Engineering Cellular Transport Systems to Enhance Lignocellulose Bioconversion

by

Gavin Kurgan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2018 by the
Graduate Supervisory Committee:

Xuan Wang, Chair
David Nielsen
Rajeev Misra
Brent Nannenga

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

Lignocellulosic biomass represents a renewable domestic feedstock that can support large-scale biochemical production processes for fuels and specialty chemicals. However, cost-effective conversion of lignocellulosic sugars into valuable chemicals by microorganisms still remains a challenge. Biomass recalcitrance to saccharification, microbial substrate utilization, bioproduct titer toxicity, and toxic chemicals associated with chemical pretreatments are at the center of the bottlenecks limiting further commercialization of lignocellulose conversion. Genetic and metabolic engineering has allowed researchers to manipulate microorganisms to overcome some of these challenges, but new innovative approaches are needed to make the process more commercially viable. Transport proteins represent an underexplored target in genetic engineering that can potentially help to control the input of lignocellulosic substrate and output of products/toxins in microbial biocatalysts. In this work, I characterize and explore the use of transport systems to increase substrate utilization, conserve energy, increase tolerance, and enhance biocatalyst performance.

DEDICATION

To my wife Erin and daughter Tatum, for their unconditional love and support.

ACKNOWLEDGMENTS

I would first like to thank my mentor, Dr. Xuan Wang, for his guidance, advice, and contributions for all of the work performed. Xuan has been extremely supportive of my decisions both professional and personal throughout my education and I could not imagine having a better mentor. I especially appreciate the hands-on mentorship environment that he facilitated, and that he constantly talked to me casually about my work and future to get me thinking about the purpose of a scientific career. He has been an incredible scientific role model, and has pushed me to achieve things I previously did not think were within my reach.

Additionally, I would like to thank my colleagues within the lab, both past and present. In Chapter 2 Larry Panyon and Lizbeth Nieves contributed significantly to the construction and testing of ASU1 *gal* system inactivation strains. I also appreciate their early mentorship in my PhD, as they helped show me the ropes of our lab and helped establish our enjoyable work culture. Larry also was a part of the team that performed the initial adaptive laboratory evolution on KJ122 to obtain the strains used in this chapter, along with Chandler Morris. I also have to thank Dr. Christian Sievert, for his help in learning to perform genomic analysis with NGS datasets and for performing much of the initial genomic characterization in Chapter 2. Additionally, undergrads Tommy Billings and Aidan Schneider helped me construct the *galP* inactivation strains in Chapter 2, and performed much of the fermentations associated with them. Finally, Eric Taylor and Logan Kurgan also helped me perform several co-sugar fermentations for succinate production in Chapter 2, and largely helped train Tommy Billings for this project.

In Chapter 3, Eric Taylor helped me do all of the strain construction for the “EG” strains, and performed almost all of the transporter enrichment experiments himself. He was a major driver of this project and I cannot thank him enough. I also would like to thank Andrew Flores for being a great co-worker and providing the AF11A strain that we used as an initial genotype in Chapter 3. Furthermore, I’d like to thank Logan Kurgan for helping direct/train both Eric and Tommy Billings to do the fermentations that helped lead to the characterization of glucose inhibition of xylose transport by Glf. I’d also like to thank Aidan Schneider for helping me

characterize the performance of the Glf mutants in succinate producing strains, and for carrying the torch of this project after my leave.

Chapter 4 had a number of contributors, the most significant of which was probably Larry Panyon. He and Eric Pacheco did a large majority of the initial plate and liquid based screening of ASKA overexpression plasmids in LY180, and were essential for finding our initial candidates. Additional Lizbeth Nieves performed much of the custom plasmid construction in Chapter 4 which helped facilitate numerous experiments. I also have to thank Yesenia Rodriguez-Sanchez for performing many of the furfural supplemented fermentations in Chapter 4 and helping to integrate transporters into LY180. She also was crucial for testing these integrated strains, as well as trying to search for additional transporters that could be synergistic with our identified candidates. Also Robert Mann helped to perform experiments which demonstrated that overexpression of *lldP* was more sensitive to overexpression toxicity than the other candidates.

For Chapter 5 I owe a big thank you to my brother Logan Kurgan. Under my supervision he helped to perform almost all of the knockouts after the initial characterization of GK153, and was crucial to getting most of the fermentations in this chapter completed. I also have to thank Aidan Schneider, since he was the one who came up with the computational portion of this chapter and managed to really challenge us to think critically about alternative means to help identify transport proteins. Additionally, I'd like to thank Eric Taylor and Yesenia Rodriguez-Sanchez for helping with final fermentations to help finalize the figures for this manuscript. Lastly, I need to thank Dr. Pablo Carbonell and Dr. Haiwei Guo for providing assistance for the computational and metabolic aspects of this work, respectively.

I'd also like to thank Michael Machas, for being an excellent collaborator on work for engineering aromatic tolerance, even though it did not manage to make it into the Appendix. I definitely enjoyed having you as someone to share my pain in the difficulties of engineering transport mechanisms, as this has caused me far too many headaches.

I would also like to thank my committee members Drs. Dave Nielsen, Rajeev Misra, and Brent Nannenga. I sincerely appreciate all of your constructive feedback at our meetings and for helping me develop these pieces of work to become a better intellectual product. I would

especially like to thank Dave for his help in guiding the experimental work in Chapter 4. I've really enjoyed our conversations over the years and hope we manage to run into each other in the future.

I would especially like to give a huge thanks to all of my undergraduates over the years. These people have been at the heart of my work, and there is no way I could have done this work at such a rapid pace without them. I hope they have all learned as much from this process as I have. I know I definitely learned a lot about myself in the mentoring process, and definitely think I came out a better scientist and person because of it. This includes: Logan and Cullen Kurgan, Aidan Schneider, Junpei Xiao, Yesenia Rodriguez-Sanchez, Eric Taylor, Robert Mann, Brittany Retallack, Rebecca Condruti and Tommy Billings. I especially have to give a huge thanks to Logan, Aidan, and Yesenia. They were my first three undergraduates I mentored and were absolutely instrumental to the training of future undergraduates and execution of almost all of my work. If Logan hadn't stepped up to the plate to help manage the flow of work during my summer at Intrexon, I may have never managed to graduate in four years. Aidan has worked an incredible amount of hours and contributed to projects in novel ways that I could not have easily trained him. These novel ideas formed the basis of the computational work in Chapter 5 and have helped to inspire me to pursue my future in bioinformatics. Yesenia has truly been one of the hardest (and most careful) workers I have ever known. Thank you for all the times you went into the lab in the middle of the night so that I could take a night off from tedious experiments. Without her, there is no way that Chapter 4 would have happened at the speed it did.

I would also like to thank my mother and step-father, Katherine and Bill. Thank you for constantly encouraging me to pursue my dreams and raising me to be independent. Without your guidance, I feel I surely would not be where I am today. Additionally, I'd like to thank my brothers Logan, Cullen, and Brian, for always being there when I need them. My good friends have also played a big role in keeping me sane, so I have to say thanks to Brandon von Hellens and Trevor Halt for getting me out of the house when I needed it most. I'd also like to thank Sheila Matthias for the love and kindness she has shown me and my family over the years. I'd also like to extend

a big thank you to my extended family (the Billings) for keeping me sane and being supportive for me and my family.

Lastly, I'd like to thank my wife and daughter, Erin and Tatum. Thank you Erin for your unconditional love and support throughout this journey. You are truly my best friend and I could not imagine having a better partner in life. I'm so proud of the mother you have become and am so glad you had the patience to help me make it through all of this. Tatum, thank you for being the most amazing little girl in the world. Sometimes it feels like your goofiness was the only thing keeping me sane. I love you and your mother so much, and can't wait to begin life after graduate school with you both.

I wish I had enough brain power to thank all of you more, but I think I'm pretty much tapped out at this point.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Lignocellulosic bioconversion and transport bottlenecks

Modern society is unsustainably dependent on petroleum. For example, approximately 120 billion gallons of gasoline were consumed in the United States in 2012 (1). Besides transportation fuels, many other bulk industrial chemicals, such as solvents, fertilizers, pesticides, and plastics, are also derived from petroleum (2). Due to escalating global energy consumption, dramatic changes have manifested in the atmosphere and global climate (3). To ensure the future advancement of human society, a sustainable supply of energy and chemicals is needed. Production of fuels and chemicals by fermentation-based manufacturing processes using a renewable feedstock is a desirable alternative to petrochemical production.

Lignocelluloses are the most abundant renewable natural organic materials present on the earth (4-6). They account for more than 60% of total biomass and are renewable due to carbon fixing photosynthetic processes of plants, with a net productivity of 155 billion tons per year (7). Converting lignocellulose into fuels and chemicals does not compete with food sources since economic plants grown for food and other commercial purposes generate millions of tons of lignocellulosic waste. A wide range and variety of agricultural, forestry and industrial wastes are available for value-added microbial conversion. For example, approximately 731 million tons of rice straw is annually produced globally as an agricultural waste (8). The stems, leaves and fibers from agricultural crops generally have sugar content higher than 50% of their dry weight, and thus these agricultural waste residues represent abundant carbon feedstocks. Additionally, ample amounts of lignocelluloses are present in forest residues, wood sulfite waste, fruit/vegetable waste, waste paper and municipal solid waste (5-7). The potential for using such wastes as a renewable carbon source is of great importance.

Lignocellulose is a complex matrix present in plant cell wall structures; composed of many different polysaccharides, phenolic polymers, and proteins. Regardless of source, most lignocellulosic biomass contains cellulose, hemicellulose, and lignin as three major polymeric components as shown in Figure 1.1. Unlike starch, lignocellulose has been evolved to resist

deconstruction. Cellulose fibers are encased in a covalently-linked mesh of lignin and hemicellulose. Cellulose (30-50% of lignocellulose dry weight) is composed of D-glucose and is highly resistant to deconstruction. Efficient degradation of cellulose generally requires cellulases (9-11). Lignin (10-25% lignocellulose dry weight) is a heterogeneous aromatic polymer and is difficult for microbes to use as carbon source. Three basic phenol derivatives (p-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol), the so-called monolignols, make up almost all types of lignin found in nature (12, 13). Hemicellulose (20-35% of lignocellulose dry weight) is composed of a mixture of pentoses and hexoses mostly with D-xylose as the major sugar (Figure 1.1) (5, 6, 14). Fermentable sugar content of lignocelluloses occupies 50 to 70% of biomass dry weight, which is comparable to corn (5). However, these sugars are covalently fixed in polymeric states, and thus require chemical and/or enzymatic pretreatment processes to release them for microbial conversion (5, 15). Simple pretreatments such as steam pretreatment with dilute mineral acids (e.g. sulfuric and phosphoric acid) are able to depolymerize hemicellulose into sugar monomers (5, 6, 14).

After pretreatment, the resulting syrups contain a mixture of hexoses and pentoses, however, conversion of this hydrolysate to valuable products is difficult for multiple reasons. First, utilization efficiency of sugar mixtures by microbes is hindered by a global regulatory mechanism called catabolite repression (16-18). This regulation is common for most microbes used in bio-based production, if not all, and is tightly controlled at transcriptional and biochemical levels (18, 19). Glucose is often the preferred substrate by industrial microbes and its presence represses the catabolism of other secondary sugars in lignocellulose such xylose and arabinose (20, 21). Under anaerobic or micro-aerobic fermentation, complete consumption of sugar mixtures at high rates is difficult, especially for high sugar concentrations (100 g/L total sugars or higher) (18). This results in sugar loss, decreased productivity, and lower product titers for lignocellulose conversion. Second, microbes that can utilize xylose, such as *Escherichia coli*, possess transport systems that expend energy for the uptake of xylose. In *E. coli,* import of xylose is catalyzed primarily by XylFGH, an ATP-binding cassette transporter that expends one ATP per molecule of D-xylose imported (22, 23). Thus, under anaerobic and microaerobic xylose fermentation conditions, the amount of ATP

spent on transport approximately doubles compared to glucose fermentation (22). This ultimately contributes to decreased fermentative production, especially under microaerobic conditions when ATP is scarce. Third, toxic side products, such as furan aldehydes, formate, acetate and soluble aromatics, are produced during sugar degradation steps in thermochemical pretreatments (6, 15). The aforementioned toxic side products can hinder or even be detrimental to cell growth (24). Furan aldehydes and aromatics are especially problematic due to both abundance and cytotoxicity (25, 26). Although efforts targeting cytosolic toxicity mechanisms of these compounds have incrementally enhanced tolerance, characterization of efflux systems for expulsion of lignocellulosic inhibitors from cells remains largely unexplored. Overliming to pH 10 with $Ca(OH)_2$ or other physical methods such as active carbon filter or vacuum treatment are able to remove inhibitors and thus reduce cytotoxicity of lignocellulosic hydrolysate syrups (27, 28). However, these extra steps increase process complexity and operational costs, thus reducing economic viability. Lastly, the metabolic pathways of hosts catabolizing lignocellulosic sugars are often altered to direct a high flux of carbon to a non-natural product that may not optimally be transported to the extracellular space. Both product and lignocellulosic inhibitor export from the cell is then potentially catalyzed by promiscuous or other non-ideal transport systems, limiting fermentative production metrics and imposing toxicity on the biocatalyst. Thus, for bioconversion of lignocellulose to valuable products to be efficient, robust biocatalysts must be developed that optimally transport and utilize the sugars present while tolerating the toxic effects of both inhibitors and products in the media. The optimization and characterization of transport systems for this purpose is the theme of my dissertation.

Overview of the bacterial cell envelope

To effectively manipulate transport systems, it is important to have a high level understanding of the cellular envelope of the organism of interest. The cellular envelope of bacteria consists of either one or two lipid membranes, a feature that discerns bacteria as either Gram positive or Gram negative, respectively (29). The cellular envelope of Gram negative bacteria consists of an outer membrane, a cell wall containing a thin layer of peptidoglycan, and an inner membrane (30, 31). In contrast, Gram positive cells lack an outer membrane and instead have a much thicker

peptidoglycan cell wall that contains covalently linked anionic polymers called teichoic acids that help maintain membrane structure and function (31, 32). Although different in structure, the cell envelope of these organisms share the same general functions to maintain cell boundaries, selectively allow the passage of molecules to the cytoplasm, and maintain critical electrochemical gradients (31). For the sake of brevity, I will only discuss the cell envelope of Gram negative bacteria in detail.

The outer membrane is composed of lipopolysaccharides (LPS), phospholipids and approximately 50% proteins by mass and serves a number of functions involving nutrient transport, protein translocation, signal transduction, and protection from the extracellular environment (30, 31, 33). Similar to the inner membrane, the phospholipid composition of the outer membrane is ~80% phosphotidylethanolamine, 5% cardiolipin, and 15% phsophotidylglycerol (34). The LPS anchored to the outer membrane generally consist of an external hydrophilic polysaccharide region followed by an internal hydrophobic core which, together, serve as an effective permeability barrier against hydrophobic and hydrophilic molecules (30, 35). For example it has been observed that LPS slow the penetration of lipophilic molecules by diffusion to ~1-2% the rate of a normal phospholipid bilayer (36). Thus for many molecules to effectively cross the outer membrane, they must enter through proteinaceous channels known as porins (37). These β-barrel proteins essentially act as both specific and non-specific molecular filters, often preventing the passage of molecules larger than 600-1,000 Da (33, 38). For instance, OmpF is a general diffusion porin that is known to prefer cation molecules less than 600 Da in size (39), whereas other ligand-specific porins, such as LamB, are induced by maltose/maltodextrins and are essential for uptake of these molecules (40). Even hydrophobic molecules such as toluene and benzene utilize porins for effective uptake due to the slow rate of diffusion through the outer membrane (41-43). Thus the outer membrane serves as an important first step in facilitating the exclusion and uptake of nutrients to the periplasm.

Once in the periplasm, molecules can enter the cytoplasm through carrier-dependent and/or independent mechanisms depending on the properties of the molecule. Small uncharged molecules (e.g. ethanol, butanol), gases (such as O_2 or CO_2), and, to some extent, water can cross

the inner membrane via carrier-independent mechanisms such as passive diffusion (44). In this process, the rate of transport largely depends on the concentration gradient across the membrane and hydrophobicity of the molecule. Due to the correlation between partition coefficients of lipophilic compounds in membrane-buffer systems and octanol-water mixtures (Log P_{ow}), the relevant hydrophobicity of a molecule is often determined by Log P_{ow} (45, 46). As Log P_{ow} increases it becomes easier for a molecule to traverse the hydrophobic core of the inner membrane, which has been linked to increases in toxicity (46). Sometimes, however, the hydrophobicity of a molecule can actually make it favorable to partition and reside in the inner membrane instead of completely diffusing into the cytoplasm (46). This is a common toxicity mechanism of many short alcohols and small lipophilic molecules, such as those found in lignocellulose hydrolysate (e.g. furfural, syringaldehyde) and as valuable metabolic end-products of biocatalysis (e.g. ethanol, phenol) (24, 46, 47). This accumulation within the inner membrane results in disruption of phospholipid-phospholipid interactions which can cause alterations in membrane fluidity (45, 48). Alterations in membrane fluidity can result in leaky membranes, dissipating important proton (49, 50) and nutrient (51, 52) gradients for maintaining the cellular energy status, intracellular pH and nutrient homeostasis (46). Although this process is free of energy expenditures, passive diffusion is a relatively slow and potentially toxic process, thus some molecules capable of passive diffusion still have carrier-dependent transport mechanisms to cross the inner membrane as well.

Channels are integral membrane proteins that transport ions across the inner membrane using concentration or electrochemical gradients. This passive transport is thus similar to diffusion, but occurs on a much faster time scale. For instance, GlpF and AqpZ are known to facilitate diffusion of water molecules at a rate could not be achieved when considering diffusion through a membrane (53, 54) . This allows cells to rapidly adjust to differences in osmotic pressure, among other things. These channels can be gated as well to provide control mechanisms to open/close the channel only when certain electrochemical, physical, or chemical signals are present. Possibly the most well-known example of this is with voltage gated channels for generating an action potential in mammalian cells (55). Although these proteins are often critical for maintaining cellular

homeostasis of water and ions, transport of organic compounds such as sugars are primarily handled by membrane proteins called transporters.

Transporters are a diverse group of membrane proteins which can transport a wide range of ions, organic, and inorganic compounds using primary active, secondary active, or passive transport mechanisms. Active transporters, as the name implies, expend some type of energy to transport chemicals against a concentration gradient. Primary active transporters are single or multi-component membrane transport complexes which hydrolyze ATP to transport a compound across the membrane (56). The free energy released from ATP hydrolysis allows these transporters to concentrate a compound against a chemical gradient, making these transporters excellent high-affinity transporters for compounds present in low concentrations in the extracellular environment. These ATP-binding cassette (ABC) transporters have been reviewed extensively and are known to carry out many key physiological activities for both import and export in cells (56-58). For instance, MsbA catalyzes transport of lipid A, the hydrophobic moiety largely present in LPS from the cytoplasm to the periplasm (59). This is a function essential to survival for many Gram negative cells, as intracellular accumulation of lipid A is known to be toxic and LPS are needed to help protect cells from the extracellular environment (35, 60). These transporters can have diverse conformational changes and mechanisms that permit transport, but all of these mechanisms depend on ATP hydrolysis in the nucleotide binding domain (NBD), a structure that is relatively widely conserved within ABC transporters (56).

Secondary active transporters are permeases which use a favorable chemical gradient of one chemical to help facilitate the transport of another chemical against a chemical gradient (61). In this manner, the co-substrate actually drives the thermodynamically favorable accumulation or export of the substrate. This co-substrate can either be transported in the same or opposite direction across the membrane, a process termed symport or antiport, respectively. Although the co-substrate for this process can widely vary, it is common for many transporters to leverage the proton (H^+) and electrochemical gradients (Na^+, K^+) of the cell to drive translocation (61, 62). For instance many multidrug resistance (MDR) transporters use the proton motive force (proton:substrate antiport) as an energy source to drive the export of a broad range of intracellular lipophilic chemicals

and drugs (63-65). Import of a proton as a co-substrate has an energetic cost however of approximately 0.33 ATP per molecule transported due to the need for ATP synthase to hydrolyze ATP to maintain the proton gradient (66, 67). However, some transport systems can actually be electrogenic due to translocation of protons to the periplasm or net intracellular consumption of protons. This can be seen in lactate efflux with many Lactic Acid Bacteria (LAB). For example, *Lactococcus lactis* can catabolize sugars to lactate and use the high intracellular lactate concentration to drive $lactate^-/2H^+$ symport from the cytoplasm to the periplasm (68). Other non-obvious transport mechanisms can be electrogenic as seen in malolactic acid fermentation, which couples $malate^{2-}_{out}/lactate^-_{in}$ antiport with a proton consuming decarboxylation reaction to generate a net decrease in intracellular protons (69). Without the functioning of this transporter, malolactic acid fermentation cannot proceed, demonstrating the importance of transport mechanism on the cellular energy status.

Passive transporters are uniporters (or facilitators) which facilitate the transfer of chemicals down a chemical concentration gradient by facilitated diffusion. In this sense, they are very similar to channels, but are not gated to simply open/close in response to chemical, electrical, or physical stimuli. Instead, they facilitate transport with conformational changes more similar to the mechanisms of symporters and antiporters (62, 70). For instance, GLUT1 is a well-studied glucose facilitator in eukaryotes which is thought to import glucose by an alternating access mechanism (70, 71). This is a mechanism common to many uniport, symport, and antiport transporters that proposes that the transporter essentially “rocks” between two major conformations: one with the ligand binding site facing the periplasm, and the other facing the cytoplasm (70). Similar to channels, this mechanism is much faster than diffusion through a membrane, and is free of energetic costs. It is important to note that there are not many well-characterized uniporters in bacteria, although there is at least one characterized sugar facilitator, *glf* from *Zymomonas mobilis*, which appears to function similarly to eukaryotic sugar facilitators (72).

Finally, group translocators are a transport system that is relatively unique to bacteria and involves the transfer of a molecular donor to facilitate and provide energy for transport. The sugar phosphotransferase system (PTS) is the only very well studied member of these transporters, and

facilitates efficient transport of sugars from the periplasm to the cytoplasm by transferring the phosphoryl from phosphoenolpyruvate (PEP) to a membrane transporter complex (73, 74). One advantage of a group translocator system, such as the PTS, is that it can immediately transform the incoming substrate by phosphorylation which limits escape from the cytoplasm by a reversal of the uptake reaction. The PTS and other sugar transport systems of bacteria are however highly complex and involved in the regulation of catabolic pathways, making them an important target in engineering efficient utilization of sugars derived from lignocellulose (73, 75). This information is all summarized in Figure 1.2.

Uptake and utilization of lignocellulose-derived sugars

In many bacterial species, the PTS facilitates the transport and concomitant phosphorylation of exogenous carbohydrates across the cytoplasmic membrane (75). Using *Escherichia coli* as an example, the PTS of *E. coli* is a multiprotein phosphorelay system consisting of two soluble and non sugar-specific enzymes Enzyme I (EI) and the histidine protein (HPr), encoded by the *ptsI* and *ptsH* genes, respectively, and the sugar-specific enzyme Enzyme II (EII) system (21) (Figure 1.3). EII is a multicomponent complex composed of two hydrophilic domains, EIIA and EIIB, and one or two carbohydrate selective transmembrane domains, EIIC and EIID (76). These mentioned domains of EII may occur as individual proteins or as a combination of subunits in variable order and number (76) (Figure 1.3). Multiple parallel EII complexes facilitate cellular uptake of different carbohydrates. The *E. coli* genome encodes for more than 20 different EII complexes, thus allowing for the transport and simultaneous phosphorylation of more than 20 different carbohydrates (77).

In the PTS, relay of the phosphoryl group initiates with the autophosphorylation of EI from phosphoenolpyruvate (PEP) and subsequently transfers the phosphoryl group to a histidine residue on the HPr (His-15 in *E. coli*) (21, 78). HPr then phosphorylates various sugar-specific EII complexes. In *E. coli*, the glucose specific EII complex comprises of the soluble enzyme $EIIA^{Glu}$ and the integral membrane permease $EIIBC^{Glu}$, encoded by *crr* and *ptsG*, respectively (78). The reported kinetic activity of EII^{Glu} with glucose as substrate is reported to have a high affinity with Km and Vmax values of 3-10 μM and 126 μmol min^{-1} g^{-1}, respectively (79, 80). Lastly, the phosphoryl group is transferred to EII's corresponding sugar during transport across the

cytoplasmic membrane (21) (Figure 1.3). With the monosaccharide phosphorylated it can now be catabolized through the respective pathways. For example, in *E. coli* glucose-6-phospate can be catabolized by the Embden-Meyerhof-Parnas (EMP) pathway or the Pentose Phosphate Pathway (PPP). Glucose transport by means other than the PTS EIIBC^{Glu} complex has been observed in *E. coli* (Figure 1.3). There is compelling evidence that the mannose sugar specific EII complex (EII^{Man}) is a promiscuous component of PTS and is able to transport glucose, fructose and N-acetylglucosamine. The EII^{Man} complex comprises of the EIIAB^{Man} homodimer enzyme encoded by *manX* and the integral membrane permease EIICD^{Man}, encoded by *manY* and *manZ*. The phosphoryl group from PEP is transferred to EIIAB^{Man} and then to EIICD^{Man} which facilitates the transport and concomitant phosphorylation of glucose. The reported kinetic activity of EII^{Man} with glucose as substrate is reported to have a high affinity with Km and Vmax values of 15 µM and 72 µmol min^{-1} g^{-1}, respectively (73, 79, 81). Glucose may also be transported via the galactose proton symporter (GalP). Upon transport of glucose via GalP, the enzyme glucokinase, encoded by *glk*, phosphorylates glucose producing glucose-6-phosphate, which can now enter EMP or PPP. The reported kinetic activity of GalP with glucose as substrate is reported to have Km and Vmax values of 10.2 µM and 15.6 µmol min^{-1} g^{-1}, respectively (73, 82). Transport of extracellular lignocellulose-derived pentoses such as xylose and arabinose across the plasma membrane in *E. coli* occurs not through the PTS but through two unique set of transport systems: an ATP-binding cassette (ABC) and a proton symporter. The ABC transporters XylFGH for xylose and AraFGH for arabinose, encoded by *xylFGH* and *araFGH*, respectively, actively transport sugars with the cost of one ATP per sugar, whereas the proton symporters XylE for xylose and AraE for arabinose, encoded by *xylE* and *araE*, respectively, uses a proton gradient to transport the monosaccharide across the plasma membrane (83-87). Both biochemical and genetic tests have suggested that XylFGH and AraE are primarily responsible for the uptake of their respective sugars in *E. coli* (22, 88). In other species such as *Zymomonas mobilis*, uniport of hexose and pentose sugars can be performed using a sugar facilitator Glf, encoded by *glf* (72, 89).

AraE also has affinity towards xylose but its expression is repressed by xylose under normal conditions (88, 90). Both ABC transporters XylFGH and AraFGH exhibit a high affinity with low Km

values between 0.2-4 μM and 4.1-6.1 μM, respectively. The proton symporters XylE and AraE possess a relatively low affinity with high Km values between 63-169 μM and 150-320 μM, respectively (84, 91, 92). Immediately following the transport of xylose and arabinose into the cell both substrates are eventually converted to xylose-5-phosphate via the xylose isomerase and arabinose isomerase pathway (93, 94) (Figure 1.3). In the xylose isomerase pathway, xylose is converted to xylulose through a reversible one-step reaction catalyzed by xylose isomerase, encoded by *xylA*. Xylulose is then converted to xyulose-5-phosphate by the xylulokinase, encoded by *xylB* (95). Catabolism of arabinose begins by converting arabinose to ribulose by arabinose isomerase, encoded by *araA*. Ribulose is then phosphorylated by ribulokinase, encoded by *araB*, and finally, converted to xylulose-5-phosphate by ribulose-5-phosphate 4epimerase, encoded by *araD* (85, 94). With both xylose and arabinose converted to xylulose-5-phosphate, substrates can now be catabolized eventually by the EMP or PPP pathways. Many bacterial biocatalysts such as *Z. mobilis*, *Pseudomonas putida*, *Streptomyces coelicolor*, *Corynebacterium glutamicum*, *Rhodococcus opacus* and *Lactococcus lactis*, do not appear to have complete catabolic pathways for either xylose, or arabinose, or both (missing at least two homologues) (96-99) (Figure 1.4). Missing genes need to be integrated into these strains for an ideal lignocellulose conversion to use up all sugar content (100, 101).

Lignocellulose-derived pentoses and hexoses are transported and preferentially selected for further catabolism based on accessibility and utilization that allows for optimal growth rate by bacteria (18, 102). Glucose commonly represses the catabolism of other secondary sugars such as xylose, arabinose and galactose, which causes hierarchical control of sugar mixture utilization (103-106). Catabolite repression is a well-studied and classic topic for bacterial global transcriptional regulation. In *E. coli* and many other enteric bacteria, catabolite repression is controlled by two non-mutually exclusive mechanisms: 1) operon-specific regulatory mechanisms, such as inducer exclusion, and 2) global regulatory mechanisms (21). Inducer exclusion is an operon-specific regulatory mechanism that controls the formation or uptake of an operon's inducer. A classic example is the glucose repression of the *lac* operon transcription through lactose permease transporter, LacY, in *E. coli* (21, 107-109). Expression of the *lac* operon requires lactose

to be isomerized by β-galactosidase forming allolactose, which can bind and inactivate the lac repressor. In a sugar mixture of glucose and lactose, the preferred sugar glucose at high concentrations causes the PTS glucose specific EIIA ($EIIA^{Glu}$) to be dominantly dephosphorylated and $EIIA^{Glu}$ binds to LacY, preventing lactose transport (Figure 1.3). At low concentrations of glucose, the $EIIA^{Glu}$ is dominantly phosphorylated and cannot bind to LacY, thereby allowing lactose to be transported across the cell membrane, isomerized, and relieve repression of the *lac* operon by binding to the *lac* repressor. The $EIIA^{Glu}$ is an essential component of the PTS not only because is it a major component of catabolite repression, but it has been shown that the same mechanism is applicable to other secondary non-PTS carbohydrates, such as maltose, melibiose and glycerol (21, 78). For abundant secondary sugars in lignocellulose such as xylose and arabinose, the catabolite repression potentially caused by inducer exclusion mechanism remains to be investigated. The catabolite repression caused by global regulatory mechanisms generally involves global transcriptional regulators to modulate the transcription of catabolic genes for secondary sugars (Figure 1.3). In *E. coli*, the main involved global regulator is CRP (cAMP receptor protein), also called catabolite gene-activator protein (CAP), which is the transcriptional activator for catabolic genes for secondary sugars such as xylose and arabinose when bound by cAMP, an important intracellular signaling molecule employed in many different organisms. The membrane-bound protein adenylate cyclase (AC) and the $EIIA^{Glu}$ component of the PTS are also essential parts of regulation (21, 73, 75, 78). At low glucose concentrations copious amounts of phosphorylated $EIIA^{Glu}$ exist and are able to bind and activate AC, leading to the synthesis of cAMP. As cAMP intracellular concentrations increase, the formation of cAMP-CRP complex activates catabolic operons such as *xylAB*, *xylFGH* and *araBAD* (Figure 1.3). The promoters for catabolic operons of secondary sugars are usually weak and require co-activation by both CRP and theirs own sugar-specific activators such as XylR and AraC to enhance binding of the RNA polymerase (Figure 1.3). In contrast, at high glucose concentrations, the cAMP level is low due to inactivation of AC, thereby limiting the CRP activity to increase the transcription of catabolic operons for secondary sugars (21, 110, 111). The global regulator CRP plays an essential role in not only regulating secondary catabolic genes, but also many other important biological processes such as

respiratory genes and multidrug resistance, with over 180 genes under its control (107, 112). The intricate catabolite repression mechanisms in other microorganisms are less understood.

The inability of bacteria to efficiently consume two or more carbon sources hinders commercial use of lignocellulosic biomass due to increased residence time, lower product titer and productivity (73). In the past few decades, considerable engineering efforts have been made and successes have been achieved to engineer bacteria for co-utilization of lignocellulose-derived sugars such as glucose, xylose and arabinose. The strategies to abolish catabolite repression by inactivating the $EIIBC^{Glu}$ complex, encoded by *ptsG*, have been explored. For example, simultaneous uptake of glucose and pentoses for conversion to lactate was achieved at some degree by expressing the lactic acid dehydrogenase from *Streptococcus bovis* in a *ptsG* mutant yielding the FBR19 strain (102, 113). This strain produced a lactic acid titer of 64 g/L with a yield (g/g) of 77%. However, inactivation of the PTS often impairs glucose uptake and thus efforts to compensate this defectiveness are needed. One common strategy is to evolve PTS mutant strains for better growth using glucose as sole carbon source. For example, the work of Hernández-Montalvo et al. showed that subjecting the PTS devoid mutant (PTS^- $glucose^-$ phenotype) NF6 to a continuous culture selective method led to revertants (PTS^- $glucose^+$ phenotype) with a specific growth rates on glucose ($\mu = 0.36\ h^{-1}$) comparable to a PTS^+ wild-type *E. coli* strain ($\mu = 0.42\ h^{-1}$) under aerobic conditions (114, 115). In these revertant mutants, catabolite repression of arabinose and xylose was alleviated and this allowed for simultaneous mixed sugar uptake (1 g/L glucose, xylose, and arabinose). Although sugar co-utilization was achieved, catabolite repression was only eradicated in a glucose-arabinose sugar mixture. In a glucose-xylose mixture, glucose exerted repression on xylose uptake. In an arabinose-xylose or a glucose-arabinose-xylose mixture, arabinose repressed xylose utilization (115). Additionally, it was shown that a plasmid overexpressing a non-PTS transport system, the galactose proton symporter (*galP*) and glucokinase, in a PTS-deficient strain (W3110 PTS^-) resulted in growth rates corresponding to 89% of the growth rate for W3110 (116). Building upon the work of Hernández-Montalvo and other colleagues, a W3110 mutant with a PTS^- $glucose^-$ phenotype was subjected to anaerobic adaptive laboratory evolution to obtain an evolved strain, VH30N4, capable of utilizing glucose (phenotype of PTS^- $glucose^+$). Upon transforming an

ethanologenic plasmid pLOI594 into VH30N4, the newly evolved strain was capable of co-utilizing a 5 g/L glucose-xylose sugar mixture with a specific growth rate μ of 0.14 and ethanol titer of 2.1 g/L (117). These adaptive laboratory evolution approaches to obtain PTS^- $glucose^+$ mutants demonstrate that catabolite repression can be alleviated to allow simultaneous uptake of lignocellulose-derived sugar. Most recently, our lab performed adaptive laboratory evolution on an *E. coli* biocatalyst for conversion of xylose to lactate. This resulted in the finding that two point mutations in the xylose operon transcriptional activator, *xylR*, led to constitutive activation of the xylose catabolic operons independent of cAMP-CRP (118). Utilization of this mutation has shown to permit glucose-xylose co-utilization under microaerobic conditions under industrially relevant batch fermentation sugar concentrations (~10% w/v), thus largely eliminating the issue of catabolite repression in glucose-xylose sugar mixtures (118).

Besides catabolite repression, ATP consumption of bacterial pentose transport systems has also been implicated as a potential limitation of achieving efficient sugar co-utilization. ATP is an important currency for 1) cell growth, 2) cell maintenance, and 3) metabolite transport systems. Under aerobic conditions, oxidative phosphorylation will generate the majority of the cellular ATP budget and cells grown on either glucose or xylose can generate approximately equal amounts of biomass even though cells growing on xylose expend approximately twice as much energy on transport (22). Production of renewable chemicals using aerobic processes can be desirable for some valuable products, but aeration significantly increases the cost of a bioprocess. This makes it economically more challenging to compete with many products traditionally derived from petroleum feedstocks, and thus microaerobic fermentation is often used to decrease operating costs. However, under microaerobic conditions, the main mechanism of ATP production is glycolysis due to inactivation of oxidative phosphorylation, and the ATP budget of cells greatly decreases (22). Under these conditions, the transport of xylose consumes approximately 71% of the ATP budget, and cells grown using xylose show decreased biomass formation and specific growth rates (22). In contrast, during growth on glucose cells only allocate approximately 34% of the ATP budget for transport (22). Using XylFGH for substrate import would theoretically cost twice

as much ATP equivalents as the glucose PTS, thus further demonstrating that XylFGH is the main xylose import mechanism in *E. coli* under aerobic and microaerobic conditions.

In comparison to catabolite repression, the burden of ATP consumption during pentose transport has been much less thoroughly investigated as a limitation for lignocellulose bioconversion, and thus robust genetic modifications have yet to be characterized. Strains constructed for homolactic acid fermentation under microaerobic conditions can convert glucose to lactate, however no growth is observed during xylose fermentation (88, 119). This is most likely due to the decrease in ATP budget under microaerobic conditions in combination with ATP consumption of XylFGH lowering the ATP/xylose yield to 0.67 (119). Work from Utrilla et al demonstrated that a simple inactivation of *xylFGH* permits growth of a homolactate strain on xylose. Furthermore, by performing adaptive laboratory after inactivation of *xylFGH,* the specific growth rate can be further improved by ~50% after a mutation in the EIIC component of the galactitol PTS transport system, *gatC,* is gained that confers xylose transport activity (119). Multiple other studies have noticed a similar growth deficiency for *E. coli* biocatalysts converting xylose to succinate, a product with an inherently low ATP yield. (120-122). Similarly, inactivation of *xylFGH* in the succinate biocatalyst KJ122 led to enhanced growth and succinate production, and upon adaptive laboratory evolution the strain exhibited greatly increased xylose and glucose-xylose co-sugar fermentative production metrics (122). Although this work never characterized the causative mutation following adaptive laboratory evolution, work by Sawisit et al. performed a similar laboratory evolution without initial inactivation of *xylFGH*. By reverse engineering the strain, they found that a mutation in the galactose permease (*galP* G236D) that was hypothesized to confer xylose transport activity caused similar increases in production metrics (123). Interestingly, *xylFGH* was never identified to be inactivated in this strain however, suggesting that either *galP* is so highly expressed that a majority of xylose is imported through the mutant *galP*, or that more complicated interactions are occurring. In **Chapter 2** I thoroughly probe the genetic determinants permitting adaptation of KJ122 for xylose fermentation and find that *galP* has unknown regulatory interactions that actually affect energy conservation during xylose fermentation. Although these pieces of work seem to provide general knowledge that *xylFGH* is an energetic burden to cells converting xylose

to a broad range of products and this has implications on lignocellulose bioconversion, no work to our knowledge has explicitly looked into engineering more energy efficient xylose transport mechanisms in bacterial cells released from catabolite repression. Work by Tang et al. optimized glucose to succinate conversion in a PTS^- strain and found that overexpression of *glf* as an alternative glucose transporter resulted in increased production metrics relative to overexpression of *galP*, which could be a result of the energy saved during transport (124). Although Glf has also been observed to be a high velocity xylose facilitator, it can also be inhibited in the presence of glucose (125). Utilization of Glf for xylose uniport represents an energy efficient uptake system that has not been thoroughly explored in strains with catabolite repression released. In **Chapter 3** I develop a transport deficient strain with catabolite repression released, and create variants of *glf* that are capable of efficient xylose transport in the presence of glucose.

Tolerance and efflux of pretreatment inhibitors

Pretreatment of lignocellulosic materials is considered to be an essential unit process of a lignocellulosic biorefinery, accounting for 16-19% of its total investment (126). In the past few decades, a variety of pretreatment technologies have been developed to overcome the recalcitrance of lignocellulose, increase cellulase efficiency, and improve the yields of monomeric sugars. These different pretreatments yield varied amounts of sugars and degradation products and have varied effects on cellulose digestibility and lignin solubilization, which ultimately affect downstream bioconversion processes. It is challenging to achieve a high yield of fermentable sugars from lignocellulose while being economically viable however. Several physical and chemical pretreatment methods have been developed including dilute acid (127, 128), alkaline (129, 130), steam explosion (131, 132), organosolv (133, 134), ozonolysis (135, 136), ammonia fiber expansion (137, 138), ionic liquid (139-141) and other pretreatment technologies. The advantages and disadvantages of these technologies are briefly summarized in Table 1.1.

In most pretreatments, three major categories of by-products are produced: furan aldehydes, small organic acids and phenolic compounds (142). Furan aldehydes include furfural and 5-hydroxymethylfurfural (HMF), which are the dehydration products of pentoses and hexoses, respectively. Most phenolic compounds (e.g vanillin, syringaldehyde, etc.) are degradation

products from lignin (13). Pretreated lignocellulosic hydrolysates can also contain large amounts of organic acids, including acetic acid, formic acid, levulinic acid and lactic acid. These acids are only toxic at high concentrations compared to furan aldehydes (24, 143). Furan aldehydes, especially furfural, are able to potentiate the toxicity of other side-products and inhibit cell growth at relatively low concentrations (24, 143-145).The above-mentioned pretreatments in Table 1 yield different side product spectra even when treating the same type of biomass. Different types of biomass will also yield different amounts of side-products after the same pretreatment. Among these techniques, dilute acid pretreatment has widely been regarded as one of the most economically promising methods of biomass pretreatment (142, 146). The main advantages of this pretreatment method are that it is effective at hydrolyzing the hemicellulose fraction and increasing enzyme accessibility to the cellulose fraction with a relatively low operational cost. However, it also generates a wide variety of inhibitors, including furan aldehydes and other phenolic compounds. To mitigate the production of these inhibitors, downstream processes can be performed to decrease the concentration of inhibitors prior to fermentation. These include water washing the pretreated solids (147), overliming (148, 149), and reverse osmosis (150). Although relatively effective, these methods all introduce large additional costs at the industrial scale which are ultimately undesirable. Since the toxicity of side-products is a limiting factor for economically viable pretreatment technologies, it is desirable to have robust organisms that are capable of tolerating these compounds in fermentation broth. The development of biological solutions for addressing hydrolysate toxicity would not only be cheaper and more sustainable, but also more scalable.

Among common toxic side-products derived from lignocellulose pretreatments, furan aldehydes are arguably the most studied due to their abundance, toxicity and the unique property to potentiate the toxicity of other side-products. Furfural and 5-HMF are formed by the dehydration of sugars (pentoses and hexoses, respectively) during some pretreatment processes with a concentration range between 0 to 5 g/L depending on the severity of the processes (25, 151). More specifically, furan aldehydes arrest the growth of most microbes when present above a threshold (~0.5-1 g/L furfural and ~2-3 g/L 5-HMF for *E. coli* in mineral salts medium, respectively). Due to the highly similar structures, furfural and 5-HMF share the same degradation pathways and

potential toxicity mechanisms. The genetic traits or methods proven useful for furfural are also effective for 5-HMF (152-154). Furfural is a potentially more important inhibitor than 5-HMF for the following reasons: first, dilute acid pretreatments and other pretreatment technologies yield more furfural than 5-HMF in hemicellulose hydrolysates (5, 14). Second, furfural is more toxic than 5-HMF to industrial catalysts including both *E. coli* and *S. cerevisiae*. Therefore, most research about bacterial toxicity of furan aldehydes is focused on furfural.

In spite of the high toxicity of furan aldehydes, a number of microbes have evolved different strategies to deal with furan aldehydes, though with varying degrees of success. Microbes including *S. cerevisiae* and *E. coli*, for example, are capable of transforming furfural to furfuryl alcohol by a reduction reaction catalyzed by oxidoreductases (Figure 1.5A) (152). Furfuryl alcohol is a less toxic compound (24, 47) and is excreted into the culture medium and remains in the fermentation broth without further degradation (152). It was reported that a few bacterial oxidoreductases such as YqhD and FucO in *E. coli* (153, 155, 156), FudC in *C. glutamicum* (157), and ZMO0976 in *Z. mobilis* (158) are able to reduce furfural into furfuryl alcohol using reducing factors NADH and/or NADPH (Figure 1.5A). This appears to be a common mechanism used by different bacteria to detoxify aldehydes including furfural and 5-HMF (Figure 1.5B). Without further degradation, however, terminal accumulation of furfuryl alcohol (especially in fed-batch operation) can itself eventually lead to growth inhibition. In contrast, certain other bacteria (e.g., *Cupriavidus basilensis* HMF14) have evolved a complete biodegradation pathway to fully metabolize furfural (when supplied as a sole carbon and energy source) under aerobic or microaerobic conditions (Figure 1.5A) (159, 160). Furfural is firstly oxidized into 2-furoic acid and then further metabolized to α-ketoglutarate that eventually enters TCA cycle to provide energy and biosynthetic building blocks (Figure 1.5A). In contrast to biotransformation, biodegradation has the potential to provide an irreversible and absolute solution to the problem of furfural toxicity (Figure 1.5).

