas pathway. The PTSs phosphorylate sugars, while sugars that enter the cell by other transport systems must be phosphorylated by sugar kinases.

Glucose is generally the most preferred sugar source in bacteria. In *Escherichia coli*, where glucose transport and utilization has been the most extensively characterized, glucose can be phosphorylated for use by glycolysis by three different routes, the glucose PTS, the mannose PTS, and glucokinase (Curtis and Epstein, 1975). The glucose PTS and mannose PTS transport and phosphorylate glucose directly, while glucokinase phosphorylates glucose that has been internalized by other transport systems, such as the galactose or maltose ABC transport systems or the galactose permease transport system.

An *E. coli glk manZ ptsG* strain lacking a functional glucose PTS and mannose PTS as well as glucokinase, cannot utilize glucose as a carbon source (Curtis and Epstein, 1975). The doubling time of this strain on minimal glucose medium was greater than 2,000 minutes as compared to a wild-type strain that had a doubling time of 63 minutes. A

strain which lacked the glucose PTS and the mannose PTS, but contained glucokinase, had a doubling time of 295 minutes on minimal glucose medium, while a strain which contained the glucose PTS and the mannose PTS, but lacked glucokinase, had a doubling time of 75 minutes (Curtis and Epstein, 1975). Thus the two PTSs were far more important in the utilization of glucose than glucokinase. Furthermore, the glucose PTS was better at utilizing glucose than the mannose PTS. A strain which contained the glucose PTS, but lacked the mannose PTS and glucokinase, had a doubling time of 75 minutes on minimal glucose medium, while a strain which contained the mannose PTS, but lacked the glucose PTS and glucokinase, had a doubling time of 198 minutes on minimal glucose medium (Curtis and Epstein, 1975). Using glucose uptake studies, Hunter and Kornberg (1979), confirmed these observations and determined that the affinity for glucose by the glucose PTS was much higher than the affinity for glucose by the mannose PTS.

The PTS has been extensively characterized in *E. coli* (for general reviews see Postma *et al*., 1996; Tchieu *et al*., 2001; and Lengeler, 2015). All PTSs rely on the cytoplasmic enzyme I (EI) and the histidine phosphocarrier protein (HPr) which phosphorylates the sugar via sugar specific EII complexes that contain one to four domains (EIIA, EIIB, EIIC, EIID) on one to four proteins. The EIIC or EIICD domains are membrane bound and transfer the sugar into the cytoplasm where it is phosphorylated in a multi-stage process that involves the EIIA and EIIB domains. The phosphoryl group from phosphoenolpyruvate is transferred from EI to HPr to EIIA to EIIB, which then transfers the phosphoryl group to the sugar that has been transported by EIIC or EIICD. The

glucose PTS consists of the EIIA^{Glc} protein coded by *crr* and the EIICB^{Glc} protein coded by *ptsG*. The mannose PTS consists of the EIIAB^{Man} protein coded by *manX*, the EIIC^{Man} protein coded by *manY* and the EIID^{Man} protein coded by *manZ*.

Fast-growing glucose^{+} revertants were found to occur spontaneously at high levels when *E. coli* Δ*glk* Δ*manZ* Δ*ptsG* glucose^{-} strains that lack a functional glucose PTS, mannose PTS and glucokinase were plated on minimal glucose medium. Based on what is known about glucose utilization, the most obvious explanation for this observation was that the revertants contained mutations which altered PTSs for other sugars. Characterization of these mutants demonstrated that all of them altered the N-acetylglucosamine PTS, allowing it to utilize glucose in addition to N-acetylglucosamine. Unlike the glucose PTS and the mannose PTS, which are multi-protein systems, the N-acetylglucosamine PTS is composed of a single EIICBA^{Nag} protein.

The genes for the transport (*nagE*) and catabolism (*nagBA*) of N-acetylglucosamine are located in the divergent *nagE-nagBACD* regulon (Plumbridge, 1989). Two operators and two divergent promoters are located in between the *nagE* and *nagB* genes and control transcription of the regulon (Plumbridge and Kolb, 1993). The *nagA* gene encodes N-acetylglucosamine-6-phosphate deacetylase, the *nagB* gene encodes N-acetylglucosamine-6-phosphate deaminase, the *nagC* gene encodes NagC repressor, which regulates transcription of the regulon, the *nagD* or *umpH* gene encodes UMP phosphatase, which is not involved in N-acetylglucosamine metabolism, and the *nagE* gene encodes N-acetylglucosamine PTS (EIICBA^{Nag}).

4.3 Materials and Methods

4.3.1 Bacterial strains and growth conditions

Lysogeny broth (LB) or M9 (Miller, 1972) were used as rich or minimal defined media, respectively. The antibiotics kanamycin, streptomycin and tetracycline were used in LB at a final concentration of 40 μg/mL, 200 μg/mL and 20 μg/mL, respectively. Sugars were added to M9 minimal medium at a final concentration of 0.2% and the antibiotic tetracycline was used in M9 minimal medium at a final concentration of 10 μg/mL. The parental *E. coli* strains used in this study are listed in Table 4.1. P1 *vir* phage were used to prepare lysates from the donor strains for transductions. Most of the deletion mutants were generated by transducing the recipient strain with the corresponding Keio (FRT)kan deletion (Baba *et al*., 2006) and then curing the kanamycin resistance using the pCP20 plasmid, which contains a temperature-inducible FLP recombinase as well as a temperature-sensitive replicon (Datsenko and Wanner, 2000).

4.3.2 Hfr mapping studies

For the Hfr mapping studies, the MG1655 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants were first transduced to streptomycin resistance using a P1 *vir* lysate prepared from MC4100. The MG1655 Δ*glk* Δ*manZ* Δ*ptsG rpsL* and the MC4100 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants were then mated with one of the seven Hfr strains described in Singer *et al*. (1989) and plated on LB streptomycin tetracycline plates. 50 recombinant colonies from each mating were then patched onto M9 glucose plates that contained any amino acid or