The exact mechanism of growth inhibition caused by furan aldehydes is not completely known, but it is thought to be multifaceted and include potential membrane damage (25, 93). Most positive genetic traits found thus far have focused on improving tolerance using cytosolic enzymes. One such approach for engineering bacterial tolerance to furan aldehydes is to improve microbial native

biotransformation or biodegradation abilities. Overexpression of furfural oxidoreductases confers tolerance to a variety of bacteria, although the actual kinetic properties could benefit from directed evolution (156, 161, 162). Second, the reducing cofactors play an important role for furfural biotransformation and cytotoxicity. Different growth conditions such as medium, aeration levels and carbon source influence the relative abundance of intracellular NADH and NADPH levels. For example, the NADPH source is relatively limited when *E. coli* is growing under anaerobic fermentative conditions using xylose as a sole carbon source. Exposure of furan aldehydes activates the transcription of furfural reductase genes such as *yqhD* (153). When *E. coli* cells overexpress the NADPH-dependent furfural reductase YqhD to reduce furfural or 5-HMF, the NADPH intracellular pool is depleted and cell growth is arrested (153, 155). Actually this depletion of NADPH by YqhD was proposed as the one of the main mechanism for furfural/5-HMF induced growth inhibition in *E. coli* under xylose fermentative conditions (153, 155, 156). With respect to the cofactor issue, NADH-dependent furfural reductases such as FucO are a better choice for anaerobic fermentation using xylose as a sole carbon source (156). The complete furfural degradation pathway discovered in *C. basilensis* HMF14 and other species (159) has not been introduced into common genetically tractable industrial hosts such as *E. coli* and *S. cerevisiae*. One caveat for this pathway is its oxygen requirement which is intrinsically conflicting with the anaerobic/microaerobic fermentation condition used in most bioproduction scenarios for biofuels and bulk chemicals. Although biological abatement can be effectively used to remove furfural from the media, this increases complexity and costs. After NADPH limitation was proposed as one toxicity mechanism for furfural, researchers have identified multiple effective methods to improve NADPH levels and thus to relieve furfural toxicity substantially. For example, deletion of *yqhD* or increased expression of *pntAB* (a transhydrogenase for interconversion of NADH and NADPH) increased tolerance to furan aldehydes in *E. coli* (154, 155, 163). DNA damage has also been evidenced to be a mechanism of furfural toxicity. Genomic libraries from three different bacteria and metagenomic libraries have been constructed and screened in separate instances that led to the discovery of *thyA/thyX* as genes which confer furfural tolerance to *E. coli* (164, 165). These genes encode thymidylate synthases, enzymes in dTMP biosynthesis, further suggesting that DNA

damage is a potential toxicity mechanism of these furans. A few genes such as *ucpA, lpcA, groESL, ahpC, yhiH, rna* and *dicA* are associated with furfural tolerance and the overexpression of these genes individually conferred furfural tolerance at varied degrees (152, 166, 167). However, it is difficult to characterize how these genes are functionally linked in furfural resistance. Therefore, further rational engineering of cellular tolerance using these traits is limited without clear understanding of the resistance mechanism. Another important and effective approach used to improve microbial tolerance for lignocellulose inhibitors is adaptive laboratory evolution (also called "metabolic evolution" or "evolutionary engineering") (155, 168-170). For example, the mutations that cause *yqhD* inactivation was found in multiple furfural-resistant mutants isolated by adaptive laboratory evolution, suggesting that NADPH starvation induced by furfural enzymatic reduction is one important toxicity mechanism (153). Transcriptomic analysis of some furfural resistant *E. coli* mutants showed that the transcription of multiple polyamine transporters was up-regulated compared to normal *E. coli* due to the gene multiplication at the chromosomal level as indicated by whole genome sequencing (171). Further tests discovered that overexpression of these polyamine transporter genes including *potE, puuP, plaP* and *potABCD* makes cells more resistant to furfural, suggesting a potential protection role of polyamine for important cellular constituents (171). Quite a few distinct beneficial genetic traits have been identified due to the complexity of toxicity modes of furan aldehydes (Table 2). It is plausible that multiple pathways need to be co-activated and recruited for optimization of furfural tolerance. The beneficial genetic traits mentioned above are summarized in Table 2.

Besides furan aldehydes, there are two other groups of toxic side-products: small organic acids and phenolic compounds (142). Acetic acid and formic acid are the major organic acids after lignocellulose pretreatments, although others may also be present (25, 142). Generally, lignocellulose hydrolysates need to be neutralized to a certain degree for the following fermentation processes if dilute acid pretreatment is used. Thus, small organic acids will be converted to salt forms after neutralization and the toxicity of these side products is greatly decreased. For example, only lower than 5 g/L acetate and formate were produced after a dilute acid pretreatment of sugarcane bagasse while *E. coli* can tolerate 12 g/L acetate or higher without significant growth

defect (14, 172). Furthermore, it has been shown that an ethanolgenic *E. coli* can tolerate up to 40 g/L acetate after adaptive laboratory evolution (173). Researchers also have identified potential acetate-resistant genetic traits in different microbes such as *E. coli* (174, 175) and *Z. mobilis* (176). So the toxicity issue of small organic acids is often not significant for industrial bacteria hosts. A variety of phenolic compounds are produced during some pretreatment processes (Table 1). Most of them are very toxic but the quantity is normally small compared to furan aldehydes and organic acids after pretreatments (142). Without clear understanding which phenolic compound is the predominant inhibitor, the research of engineering bacterial tolerance to phenolic compounds is limited. Recent research in *Z. mobilis* ZM4 showed that phenolic aldehydes were reduced into alcohol forms by oxidoreductases to decrease toxicity, analogous to how cells cope with toxic furan aldehydes by biotransformation (Figure 1.5A) (177). Conversion of these aldehydes into alcohol form is beneficial due to the reduced toxicity of the functional group and increasing culture inoculum has been shown to enhance tolerance of phenolic aldehydes (24, 47, 143). It seems reasonable to assume that the same approaches to engineering furfural tolerance, such as degradation, might be effective in engineering phenolic tolerance. However, one approach that has been widely overlooked for enhancing tolerance to furan aldehydes, phenolics, and even final metabolic end-products of engineered strains has been the engineering of the cell envelope.

In the context of lignocellulosic bioconversion, protection from the extracellular environment is an especially important function due to all of the lipophilic compounds that can be produced upon deconstruction of lignocellulosic biomass. A series of studies about the toxicity of side-products derived from lignocellulose suggest that the hydrophobicity of these chemicals is correlated with toxicity (24, 47, 143). The membrane damage caused by aromatic compounds was also proposed as an important toxicity mechanism (24, 25, 47, 143). Since membrane stress is a toxicity mechanism of potentially multiple lignocellulose-derived inhibitors, biocatalysts can benefit from strategies that decrease intracellular inhibitor concentrations and minimize membrane fluidity changes and the resulting disruptive effects (178). This can be accomplished by a broad range of changes, such as altering the conformation of phospholipid chains (179, 180), altering acyl chain length (181), modifying phospholipid head groups (182), and expressing transport proteins capable

of inhibitor export (183). However, very few of these have ever actually been explored in any detail for these molecules. One of the only pieces of work that has attempted to engineer the cell envelope for lignocellulose inhibitor tolerance modified phospholipid head groups to enhance tolerance to multiple compounds including furan aldehydes and multiple phenolics derived from lignocellulose (178, 182). By overexpressing a phosphotidylserine synthase, *pssA*, to increase the abundance of phosphotidylserine in the inner membrane, they managed to increase cell hydrophilicity and maintain a higher membrane potential than a control exposed to membrane damaging compounds (182). Multiple other works have demonstrated methods to alter acyl chain length and composition of phospholipids in the membrane to increase tolerance. However, these have been studied mainly with short alcohols, so it is difficult to tell how directly this will translate into tolerance for lignocellulose inhibitors.

Transporters capable of lignocellulose inhibitor efflux also have yet to be discovered. MDR transporters are known to have broad substrate specificity for efflux of lipophilic compounds, yet the affinity for many industrially relevant compounds is relatively unknown (Figure 1.6). MDR proteins include transporters from the Resistance Nodulation Cell-Division (RND), Major Facilitator (MF), Multidrug and Toxin Extrusion (MATE), ATP-binding cassette (ABC), and Small Multidrug Resistance (SMR) families. Since most lignocellulose inhibitors are relatively lipophilic it seems probable that an endogenous MDR transporter may be capable of efflux. In **Chapter 4** I construct a transporter library largely comprised of MDR transporters and screen for proteins capable of furfural efflux. I further extend the use of this library in **Appendix B** by screening for transporters capable of transporting a broad range of lipophilic lignocellulose inhibitors and aromatic bioproducts.

Enhancing production by engineering bioproduct efflux

Genetic systems have advanced rapidly to permit tuning of biological systems for the production of a diverse portfolio of compounds (184-186). In the past it was thought that endogenous transporters could facilitate export of a diverse array of chemicals of interest, and that most metabolic bottlenecks must be in the biosynthetic pathways. Although product export can be facilitated by promiscuous transport systems at low production metrics, as production parameters

increase the potential for increased intracellular accumulation of the product becomes more likely (187). This is supported by an increasing number of examples that have shown that modulation of endogenous or heterologous transport systems can increase production metrics (187-189). High intracellular accumulation of the metabolic end-product can be cytotoxic and potentially lead to regulatory changes through feedback inhibition (188). Thus, product export has been previously observed as a bottleneck for the bioproduction of both native and non-native products.

Engineering efficient export systems capable of enhancing production of non-native products is challenging for a number of reasons. One challenge is that export systems for many non-native bioproducts are uncharacterized and thus significant efforts are required just to identify a candidate transporter (190). Secondly, even once candidate proteins are identified, they may have sub-optimal kinetics due to a lack of selective pressure from the environment to couple the transporter structure to the desired function (191, 192). For instance, previous work has identified the RND pump AcrAB-TolC as a candidate exporter for multiple non-native bioproducts, including n-butanol, styrene, limonene, fatty acids, and more (191, 193-195). This activity has been attributed to the broad substrate specificity of the inner membrane transporter, AcrB, and thus it has been proposed that efflux and expression may be sub-optimal for non-native products. However, the components of complexes like AcrAB-TolC often exist in functional ratios (3:6:3) necessitating fine tuning of expression for an optimal phenotype (196-198). This is demonstrated by the work of Wang et al. in which expression modulation of individual components of *acrAB-tolC* led to enhanced production of two separate isoprenoid molecules (198). Interestingly, the optimal ratio may not have necessarily been the defined 3:6:3 ratio the pump exists in, suggesting a more complex relationship between these genes and other tolerance genes/phenotypes (198). The expression of these pumps can also be modulated through native transcriptional regulators to enhance tolerance, which may represent a more natural way to maintain the protein functional ratios. The *acrAB* genes are known to be regulated by multiple transcription factors including *marA, marR, acrR,* and *soxR* (199-201). Recent work showed that simultaneous introduction of a nonsense mutation in *marR* and insertion into *acrR* could be used to enhance tolerance to multiple solvents including cyclohexane and p-xylene, likely by upregulating the expression of *acrAB-tolC* (202). However, for an optimal

phenotype, transport proteins must be optimally expressed and have sufficient activity for the desired chemical. Directed evolution has been performed on AcrB to enhance tolerance to non-native products such as n-butanol, n-heptanol, n-octane, 1-hexene, and α-pinene (191, 195, 203). In addition, overexpression of variants that conferred enhanced n-butanol tolerance increased titer in a production strain using a membrane stress sensitive expression system by ~40% (204). Success of these transporter engineering strategies vary widely from chemical to chemical however, as the same approach that was successful in isolating a mutant protein conferring enhanced tolerance to 1-hexene was unsuccessful for other compounds that are substrates of AcrB, such as styrene (195).

For many native metabolites, endogenous transporters may be present for product efflux but under sub-optimal transcriptional regulation. For instance, amino acids are a commonly produced valuable metabolic end-product with a global market volume of more than 2 million tons annually (205). Due to the highly regulated metabolic network surrounding amino acid biosynthesis, accumulation of the product of interest actually begins to inhibit the biosynthetic pathway, necessitating that biocatalysts be resistant to inhibition (205-207). Pathway optimization has been facilitated by numerous random and rational engineering approaches to achieve high titers using microbes for production of L-alanine, L-valine, L-phenylalanine, L-threonine, and more (206, 208-210). Within these rational approaches, multiple pieces of work have incorporated feedback-resistant biosynthetic enzymes to further decrease inhibitory effects during amino acid production (211, 212). This strategy has been wildly successful in enhancing production of multiple amino acids, but can potentially be limited by incomplete knowledge of all the different genetic targets that are affected by allosteric inhibition. Feedback inhibition has thus been further decreased non-specifically by modulating the expression of amino acid exporters (188). To identify amino acid exporters, some previous studies have leveraged the toxicity of high concentration of amino acids or analogs to select mutants that aid in identification of the genetic determinants. For instance, to identify the alanine exporter, *alaE,* an *E. coli* strain incapable of L-alanine degradation was mutagenized and mutants with a growth deficiency were selected for in media with a toxic analog, L-alanyl-L-alanine (213). By creating a genomic library of the parent strain, they were able to prove

that *alaE* was capable of complementing the L-alanyl-L-alanine growth deficiency and furthermore that overexpression of this transporter could enhance alanine production 1.5 fold (214). Multiple other amino acid pathways have achieved enhanced titers by identifying and engineering expression of native exporters, including threonine, cysteine, arginine, and glutamate (188, 215-218). Modulation of exporters has also been shown to be synergistic with other rational engineering approaches. For example, one group engineered a threonine producing *E. coli* with a reduced genome (~14% less) that overexpressed a feedback resistant biosynthetic pathway and had relevant side product pathways inactivated. By inactivating threonine importers, and overexpressing the threonine exporter *rhtA23,* they achieved an ~80% increase in titer (219).

Similar work has also been done for the modulation of transporters for short-chain carboxylic acids for bioproduction, especially those derived from the Tricarboxylic Acid Cycle (TCA). This includes products such as malate, succinate, and citrate that have multi-billion dollar markets with applications in agricultural, food, and pharmaceutical industries (220). Unlike the metabolic pathways for amino acid production, many of these pathways are actually regulated in a positive feedback manner, such that increases in the product actually stimulates relevant product biosynthetic pathways and transporters (221, 222). For instance, the expression of TCA cycle enzymes *mdh, fumABC,* and *frdABCD* are actually enhanced by increasing concentrations of the C_4-dicarboxylates that they form, such as succinate (223, 224). Relevant succinate exporters such as *dcuB* and *dcuC* have also been demonstrated to be upregulated in the presence of C_4-dicarboxylates (225, 226). Although it might seem that positive feedback may be sufficient to adequately express transporters during production, modulation of actively expressed C_4-dicarboxylate transporters has been shown to further enhance production. For instance, Chen et al modulated the expression of *dcuB* and *dcuC* to enhance succinate titer by ~34% in an *E. coli* succinate production biocatalyst (227). In other hosts, such as fungi, this has been seen to be even more dramatic. For instance, the production of L-malate has been enhanced by >2-fold in *S. cerevisiae* by incorporation of a heterologous malate exporter, MAE1p from *Schizosaccharomyces pombeii* (228). This appears to be due to a lack of energy-efficient short-chain carboxylic acid exporters in bakers yeast, as production of lactate has also been proposed to be bottlenecked by

export (229). Other filamentous fungi such as *Aspergillus oryzae* have been able to increase malate production by more than two-fold just by overexpressing a native transporter homolog to MAE1p (189). By combining this approach along with overexpression of other important biosynthetic enzymes, they achieved an impressive titer of 154 g/l malate after approximately 164 h fermentation (189). Although *E. coli* has also been metabolically engineered to produce malate, there are no known malate exporters in Gram negative bacteria so this strategy has not been explored. This further highlights one of the most important challenges in engineering efflux systems for bioproducts, as transporters with affinity for relevant chemicals are largely unknown.

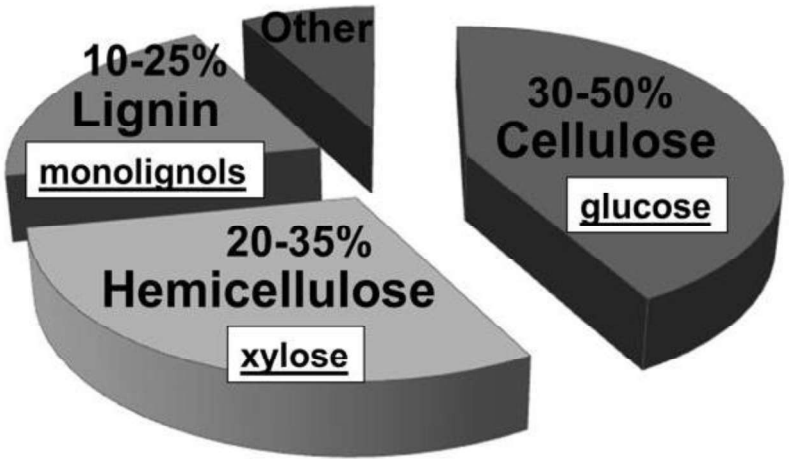
The identification of transporters for product efflux is a particularly challenging task due to difficulties in performing genetic screens. One approach is to perform a genome-wide screen on the native cellular transport system, inactivating or overexpressing genes on an individual basis (190). However, if cellular efflux is not strongly linked to growth in the screening environment or if there are multiple transporters participating in efflux, this methodology can be ineffective. Transcriptomic assays can also be used to probe the native regulation of efflux pumps upon exposure to the bioproduct. In this screening method, transport proteins regulated by the chemical serve as potential candidates. For example, this method has been used to identify that expression of an ABC transporter, MdlB, is modulated by isopentanol and overexpression of this transporter can be used to moderately enhance isopentanol tolerance, suggesting a potential role in efflux (230). However, this type of screen relies on native transport proteins being transcriptionally regulated by the molecule of interest, which may not be the case for products not naturally encountered in the evolutionary history (66). Additionally, both of these methodologies assume that the host genome has an ideal efflux protein already present. Other screens have sought to overcome this issue by performing heterologous screens of efflux proteins. For instance one group screened a heterologous *acrB* homolog library and found multiple transporters capable of complementing the native function of *acrB* for tolerance to terpenes (193). They also demonstrated with this approach that *acrB* is important for production of terpenes, such as limonene (193). Other work has screened a library of heterologous ABC transporters and found that multiple homologs of MsbA can facilitate efflux of and enhance production of multiple carotenoids, such a β-carotene

(231). In **Chapter 4,** I use this same screening methodology to characterize transporters for furfural efflux. However, in **Chapter 5** I use a more targeted methodology to perform a negative screen by rationally selecting transporters with affinity to molecules of similar chemical structures.

Figures.

Table 1.1. A summary of representative lignocellulose pretreatment technologies. Figure extracted from our book chapter (4).

Methods	Potential inhibitors	Advantages	Disadvantages
Diluted acid	Furans; phenolic and small organic acids at low levels	High hemicellulose monosaccharides production; increasing cellulose digestibility; relatively low cost	Inhibitors; equipment corrosion; low sugar concentration in exit stream
Alkaline	Furans at low levels	Some lignin degradation; increasing cellulose digestability	Chemical reuse efficiency is low; low hemicellulose monosaccharides production
Steam explosion	Furans, acetate and other acids	Increasing cellulose digestibility; low environmental impact	Incomplete disruption of lignin-carbohydrate complex; inhibitors
Organosolv	lignin degradation products	High efficiency for lignin degradation	High operational costs
Ozonolysis	Short-chain carboxylic acids	Low furan aldehydes; some lignin degradation; ambient temperature and pressure	Highly reactive and flammable; high energy demand and cost
AFEX	Small amounts of inhibitors; amide-containing chemicals	Decrystalization of cellulose; lignin removal	Low hemicellulose monosaccharides production; high operational cost
Ionic liquid	minimum inhibitors	Reducing cellulose crystallinity and lignin content	High chemical costs



Representative sugar composition (% dry weight) of lignocellulose

Glucose	Xylose	Arabinose + mannose	galactose	total
43.8 ± 1.4	23.4 ± 2.1	5.6 ± 0.7	3.5 ± 0.4	76.3 ± 2.7

Figure 1.1. Composition of lignocellulose. The approximate lignocellulose composition is given as a percentage of total dry weight. The major carbon monomers of the main polymeric components for typical lignocelluloses are underlined in a white box. The representative sugar composition shown in the table was obtained from a sugarcane bagasse sample (14). Note that the composition of soft wood materials such as gymnosperm trees is usually different from this graph with higher lignin content and lower xylose content in the hemicellulose portion. Extracted from our book chapter (4).

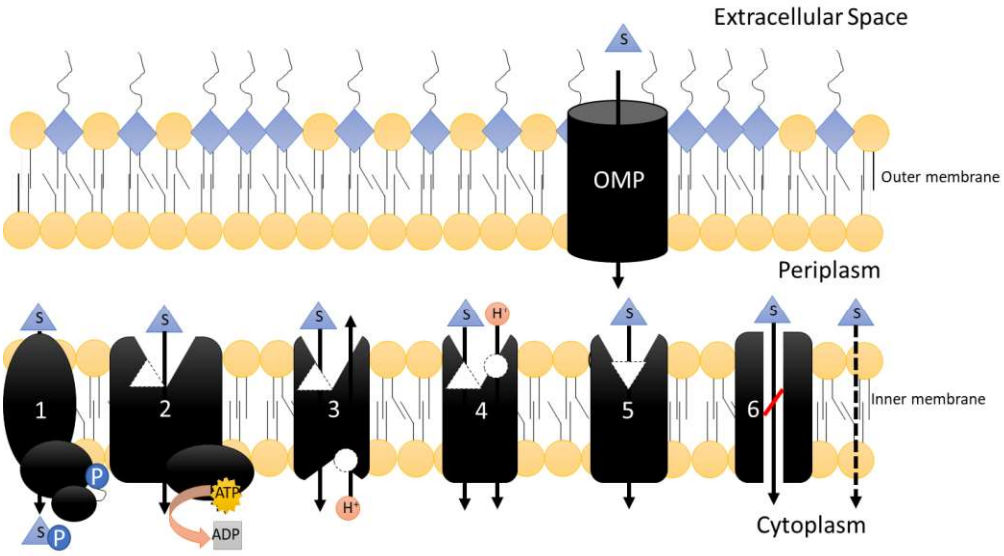
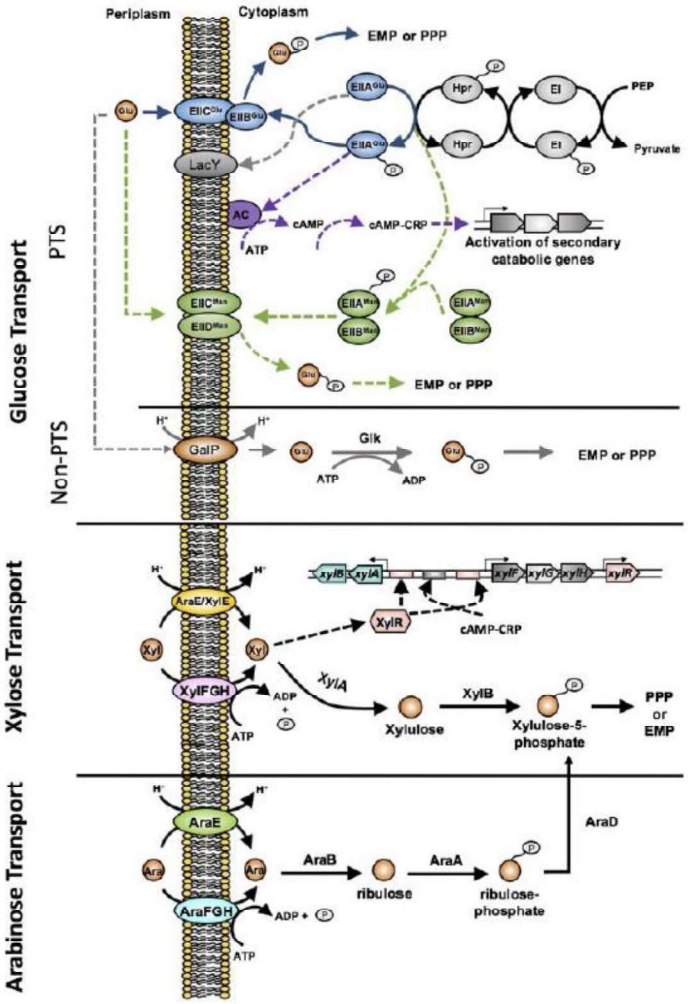
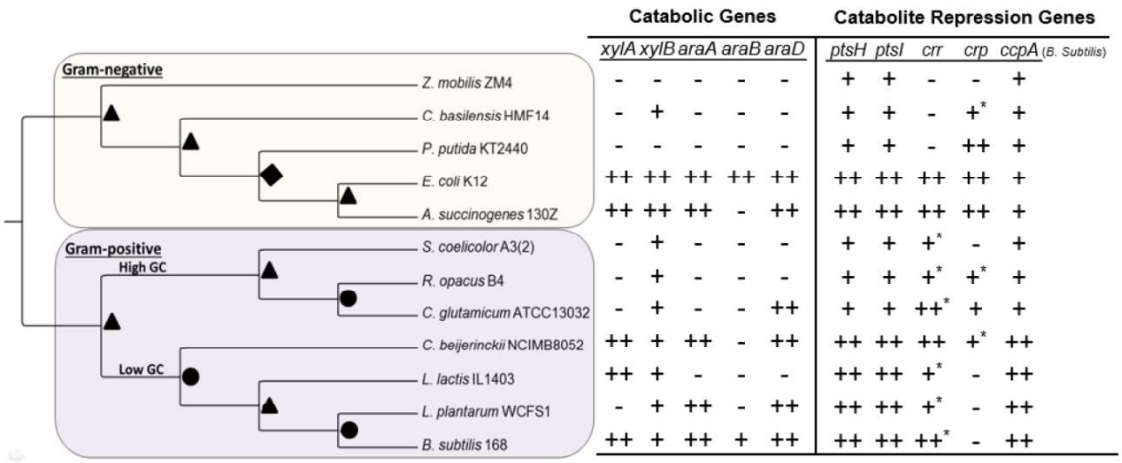


Figure 1.2. Mechanisms of membrane transport. The outer membrane serves as the first layer of selective permeability in which diffusion of chemicals (blue triangle) through the outer membrane is difficult due to lipopolysaccharides (blue diamonds) and thus facilitated diffusion is mainly mediated through outer membrane porins (OMP). Upon entrance into the periplasmic space, cells can catalyze transport using 1) group translocators 2) primary activate transport 3) antiport, 4) symport, 5) uniport, 6) channels or diffusion (dotted line).



Enzyme Abbreviation	Enzyme Name	Gene Name
AC	Adenylate Cyclase	*cyaA*
AraE	Arabinose Proton Symporter	*araE*
AraFGH	Arabinose ABC Transporter ATPase	*araFGH*
AraB	Ribulokinase	*araB*
AraA	Arabinose isomerase	*araA*
AraD	Ribulose-5-phosphate 4-epimerase	*araD*
CRP	Catabolite Repression Protein	*crp*
EI	Enzyme I	*ptsI*
$EIIA^{Glc}$	Enzyme IIA Glucose Specific	*crr*
$EIIBC^{Glc}$	Enzyme IIBC Glucose Specific	*ptsG*
$EIIAB^{Man}$	Enzyme IIAB Mannose Specific	*manX*
$EIICD^{Man}$	Enzyme IICD Mannose Specific	*manYZ*
GalP	Galactose Proton Symporter	*galP*
Glk	Glucokinase	*glk*
HPr	Histidine Protein	*ptsH*
LacY	Lactose permease transporter	*lacY*
XylE	Xylose Proton Symporter	*xylE*
XylFGH	Xylose ABC Transporter ATPase	*xylFGH*
XylA	Xylose Isomerase	*xylA*
XylB	Xylulose kinase	*xylB*
XylR	Transcriptional activator	*xylR*

Figure 1.3. Transport, catabolism and catabolite repression mechanisms of major lignocellulose-derived sugars in *E. coli.* In *E. coli*, there are three known mechanisms for glucose transport including EII^{Glu}-based PTS, EII^{Man}-based PTS and GalP. EII^{Glu}-based PTS is the predominant mechanism for glucose transport. Glucose induced-repression is mainly caused by low levels of cAMP which leads to nonfunctional CRP. Without CRP activation, the transcriptional activation of most secondary sugar catabolism pathways cannot be achieved. Phosphorylated sugar intermediates from glucose, xylose and arabinose catabolism enter PPP or EMP pathways for full degradation. The details of the important components in *E. coli* are summarized in the table. Figure extracted from our book chapter (4).



Catabolic Genes					Catabolite Repression Genes				
xylA	*xylB*	*araA*	*araB*	*araD*	*ptsH*	*ptsI*	*crr*	*crp*	*ccpA* (*B. Subtilis*)
-	-	-	-	-	+	+	-	-	+
-	+	-	-	-	+	+	-	+*	+
-	-	-	-	-	+	+	-	++	+
++	++	++	++	++	++	++	++	++	+
++	++	++	-	++	++	++	++	++	+
-	+	-	-	-	+	+	+*	-	+
-	+	-	-	-	+	+	+*	+*	+
-	+	-	-	++	+	+	++*	+	+
++	+	++	-	++	++	++	++	+*	++
++	+	-	-	-	++	++	+*	-	++
-	+	++	-	++	++	++	+*	-	++
++	+	++	+	++	++	++	++*	-	++

Figure 1.4. Bacterial genes for xylose and arabinose catabolism, and major catabolite repression. Midpoint rooted cladogram of some important industrial prokaryotic biocatalysts is constructed using 16s rRNA sequences from the Greengenes database (96). Sequences were aligned using MAFFT (97), and tree construction was performed using the default PhyML parameters (98). One hundred bootstrap replicates were performed and confidence values of 60-69% (circles) 70-79% (diamonds) 80-89% (squares) and 90-100% (triangles) are listed at each respective node. The presence of potential gene homologs were tested in each biocatalyst using blastp (99). Proteins that exhibited >90% query coverage and identity of 40-100% (++), 25-39% (+) and <25% (-) compared to *E.coli* and *B. subtilis* homologues (only CcpA is from *B. subtilis*) were scored in the table for each respective gene. Proteins that had 80-89% query coverage are marked with an asterisk (*). Figure extracted from our book chapter (4).

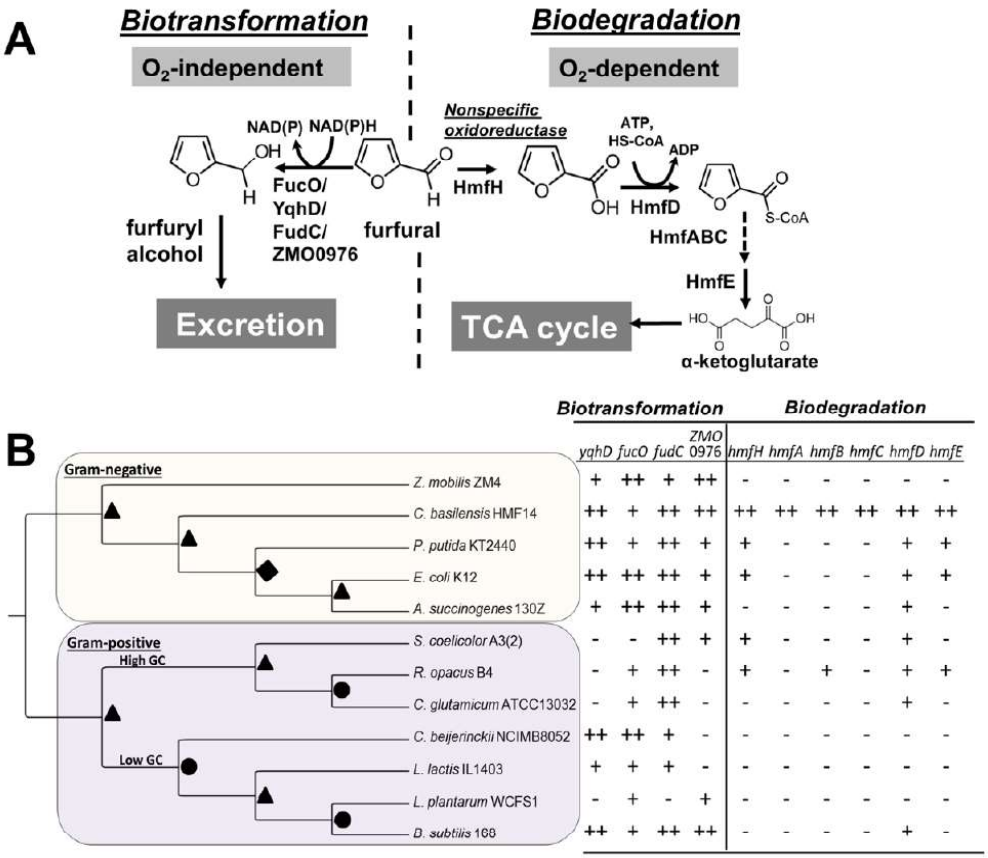


Figure 1.5. Naturally-evolved mechanisms used by different microbes to detoxify furfural. A) Two native furfural detoxification mechanisms are used by different bacteria to relieve the toxicity. B) The presence of potential furfural detoxification protein homologs were tested in each biocatalyst using blastp (99). Proteins that exhibited >90% query coverage and identity of 40-100% (++), 25-39% (+) and <25% (-) were scored in the table for each respective gene. Figure extracted from our book chapter (4).

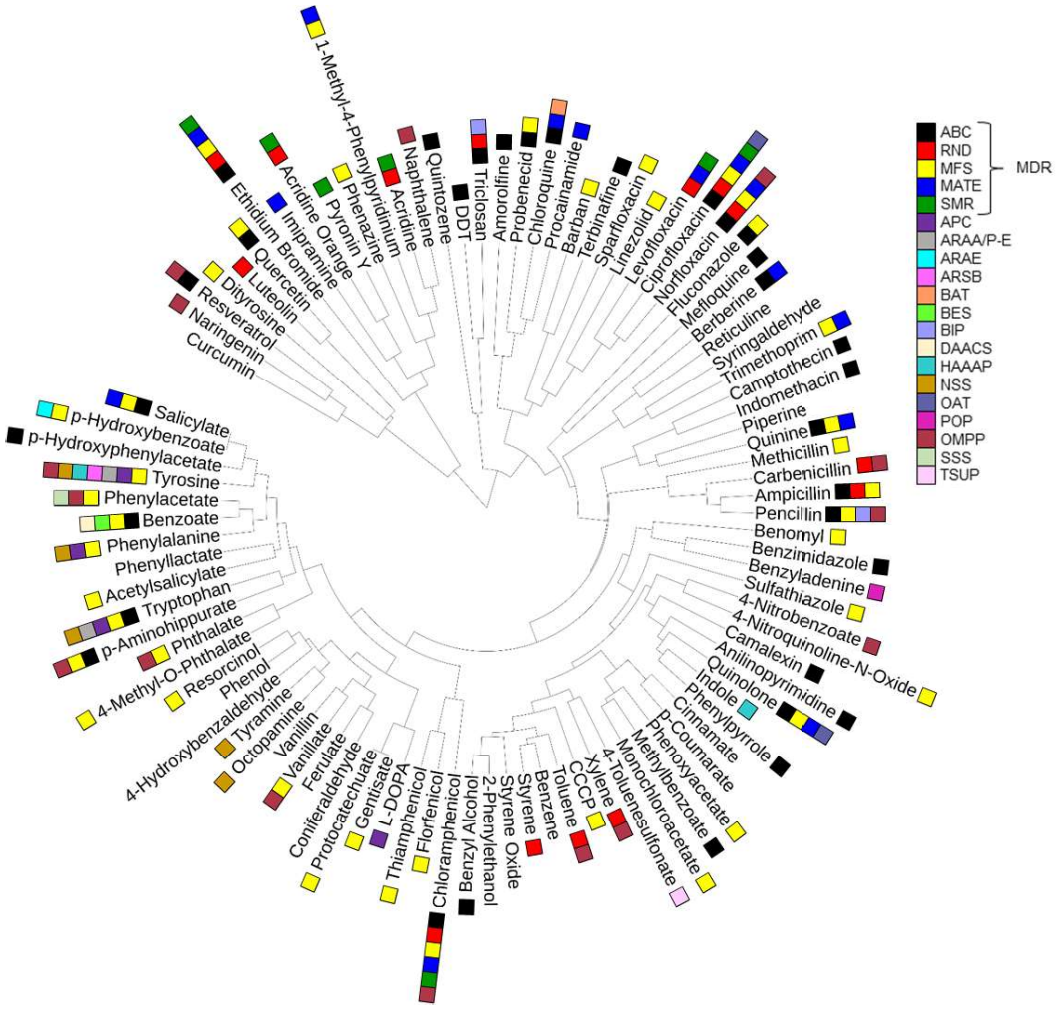


Figure 1.6. Hierarchical clustering of aromatics with a MW <400 g/mole and their respective transporters in the Transporter Classification Database (232) in addition to select aromatics of interest. Chemical similarity was compared using the Signature molecular descriptor (h=0,1) (233) to generate an all-against-all cosine angle distance (234) matrix that was clustered using complete linkage agglomeration with hclust (235) . Transporter family abbreviations: ABC, ATP-Binding Casette; RND, Resistance-Nodulation-Cell Division; MFS, Major Facilitator Superfamily; MATE, Multi Antimicrobial Extrusion; SMR, Small Multidrug Resistance; APC, Amino Acid Polyamine-Organocation; ARAA/P-E, Aromatic Amino Acid/Paraquat Exporter; ARAE, Aromatic Acid Exporter; ARSB, Arsenite-Antimonite Efflux; BAT, 5-TMS Bacterial/Archaeal Transporter; BES, Bestrophin Family; BIP, Bacteriocin Immunity Protein; DAACS, Dicarboxylate/Amino Acid:Cation Symporter; HAAAP, Hydroxy/Aromatic Amino Acid Permease; NSS, Neurotransmitter:Sodium Symporter; OAT, Organo Anion Transporter; POP, Plant Organocation Permease; OMPP, Outer Membrane Pore-forming Protein; SSS, Solute:Sodium Symporter; TSUP, 4-Toluene Sulfonate Uptake Permease; MDR, Multidrug Resistance. Figure extracted from a co-authored review article (in revision).

References

1. Choi DG, Kreikebaum F, Thomas VM, Divan D. 2013. Coordinated EV adoption: double-digit reductions in emissions and fuel use for $40/vehicle-year. Environ Sci Technol 47:10703-7.

2. Service RF. 2007. Cellulosic ethanol. Biofuel researchers prepare to reap a new harvest. Science 315:1488-91.

3. Karl TR, Trenberth KE. 2003. Modern global climate change. Science 302:1719-23.

4. Clark JH, Luque R, Matharu AS. 2012. Green chemistry, biofuels, and biorefinery. Annu Rev Chem Biomol Eng 3:183-207.

5. Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30:279-91.

6. Girio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Lukasik R. 2010. Hemicelluloses for fuel ethanol: A review. Bioresour Technol 101:4775-800.

7. Mishra P, Singh A. 1993. Microbial pentose utilization. Adv Appl Microbiol 39:91-152.

8. Saini JK, Saini R, Tewari L. 2015. Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. 3 Biotech 5:337-353.

9. Garvey M, Klose H, Fischer R, Lambertz C, Commandeur U. 2013. Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. Trends Biotechnol 31:581-93.

10. Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. 2013. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol 135:513-22.

11. Bommarius AS, Sohn M, Kang Y, Lee JH, Realff MJ. 2014. Protein engineering of cellulases. Curr Opin Biotechnol 29:139-45.

12. Suhas, Carrott PJ, Ribeiro Carrott MM. 2007. Lignin--from natural adsorbent to activated carbon: a review. Bioresour Technol 98:2301-12.

13. Eudes A, Liang Y, Mitra P, Loque D. 2014. Lignin bioengineering. Curr Opin Biotechnol 26:189-98.

14. Geddes CC, Peterson JJ, Roslander C, Zacchi G, Mullinnix MT, Shanmugam KT, Ingram LO. 2010. Optimizing the saccharification of sugar cane bagasse using dilute phosphoric acid followed by fungal cellulases. Bioresour Technol 101:1851-7.

15. Hendriks AT, Zeeman G. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresour Technol 100:10-8.

16. Kim SR, Ha SJ, Wei N, Oh EJ, Jin YS. 2012. Simultaneous co-fermentation of mixed sugars: a promising strategy for producing cellulosic ethanol. Trends Biotechnol 30:274-82.

17. Doran-Peterson J, Jangid A, Brandon SK, DeCrescenzo-Henriksen E, Dien B, Ingram LO. 2009. Simultaneous saccharification and fermentation and partial saccharification and co-

fermentation of lignocellulosic biomass for ethanol production. Methods Mol Biol 581:263-80.

18. Kim JH, Block DE, Mills DA. 2010. Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass. Appl Microbiol Biotechnol 88:1077-85.

19. Kolb A, Busby S, Buc H, Garges S, Adhya S. 1993. Transcriptional regulation by cAMP and its receptor protein. Annu Rev Biochem 62:749-95.

20. Deutscher J. 2008. The mechanisms of carbon catabolite repression in bacteria. Curr Opin Microbiol 11:87-93.

21. Gorke B, Stulke J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol 6:613-24.

22. Gonzalez JE, Long CP, Antoniewicz MR. 2017. Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by (13)C metabolic flux analysis. Metab Eng 39:9-18.

23. Khankal R, Chin JW, Cirino PC. 2008. Role of xylose transporters in xylitol production from engineered *Escherichia coli*. J Biotechnol 134:246-52.

24. Zaldivar J, Martinez A, Ingram LO. 1999. Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng 65:24-33.

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

26. Geddes CC, Nieves IU, Ingram LO. 2011. Advances in ethanol production. Curr Opin Biotechnol 22:312-9.

27. Martinez A, Rodriguez ME, York SW, Preston JF, Ingram LO. 2000. Effects of Ca(OH)(2) treatments ("overliming") on the composition and toxicity of bagasse hemicellulose hydrolysates. Biotechnol Bioeng 69:526-36.

28. Geddes R, Shanmugam KT, Ingram LO. 2015. Combining treatments to improve the fermentation of sugarcane bagasse hydrolysates by ethanologenic Escherichia coli LY180. Bioresour Technol 189:15-22.

29. Beveridge TJ, Graham LL. 1991. Surface layers of bacteria. Microbiol Rev 55:684-705.

30. Nikaido H. 2003. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev 67:593-656.

31. Silhavy TJ, Kahne D, Walker S. 2010. The bacterial cell envelope. Cold Spring Harb Perspect Biol 2:a000414.

32. Neuhaus FC, Baddiley J. 2003. A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. Microbiol Mol Biol Rev 67:686-723.

33. Koebnik R, Locher KP, Van Gelder P. 2000. Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol Microbiol 37:239-53.

34. Dowhan W. 1997. Molecular basis for membrane phospholipid diversity: why are there so many lipids? Annu Rev Biochem 66:199-232.

35. Delcour AH. 2009. Outer membrane permeability and antibiotic resistance. Biochim Biophys Acta 1794:808-16.

36. Vaara M, Plachy WZ, Nikaido H. 1990. Partitioning of hydrophobic probes into lipopolysaccharide bilayers. Biochim Biophys Acta 1024:152-8.

37. Nikaido H, Rosenberg EY, Foulds J. 1983. Porin channels in Escherichia coli: studies with beta-lactams in intact cells. J Bacteriol 153:232-40.

38. Nikaido H, Rosenberg EY. 1981. Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of Escherichia coli. J Gen Physiol 77:121-35.

39. Cowan SW, Schirmer T, Rummel G, Steiert M, Ghosh R, Pauptit RA, Jansonius JN, Rosenbusch JP. 1992. Crystal structures explain functional properties of two E. coli porins. Nature 358:727-33.

40. Luckey M, Nikaido H. 1980. Specificity of diffusion channels produced by lambda phage receptor protein of Escherichia coli. Proc Natl Acad Sci U S A 77:167-71.

41. Kahng HY, Byrne AM, Olsen RH, Kukor JJ. 2000. Characterization and role of tbuX in utilization of toluene by Ralstonia pickettii PKO1. J Bacteriol 182:1232-42.