Table 4.1. Parental bacterial strains used in this study

Strain	Genotype	Reference
CAG5051	PO1 *thi-1 relA1 spoT1 supQ80 nadA57*::Tn*10*	Singer *et al*., 1989
CAG5052	PO3 *metB1 relA1 btuB3191*::Tn*10*	Singer *et al*., 1989
CAG5053	PO43 *relA1 zbc-280*::Tn*10*	Singer *et al*., 1989
CAG5054	PO44 *relA1 thi-1 trpB83*::Tn*10*	Singer *et al*., 1989
CAG5055	PO45 *relA1 thi-1 zed-3069*::Tn*10*	Singer *et al*., 1989
CAG8160	PO68 *leu relA1 thi-39*::Tn*10*	Singer *et al*., 1989
CAG8209	PO13 *thi-1 leu-6 gal-6 lacY1* or *lacZ4 supE44 zgh-3075*::*Tn10*	Singer *et al*., 1989
CAG18433	*zbf-3057*::Tn*10*	Singer *et al*., 1989
JW0662-2	Δ*nagC725*::(FRT)kan Δ(*araD-araB*)567Δ*lacZ4787* (::*rrnB-3*) Δ(*rhaD-rhaB*)568 *hsdR514 rph-1* λ-	Baba *et al*., 2006
JW0665-1	Δ*nagE728*::(FRT)kan Δ(*araD-araB*)567Δ*lacZ4787* (::*rrnB-3*) Δ(*rhaD-rhaB*)568 *hsdR514 rph-1* λ -	Baba *et al*., 2006
JW1087-2	Δ*ptsG763*::(FRT)kan Δ(*araD-araB*)567 Δ*lacZ4787* (::*rrnB-3*) Δ(*rhaD-rhaB*)568 *hsdR514 rph-1* λ -	Baba *et al*., 2006
JW1586	Δ*mlc742*::(FRT)kan[1] Δ(*araD-araB*)567 Δ*lacZ4787* (::*rrnB-3*) Δ(*rhaD-rhaB*)568 *hsdR514 rph-1* λ -	Baba *et al*., 2006
JW1808-1	Δ*manZ743*::(FRT)kan Δ(*araD-araB*)567 Δ*lacZ4787* (::*rrnB-3*) Δ(*rhaD-rhaB*)568 *hsdR514 rph-1* λ -	Baba *et al*., 2006
JW2385-1	Δ*glk-726*::(FRT)kan Δ(*araD-araB*)567 Δ*lacZ4787* (::*rrnB-3*) Δ(*rhaD-rhaB*)568 *hsdR514 rph-1* λ -	Baba *et al*., 2006
MC4100	*araD139* Δ(*lac*)U169 *rpsL150 thi f1bB5301 deoC7 ptsF25*	Casadaban, 1976
MG1655	F- λ - (wild-type)	Guyer *et al*., 1981

vitamin supplements required by the Hfr strains. Glucose$^-$ recombinants indicated revertants that were converted from glucose$^+$ back to glucose$^-$ due to the transfer of DNA from the Hfr donor.

4.3.3 P1 transduction mapping studies

The MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants were definitively mapped using the set of Tn*10* insertions described in Singer *et al*. (1989). P1 lysates were prepared from Tn*10* insertions that mapped between 13.10 and 16.84 minutes on the *E. coli* chromosome and used to transduce the glucose$^+$ revertants to tetracycline resistant, and transductants were selected on LB tetracycline plates. 50 transductant colonies were then patched onto M9 glucose plates that contained any amino acid or vitamin supplements required due the Tn*10* insertions. Glucose$^-$ recombinants indicated revertants that were converted from glucose$^+$ back to glucose$^-$ due to the transfer of DNA from the donor strain.

To confirm that the glucose$^+$ revertants did not contain any extraneous mutations unlinked to the *nag* locus, a P1 lysate from *zbf-3057*::Tn*10*, which is 99% linked to the *nag* locus, was used to transduce each revertant to tetracycline resistance and transductants were selected on a M9 glucose tetracycline plate. A P1 lysate from each tetracycline resistant glucose$^+$ revertant was then used to transduce the *nag* locus into either MG1655 or MC4100 Δ*glk* Δ*manZ* Δ*ptsG* and transductants were selected on a LB tetracycline plate. 50 transductant colonies from each glucose$^+$ revertant were then patched onto M9 glucose plates. Glucose$^+$ recombinants indicated that MG1655 or

MC4100 Δ*glk* Δ*manZ* Δ*ptsG* was converted from glucose$^-$ to glucose$^+$ due to the transfer of the *nag* locus from the revertant.

4.3.4 Sequencing the *nagE* and *nagC* genes from the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants

The *nagE* and *nagC* genes were sequenced in each of the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants by PCR amplifying both the *nagE* and *nagC* genes in each of the revertants

using *Pfu* polymerase and then using overlapping primers to sequence both strands of the *nagE* and *nagC* genes.

4.3.5 Growth rate studies

LB overnights of the glucose$^+$ revertants and control strains were pelleted and washed twice in M9 medium without sugar and then diluted 1:200 in fresh M9 medium with the appropriate sugar added. OD_{550} readings were taken every hour until the OD_{550} reached 0.75. The growth rates (/h) were determined by calculating the slope of a plot of the growth time in hours versus the natural logarithm of the OD_{550} reading.

4.3.6 Generating the Δ*nagC*::Tet knockout

A Δ*nagC*::Tet knockout was generated using the lambda Red recombination system. Primers were designed which could amplify the *tetA* gene, which encodes the tetracycline resistance protein, along with its promoter from pWM41 (Metcalf *et al*., 1996) bracketed

by the first 50 bases of the *nagC* coding sequence and the TAA stop codon of the *nagC* gene followed by the subsequent 47 bases. The forward primer 5'-**ATG**ACACCAGGCGGACAAGCTCAGATAGGTAATGTTGATCTCGTAAAACAAC ATCTCAATGGCTAAGGCG-3' contained the first 50 bases of the *nagC* coding sequence followed by bases 1349 - 1368 of TRN10TETR (Accession Number J01830) from pWM41 while the reverse primer 5'-GGACTACCCAGAATATTGACAACAATAAGCGCCACTATAAAAGCACA**TTA**G GCTGGTTTATGCATATCGC-3' contained the TAA stop codon of the *nagC* gene and the subsequent 47 bases followed by bases 3020 - 3039 of TRN10TETR from pWM41. The bases from pWM41 are underlined and the ATG and TAA stop codons of the nagC coding sequence are indicated in bold. The two primers were used to amplify a 1,791 bp fragment from pWM41 DNA using the polymerase chain reaction (PCR) with KAPA HiFi DNA polymerase. The resulting DNA was gel-isolated and electroporated into DY330 electrocompetent cells which were prepared as described by Yu *et al*. (2000). Tetracycline-resistant colonies were selected and the presence of the Δ*nagC*::Tet knockout was confirmed by DNA sequencing. The Δ*nagC*::Tet knockout from DY330 was then moved into MG1655 by P1 transduction to create ALS2817.

4.3.7 RT-qPCR analysis of *nagE* mRNA levels

The quantification of *nagE* mRNA levels in the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants was determined using RT-qPCR with the moderately expressed *trmA* gene serving as a control. Strains were either grown in LB or M9 glucose to mid-log phase (0.50 - 0.55

OD_{550}), total RNA was isolated from 0.25 OD_{550} cell equivalents, quantified and then complementary DNA (cDNA) was prepared using DNA primers designed to initiate replication as close to the end of the coding sequence as possible. The deviance of *trmA* cDNA levels between the different LB or M9 glucose samples was less than 5%. Amplicons were designed to be as close to 100 bp as possible and forward primers were designed to initiate replication as close to the start codon of the coding sequence as possible. For *nagE* the 5'-CACAATATGGCCCTGAGCTT-3' primer was used for cDNA synthesis, while the forward 5'-AGGTTTTTTCCAGCGACTCG-3' and reverse 5'-CTGGGCAATAAACGCAACGT-3' DNA primers were used to generate a 118 bp amplicon. For *trmA* the 5'-ACTTCGCGGTCAGTAATACG-3' DNA primer was used for cDNA synthesis, while the forward 5'-GCAAAGTATGATGGCACCGT-3' and reverse 5'-GTGATACAGGTCATCGCCAT-3' DNA primers were used to generate a 118 bp amplicon.