42. Wang Y, Rawlings M, Gibson DT, Labbe D, Bergeron H, Brousseau R, Lau PC. 1995. Identification of a membrane protein and a truncated LysR-type regulator associated with the toluene degradation pathway in Pseudomonas putida F1. Mol Gen Genet 246:570-9.

43. Hearn EM, Patel DR, van den Berg B. 2008. Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation. Proc Natl Acad Sci U S A 105:8601-6.

44. Kurniawan Y, Venkataramanan KP, Scholz C, Bothun GD. 2012. n-Butanol partitioning and phase behavior in DPPC/DOPC membranes. J Phys Chem B 116:5919-24.

45. Sikkema J, de Bont JA, Poolman B. 1994. Interactions of cyclic hydrocarbons with biological membranes. J Biol Chem 269:8022-8.

46. Sikkema J, de Bont JA, Poolman B. 1995. Mechanisms of membrane toxicity of hydrocarbons. Microbiol Rev 59:201-22.

47. Zaldivar J, Martinez A, Ingram LO. 2000. Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic Escherichia coli. Biotechnol Bioeng 68:524-30.

48. Cornell BA, Separovic F. 1983. Membrane thickness and acyl chain length. Biochim Biophys Acta 733:189-93.

49. Juroszek JR, Feuillat M, Charpentier C. 1987. Effect of ethanol on the glucose-induced movements of protons across the plasma membrane of Saccharomyces cerevisiae NCYC 431. Can J Microbiol 33:93-7.

50. Dombek KM, Ingram LO. 1984. Effects of ethanol on the Escherichia coli plasma membrane. J Bacteriol 157:233-9.

51. Booth IR. 1985. Regulation of cytoplasmic pH in bacteria. Microbiol Rev 49:359-78.

52. Jenkins RO, Stephens GM, Dalton H. 1987. Production of toluene cis-glycol by Pseudomonas putida in glucose feb-batch culture. Biotechnol Bioeng 29:873-83.

53. Calamita G. 2000. The Escherichia coli aquaporin-Z water channel. Mol Microbiol 37:254-62.

54. Borgnia MJ, Agre P. 2001. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. Proc Natl Acad Sci U S A 98:2888-93.

55. Catterall WA. 1995. Structure and function of voltage-gated ion channels. Annu Rev Biochem 64:493-531.

56. Moussatova A, Kandt C, O'Mara ML, Tieleman DP. 2008. ATP-binding cassette transporters in Escherichia coli. Biochim Biophys Acta 1778:1757-71.

57. Wang B, Dukarevich M, Sun EI, Yen MR, Saier MH, Jr. 2009. Membrane porters of ATP-binding cassette transport systems are polyphyletic. J Membr Biol 231:1-10.

58. Linton KJ, Higgins CF. 1998. The Escherichia coli ATP-binding cassette (ABC) proteins. Mol Microbiol 28:5-13.

59. Zhou Z, White KA, Polissi A, Georgopoulos C, Raetz CR. 1998. Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. J Biol Chem 273:12466-75.

60. Raetz CR, Whitfield C. 2002. Lipopolysaccharide endotoxins. Annu Rev Biochem 71:635-700.

61. Boudker O, Verdon G. 2010. Structural perspectives on secondary active transporters. Trends Pharmacol Sci 31:418-26.

62. Shi Y. 2013. Common folds and transport mechanisms of secondary active transporters. Annu Rev Biophys 42:51-72.

63. Higgins CF. 2007. Multiple molecular mechanisms for multidrug resistance transporters. Nature 446:749-57.

64. Saier MH, Jr., Paulsen IT. 2001. Phylogeny of multidrug transporters. Semin Cell Dev Biol 12:205-13.

65. Paulsen IT, Brown MH, Skurray RA. 1996. Proton-dependent multidrug efflux systems. Microbiol Rev 60:575-608.

66. Konings WN. 2006. Microbial transport: adaptations to natural environments. Antonie Van Leeuwenhoek 90:325-42.

67. Jiang W, Hermolin J, Fillingame RH. 2001. The preferred stoichiometry of c subunits in the rotary motor sector of Escherichia coli ATP synthase is 10. Proc Natl Acad Sci U S A 98:4966-71.

68. Otto R, Sonnenberg AS, Veldkamp H, Konings WN. 1980. Generation of an electrochemical proton gradient in Streptococcus cremoris by lactate efflux. Proc Natl Acad Sci U S A 77:5502-6.

69. Poolman B, Molenaar D, Smid EJ, Ubbink T, Abee T, Renault PP, Konings WN. 1991. Malolactic fermentation: electrogenic malate uptake and malate/lactate antiport generate metabolic energy. J Bacteriol 173:6030-7.

70. Yan N. 2015. Structural Biology of the Major Facilitator Superfamily Transporters. Annu Rev Biophys 44:257-83.

71. Gorga FR, Lienhard GE. 1981. Equilibria and kinetics of ligand binding to the human erythrocyte glucose transporter. Evidence for an alternating conformation model for transport. Biochemistry 20:5108-13.

72. Parker C, Barnell WO, Snoep JL, Ingram LO, Conway T. 1995. Characterization of the Zymomonas mobilis glucose facilitator gene product (glf) in recombinant Escherichia coli: examination of transport mechanism, kinetics and the role of glucokinase in glucose transport. Mol Microbiol 15:795-802.

73. Gosset G. 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb Cell Fact 4:14.

74. Chen LQ, Cheung LS, Feng L, Tanner W, Frommer WB. 2015. Transport of sugars. Annu Rev Biochem 84:865-94.

75. Stulke J, Hillen W. 1999. Carbon catabolite repression in bacteria. Curr Opin Microbiol 2:195-201.

76. Kotrba P, Inui M, Yukawa H. 2001. Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. J Biosci Bioeng 92:502-17.

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

78. Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57:543-94.

79. Stock JB, Waygood EB, Meadow ND, Postma PW, Roseman S. 1982. Sugar transport by the bacterial phosphotransferase system. The glucose receptors of the Salmonella typhimurium phosphotransferase system. J Biol Chem 257:14543-52.

80. Misset O, Blaauw M, Postma PW, Robillard GT. 1983. Bacterial phosphoenolpyruvate-dependent phosphotransferase system. Mechanism of the transmembrane sugar translocation and phosphorylation. Biochemistry 22:6163-70.

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

82. McDonald TP, Walmsley AR, Henderson PJ. 1997. Asparagine 394 in putative helix 11 of the galactose-H+ symport protein (GalP) from Escherichia coli is associated with the internal binding site for cytochalasin B and sugar. J Biol Chem 272:15189-99.

83. Horazdovsky BF, Hogg RW. 1987. High-affinity L-arabinose transport operon. Gene product expression and mRNAs. J Mol Biol 197:27-35.

84. Sumiya M, Davis EO, Packman LC, McDonald TP, Henderson PJ. 1995. Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in Escherichia coli. Receptors Channels 3:117-28.

85. Jojima T, Omumasaba CA, Inui M, Yukawa H. 2010. Sugar transporters in efficient utilization of mixed sugar substrates: current knowledge and outlook. Appl Microbiol Biotechnol 85:471-80.

86. Davis EO, Henderson PJ. 1987. The cloning and DNA sequence of the gene xylE for xylose-proton symport in Escherichia coli K12. J Biol Chem 262:13928-32.

87. Maiden MC, Jones-Mortimer MC, Henderson PJ. 1988. The cloning, DNA sequence, and overexpression of the gene araE coding for arabinose-proton symport in Escherichia coli K12. J Biol Chem 263:8003-10.

88. Hasona A, Kim Y, Healy FG, Ingram LO, Shanmugam KT. 2004. Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of Escherichia coli on xylose. J Bacteriol 186:7593-600.

89. Snoep JL, Arfman N, Yomano LP, Fliege RK, Conway T, Ingram LO. 1994. Reconstruction of glucose uptake and phosphorylation in a glucose-negative mutant of Escherichia coli by using Zymomonas mobilis genes encoding the glucose facilitator protein and glucokinase. J Bacteriol 176:2133-5.

90. Koirala S, Wang X, Rao CV. 2016. Reciprocal Regulation of l-Arabinose and d-Xylose Metabolism in Escherichia coli. J Bacteriol 198:386-93.

91. Henderson PJ. 1990. Proton-linked sugar transport systems in bacteria. J Bioenerg Biomembr 22:525-69.

92. Daruwalla KR, Paxton AT, Henderson PJ. 1981. Energization of the transport systems for arabinose and comparison with galactose transport in Escherichia coli. Biochem J 200:611-27.

93. Nieves LM, Panyon LA, Wang X. 2015. Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. Front Bioeng Biotechnol 3:17.

94. Schleif R. 2000. Regulation of the L-arabinose operon of Escherichia coli. Trends Genet 16:559-65.

95. Jeffries TW. 1983. Utilization of xylose by bacteria, yeasts, and fungi. Adv Biochem Eng Biotechnol 27:1-32.

96. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069-72.

97. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-66.

98. Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557-9.

99. Gish W, States DJ. 1993. Identification of protein coding regions by database similarity search. Nat Genet 3:266-72.

100. Zhang M, Eddy C, Deanda K, Finkelstein M, Picataggio S. 1995. Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic Zymomonas mobilis. Science 267:240-3.

101. Kawaguchi H, Vertes AA, Okino S, Inui M, Yukawa H. 2006. Engineering of a xylose metabolic pathway in Corynebacterium glutamicum. Appl Environ Microbiol 72:3418-28.

102. Dien BS, Nichols NN, Bothast RJ. 2002. Fermentation of sugar mixtures using Escherichia coli catabolite repression mutants engineered for production of L-lactic acid. J Ind Microbiol Biotechnol 29:221-7.

103. 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 280:15921-7.

104. Blencke HM, Homuth G, Ludwig H, Mader U, Hecker M, Stulke J. 2003. Transcriptional profiling of gene expression in response to glucose in Bacillus subtilis: regulation of the central metabolic pathways. Metab Eng 5:133-49.

105. Moreno MS, Schneider BL, Maile RR, Weyler W, Saier MH, Jr. 2001. Catabolite repression mediated by the CcpA protein in Bacillus subtilis: novel modes of regulation revealed by whole-genome analyses. Mol Microbiol 39:1366-81.

106. Yoshida K, Kobayashi K, Miwa Y, Kang CM, Matsunaga M, Yamaguchi H, Tojo S, Yamamoto M, Nishi R, Ogasawara N, Nakayama T, Fujita Y. 2001. Combined transcriptome and proteome analysis as a powerful approach to study genes under glucose repression in Bacillus subtilis. Nucleic Acids Res 29:683-92.

107. Escalante A, Salinas Cervantes A, Gosset G, Bolivar F. 2012. Current knowledge of the Escherichia coli phosphoenolpyruvate-carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation. Appl Microbiol Biotechnol 94:1483-94.

108. Nelson SO, Wright JK, Postma PW. 1983. The mechanism of inducer exclusion. Direct interaction between purified III of the phosphoenolpyruvate:sugar phosphotransferase system and the lactose carrier of Escherichia coli. EMBO J 2:715-20.

109. Osumi T, Saier MH, Jr. 1982. Regulation of lactose permease activity by the phosphoenolpyruvate:sugar phosphotransferase system: evidence for direct binding of the glucose-specific enzyme III to the lactose permease. Proc Natl Acad Sci U S A 79:1457-61.

110. Malan TP, Kolb A, Buc H, McClure WR. 1984. Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. J Mol Biol 180:881-909.

111. Tagami H, Aiba H. 1998. A common role of CRP in transcription activation: CRP acts transiently to stimulate events leading to open complex formation at a diverse set of promoters. EMBO J 17:1759-67.

112. Geng H, Jiang R. 2015. cAMP receptor protein (CRP)-mediated resistance/tolerance in bacteria: mechanism and utilization in biotechnology. Appl Microbiol Biotechnol 99:4533-43.

113. Dien BS, Nichols NN, Bothast RJ. 2001. Recombinant Escherichia coli engineered for production of L-lactic acid from hexose and pentose sugars. J Ind Microbiol Biotechnol 27:259-64.

114. Flores N, Xiao J, Berry A, Bolivar F, Valle F. 1996. Pathway engineering for the production of aromatic compounds in Escherichia coli. Nat Biotechnol 14:620-3.

115. Hernandez-Montalvo V, Valle F, Bolivar F, Gosset G. 2001. Characterization of sugar mixtures utilization by an Escherichia coli mutant devoid of the phosphotransferase system. Appl Microbiol Biotechnol 57:186-91.

116. Hernandez-Montalvo V, Martinez A, Hernandez-Chavez G, Bolivar F, Valle F, Gosset G. 2003. Expression of galP and glk in a Escherichia coli PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol Bioeng 83:687-94.

117. Balderas-Hernandez VE, Hernandez-Montalvo V, Bolivar F, Gosset G, Martinez A. 2011. Adaptive evolution of Escherichia coli inactivated in the phosphotransferase system operon improves co-utilization of xylose and glucose under anaerobic conditions. Appl Biochem Biotechnol 163:485-96.

118. Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, Wang X. 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc Natl Acad Sci U S A 114:7349-7354.

119. Utrilla J, Licona-Cassani C, Marcellin E, Gosset G, Nielsen LK, Martinez A. 2012. Engineering and adaptive evolution of Escherichia coli for D-lactate fermentation reveals GatC as a xylose transporter. Metab Eng 14:469-76.

120. Zhang X, Jantama K, Moore JC, Jarboe LR, Shanmugam KT, Ingram LO. 2009. Metabolic evolution of energy-conserving pathways for succinate production in Escherichia coli. Proc Natl Acad Sci U S A 106:20180-5.

121. Wang X, Yomano LP, Lee JY, York SW, Zheng H, Mullinnix MT, Shanmugam KT, Ingram LO. 2013. Engineering furfural tolerance in Escherichia coli improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 110:4021-6.

122. Khunnonkwao P, Jantama SS, Kanchanatawee S, Jantama K. 2018. Re-engineering Escherichia coli KJ122 to enhance the utilization of xylose and xylose/glucose mixture for efficient succinate production in mineral salt medium. Appl Microbiol Biotechnol 102:127-141.

123. Sawisit A, Jantama K, Zheng H, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2015. Mutation in galP improved fermentation of mixed sugars to succinate using engineered Escherichia coli AS1600a and AM1 mineral salts medium. Bioresour Technol 193:433-41.

124. Tang J, Zhu X, Lu J, Liu P, Xu H, Tan Z, Zhang X. 2013. Recruiting alternative glucose utilization pathways for improving succinate production. Appl Microbiol Biotechnol 97:2513-20.

125. Chen T, Zhang J, Liang L, Yang R, Lin Z. 2009. An in vivo, label-free quick assay for xylose transport in Escherichia coli. Anal Biochem 390:63-7.

126. da Costa Sousa L, Chundawat SP, Balan V, Dale BE. 2009. 'Cradle-to-grave' assessment of existing lignocellulose pretreatment technologies. Curr Opin Biotechnol 20:339-47.

127. Lloyd TA, Wyman CE. 2005. Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. Bioresour Technol 96:1967-77.

128. Geddes CC, Peterson JJ, Mullinnix MT, Svoronos SA, Shanmugam KT, Ingram LO. 2010. Optimizing cellulase usage for improved mixing and rheological properties of acid-pretreated sugarcane bagasse. Bioresour Technol 101:9128-36.

129. Karagoz P, Rocha IV, Ozkan M, Angelidaki I. 2012. Alkaline peroxide pretreatment of rapeseed straw for enhancing bioethanol production by Same Vessel Saccharification and Co-Fermentation. Bioresour Technol 104:349-57.

130. Banerjee G, Car S, Scott-Craig JS, Hodge DB, Walton JD. 2011. Alkaline peroxide pretreatment of corn stover: effects of biomass, peroxide, and enzyme loading and composition on yields of glucose and xylose. Biotechnol Biofuels 4:16.

131. Brownell HH, Yu EK, Saddler JN. 1986. Steam-explosion pretreatment of wood: Effect of chip size, acid, moisture content and pressure drop. Biotechnol Bioeng 28:792-801.

132. Taniguchi M, Takahashi D, Watanabe D, Sakai K, Hoshino K, Kouya T, Tanaka T. 2010. Effect of steam explosion pretreatment on treatment with Pleurotus ostreatus for the enzymatic hydrolysis of rice straw. J Biosci Bioeng 110:449-52.

133. Yu H, Xing Y, Lei F, Liu Z, Liu Z, Jiang J. 2014. Improvement of the enzymatic hydrolysis of furfural residues by pretreatment with combined green liquor and ethanol organosolv. Bioresour Technol 167:46-52.

134. Zhao X, Cheng K, Liu D. 2009. Organosolv pretreatment of lignocellulosic biomass for enzymatic hydrolysis. Appl Microbiol Biotechnol 82:815-27.

135. Travaini R, Otero MD, Coca M, Da-Silva R, Bolado S. 2013. Sugarcane bagasse ozonolysis pretreatment: effect on enzymatic digestibility and inhibitory compound formation. Bioresour Technol 133:332-9.

136. Travaini R, Martin-Juarez J, Lorenzo-Hernando A, Bolado-Rodriguez S. 2016. Ozonolysis: An advantageous pretreatment for lignocellulosic biomass revisited. Bioresour Technol 199:2-12.

137. Lau MW, Dale BE. 2009. Cellulosic ethanol production from AFEX-treated corn stover using Saccharomyces cerevisiae 424A(LNH-ST). Proc Natl Acad Sci U S A 106:1368-73.

138. Lau MW, Dale BE, Balan V. 2008. Ethanolic fermentation of hydrolysates from ammonia fiber expansion (AFEX) treated corn stover and distillers grain without detoxification and external nutrient supplementation. Biotechnol Bioeng 99:529-39.

139. Hou Q, Ju M, Li W, Liu L, Chen Y, Yang Q. 2017. Pretreatment of Lignocellulosic Biomass with Ionic Liquids and Ionic Liquid-Based Solvent Systems. Molecules 22.

140. Nguyen TA, Kim KR, Han SJ, Cho HY, Kim JW, Park SM, Park JC, Sim SJ. 2010. Pretreatment of rice straw with ammonia and ionic liquid for lignocellulose conversion to fermentable sugars. Bioresour Technol 101:7432-8.

141. Ninomiya K, Omote S, Ogino C, Kuroda K, Noguchi M, Endo T, Kakuchi R, Shimizu N, Takahashi K. 2015. Saccharification and ethanol fermentation from cholinium ionic liquid-pretreated bagasse with a different number of post-pretreatment washings. Bioresour Technol 189:203-9.

142. van der Pol EC, Bakker RR, Baets P, Eggink G. 2014. By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. Appl Microbiol Biotechnol 98:9579-93.

143. Zaldivar J, Ingram LO. 1999. Effect of organic acids on the growth and fermentation of ethanologenic Escherichia coli LY01. Biotechnol Bioeng 66:203-10.

144. Gorsich SW, Dien BS, Nichols NN, Slininger PJ, Liu ZL, Skory CD. 2006. Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in Saccharomyces cerevisiae. Appl Microbiol Biotechnol 71:339-49.

145. Liu ZL, Slininger PJ, Dien BS, Berhow MA, Kurtzman CP, 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.

146. Ko JK, Um Y, Park YC, Seo JH, Kim KH. 2015. Compounds inhibiting the bioconversion of hydrothermally pretreated lignocellulose. Appl Microbiol Biotechnol 99:4201-12.

147. Nagle NJ, Elander RT, Newman MM, Rohrback BT, Ruiz RO, Torget RW. 2002. Efficacy of a hot washing process for pretreated yellow poplar to enhance bioethanol production. Biotechnol Prog 18:734-8.

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

149. Martinez A, Rodriguez ME, Wells ML, York SW, Preston JF, Ingram LO. 2001. Detoxification of dilute acid hydrolysates of lignocellulose with lime. Biotechnol Prog 17:287-93.

150. Grzenia DL, Schell DJ, Wickramasinghe SR. 2012. Membrane extraction for detoxification of biomass hydrolysates. Bioresour Technol 111:248-54.

151. Almeida JR, Bertilsson M, Gorwa-Grauslund MF, Gorsich S, Liden G. 2009. Metabolic effects of furaldehydes and impacts on biotechnological processes. Appl Microbiol Biotechnol 82:625-38.

152. Wang X, Miller EN, Yomano LP, Shanmugam KT, Ingram LO. 2012. Increased furan tolerance in Escherichia coli due to a cryptic ucpA gene. Appl Environ Microbiol 78:2452-5.

153. Turner PC, Miller EN, Jarboe LR, Baggett CL, Shanmugam KT, Ingram LO. 2011. YqhC regulates transcription of the adjacent Escherichia coli genes yqhD and dkgA that are involved in furfural tolerance. J Ind Microbiol Biotechnol 38:431-9.

154. Miller EN, Turner PC, Jarboe LR, Ingram LO. 2010. Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing Escherichia coli LY180. Biotechnol Lett 32:661-7.

155. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2009. Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic Escherichia coli. Appl Environ Microbiol 75:4315-23.

156. Wang X, Miller EN, Yomano LP, Zhang X, Shanmugam KT, Ingram LO. 2011. Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in Escherichia coli strains engineered for the production of ethanol and lactate. Appl Environ Microbiol 77:5132-40.

157. Tsuge Y, Kudou M, Kawaguchi H, Ishii J, Hasunuma T, Kondo A. 2016. FudC, a protein primarily responsible for furfural detoxification in Corynebacterium glutamicum. Appl Microbiol Biotechnol 100:2685-92.

158. Agrawal M, Chen RR. 2011. Discovery and characterization of a xylose reductase from Zymomonas mobilis ZM4. Biotechnol Lett 33:2127-33.

159. 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 U S A 107:4919-24.

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

161. Zheng H, Wang X, Yomano LP, Geddes RD, Shanmugam KT, Ingram LO. 2013. Improving Escherichia coli FucO for furfural tolerance by saturation mutagenesis of individual amino acid positions. Appl Environ Microbiol 79:3202-8.

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

163. Miller EN, Jarboe LR, Turner PC, Pharkya P, Yomano LP, York SW, Nunn D, Shanmugam KT, Ingram LO. 2009. Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. Appl Environ Microbiol 75:6132-41.

164. Zheng H, Wang X, Yomano LP, Shanmugam KT, Ingram LO. 2012. Increase in furfural tolerance in ethanologenic *Escherichia coli* LY180 by plasmid-based expression of thyA. Appl Environ Microbiol 78:4346-52.

165. Forsberg KJ, Patel S, Witt E, Wang B, Ellison TD, Dantas G. 2016. Identification of Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes. Appl Environ Microbiol 82:528-37.

166. Glebes TY, Sandoval NR, Reeder PJ, Schilling KD, Zhang M, Gill RT. 2014. Genome-wide mapping of furfural tolerance genes in *Escherichia coli*. PLoS One 9:e87540.

167. Glebes TY, Sandoval NR, Gillis JH, Gill RT. 2015. Comparison of genome-wide selection strategies to identify furfural tolerance genes in *Escherichia coli*. Biotechnol Bioeng 112:129-40.

168. Shui ZX, Qin H, Wu B, Ruan ZY, Wang LS, Tan FR, Wang JL, Tang XY, Dai LC, Hu GQ, He MX. 2015. Adaptive laboratory evolution of ethanologenic *Zymomonas mobilis* strain tolerant to furfural and acetic acid inhibitors. Appl Microbiol Biotechnol 99:5739-48.

169. Mohagheghi A, Linger JG, Yang S, Smith H, Dowe N, Zhang M, Pienkos PT. 2015. Improving a recombinant Zymomonas mobilis strain 8b through continuous adaptation on dilute acid pretreated corn stover hydrolysate. Biotechnol Biofuels 8:55.

170. Geddes CC, Mullinnix MT, Nieves IU, Peterson JJ, Hoffman RW, York SW, Yomano LP, Miller EN, Shanmugam KT, Ingram LO. 2011. Simplified process for ethanol production from sugarcane bagasse using hydrolysate-resistant *Escherichia coli* strain MM160. Bioresour Technol 102:2702-11.

171. Geddes RD, Wang X, Yomano LP, Miller EN, Zheng H, Shanmugam KT, Ingram LO. 2014. Polyamine transporters and polyamines increase furfural tolerance during xylose fermentation with ethanologenic *Escherichia coli* strain LY180. Appl Environ Microbiol 80:5955-64.

172. Takahashi CM, Takahashi DF, Carvalhal ML, Alterthum F. 1999. Effects of acetate on the growth and fermentation performance of *Escherichia coli* KO11. Appl Biochem Biotechnol 81:193-203.

173. Fernandez-Sandoval MT, Huerta-Beristain G, Trujillo-Martinez B, Bustos P, Gonzalez 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.

174. 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.

175. 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.

176. Yang S, Land ML, Klingeman DM, Pelletier DA, Lu TY, Martin SL, Guo HB, Smith JC, Brown SD. 2010. Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A 107:10395-400.

177. Yi X, Gu H, Gao Q, Liu ZL, Bao J. 2015. Transcriptome analysis of *Zymomonas mobilis* ZM4 reveals mechanisms of tolerance and detoxification of phenolic aldehyde inhibitors from lignocellulose pretreatment. Biotechnol Biofuels 8:153.

178. Sandoval NR, Papoutsakis ET. 2016. Engineering membrane and cell-wall programs for tolerance to toxic chemicals: Beyond solo genes. Curr Opin Microbiol 33:56-66.

179. Junker F, Ramos JL. 1999. Involvement of the cis/trans isomerase Cti in solvent resistance of *Pseudomonas putida* DOT-T1E. J Bacteriol 181:5693-700.

180. Tan Z, Yoon JM, Nielsen DR, Shanks JV, Jarboe LR. 2016. Membrane engineering via trans unsaturated fatty acids production improves *Escherichia coli* robustness and production of biorenewables. Metab Eng 35:105-113.

181. Bui le M, Lee JY, Geraldi A, Rahman Z, Lee JH, Kim SC. 2015. Improved n-butanol tolerance in *Escherichia coli* by controlling membrane related functions. J Biotechnol 204:33-44.

182. Tan Z, Khakbaz P, Chen Y, Lombardo J, Yoon JM, Shanks JV, Klauda JB, Jarboe LR. 2017. Engineering *Escherichia coli* membrane phospholipid head distribution improves tolerance and production of biorenewables. Metab Eng 44:1-12.

183. Ruegg TL, Kim EM, Simmons BA, Keasling JD, Singer SW, Lee TS, Thelen MP. 2014. An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. Nat Commun 5:3490.

184. Si T, Chao R, Min Y, Wu Y, Ren W, Zhao H. 2017. Automated multiplex genome-scale engineering in yeast. Nat Commun 8:15187.

185. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894-898.

186. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol 8:536-46.

187. Kind S, Kreye S, Wittmann C. 2011. Metabolic engineering of cellular transport for overproduction of the platform chemical 1,5-diaminopentane in Corynebacterium glutamicum. Metab Eng 13:617-27.

188. Jones CM, Hernandez Lozada NJ, Pfleger BF. 2015. Efflux systems in bacteria and their metabolic engineering applications. Appl Microbiol Biotechnol 99:9381-93.

189. Brown SH, Bashkirova L, Berka R, Chandler T, Doty T, McCall K, McCulloch M, McFarland S, Thompson S, Yaver D, Berry A. 2013. Metabolic engineering of Aspergillus oryzae NRRL 3488 for increased production of L-malic acid. Appl Microbiol Biotechnol 97:8903-12.

190. Boyarskiy S, Tullman-Ercek D. 2015. Getting pumped: membrane efflux transporters for enhanced biomolecule production. Curr Opin Chem Biol 28:15-9.

191. Fisher MA, Boyarskiy S, Yamada MR, Kong N, Bauer S, Tullman-Ercek D. 2014. Enhancing tolerance to short-chain alcohols by engineering the *Escherichia coli* AcrB efflux pump to secrete the non-native substrate n-butanol. ACS Synth Biol 3:30-40.

192. Martinez JL, Sanchez MB, Martinez-Solano L, Hernandez A, Garmendia L, Fajardo A, Alvarez-Ortega C. 2009. Functional role of bacterial multidrug efflux pumps in microbial natural ecosystems. FEMS Microbiol Rev 33:430-49.

193. Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, Keasling JD, Hadi MZ, Mukhopadhyay A. 2011. Engineering microbial biofuel tolerance and export using efflux pumps. Mol Syst Biol 7:487.

194. Lennen RM, Politz MG, Kruziki MA, Pfleger BF. 2013. Identification of transport proteins involved in free fatty acid efflux in *Escherichia coli*. J Bacteriol 195:135-44.

195. Mingardon F, Clement C, Hirano K, Nhan M, Luning EG, Chanal A, Mukhopadhyay A. 2015. Improving olefin tolerance and production in E. coli using native and evolved AcrB. Biotechnol Bioeng 112:879-88.

196. Shah AA, Wang C, Chung YR, Kim JY, Choi ES, Kim SW. 2013. Enhancement of geraniol resistance of *Escherichia coli* by MarA overexpression. J Biosci Bioeng 115:253-8.

197. Du D, Wang Z, James NR, Voss JE, Klimont E, Ohene-Agyei T, Venter H, Chiu W, Luisi BF. 2014. Structure of the AcrAB-TolC multidrug efflux pump. Nature 509:512-5.

198. Wang JF, Xiong ZQ, Li SY, Wang Y. 2013. Enhancing isoprenoid production through systematically assembling and modulating efflux pumps in *Escherichia coli*. Appl Microbiol Biotechnol 97:8057-67.

199. Ma D, Alberti M, Lynch C, Nikaido H, Hearst JE. 1996. The local repressor AcrR plays a modulating role in the regulation of *acrAB* genes of *Escherichia coli* by global stress signals. Mol Microbiol 19:101-12.

200. White DG, Goldman JD, Demple B, Levy SB. 1997. Role of the acrAB locus in organic solvent tolerance mediated by expression of marA, soxS, or robA in *Escherichia coli*. J Bacteriol 179:6122-6.

201. Martin RG, Gillette WK, Rhee S, Rosner JL. 1999. Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in *Escherichia coli*: sequence, orientation and spatial relationship to the core promoter. Mol Microbiol 34:431-41.

202. Watanabe R, Doukyu N. 2012. Contributions of mutations in acrR and marR genes to organic solvent tolerance in *Escherichia coli*. AMB Express 2:58.

203. Foo JL, Leong SS. 2013. Directed evolution of an E. coli inner membrane transporter for improved efflux of biofuel molecules. Biotechnol Biofuels 6:81.

204. Boyarskiy S, Davis Lopez S, Kong N, Tullman-Ercek D. 2016. Transcriptional feedback regulation of efflux protein expression for increased tolerance to and production of n-butanol. Metab Eng 33:130-137.

205. Park JH, Lee SY. 2008. Towards systems metabolic engineering of microorganisms for amino acid production. Curr Opin Biotechnol 19:454-60.

206. Rodriguez A, Martinez JA, Flores N, Escalante A, Gosset G, Bolivar F. 2014. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. Microb Cell Fact 13:126.

207. Liu Q, Cheng Y, Xie X, Xu Q, Chen N. 2012. Modification of tryptophan transport system and its impact on production of L-tryptophan in *Escherichia coli*. Bioresour Technol 114:549-54.

208. Zhang X, Jantama K, Moore JC, Shanmugam KT, Ingram LO. 2007. Production of L-alanine by metabolically engineered *Escherichia coli*. Appl Microbiol Biotechnol 77:355-66.

209. Oldiges M, Eikmanns BJ, Blombach B. 2014. Application of metabolic engineering for the biotechnological production of L-valine. Appl Microbiol Biotechnol 98:5859-70.

210. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. 2007. Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol 3:149.

211. Chavez-Bejar MI, Lara AR, Lopez H, Hernandez-Chavez G, Martinez A, Ramirez OT, Bolivar F, Gosset G. 2008. Metabolic engineering of *Escherichia coli* for L-tyrosine

production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from *Zymomonas mobilis*. Appl Environ Microbiol 74:3284-90.

212. Takpho N, Watanabe D, Takagi H. 2018. High-level production of valine by expression of the feedback inhibition-insensitive acetohydroxyacid synthase in *Saccharomyces cerevisiae*. Metab Eng 46:60-67.

213. Hori H, Ando T, Isogai E, Yoneyama H, Katsumata R. 2011. Identification of an L-alanine export system in *Escherichia coli* and isolation and characterization of export-deficient mutants. FEMS Microbiol Lett 316:83-9.

214. Hori H, Yoneyama H, Tobe R, Ando T, Isogai E, Katsumata R. 2011. Inducible L-alanine exporter encoded by the novel gene *ygaW* (*alaE*) in *Escherichia coli*. Appl Environ Microbiol 77:4027-34.

215. Yamada S, Awano N, Inubushi K, Maeda E, Nakamori S, Nishino K, Yamaguchi A, Takagi H. 2006. Effect of drug transporter genes on cysteine export and overproduction in *Escherichia coli*. Appl Environ Microbiol 72:4735-42.

216. Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, Kim SC. 2009. Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. Microb Cell Fact 8:2.

217. Nakamura J, Hirano S, Ito H, Wachi M. 2007. Mutations of the *Corynebacterium glutamicum* NCgl1221 gene, encoding a mechanosensitive channel homolog, induce L-glutamic acid production. Appl Environ Microbiol 73:4491-8.

218. Ginesy M, Belotserkovsky J, Enman J, Isaksson L, Rova U. 2015. Metabolic engineering of *Escherichia coli* for enhanced arginine biosynthesis. Microb Cell Fact 14:29.

219. Posfai G, Plunkett G, 3rd, Feher T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR. 2006. Emergent properties of reduced-genome *Escherichia coli*. Science 312:1044-6.

220. Werpy T, Petersen G. 2004. Top value added chemicals from biomass. US Department of Energy, Washington D.C., Maryland.

221. Janausch IG, Zientz E, Tran QH, Kroger A, Unden G. 2002. C4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta 1553:39-56.

222. Zientz E, Bongaerts J, Unden G. 1998. Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (*dcuSR* genes) two-component regulatory system. J Bacteriol 180:5421-5.

223. Jones HM, Gunsalus RP. 1987. Regulation of *Escherichia coli* fumarate reductase (frdABCD) operon expression by respiratory electron acceptors and the fnr gene product. J Bacteriol 169:3340-9.

224. Park SJ, Cotter PA, Gunsalus RP. 1995. Regulation of malate dehydrogenase (*mdh*) gene expression in *Escherichia coli* in response to oxygen, carbon, and heme availability. J Bacteriol 177:6652-6.

225. Golby P, Kelly DJ, Guest JR, Andrews SC. 1998. Transcriptional regulation and organization of the dcuA and dcuB genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli*. J Bacteriol 180:6586-96.

226. Zientz E, Janausch IG, Six S, Unden G. 1999. Functioning of DcuC as the C4-dicarboxylate carrier during glucose fermentation by *Escherichia coli*. J Bacteriol 181:3716-20.

227. Chen J, Zhu X, Tan Z, Xu H, Tang J, Xiao D, Zhang X. 2014. Activating C4-dicarboxylate transporters DcuB and DcuC for improving succinate production. Appl Microbiol Biotechnol 98:2197-205.

228. Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JM, van Dijken JP, Pronk JT, van Maris AJ. 2008. Malic acid production by *Saccharomyces cerevisiae*: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. Appl Environ Microbiol 74:2766-77.

229. van Maris AJ, Winkler AA, Porro D, van Dijken JP, Pronk JT. 2004. Homofermentative lactate production cannot sustain anaerobic growth of engineered Saccharomyces cerevisiae: possible consequence of energy-dependent lactate export. Appl Environ Microbiol 70:2898-905.

230. Foo JL, Jensen HM, Dahl RH, George K, Keasling JD, Lee TS, Leong S, Mukhopadhyay A. 2014. Improving microbial biogasoline production in Escherichia coli using tolerance engineering. MBio 5:e01932.

231. Doshi R, Nguyen T, Chang G. 2013. Transporter-mediated biofuel secretion. Proc Natl Acad Sci U S A 110:7642-7.

232. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res. 44(D1):D372-9.

233. Carbonell P, Carlsson L, Faulon JL. 2013. Stereo signature molecular descriptor. J Chem Inf Model. 53(4):887-97.

234. Gang Q, Shamik S, Yuelong G, Sakti P. 2004 Similarity between Euclidean and cosine angle distance for nearest neighbor queries. ACS Symposium on Applied Computing.

CHAPTER 2

REVERSE ENGINEERING *ESCHERICHIA COLI* FOR CONVERSION OF LIGNOCELLULOSIC SUGARS TO SUCCINATE BY MODULATING THE GENETIC CONTROL OF GALP

Abstract

Microbial conversion of lignocellulosic substrates into valuable products requires biocatalysts to use a mixture of sugars efficiently, especially glucose and xylose. *Escherichia coli* KJ122 has previously been engineered to produce succinate from glucose at high titer (85 g l^{-1}), yield (0.9 g g^{-1}), and productivity (0.8 g l^{-1} h^{-1}), but xylose utilization is surprisingly inefficient under fermentative conditions. To probe the potential underlying mechanism inhibiting xylose fermentation, KJ122 was adapted to enhance xylose utilization multiple times independently in xylose fermentation media. A quick adaptation repeatedly occurred within 20 generations and strains isolated after 60 generations showed efficient succinate production from xylose (17-fold increase). Genome sequencing analysis of the evolved strains revealed that the convergent mutations occurred in the galactose regulon during laboratory adaptive evolution potentially decreasing the level or the activity of GalP, a galactose permease. GalP is the main substitute glucose transporter in KJ122 since the phosphoenolpyruvate (PEP)-dependent phosphotransferase system (PTS) was inactivated in KJ122 to improve PEP levels for succinate production. We showed that deletion of *galP* increased xylose utilization in KJ122 and wild-type *E. coli*, suggesting a common repressive role of GalP for xylose fermentation. Concomitantly, induced expression of *galP* from a plasmid repressed xylose fermentation in wild-type *E. coli*. Transcriptome analysis using RNA sequencing indicates that *galP* inactivation increases transcription levels of many catabolic genes for secondary sugars including xylose and arabinose. In addition, many transcriptional changes facilitating succinate fermentative production were observed including genes involved in the anaplerotic reaction, the reductive branch of TCA cycle, and high-sugar osmotic stress tolerance. The

discovery of the repressive role of GalP for secondary sugars in *E. coli* suggests that utilization of GalP as a substitute glucose transporter may be undesirable for conversion of lignocellulosic sugar mixtures.

Importance

Lignocellulose is an abundant renewable feedstock that can be used for microbial conversion to produce renewable fuels and bulk chemicals. However, simultaneous co-utilization of heterogeneous sugar mixtures present in lignocellulose hydrolysates is a challenging task, especially when sugar concentrations are high. In this work, we discover that a commonly used substitute glucose transporter, GalP, represses the utilization of secondary sugars, such as xylose, in *E. coli* through transcriptional downregulation of many catabolic genes for secondary sugars, suggesting the presence of a novel repressive mechanism. Therefore, utilization of GalP as an alternative glucose transporter in PTS defective strains may hinder xylose conversion, an important consideration in engineering strains for bioconversion of lignocellulose to products that require the precursor PEP, such as C_4-dicarboxylates and aromatics derived from the shikimate pathway.

Introduction

Lignocellulose is a complex matrix present in the cell wall of plants that makes up more than half of the earth's total biomass, making it an abundant renewable feedstock (1). Independent of source, lignocellulose usually contains cellulose and hemicellulose, which can be degraded to glucose and pentoses for use as a feedstock for microbial conversion (2, 3). Glucose is the only sugar monomer in cellulose and xylose is the major sugar component for common hemicellulose fractions (2, 4). Simultaneous co-utilization of both sugars is desired for an efficient microbial lignocellulose bioconversion (5, 6). However, co-fermentation of mixed sugars often presents an obstacle to model microbial biocatalysts such as *Escherichia coli* due to a global regulation mechanism known as carbon catabolite repression (CCR) (6-8). For using secondary sugars such as xylose and arabinose in *E. coli*, transcriptional activation of the relevant catabolic operons needs 1) the activated global transcriptional regulator CRP (cAMP receptor protein) bound by cAMP, 2) the activated transcriptional regulators specific for the secondary sugars, such as XylR for xylose

and AraC for arabinose, respectively (7, 8). In the presence of glucose, CRP is not activated due to the low cAMP levels, thereby causing glucose-induced transcriptional repression of catabolic genes for secondary sugars (7, 8).

Glucose uptake and phosphorylation is mainly achieved by the phosphoenolpyruvate: sugar phosphotransferase system (PTS) in *E. coli* which is also an important component of CCR to control cAMP levels (6, 9, 10). Disruption of the PTS system (deletion of *ptsG* or *ptsI*) is an effective approach to release CCR, but glucose uptake is significantly impaired (11, 12), which necessitates alternative glucose transporter for optimal productivity (13-15). In addition to releasing CCR, inactivation of PTS also leads to more abundant PEP accumulation as PEP is used as the substrate by PTS to phosphorylate sugars through a phosphorylation cascade (9, 11), which is a useful approach to increase production of chemicals using PEP as the precursor, such as C_4-dicarboxylates using the reductive branch of TCA cycle (16, 17), and aromatic compounds derived from the shikimate pathway (18, 19). The galactose permease GalP, is a galactose:proton symporter that has glucose transport activity and has been often used as an alternative glucose uptake system in PTS deficient strains (13, 15, 19, 20).

Interestingly, the mutations that inactivate PTS and upregulate of *galP* were acquired during adaptive evolution of an *E. coli* biocatalyst to enhance glucose-succinate conversion (17). Although the strain derived from this laboratory evolution, KJ122, has high succinate production metrics with a high titer (~85 g l^{-1}), yield (~0.9 g g^{-1}), and productivity (~0.8 g l^{-1} h^{-1}) (21), the xylose fermentative growth of this strain is stunted with much lower production metrics (22). Here, to probe underlying genetic mechanisms inhibiting xylose utilization, we evolved KJ122 to enhance its xylose fermentation abilities in parallel and characterized the potential convergent genetic changes shared by multiple independently evolved strains. We discovered that GalP represses xylose fermentation in KJ122 and even in wild-type *E. coli*, suggesting that utilization of GalP as a substitute glucose transporter may be undesired for conversion of lignocellulosic sugar mixtures.

Results

Quick adaptation for xylose utilization in a succinate producing *E. coli* biocatalyst

KJ122 was previously developed for efficient conversion of glucose into succinate (21), but suffers from deficient xylose catabolism (22). Since this strain has all other fermentation and competing pathways disrupted, the growth and ATP production of this strain are linked to succinate fermentative production, thereby providing very strong growth-based selection. A quick adaption for using xylose was previously observed, and two evolved strains XW055 and AS1600a were isolated (22, 23). XW055 was isolated after ~40 generations, but the mechanism improving xylose fermentation remained unknown (22). For AS1600a, the causative mutation was identified as a point mutation in the coding region of galactose permease gene, *galP* (G236D). However, the working mechanism remains elusive. To gain deep mechanistic insights of the potential genetic changes inhibiting xylose fermentation, we hypothesized that characterization of independently evolved strains would reveal the convergent causative mechanism for this adaptation. We repeated the evolution of KJ122 in two independent experiments for enhanced xylose catabolism as previously described (22) by transferring cells during exponential growth phase into new AM1 medium containing 10% (w/v) xylose (Figure 2.1A and 2.1B). In both evolutionary trajectories, a rapid adaptation occurred even at the second or third transfer that simultaneously increased xylose catabolism and cell growth (Figure 2.1A, 2.1B and Figure 2.2). From these two evolved populations (approximately 60 generations), strains LP001 and CM001 were isolated and confirmed with increased xylose fermentation capabilities (Table 2.2).

Identification and characterization of convergent genetic changes occurring at the galactose regulon in the ancestor and evolved strains

The genomic DNAs of the ancestor KJ122 as well as three evolved strains LP001, CM001, and XW055 were extracted and sequenced using Illumina paired-end sequencing. Interestingly, all three evolved strains have genetic differences in *gal* regulon including *galR*, *galS* and *galP*, compared to the ancestor KJ122 (Figure 2.1C), suggesting a result of convergent evolution to relieve the inhibitory mechanism for xylose catabolism in the ancestor strain. We hypothesized that these genetic changes in *gal* regulon are responsible for the increased xylose fermentation.

KJ122 has an IS1 insertion sequence at the position 261 of *galR* ORF and an adenine insertion at the position 231 *galS* ORF causing frame-shift mutation, which likely inactivate both proteins

(Figure 2.1C). Both GalR and GalS are known to be repressors of genes needed for the transport and utilization of galactose, including *galP*, *galETKM*, and *mglBAC* (24) (Figure 2.1D). It was reported that expression of *galP* was increased to compensate the PTS deficiency in the precursor strain of KJ122 due to adaptive evolution for increased glucose to succinate bioconversion. However, the exact genetic mechanism remains uncharacterized (17). We hypothesized that these two mutations are responsible for enhanced *galP* expression that leads to higher glucose-succinate conversion. To test this, we reconstructed succinate producing *E. coli* XW01 derived from a previously engineered strain XZ721 with only four defined mutations (Δ*ptsI* Δ*pflB pck::pck** Δ*ldhA*) (16) and then tested of effect of the deletion of *galR*, *galS* or both on succinate production. In the strain XW01, *pflB* and *ldhA* were deleted to eliminate competing fermentation pathways for ethanol and D-lactate, respectively. The *ptsI* gene was deleted to disrupt PTS for increased levels of PEP, a precursor for succinate (16). In addition, *pck* expression was enhanced with an upstream mutation (*pck::pck**; G to A at position - 64 relative to the ATG start codon) (16, 17). Fermentation of XW01 in glucose resulted in a succinate titer of 28 g l^{-1} after 96 h (Figure 2.3A). Interestingly, single deletion of either *galR* or *galS* decreased the succinate titer at 96 h to 24 g l^{-1} and 9 g l^{-1}, respectively (Figure 2.3B and 2.3C). Deletion of both *galR* and *galS* enhanced production, increasing the titer from 28 g l^{-1} to 46 g l^{-1} after 96 h and from 62 g l^{-1} to 71 g l^{-1} after 144 h (Figure 2.3D). Overall productivity, specific growth rate, and titer were all increased when both *galR* and *galS* were deleted, but all decreased by inactivation of either *galR* or *galS* (Figure 2.3). This result supports that *galRS* inactivation was the causative mutations leading to increased *galP* expression in KJ122.

For the evolved strains, both XW055 and LP001 reverted *galR* back to the wild-type sequence during adaptive evolution while CM001 gained a nonsynonymous mutation in *galP* (A392D) (Figure 2.1C). AS1600a that was evolved independently in another group also was reported to have point mutation in *galP* (G236D) (23). Given the results of single deletions of *galR* and *galS* regarding their effect on succinate production (Figure 2.3), the restored *galR* in LP001 and XW055 will potentially decrease *galP* levels, which is supported by the defective performance of glucose fermentation (Table 2.2). Similarly, as shown in Table 2.2 and the previous report (23), CM001 and

AS1600a also showed decreased both glucose utilization and succinate production, suggesting these two point mutations probably negatively influence GalP function. It is plausible that decreased levels or activities of GalP are the underlying mechanism for enhanced xylose utilization by adaptive evolution.

Confirmation of the inhibiting role of *galP* on xylose fermentation in succinate production *E. coli* biocatalysts

To test if GalP represses xylose fermentation and also to directly prove that the inactivation of GalP is the causative mechanism for xylose adaption in KJ122, we deleted *galP* in KJ122 and compared the resulting strain AG055 to KJ122, CM001 and LP001 in terms of their fermentation performance using mono- and co-sugar conditions (Table 2.2). With xylose as a substrate, AG055 produced 70 g l^{-1} succinate after 96 hours, approximately 14-fold that of KJ122 (Table 2.2). This was comparable to the titer of the evolved strains XW055, CM001, and LP001 in the same period of time, producing 76 g l^{-1}, 80 g l^{-1}, and 77 g l^{-1}, respectively (Table 2.2). This demonstrated that inactivation/repression of *galP* was primarily responsible for the observed phenotype in the evolved strain. With a mixture of glucose and xylose as substrate, titer and productivity increased by approximately 25% in AG055 compared to KJ122 (Table 2.2). However, when glucose was used as the sole carbon source, similar to evolved strains AG055 had decreased succinate production (Table 2.2). This is consistent with the finding that GalP is a major glucose transporter in KJ122 which has a disrupted PTS (17). To directly test the repressive effect of *galP* on xylose fermentation, wild-type *galP* was cloned into a plasmid pTrc99A and *galP* expression was induced by 10 μM IPTG in AG055 to test the effect on xylose fermentation. Induced expression of *galP* resulted in a 3-day lag for AG055 before any succinate production and a 3-fold decrease in succinate titer after 96 hours (Figure 2.4).

Transcriptomic changes caused by *galP* inactivation

To understand the transcriptomic changes upon *galP* inactivation, we isolated RNAs for both KJ122 and AG055 (KJ122 Δ*galP*) in early exponential growth during xylose fermentation and used RNA sequencing to quantify transcriptomic differences. There were 17 genes downregulated and 75 genes upregulated at least 2-fold in AG055 compared to KJ122 (Table 2.3). First, it is noticeable

that many catabolic genes, especially for secondary sugars such as xylose, arabinose, ribose and galactose are upregulated upon *galP* inactivation (Table 2.3). The xylose (*xylAB*, *xylFGH*, *xylE* and *xylR*), arabinose (*araBAD*, *araFGH* and *araC*), ribose (*rbsDACB*, *rbsK* and *rbsR*), and galactose (*mglBAC* and *galETKM*) catabolic genes were upregulated 2.2, 2.0, 5.5, and 2.4-fold on average estimated by relative ratio of transcripts per million (TPM) (Figure 2.5). Second, many transcriptional changes facilitating succinate fermentative production were observed including genes involved in the anaplerotic reaction, the reductive branch of TCA cycle, succinate transport and high-sugar osmotic stress tolerance (Figure 2.5). In particular, there was a 3.5-fold increase in the expression of *pck* that encodes PEP carboxykinase known to be crucial for conserving energy in succinate production (17), while the competing carboxylation enzyme, *ppc*, was downregulated by 2-fold. This resulted in an increase in the ratio of TPM of *pck* over *ppc* from 1.5 to 8.5 in AG055. The transcripts of fumarate reductase, catalyzing the conversion of fumarate to succinate production (encoded by *frdABCD*) was upregulated 2.5-fold as well. These changes may benefit xylose to succinate fermentative conversion by enhancing xylose catabolism and carbon flux to succinate through the TCA cycle (Figure 2.5).

GalP represses xylose catabolism in wild-type *E. coli*

To test if the negative regulatory effect of *galP* is a common scenario not only specific to succinate producers, we deleted *galP* in wild-type *E. coli* W (ATCC 9637) and compared the resulting strain GK501 to wild-type strain for their xylose fermentation performance. Inactivation of *galP* enhanced the initial xylose consumption rate (0-24 h) by 140%, suggesting that the repressive effect of *galP* is commonly present in wild-type *E. coli* (Figure 2.6A). Cell growth of GK501 was also increased compared to wild-type with 66% more biomass at 24 hours (Figure 2.6A). As a complementary test, *galP* expression was induced from a plasmid in wild-type *E. coli* W to test the effect on xylose fermentation. Induced expression of *galP* resulted in much slower growth and xylose utilization with only approximately half biomass accumulated for 24 and 48 hours compared to the empty vector control (Figure 2.6B).

Discussion

Characterization of microbial control mechanisms for sugar preference will help engineering efficient sugar co-utilization, which is desired for lignocellulose microbial conversion. In this work, we discovered a novel mechanism that natively represses xylose fermentation through a galactose permease, GalP. The reverse engineering of xylose adaptation mechanisms for multiple experimental evolutionary trajectories of an *E. coli* succinate production strain revealed that a set of convergent causative mutations disrupting *galP* enhanced xylose fermentation. The transcriptomic analysis suggests that the improved fermentation is likely due to the increased expression of the genes involved in secondary sugar catabolism and succinate fermentative production.

Succinate, a C_4-dicarboxylic acid with a multi-billion dollar market, can be used to make many commodity chemicals in plastics and solvents (25). Many efforts have been made to both isolate and engineer biocatalysts for succinate production with varying success (26). A series of efficient *E. coli* strains, such as KJ73 and KJ122, have been engineered using a combination of metabolic evolutions and chromosomal deletions of competing pathways, ultimately producing succinate at a high titer (85 g l^{-1}), yield (0.9 g g^{-1}), and productivity (0.8 g l^{-1} h^{-1}) from glucose (21, 27). In these strains, GalP was found to be more abundant (a 20-fold increase at transcriptional levels) compared to wild-type strain and serve as the dominant glucose transporter in evolved strains (17). However, the mutations increasing *galP* expression remain elusive until this reported work: inactivation of both *galR* and *galS* is required for increased GalP activities serving an effective glucose uptake system in the PTS deficient background (Figure 2.1 and 2.3). Six defined chromosomal modifications (Δ*ptsI* Δ*pflB pck::pck** Δ*ldhA* Δ*galR* Δ*galS*) in wild-type *E. coli* ATCC8739 without any experimental adaptation yielded efficient succinate production with a titer at 71 g l^{-1}, a yield at 1.0 g g^{-1}, a productivity at 0.50 g/ l^{-1} h^{-1} for 6-day simple batch fermentations (Figure 2.3).

GalR and GalS are homologous dimeric repressors of the galactose regulon in *E. coli* (28). Genes encoding the ATP-dependent galactose transport system *mglBAC,* the Leloir pathway operon *galETKM,* and the galactose:proton symporter *galP* are repressed by GalR and GalS, and meanwhile activated by the presence of CRP bound with cAMP (24, 29). Our data suggest the presence of either GalR or GalS was sufficient to repress *galP* at some degree, leading to low

glucose consumption (Figure 2.3), which agrees with the evolved mechanism to restore only one copy of repressors, *galR*, in LP001 and XW055 (Figure 2.1). The restored *galR* is due to the removal of IS*1* element which was strongly selected for higher fitness during xylose adaptation. In general, IS elements increase genome dynamics and cause genome rearrangements, which could be beneficial for cells to adapt to new environments (30).

A GalP mutation (G236D) was previously identified using a similar approach, but the exact working mechanism is unclear (23). Although it is possible that *galP* G239D is a 'gain of function' mutation making GalP an efficient xylose transporter, the strain with a complete *galP* deletion in the same background (KJ122) showed very similar fermentation performance compared to the strain with *galP* (G236D) mutation (Table 2.2) (23). In addition, the glucose transport function of GalP (G236D) is also disrupted (23), which is also similar to the strain with the *galP* deletion and all other three evolved strains (Table 2.2). Characterization of the convergent genetic basis for three independently evolved strains from the same ancestor KJ122 suggests that the 'loss-of-function' of *galP* is likely the convergent mechanism for this quick xylose adaptation. This strategy of investigating parallel evolutionary trajectories is very effective to comprehensively understand underlying molecular mechanisms for the improved phenotypes.

Besides CCR occurring at the transcriptional level, the activities of some sugar transporters can be biochemically regulated through protein-protein interactions to achieve the control of sugar utilization preference (8, 31-33). Different from this type of biochemical regulation, the *galP*-induced repression seems to modulate global transcriptional regulation of many catabolic genes for secondary sugars as characterized by the RNA sequencing (Figure 2.5). The exact mechanism is still elusive, but global regulators such as CRP and Cra in control of secondary sugar catabolism may play a role in this process since many relevant catabolic genes are transcriptionally changed due to deletion of *galP*.

Transcriptomic data also suggests that there are beneficial changes for succinate production when *galP* is deleted. First, expression differences were observed in *ppc* and *pck* which encode two anaplerotic enzymes catalyzing carboxylation of PEP to form oxaloacetate (Figure 2.5). Compared to Ppc, Pck was found to conserve energy (generating one net ATP) and increased *pck*

expression enhanced succinate production in *E. coli* (16, 17). The relative ratio of *pck* to *ppc* transcripts between KJ122 and its *galP* deletion mutant was changed from 1.5:1 to 8.5:1, presumably conserving more energy during the conversion of PEP to oxaloacetate. Modulating the relative abundance of Pck and Ppc has been shown to enhance glucose to succinate conversion (34). Similarly, overexpression of *pck* in combination with a *ppc* inactivation was needed to effectively convert xylose to succinate (35, 36). The net energy gained from utilization of xylose is less than that of glucose (37). The distribution of metabolic flux between these two enzymes to maximize energy conservation could be important for xylose to succinate conversion. The need for conservation of energy for xylose to succinate conversion is further supported by a recent finding that inactivation of *xylFGH* (an energy intensive xylose transporter) enhances xylose to succinate conversion (38). Second, the expression of fumarase and fumarate reductase genes was increased upon *galP* inactivation, which would enhance carbon flux in the reductive branch of TCA cycle. Third, upregulation of a betaine ABC transporter, *proVWX*, was also observed upon *galP* inactivation. Osmotic stress is present in fermentation with 10% (w/v) sugar and previous evidence has shown that addition of the osmolyte betaine to AM1 mineral salts media can significantly increase production metrics (39). Thus, enhanced import of betaine could also play a role in enhancing productivity in AM1 mineral salts medium during succinate production by reducing osmotic stress.

Inactivation of the PTS has been widely used as an engineering strategy in bacteria to increase metabolic flux downstream of PEP and relieve carbon catabolite repression for a variety of products, but results in a deficiency in glucose uptake and utilization (12, 13, 15, 19). In many studies, *galP* was overexpressed to compensate the PTS defectiveness and enhance substrate uptake (13, 40). However, these strains are not often tested for xylose or glucose-xylose fermentation yet, and reported strains may be sub-optimally designed for conversion of sugar mixtures derived from lignocellulose because of the repressive effect of *galP*, especially under a *galP* overexpression scenario. Although *galP* inactivation can increase xylose and co-sugar utilization, glucose consumption remains low in these strains due to the defective PTS. Thus,

further work may look into the use of heterologous glucose transporters presumably without repressive effect on secondary sugars, such as Glf from *Zymomonas mobilis* (41).

Materials and Methods

Strains, plasmids and Cultivation Conditions

All strains and plasmids used in this work are listed in Table 2.1. The plasmid encoding *galP* was constructed by assemble the fragment containing the native ribosomal binding site, coding region and terminator of *galP* with the backbone of pTrc99A at the multiple cloning site using the circular polymerase extension cloning (CPEC) method (42). Primers for plasmid construction are listed in Table 2.4. The constructed plasmid was verified using Sanger sequencing. For strain construction, all manipulations were done in Luria Broth (10 g l^{-1} Difco tryptone, 5 g l^{-1} yeast extract, and 5 g l^{-1} NaCl) at 30°C, 37°C, or 39°C as needed with rotation at 180 rpm when in liquid culture. During genetic manipulations, 5% arabinose (w/v) was added to induce λ-red recombinase expression. 100 mg l^{-1} ampicillin, and/or 50 mg l^{-1} chloramphenicol were supplemented as needed.

Genetic Methods

Deletion of genes was performed using a λ-red recombinase based two-step recombination method as previously described (21, 43, 44). Briefly, strains transformed with pKD46 were grown in LB media supplemented with 5% arabinose (w/v) to an OD ~0.3 - 0.5 before being washed with ice-cold water and electroporated with a linear DNA fragment with homology to the region of interest. The DNA fragment used in the first-step integration was made by amplifying the *cat-sacB* cassette from pXW1 (44) with 50 bp of homology to the target gene. The second-step integration was performed using linear fragments generated using fusion PCR (45) with 500 bp homology at each end. Chloramphenicol resistance and colony PCR were used to select the successful clones for the first step integration. For the second step integration, counter selection was performed in media supplemented with 10% (w/v) sucrose. Gene deletion then was verified using colony PCR. Primers for generation for all cassettes are listed in Table 2.4.

Fermentation

Pre-inoculum for fermentation was generated by transferring cells freshly grown on AM1 mineral salts (46) agar plates supplemented with 2% glucose into a 250 ml flask containing 100 ml AM1 media supplemented with the appropriate sugar (2% w/v) and incubating for approximately 18 hours (37ºC, 120 rpms). All batch fermentations were performed using AM1 mineral salts media supplemented with the appropriate concentrations of sugar in fermentation vessels with 300 ml working volume as previously described (21). 100 mM potassium bicarbonate was included in the fermentation medium for succinate fermentative production (21). Co-sugar fermentation was performed using 5% glucose and 5% xylose (w/v). An initial inoculum of 0.022 g cell dry wt (CDW) l^{-1} (0.05 as OD550nm) was used for all fermentation tests. 100 mg l^{-1} Ampicillin and 10 µM Isopropyl β-D-1-thiogalactopyranoside (IPTG) were included if fermentation used the strains with plasmids. Fermentations were performed at 37°C and pH was controlled at 7.0 by automatic addition of base solution (2.4 M potassium carbonate and 1.2 M potassium hydroxide) as previously described (16, 21).

Laboratory Evolution

The strain KJ122 was consecutively transferred 10 times (~60 generations) during approximately logarithmic growth into new fermentation vessels with fresh AM1 medium supplemented with 10% xylose and 100mM potassium bicarbonate. An initial inoculum of 0.022 g CDW l^{-1} was used for all transfers. This was performed for two independent trajectories until xylose utilization became stable (~10 transfers for each trajectory). The final population was saved as in a cryogenic tube (-80 °C) and individual isolates were re-fermented to confirm stability of the phenotype. CM001 and LP001 were selected from each evolved population for further investigation along with another previously evolved strain XW055 (22).

Genome sequencing and variant calling

Genomic DNAs were extracted from KJ122 and all evolved strains (CM001, LP001, XW055) using the Promega Wizard genomic DNA purification kit according to manufacturer instructions. Purified DNAs were fragmented to an average size of 500bp and libraries were generated with a TruSeq DNA sample preparation kit (Illumina). Sample preparation and paired-end sequencing was performed with technical duplicates using an Illumina MiSeq (2 x 300 bp) by the DNASU

Sequencing Core at Arizona State University. Reads were trimmed with Trim Galore (https://github.com/FelixKrueger/TrimGalore), aligned to the *Escherichia coli* ATCC 8739 reference genome using BWA (47). Sequencing duplicates were removed with Picard (https://github.com/broadinstitute/picard), and variants were called using GATK HaplotypeCaller (48). Duplications, deletions, and other sequence junctions were called using CNVnator (49) and breseq (50).

Transcriptomic profiling by RNA sequencing

Strains KJ122 and AG055 were grown to an OD ~0.7 under 10% xylose (w/v) fermentation conditions and had total RNA extracted using a Qiagen RNAeasy kit according to the manufacturer instructions. Two pools (2 biological replicates for each pool) of total RNAs for each strain were prepared by combining equal amounts of RNA for each replicate. Samples were depleted of ribosomal RNA using a RiboZero kit (Illumina). Random hexamer priming was used to generate cDNA and libraries were prepared using a Nextera library prep kit (Illumina). Sample preparation and sequencing using an Illumina NextSeq (2x150bp) was performed by the DNASU Sequencing Core at Arizona State University. Reads were trimmed with Trim Galore, aligned to an *Escherichia coli* ATCC 8739 reference using STAR (51) and had differential gene expression analysis performed using edgeR (52).

Analyses

Sugars and organic acids in fermentation broth were measured by high-performance liquid chromatography (HPLC) using an Aminex® HPX-87H column (Bio-Rad) and 4 mM sulfuric acid as the mobile phase, as previously described (44). Cell dry weight was calculated from the measured optical density at 550 nm (0.44 g cell dry weight l^{-1} when OD550nm is 1.0).

Figures

Table 2.1. Strains and plasmids used in Chapter 2

Strain/Plasmid	Relevant characteristics	Reference
Strains		
KJ122	ATCC 8739, *pck**, *ptsI**, Δ*ldhA*, Δ*adhE*, Δ*ackA*, Δ(*focA-pflB*) Δ*mgsA*, Δ*poxB*, Δ*tdcDE*, Δ*citF*, Δ*aspC*, Δ*sfcA*	Wang et al 2013
AG055	KJ122, Δ*galP*	This study
XW055	KJ122 isolate adapted for ~40 generations in 10% (w/v) xylose	Wang et al 2013
CM1	KJ122 isolate adapted for ~60 generations in 10% (w/v) xylose	This study
LP1	KJ122 isolate adapted for ~60 generations in 10% (w/v) xylose	This study
XZ721	ATCC 8739, *pck**, Δ*ptsI,* Δ*pflB*	Zhang et al 2009
XW01	XZ721 Δ*ldhA*	This study
LP03	XW01, Δ*galR*	This study
LP05	XW01, Δ*galS*	This study
LP07	XW01, Δ*galR,* Δ*galS*	This study
ATCC 9637	Wildtype *E. coli*	ATCC
GK501	ATCC 9637, Δ*galP*	This study
Top10F'	F´{lacIq, Tn10(TetR)} mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG	Invitrogen™
Plasmids		
pKD46	*bla*, *γ β exo* (Red recombinase)	Datsenko et al 2000
pXW1	*The cat-sacB* cassette with the *sacB* native terminator cloned into a modified vector pLOI4162	Sievert et al 2017
pTrc99A	P_{trc}, *bla*, *lacI*^{q}	This study
pTrc99A-*galP*	*galP* in pTrc99A	This study

Table 2.2. Fermentation parameters KJ122 and its derived strains in AM1 supplemented with indicated carbon sources after 96 hours

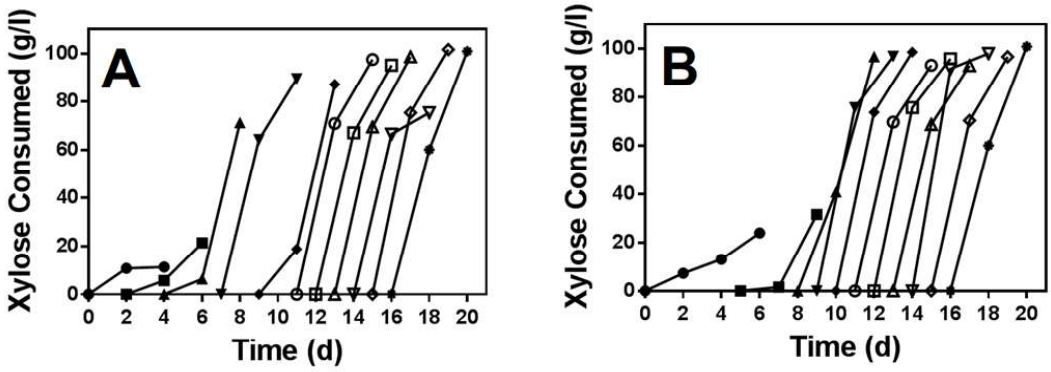
Strain	Carbon source	Maximum Biomass (g/l)	Succ Titer (g/l)	Yield (g/g)	Productivity (g/l/h)
KJ122	10% Glc	2.9 ± 0.3	87 ± 3	0.91 ± 0.02	0.91 ± 0.03
	10% Xyl	0.2 ± 0.1	5 ± 2	0.41 ± 0.15	0.05 ± 0.02
	5% Glc 5% Xyl	2.2 ± 0.1	61 ± 3	0.86 ± 0.04	0.64 ± 0.03
AG055	10% Glc	2.9 ± 0.1	45 ± 1	0.93 ± 0.01	0.47 ± 0.01
	10% Xyl	2.0 ± 0.2	70 ± 5	0.90 ± 0.05	0.73 ± 0.05
	5% Glc 5% Xyl	1.9 ± 0.1	76 ± 3	0.94 ± 0.05	0.80 ± 0.03
CM001	10% Glc	2.5 ± 0.1	66 ± 2	0.93 ± 0.04	0.68 ± 0.02
	10% Xyl	2.3 ± 0.1	80 ± 3	0.85 ± 0.03	0.80 ± 0.01
XW055	10% Glc	3.2 ± 0.2	62 ± 3	0.81 ± 0.07	0.64 ± 0.05
	10% Xyl	2.2 ± 0.1	76 ± 1	0.91 ± 0.01	0.83 ± 0.03
LP001	10% Glc	2.3 ± 0.2	45 ± 1	1.11 ± 0.01	0.46 ± 0.02
	10% Xyl	2.5 ± 0.2	77 ± 6	0.93 ± 0.09	0.79 ± 0.04

Table 2.3. Differentially expressed genes with greater than a 2-fold change in expression and a false discovery rate (FDR) < 0.1.

Functional categories	Downregulated genes		Upregulated genes			
Carbon metabolic genes	*maa*	*dmlA*	*pck*	*ygcE*	*ydjI*	
	glgC	*ppc*	*lpxL*	*agp*	*sucD*	*ykgEFG*
			frdA,B,D	*aspA*	*rbsK*	*rbsD*
Regulator for carbon utilization			*cstA*	*araC*	*rbsR*	
Sugar and acid transporters			*xylH*	*yqcE*	*glpT*	*yedE*
			rbsABC	*mglBC*	*btsT*	*dcuA*
Amino acid metabolism	*nepI*	*lysA*	*ilvN*	*mtr*	*ansB*	pepE
	asd		*fucO*	*trpE*	*oppB*	
Regulatory function unrelated to carbon metabolism						
	ada	*yihI*	*yrbL*	*ygiM*	*nuoHM*	*rpoE*
			ychH	*caiF*		
Other unrelated functions	*mutM*	*truC*	proVWX (betaine uptake)			
	yqcC	*alkB*	*alkA* *ssuABCDE* *pspE*			
	phoE					
	(porins)		(sulfonate-sulfur utilization)			
			abgA,ybcF, ylbF, fdrA, allCD			
			(allantoin/nitrogen catabolism)			
			ompW, ompF (porins)			*uspF*
			aslA	*cutC*	*pstA*	*ydeM*
Unclassified	*ydjO*	*yigI*	*ucpA*	*preAT*	*ydeP*	*ybhG*
	gspG		*ybhF*	*ybhR*	*ybhQ*	*ylbE*
			ybaE			

Table 2.4. Primers used in Chapter 2

Primer	Sequence (5' to 3')
galP deletion	GCTGCCGGTCTGAAGTAATC
	CCACTTCCACTCTTTGCCAT
	TACTCACCTATCTTAATTCACAATAAAAAATAACCATATTGGAGGGCTCGAGTGTGACGGAAGATCA
	GATGACTGCAAGAGGTGGCTTCCTCCGCGATGGGAGGAAGCCTGAGGAGACTTAGCCATTTGCCTGCTTTT
	TCACAATAAAAAATAACCATATTGGAGGGCATCTCTCCCCAAGCTTCCTCCCATCGCGGAGGA
	TCCTCCGCGATGGGAGGAAGCTTGGGGAGAGATGCCCTCCAATATGGTTATTTTTTATTGTGA
galR deletion	AGGTTGGTGCAAGAGACAGG
	TAGCGCAGTCTGGCGTATTA
	TTTTCCGTAACACTGAAAGAATGTAAGCGTTTACCCACTAAGGTATTTTCTCGAGTGTGACGGAAGATCA
	GTCGCCAGACCATCGAAGAATTACTGGCGCTGGAATTGCTTTAACTGCGGTTAGCCATTTGCCTGCTTTT
	TGTAAGCGTTTACCCACTAAGGTATTTTCCCGCAGTTAAAGCAATTCCAGC
	GCTGGAATTGCTTTAACTGCGGGAAAATACCTTAGTGGGTAAACGCTTACA
galS deletion	ATAGGCTATGCGTCGCTGTT
	TGATCGTTCTGCTTGGACTG
	TAGCCGCCAGCAAGCAGTCATTTACTGCAATCTCATAACAGGTAGTGAATTCGAGTGTGACGGAAGATCA
	AGCCATCACATTGTTAAGATACTGTGAAATCACTCACAGATTGAAAGCGGTTAGCCATTTGCCTGCTTTT
	TCTCATAACAGGTAGTGAATCCGCTTTCAATCTGTGAGTG
	CACTCACAGATTGAAAGCGGATTCACTACCTGTTATGAGATTGCAG
galP cloning into pTrc99a	GGATCCTCTAGAGTCGACCTGCAGG
	CGGGTACCGAGCTCGAATTCCATG
	CATGGAATTCGAGCTCGGTACCCGTAACCATATTGGAGGGCATCATGCCT
	CCTGCAGGTCGACTCTAGAGGATCCGCGGCAGAGGATAGAGCGAAG
ptsI deletion	ACTGTATTGCGCTCTTCGTG
	CCGGAGTCAGGGTAGACTTG
	GAGTAATTTCCCGGGTTCTTTTAAAAATCAGTCACAAGTAAGGTAGGGTTTCGAGTGTGACGGAAGATCA
	ACAAACCCATGATCTTCTCCTAAGCAGTAAATTGGGCCGCATCTCGTGGATTAGCCATTTGCCTGCTTTT
	AATCAGTCACAAGTAAGGTAGGGTTTCCACGAGATGCGGCCCAATT
	AATTGGGCCGCATCTCGTGGAAACCCTACCTTACTTGTGACTGAT
pflB deletion	CCTGGCAAACCTGATGGTAT
	ACTCAGCTTGCAGGATTGCT
	TGTCGAAGTACGCAGTAAATAAAAAATCCACTTAAGAAGGTAGGTGTTACTCGAGTGTGACGGAAGATCA
	GTGGAGCCTTTATTGTACGCTTTTTACTGTACGATTTCAGTCAAATCTAATTAGCCATTTGCCTGCTTTT
	ATCCACTTAAGAAGGTAGGTGTTACTTAGATTTGACTGAAATCGTACAGTAAAAAGC
	GCTTTTTACTGTACGATTTCAGTCAAATCTAAGTAACACCTACCTTCTTAAGTGGAT
ldhA deletion	CTCCCACCAGATAACGGAGA
	CTTTGGCTGTCAGTTCACCA
	AAATATTTTTAGTAGCTTAAATGTGATTCAACATCACTGGAGAAAGTCTTTCGAGTGTGACGGAAGATCA
	ATTGGGGATTATCTGAATCAGCTCCCCTGGGTTGCAGGGGAGCGGCAAGATTAGCCATTTGCCTGCTTTT
	ACATCACTGGAGAAAGTCTTTCTTGCCGCTCCCCTGCA
	TGCAGGGGAGCGGCAAGAAAGACTTTCTCCAGTGATGTTGAATC
*pck** integration	GGTAGCAAACCAGACCCTGA
	GCTATGAGCGGGAAATTGAA
	TACATATTGGCTAAGGAGCAGTGAATCGAGTGTGACGGAAGATCA
	TGATCTTCCGTCACACTCGATTCACTGCTCCTTAGCCAATATGTA
	AAAAGCAGGCAAATGGCTAATGATTTGAAGCTGGAGAATATCTATCCAG
	CTGGATAGATATTCTCCAGCTTCAAATCATTAGCCATTTGCCTGCTTTT
	GAAAAACGTGATGCCAACAA
	TTCATCAGGGCGATTTCTCT



C

	KJ122	CM1	LP1	XW55	AS1600a
galR	-	-	+	+	-
galS	-	-	-	-	-
galP	+	A392D	+	+	G236D

D

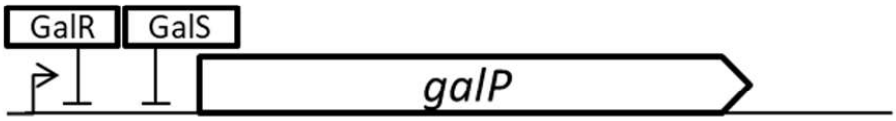


Figure 2.1. KJ122 was adapted for xylose fermentation by serial transfers of cultures in AM1 supplemented with 10% (w/v) xylose. This adaptive laboratory evolution was performed independently two times, from which A) LP001 and B) CM001 were isolated. Used xylose was determined for each transfer during fermentation. C) Primary convergent mutations were characterized using genome sequencing for CM001, LP001 and XW055, a previously evolved strain (24). In comparison to these three strains, a previously evolved and characterized strain AS1600a (30) was also summarized. D) Schematic drawing of the transcriptional repression of *galP* by both GalR and GalS. Inactivation of *galS* was by an adenine insertion at the position 231 in *galS* ORF, whereas inactivation of *galR* was by an IS1 element inserted at the position 261 in *galR* ORF. The given size does not reflect real proportions

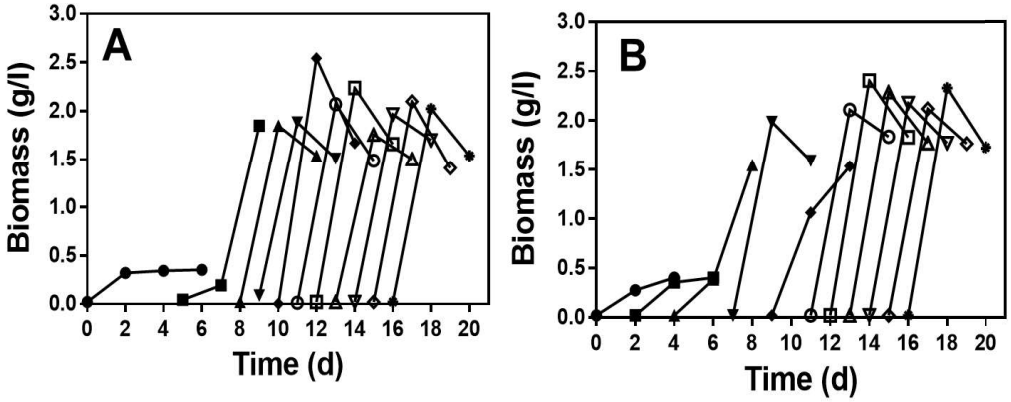


Figure 2.2. KJ122 was adapted for xylose fermentation by serial transfers of cultures in AM1 supplemented with 10% (w/v) xylose. This adaptive laboratory evolution was performed independently two times, from which A) LP001 and B) CM001 were isolated. Adaptive evolution increased biomass for both evolutionary trajectories.

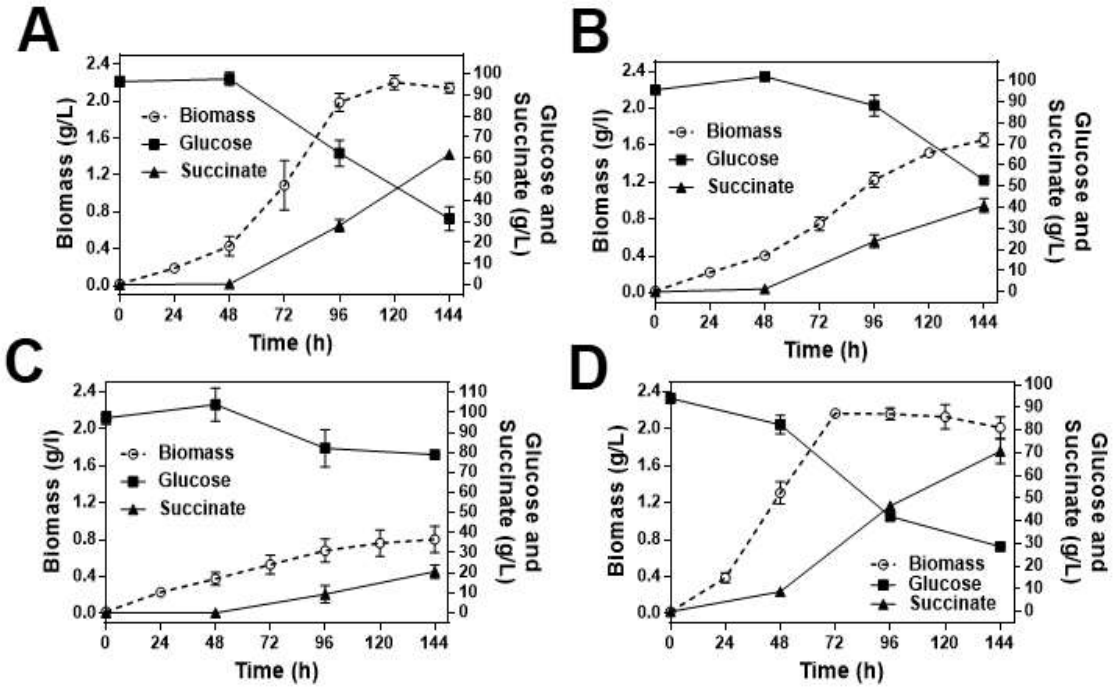


Figure 2.3. Fermentation of a rationally designed succinate producer A) XW01 and its derivatives with inactivation of B) *galR,* C) *galS,* and D) *galRS* in AM1 medium supplemented with 10% glucose (w/v). Symbols for all: biomass (open circle), glucose (filled square), succinate (filled triangle).

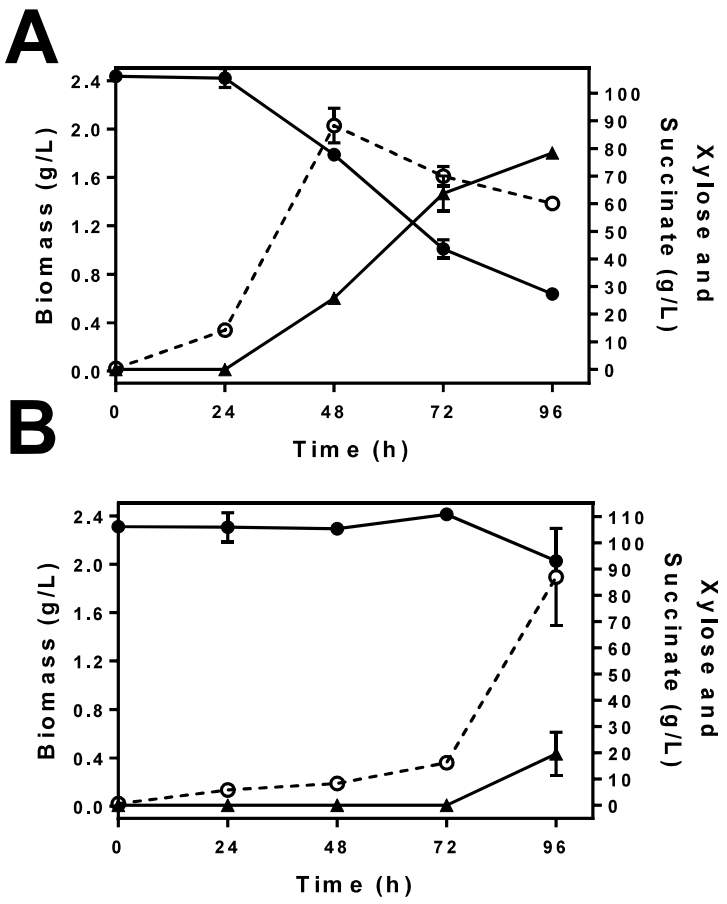


Figure 2.4. Fermentation of AG055 transformed with A) an empty vector pTrc99A or B) a plasmid encoding GalP in AM 1 medium containing 10% (w/v) xylose and 10 µM IPTG . Symbols for all: biomass (open circle; dotted lines), xylose (filled circle ; solid lines), succinate (filled triangle ; solid lines).

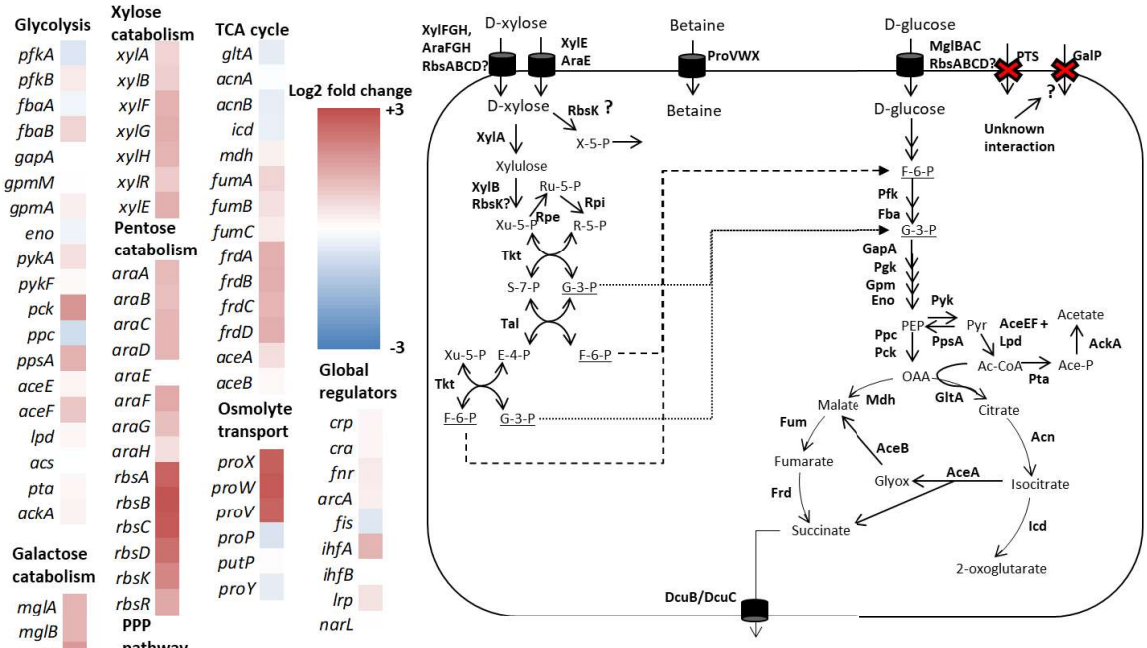


Figure 2.5. Differential expression analysis of relevant pathways relevant to succinate bioproduction in AG055. The genome-wide expression of AG055 is compared to KJ122 and relative fold change is expressed using a Log2 scale. Abbreviations: F-6-P, β-D-fructose-6-phosphate ; G-3-P, D-glyceraldehyde-3-phosphate; X-5-P, xylose-5-phosphate; Xu-5-P, xylulose-5-phosphate; Ru-5-P, ribulose-5-phosphate ; R-5-P, ribose-5-phosphate; S-7-P, sedoheptulose-7-phosphate; E-4-P, erythrose-4-phosphate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; Pyr, pyruvate; Ac-CoA, acetyl-CoA; Ace-P, acetyl phosphate; Suc-CoA, succinyl-CoA; PTS, phosphotransferase system.

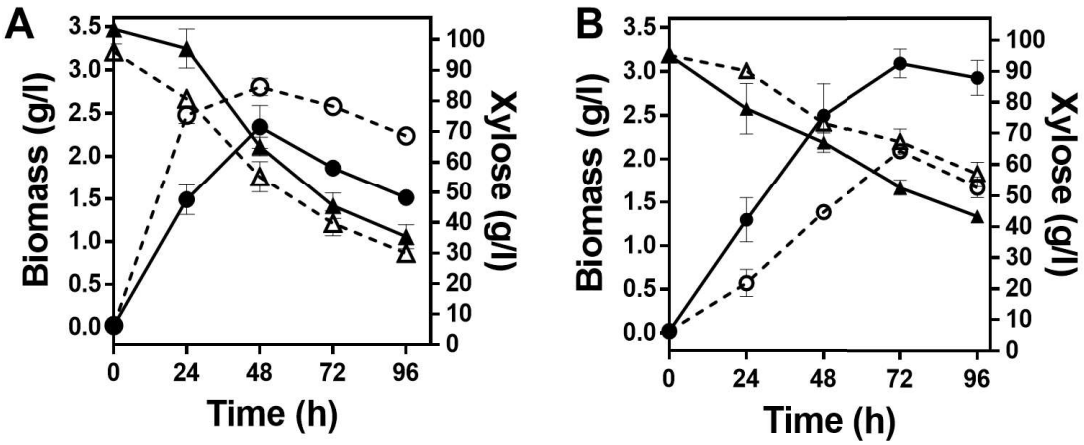


Figure 2.6. Effect of *galP* in wild-type background on xylose fermentation using AM1 medium supplemented with 10% xylose (w/v). A) wild-type *E. coli* W (solid lines) and GK501 (*E. coli* W Δ*galP*; dotted lines). B) *E. coli* W with empty vector pTrc99A (solid lines) and with a plasmid encoding GalP (dotted line). 10 μM IPTG was used to induce *galP* expression. Symbols for all: biomass (circle), xylose (triangle).

References

1. Mishra P, Singh A. 1993. Microbial pentose utilization. Adv Appl Microbiol 39:91-152.

2. Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30:279-91.

3. Garvey M, Klose H, Fischer R, Lambertz C, Commandeur U. 2013. Cellulases for biomass degradation: comparing recombinant cellulase expression platforms. Trends Biotechnol 31:581-93.

4. Flores A, Kurgan G, Wang X. 2016. Engineering bacterial sugar catabolism and tolerance toward lignocellulose conversion. *In* Gosset G (ed), Engineering of Microorganisms for the Production of Chemicals and Biofuels from Renewable Resources doi:10.1007/978-3-319-51729-2. Springer Nature.