4.3.8 Calculating protein homologies

Protein homologies were determined using the blastp program on the National Institutes of Health (NIH) National Center for Biotechnology Information (NCBI) website.

4.4 Results

4.4.1 Isolation, mapping, and sequencing Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants

As shown in Figure 4.1, *E. coli* Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains yielded a large number of glucose$^+$ revertants when plated on minimal glucose medium. This phenomenon occurred in either the wild-type MG1655 *E. coli* strain (Guyer *et al*., 1981) or the widely-utilized faster growing MC4100 *E. coli* strain (Casadaban, 1976). In both *E. coli* Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains fast-growing revertants appeared in four to five days, while slow-growing revertants appeared in seven to nine days. The revertant rates of the two strains were similar, 2.55 x 10^{-7} fast-growing revertants per cell for MG1655 Δ*glk* Δ*manZ* Δ*ptsG* and 3.75 x 10^{-7} fast-growing revertants per cell for MC4100 Δ*glk* Δ*manZ* Δ*ptsG*. When re-streaked on plates, the revertants were able to grow on minimal glucose medium in two to three days, although they did grow slower than the wild-type strain (Figure 4.2).

15 independent fast-growing revertants from the MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains were isolated for further study. Using the Hfr Tn*10* mapping strains described in Singer *et al*. (1989), all 15 revertants were found mapped to between 13.10 and 16.84 minutes on the *E. coli* chromosome. Using P1 transductions and the set of Tn*10* insertions described in Singer *et al*. (1989) that were in between 13.10 and 16.84 minutes, it was shown that all 15 revertants were 99% linked to *zbf-3057*::Tn*10*, which has subsequently been determined to be an *asnB* insertion at 15.02 minutes (Nichols *et*

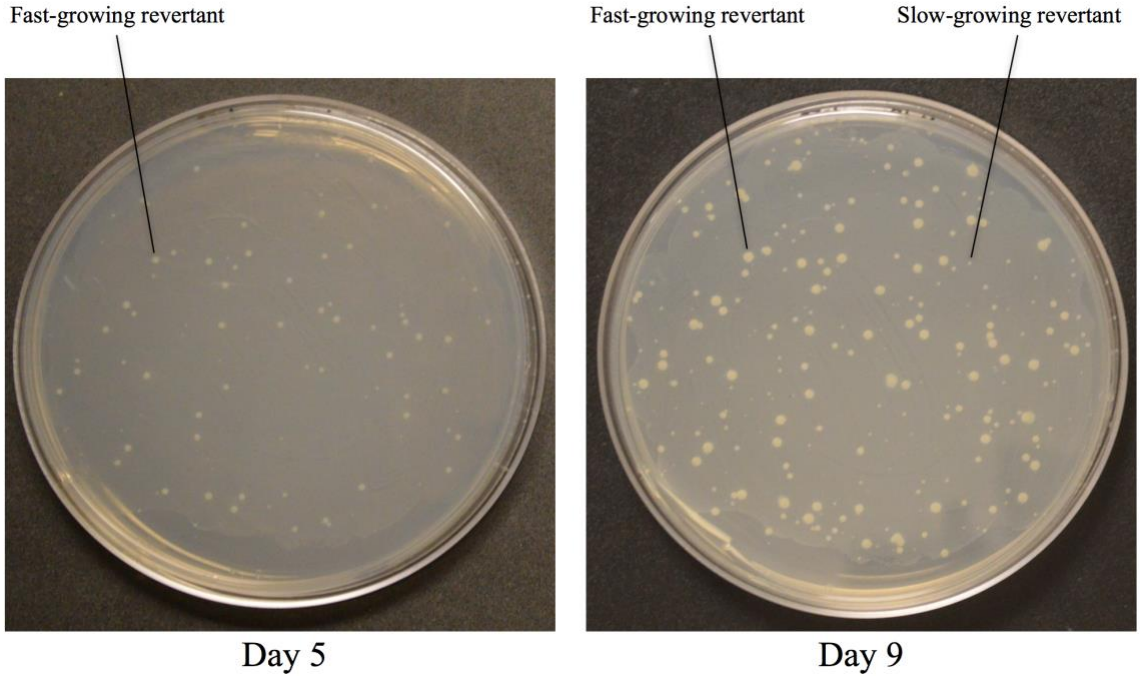

Figure 4.1. Isolation of glucose$^+$ revertants from a MG1655 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strain

For demonstration purposes 2 x 10^8 cells from a LB overnight were pelleted, washed twice in M9 glucose medium, plated on a M9 glucose plate and incubated at 37ºC. A. Picture taken at five days showing the accumulation of fast-growing revertants. B. Picture taken at nine days showing the accumulation of slow-growing revertants.

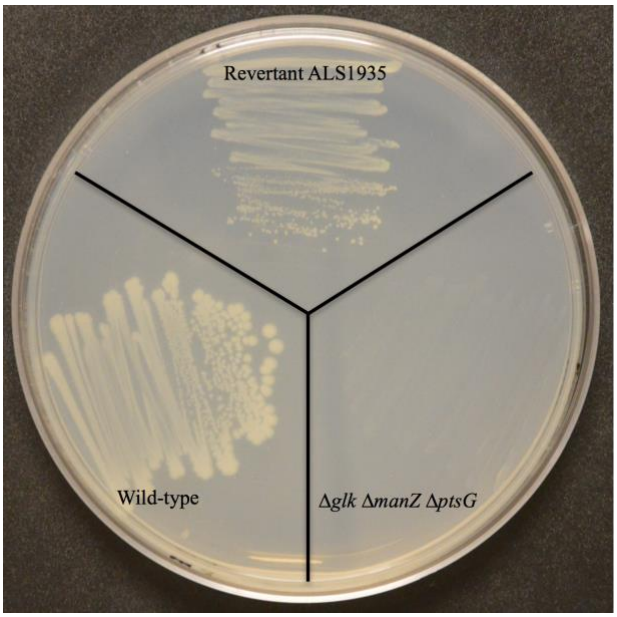

Figure 4.2. Growth of a MG1655 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertant versus the MG1655 Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ parent or the wild-type MG1655 strain on a minimal glucose plate

The strains were streaked for single colonies on a M9 glucose plate and incubated at 37ºC for three days.

al., 1998). To confirm the revertants did not contain any extraneous mutations unlinked to the *nag* locus, the *zbf-3057*::Tn*10* was transduced into each revertant and then used to transduce the *nag* locus of each revertant into MG1655 or MC4100 Δ*glk* Δ*manZ* Δ*ptsG* and recreate the glucose$^+$ phenotype. How the Hfr mapping and P1 transduction studies were conducted are described in detail in the *Methods* section.

Because the *nagE* gene, which encodes the N-acetylglucosamine PTS (EIICBA^{Nag}), and its negative regulator *nagC*, which encodes the NagC repressor, are located at 15.16 and 15.08 minutes, respectively, and thus were the most likely candidates for the mutation that occurred in the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants, both the *nagE* and *nagC* genes as well as their upstream promoter regions in each of the 15 revertants were sequenced. Each of the 15 revertants contained a single mutation in either the *nagE* or *nagC* gene. Table 4.2 lists the mutations that occurred in each of the revertants as well as the associated change in the coded protein. Only one revertant contained a mutation in the *nagE* gene, while the rest of the revertants contained a mutation in the *nagC* gene.

4.4.2 Growth rates of the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants in minimal glucose and minimal N-acetylglucosamine media

To determine how quickly the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants could grow in minimal glucose medium and whether the revertants affected the utilization of N-acetylglucosamine, the usual substrate of the N-acetylglucosamine PTS, the growth rates of the revertants were determined in both minimal glucose and minimal N-

Table 4.2. Mutations in the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants

Strain	Mutation in *nagE*	Protein change
ALS1945	788 (T to A)	M263K
	Mutation in *nagC*	
ALS1935	937 (G to T)	Stop at 313
ALS1936	788 (G to A)	G263D
ALS1937	12 bp deletion from 903 - 914	In-frame deletion from 301-304
ALS1938, 1943	794 (T to G)	L265R
ALS1939	1329 bp IS2A insertion at 804	Frameshift at 269; Stop at 327
ALS1940	769 bp IS1A insertion at 987	Frameshift at 330; Stop at 336
ALS1941	13 bp deletion from 204 - 216	Frameshift at 70; Stop at 126
ALS1946, 1948 1949, 1950	AACCGTCGC insertion at 804	NRR in-frame insertion at 268
ALS1951	226 (G to A)	A76T
ALS1952	769 bp IS1H insertion at 1080	Frameshift at 360; Stop at 367

Revertant ALS1943 was identical to ALS1938 and revertants ALS1948, ALS1949 and ALS1950 were identical to ALS1946 and thus not further characterized. The *nagE* coding sequence is 1,944 bp in length and codes for a 648 aa protein. The *nagC* coding sequence is 1,218 bp in length and codes for a 406 aa protein.

Table 4.3. Growth rates in minimal glucose or N-acetylglucosamine media

MG1655 strains	**Growth rate (/h) in M9-Glucose**	**Growth rate (/h) in M9-N-acetylglucosamine**
MG1655	0.446	0.468
ALS1935 (*nagC*)	0.110	0.483
ALS1936 (*nagC*)	0.108	0.497
ALS1937 (*nagC*)	0.106	0.474
ALS1938 (*nagC*)	0.114	0.508
ALS1939 (*nagC*)	0.123	0.483
ALS1940 (*nagC*)	0.125	0.540
ALS1941 (*nagC*)	0.116	0.479
Δ*glk* Δ*manZ* Δ*ptsG*	0.002	0.476
Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*	0.130	0.502
MC4100 strains		
MC4100	0.702	0.721
ALS1945 (*nagE*)	0.307	0.673
ALS1946 (*nagC*)	0.207	0.693
ALS1951 (*nagC*)	0.192	0.749
ALS1952 (*nagC*)	0.190	0.680
Δ*glk* Δ*manZ* Δ*ptsG*	0.008	0.701
Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*	0.221	0.745

All growth rate studies were conducted in triplicate and the standard deviation was less than 10%.

acetylglucosamine media. The revertants restored the growth of a Δ*glk* Δ*manZ* Δ*ptsG* parent to an appreciable level in minimal glucose medium, although they did not grow at the wild-type growth rate (Table 4.3).

All of the revertants were able to grow in minimal N-acetylglucosamine medium at rates equal to both the parent Δ*glk* Δ*manZ* Δ*ptsG* and wild-type strains; thus the revertants did not impact the ability of the N-acetylglucosamine PTS to utilize its preferred sugar. Contrary to Curtis and Epstein (1975), who reported a strain containing knockouts in *glk*, *manZ*, and *ptsG* could not grow on glucose as a sole carbon source, it was found that the MG1655 Δ*glk* Δ*manZ* Δ*ptsG* strain had a growth rate of 0.002/h in minimal glucose medium, while the faster growing MC4100 Δ*glk* Δ*manZ* Δ*ptsG* strain had a growth rate of 0.008/h in minimal glucose medium. Thus Δ*glk* Δ*manZ* Δ*ptsG* strains did grow, albeit very slowly, in minimal glucose medium. This observation could explain the high number of spontaneous glucose$^+$ revertants that can be obtained from either the MG1655 or MC4100 Δ*glk* Δ*manZ* Δ*ptsG* strains.

Because a significant number of the glucose$^+$ revertants were predicted to code for a truncated NagC repressor protein (Table 4.2), MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC* deletion strains were constructed using the Keio collection (Baba *et al*., 2006). The MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC* deletion strains were able to grow on minimal glucose medium although on average their growth rates were 25% lower than the *nagC* glucose$^+$ revertants, at 0.082 and 0.160/h for the strain in MG1655 and

MC4100, respectively. Given this unexpected result with the Δ*nagC* Keio knockout, a new Δ*nagC*::Tet knockout was constructed as described in *Methods* and moved it into the MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* deletion strains. Consistent with expectations, the MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*::Tet deletion strains were able to grow on minimal glucose medium with growth rates very similar to the *nagC* glucose$^+$ revertants (Table 4.3). Since the Mlc repressor protein, which regulates the glucose PTS component EIICB^{Glc} coded by *ptsG*, the mannose PTS component EIIABCD^{Man} coded by *manXYZ* as well as HPr and E1 coded by *ptsHI*, is 40% identical to NagC and has been predicted to recognize similar binding sites (Pennetier *et al*., 2008), MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* Δ*mlc* deletion strains were also constructed. The Δ*glk* Δ*manZ* Δ*ptsG* Δ*mlc* deletion strains were not able to grow in minimal glucose medium (data not shown).

4.4.3 Expression levels of *nagE* in the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants

Since most of the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants contained mutations in the NagC repressor protein, and based on the growth rate data in Table 4.3 would be predicted to have increased transcription of *nagE*, which codes for the EIICBA^{Nag} protein, quantitative real time polymerase chain reaction (qPCR) was performed to determine the expression levels of *nagE* mRNA. As seen in Table 4.4, all of the glucose$^+$ revertants that contained mutations in *nagC* as well as the Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC* deletion strains exhibited increased expression levels of *nagE* mRNA when the cells were grown in LB.