5. Gorke B, Stulke J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol 6:613-24.

6. Deutscher J. 2008. The mechanisms of carbon catabolite repression in bacteria. Curr Opin Microbiol 11:87-93.

7. Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol Rev 57:543-94.

8. Gosset G. 2005. Improvement of *Escherichia coli* production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb Cell Fact 4:14.

9. Hernandez-Montalvo V, Valle F, Bolivar F, Gosset G. 2001. Characterization of sugar mixtures utilization by an *Escherichia coli* mutant devoid of the phosphotransferase system. Appl Microbiol Biotechnol 57:186-91.

10. Escalante A, Salinas Cervantes A, Gosset G, Bolivar F. 2012. Current knowledge of the *Escherichia coli* phosphoenolpyruvate-carbohydrate phosphotransferase system: peculiarities of regulation and impact on growth and product formation. Appl Microbiol Biotechnol 94:1483-94.

11. Hernandez-Montalvo V, Martinez A, Hernandez-Chavez G, Bolivar F, Valle F, Gosset G. 2003. Expression of galP and glk in a *Escherichia coli* PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol Bioeng 83:687-94.

12. Balderas-Hernandez VE, Hernandez-Montalvo V, Bolivar F, Gosset G, Martinez A. 2011. Adaptive evolution of *Escherichia coli* inactivated in the phosphotransferase system operon improves co-utilization of xylose and glucose under anaerobic conditions. Appl Biochem Biotechnol 163:485-96.

13. Lu J, Tang J, Liu Y, Zhu X, Zhang T, Zhang X. 2012. Combinatorial modulation of galP and glk gene expression for improved alternative glucose utilization. Appl Microbiol Biotechnol 93:2455-62.

14. McDonald TP, Walmsley AR, Henderson PJ. 1997. Asparagine 394 in putative helix 11 of the galactose-H+ symport protein (GalP) from *Escherichia coli* is associated with the internal binding site for cytochalasin B and sugar. J Biol Chem 272:15189-99.

15. Flores N, Xiao J, Berry A, Bolivar F, Valle F. 1996. Pathway engineering for the production of aromatic compounds in *Escherichia coli*. Nat Biotechnol 14:620-3.

16. Zhang X, Jantama K, Shanmugam KT, Ingram LO. 2009. Reengineering *Escherichia coli* for Succinate Production in Mineral Salts Medium. Appl Environ Microbiol 75:7807-13.

17. Davis EO, Henderson PJ. 1987. The cloning and DNA sequence of the gene *xylE* for xylose-proton symport in *Escherichia coli* K12. J Biol Chem 262:13928-32.

18. Sumiya M, Davis EO, Packman LC, McDonald TP, Henderson PJ. 1995. Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in *Escherichia coli*. Receptors Channels 3:117-28.

19. T. W, G P. 2004. Top value added chemicals from biomass. U.S. Department of Energy, Washington, DC.

20. Ahn JH, Jang YS, Lee SY. 2016. Production of succinic acid by metabolically engineered microorganisms. Curr Opin Biotechnol 42:54-66.

21. Jantama K, Zhang X, Moore JC, Shanmugam KT, Svoronos SA, Ingram LO. 2008. Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. Biotechnol Bioeng 101:881-93.

22. Zhang X, Jantama K, Moore JC, Jarboe LR, Shanmugam KT, Ingram LO. 2009. Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. Proc Natl Acad Sci U S A 106:20180-5.

23. Jantama K, Haupt MJ, Svoronos SA, Zhang X, Moore JC, Shanmugam KT, Ingram LO. 2008. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. Biotechnol Bioeng 99:1140-53.

24. Wang X, Yomano LP, Lee JY, York SW, Zheng H, Mullinnix MT, Shanmugam KT, Ingram LO. 2013. Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 110:4021-6.

25. Gonzalez JE, Long CP, Antoniewicz MR. 2017. Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by (13)C metabolic flux analysis. Metab Eng 39:9-18.

26. Singh A, Cher Soh K, Hatzimanikatis V, Gill RT. 2011. Manipulating redox and ATP balancing for improved production of succinate in *E. coli*. Metab Eng 13:76-81.

27. Liu R, Liang L, Chen K, Ma J, Jiang M, Wei P, Ouyang P. 2012. Fermentation of xylose to succinate by enhancement of ATP supply in metabolically engineered *Escherichia coli*. Appl Microbiol Biotechnol 94:959-68.

28. Anonymous. 1998. News Briefs: Half of the top 12 greenest vehicles are electric. Environ Sci Technol 32:213A.

29. Khunnonkwao P, Jantama SS, Kanchanatawee S, Jantama K. 2018. Re-engineering *Escherichia coli* KJ122 to enhance the utilization of xylose and xylose/glucose mixture for efficient succinate production in mineral salt medium. Appl Microbiol Biotechnol 102:127-141.

30. Sawisit A, Jantama K, Zheng H, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2015. Mutation in *galP* improved fermentation of mixed sugars to succinate using engineered *Escherichia coli* AS1600a and AM1 mineral salts medium. Bioresour Technol 193:433-41.

31. Weickert MJ, Adhya S. 1993. The galactose regulon of *Escherichia coli*. Mol Microbiol 10:245-51.

32. Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, Wang X. 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc Natl Acad Sci U S A 114:7349-7354.

33. Geanacopoulos M, Adhya S. 1997. Functional characterization of roles of GalR and GalS as regulators of the gal regulon. J Bacteriol 179:228-34.

34. Semsey S, Krishna S, Sneppen K, Adhya S. 2007. Signal integration in the galactose network of *Escherichia coli*. Mol Microbiol 65:465-76.

35. Laikova ON, Mironov AA, Gelfand MS. 2001. Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria. FEMS Microbiol Lett 205:315-22.

36. Shimada T, Kori A, Ishihama A. 2013. Involvement of the ribose operon repressor RbsR in regulation of purine nucleotide synthesis in *Escherichia coli*. FEMS Microbiol Lett 344:159-65.

37. Mauzy CA, Hermodson MA. 1992. Structural and functional analyses of the repressor, RbsR, of the ribose operon of *Escherichia coli*. Protein Sci 1:831-42.

38. Song S, Park C. 1998. Utilization of D-ribose through D-xylose transporter. FEMS Microbiol Lett 163:255-61.

39. Ginsburg A. 1959. A deoxyribokinase from Lactobacillus plantarum. J Biol Chem 234:481-7.

40. Chuvikovsky DV, Esipov RS, Skoblov YS, Chupova LA, Muravyova TI, Miroshnikov AI, Lapinjoki S, Mikhailopulo IA. 2006. Ribokinase from E. coli: expression, purification, and substrate specificity. Bioorg Med Chem 14:6327-32.

41. Zhou S, Grabar TB, Shanmugam KT, Ingram LO. 2006. Betaine tripled the volumetric productivity of D(-)-lactate by *Escherichia coli* strain SZ132 in mineral salts medium. Biotechnol Lett 28:671-6.

42. Long CP, Au J, Sandoval NR, Gebreselassie NA, Antoniewicz MR. 2017. Enzyme I facilitates reverse flux from pyruvate to phosphoenolpyruvate in *Escherichia coli*. Nat Commun 8:14316.

43. Wei T, Cheng BY, Liu JZ. 2016. Genome engineering *Escherichia coli* for L-DOPA overproduction from glucose. Sci Rep 6:30080.

44. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-5.

45. Shevchuk NA, Bryksin AV, Nusinovich YA, Cabello FC, Sutherland M, Ladisch S. 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res 32:e19.

46. Quan J, Tian J. 2014. Circular polymerase extension cloning. Methods Mol Biol 1116:103-17.

47. Martinez A, Grabar TB, Shanmugam KT, Yomano LP, York SW, Ingram LO. 2007. Low salt medium for lactate and ethanol production by recombinant *Escherichia coli* B. Biotechnol Lett 29:397-404.

48. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.

49. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11 10 1-33.

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

51. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165-88.

52. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21.

53. Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139-40.

54. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK. 2011. The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of E. coli. BMC Genomics 12:9.

CHAPTER 3

DIRECTED EVOLUTION TO RELEASE GLUCOSE INHIBITION OF GLF FOR LOW-ENERGY IMPORT OF LIGNOCELLULOSIC SUGARS

Abstract

Cellular import of D-xylose, the second most abundant sugar in lignocellulosic biomass, has been evidenced to be an energetic burden using the native transport systems of multiple bacterial biocatalysts. The glucose facilitator of *Zymomonas mobilis*, *glf*, is capable of high velocity xylose uniport, but is known to be inhibited by D-glucose, potentially limiting the utility of this transporter in sugar mixtures derived from lignocellulose. In this work we developed an *E. coli* transport deficient platform strain that is incapable of glucose catabolism (Glc^-) for the screening of glucose resistant variants of Glf. Using this platform, we isolate variants by enrichment derived from both random and rational mutagenesis with a 60-210% increase in the D-xylose consumption rate in co-sugar medium, demonstrating the utility of this approach. Mutations relieving glucose repression from this approach are diverse and include previously described mutations such as a conserved asparagine of sugar porters in TM8, as well as novel point mutations that target residues specific to glucose coordination, cause subtle changes in TM12, and more. Furthermore, fermentation of these variants in a Glc^+ background demonstrates that these mutations do not inactivate glucose import but appear to increase the preference for xylose in co-sugar fermentation. These findings suggest that these transporters may be energy-conservative alternatives to the native transport systems of many bacterial biocatalysts for fermentation of lignocellulose-derived sugars.

Importance

Glucose and xylose are the two most prevalent sugars in lignocellulosic biomass, a renewable carbon feedstock that can potentially be used for bioconversion to valuable compounds. Xylose transport, one of the first steps in xylose catabolism, is known to often be catalyzed by transporters that unnecessarily expend ATP and/or are inhibited by glucose. This work develops a multitude of variants of a glucose facilitator, *glf,* for xylose uniport that has been relieved of glucose inhibition. The use of this transport system theoretically allows an *E. coli* biocatalyst to reduce xylose

catabolism ATP expenditures by 50% during consumption of glucose-xylose mixtures reminiscent of lignocellulose hydrolysate. Additionally, mutations gained that confer this phenotype provide interesting insights into the similarity of structure/function mechanisms of the MFS sugar porters.

Introduction

Bioconversion of lignocellulose represents a renewable production route for many petroleum derived chemicals that does not compete with food crop production. However, the utilization of lignocellulose by microbes is limited by both the toxicity and sugar composition of lignocellulose biomass following chemical pretreatment strategies (1). Independent of feedstock, this sugar composition mainly consists of glucose and xylose, and co-consumption of these sugars impose challenges for D-xylose uptake systems and microbial regulatory pathways through carbon catabolite repression (CCR) for a broad range of biocatalysts (2-4). This results in sequential and mechanistically sub-optimal utilization of these sugars, which ultimately decreases important production metrics such as titer, productivity and yield (5). Recent findings in our lab discovered mutations in a xylose transcriptional activator, XylR, which essentially eliminates CCR for co-utilization of glucose/xylose sugar mixtures and is transferrable to multiple production backgrounds (6). Thus, circumventing energy expenditures and glucose inhibition during D-xylose uptake is one of the major remaining challenges in utilizing lignocellulose sugars.

Transport of D-xylose in bacteria is generally catalyzed by either primary or secondary active transporters, which results in unnecessary energy consumption. In *E. coli* the cell has two major xylose uptake systems that can be induced by xylose, XylFGH and XylE (7-9). XylFGH is a high affinity ATP-binding cassette transporter (ABC) which hydrolyzes ATP for substrate import and is known to be the main importer of xylose under fermentation conditions (8, 10, 11). XylE is a low affinity xylose:H^+ symporter and although more energy efficient it appears to catalyze much less of the physiologically relevant transport (7, 11). Following import, xylose is isomerized and phosphorylated before proceeding into the pentose phosphate pathway. Utilization of this native transport system however has been shown to approximately double the amount of ATP spent on substrate uptake relative to other efficient sugar uptake systems, such as glucose import (10, 12).

Since ATP is limited under microaerobic fermentation conditions due to inactivation of oxidative phosphorylation, this deprives the biocatalyst of energy, decreasing biomass and specific growth rates (10). This has been shown to be particularly problematic in biocatalysts with pathways that have been engineered to remove side products (e.g. acetate) which help to increase ATP yields. For instance, xylose to lactate conversion in *E. coli* has been shown to be inhibited due to the low ATP/xylose molar yield (0.67) of the pathway using XylFGH (11, 13). Inactivation of XylFGH immediately improves fermentative production, and by performing adaptive laboratory evolution (ALE) the cells activate a new transport system which presumably does not use ATP for import (13). Other fermentation products, such as succinate, have low a ATP yield which require energy conserving mechanisms in the pathway to even facilitate production from glucose (14). Thus mutations that increase ATP conservation, such as activation of the phosphoenolpyruvate carboxykinase, *pck,* through both direct and indirect methods have been shown to enhance both xylose and glucose-xylose to succinate conversion (14-16). The previous strategy of inactivating XylFGH coupled to ALE has been used to enhance xylose to succinate conversion as well, although the exact transport mechanism of these newly activated transporters remains unknown (17). More ideal would be the incorporation of a xylose uniporter to catalyze uptake without expending cellular energy, although these types of transporters appear to be rare in bacteria.

Some organisms such as *Zymomonas mobilis* and many eukaryotes have uniporters from the Major Facilitator Superfamily (MFS) which catalyze facilitated diffusion of monosaccharides (18). These transporters belong to the same family as many bacterial sugar symporters such as XylE and contain 12 transmembrane spans (TMS) that potentially recognize sugars at the C-terminal domain and facilitate translocation by a rocking mechanism in the N-terminal domain (19, 20). The substrate spectra of a broad range of eukaryotic uniporters has been characterized (21), however many of these transporters have low transport kinetics for xylose, are difficult to express in bacterial hosts, and are in general inhibited by glucose (19, 21, 22). The kinetics of these transporters have previously been improved in work by Young et al. who enhanced xylose uptake of multiple importers by modifying residues in a G-G/F-XXX- G motif that is widely conserved among sugars

porters (SP) in the MFS (22). In particular, this elucidated the importance of hydrophobic residues in this motif and found that a G-G/F-XX-F-G motif was particularly enriched in uniporters that effectively translocated xylose (22). However, these transporters still were inhibited by glucose. This inhibition results in sequential sugar utilization and decreased production metrics for multiple biocatalysts in glucose-xylose mixtures, and thus has been the target of protein engineering techniques (9, 23, 24).

The glucose facilitator, *glf*, of *Z. mobilis* is a desirable candidate to replace the native bacterial xylose import system since it has been functionally expressed in *E. coli,* already has the G-G/F-XX-F-G motif associated with xylose uptake, and has a high velocity for xylose import (18, 25). Additionally, previous work has found that mutations Glf-A18T and Glf-V275F can be used to enhance pentose import, suggesting that the specificity for xylose can be easily modified (26). However, like other MFS uniporters, it too is inhibited by glucose (27, 28). Previous work has found that structural interactions in TMS4-8 may be important in preventing efficient xylose and glucose-xylose co-utilization in this protein (28). Recent work has elucidated a high resolution of structure of XylE bound to both xylose and glucose (29). Within this structure, residues involved in coordination of the 6-hydroxymethyl group of glucose that are not involved with xylose coordination were found that reside with TM5, TM8, and TM10 (29). Additionally work by Farwick et al. and others have found mutations in TM5 and a conserved asparagine in TM8 that are important for releasing glucose inhibition of fungal xylose uniporters potentially by altering the binding pocket near the 6-hydroxymethyl group of D-glucose (30). However, many of these mutants have not often been tested under relevant fermentation conditions, or with biocatalysts that are relieved from catabolite repression (28, 30, 31). This may result in sub-optimal performance, thus emphasizing the importance of an optimized catabolic pathway released from CCR in the optimization of sugar porters for xylose uptake.

To create a bacterial xylose uniporter released from glucose inhibition we first created a platform strain to facilitate effective screening. We thus inactivated the known glucose and xylose transport systems in a strain relieved from CCR (*xylR::xylR**). By further inactivating glucose

catabolism (-*glk*) we created a strain with a selective advantage for xylose transport that is relieved from glucose inhibition. We then created multiple transporter libraries both by random mutagenesis and by targeting residues involved in glucose coordination with site saturation mutagenesis. By performing a pooled enrichment, we isolated 10 transporter variants that enhance xylose utilization under co-sugar fermentation conditions. We then find two potentially distinct mechanisms by which cells are relieved from glucose repression and that these alterations enhance xylose transport preference in mutant transporters.

Results

Construction of a screening platform for glucose resistant *glf*

To create a strain background to facilitate rapid identification of glucose resistant Glf variants, we sought to create a glucose and xylose transport deficient strain that would only be able to catabolize xylose, as has been done previously in yeast (30). We began by using a xylose specialist, AF11A, derived from *E. coli* ATCC 9637 that had been developed in previous work by our lab (publication pending). This strain has the major glucose uptake systems inactivated (Δ*ptsI*Δ*ptsG*Δ*galP*), resulting in almost no glucose consumption over 96h in a 10% (w/v) glucose fermentation even while catabolic pathways are present. Additionally, this strain has point mutations in *xylR* (*xylR**) which relieves the cell from CCR of xylose in the presence of glucose. We further developed this strain by inactivating known xylose importers *xylFGH, xylE,* and the arabinose importer, *araE*. Inactivation of *xylFGH* decreased the initial xylose consumption rate (0-24h) by ~50% (Figure 3.1A). By inactivating *xylE* in this background, we further decreased the xylose consumption rate to ~25% the parent strain (Figure 3.1A). Finally, inactivation of *araE* in this background decreased initial xylose consumption to ~8% of the parent strain (Figure 3.1A). Although inactivation of these transporters provided a strong selective pressure for efficient xylose transport, we found that additional xylose transporters facilitated uptake after 24h. To attempt to identify these transporters, we inactivated other potential xylose importers, *araFGH* and *gatC*. Fermentation of this strain, EG29, using xylose or arabinose as a carbon source resulted in no change in xylose uptake and demonstrated that there are uncharacterized pentose transporters

that facilitate a significant amount of sugar uptake, especially for arabinose (Figure 3.1). We then inactivated glucose kinase in this strain to prevent catabolism of glucose for transporters acting preferentially on this sugar. By fermenting this strain, EG51A, in the presence and absence of glucose while overexpressing *glf,* we find that overexpression of *glf* results in complementation of the xylose utilization rates to be equivalent to AF11A (Figure 3.2). However, xylose transport/catabolism catalyzed by Glf is almost completely inactive in the presence of glucose, and thus cells consume a similar amount of xylose as those carrying an empty vector (Figure 3.2). In the absence of glucose, cells have a greater than 2-fold increase in biomass at 24 h, thus we determined this co-sugar fermentation to be an ideal selective condition for enrichment (Figure 3.2).

Identification of Glf variants resistant to glucose

To create variants of Glf, we took a two pronged approach employing both random and site directed mutagenesis. For random mutagenesis, two different variant pools of ~12,000 variants were created using error prone PCR (ePCR). To create a site saturation mutagenesis library, we specifically targeted sites that either were involved in the coordination of glucose, but not xylose in the XylE crystal structure. By performing a an alignment of XylE, and Glf using the Clustal Omega algorithm (32) and constructing a homology model of Glf we selected residues that corresponded to glucose coordinating residues Q175, F383, and G388 in XylE in Glf (A165, F374, and G379) (Figure 3.3). We also chose I172 in between the glucose binding residues I171 and Q175 (in XylE) due to the finding that mutation of this corresponding residue was capable of releasing glucose repression in multiple yeast uniporters (30, 33). This resulted in targeting V162 in Glf as well for site saturation mutagenesis (Figure 3.3). We also constructed and tested two variants that had been previously identified to enhance pentose utilization, Glf-A18T and Glf-V275F (26). Following the creation of these mutant constructs we pooled all site saturation mutagenesis constructs into a single tube at equimolar amounts to create a pool. Each pool was then transformed into EG51A to at least 5x coverage and immediately transferred from plates into fermentation media supplemented with a mixture of glucose and xylose. By transferring these fermentation cultures into a new vessel with fresh media every ~24h we enriched for mutants that could transport xylose

in the presence of glucose. This resulted in the identification of 10 unique variants after enrichment, in addition to the previously described mutants (Figure 3.4A). We retransformed these variants into EG51A to eliminate potential beneficial mutations gained during enrichment and fermented them in the same media used for enrichment. Of these, nine variants demonstrated enhanced xylose utilization by 60-210% in the presence of glucose, suggesting glucose inhibition was relieved (Figure 3.4B). GK14, GK3 and GK7 appeared to have the highest xylose consumption rate in the presence of glucose with ~190-210% increases in consumption compared to wildtype Glf (Figure 3.4B). GK2 and SM1 had the lowest increase in xylose utilization rate, with only a 60-80% increase in utilization between 0 – 24h (Figure 3.4B). Variants A18T, V275F, and GK6 did not release Glf from glucose inhibition, and thus were not analyzed further (Figure 3.4B).

Glf variants have enhanced preference for xylose transport

To test if the in-vivo substrate preference of these transporters was altered we tested each variant in EG29 to permit simultaneous consumption of glucose and xylose in a co-sugar fermentation. Simply expressing wildtype Glf in the transport deficient strain greatly increased both glucose and xylose utilization, indicating that the transporter still does actively transport xylose during glucose import (Figure 3.4C). However, cells expressing Glf variants had an increased xylose utilization rate relative to a strain expressing the wildtype Glf, ranging from a 22% - 74% increase (Figure 3.4C). Of these GK14 and SM1 had the lowest increases in xylose utilization of approximately 20%. The best performing variants corresponded to SM3 and GK8 for total xylose utilization in 24h, with an increase of approximately 70% each (Figure 3.4C). Additionally, the substrate preference of these transporters appeared to be altered. Cells expressing wildtype Glf consumed on average approximately 0.7 g xylose per g glucose (Figure 3.5). All variants had increased this ratio to be between 0.9 - 1.4 g xylose per g glucose. The lowest of these was SM1 who only had a 0.9 g xylose per g glucose ratio (Figure 3.5). This suggests that SM1 is only partially relieved of glucose inhibition of xylose transport, as was seen previously (Figure 3.4B). Other mutants such as GK1 and GK4 also had lower sugar preference for xylose, suggesting incomplete release for glucose inhibition or sub-optimal xylose transport kinetics (Figure 3.5). Variants GK14,

GK2, GK3, GK7, GK8, and SM3 all appeared to have the greatest relief from glucose inhibition with a 1.3 – 1.4 g xylose per g glucose ratio (Figure 3.5). According to these fermentation results and structural investigations, we determined GK3, SM3, and GK8 to be the most ideal xylose uniporters to investigate further since they maintain high preference and kinetics for xylose and are composed of only 1-3 amino acid changes.

Structural locations of amino acid mutations suggest multiple mechanisms of overcoming glucose inhibition

To gain more information on the potential structural consequences of these mutations, we constructed a homology model of Glf. Since the crystal structure of XylE was recently solved with both glucose and xylose bound and has a high degree of identity to Glf (~43%) we selected this structure as a template (PDB ID: 4GBY). The structure of Glf appears to consist of twelve TMS, two extracellular helices (EC), five intracellular helices (IC), in addition to one kinked region of TM1 that may reside on the extracellular side (TM1e) (Figure 3.6A). Similar to XylE most putative sugar coordinating residues are located in the C-terminal region of the transporter. However, variants released from repression were found to have a wide range of mutations with no clear bias for enrichment of mutations in the N- or C-terminal domain (Figure 3.6B). We investigated if mutations in TM helices that are known to coordinate the 6-hydroxymethyl group of D-glucose in XylE are enriched in the nine variants. These include TM5, TM8, and TM10. Variants in these locations account for 66.6% of variants suggesting that this region is important. Mutations in TM5 and TM8 were the most prevalent with 44.4% and 33.3% of variants having mutations in these features, respectively (Figure 3.6B). Notably, almost all residue contacts in XylE for xylose or glucose recognition are conserved between XylE and Glf (Figure 3.6C). The only exception is Q175 (A165 for Glf) which is potentially capable of forming a hydrogen bond with the 6-hydroxyl group of glucose in XylE (Figure 3.3). Mutant SM3 has a mutation in A165 (A165M), which is predicted to extend directly into the sugar coordinating pocket near the 6-hydroxymethyl group of glucose. Since this is a larger side chain than alanine, this may potentially hinder glucose from preferentially entering the binding pocket. Besides SM3 multiple other mutants were isolated with mutations that

potentially alter the size or structure of the binding pocket. Among the many mutations in GK7 is a mutation in a phenylalanine (F374S) that is involved in glucose coordination and is structurally conserved in Glf (Figure 3.3 and 3.7). Additionally, two variants (GK1 and GK8) had mutations in both a conserved asparagine residue in TM8 of sugar porters (N316→S/D) and a C-terminal lysine (K458→R/I) residue. This asparagine presumably protrudes into the binding pocket and although not directly involved in binding, has been shown to alter sugar substrate specificity in multiple transporters. Similarly to previous reports, the N316 side chain of Glf appears to extend into the translocation pocket near the 6-hydroymethyl coordinating region of glucose in our model, and thus mutations in this residue may alter the size or structure of the pocket (30) (Figure 3.7). The C-terminal lysine residue however appears to be positioned directly below the cytoplasmic extending junction between TM8 and TM9 which may be indicative of an interaction between this residue and residues in TM8-9. It is not clear if these two mutations are needed however to permit release of glucose repression. Other mutants had less obvious mutations that may be involved in interactions difficult to predict. GK2 had a mutation in TM8 (G313S) that is homologous to a described mutation in Glut1 (G314S) known to cause decreased glucose transport without affecting cation permeability (34) (Figure 3.8). Based on our model, this residue extends towards TM5 near the binding cavity (Figure 3.7). SM1 was isolated with mutation (V162G) that is positioned similarly in TM5 extending towards TM8 (Figure 3.3 and 3.7). These two variants may also potentially affect the size or structure of the translocation pocket. The L445I mutation in GK3 is possibly one of the most subtle mutations with no obvious structural consequences. On the basis of the homology model, this residue points outward from the protein core on TM12 and does not have any obvious interactions that may be disrupted by a mutation to isoleucine (Figure 3.7). No known mutations affecting glucose transport have been described for any of the variations in GK4. The likely causative mutations of these variants seem to point to two main potential mechanisms 1) alteration of the translocation pore, particularly in the area around residues coordinating the 6-hyroxymethyl group of D-glucose or 2) unknown interactions in the C-terminus (Figure 3.7).

Effects of Glf on co-sugar to succinate conversion

Our lab previously created AG055, a strain capable of presumably conserving energy during xylose to succinate fermentation (Chapter 2). This ATP conservation results in an increase in fermentative production using glucose/xylose mixtures as a substrate as well. However, it is unknown if additional ATP conservation may further increase succinate titer and yield by decreasing the need for generation of unwanted side products, such as acetate. Thus, we fermented AG055 overexpressing an empty vector, wildtype Glf or variant GK3 for 120h. AG055 overexpressing GK3 has an approximately 24h lag in growth and production compared to the other strains (Figure 3.9). This may be due to a sub-optimal xylose utilization pathway, as GK3 may preferentially transport xylose but release of catabolite repression is incomplete in this strain, as it does not possess *xylR**. After 120h the titer and yield of all strains is approximately comparable (Figure 3.9).

Discussion

For effective lignocellulose bioconversion, biocatalysts need to be able to efficiently transport and utilize the main sugars present in woody biomass, glucose and xylose. Previous work has largely focused on the issue of CCR, which now has multiple robust solutions, while ignoring issues in energetic expenditures during the uptake of xylose using ABC transporters like XylFGH (1, 3). Although the use of xylose uniporters would be more efficient, these proteins are known to be inhibited by glucose thus constraining the effectiveness of overexpression of heterologous transporters in co-sugar mixtures. In this work, we sought to create a glucose-resistant xylose facilitator, Glf, to overcome sub-optimal transport mechanisms in the presence of sugar mixtures. By deleting all known glucose/xylose uptake proteins and inactivating glucose kinase, we were able to create a strain that confers a selective growth advantage for xylose transporters resistant to glucose. This resulted in the isolation of nine Glf variants that were at least partially released from glucose inhibition and had an enhanced preference for xylose uptake compared to the wildtype protein. This is likely due to alterations in the sugar translocation pocket that are responsible for coordinating the 6-hydroxymethyl group of D-glucose, although it appears other more complex mechanisms may also be present. These findings suggest that these proteins may be useful for

incorporation into strain engineering strategies to help conserve ATP during lignocellulose bioconversion.

In general, many MFS transporters are known to be inhibited by glucose, including the *E. coli* xylose symporter XylE (21, 29). Previous work has looked at releasing this inhibition mainly in eukaryotic transporters in fungi, since it is known that xylose catabolism can be bottlenecked by transport in these organisms (35, 36). Our work finds that many of the residues used to release glucose inhibition in these eukaryotic uniporters also works to release inhibition in Glf, suggesting broadly conserved mechanisms of inhibition. For instance, in fungal MFS transporters Gxs1 (from *Candida intermedia),* Gal2, Hxt7, Hxt5, Hxt11, and Hxt3, mutation of the conserved asparagine in TM8 (Glf-N316) has been observed to release glucose inhibition, with varying success (24, 30, 33, 37). Similar to our observation, this residue has been predicted to release glucose inhibition in variants GK1 and GK8 by potentially altering the structure of the sugar coordinating pocket, especially near the 6-hydroxymethl group of D-glucose. We also find that Glf-V162G partially releases glucose repression, a discovery that is corroborated by findings that mutations in the corresponding residue in the Gxs1, Gal2, Hxt7 and Hxt5 results in a similar effect (30, 33). This work also finds that a mutation of Glf-A165M, which is normally a glucose coordinating residue in other proteins (such as XylE) can be used to release glucose inhibition while maintaining high xylose utilization efficiency *in vivo*. Other mutations we find that may have a role in releasing glucose inhibition mainly seem to be located in a hotspot of TM5, TM8 and TM10, all of which are potentially involved in disruption of the coordination of the 6-hydroxymethyl group of glucose (Figure 3.3, 3.6, and 3.7). This seems to suggest that disruption of the interactions coordinating the unique molecular side chain of D-glucose relative to D-xylose is a common mechanism to release glucose inhibition. Other mutants are still present however that do not appear to have a direct role in this mechanism such as GK3 (Glf-L445I). Work by Li et al has observed that truncation of the C-terminal region of Gxs1 can decrease glucose inhibition and can be combined with a mutation of the conserved asparagine in TM8 for an additive effect (33). This may further suggest that there at

two distinct mechanisms which can release glucose inhibition, or that this subtle mutation might be capable of creating more drastic structural changes than may be predicted.

It has been observed that incorporation of these mutations only enhances co-utilization significantly when strains have been released from CCR, emphasizing the importance of this quality in a testing strain (31). Utilization of a strain free from CCR allowed us to probe the efficacy of these transporter variants without being constrained by metabolic regulation. This allowed us to discover that multiple of these transporters such as GK14, GK2, GK3, GK7, GK8, and SM3 actually prefer xylose to glucose now when given the option for consumption. Another interesting finding is that none of these mutations disrupt glucose utilization. A matter of fact, all of the mutants except GK14 appear to retain relatively wildtype glucose utilization rates. This finding further provides evidence that the mechanism of glucose inhibition is independent of transport function. This is corroborated by findings by Young et al that eliminating glucose transport but enabling xylose transport does not simply release glucose inhibition in Gxs1 (22). This is also observed in XylE, which is inhibited by glucose even though it does not actively transport the molecule (29). Farwick et al was able to show however that by changing the conserved aspargine in TM8 to a phenylalanine, glucose transport can be blocked presumably by hindering entrance in the binding pocket while still releasing glucose inhibition (30). One interesting idea might introduce a mutation of N316F into mutant GK3 to attempt to decouple glucose and xylose transport and thus create a more specific xylose uniporter.

Since it is known that ATP conservation during microaerobic fermentation of xylose is important for efficient bioproduction (11, 13), we tried to overexpress one mutant transport (GK3) in an optimized succinate producer, AG055, to attempt to increase production. However, this strain already gained a mutation that activated energy conserving pathways, making it unclear whether ATP conservation is still a metabolic bottleneck. Our results suggest that ATP conservation may not be a bottleneck anymore in this strain, but also that strains may require a *xylR** mutation to effectively use these glucose-resistant uniporters. This is due to the 24h lag that is observed upon overexpression of GK3, possibly due to a lack of adequate expression of xylose catabolism genes (Figure 3.9). Also, these cells still retain a native copy of XylFGH, which may interfere with energy

conservation depending on the amount of xylose that is taken up by the transporter. More ideal to test our hypothesis may be to overexpress these mutant transporters in the precursor strain KJ122 with a *xylFGH* inactivation and *xylR** mutation to eliminate influence of XylFGH on energy conservation and enable efficient co-utilization.

In conclusion, this work describes multiple variants of Glf that represent effective bacterial xylose uniporters that can be used in the presence of mixed sugars, similar to lignocellulose hydrolysate. The exact causative mutations of some mutants such as GK1, GK2, GK7, GK8 and GK14 remain unknown, however we hypothesize that many of them may be involved with mutations that alter the structure of TM5, TM8, and TM10 coordinating the 6-hydroxymethyl group of glucose. Future directions may improve this work by combining mutations found in this work, especially those targeting potentially distinct mechanisms (e.g. N316 and L445). Additionally, this work could be expanded to further characterize the single point mutations required to release glucose inhibition in some of these variants, and to further characterize the molecular mechanisms which lead to glucose inhibition of MFS transporters.

Materials and Methods

Strains and Media

All strains used in this work are listed in Table 3.1. All genetic manipulations were performed by growing cells in Luria Broth (10 g l^{-1} Difco tryptone, 5 g l^{-1} yeast extract, and 5 g l^{-1} NaCl). Cells in liquid culture were grown at 30°C, 37°C, or 39°C with rotation at 180 rpm. During genetic manipulations, 5% arabinose (w/v) was added to induce recombination and ampicillin (100 mg l^{-1}), chloramphenicol (50 mg l^{-1}), or kanamycin (50 mg l^{-1}) were supplemented as needed.

Genetic Methods

Gene inactivation was performed using either one-step inactivation or two-step recombination as previously described (38, 39). After cells were transformed with pKD46, 10-12 colonies were inoculated into 25 ml activation media incubated until OD_{550} reached 0.3-0.5. Cells were washed three times with nanopure water before being electroporated with ~100 ng of linear DNA for

integration into the region of interest. Linear DNA for integration was amplified from pKD4 for primary integration (FRT-*kan*-FRT) or pXW01 (*cat-sacB*) for secondary integration. All linear DNA fragments were flanked by 50 bp of homology to the region of interest. The *kan* cassette was remove from integrated cells by transformation with pCP20. The *cat-sacB* cassette was removed by repeating the integration with a linear fragment created using Fusion PCR to encode ~1 kb of homology to the flanking regions (40). Cells were then incubated in LB supplemented with 10% (w/v) sucrose to select for positive integrations. All genetic manipulations were verified using colony PCR. All primers can be found in Table 3.2.

Plasmid variant pool construction and enrichment

Circular polymerase extension cloning (CPEC) (41) was used to create all plasmids in Table 1. All primers used are in Table 2. The *glf* gene was amplified from *Zymomonas mobilis* CP4 with ~20 bp homology overhangs for a linearized pTrc99a backbone. To mutagenize *glf* we performed error-prone PCR using the GeneMorph II (Agilent), according to the manufacturer instructions to create two separate libraries (~20,000 variants) with low (0-3 mutations/kb) and medium (4-7 mutations/kb) mutation frequencies. To generate mutants for site saturation mutagenesis, we used Fusion PCR to generate NNK variants at codons that corresponded to Glf162, 165, 332, and 374. Glf variants relieved of glucose inhibition were enriched by transforming one of the three plasmid variant pools (low, medium, NNK) into EG51A and fermenting in TB2 (publication in preparation) supplemented with 6.6% glucose (w/v) and 3.4% xylose (w/v). Fermenting cells were transferred into fresh media every ~24h for approximately 20 generations before being plated onto LB agar supplemented with ampicillin and isolated for individual colonies. Plasmid was isolated from these variants and Sanger sequencing used to confirm unique variants. All unique variants were retransformed into selected strains for further testing.

Fermentation

Colonies were incubated for ~18 h under microaerobic conditions (gassed with argon) on LB agar before being transferred into 100 ml LB in supplemented with 100 mM MOPS in 250 ml flasks.

These seed flasks were incubated for ~18h before being transferred into 300 ml AM1 (single sugar fermentation) (42) or TB2 (co-sugar fermentation), supplemented with 10% total (w/v) of the appropriate sugar(s) in 500 ml fermentation vessels (37°C, 150 rpm). When performing co-sugar fermentations a ratio of 6.6% glucose and 3.4% xylose was used. Initial inoculum for fermentation was always OD_{550} = 0.05. All fermentations had pH controlled at 7.0 with automatic additions of 6M potassium hydroxide.

Homology modeling and structural analysis

Homology models of Glf were constructed using SWISS-MODEL (43) with the crystal structures of XylE bound to glucose (PDB ID= 4GBZ) or xylose (PDB ID= 4GBY) as template. Structures were viewed and edited using the Chimera software suite (44). The location of predicted structural features was overlaid onto an Clustal Omega alignment of Glf and XylE. Visualizations of the alignment similarity and identity were generated using the Multiple Alignment Show tool from the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/).

Analysis

Quantification of sugars was performed using high-performance liquid chromatography (HPLC) using an Ultimate3000 equipped with an Aminex HPX-87H column (Bio-Rad). As described previously (6), we used 4mM sulfuric acid as mobile phase at a flow rate of 0.4 ml $minute^{-1}$. Optical density was quantified using a Beckman Coulter DU 730 spectrophotometer.

Figures.

Table 3.1. Strains and plasmids used in Chapter 3

	Relevant characteristics	Reference
Strains		
Top 10F'	F´lacIq, Tn10(TetR) mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG	Pharmacia
ATCC 9637	Wildtype	(45)
AF11A	ATCC 9637 Δ*ptsI*, Δ*ptsG*, Δ*galP, xylR::xylR**	(In prep)
EG21	AF11A Δ*xylFGH*	This study
EG23	AF11A Δ*xylFGH*, Δ*xylE*	This study
EG25	AF11A Δ*xylFGH*, Δ*xylE*, Δ*araE*	This study
EG27	AF11A Δ*xylFGH*, Δ*xylE*, Δ*araE*, Δ*araFGH*	This study
EG29	AF11A Δ*xylFGH*, Δ*xylE*, Δ*araE*, Δ*araFGH*, Δ*gatC*	This study
EG51A	EG29 *glk*::*kan*	This study
KJ122	ATCC 8739, *pck**, *ptsI**, Δ*ldhA*, Δ*adhE*, Δ*ackA*, Δ(*focA-pflB*) Δ*mgsA*, Δ*poxB*, Δ*tdcDE*, Δ*citF*, Δ*aspC*, Δ*sfcA*	(39)
AG055	KJ122 Δ*galP*	(In prep)
Plasmids		
pKD46	bla, γ β exo (Red recombinase)	(38)
pCP20	*bla, cat,* FLP+, ts-rep, cI857λts	(46)
pKD4	bla, FRT-kan-FRT	(38)
pXW001	*cat-sacB* cassette with the *sacB* native terminator cloned into a modified vector pLOI4162	(6)
pTrc99A	P*trc bla oriR rrnB lacI*^{q}	(47)
pETG1	*glf* from *Z. mobilis* ZM4 in pTrc99A	This study

Table 3.2. Primers used in Chapter 3

Primer	Sequence (5' to 3')
xylFGH deletion	GCTGGTCAAAATAGGCTTGC
	GACCATAAAAAGCAAAGCGG
	GTGATTGTTACTTATTAAAGCTGTCCTCTAACTACAGAAGGCCCTACACCGTGTAGGCTGGAGCTGCTTC
	AACTCAAAACCGGTAATACGTAACCGGCTTTGAGAAAATTTTTATCAAAACATATGAATATCCTCCTTAGT
xylE deletion	TCACCATCGTCTTCTTGCTG
	ACTCAATCTTGGCCTGCTGT
	CCAACATCAATGCACTGATAAAAGATCAGAATGGTCTAAGGCAGGTCTGAGTGTAGGCTGGAGCTGCTTC
	GGAGTAAAAAGACAGCACGCCGAACTGGCGCGGCGTGCTGGACAGGAAGACATATGAATATCCTCCTTAGT
araE deletion	CCTCCAGCGCTTCAGATTAT
	AGAAAGGCAGAGTGCAACGT
	ATTGTTCACGTATTTTTTCACTATGTCTTACTCTCTGCTGGCAGGAAAAAGTGTAGGCTGGAGCTGCTTC
	CTCTATTAACGAAAAAAGGGCCGGATGTACAGCACATCCGGCCCGTGAAACATATGAATATCCTCCTTAGT
araFGH deletion	AATGACGTGCATTTTCCACA
	GAAGCGATAAGGTTTTGCCA
	TGTCATTCGTTTTTGCCCTACACAAAACGACACTAAAGCTGGAGAGAACCGTGTAGGCTGGAGCTGCTTC
	GGCGGCTGGCTGTGGTGGGAAAAAACGTTAAATTGTTGTGGAAAAAAGCACATATGAATATCCTCCTTAGT
gatC deletion	CCCGATATCGTCGAGACACT
	CAATAAAAGCCCCATCCTCA
	GTATCGAAGCATTACAAAATAAAATTCTGACTATCTTACAGGGGTGACCTGTGTAGGCTGGAGCTGCTTC
	CATAAAAACTCCTGATTGTTAAGGGGGATAACCTCCCCCTCAGGTAAAAACATATGAATATCCTCCTTAGT
glk deletion	ACCGATGATCCAGAGAGGTG
	TAATTGCGTCCTGCAAAGTG
	ATTTACAGTGTGAGAAAGAATTATTTTGACTTTAGCGGAGCAGTTGAAGAGTGTAGGCTGGAGCTGCTTC
	TGATTTAAAAGATTATCGGGAGAGTTACCTCCCGATATAACAGGAAGGATCATATGAATATCCTCCTTAGT
glf cloning into	GGATCCTCTAGAGTCGACCTGCAGG
pTrc99a	CGGGTACCGAGCTCGAATTCCATG
	CATGGAATTCGAGCTCGGTACCCGGCCATGAGTTCTGAAAGTAGTCAGGG
	CCTGCAGGTCGACTCTAGAGGATCCCTACTTCTGGGAGCGCCACATC
A18T mutagenesis	CGCGACTAGCCCTAATCGCTACCATAGGCGGCTTGCTTTTCGGT
	ACCGAAAAGCAAGCCGCCTATGGTAGCGATTAGGGCTAGTCGCG
V275F	CCGTTGTTTTTGCCGGTGTATCCTTTGCTGCCTTCCAGCAGTTAGTCG
mutagenesis	CGACTAACTGCTGGAAGGCAGCAAAGGATACACCGGCAAAAACAACGG
162NNK	CAGCAGATGGCCATTNNKACGGGTGCTTTAACC
mutagenesis	GGTTAAAGCACCCGTMNNAATGGCCATCTGCTG
165NNK	GCCATTGTGACGGGTNNKTTAACCGGTTATATC
mutagenesis	GATATAACCGGTTAAMNNACCCGTCACAATGGC
332NNK	GTTGACCGCTTCGGCNNKAAACCTCTGCTTATT
mutagenesis	AATAAGCAGAGGTTTMNNGCCGAAGCGGTCAAC
374NNK	CTTTATATTGCAGTCNNKGGTATGTCATGGGGC
mutagenesis	GCCCCATGACATACCMNNGACTGCAATATAAAG

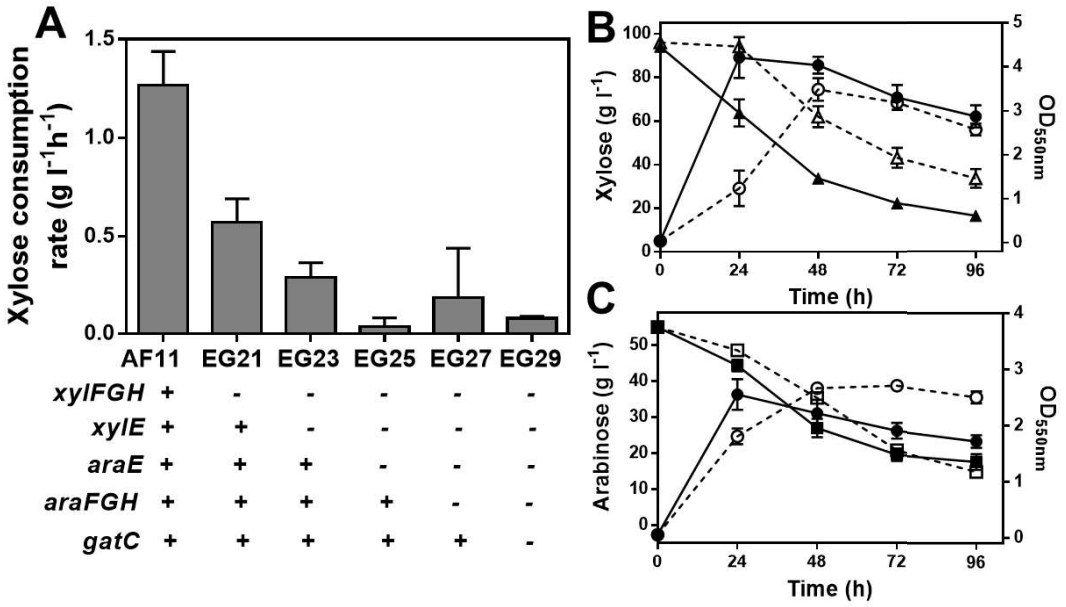


Figure 3.1. Construction of a screening platform deficient in both glucose and xylose transport. A) Xylose consumption rates of AF11 and its derivatives engineered by consecutive inactivation of known and putative xylose transporters during the initial 24 hours fermentation. Fermentation of AF11 (solid lines with filled symbols) and EG29 (dotted lines with open symbols) using AM1 media supplemented with B) 100 g/L xylose or C) 50 g/L arabinose. Symbols for all: OD550nm (circle), xylose (triangle), arabinose (square).