Table 4.4. Expression levels of *nagE* in the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants

MG1655 strains	Fold change of *nagE* expression in LB[1]	Fold change of *nagE* in M9-glucose[1]
ALS1935 (*nagC*)	40.8	149.1
ALS1936 (*nagC*)	13.4	106.2
ALS1937 (*nagC*)	56.9	237.2
ALS1938 (*nagC*)	42.8	118.6
ALS1939 (*nagC*)	33.4	238.9
ALS1940 (*nagC*)	11.6	105.4
ALS1941 (*nagC*)	39.9	151.2
Δ*glk* Δ*manZ* Δ*ptsG*	0.4	ND[2]
Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*	42.8	208.1
MC4100 strains		
ALS1945 (*nagE*)	0.7	13.9
ALS1946 (*nagC*)	9.1	212.4
ALS1951 (*nagC*)	19.6	310.3
ALS1952 (*nagC*)	17.4	277.7
Δ*glk* Δ*manZ* Δ*ptsG*	0.8	ND[†]
Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*	16.4	294.1

[1] The expression levels of *nagE* mRNA in the revertants or other strains were normalized to the expression levels of *nagE* mRNA in MG1655 or MC4100, which was set at 1.0.
[2] The expression levels of *nagE* mRNA were not determined in M9 glucose for the Δ*glk* Δ*manZ* Δ*ptsG* strains because of their very slow growth rate in M9 glucose.

As expected, the one revertant in *nagE* had no significant difference in *nagE* expression compared to the parent strain. Consistent with the results of the growth rate studies the Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC* deletion strains expressed lower levels of *nagE* mRNA compared to the Δ*glk* Δ*manZ* Δ*ptsG nag*C glucose$^+$ revertants, while the Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*::Tet deletion strains expressed similar levels of *nagE* mRNA compared to the Δ*glk* Δ*manZ* Δ*ptsG nagC* glucose$^+$ revertants.

Because the ability of the ALS1945 Δ*glk* Δ*manZ* Δ*ptsG nagE* glucose$^+$ revertant strain to grow in minimal glucose medium was paradoxical, given that it still had a functional NagC repressor, expression levels of *nagE* mRNA in all of the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants grown in minimal glucose medium were determined. As seen in Table 4.4, the expression level of *nagE* mRNA was significantly elevated when the ALS1945 Δ*glk* Δ*manZ* Δ*ptsG nagE* glucose$^+$ revertant was grown in minimal glucose medium versus the MC4100 control and thus induction of *nagE* expression occurs when the ALS1945 Δ*glk* Δ*manZ* Δ*ptsG nagE* glucose$^+$ revertant is grown in minimal glucose medium. Table 4.4 also shows that the expression levels of *nagE* increased in all of the Δ*glk* Δ*manZ* Δ*ptsG* glucose$^+$ revertants when they were grown in minimal glucose medium versus LB and that the differences in *nagE* expression levels amongst the different glucose$^+$ revertants was consistent whether they were grown in LB or minimal glucose medium.

4.4.4 The deletion of *nagE* from Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains prevents the isolation of fast-growing glucose$^+$ revertants

The results suggested that only mutations which altered the structure or expression of NagE would allow Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains to yield fast-growing glucose$^+$ revertants. To find out whether this was the case, the frequency at which glucose$^+$ revertants could be isolated in both Δ*glk* Δ*manZ* Δ*ptsG* and Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagE* deletion strains were determined. Table 4.5 shows the results of this study. In both the MG1655 and MC4100 Δ*glk* Δ*manZ* Δ*ptsG* deletion strains, the additional deletion of *nagE* prevented the isolation of fast-growing glucose$^+$ revertants that occurred at four to five days. The slower growing glucose$^+$ revertants that occurred at seven to nine days were not prevented by the deletion of *nagE*.

4.5 Discussion

Both fast-growing and slow-growing glucose$^+$ revertants were obtained from *E. coli* Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains, which lacked a functional glucose PTS and mannose PTS as well as glucokinase. All of the fast-growing revertants contained mutations that affected the N-acetylglucosamine PTS. In the 15 revertants that were characterized, one contained a mutated *nagE* gene, while the rest contained mutated *nagC* genes. The *nagE* gene encodes the EIICBA^{Nag} N-acetylglucosamine PTS, while the *nagC* gene encodes the

Table 4.5. Revertant colony counts of Δ*glk* Δ*manZ* Δ*ptsG* glucose$^-$ strains with or without the additional deletion of *nagE* on minimal glucose plates

MG1655 strains	**4 days (fast-growing revertants)**	**9 days (slow-growing revertants)**
Δ*glk* Δ*manZ* Δ*ptsG*	51	84
Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagE*	0	199
MC4100 strains	**5 days (fast-growing revertants)**	**7 days (slow-growing revertants)**
Δ*glk* Δ*manZ* Δ*ptsG*	75	100
Δ*glk* Δ*manZ* Δ*ptsG* Δ*nagC*	0	109

2×10^8 cells from LB overnights were pelleted, washed twice in M9 glucose medium, plated on M9 glucose plates and incubated at 37ºC.

NagC repressor protein which regulates the expression of *nagE*. The *nagE* mutant coded for an altered EIICBA^{Nag} protein that could transport both glucose and N-acetylglucosamine, and its expression was elevated when the mutant was grown in minimal glucose medium. All of the *nagC* mutants caused increased expression of *nagE* mRNA and presumably greater amounts of the IICBA^{Nag} protein, which facilitated the transport and utilization of glucose.

While most of the sugar PTSs in *E. coli* contain multiple proteins that compose the EIIC, EIIB and EIIA domains, the N-acetylglucosamine PTS consists of a single protein that contains the EIIC, EIIB and EIIA domains. Interestingly, *nagE* has been shown to be the closest related paralog to the glucose PTS genes, *ptsG*, which codes for EIICB^{Glc} and *crr* which codes for EIIA^{Glc} in *E. coli* (Tchieu *et al*., 2001). Thus it is not surprising that a modified or overproduced N-acetylglucosamine PTS can substitute for the glucose PTS.

Figure 4.3 shows the locations of the EIIC, EIIB and EIIA domains in the EIICBA^{Nag} protein, based on the highly homologous EIICB^{Glc} and EIIA^{Glc} proteins, and depicts the location of the NagE M263K mutant that was isolated in this study. Several analogous studies have been conducted by other researchers in which mutations scattered throughout EIICB^{Glc} have enabled the glucose PTS to utilize other sugars as carbon sources. A G320V mutant enabled the glucose PTS to utilize mannitol (Begley *et al*., 1996), while F37Y, G176D, G281D, I283T and L289Q mutants enabled the glucose PTS to utilize ribose (Oh *et al*., 1999). Similarly, V12F, V12G, G13C, G176D, A288V, G320S and P384R mutants (Notley-McRobb and Ferenci, 2000) and S169F, S169P,

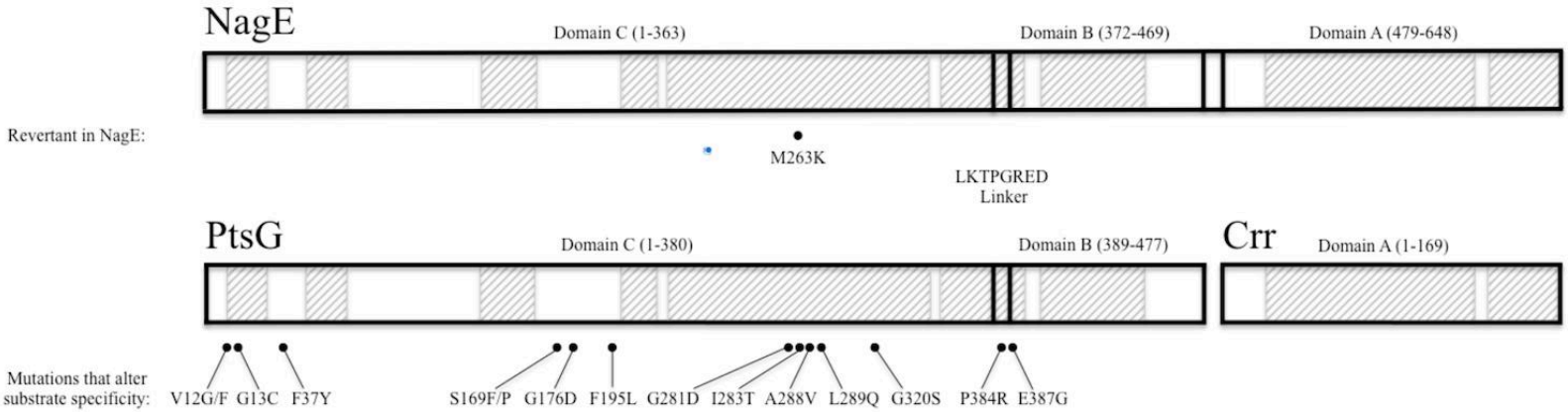

Figure 4.3. Comparison of EIICBA^{Nag} (*nagE*) with EIICB^{Glc} (*ptsG*) and EIIA^{Glc} (*crr*) indicating mutants that affect substrate specificity

EIICB^{Nag} and EIICB^{Glc} are 42% identical and 65% similar, while EIIA^{Nag} and EIIA^{Glc} are 47% identical and 66% similar and the LKTPGRED linker that separates the C and B domains are identical in EIICB^{Nag} and EIICB^{Glc}. The *nagE*, *ptsG* and *crr* sequences were obtained from the *E. coli* genome sequence (Blattner *et al*., 1997) (GenBank accession number u00096.3) and the C and B domains, which differ slightly in length, were scaled proportionally. The regions that are highly homologous are indicated by shaded hatched boxes.

F195L and E387G mutants (Zeppenfeld *et al*., 2000) enabled the glucose PTS to utilize mannose and glucosamine. With the exception of the P384R and the E387G mutants that map to the highly conserved LKTPGRED linker, which links the EIIC and EIIB domains, all of the mutants were located in the EIIC domain. The crystal structure of the EIIC domain of the diacetylchitobiose PTS from *Bacillus cereus* has been determined (Cao *et al*., 2011) and compared to the $EIIC^{Glc}$ domain (McCoy *et al*., 2015). This analysis predicts that there are 10 transmembrane spanning segments in $EIIC^{Glc}$ with a large periplasmic loop from amino acid 102 through 115 between transmembrane spanning segments three and four and a large cytoplasmic loop from amino acid 134 through 152 between transmembrane spanning segments four and five. Most of the PtsG mutants shown in Figure 4.3 are located in amino acids that are in transmembrane spanning segments of the $EIIC^{Glc}$ domain. The M263K mutant in $EIICBA^{Nag}$ that allows the N-acetylglucosamine PTS to utilize both glucose and N-acetylglucosamine would be located at amino acid 283 of $EIICB^{Glc}$ based on homology. Interestingly, several of the mutants that change the specificity of $EIICB^{Glc}$, namely, G281D, I283T, A288V and L289Q, are also located in this region.

The ability of the NagE M263K mutant to grow on minimal glucose medium in the presence of functional NagC repressor is somewhat surprising; however, the qPCR experiments clearly demonstrate that *nagE* mRNA expression increased in the NagE M263K mutant when it was grown in minimal glucose medium. Mutants similar to the NagE M263K mutant isolated in this study have been isolated in *Cupriavidus necator*, formerly called *Ralstonia eutropha*, by other researchers. *C. necator* normally can only

utilize the sugars fructose and N-acetylglucosamine as a carbon source, but mutants can be isolated that are also able to utilize glucose as a carbon source. Raberg *et al*. (2011) isolated a *C. necator* glucose$^+$ mutant with two mutations, an A153T mutation in NagE, which contains the EIICB^{Nag} domain, and a nonsense mutation in the NagR repressor. Orita *et al*. (2012) isolated a *C. necator* glucose$^+$ mutant that also contained two mutations, a G265R mutation in NagE and a nonsense mutation in NagR. The *C. necator* EIICB^{Nag} domain is highly homologous to both the *E. coli* EIICB^{Nag} and EIICB^{Glc} domains and the *C. necator* A153 and G265 amino acids in NagE correspond to amino acids 151 and 262 in *E. coli* NagE and amino acids 171 and 282 in *E. coli* PtsG, respectively, based on homology. Interestingly both of the glucose$^+$ mutants isolated in *C. necator* required inactivation of the repressor, while the *E.coli* NagE M263K mutant did not.

Figure 4.4 shows the three domains of the NagC repressor protein, a helix-turn-helix DNA binding domain formed by amino acids 1 - 81 and 395 - 406, a smaller α/β domain formed by amino acids 82 - 194 and 381 - 394 and a larger α/β domain formed by amino acids 195 - 380. The key amino acids of the potential N-acetylglucosamine-6-phosphate binding site known to prevent the binding of NagC to the *nagE* operator are also shown in Figure 4.4 along with the *nagC* mutants obtained in this study which cause increased expression of *nagE* and allow glucose$^-$ strains that lack the glucose PTS, mannose PTS and glucokinase to grow minimal glucose medium. 10 out of the 14 *nagC* mutants that were characterized in this study either caused a single amino acid change, a small four amino acid deletion or a small three amino acid insertion in the NagC protein, and seven