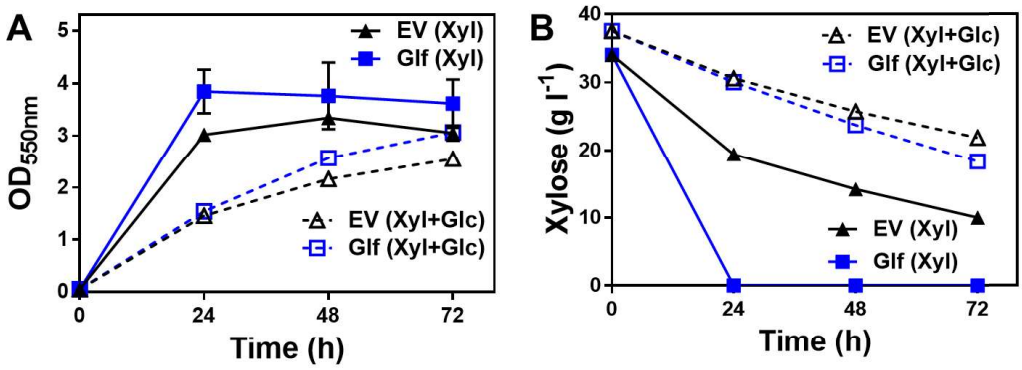


Figure 3.2. Xylose transport of Glf is inhibited by glucose. Fermentation of EG51A (EG29 *glk*$^{-}$) transformed with empty vector (EV) pTrc99A (triangle symbols; black lines) or pTrcGlf (square symbols; blue lines) using AM1 media containing 34 g/L xylose only (solid lines and filled symbols) or sugar mixtures (dotted lines and open symbols) containing 66 g/L glucose and 34 g/L xylose, as well as 10 μM IPTG to induce *glf* expression. A) Cell optical density (OD550nm), and B) xylose concentrations were determined.

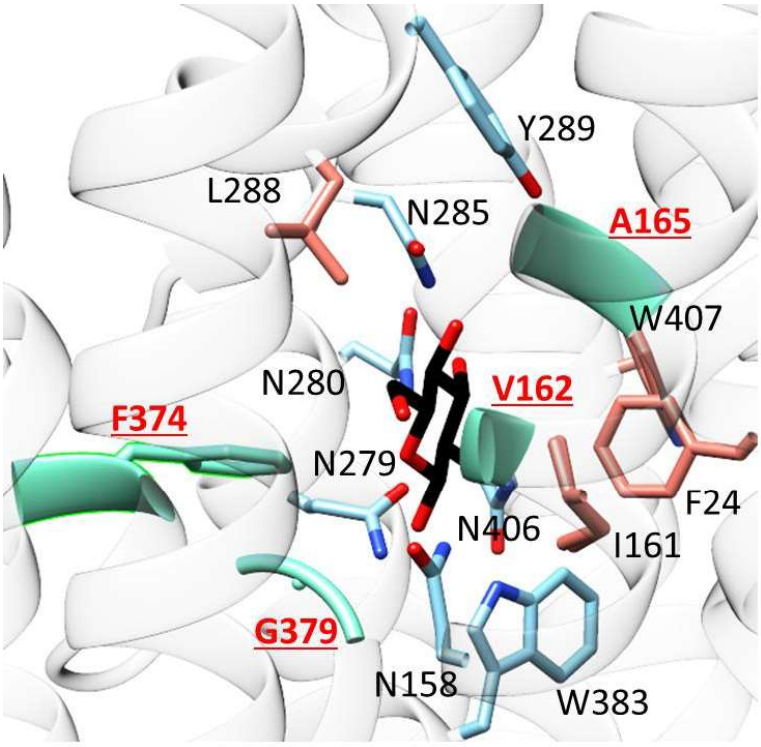


Figure 3.3. Glucose coordination of Glf. Contacts forming hydrogen bonds (residues in sky blue) or hydrophobic contacts (residues in salmon) with D-glucose (black) according to the XylE crystal structure (PDB ID: 4GBZ) are displayed in the image. All glucose coordinating sites except A165 are conserved in the D-glucose binding pocket of XylE. Residues V162, A165, F374 and G379 were selected for site saturation mutagenesis to potentially disrupt glucose coordination are in aquamarine.

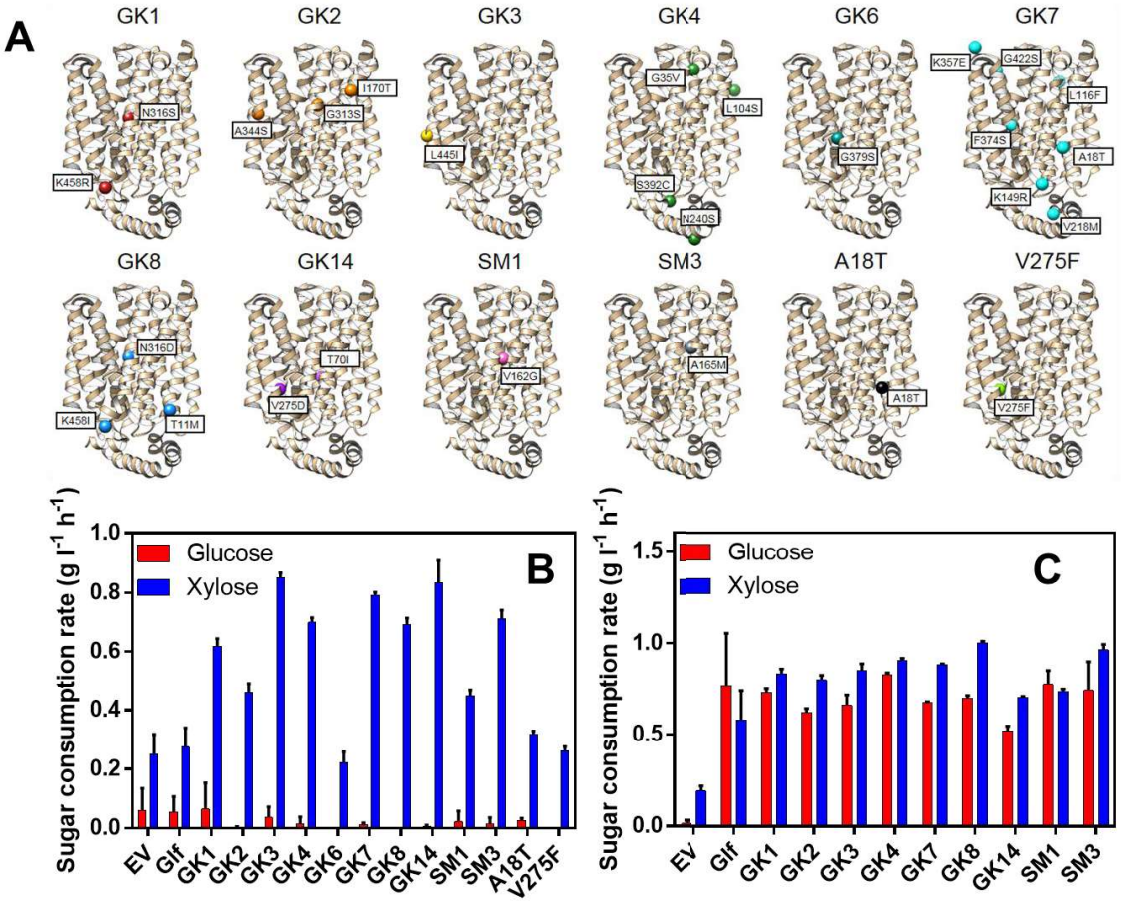


Figure 3.4. A) Glf variants isolated after fermentation enrichment in TB2 supplemented with 6.6% glucose and 3.4% xylose (w/v). Sugar consumption rate (0-24 h) of B) EG51A or C) EG29 overexpressing Glf variants fermented in TB2 supplemented with 6.6% glucose and 3.4% xylose was evaluated(w/v).

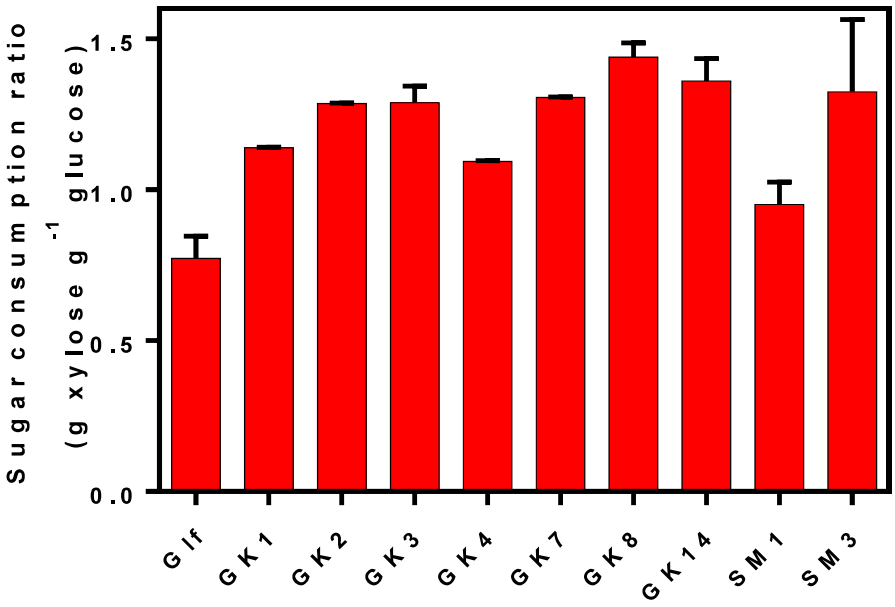


Figure 3.5. Ratio of sugars consumed between 0-24h of sugar transport deficient *E. coli* (EG29) expressing Glf variants fermented in TB2 supplemented with 6.6% glucose and 3.4% xylose (w/v).

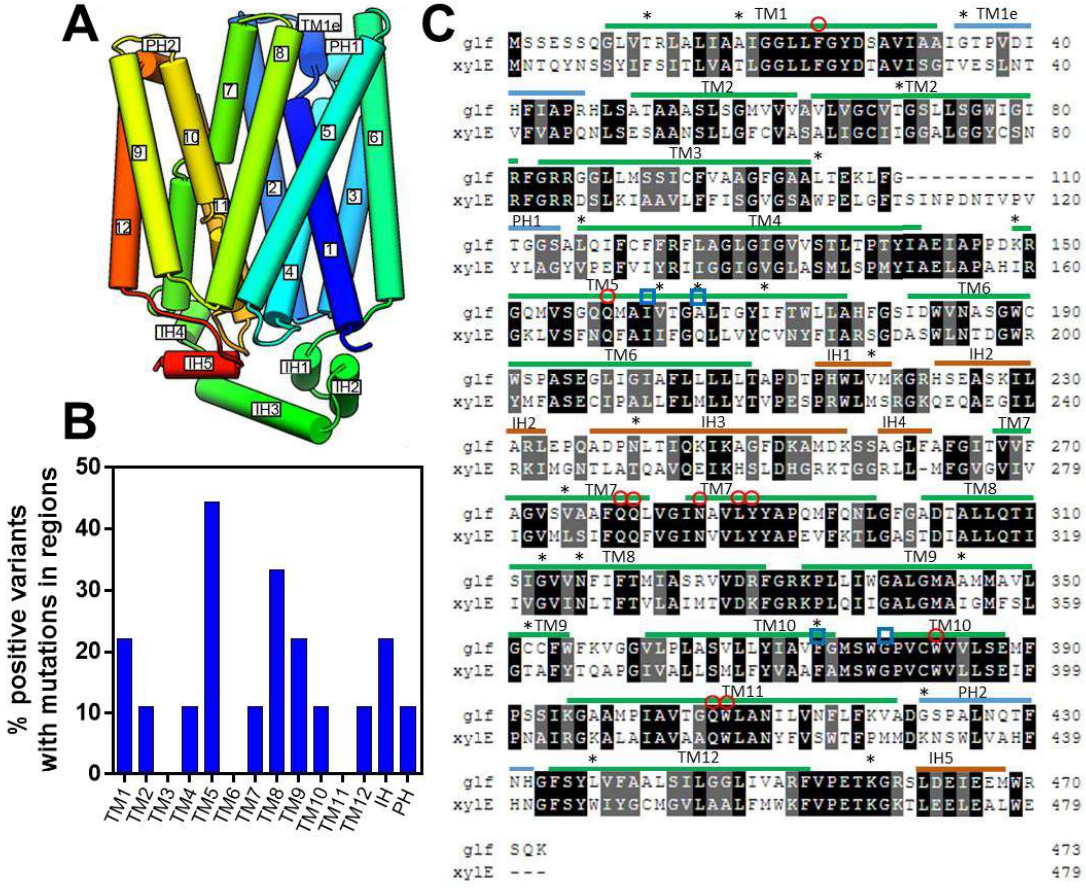


Figure 3.6. Structural information for mutagenesis of Glf. A) A homology model of Glf depicts a similar structural organization to XylE consisting of 12 TMS, 2 periplasmic helices (PH), and 5 intracellular helices (IH). B) Positional distributions of point mutations from identified positive Glf variants among the structural components in the homology model. C) Protein sequence alignment of Glf and XylE using Clustal Omega with mutations in positive Glf variants labeled (*), glucose coordinating (□), and glucose/xylose coordinating (○) residues highlighted.

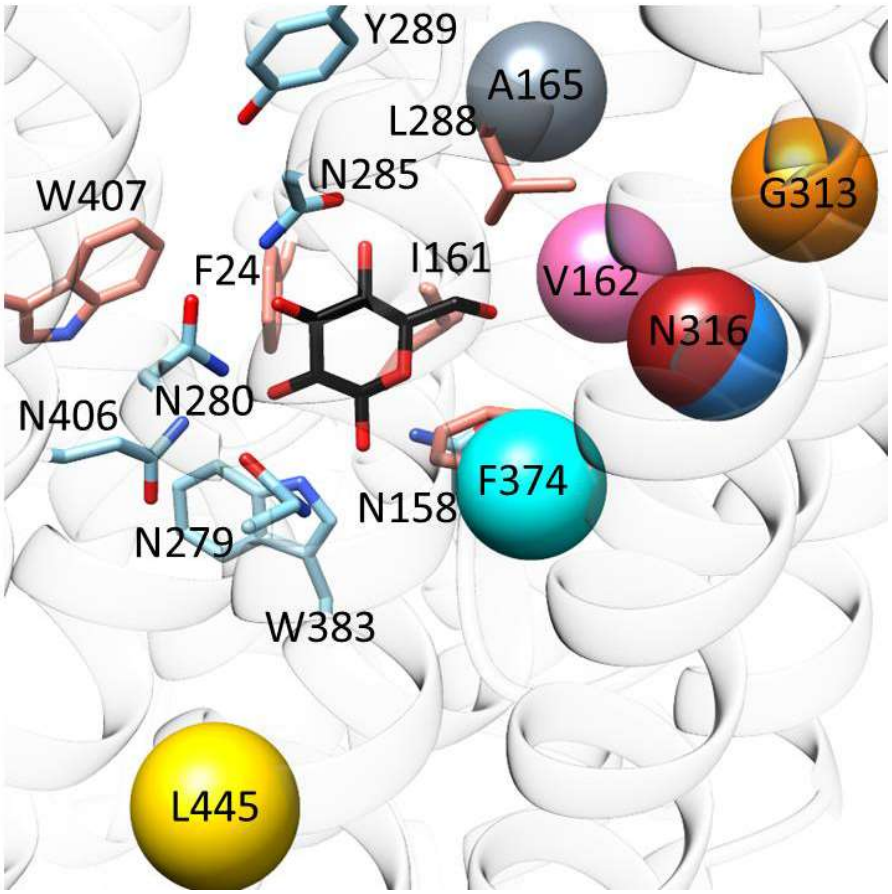


Figure 3.7. Predicted or causative mutations of variants that enhance xylose consumption in glucose/xylose mixtures.

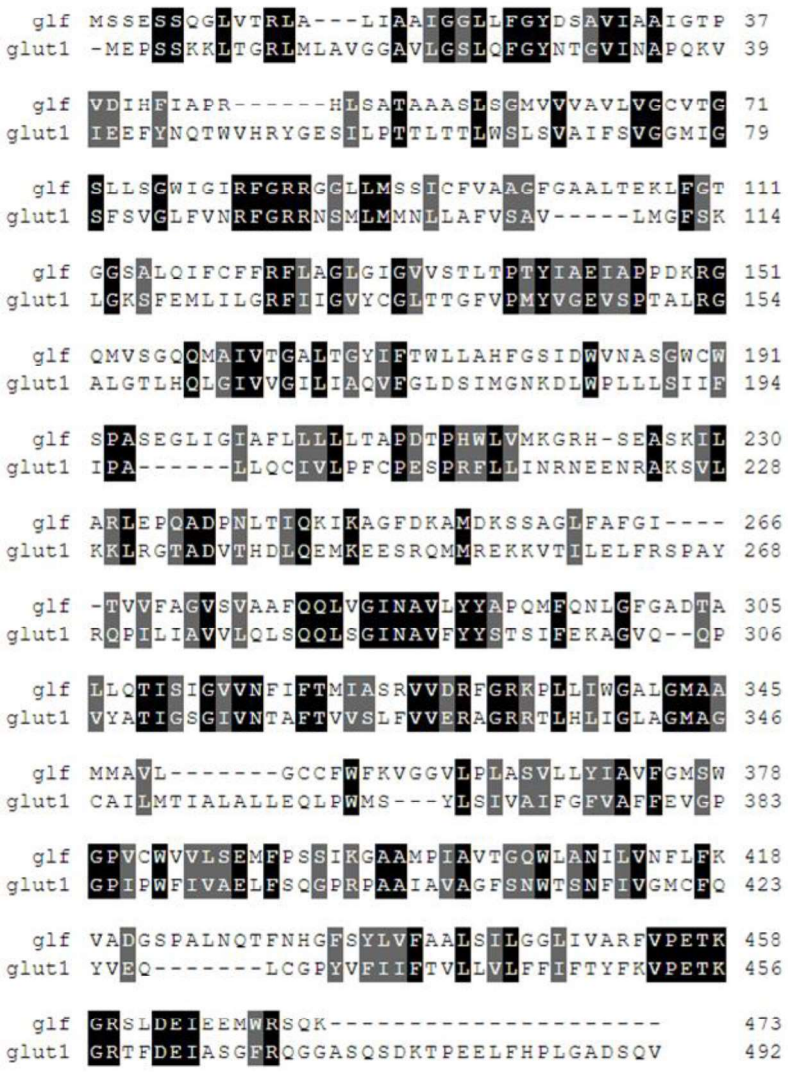


Figure 3.8. Protein alignment of Glut1 and Glf using Clustal Omega

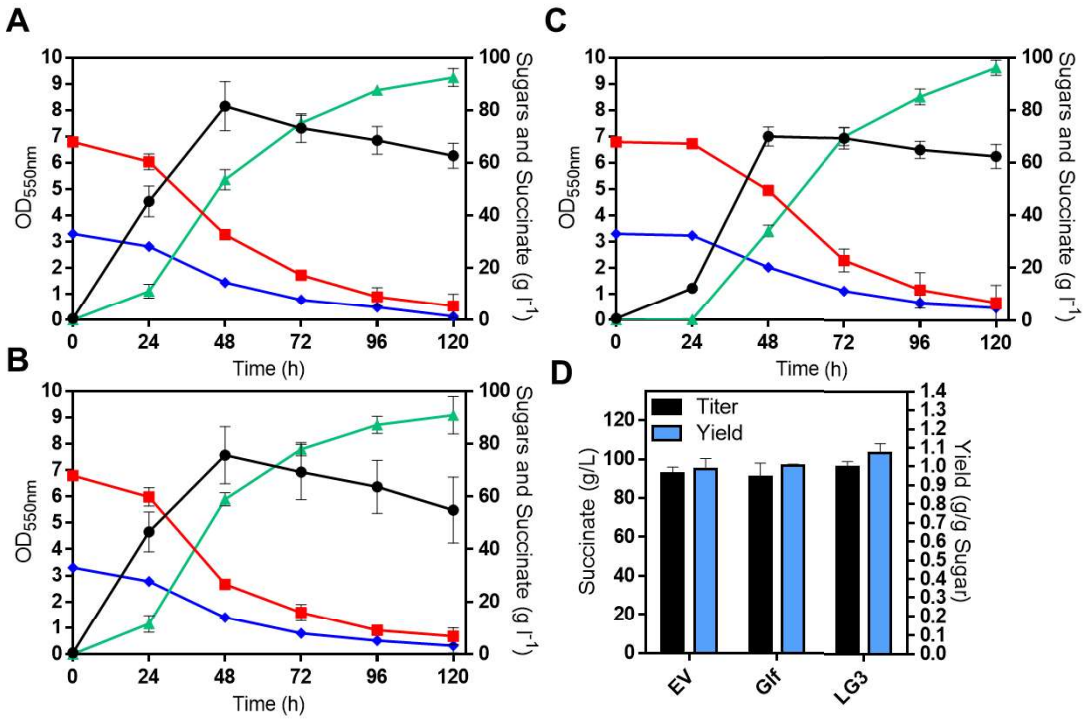


Figure 3.9. Fermentation of AG055 in TB2 supplemented with 6.6% glucose 3.4% xylose and 100mM potassium bicarbonate overexpressing either A) an empty vector B) pETG1 or C) pGK3. Optical density (●), glucose (■), xylose (♦), and succinate (▲) were measured over time and D) titer and yield were compared at 120h.

References

1. Nieves LM, Panyon LA, Wang X. 2015. Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. Front Bioeng Biotechnol 3:17.

2. Mishra P, Singh A. 1993. Microbial pentose utilization. Adv Appl Microbiol 39:91-152.

3. Gorke B, Stulke J. 2008. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol 6:613-24.

4. Deutscher J. 2008. The mechanisms of carbon catabolite repression in bacteria. Curr Opin Microbiol 11:87-93.

5. Kim JH, Block DE, Mills DA. 2010. Simultaneous consumption of pentose and hexose sugars: an optimal microbial phenotype for efficient fermentation of lignocellulosic biomass. Appl Microbiol Biotechnol 88:1077-85.

6. Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, Wang X. 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc Natl Acad Sci U S A 114:7349-7354.

7. Henderson PJ. 1990. Proton-linked sugar transport systems in bacteria. J Bioenerg Biomembr 22:525-69.

8. Khankal R, Chin JW, Cirino PC. 2008. Role of xylose transporters in xylitol production from engineered Escherichia coli. J Biotechnol 134:246-52.

9. Jojima T, Omumasaba CA, Inui M, Yukawa H. 2010. Sugar transporters in efficient utilization of mixed sugar substrates: current knowledge and outlook. Appl Microbiol Biotechnol 85:471-80.

10. Gonzalez JE, Long CP, Antoniewicz MR. 2017. Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by (13)C metabolic flux analysis. Metab Eng 39:9-18.

11. Hasona A, Kim Y, Healy FG, Ingram LO, Shanmugam KT. 2004. Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of Escherichia coli on xylose. J Bacteriol 186:7593-600.

12. Gosset G. 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb Cell Fact 4:14.

13. Utrilla J, Licona-Cassani C, Marcellin E, Gosset G, Nielsen LK, Martinez A. 2012. Engineering and adaptive evolution of Escherichia coli for D-lactate fermentation reveals GatC as a xylose transporter. Metab Eng 14:469-76.

14. Zhang X, Jantama K, Moore JC, Jarboe LR, Shanmugam KT, Ingram LO. 2009. Metabolic evolution of energy-conserving pathways for succinate production in Escherichia coli. Proc Natl Acad Sci U S A 106:20180-5.

15. Wei LN, Zhu LW, Tang YJ. 2016. Succinate production positively correlates with the affinity of the global transcription factor Cra for its effector FBP in Escherichia coli. Biotechnol Biofuels 9:264.

16. Tan Z, Zhu X, Chen J, Li Q, Zhang X. 2013. Activating phosphoenolpyruvate carboxylase and phosphoenolpyruvate carboxykinase in combination for improvement of succinate production. Appl Environ Microbiol 79:4838-44.

17. Khunnonkwao P, Jantama SS, Kanchanatawee S, Jantama K. 2018. Re-engineering Escherichia coli KJ122 to enhance the utilization of xylose and xylose/glucose mixture for efficient succinate production in mineral salt medium. Appl Microbiol Biotechnol 102:127-141.

18. Parker C, Barnell WO, Snoep JL, Ingram LO, Conway T. 1995. Characterization of the Zymomonas mobilis glucose facilitator gene product (glf) in recombinant Escherichia coli: examination of transport mechanism, kinetics and the role of glucokinase in glucose transport. Mol Microbiol 15:795-802.

19. Chen LQ, Cheung LS, Feng L, Tanner W, Frommer WB. 2015. Transport of sugars. Annu Rev Biochem 84:865-94.

20. Pao SS, Paulsen IT, Saier MH, Jr. 1998. Major facilitator superfamily. Microbiol Mol Biol Rev 62:1-34.

21. Young E, Poucher A, Comer A, Bailey A, Alper H. 2011. Functional survey for heterologous sugar transport proteins, using *Saccharomyces cerevisiae* as a host. Appl Environ Microbiol 77:3311-9.

22. Young EM, Tong A, Bui H, Spofford C, Alper HS. 2014. Rewiring yeast sugar transporter preference through modifying a conserved protein motif. Proc Natl Acad Sci U S A 111:131-6.

23. Young EM, Comer AD, Huang H, Alper HS. 2012. A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. Metab Eng 14:401-11.

24. Shin HY, Nijland JG, de Waal PP, de Jong RM, Klaassen P, Driessen AJ. 2015. An engineered cryptic Hxt11 sugar transporter facilitates glucose-xylose co-consumption in *Saccharomyces cerevisiae*. Biotechnol Biofuels 8:176.

25. Snoep JL, Arfman N, Yomano LP, Fliege RK, Conway T, Ingram LO. 1994. Reconstruction of glucose uptake and phosphorylation in a glucose-negative mutant of *Escherichia coli* by using *Zymomonas mobilis* genes encoding the glucose facilitator protein and glucokinase. J Bacteriol 176:2133-5.

26. Dunn KL, Rao CV. 2015. High-throughput sequencing reveals adaptation-induced mutations in pentose-fermenting strains of *Zymomonas mobilis*. Biotechnol Bioeng 112:2228-40.

27. Chen T, Zhang J, Liang L, Yang R, Lin Z. 2009. An in vivo, label-free quick assay for xylose transport in *Escherichia coli*. Anal Biochem 390:63-7.

28. Ren C, Chen T, Zhang J, Liang L, Lin Z. 2009. An evolved xylose transporter from *Zymomonas mobilis* enhances sugar transport in *Escherichia coli.* Microb Cell Fact 8:66.

29. Sun L, Zeng X, Yan C, Sun X, Gong X, Rao Y, Yan N. 2012. Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. Nature 490:361-6.

30. Farwick A, Bruder S, Schadeweg V, Oreb M, Boles E. 2014. Engineering of yeast hexose transporters to transport D-xylose without inhibition by D-glucose. Proc Natl Acad Sci U S A 111:5159-64.

31. Reider Apel A, Ouellet M, Szmidt-Middleton H, Keasling JD, Mukhopadhyay A. 2016. Evolved hexose transporter enhances xylose uptake and glucose/xylose co-utilization in Saccharomyces cerevisiae. Sci Rep 6:19512.

32. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539.

33. Li H, Schmitz O, Alper HS. 2016. Enabling glucose/xylose co-transport in yeast through the directed evolution of a sugar transporter. Appl Microbiol Biotechnol 100:10215-10223.

34. Weber YG, Storch A, Wuttke TV, Brockmann K, Kempfle J, Maljevic S, Margari L, Kamm C, Schneider SA, Huber SM, Pekrun A, Roebling R, Seebohm G, Koka S, Lang C, Kraft E, Blazevic D, Salvo-Vargas A, Fauler M, Mottaghy FM, Munchau A, Edwards MJ, Presicci A, Margari F, Gasser T, Lang F, Bhatia KP, Lehmann-Horn F, Lerche H. 2008. GLUT1 mutations are a cause of paroxysmal exertion-induced dyskinesias and induce hemolytic anemia by a cation leak. J Clin Invest 118:2157-68.

35. Gardonyi M, Jeppsson M, Liden G, Gorwa-Grauslund MF, Hahn-Hagerdal B. 2003. Control of xylose consumption by xylose transport in recombinant Saccharomyces cerevisiae. Biotechnol Bioeng 82:818-24.

36. Runquist D, Hahn-Hagerdal B, Radstrom P. 2010. Comparison of heterologous xylose transporters in recombinant Saccharomyces cerevisiae. Biotechnol Biofuels 3:5.

37. Nijland JG, Vos E, Shin HY, de Waal PP, Klaassen P, Driessen AJ. 2016. Improving pentose fermentation by preventing ubiquitination of hexose transporters in *Saccharomyces cerevisiae*. Biotechnol Biofuels 9:158.

38. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-5.

39. Jantama K, Zhang X, Moore JC, Shanmugam KT, Svoronos SA, Ingram LO. 2008. Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. Biotechnol Bioeng 101:881-93.

40. Shevchuk NA, Bryksin AV, Nusinovich YA, Cabello FC, Sutherland M, Ladisch S. 2004. Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. Nucleic Acids Res 32:e19.

41. Quan J, Tian J. 2011. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242-51.

42. Martinez A, Grabar TB, Shanmugam KT, Yomano LP, York SW, Ingram LO. 2007. Low salt medium for lactate and ethanol production by recombinant *Escherichia coli* B. Biotechnol Lett 29:397-404.

43. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Gallo Cassarino T, Bertoni M, Bordoli L, Schwede T. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 42:W252-8.

44. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-12.

45. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK. 2011. The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. BMC Genomics 12:9.

46. Cherepanov PP, Wackernagel W. 1995. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158:9-14.

47. Amann E, Ochs B, Abel KJ. 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301-15.

CHAPTER 4

BIOPROSPECTING OF NATIVE EFFLUX PUMPS TO ENHANCE FURFURAL TOLERANCE IN ETHANOLOGENIC ESCHERICHIA COLI

Abstract

Efficient microbial conversion of lignocellulose into valuable products is often hindered by the presence of furfural, a dehydration product of xylose in hemicellulose sugar syrups derived from woody biomass. For a cost-effective lignocellulose microbial conversion, robust biocatalysts are needed that can tolerate toxic inhibitors while maintaining optimal metabolic activities. A comprehensive plasmid-based library encoding native multidrug resistance (MDR) efflux pumps, porins and select exporters from *Escherichia coli* was screened for furfural tolerance in an ethanologenic *E. coli*. Small multidrug resistance (SMR) pumps, such as SugE and MdtJI, as well as a lactate/glycolate:H^+ symporter LldP, conferred furfural tolerance in liquid culture tests. Expression of theses SMR pumps potentially increased furfural efflux and cellular viability upon furfural assault, suggesting novel activities for SMR pumps as furfural efflux proteins. Furthermore, induced expression of *mdtJI* in the presence of furfural and 5-hydroxymethylfurfural enhanced ethanol fermentative production of LY180, further demonstrating the applications of these pumps. This work described an effective approach to identify useful efflux systems with desired activities for non-native toxic chemicals and provides a platform to further enhance furfural efflux by protein engineering and mutagenesis.

Importance

Lignocellulosic biomass, especially agricultural residues, represents an important potential feedstock for microbial production of renewable fuels and chemicals. During the deconstruction of hemicellulose by thermochemical processes, side products such as furan aldehydes are generated which inhibit cell growth and production, limiting cost-effective lignocellulose conversion. Here, we developed a new approach to increase cellular resistance by identifying MDR pumps with putative efflux activities for furan aldehydes. The developed plasmid library and screening methods may facilitate new discoveries of MDR pumps for diverse toxic chemicals important for microbial conversion.

Introduction

Lignocellulose represents one of the most abundant carbon resources on the planet, with an estimated net photosynthetic productivity of 155 billion tons per year in the biosphere (1). Since 50-70% of lignocellulose dry weight is composed of fermentable sugars, the abundance and low cost of lignocellulose has made it an attractive renewable feedstock for bioconversion (2-4), although multiple bioprocessing barriers currently impede its cost effective conversion. Most of the utilizable monosaccharides of lignocellulose are locked up in recalcitrant structures such a cellulose and hemicellulose which need to be denatured and depolymerized prior to fermentation. Chemical pretreatment of lignocellulose to release fermentable sugars is an essential process of typical lignocellulosic biorefineries (5), but cost-effective pretreatment strategies (e.g. dilute acid hydrolysis) are known to release multiple inhibitors into the resulting broth (3, 6). Inhibitors released during this process include furan-aldehydes (e.g. furfural, 5-hydroxymethylfurfural), small organic acids, and soluble aromatics which have a broad range of toxicity for multiple biocatalysts (7-9). To reduce additional costs to the pretreatment unit process, it is desirable to engineer robust biocatalysts that can tolerate relevant concentrations of these inhibitors. Of the inhibitors produced during this process, furfural (a dehydration product of pentoses) is particularly important due to its abundance in hemicellulose hydrolysates, strong cytotoxicity, and the unique property to potentiate the toxicity of other side-products (10-12).

Although the exact mechanism of furfural toxicity remains elusive, it is thought to be multifaceted (9, 10). Some evidenced toxicity mechanisms include co-factor starvation during furfural reduction to the less toxic furfuryl alcohol (13-15), membrane damage (16) and DNA damage due to oxidative stress (17-19). Thus, previous work on engineering microbial tolerance to furan aldehydes has focused on the use of alternative oxidoreductases (20-22), increasing the cellular pool of NADPH (14, 23), altering membrane properties (24), and enhancing DNA repair systems (25, 26) as the main mechanisms among other general strategies for increasing stress tolerance (9, 27). However, many of these beneficial traits do not have a simple cumulative synergy (9, 28), thus making further combinatorial optimization of furfural resistance traits a challenging task.

Microbial efflux is an important mechanism utilized by many bacteria to acquire antimicrobial resistance (29, 30), heavy metal tolerance (31, 32), and recently enhanced tolerance for renewable chemical production (33, 34). Engineering microbial efflux systems represents a novel route to reduce cytotoxicity of hydrophobic molecules like furfural (Log Kow= 0.41) and may have a positive epistatic interaction with other tolerance traits due to a distinct working mechanism. However, no exporters or efflux pumps have been described with activity for furfural. Multidrug resistance (MDR) efflux pumps are a group of active transporters that that include proteins from the major facilitator superfamily (MFS), small multidrug resistance (SMR), resistance-nodulation-cell division (RND), ATP-binding cassette (ABC), and multidrug and toxin extrusion (MATE) families (35, 36). Each of these families are known for their ability to expel a broad range of lipophilic compounds such as antimicrobial agents and solvents (32, 36, 37).

Despite the great potential of MDR pumps in industrial applications, only some members from RND and ABC families have been thoroughly investigated as tools for improving tolerance to renewable fuels and/or chemicals in heterologous systems (33, 34, 38). In this work, we comprehensively screened a library containing all known native MDR efflux pumps, porins, and other selected transporters in *E. coli* for their ability to enhance furfural tolerance. Using a two-step screening method we discovered two SMR transporters (*sugE* and *mdtJI*) and one lactate/glycolate permease (*lldP*) which conferred tolerance to furfural. Increased expression of these SMR genes can enhance furfural tolerance and thus ethanol fermentative production in an ethanolgenic *E. coli*. Furthermore, furfural tolerance conferred by the SMR pump is evidenced to be due to its efflux activity.

Results

Screening and identification of native export systems in *E. coli* to enhance furfural tolerance

A small plasmid library (n = 103) encoding all reported MDR pumps, selected exporters of interest (defined or putative exporters for various substrates based on EcoCyc database (39)), and porins from *E. coli* was constructed based on the ASKA plasmid collection which includes presumably all single ORFs from *E. coli* K-12 (40) (Figure 4.1A). For multi-component transporters, individual genes were added on the existing ASKA plasmid in the same manner using circular

polymerase extension cloning (CPEC) method (41). Plasmids were individually transformed into ethanologenic *E. coli* LY180 and a two-step screening process was employed to identify candidates that confer furfural tolerance during fermentative growth (Figure 4.1B).

As the first step, a plate-based assay was used to thoroughly evaluate all individual members from the constructed plasmid library for their abilities to confer furfural tolerance by comparing visible differences in colony size upon exposure to 0.9 gL^{-1} furfural. Cells with higher furfural tolerance will grow faster and thus yield a larger colony size (Figure 4.1C). LY180 with empty vector or a plasmid encoding *pntAB* (14) was used as a negative and a positive control, respectively (Figure 4.1C). Two concentrations of inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG) were used to test the effect of these candidate genes under different expressional levels since there is a known trade-off between the beneficial effect of membrane proteins and cytotoxicity associated with their overexpression (42). Similar to the positive control, the colony size of strains with constructs encoding two SMR pumps (*sugE* and *mdtJI),* two glycolate/lactate permeases (*glcA* and *lldP*) and two RND pumps (*acrA* and *acrEF-tolC*) were larger than that of the strain with the empty vector pCA24N in the presence of 10 μM IPTG (Figure 4.1A). Enhanced growth was not observed when the IPTG concentration was increased to 100 μM for strains carrying *acrA, acrEF-tolC,* or *glcA* constructs.

As the second step of the screening process, LY180 transformed with the empty vector pCA24N or the potential positive constructs was grown in liquid cultures using AM1 mineral salt medium supplemented with 5% xylose (w/v) and IPTG at 10 or 100 μM. Since it is known that gelling agents for solid media can contain variable elemental contaminants that can potentially effect transcriptional activity (43), this second assay is more representative of relevant fermentation growth conditions. Induced expression of *sugE, mdtJI,* and *lldP* by 10 μM IPTG increased cell growth after 48 hours in the presence of 1.0 g L^{-1} furfural by ~4-5 fold compared to the empty vector control (Figure 4.2B). At 100 μM IPTG induction, only overexpression of *sugE* conferred a minor increase in cell growth (Figure 4.2C). The increased IPTG concentration possibly resulted in cytotoxicity associated with membrane protein overexpression and thus offset the beneficial effect. In addition, increased IPTG concentrations were observed to influence cellular furfural sensitivity

even for the cells with empty vector (Figure 4.2C), which was consistent with previous results (44). Constructs from the ASKA collection have short stretches of artificial sequences at both the N- and C-terminus (Figure 4.2A). Although some scattered evidence suggests that these artificial sequences may not disrupt the localization and function of membrane proteins (40), we cloned the wild-type ORFs of positive candidate genes *sugE*, *mdtJI* and *lldP* omitting the artificial sequences into a pTrc99A (different replication origin and antibiotic resistance from pCA24N) to test the potential influence of ASKA artificial sequences and plasmid backbone. Induced expression of native *sugE, mdtJI,* and *lldP* using 10 μM IPTG also enhanced biomass production in the presence of furfural after 48h by approximately 3.4, 5.4, and 5.3-fold, respectively (Figure 4.2D). Similar to the ASKA-based constructs, overexpression of these genes induced by 100 μM IPTG showed much smaller beneficial effects for SMR pumps or even a negative effect for LldP compared to results using 10 μM IPTG (Figure 4.2D and 4.2E), suggesting that optimal beneficial effects require suitable expression levels for these transporter genes due to known trade-off (34). Cells with plasmids pTrc99A and pCA24N showed different baseline furfural sensitivity likely due to different copy numbers and antibiotics used in the medium (Figure 4.2).

Since overexpression of *lldP* was more toxic than *sugE* and *lldP* under different culture conditions (Figure 4.2E and 4.3) and wild-type *sugE* had a consistently smaller beneficial effect than *mdtJI* (Figure 4.2D and 4.2E), *mdtJI* was selected for further investigation.

Induced expression of *mdtJI* shortens fermentation lag phase caused by furan aldehydes

To investigate if increased expression of *mdtJI* could enhance ethanolgenic fermentation under furfural stress, we fermented LY180 containing empty vector or the plasmid encoding wild-type *mdtJI* with IPTG induction at 10 μM in the presence or absence of 1.25 g L^{-1} furfural. Without furfural, there was essentially no difference between cells with and without *mdtJI* in terms of cell growth and ethanol production metrics upon 10 μM IPTG induction (Figure 4.4). With furfural, the increased expression of *mdtJI* led to significant biomass production after 72 hours, whereas cells with the empty vector took 96 hours to accumulate significant biomass (Figure 4.4A). Thus, overexpression of *mdtJI* shortened the furfural-induced lag phase by approximately 24 hours. Due to the decreased lag time, productivity was higher over 96 h in *mdtJI* overexpressing cells at 0.42 g L^{-1} h^{-1}, in

comparison to 0.19 g L^{-1} h^{-1} in the empty vector control (Figure 4.4B). Since many genetic traits that enhance furfural tolerance also enhance tolerance to 5-hydroxymethylfurfural (HMF) (15, 45), another important inhibitor derived from lignocellulose, we tested if induced expression of *mdtJI* increased tolerance to this chemical as well. Similarly, overexpression of *mdtJI* enhanced ethanol productivity and decreased fermentation lag time when exposed to 1.75 g L^{-1} HMF (Figure 4.5).

Functional overexpression of *mdtJI* enhances survival under furfural stress and furfural extracellular accumulation

To test if the tolerance phenotype was actually due to efflux activity of MdtJI, we disrupted its transport function by introducing a previously described point mutation (E19Q) into *mdtI*, This disrupts the conserved glutamate residue at position 19 of MdtI that is essential for proton-coupled transport activity of the heterodimer MdtJI transporter (46). It was reported that disruption of this conserved glutamate residue in *E. coli* native SMR pumps yielded stable mutants with severely hindered transporter activity (46-49). Introducing this mutation into *mdtJI* eliminated the furfural tolerance phenotype, suggesting the transporter activity is involved in enhanced furfural tolerance (Figure 4.6A). We then hypothesized that the efflux activities of MdtJI may increase cell viability upon furfural exposure. Cells were incubated at toxic concentrations of furfural and their viabilities were measured over time. Induced expression of wild-type *mdtJI* increased the cell viability during 48 and 72 hours of incubation, with a 4.5-fold and 2.6-fold increase in colony viable counts, respectively (Figure 4.6B). Induced *mdtJI* mutant (E19Q in *mdtI*) did no increase cell viability upon furfural exposure (Figure 4.6B). This increase in viability caused by MdtJI is likely due to a decrease in intracellular furfural if *mdtJI* functions as a furfural efflux pump. To further test this hypothesis, cells with empty vector or plasmids encoding wild-type MdtJI and its E19Q mutant were incubated with furfural and the extracellular furfural concentrations were measure at different time intervals. The main metabolic route for furfural in *E. coli* is to be reduced to furfuryl alcohol by native furfural reductases (45). To prevent any furfural-induced changes in proteome that may interfere with our assay, chloramphenicol was added to stop protein synthesis before furfural exposure. If furfural enters at the same rates for different cells, the same reduction rates will be observed because the native furfural reductases such as YqhD (50) and FucO (20) should be at equivalent levels due to

the addition of chloramphenicol. If furfural efflux pumps are active, slower furfural reduction rates will be observed as more furfural will be expelled and accumulated in the medium. Over 60-min incubation time, furfural reduction is slower for cells with induced expression of wild-type *mdtJI*. After sixty minutes incubation, the reduced amount of furfural in the broth was 225% greater in cells with an empty vector compared to those overexpressing wild-type *mdtJI* (Figure 4.6C). Consistent with all previous results, no significant difference was found in extracellular furfural reduction experiments between the empty vector and plasmid containing *mdtJI* E19Q (Figure 4.6).

Chromosomal integration of *mdtJI* enhances furfural tolerance and potential synergy of *mdtJI* and other beneficial genetic traits

Since plasmid-based expression is not economically desirable at the industrial scale due to genetic instability, metabolic burden, and the costly requirement of antibiotics and inducers (51-53), we integrated *mdtJI* into the *frdBC* locus in LY180 to develop a plasmid-free system. The *frdBC* genes are already deleted in LY180 and *mdtJI* expression will theoretically be active under the promoter of the *frdABCD* operon, which is known to be upregulated under fermentation conditions (54, 55). This also sought to remove any influence that the plasmid, antibiotic, or inducer may have on furfural tolerance. The integrated strain, YR105, was then assessed for furfural tolerance using the same liquid assay for plasmid-based tests. After 48 h, YR105 demonstrated a 5.7-fold increase in biomass compared to LY180 upon furfural stress (Figure 4.7A). We then tested to see if *mdtJI* could be combined with other known furfural tolerance genes, using a previously developed furfural-tolerant ethanologenic *E. coli* XW129 containing three beneficial genetic traits (deletion of *yqhD*, overexpression of *fucO* and *ucpA*) (28). Induced expression of *mdtJI* increased biomass after 48h in the presence of 1.75 g L^{-1} furfural by 50% compared to an empty vector control (Figure 4.7B). A similar improvement of cell growth was observed during fermentation, especially at 48 h (Figure 4.8). However, the improvement of cell growth at small degree did not significantly improve ethanol production (Figure 4.8), suggesting that further improvement of efflux activity may be needed to achieve a significant synergy for furfural tolerance.

Discussion

In this work, we comprehensively interrogated a plasmid library of all known *E. coli* MDR efflux pumps as well as other selected exporters and porins for their abilities to enhance tolerance to furfural using a two-step screening process. Two SMR pumps (*mdtJI* and *sugE*) and one glycolate/lactate permease (*lldP*) were identified to enhance furfural tolerance in an ethanologenic *E. coli* strain. We further demonstrate that the transport function of SMR pump MdtJI is required for enhanced furfural tolerance and its increased expression shortens the lag time for ethanol fermentative production during batch fermentations.

Researchers have performed a variety of genome-scale screening methods to identify furfural-resistant genetic traits. For example, genomic libraries from different bacteria and metagenomic libraries have been constructed and screened that led to the discovery genes such as *thyA* that confer furfural tolerance to *E. coli* (25, 26). MDR pumps and other membrane proteins were rarely identified as positive traits at least partially because their beneficial effect is sensitive to the expression level, and often eliminated when the expression level is high as observed for this work (Figure 4.2). However, most reported work employed the screening methods with high copy plasmids and high induction conditions (25, 26). Careful evaluation of transporter genes with different expression levels may yield new candidates for microbial tolerance engineering.

The SMR family is a group of small (~100 amino acids) inner membrane transporters that in *E. coli* has four members, SugE, EmrE, MdtJ, and MdtI (56). Of these, EmrE and SugE are thought to form functional homodimers and MdtJI a heterodimer to enhance tolerance to a number of lipophilic, quaternary ammonium compounds (QACs), and cationic drugs (57, 58). The SMR pumps are proposed to be H^+/substrate alternating access antiporters (59) and their efflux activity is dependent on a conserved glutamate residue (E14 of EmrE, E19 in MdtI*)* (47, 48). The introduction of a well-studied point mutation (E19Q) (46, 49) in *mdtI* disrupted transport mechanism and thus the furfural tolerance phenotype, indicating that transport is necessary for the furfural tolerance phenotype. Besides sharing several of the substrates with other SMR pumps, MdtJI is also an evidenced H^+/spermidine antiporter in *E. coli* (46). This is important in the context of furfural resistance, as it has been discovered previously that import and supplementation of polyamines such as putrescene enhances furfural tolerance (44). Although it is possible that the enhanced

furfural tolerance phenotype is due to polyamine transporter activity of MdtJI, it is unlikely to be the main mechanism. This is evidenced by the observation that *sugE* overexpression enhanced furfural tolerance as well (Figure 4.2), and was not found to have polyamine transport activities (46). In addition, the extracellular concentration of furfural was increased in cells overexpressing *mdtJI*, suggesting that MdtJI is actively transporting furfural out of the cytoplasm to confer the tolerance (Figure 4.6). In addition, recent work has also found that upregulation of *mdtJI* may confer tolerance to ionic liquids, suggesting that SMR pumps may be useful for tolerance to different toxic chemical associated with lignocellulose conversion (60). It is plausible that furfural (uncharged, hydrophobic) is not the preferred substrate for SMR proteins (QACs and cationic chemicals as defined substrates). Improvement of the activities of SMR pumps towards furfural by mutagenesis will further enhance the furfural tolerance of biocatalysts.

The lactate permease family is a group of inner membrane transporters (~550aa) that is known to consist of LldP and GlcA in *E. coli* (61). Induced *lldP* overexpression enhanced furfural tolerance to a similar extent as *sugE* or *mdtJI* (Figure 4.2). However, *lldP* overexpression was more toxic to the cells than *sugE* or *mdtJI* (Figure 4.3). Overexpression of membrane proteins can overwhelm membrane translocation machinery and quickly dilute essential membrane proteins being integrated into the membrane (62). The increased toxicity of LldP is most likely due to its 4-fold larger size than SMR transporters, thus interfering with membrane translocation machinery at a higher degree.

We demonstrate here that the increased expression of native SMR enhanced tolerance to furfural. Interestingly, the microarray data of the same ethanologenic strain upon furfural exposure in the previous reports (14, 50) shows no significant change in expression of *sugE, mdtJI,* and *lldP* upon exposure to furfural (less than 2-fold changes). This suggests that the MDR pumps or other exporters are able to transport non-native toxic chemicals due to the promiscuous substrate specificity although the transcriptional regulation has not been correctly "wired" yet by evolution. Indeed, the bioprospecting strategy presented here may be broadly applicable to a wide range of non-native inhibitors, and this plasmid library can be expanded to other microbial exporters and

even metagenomic libraries. Other biocatalysts may benefit from similar investigations for other industrially relevant chemicals.

Materials and Methods

Strains and Cultivation

All strains used in this work listed in Table 4.1. During strain construction, transformation and genetic manipulations were performed in Luria Broth (10 g L^{-1} Difco tryptone, 5 g L^{-1} Difco tryptone, and 5 g L^{-1} NaCl) at 30°C, 37°C, or 39°C as needed and with rotation at 180 rpm when in liquid culture. During chromosomal integration experiments, 5% (w/v) arabinose and 10% (w/v) sucrose were used for λ-red recombinase induction from pKD46 and counterselection of *sacB* negative strains, respectively. Ampicillin (100 mg L^{-1}) and/or chloramphenicol (50 mg L^{-1}) were supplemented if selection was needed.

Strain, Plasmid and Library Construction

A λ-red recombinase-mediated two-step method was used as previously described (63, 64) to integrate *mdtJI* into *frdBC* locus in LY180. Briefly, a *cat-sacB* counter-selection cassette from pXW001 (64) and *mdtJI* from *E. coli* LY180 flanked with 50 bp of homology to the *frdBC* gene locus were amplified by PCR using the primers in Table 4.2. This *cat-sacB* fragment was integrated into the *frdBC* locus and validated by colony PCR before being subjected to a second round of integration. After integration of the *mdtJI* cassette, the cells were incubated in media supplemented with 10% (w/v) sucrose to enrich for the correct clones (*sacB* negative). The final positive clones were verified by colony PCR and Sanger sequencing.

To construct a plasmid library containing target genes, single ORF plasmids were isolated from the previously constructed ASKA collection (40). For multi-component transporters, genes were assembled into the ASKA vector backbone pCA24N in approximately the same fashion as all single ORF plasmids using the circular polymerase extension cloning (CPEC) method (65), with assembly specifications included in the primer description (Table 4.2). Positive hits identified from the screening process were re-cloned into EcoRI/XbaI sites of pTrc99A using conventional cloning method. LY180 genomic DNAs were used as PCR template to get native ribosomal binding sites, coding sequences and terminator regions without ASKA artificial sequences. The plasmid encoding

mdtJI (E19Q) was made by site-specific mutagenesis using overlapping PCR extension. All constructed plasmids were verified using Sanger sequencing.

Two-step Screening Process

All library constructs including empty vector control were transformed individually into ethanolgenic *E. coli* LY180 (50). These transformants were streaked to form individual colonies on AM1 2% (w/v) xylose agar plates supplemented with 0.9 g L^{-1} furfural, and IPTG at 10 μM or 100 μM. Plates were incubated at 37 °C for 48 to 72 h in a container filled with argon gas. Each transformant was compared to empty vector control independently at least twice. Only if the colony size of transformants was consistently larger than that of empty control, they were scored as positive candidates for the plate-based assay and further tested using liquid a culture test.

Testing furfural toxicity in liquid cultures was the second-step screening process and was performed as previously described (20). Tube cultures (13 mm by 100 mm) contain 4.0 mL of AM1 medium with 50 g L^{-1} xylose, 50 mg L^{-1} chloramphenicol for pCA24N-based constructs or 12.5 mg L^{-1} ampicillin for pTrc99A-based constructs, various concentrations of furfural, and 10 or 100 μM IPTG. An initial inoculum of 44 mg dry cell weight (dcw) L^{-1} (0.1 as OD_{550}) was used for all liquid assays and optical density of cultures was measured at 550 nm after incubation for 48 h with 120 rpm shaking speed at 37°C.

Fermentation

Fresh colonies were grown microaerobically on AM1 agar plates (66) supplemented with 2% xylose (w/v) and antibiotics in a container filled with argon gas, and then were transferred into 100 mL AM1 2% xylose (w/v) medium with antibiotics in a 250mL flask. After growth at 37°C 120 rpm for approximately 16 h, this culture was used as the seed inoculum for fermentation. Fermentation was performed in 500 mL fermentation vessels containing 300 mL AM1 medium supplemented with 10% xylose (w/v) (37°C, 120 rpm) as previously described (20, 45, 50). Fermentations were maintained at pH 7.0 by the automatic addition of 2.0 M potassium hydroxide and an initial inoculum of 44 mg dcw L^{-1} (0.1 as OD_{550}) was used for all fermentations. Furfural or HMF were added at the beginning of the fermentation in addition to 10 μM IPTG. 12.5 mg L^{-1} ampicillin was included in both seed cultures and batch fermentations.

Analysis

All quantification of xylose and ethanol was performed using high performance liquid chromatography (HPLC) as previously described (64, 67). More specifically, compounds were separated and quantified with an UltiMate3000 (Thermo Fisher) using an Aminex HPX-87H column (Bio-Rad) and 4 mM sulfuric acid as the mobile phase. Furfural was quantified using a described UV absorbance method on a Beckman Coulter DU 730 spectrophotometer (68).

Cell Viability Assay

Cells were grown on AM1 2% xylose ampicillin agar plates before being transferred into 100 mL AM1 2% xylose medium supplemented with 10 µM IPTG and 12.5 mg L^{-1} ampicillin and incubated for 16 hours. 6.6 mg cells were harvested from this liquid culture and was then resuspended in 30 mL fresh AM1 2% xylose supplemented with 3.5 g L^{-1} furfural, 10 µM IPTG and 12.5 mg L^{-1} ampicillin in a sealed glass tube (16 x 150mm). This culture was then incubated at 37°C 120 rpm for 72 h with viable cell counts being performed at selected intervals using AM1 2% xylose plates.

Furfural accumulation

Cells were grown microaerobically on AM1 2% xylose ampicillin agar plates before being transferred into 4 mL AM1 2% xylose medium supplemented with 10 µM IPTG and 12.5 mg L^{-1} ampicillin. After 16 hour incubation at 37°C 120 rpm, 50 mg L^{-1} chloramphenicol was added and cells were incubated for an additional 1.5 hours to stop protein synthesis. 2.2 mg cells were harvested from this culture by centrifugation and resuspended in AM1 2% xylose medium supplemented with 12.5 mg L^{-1} ampicillin, 50 mg L^{-1} chloramphenicol, and 3.5 g L^{-1} furfural. At designated intervals, supernatants were collected and extracellular furfural concentrations were measured.

Figures

Table 4.1: Bacterial strains and plasmids used in this study

Strain/Plasmid	Relevant characteristics	Reference
Strains		
LY180	Δ*frdBC*::(*Zm frg celY*$_{Ec}$ FRT) Δ*ldhA*::(*Zm frg casAB*$_{Ko}$) *adhE*::(*Zm frg estZ*$_{Pp}$ FRT) Δ*ackA*::FRT *rrlE*::(*pdc adhA adhB* FRT) Δ*mgsA*::FRT	(50)
YR105	LY180 *frdBC*::*mdtJI*	This study
XW129	LY180 Δ*yqhD ackA*::P$_{yadC'}$*fucO-ucpA*	(28)
Top10F'	F´{*lacIq*, Tn10(TetR)} *mcrA* Δ(*mrr-hsdRMS-mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *ara*D139 Δ(*ara leu*) 7697 *galU galK rpsL* (StrR) *endA1 nupG*	Invitrogen™
Plasmids		
pKD46	*bla*, *γ β exo* (Red recombinase)	(69)
pTrc99A	P$_{trc}$, *bla*, *lacI*^{q}	(70)
pTrc99A-sugE	*sugE* cloned into pTrc99A	This study
pTrc99A-mdtJI	*mdtJI* cloned into pTrc99A	This study
pTrc99A-lldP	*lldP* cloned into pTrc99A	This study
pTrc99A-mdtJIE19Q	*mdtJI (mdtI*E19Q) cloned into pTrc99A	This study
pCA24N	Empty vector for ASKA collection; P$_{T5\text{-}lac}$, *cat*, *lacI*^{q}	(40)
pCA24N-emrKY	*emrKY* cloned into pCA24N	This study
pCA24N-emrAB	*emrAB* cloned into pCA24N	This study
pCA24N-macAB	*macAB* cloned into pCA24N	This study
pCA24N-mdtJI	*mdtJI* cloned into pCA24N	This study
pCA24N-aaeAB	*aaeAB* cloned into pCA24N	This study
pCA24N-cusBA	*cusBA* cloned into pCA24N	This study
pCA24N-mdtEF	*mdtEF* cloned into pCA24N	This study
pCA24N-acrAB	*acrAB* cloned into pCA24N	This study
pCA24N-acrAD	*acrAD* cloned into pCA24N	This study
pCA24N-acrEF	*acrEF* cloned into pCA24N	This study
pCA24N-mdtAB	*mdtAB* cloned into pCA24N	This study
pCA24N-mdtABC	*mdtABC* cloned into pCA24N	This study
pCA24N-acrAB-tolC	*acrAB-tolC* cloned into pCA24N	This study
pCA24N-acrAD-tolC	*acrAD-tolC* cloned into pCA24N	This study
pCA24N-acrEF-tolC	*acrEF-tolC* into pCA24N	This study
pCA24N-mdtH	*acrAB-tolC* into pCA24N	This study
pCA24N-mdtEF-tolC	*acrAD-tolC* into pCA24N	This study
pCA24N-slp	*slp* into pCA24N	This study
ASKA Plasmids	Indicated genes in Fig. 1A cloned into pCA24N	(40)

Table 4.2: Primers used in this study

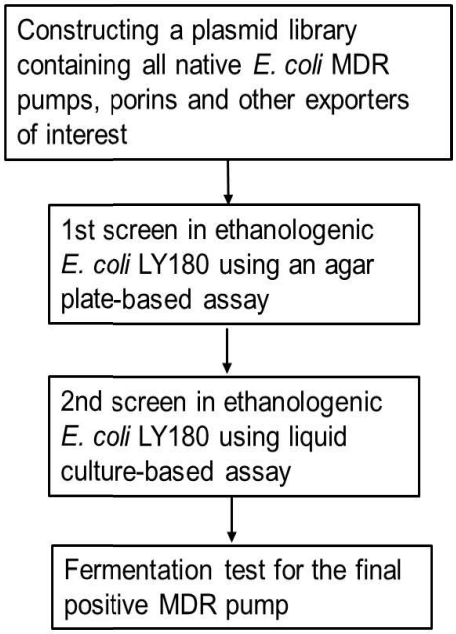
Primer	Sequence
pCA24n CPEC cloning	5' GGGTCGACCTGCAGCCAA
	5' TTAGCGGCCGCATAGGC
	5' TTGGCTGCAGGTCGACCC
	5' TCAGAACTCCATCTGGATTTGTTC
	5' GTCGACCCTTAGCGGCCGC
	5' GCTGCAGGTCGACCCTTAGCG
	5' CTGCAGCCAAGCTTAATTAGCTGAGC
emrKY ASKA cloning	5' GCCTATGCGGCCGCTAACAACATTATTTCGCATAATGGACAACTTTAATGAGAGGATCTCACCATCACCATC
emrAB ASKA cloning	5' GCCTATGCGGCCGCTAACTAACGCTGGCTAATCCAGAGGTGCGTGTGATGAGAGGATCTCACCATCACCATC
macAB ASKA cloning	5' GCCTATGCGGCCGCTAAGATTGGTGAGGCCAAACCAGGAGCTGCACAATGAGAGGATCTCACCATCACCATC
mdtJI ASKA cloning	5' GCCTATGCGGCCGCTAATTAAAATAATTCTCTTGCAGGAGAAGGACAATGAGAGGATCTCACCATCACCATC
aaeAB ASKA cloning	5' CATCGCCTGCGTGAGTTTGGTTAATCACGATGGGTATTTTCTCCATTGCTAACCAACATATTC
cusBA ASKA cloning	5' GAAAGTGCTACCCATGCGCATTGAGGGAATAACCAATGATTGAATGGATTATTCGTCGCTCGG
mdtEF ASKA cloning	5' GCCTATGCGGCCGCTAACAATAACGTTGCAGGCTTAAGGGGACTTTCATGAGAGGATCTCACCATCACCATC
acrAB ASKA cloning	5' GCCTATGCGGCCGCTAACTTAAACAGGAGCCGTTAAGACATGAGAGGATCTCACCATCACCATC
acrAD ASKA cloning	5' GCCTATGCGGCCGCTAACGATACGCAGAAACACGAGGTCCTCTTTTAATGAGAGGATCTCACCATCACCATC
acrEF ASKA cloning	5' GCCTATGCGGCCGCTAAAGATACTGCATCGAAGTAAGGTAATCTGACATGAGAGGATCTCACCATCACCATC
mdtABC ASKA cloning	5' CGAATACGCGAAAAAAGGAGCACGCTCCTGATGAGAGGATCTCACCATCACCATC
	5' CGCTAAGGGTCGACCTGCAGCGCTTTGCCCGTCATGAAGAGGAGGCGTAAATGAGAG GATCTCACCATCACCATC
	5' GGAGTCCAAGCTCAGCTAATTAAGCGCCGCATAGGCTCGGTTACC
tripartite tolC ASKA cloning	5' CTAAATACTGCTTCACCACAAGGAATGCAAATGAGAGGATCTCACCATCACCATC
	5' GCATTCCTTGTGGTGAAGCAGTATTTAGGCTGCAGGTCGACCCTTAG
	5' CTGCAGCCAAGCTTAATTAGCTGAGCTTGGACTCCCTAAATACTGCTTCACCACAAGGAATGC AAATGAGAGGATCTCACCATCACCATC
	5' GGCGTGCTTGCTGATAAACTT (for sequencing)
sugE pTrc99a cloning	5' CGTACGAATTCTGCAAACGCCTCTTTTCACCG
	5' ACACTCTAGACCGCGACTACCTGTGGATACTC
lldP pTrc99a cloning	5' CGTACGAATTCGATGAGCAACAGACTCATTACACG
	5' GAGATCTAGAAGTTGCATCGCCAGTTGGC
mdtJI pTrc99a cloning	5' CGTACGAATTCACTTTGGTTTCGCTGAATTAAGC
	5' ACACTCTAGAAACTGTCGCAAAAATAATGTTCAAACG
for *cat-sacB* integration	5' AGCGGATGCAGCCGATAAGGCGGAAGCAGCCAATAAGAAGGAGAAGGCGATCGAGTGTGACGGAAGATCA
	5' ATACCGGTTCGTCAGAACGCTTTGGATTTGGATTGATCATCTCAGGCTCCTTAGCCATTTGCCTGCTTTT
for mdtJI integration	5' AAACGCGTTTACGGTGGCGAAGCGGATGCAGCCGATAAGGCGGAAGCAGCACTTTGGTTTCGCTGAATTA
	5' ATACCGGTTCGTCAGAACGCTTTGGATTTGGATTAATCATCTCAGGCTCCCAACTGTCGCAAAAATAATG
	5' TATCTACCGTACGCCGGAAC
	5' AGCAAATGTGGAGCAAGAGG

A

The pump, transporter and porin genes screened

Types	Gene names				
ABC	*macA*	*macB*	*macAB**		
	mdlA	*mdlB*	*yddA*	*yojI*	
MATE	*mdtK*				
SMR	*emrE*	***sugE***	*mdtI*	*mdtJ*	***mdtJI****
MF	*fsr*	*mdfA*	*mdtG*	*mdtH**	*ydhC*
	emrD	*mdtD*	*ydiM*	*mdtM*	*yajR*
	ydeA	*ydeE*	*yebQ*	*ynfM*	*hsrA*
	yqjA	*yghB*	*mdtL*	*bcr*	*yeaN*
	emrK	*emrY*	*emrKY**	*emrA*	*emrB*
	*emrAB**				
RND	***acrA***	*acrB*	*acrAB**	*acrAB-tolC**	
	acrD	*acrAD**	*acrAD-tolC**		
	acrE	*acrF*	*acrEF**	***acrEF-tolC****	
	cusA	*cusB*	*cusAB**	*mdtE*	
	mdtF	*mdtEF**	*mdtA*	*mdtB*	*mdtABC**
	*mdtEF-tolC**				
Other exporters	*setA*	*yahN*	*rhtA*	*cydC*	*cydD*
	ybjE	*tehA*	*tqsA*	*yddG*	*yeaS*
	setB	*atoE*	*ygaZ*	*yfiK*	*ygaH*
	aaeB	***lldP***	***glcA***	*hsrA*	*rhtB*
	aaeA	*alaE*	*dauA*	*dcuA*	*dcuB*
	dcuC	*dcuD*	*dctA*	*aaeAB**	
Porins	*pal*	*wza*	*ompX*	*ompC*	*slp**
	ompF	*ompA*	*bglH*	*ompW*	*tolC*
	ompN	*lamB*	*fadL*		

B



C

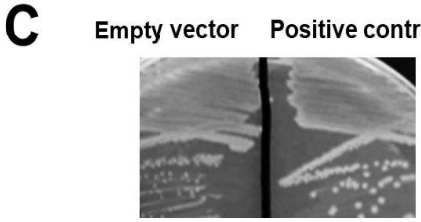


Figure 4.1. Screening procedure and identified positive candidate transporters. A) List of genes tested. Positive plasmids identified in the plate-based screen with 10 μM IPTG induction are indicated in bold. The genes are underscored if the plate-based results are still positive with 100 μM IPTG induction. Candidates are displayed with boxes if the following liquid culture screen results are positive. *Cassettes were assembled manually in the same manner as single ORF plasmids in ASKA collection. B) Ethanologenic *E.coli* LY180 was transformed with individual members of a plasmid library encoding MDR pumps, porins and exporters of interest, and then subjected to two screens under furfural stress conditions. C) Plate based screening was scored according to an empty vector control (negative control) and a known furfural resistance trait (overexpression of *pntAB*) (positive control). Strains displaying larger colonies in two independent tests were scored positive

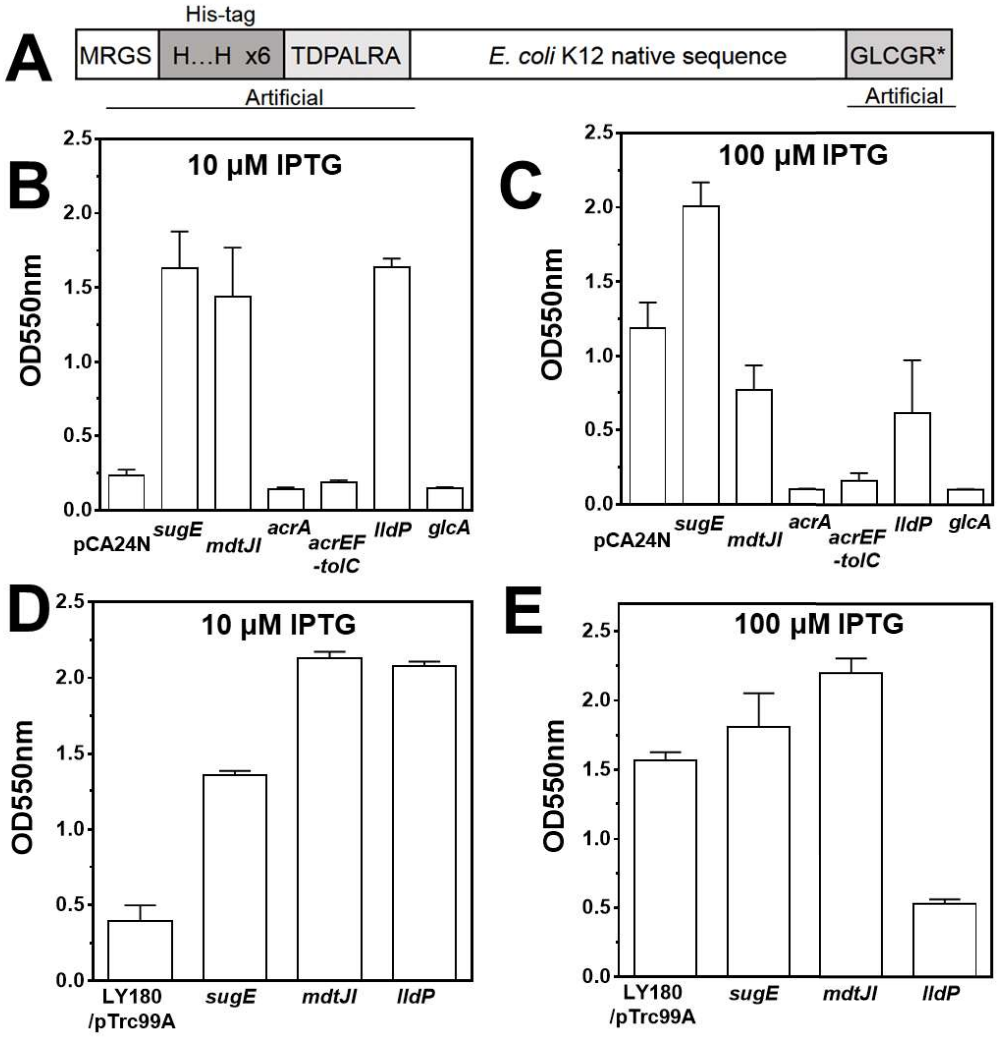


Figure 4.2. Identified genes confer furfural tolerance to LY180. A) Schematic drawing of a typical construct in ASKA collection shows artificial N and C-terminal sequences. The given size does not reflect real proportions. The optical density at 550 nm of LY180 with the positive ASKA plasmids after 48 hours growth in AM1 medium containing 5% xylose (w/v) and 1.0 g liter^{-1} furfural in the presence B) 10 µM or C) 100 µM IPTG. To test the effect of the wild-type candidate genes, the constructed pTrc99A-based plasmids were transformed into LY180 and the optical density at 550 nm of LY180 with native genes were measured after 48 hours growth in AM1 medium containing 5% xylose (w/v) and 1.25 g liter^{-1} furfural in the presence D) 10 µM or E) 100 µM IPTG. Note that different plasmids and antibiotics influenced furfural sensitivity of LY180.

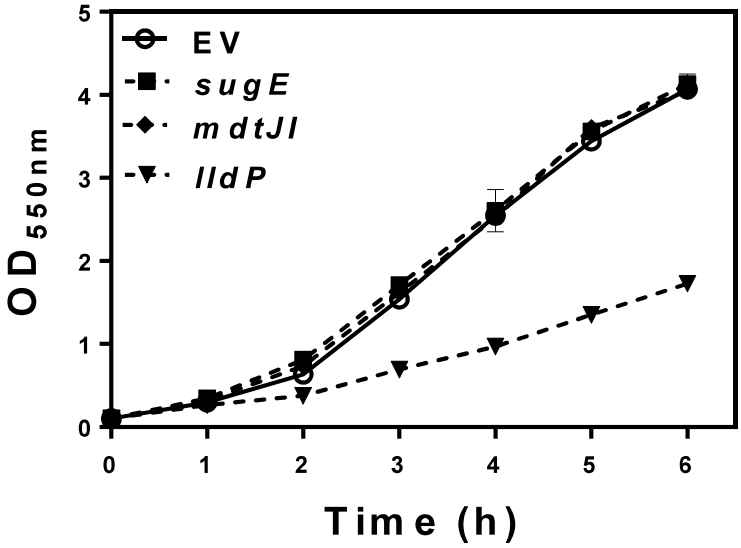


Figure 4.3. Effects of overexpression of *sugE*, *mdtJI* and *lldP* on cell growth using LB medium. The cultures were grown in 25 mL LB broth supplemented with 50 mg L^{-1} ampicillin with the starting OD is 0.1 in shake flasks at 37 C° and overexpression of tested genes were induced by 100 µM IPTG .

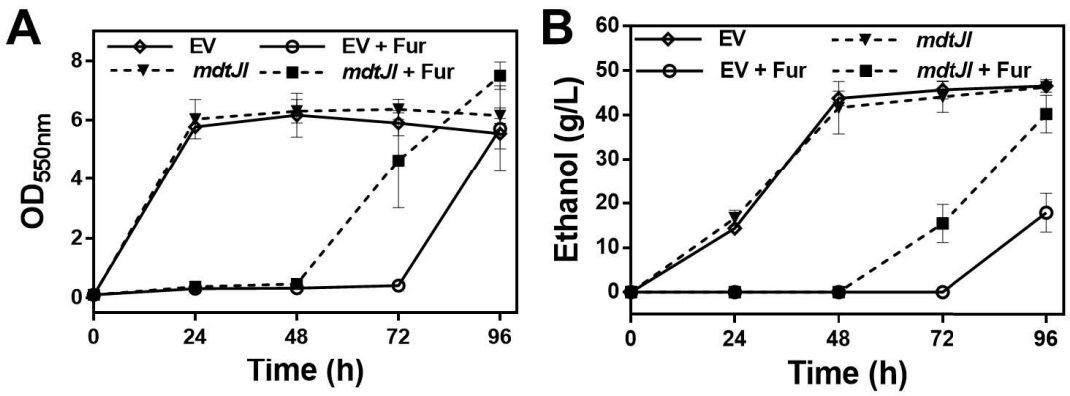


Figure 4.4. Induced expression of *mdtJI* in LY180 increases furfural resistance during ethanol fermentation. Batch fermentations were conducted in pH-controlled vessels in the absence and presence of 1.25 g liter^{-1} furfural. Expression of *mdtJI* from pTrc99A-*mdtJI* was induced by 10 µM IPTG (dotted lines) was compared to the empty vector (EV) controls (pTrc99A; solid lines) using LY180 as the host. A) Cell growth indicated by OD550nm. B) ethanol production was measured by HPLC and compared to controls without furfural.

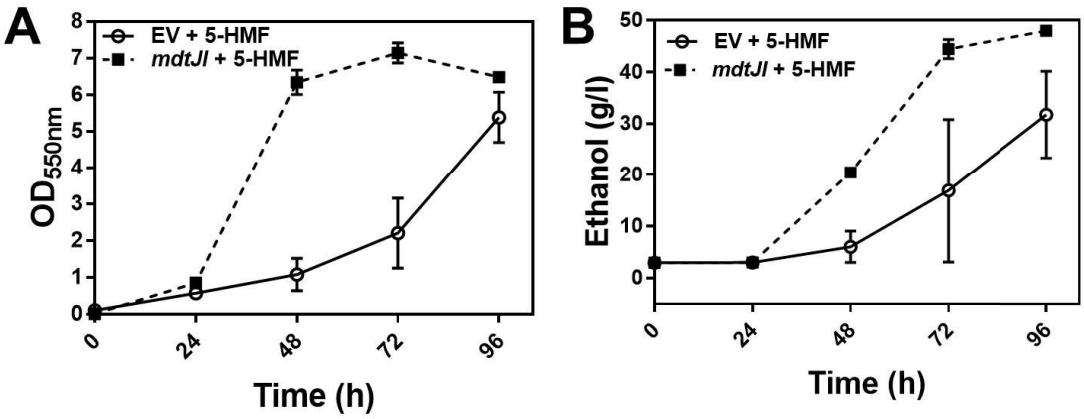


Figure 4.5. Induced expression of *mdtJI* in LY180 increases 5-hydroxylmethyl furfural (5-HMF) resistance during ethanol fermentation. Batch fermentations were conducted in pH-controlled vessels in the absence and presence of 1.75 g liter^{-1} 5-HMF. Expression of *mdtJI* from pTrc99A-*mdtJI* was induced by 10 µM IPTG (dotted lines) was compared to the empty vector (EV) controls (pTrc99A; solid lines) using LY180 as the host. A) Cell growth indicated by OD550nm. B) ethanol production was measured by HPLC.

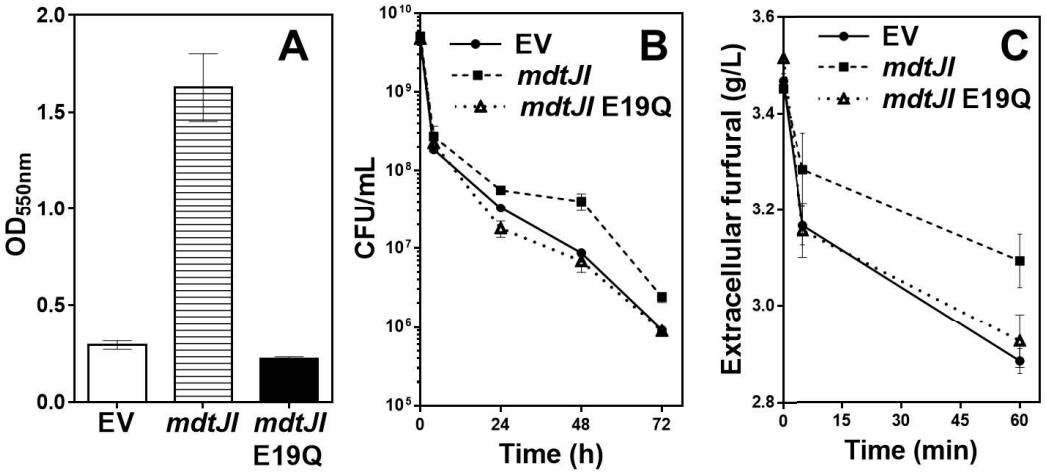


Figure 4.6. Comparison of wild-type MdtJI and its nonfunctional mutant (E19Q in *mdtI*) in terms of furfural resistance, survival ability under furfural stress and potential furfural export. A) Growth of LY180 transformed with indicated plasmids after 48 h in AM1 medium containing 5% xylose, 1.25 g L^{-1} furfural, and 10 μM IPTG. B) viabilities upon furfural exposure (3.5 g L^{-1}) were measured by colony forming unit, and C) extracellular furfural were evaluated upon exposure to furfural (3.5 g L^{-1}).

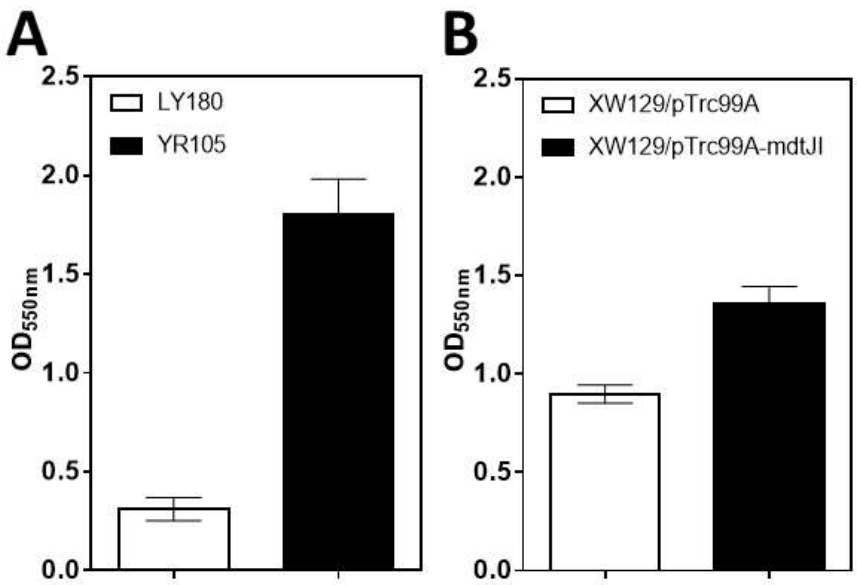


Figure 4.7. The chromosomal integration of *mdtJI* and potential synergy confer furfural resistance. A) Growth of LY180 and YR105 (LY180 *frdBC*::*mdtJI*) after 48 h in AM1 medium containing 5% xylose and 1.0 g L^{-1} furfural. B) Growth of XW129 transformed with either empty vector pTrc99A or plasmid encoding *mdtJI* after 48 h in AM1 medium containing 5% xylose, 1.75 g L^{-1} furfural and 10 µM IPTG..

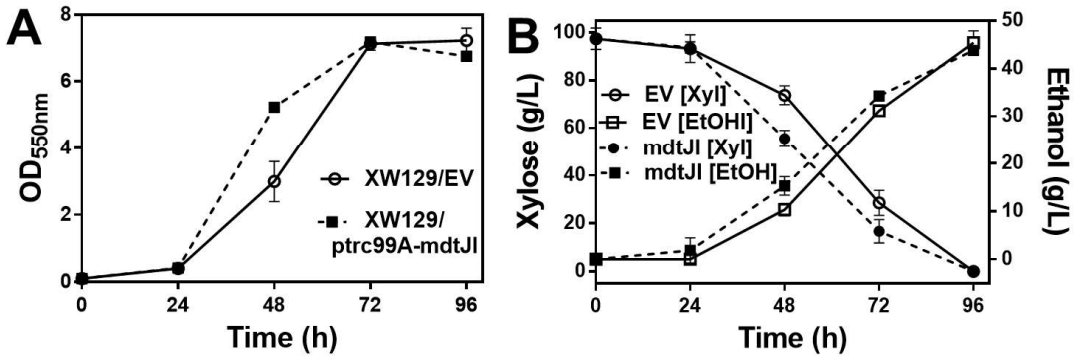


Figure 4.8. Induced expression of *mdtJI* in XW129 increases furfural resistance during ethanol fermentation. Batch fermentations were conducted in pH-controlled vessels in the absence and presence of 1.75 g liter^{-1} furfural. Expression of *mdtJI* from pTrc99A-*mdtJI* was induced by 10 μM IPTG (dotted lines) was compared to the empty vector (EV) controls (pTrc99A; solid lines) using XW129 as the host. A) Cell growth indicated by OD550nm. B) xylose and ethanol in fermentation broth were measured by HPLC.

References

1. Mishra P, Singh A. 1993. Microbial Pentose Utilization. Advances in Applied Microbiology, Vol 39 39:91-152.

2. Clark JH, Budarin V, Deswarte FEI, Hardy JJE, Kerton FM, Hunt AJ, Luque R, Macquarrie DJ, Milkowski K, Rodriguez A, Samuel O, Tavener SJ, White RJ, Wilson AJ. 2006. Green chemistry and the biorefinery: A partnership for a sustainable future. Green Chemistry 8:853-860.

3. Girio FM, Fonseca C, Carvalheiro F, Duarte LC, Marques S, Bogel-Lukasik R. 2010. Hemicelluloses for fuel ethanol: A review. Bioresour Technol 101:4775-800.

4. Saha BC. 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30:279-91.

5. Sousa LD, Chundawat SPS, Balan V, Dale BE. 2009. 'Cradle-to-grave' assessment of existing lignocellulose pretreatment technologies. Current Opinion in Biotechnology 20:339-347.

6. Hendriks ATWM, Zeeman G. 2009. Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresource Technology 100:10-18.

7. Lin FM, Qiao B, Yuan YJ. 2009. Comparative proteomic analysis of tolerance and adaptation of ethanologenic Saccharomyces cerevisiae to furfural, a lignocellulosic inhibitory compound. Appl Environ Microbiol 75:3765-76.

8. He MX, Wu B, Shui ZX, Hu QC, Wang WG, Tan FR, Tang XY, Zhu QL, Pan K, Li Q, Su XH. 2012. Transcriptome profiling of Zymomonas mobilis under furfural stress. Appl Microbiol Biotechnol 95:189-99.

9. Nieves LM, Panyon LA, Wang X. 2015. Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. Front Bioeng Biotechnol 3:17.

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

11. Geddes CC, Nieves lU, Ingram LO. 2011. Advances in ethanol production. Curr Opin Biotechnol 22:312-9.

12. Zaldivar J, Martinez A, Ingram LO. 1999. Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng 65:24-33.

13. Turner PC, Miller EN, Jarboe LR, Baggett CL, Shanmugam KT, Ingram LO. 2010. YqhC regulates transcription of the adjacent *Escherichia coli* genes *yqhD* and *dkgA* that are involved in furfural tolerance. J Ind Microbiol Biotechnol: doi:10.1007/s10295-010-0787-5.

14. Miller EN, Jarboe LR, Turner PC, Pharkya P, Yomano LP, York SW, Nunn D, Shanmugam KT, Ingram LO. 2009. Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. Appl Environ Microbiol 75:6132-41.

15. Miller EN, Turner PC, Jarboe LR, Ingram LO. 2010. Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing *Escherichia coli* LY180. Biotechnol Lett 32:661-7.

16. Allen SA, Clark W, McCaffery JM, Cai Z, Lanctot A, Slininger PJ, Liu ZL, Gorsich SW. 2010. Furfural induces reactive oxygen species accumulation and cellular damage in Saccharomyces cerevisiae. Biotechnol Biofuels 3:2.

17. Hadi SM, Shahabuddin, Rehman A. 1989. Specificity of the interaction of furfural with DNA. Mutat Res 225:101-6.

18. Khan QA, Shamsi FA, Hadi SM. 1995. Mutagenicity of furfural in plasmid DNA. Cancer Lett 89:95-9.

19. Shahabuddin, Rahman A, Hadi SM. 1991. Reaction of furfural and methylfurfural with DNA: use of single-strand-specific nucleases. Food Chem Toxicol 29:719-21.

20. Wang X, Miller EN, Yomano LP, Zhang X, Shanmugam KT, Ingram LO. 2011. Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in *Escherichia coli* strains engineered for the production of ethanol and lactate. Appl Environ Microbiol 77:5132-40.

21. Liu ZL, Slininger PJ, Dien BS, Berhow MA, Kurtzman CP, 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.

22. Liu ZL, Slininger PJ, Gorsich SW. 2005. Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains. Appl Biochem Biotechnol 121-124:451-60.

23. Gorsich SW, Dien BS, Nichols NN, Slininger PJ, Liu ZL, Skory CD. 2006. Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in Saccharomyces cerevisiae. Appl Microbiol Biotechnol 71:339-49.

24. Tan Z, Khakbaz P, Chen Y, Lombardo J, Yoon JM, Shanks JV, Klauda JB, Jarboe LR. 2017. Engineering *Escherichia coli* membrane phospholipid head distribution improves tolerance and production of biorenewables. Metab Eng 44:1-12.

25. Zheng H, Wang X, Yomano LP, Shanmugam KT, Ingram LO. 2012. Increase in furfural tolerance in ethanologenic *Escherichia coli* LY180 by plasmid-based expression of *thyA*. Appl Environ Microbiol 78:4346-52.