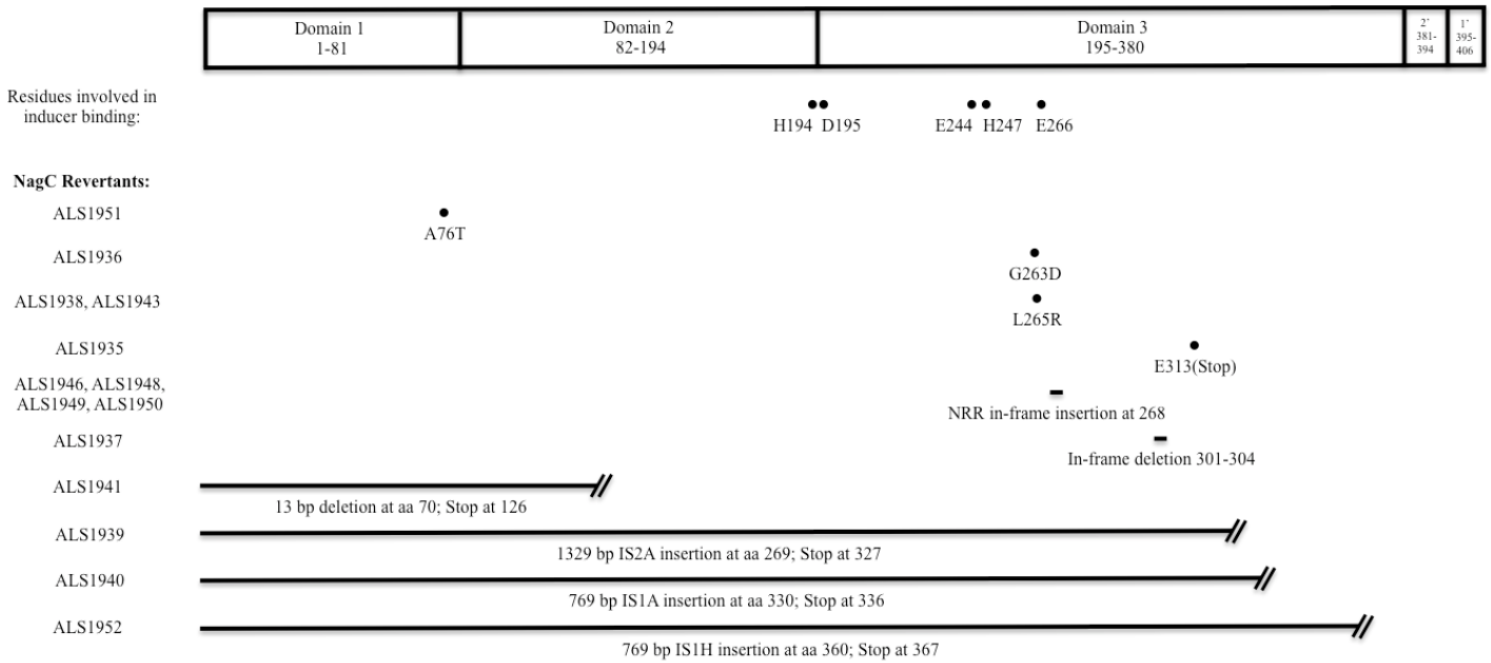

Figure 4.4. Diagram of the three domains of NagC indicating the location of the N-acetylglucosamine-6-phosphate binding site and the *nagC* mutants obtained in this study

The C-terminus of NagC forms two α-helices, 2' and 1', which fold back onto domains 2 and 1, respectively. The NagC structure is based on the Mlc structure (Schiefner *et al*., 2005). NagC is 40% identical and 62% similar to Mlc. The amino acids of the potential N-acetylglucosamine-6-phosphate binding site in NagC, H194, D195, E244, H247 and E266, are identical to the amino acids of the glucose binding site in Mlc. The *nagC* and *mlc* sequences were obtained from the *E. coli* genome sequence (Blattner *et al.*, 1997) (GenBank accession number u00096.3).

of these mutants most likely affect the N-acetylglucosamine-6-phosphate binding site. All of the *nagC* mutants caused a significant increase in *nagE* mRNA levels and thus appear to cause overproduction of the N-acetylglucosamine PTS by deactivating the NagC repressor.

Several studies have shown that sugar PTSs other than the glucose PTS and the mannose PTS, the primary transporters of glucose, can be overproduced to facilitate the transport and utilization of glucose by *E. coli*. Overproduction of the β-glucoside PTS $EIIBCA^{Bgl}$ via a multicopy plasmid allowed an *E. coli* strain that lacked a functional glucose PTS, mannose PTS and glucokinase to grow on minimal glucose medium (Schnetz *et al*., 1990). Overproduction of the maltose PTS domain $EIICB^{Mal}$ via a multicopy plasmid allowed an *E. coli* strain that lacked a functional glucose PTS, mannose PTS and glucokinase to grow on minimal glucose medium (Reidl and Boos, 1991). However, overproduction of $EIICB^{Mal}$ via the deletion of the MalI repressor protein did not. The strain used in the maltose PTS studies (Reidl and Boos, 1991) contained *crr*, which serves as the EIIA domain for both the glucose PTS and maltose PTS. Hummel *et al*. (1992) showed that a glucose/N-acetylglucosamine PTS hybrid on a multicopy plasmid containing the $EIIC^{Glc}$ domain fused to the $EIIBA^{Nag}$ domains could phosphorylate glucose but not N-acetylglucosamine, which suggested that the $EIIB^{Nag}$ domain could replace $EIIB^{Glc}$.

Our two findings: (1) that all of the mutations in glucose^{+} revertants isolated from glucose^{-} strains, which lacked the glucose PTS, mannose PTS and glucokinase, affected

the N-acetylglucosamine PTS and (2) that fast-growing glucose$^+$ revertants could not be isolated from glucose$^-$ strains, which lacked the glucose PTS, the mannose PTS, glucokinase as well as the N-acetylglucosamine PTS, strongly suggests that augmentation of the N-acetylglucosamine PTS is the most effective way to restore the growth of glucose$^-$ strains, which lacked the glucose PTS, mannose PTS and glucokinase, on minimal glucose medium. While the maltose PTS and the β-glucoside PTS are also able to facilitate the transport and utilization of glucose to some degree, our results show that the N-acetylglucosamine PTS is clearly more effective. Table 4.6 shows the identity and similarity scores for EIICB^{Nag}, EIICB^{Mal} and EIIBC^{Bgl} versus EIICB^{Glc}. EIICB^{Nag} is significantly more homologous to EIICB^{Glc} than EIICB^{Mal} or EIIBC^{Bgl} and supports our observations that the N-acetylglucosamine PTS is the most likely PTS to be augmented to allow the utilization of glucose when the glucose PTS and the mannose PTS, the primary transport systems for glucose, are not available.

Although the slow-growing glucose$^+$ revertants were not characterized, there are several types of mutations that could cause them. They could be mutants that affect the maltose PTS and β-glucoside PTS, since either of these PTSs have the potential to transport and utilize glucose (Schnetz *et al*., 1990; Reidl and Boos, 1991). They could also be mutants that affect sugar kinases, which can substitute for glucokinase to phosphorylate glucose that has been transported by the galactose or maltose ABC transport systems, or the galactose permease transport system (Steinsiek and Bettenbrock, 2012). Miller and Raines (2004; 2005) have demonstrated that overproduction of the AlsK, NanF, YajF or

Table 4.6. Comparison of the homologies between EIICB^{Nag}, EIICB^{Mal} and EIIBC^{Bgl} to EIICB^{Glc}

MG1655 strains	% Identity	% Similarity
EIICB^{Nag} (*nagE*)	42.0	64.5
EIICB^{Mal} (*malX*)	38.0	54.0
EIIBC^{Bgl} (*bglF*)	23.0	33.5

The *nagE*, *malX*, *bglF*, and *ptsG* sequences were obtained from the *E. coli* genome sequence (Blattner *et al*., 1997) (GenBank accession number U00096.3).

YcfX sugar kinases via multicopy plasmids allow an *E. coli* strain, which lacks a functional PTS due to the deletion of EI, HPr and glucokinase, to grow on minimal glucose medium.

REFERENCES

Begley GS, Warner KA, Arents JC, Postma PW, Jacobson GR. 1996. Isolation and characterization of a mutation that alters the substrate specificity of the *Escherichia coli* glucose permease. J Bacteriol. 178:940-42.

Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J. 1997. The complete genome sequence of *Escherichia coli* K-12. Science. 277:1453-62.

Cao Y, Jin X, Levin EJ, Huang H, Zong Y, Quick M, Weng J, Pan Y, Love J, Punta M, Rost B. 2011. Crystal structure of a phosphorylation-coupled saccharide transporter. Nature. 473:50-54.

Casadaban MJ. 1976. Transposition and fusion of the lac genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J Mol Biol. 104:541-55.

Curtis SJ, Epstein W. 1975. Phosphorylation of D-glucose in *Escherichia coli* mutants defective in glucosephosphotransferase, mannosephosphotransferase, and glucokinase. J Bacteriol. 122:1189-99.

Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA. 97:6640-45.

Guan L, Kaback HR. 2013. Glucose/sugar transport in bacteria. In: Lennarz WJ, Lane MD (editors). Encyclopedia of Biological Chemistry, 2nd ed. Oxford: Elsevier. 387-90.

Guyer MS, Reed RR, Steitz JA, Low KB. 1981. Identification of a sex-factor-affinity site in *E. coli* as γδ. Cold Spring Harb Symp Quant Biol. 45:135-40.

Hummel U, Nuoffer C, Zanolari B, Erni B. 1992. A functional protein hybrid between the glucose transporter and the N-acetylglucosamine transporter of *Escherichia coli*. Protein Sci. 1:356-62.

Hunter IS, Kornberg HL. 1979. Glucose transport of *Escherichia coli* growing in glucose-limited continuous culture. Biochem J. 178:97-101.

Lengeler JW. 2015. PTS 50: Past, present and future, or diauxie revisited. J Mol Microbiol Biotechnol. 25:79-93.

McCoy JG, Levin EL, Zhou M. 2015. Structural insight into the PTS sugar transporter EIIC. Biochim Biophys Acta. 1850:577-85.

Metcalf WW, Jiang W, Daniels LL, Kim SK, Haldimann A, Wanner BL. 1996. Conditionally replicative and conjugative plasmids carrying lacZ alpha for cloning, mutagenesis, and allele replacement in bacteria. Plasmid. 35:1-13.

Miller BG, Raines RT. 2004. Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. Biochemistry. 43:6387-92.

Miller BG, Raines RT. 2005. Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases. Biochemistry. 44:10776-83.

Nichols BP, Shafiq O, Meiners V. 1998. Sequence analysis of Tn*10* insertion sites in a collection of *Escherichia coli* strains used for genetic mapping and strain construction. J Bacteriol 1998. 180:6408-11.

Notley-McRobb L, Ferenci T. 2000. Substrate specificity and signal transduction pathways in the glucose-specific enzyme II (EII(Glc)) component of the *Escherichia coli* phosphotransferase system. J Bacteriol. 182:4437-42.

Oh H, Park Y, Park C. 1999. A mutated PtsG, the glucose transporter, allows uptake of D-ribose. J Biol Chem. 274:14006-11.

Orita I, Iwazawa R, Nakanura S, Fukui T. 2012. Identification of mutation points in *Cupriavidus necator* NCIMB 11599 and genetic reconstitution of glucose-utilization ability in wild strain H16 for polyhydroxyalkanoate production. J Biosci Bioeng. 113:63-69.

Pennetier C, Domínguez-Ramírez L, Plumbridge J. 2008. Different regions of Mlc and NagC, homologous transcriptional repressors controlling expression of the glucose and N-acetylglucosamine phosphotransferase systems in *Escherichia coli*, are required for inducer signal recognition. Mol Microbiol. 67:364-77.

Plumbridge J, Kolb A. 1993. DNA loop formation between Nag repressor molecules bound to its two operator sites is necessary for repression of the nag regulon of *Escherichia coli in vivo*. Mol Microbiol. 10:973-81.

Plumbridge JA. 1989. Sequence of the nagBACD operon in *Escherichia coli* K12 and pattern of transcription within the *nag* regulon. Mol Microbiol. 3:505-15.

Postma PW, Lengeler JW, Jacobson GR. 1996. Phosphoenolpyruvate:carbohydrate phosphotransferase systems. In: Neidhardt FC (ed.). *Escherichia coli* and *Salmonella:* Cellular and Molecular Biology, 2nd ed. Washington, DC: ASM Press. 1149-74.

Raberg M, Peplinski K, Heiss S, Ehrenreich A, Voigt B, Döring C, Bömeke M, Hecker M, Steinbüchel A. 2011. Proteomic and transcriptomic elucidation of the mutant *Ralstonia eutropha* G+1 with regard to glucose utilization. Appl Environ Microbiol. 77:2058-70.

Reidl J, Boos W. 1991. The *malX malY* operon of *Escherichia coli* encodes a novel enzyme II of the phosphotransferase system recognizing glucose and maltose and an enzyme abolishing the endogenous induction of the maltose system. J Bacteriol. 173:4862-76.

Schiefner A, Gerber K, Seitz S, Welte W, Diederichs K, Boos W. 2005. The crystal structure of Mlc, a global regulator of sugar metabolism in *Escherichia coli*. J Biol Chem. 280:29073-79.

Schnetz K, Sutrina SL, Saier MH, Rak B. 1990. Identification of catalytic residues in the β-glucoside permease of *Escherichia coli* by site-specific mutagenesis and demonstration of interdomain cross-reactivity between the β-glucoside and glucose systems. J Biol Chem. 265:13464-71.

Steinsiek S, Bettenbrock K. 2012. Glucose transport in *Escherichia coli* mutant strains with defects in sugar transport systems. J Bacteriol. 194:5897-908.

Tchieu JH, Norris V, Edwards JS, Saier MH. 2001. The complete phosphotransferase system in *Escherichia coli*. J Mol Microbiol Biotechnol. 3:329-46.

Zeppenfeld T, Larisch C, Lengeler JW, Jahreis K. 2000. Glucose transporter mutants of *Escherichia coli* K-12 with changes in substrate recognition of IICB(Glc) and induction behavior of the *ptsG* gene. J Bacteriol. 182:4443-52.