26. Forsberg KJ, Patel S, Witt E, Wang B, Ellison TD, Dantas G. 2016. Identification of Genes Conferring Tolerance to Lignocellulose-Derived Inhibitors by Functional Selections in Soil Metagenomes. Appl Environ Microbiol 82:528-37.

27. Flores A, Kurgan G, Wang X. 2016. Engineering Bacterial Sugar Catabolism and Tolerance Toward Lignocellulose Conversion. *In* Gosset G (ed), Engineering of Microorganisms for the Production of Chemicals and Biofuels from Renewable Resources doi:10.1007/978-3-319-51729-2. Springer Nature.

28. Wang X, Yomano LP, Lee JY, York SW, Zheng H, Mullinnix MT, Shanmugam KT, Ingram LO. 2013. Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 110:4021-6.

29. Pu YY, Zhao ZL, Li YX, Zou J, Ma Q, Zhao YN, Ke YH, Zhu Y, Chen HY, Baker MAB, Ge H, Sun YJ, Xie XS, Bai F. 2016. Enhanced Efflux Activity Facilitates Drug Tolerance in Dormant Bacterial Cells. Molecular Cell 62:284-294.

30. Harms A, Maisonneuve E, Gerdes K. 2016. Mechanisms of bacterial persistence during stress and antibiotic exposure. Science 354.

31. Outten FW, Huffman DL, Hale JA, O'Halloran TV. 2001. The independent cue and cus systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. Journal of Biological Chemistry 276:30670-30677.

32. Delmar JA, Su CC, Yu EW. 2014. Bacterial Multidrug Efflux Transporters. Annual Review of Biophysics, Vol 43 43:93-117.

33. Doshi R, Nguyen T, Chang G. 2013. Transporter-mediated biofuel secretion. Proceedings of the National Academy of Sciences of the United States of America 110:7642-7647.

34. Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, Keasling JD, Hadi MZ, Mukhopadhyay A. 2011. Engineering microbial biofuel tolerance and export using efflux pumps. Molecular Systems Biology 7.

35. Saier MH, Jr., Paulsen IT. 2001. Phylogeny of multidrug transporters. Semin Cell Dev Biol 12:205-13.

36. Nishino K, Nikaido E, Yamaguchi A. 2009. Regulation and physiological function of multidrug efflux pumps in *Escherichia coli* and *Salmonella*. Biochim Biophys Acta 1794:834-43.

37. Ramos JL, Duque E, Gallegos MT, Godoy P, Ramos-Gonzalez MI, Rojas A, Teran W, Segura A. 2002. Mechanisms of solvent tolerance in gram-negative bacteria. Annual Review of Microbiology 56:743-768.

38. Foo JL, Leong SS. 2013. Directed evolution of an *E. coli* inner membrane transporter for improved efflux of biofuel molecules. Biotechnol Biofuels 6:81.

39. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martinez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muniz-Rascado L, Ong Q, Paley S, Peralta-Gil M, Subhraveti P, Velazquez-Ramirez DA, Weaver D, Collado-Vides J, Paulsen I, Karp PD. 2017. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res 45:D543-D550.

40. Kitagawa M, Ara T, Arifuzzaman M, Ioka-Nakamichi T, Inamoto E, Toyonaga H, Mori H. 2005. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res 12:291-9.

41. Quan J, Tian J. 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4:e6441.

42. Turner WJ, Dunlop MJ. 2015. Trade-Offs in Improving Biofuel Tolerance Using Combinations of Efflux Pumps. ACS Synth Biol 4:1056-63.

43. Jain A, Poling MD, Smith AP, Nagarajan VK, Lahner B, Meagher RB, Raghothama KG. 2009. Variations in the composition of gelling agents affect morphophysiological and molecular responses to deficiencies of phosphate and other nutrients. Plant Physiol 150:1033-49.

44. Geddes RD, Wang X, Yomano LP, Miller EN, Zheng H, Shanmugam KT, Ingram LO. 2014. Polyamine Transporters and Polyamines Increase Furfural Tolerance During Xylose

Fermentation with Ethanologenic *Escherichia coli* strain LY180. Appl Environ Microbiol doi:10.1128/AEM.01913-14.

45. Wang X, Miller EN, Yomano LP, Shanmugam KT, Ingram LO. 2012. Increased furan tolerance in *Escherichia coli* due to a cryptic *ucpA* gene. Appl Environ Microbiol 78:2452-5.

46. Higashi K, Ishigure H, Demizu R, Uemura T, Nishino K, Yamaguchi A, Kashiwagi K, Igarashi K. 2008. Identification of a spermidine excretion protein complex (MdtJI) in *Escherichia coli*. J Bacteriol 190:872-8.

47. Muth TR, Schuldiner S. 2000. A membrane-embedded glutamate is required for ligand binding to the multidrug transporter EmrE. EMBO J 19:234-40.

48. Yerushalmi H, Schuldiner S. 2000. An essential glutamyl residue in EmrE, a multidrug antiporter from *Escherichia coli*. J Biol Chem 275:5264-9.

49. Dastvan R, Fischer AW, Mishra S, Meiler J, McHaourab HS. 2016. Protonation-dependent conformational dynamics of the multidrug transporter EmrE. Proc Natl Acad Sci U S A 113:1220-5.

50. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2009. Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic *Escherichia coli*. Appl Environ Microbiol 75:4315-23.

51. Ow DSW, Nissom PM, Philp R, Oh SKW, Yap MGS. 2006. Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5 alpha during batch fermentation. Enzyme and Microbial Technology 39:391-398.

52. Keasling JD. 2008. Synthetic biology for synthetic chemistry. Acs Chemical Biology 3:64-76.

53. Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, Ingram LO. 2010. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. J Biomed Biotechnol 2010:761042.

54. Ruch FE, Kuritzkes DR, Lin ECC. 1979. Use of 1ac Operon Fusions to Isolate Escherichia-Coli Mutants with Altered Expression of the Fumarate Reductase System in Response to Substrate and Respiratory Controls. Biochemical and Biophysical Research Communications 91:1365-1370.

55. Lohmeier E, Hagen DS, Dickie P, Weiner JH. 1981. Cloning and Expression of the Fumarate Reductase Gene of Escherichia-Coli. Canadian Journal of Biochemistry 59:158-164.

56. Bay DC, Rommens KL, Turner RJ. 2008. Small multidrug resistance proteins: a multidrug transporter family that continues to grow. Biochim Biophys Acta 1778:1814-38.

57. Nishino K, Yamaguchi A. 2001. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. J Bacteriol 183:5803-12.

58. Chung YJ, Saier MH, Jr. 2002. Overexpression of the *Escherichia coli sugE* gene confers resistance to a narrow range of quaternary ammonium compounds. J Bacteriol 184:2543-5.

59. Schuldiner S. 2014. Competition as a way of life for H(+)-coupled antiporters. J Mol Biol 426:2539-46.

60. Mohamed E, Wang S, Lennen R, Herrgard M, Simmons B, Singer S, Feist A. 2017. Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. Microb Cell Factories.

61. Prakash S, Cooper G, Singhi S, Saier MH, Jr. 2003. The ion transporter superfamily. Biochim Biophys Acta 1618:79-92.

62. Wagner S, Baars L, Ytterberg AJ, Klussmeier A, Wagner CS, Nord O, Nygren PA, van Wijk KJ, de Gier JW. 2007. Consequences of membrane protein overexpression in *Escherichia coli*. Mol Cell Proteomics 6:1527-50.

63. Jantama K, Zhang X, Moore JC, Shanmugam KT, Svoronos SA, Ingram LO. 2008. Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. Biotechnol Bioeng 101:881-93.

64. Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, Wang X. 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proceedings of the National Academy of Sciences of the United States of America 114:7349-7354.

65. Quan J, Tian J. 2011. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. Nat Protoc 6:242-51.

66. Martinez A, Grabar TB, Shanmugam KT, Yomano LP, York SW, Ingram LO. 2007. Low salt medium for lactate and ethanol production by recombinant *Escherichia coli* B. Biotechnol Lett 29:397-404.

67. Beall DS, Ohta K, Ingram LO. 1991. Parametric studies of ethanol production form xylose and other sugars by recombinant *Escherichia coli*. Biotechnol Bioeng 38:296-303.

68. Martinez A, Rodriguez ME, York SW, Preston JF, Ingram LO. 2000. Use of UV absorbance To monitor furans in dilute acid hydrolysates of biomass. Biotechnol Prog 16:637-41.

69. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-5.

70. Amann E, Ochs B, Abel KJ. 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301-15.

CHAPTER 5

CHEMICAL SIMILARITY SEARCHING USING THE SIGNATURE MOLECULAR DESCRIPTOR IDENTIFIES A NETWORK OF NATIVE MALATE EXPORTERS IN ESCHERICHIA COLI

Abstract

Identification of transporters of valuable end-products is essential to metabolic engineering efforts for biocatalysis of renewable chemicals, especially as production parameters increase. The identification of transport proteins for various metabolites is an arduous task however, often involving manual combing of literature/biological databases, and performing genetic screens. One way to rationally identify transport proteins of a desired function is to target proteins with chemically similar substrates in genetic screens, as many transporters are known to exhibit substrate promiscuity. In this work, we prospected for malate exporters using a cheminformatics approach in a previously developed *Escherichia coli* L-malate biocatalyst, XZ658. To do this, we first took a traditional approach and inactivated *dcuABC* to demonstrate that the major L-malate uptake system only contributes minimally to efflux. Using the Signature molecular descriptor to compare chemical similarity to our target molecule, we found citrate, tartrate, and succinate to be the most similar known transportable C_3-C_7 di- or tricarboxylate metabolites to malate in *E. coli*. Inactivating known transporters of these compounds in addition to *dcuABC* revealed that *ttdT* and *citT* decrease extracellular malate titer after 48h by 56% and 65%, respectively. Inactivation of *dcuA, citT,* and *ttdT* decreased malate titer, growth rate, and resulted in a swollen cell physiology, suggesting these transporters form a malate exporter network. Although export of malate does not appear to be an immediate bottleneck in this biocatalyst based on biochemical analysis, this work demonstrates the utility of chemical similarity methods to hasten identification of transporters capable of efflux of target products of metabolism.

Importance

With the advent of synthetic biology, incorporation of metabolic pathways leading to a diverse non-native product portfolio has outpaced our ability to identify cellular export systems for molecules of interest. Transporter promiscuity however often permits the export of products from biocatalysts, but as production parameters increase, the probability of an export-based bottleneck

also increases. This work investigates if comparison of target compound similarity coupled to biological database mining can be used to identify transporters with export activity for a non-natural accumulation product of *E. coli,* L-malate. We find that annotated transporters for metabolites with similar molecular signatures to L-malate have promiscuous transport activity for malate export, demonstrating the utility of this approach. Additionally, we show that the major uptake system of L-malate is not the same as export, implying that a literature search may have led to only partial identification of transporters catalyzing export in this case.

Introduction

Over the last decade technological advances in the field of synthetic biology have permitted rapid and complex modification of metabolic pathways in a wide variety of hosts. Tools such as multiplex automated genome engineering (1) and CRISPR-based genome editing (2) now permit high-throughput genome manipulations, the biological effects of which being quantifiable using "omics" technologies (3, 4). Additionally, computational tools and metabolic models have been developed to probe biochemical pathways *in silico*, permitting optimization of metabolic flux (5), design of heterologous biosynthetic pathways (6), and more with less experimental cost. Utilizing a combination of these tools, it is now possible to create biocatalysts capable of a diverse product portfolio and quickly optimize metabolic pathways for enhanced production of many target molecules. However, many of these valuable products are not normally overproduced by the desired host, leading to various potential shortcomings in the native host machinery for optimal production metrics.

One of these potential genetic shortcomings is a lack of transport systems for efficient export of the non-native products to the extracellular space. Native transport systems can facilitate product export at low production metrics, but the kinetics and/or expression may be sub-optimal at economically relevant metrics. Thus, export has been previously observed as a bottleneck for the bioproduction of non-native products, and engineering export systems has enhanced production metrics for multiple compound classes including biofuels (7-9), diamines (10), amino acids (11, 12), and C_4 dicarboxylic acids (13-15). Although elucidation of bioproduct transporters is important to enhance rational metabolic engineering strategies, it remains a relatively understudied area due to

difficulties in rationally predicting transport activity and designing genetic screens. However, since many transporters exhibit substrate promiscuity for chemically similar substrates, it may be possible to identify transporters by choosing genetic targets based on ligand similarity to the target molecule.

Chemical similarity searching has been widely applied to elucidate molecules of similar biological activity to target compounds in the pharmaceutical industry (16, 17). This similarity searching is performed by using a molecular descriptor to transform structural and chemical information of a compound into a quantifiable format which can then be compared to a database of compounds transformed in the same manner. Although there are a multitude of potential descriptors to represent chemicals, fingerprint methods have been shown to yield biologically meaningful results for prediction of enzyme-substrate promiscuity (18), and thus are of interest in predicting transporter promiscuity. The Signature molecular descriptor (19) is a fingerprint method that has previously been successfully used to design and characterize small molecule protein inhibitors (20), "green" industrial solvents (21), polymers (22), as well as to investigate the potential for substrate promiscuity of enzymes (23). Thus, this descriptor may be useful in predicting transporter promiscuity for a cheminformatics-based targeted genetic screen.

L-malate is a C_4-dicarboxylate that has been identified by the Department of Energy as a compound with a multibillion dollar market that can be renewably derived using biomass sugars (24). For many hosts, L-malate is a non-native metabolic end-product and thus may be subject to export bottlenecks. Although multiple biocatalysts have been engineered for L-malate production (14, 25-27), high titer formation has only been achieved in multiple hosts after overexpression of a fungal L-malate exporter, MAE1p (13, 14). In *E. coli*, multiple transporters have been characterized to be responsible for promiscuous uptake, exchange and export of C_4-dicarboxylates under microaerobic conditions, however no L-malate exporters have been identified. These often promiscuous carriers include proteins from the Dicarboxylate uptake (Dcu), Dicarboxylate uptake C (DcuC), Divalent anion-sodium symporter (DASS), Auxin efflux carrier (AEC), Tripartite ATP-independent periplasmic (TRAP), Acetate uptake transporter (AceTr), and Dicarboxylate/amino acid cation symporter (DAACS) families (28). Of these, Dcu transporters are responsible for L-malate uptake during malate respiration (29), and are postulated to contribute to export via a

reversal of the uptake mechanism (28, 30, 31). Thus, these transporters represent targets for a traditional literature based genetic screen to discover the malate export system.

In this work, we used a reverse genetics approach to characterize the L-malate export system of *Escherichia coli* using an engineered L-malate producer. To begin we inactivated the major microaerobic L-malate uptake system (*dcuABC*) to demonstrate that the uptake system is not solely responsible for L-malate export. Thus five other candidate L-malate exporters were chosen by evaluating chemical similarity of malate to other transported metabolites. By inactivating transporters of compounds with similar molecular fingerprints to malate, DcuA, CitT, and TtdT were characterized to play a role in the malate export system, and inactivation of all three transporters led to reductions in extracellular malate. We further demonstrate that L-malate export does not appear to be a bottleneck for malate production in this host.

Results

Identification of candidate genes for native malate export.

To assess the role that the major anaerobic malate importers play in L-malate export, the genes *dcuA, dcuB,* and *dcuC* were inactivated both individually and in combination in the L-malate biocatalyst XZ658 (25). Inactivation of these three genes resulted in an ~20% decrease in extracellular malate production after 48 h, and a decrease in biomass in the first 48 h (Figure 5.1). This suggests that DcuA, DcuB, and DcuC only contribute minimally to malate export, and that a reversal of uptake may not be the main mechanism for malate export. Thus, we decided to use chemical similarity searching to select new candidates for our reverse genetics approach.

To generate a new list of candidate exporters, we prospected for transporters capable (or predicted) of having catalytic activity for metabolites that are the most chemically similar to malate. We chose to use the Signature molecular descriptor for chemical comparisons (19) and developed a pipeline to compare all transportable C_3-C_7 di- or tricarboxylate metabolites to malate, according to the Transporter Classification Database (TCDB) (32). This analysis found citrate, tartrate and succinate to be the most similar compounds to malate with *E. coli* transporters using this descriptor (Figure 5.2). Of these, citrate had a similarity of 95.5%, tartrate 95.4% and succinate 91.8%. Other compounds such as isocitrate had a higher similarity to malate, but no known *E. coli* transporters

(Figure 5.2A). By allowing transporters with unknown substrates to inherit the substrates of *E. coli* proteins in the same TCDB family we found nine transporters that had known/inherited catalytic activity for the top three metabolites, including *dcuA, dcuB, dcuC, dcuD, dauA, dctA, ybhI, citT,* and *ttdT* (Figure 5.2B). Expression and transport of C_4-dicarboxylates by *dctA* is not active under microaerobic conditions, and it was therefore removed from the target list (33). We then inactivated the remaining transporters in this list (*citT, ttdT, ybhI, dauA,* and *dcuD*) in a *dcuABC* knockout background. An additional succinate transporter, *satP,* was discovered and incorporated into the TCDB after our initial evaluation, but for experimental purposes was not included in our analysis (34).

DcuA, CitT, and TtdT contribute to malate export.

To first discern which member of DcuA, DcuB, and DcuC was responsible for the 20% decrease in the malate titer seen initially (Figure 5.1), we inactivated each gene individually in XZ658. After 48 h only DcuA resulted in a 23% decrease in malate titer, approximately equivalent to GK153 (Figure 5.3). We then proceeded to inactivate all candidate malate exporters individually in GK153 and fermented all strains. This demonstrated that inactivation of *citT* or *ttdT* resulted in 56% and 65% decreases of extracellular malate titer after 48h compared to XZ658, respectively (Figure 5.3). However, most of this deficiency recovered by 96h, suggesting that additional malate exporters were still present (Figure 5.4). All other candidate malate exporters exhibited no significant difference in malate production compared to GK153 (Figure 5.3 and 5.4). To investigate which transporter had the greatest physiological relevance, all potential combinations of *dcuA, citT,* and *ttdT* inactivations were evaluated. All individual inactivations presented deficient phenotypes at 48 h suggesting each transporter contributes to malate export (Figure 5.5A). However, inactivation of *ttdT, dcuA,* or *citT* resulted in a 31%, 18%, and 10% decrease in titer at 96h, respectively, suggesting *ttdT* is most important for achieving final malate titer. Furthermore, simultaneous inactivation of *dcuA* and *ttdT* was found to generate a deficient phenotype similar to the triple knockout (GK1TC), with a 73% decrease in titer (Figure 5.5B). This is greater than the titer decrease attributed to the individual inactivation of either *dcuA* or *ttdT*, suggesting that there is a synergistic relationship between these two transporters. Individual inactivation of *citT* only

presented deficient phenotypes at 48h (Figure 5.5A), suggesting that it may only be important for early growth/production. These results demonstrate that relevant exporters for malate may follow a hierarchy from greatest to lowest impact of *ttdT* > *dcuA* > citT.

Transport deficient mutants have defects consistent with export deficient cells.

To further understand the effects of transporter inactivation in GK1TC, we fermented this strain and compared the cell physiology, growth, and intracellular metabolite profile to XZ658. This transporter deficient mutant was found to have a large decrease in growth rate under microaerobic conditions and a 50% decrease in maximum biomass accumulation after 96 h (Figure 5.6A and B). However, this growth deficiency was not observed during aerobic growth in rich or minimal media (Figure 5.7), when metabolism favors pyruvate accumulation. We hypothesized that this deficiency may be due to an accumulation of malate which may be capable of altering metabolic flux to less desirable fermentation end-products under anaerobic conditions, such as citrate, isocitrate, and pyruvate (Figure 5.6C). To better understand the metabolic changes upon inactivation of these transporters, we extracted and quantified intracellular metabolites of both XZ658 and GK1TC at 48 h. We found that intracellular malate concentrations were equivalent between the two strains, however an ~2-fold increase in the concentration of pyruvate, citrate, and isocitrate could be observed in the transport deficient cells (Figure 5.6D). During fermentation, transporter deficient cells were also observed to have an enlarged phenotype at 48 h, corresponding to an approximately 50% increase in cell length relative to XZ658 (Figure 5.8). This increase in cell size also corresponded to a decrease in cell viability (Figure 5.8). To investigate if these putative malate exporters could be applied to enhance malate production in XZ658, we overexpressed each of them individually on a plasmid construct in XZ658. Overexpression of each transporter individually did not improve L-malate titer, productivity or yield in XZ658 when fermenting in NBS supplemented with 5% (w/v) glucose (Figure 5.9).

Discussion

In this work, we originally prospected for the L-malate exporter in *E. coli* by taking a limited literature-based reverse genetics approach and targeting active importers of malate under microaerobic conditions. This essentially investigated if a reversal of the malate uptake reaction

was the main mechanism of export. We found that this only identified *dcuA* as a putative malate exporter. A cheminformatics approach of selecting candidates based on chemical similarity to the target compound was effective in finding other transporters, *citT* and *ttdT*. A triple mutant of *dcuA*, *citT*, and *ttdT* was found to be growth deficient under microaerobic conditions, had decreased extracellular malate, and had altered cell physiology. Furthermore, it appeared that potentially intracellular malate may be being shuttled to other metabolic end-products in these transport deficient cells, suggesting that accumulation of malate may be toxic to these cells. This all suggests that these transporters function as malate export proteins in *E. coli*.

Of the Dcu and DcuC family proteins, only DcuA was found to have significant catalytic activity for malate export. This protein is known to catalyze the electroneutral exchange of L-malate, fumarate, succinate and aspartate by antiport of succinate from the cytoplasm, but has been shown to be capable of proton mediated uptake as well in the absence of a counter-substrate (29, 35). Interestingly, DcuB is a paralog of DcuA known to share these same transportable substrates, however only DcuB and DcuC are responsible for physiologically relevant succinate export (15). Since *dcuA* is known to be constitutively expressed (36), this difference in substrate specificity is hypothesized to not be due to differences in expression. Likewise, we find in our work that DcuA is capable of putative malate export, but not DcuB, further suggesting that these two proteins have significant differences in substrate specificity. We do not believe this is due to glucose repression of *dcuB* (36), as previous work has observed that both DcuB and DcuC are functionally active in glucose fermentation conditions with a C4 dicarboxylate production strain (15).

CitT and TtdT are both proteins belonging to the DASS protein family, a group known to be capable of transport of multiple organic di- and tricarboxylate substrates. More specifically, CitT has been shown to function as a citrate/succinate antiporter during citrate fermentation, and can potentially use citrate or fumarate as alternative counter-substrates (37). TtdT has been described as a L-tartrate/succinate antiporter, but can potentially use fumarate or succinate as counter-substrates as well (38). This work demonstrates that CitT and TtdT both likely function in L-malate export during glucose fermentation, although the mechanism of export remains uncharacterized. CitT, TtdT, YbhI and SOT1 (a spinach 2-oxoglutarate:malate translocator), form a cluster within the

DASS family and maintain a sequence identity >33% between each other (37). SOT1 has been characterized to catalyze exchange of 2-oxoglutarate for stromal malate, succinate, 2-oxoglutarate, or fumarate from the inner membrane of chloroplasts (39, 40), further supporting that these transporters may all function in malate export. Although YbhI had no evidenced role in malate export and has no evidenced di-/tricarboxylate substrates, this may be due to a lack of expression.

In the malate producer XZ658, *dcuA, citT,* and *ttdT* all appear to play a functional role in malate export. Our combinatorial deletions suggest that malate export is mainly performed by *ttdT* and *dcuA*, while *citT* is of lesser importance. Although engineering export of L-malate has previously been demonstrated to be an effective strategy for increasing production in fungi (13, 14), we observed no improvement in production while overexpressing *dcuA, citT,* or *ttdT*. This suggests that either the L-malate transport system of *E. coli* is sufficient for current production metrics, or there is a limitation of counter ions for antiport. However, it is notable that cells are capable of maintaining ~45% final titer (144 h) even with inactivated transport systems. This might suggest that there are still multiple other promiscuous transporters for L-malate in *E. coli* and that export may not be the immediate bottleneck.

For metabolic engineering, the use of chemical similarity searching provides another tool to help elucidate enzymes with promiscuity for desired substrates. We demonstrate here that annotated transport function for the most chemically similar substrates to malate reveals *citT* and *ttdT* to be top candidates for a genetic screen, and that these candidates have activity for the target compound. One limitation of this approach for transporter discovery however is that the tested substrates during enzyme characterization of transporters are often few and of a relatively narrow chemical space (18). This emphasizes the importance of using algorithms to intelligently select substrates that will maximize information gained on enzyme versatility during characterization. Additionally, this approach may also be further improved by incorporating sequence data to allow comparisons of both chemical and sequence similarity of a global dataset to the desired organism proteome.

Materials and Methods

Strains, and Cultivation Conditions

All strains used in this work are listed in Table 5.1. A previously engineered *Escherichia coli* malate producer, XZ658 (25), was used for genetic manipulations. For strain construction, all manipulations were done in Luria Broth (10 gL^{-1} Difco tryptone, 5 gL^{-1} yeast extract, and 5 gL^{-1} NaCl) at 30°C, 37°C, or 39°C with rotation at 180rpm when in liquid culture. During genetic manipulations, 5% arabinose (w/v) was added and ampicillin (100 mg L^{-1}), kanamycin (50 mg L^{-1}), and/or chloramphenicol (50 mg L^{-1}) were supplemented as needed.

Genetic Methods

Gene inactivation was performed as previously described using Red recombinase technology and either two-step recombination (41) or one step inactivation (42). Chromosomal integration was performed using PCR products amplified to encode either a *cat-sacB* counter selection cassette from pXW1(43) or a FRT-*kan*-FRT casette from pKD4 flanked with 50bp of homology to the target gene of interest. Primers for the generation of all inactivation fragments and plasmid constructs can be found in Table 5.2. For plasmid construction, circular polymerase extension cloning (CPEC) (44) was used to generate complete plasmids from linear PCR products using a pTrc99a vector backbone. All plasmids had the insert sequence verified using Sanger sequencing. Chemical transformation of constructs into Top10F' was performed using 50mM CaCl to generate competent cells.

Chemical similarity comparisons

To compare compound similarity to malate, we used the Signature molecular descriptor. Generating a molecular signature decomposes each metabolite into a vector of atomic signatures. These atomic signatures are a representation of the subgraph of each atom and all of its connected bonds/atoms up to a predefined distance "h" (19, 45). Thus the vector for comparison is a count of the occurrences of each atomic signature (both h=0 and h=1) within the molecule. We retrieved all data from the Transporter Classification Database (32) and extracted all substrate:transporter pairs where the substrate contained a substructure of a C_3-C_7 backbone and two or three terminal carboxyl groups (excluding malate). If a transporter in TCDB had listed unknown substrates or a

molecule class substrate related to our target substrates (e.g. dicarboxylates), we used transporter accession numbers to allow the transporter to inherit all substrates of *E. coli* transporters within the same family. This metabolite list then had their signature vectors compared to malate and similarity scored using a cosine similarity score. We considered the top three most similar metabolites (with designated substrates for an *E. coli* transporter) to be our top hits and used the transporters linked to these metabolites as new targets for putative malate exporters.

Fermentation

Fresh colonies grown microaerobically on NBS (New Brunswick Scientific) mineral salts (47) agar supplemented with 2% glucose (w/v) were transferred into 100 mL NBS 2% glucose (w/v) in a 250 mL flask. After growth for approximately 16 h (37°C, 120 rpm), this culture was used as the seed inoculum for fermentation. Fermentation was performed in continuously stirred 500 mL fermentation vessels containing 300mL NBS supplemented with 5% glucose (w/v), 100 mM potassium bicarbonate, and 10 mM sodium acetate (37°C, 120rpm). An initial inoculum of 0.1 OD_{550} was used for all fermentations. Medium pH was maintained at 7.0 during fermentation by automatic additions of base (2.4M potassium carbonate and 1.2M potassium hydroxide).

Light Microscopy

During fermentation, 1.25mL of cells from the bioreactor were extracted and centrifuged for 5 minutes at 7,000 x g. The resulting pellet was resuspended in approximately 100 µL of the supernatant and a drop transferred onto the surface of a glass slide. Cells were covered with a 22 x 22-mm glass cover slip before being viewed with a Axioskop microscope (Carl Zeiss, Thornwood, NY) using differential interference contrast optics (Plan-Neofluar 100 ; 1.3-numerical-aperture oil immersion objective lens and 1.4-numerical-aperture oil immersion condenser lens). Scale bars were added prior to taking images using a Hamamatsu C3200-07 video camera coupled to an analog camera control unit (Hamamatsu Photonic Systems Corporation, Bridgewater, NJ) and an Argus 10 image processor digitally enhanced the contrast in real time. Using ImageJ, a scale bar was calibrated to the original scale bar and bacteria in random fields of multiple samples from each culture were selected for cell measurements of length and width (n =100 cells).

Viable cell counts

To determine viable cell counts 1.0 ml of fermentation broth was extracted over the course of fermentation and normalized to an OD = 0.1 before being subjected to a 10-fold serial dilution in Luria broth. Serially diluted cells were plated on LB and incubated for 24 h at 37°C prior to being manually counted. To account for difference in cell size influencing optical density, cells density and diameter were evaluated using a Multisizer 3 particle counter (Beckman Coulter). Duplicates of XZ658 and GK1TC were normalized to an OD_0 = 0.6 and placed on ice. Subsequently, 50 µL of each normalized sample was diluted in 10 mL of ISOTON® II Diluent (Beckman Coulter) and cell density and average cell diameter were quantified with a 20 µm aperture tube.

Intracellular Extraction

To quantify intracellular metabolites, we used a modified version of a previously described protocol (48). Briefly, 1 OD of cells was extracted from the fermentation and centrifuged at 7,000 x g for 5 minutes at 4°C. The supernatant was removed and cells were washed once with 1 ml of 1X phosphate buffered saline before being resuspended in a methanol-water mixture (80/20) chilled at -80°C and mixed vigorously. After 30 minutes of incubation at -80°C the mixture was centrifuged at 7,000 x g for 10 minutes and the soluble fraction was transferred into a new tube. We then resuspended the pellet in another 500 µL methanol-water mixture and incubated for 15 minutes at 80°C before the cells were centrifuged at 7,000 x g for 10 minutes. The remaining soluble fraction was then transferred to the same tube used previously and dried prior to GC-MS analysis.

HPLC

All quantification of glucose and organic acids was performed using high performance liquid chromatography (HPLC), as previously described (43). More specifically, compounds were separated and quantified with an UltiMate3000 (Thermo Fisher) using an Aminex HPX-87H column (Bio-Rad) and 4mM sulfuric acid as the mobile phase.

GC-MS

For quantification of intracellular metabolites we used GC-MS and followed the general procedures for the Agilent Fiehn GC-MS Metabolomics RTL library (51, 52), with minor changes incorporated to improve detection sensitivity. A solution of myristic acid-d_{27} (5 µL) from the Fiehn GC/MS Metabolomics Standards Kit was added as an internal standard for retention time locking.

For the derivatization process, the samples were first oximated by adding 30 µL of O-methylhydroxylamine hydrochloride solution (in pyridine) at 30°C for 90 min. The samples were then derivatized using 70 µL N-methyl-n-trimethylsilyltrifluoroacetamine with 1% chlorotrimethylsilane (MSTFA+1% TMCS) at 37 °C for 30 min. Subsequently, the solution was gently vortexed and transferred to a GC-MS glass vial for analysis.

GC-MS experiments were performed on an Agilent 7820 GC-5977 MSD system (Agilent Technologies, Santa Clara, CA) by injecting 1 µL of the prepared samples in the splitless mode. Helium was used as the carrier gas. The separation of metabolites was achieved using an Agilent DB5-MS+10m Duraguard Capillary Column (30 m x 250 µm x 0.25 µm). The column temperature was maintained at 60 °C for 1.00 min, then increased at a rate of 10 °C/min to 325 °C, and held at this temperature for 10 min. Mass spectral signals were recorded after a 4.90 min solvent delay.

Figures

Table 5.1. Strains and plasmids used in Chapter 5

	Relevant characteristics	Reference
Strain		
Top 10F'	F´lacIq, Tn10(TetR) mcrA Δ(mrr-hsdRMS-mcrBC) Φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara leu) 7697 galU galK rpsL (StrR) endA1 nupG	Pharmacia
XZ658	*E. coli* ATCC8739 Δ*ldhA* Δ*ackA* Δ*adhE* Δ*pflB* Δ*mgsA* Δ*poxB* Δ*frdBC* Δ*maeB* Δ*fumABC* Δ*sfcA*	(25)
GK001	XZ658Δ*dcuA*	This study
GK003	XZ658 *dcuC*::FRT	This study
GK005	XZ658Δ*dcuB*	This study
GK015	GK001 *dcuB*::FRT	This study
GK013	GK001 *dcuC*::FRT	This study
GK053	GK005 *dcuC*::FRT	This study
GK153	GK015 *dcuC*::FRT	This study
GK153T	GK153 *ttdT*::FRT	This study
GK153C	GK153 *citT*::FRT	This study
GK153D	GK153 *dcuD*::FRT	This study
GK153A	GK153 *dctA*::FRT	This study
GK153U	GK153 *dauA*::FRT	This study
GK153Y	GK153 *ybhI*::FRT	This study
GK153TC	GK153T *citT*::FRT	This study
GK00T	XZ658 *ttdT*::FRT	This study
GK00C	XZ658 *citT*::FRT	This study
GK01T	GK001 *ttdT*::FRT	This study
GK01C	GK001 *citT*::FRT	This study
GK0TC	GK00T *citT*::FRT	This study
GK1TC	GK01T *citT*::FRT	This study
Plasmids		
pKD46	bla, γ β exo (Red recombinase)	(42)
pCP20	*bla, cat,* FLP+, ts-rep, cI857λts	(49)
pKD4	bla, FRT-kan-FRT	(42)
pXW001	*cat-sacB* cassette with the *sacB* native terminator cloned into a modified vector pLOI4162	(43)
pTrc99A	P*trc bla oriR rrnB lacI*^{q}	(50)
pTrc99A-dcuA	*dcuA* in pTrc99A	This study
pTrc99A-citT	*citT* in pTrc99A	This study
pTrc99A-ttdT	*ttdT* in pTrc99A	This study

Table 5.2. Primers used in Chapter 5

Function	Sequence
CPEC cloning	5' CAATTAATCATCCGGCTCGT (for sequencing)
into pTrc99A	5' ATTTAGTCTTGCGTCTTCGC (for sequencing)
	5' GGATCCTCTAGAGTCGACCTGCAGG
	5' GTACCTTAAGCTCGAGCCATGGGC
cloning *dcuA*	5' CATGGAATTCGAGCTCGGTACCCGTCAACTTGTTAAAAAACAAGGAAGGC
into pTrc99A	5' AGGGCGGCTGATTGTAGGAAGCCTAGGAGATCTCAGCTGGACGTCC
cloning *citT* into	5' CATGGAATTCGAGCTCGGTACCCGTGGAAATATTAATTTTCGGAGAACCCG
pTrc99A	5' GGTGGACCGCCCGAAAAAATACTAAATTACCTAGGAGATCTCAGCTGGACGTCC
cloning *ttdT* into	5' CATGGAATTCGAGCTCGGTACCCGCCCTCTCCTTTTTCGCAGGCATA
pTrc99A	5' CGGACGAAATTTGGCGTGGTACCTAGGAGATCTCAGCTGGACGTCC
dcuA deletion	5' TTAAAAATAACAATCTAATATCAACTTGTTAAAAAACAAGGAAGGCTAATTCGAGTGTGACGGAAGATCA
	5' TTTTCGTCCGTTTACCGATTCACTAGTTAGCGCCCCGCAAGTGCGGGGCGAAAGAAAGGGCGGCTGATTGT
	5' CAACTTGTTAAAAAACAAGGAAGGCTAATTGATCAATCGCGGGGCGT
	5' GTTGAACAATTTTTTGTTCCTTCCGATTAACTAGTTAGCGCCCCGCA
dcuB deletion	5' TGTGATCTATTCAGCAAAAATTTAAATAGGATTATCGCGAGGGTTCACACTCGAGTGTGACGGAAGATCA
	5' TTTTCGTCCGTTTACCGATTCACGTGAAACGCACGGCGGGCACTGATGCGCCGTGCGGTAAAAGCTTATT
	5' TAGGATTATCGCGAGGGTTCACACATGCACTTTGCGTGCCGC
	5' ATCCTAATAGCGCTCCCAAGTGTGTACGTGAAACGCACGGCG
dcuC deletion	5' CTTTAGGAATTTTTTATTATTTACTTTGGGGCCTGGAGACAGGAAAAATTTCGAGTGTGACGGAAGATCA
	5' TTTTCGTCCGTTTACCGATTCTATCATTACGGCCTCCGTTCAAAGGAGGCCGAAATAGTACAGTGGGTGAC
	5' GGCCTGGAGACAGGAAAAATTATAGTAATGCCGGAGGCAAGTTT
	5' CCGGACCTCTGTCCTTTTTAATATCATTACGGCCTCCGTTCAAA
dauA FRT KO	5' TCACCTGCCATTTAGTTGACAGATGATGCGCTCACGGATGAAACATTATTGTGTAGGCTGGAGCTGCTTC
	5' TGATTCCTCCTATAAGTATACCATGCTCTAACTGGTCAGTCGCGTTTGACTGACCAGTCGTTTGACGTAAT
dcuD FRT KO	5' AATCTGCCGTTTATGGGATTGACCGTTTTCTTTTGACACGGAGTTCAACAGTGTAGGCTGGAGCTGCTTC
	5' TGATTCCTCCTATAAGTATACCGGTAATTTTCTGTTTTGTCCGGTGGTCCCGGACAAAAACATAATGAAGT
ybhI FRT KO	5' AATCAATAAAGGACTTCTGTATGAGTCATACAGAAAGAACAGGATTTTAAGTGTAGGCTGGAGCTGCTTC
	5' TGATTCCTCCTATAAGTATACTATATGAGCAGTATGAAGTTCAACGTACACGACGCAAACGCAAGCGAGTG
citT FRT KO	5' TTATTTAAGCACTTGATAAATTTGGAAATATTAATTTTCGGAGAACCCGTGTGTAGGCTGGAGCTGCTTC
	5' TGATTCCTCCTATAAGTATACATGCAAGCGGTTTGCGACTTATTTAGTAGTGCGGGCGGTCCATTCGGGTG
ttdT FRT KO	5' TAACCCTCCCGGAGAGGCTCACCCCTCTCCTTTTTCGCAGGCATAACACGGTGTAGGCTGGAGCTGCTTC
	5' TGATTCCTCCTATAAGTATACCATTCCTGCCACGGCCTTTAACGAGAGGCCGTGGGCTATCAAAATGCGCC
dcuA integration	5' GCCGATAAGGCGGAAGCAGCACACATCAGAAATTCGTAATAAGGAGGTAGGTTATGCTAGTTGTAGAACTCAT
	5' ACCCATCGAAGTACGACATTCCTCGGACTCTACTAATTAGGTTTAGGTTTCGCAAGACTGCTTGGCCATA
citT integration	5' AGCGGATGCAGCCGATAAGGCGGAAGCAGCCAATAAGAAGGAGAAGGCGAATGACAGCGACACGCGGTC
	5' AAGCAAGGCTAACTTCTTTCTCGGATTCCTCGGACTCTACTAGTTAGGTTTAGGTTTCGCAAGACTGCTTGGCCAT
ttdT integration	5' GCCGATAAGGCGGAAGCAGCAAAGTACGCGGATTAGGGAGGTTAGGAATGTCTTTAGCAAAAGATAA
	5' ACGGGCACCACAACGAAATTCCTCGGACTCTACTAATTAGGTTTAGGTTTCGCAAGACTGCTTGGCCATA

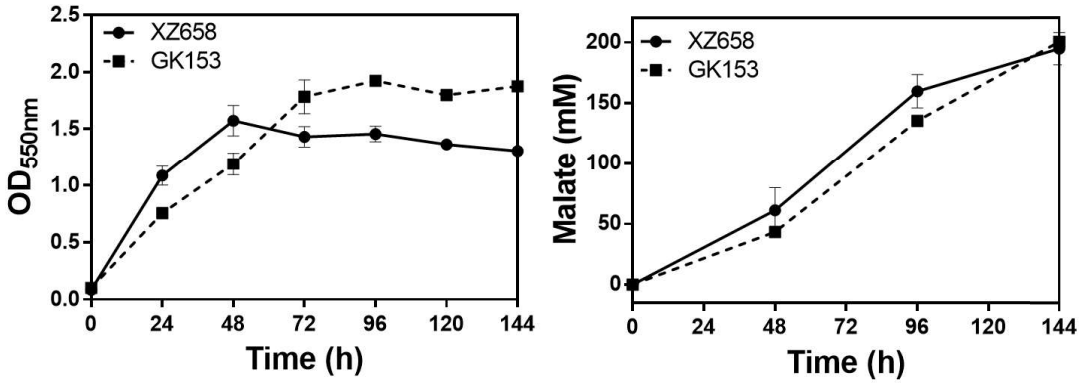
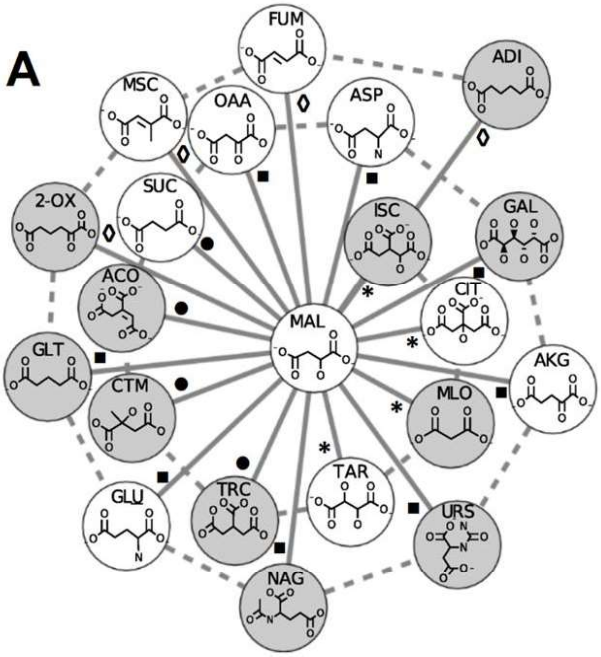


Figure 5.1. Batch fermentation of XZ658 and GK153 (XZ658 Δ*dcuA* Δ*dcuB* Δ*dcuC*) using 5% (w/v) glucose in NBS mineral salts medium. A) Cell optical density (OD550nm), and B) L-malate concentrations were determined. Error bars represent SEM with at least three independent experiments.



B

Gene	Substrates/Info	Citation
dcuD	Cryptic Dcu transporter; C4 dicarboxylates including **succinate**	(28)
dauA	**Succinate**/fumarate/asparatate uptake at acidic pH	(49)
ybhI	**Citrate/tartrate**; unknown; high similarity to *citT* and *ttdT*	(37)
ttdT	**L-Tartrate/succinate:succinate** antiporter	(38)
citT	**Citrate/succinate**/fumarate:***succin ate*** antiporter	(37)

Figure 5.2. Chemical similarity of all C_3-C_7 di- or tricarboxylate substrates in TCDB with known transporters compared to malate using the Signature Molecular Descriptor. A) Grey circles indicate substrates with no known *E. coli* transporter. Similarity based on pairwise distances between the chemical signatures of malate and all indicated chemicals is shown at each note with values of >95% (asterisk), 95-90% (circle), 90-85% (square), and 85%> (diamond) (Detailed data in Table S1). The length of solid lines also represents the relative similarity to malate. Abbreviations: MAL (malate), ISC (isocitrate), CIT (citrate), MLO (malonate), TAR (tartrate), TRC (tricarballylate), CTM (citramalate), ACO (trans-aconiate), SUC (succinate), OAA (oxaloacetate), ASP (aspartate), GAL (galactarate), AKG (2-oxoglutarate), URS (ureidosuccinate), NAG (N-acetylglutamate), GLU (glutamate), GLT (glutarate), 2-OX (2-oxoadipate), MSC (mesaconate), FUM (fumarate), and ADI (adipate). B) Candidate malate exporters selected based on substrate similarity to malate using Signature Molecular Descriptor and information from the previous literature. Three di/tri-carboxylates most similar to malate are highlighted in bold.

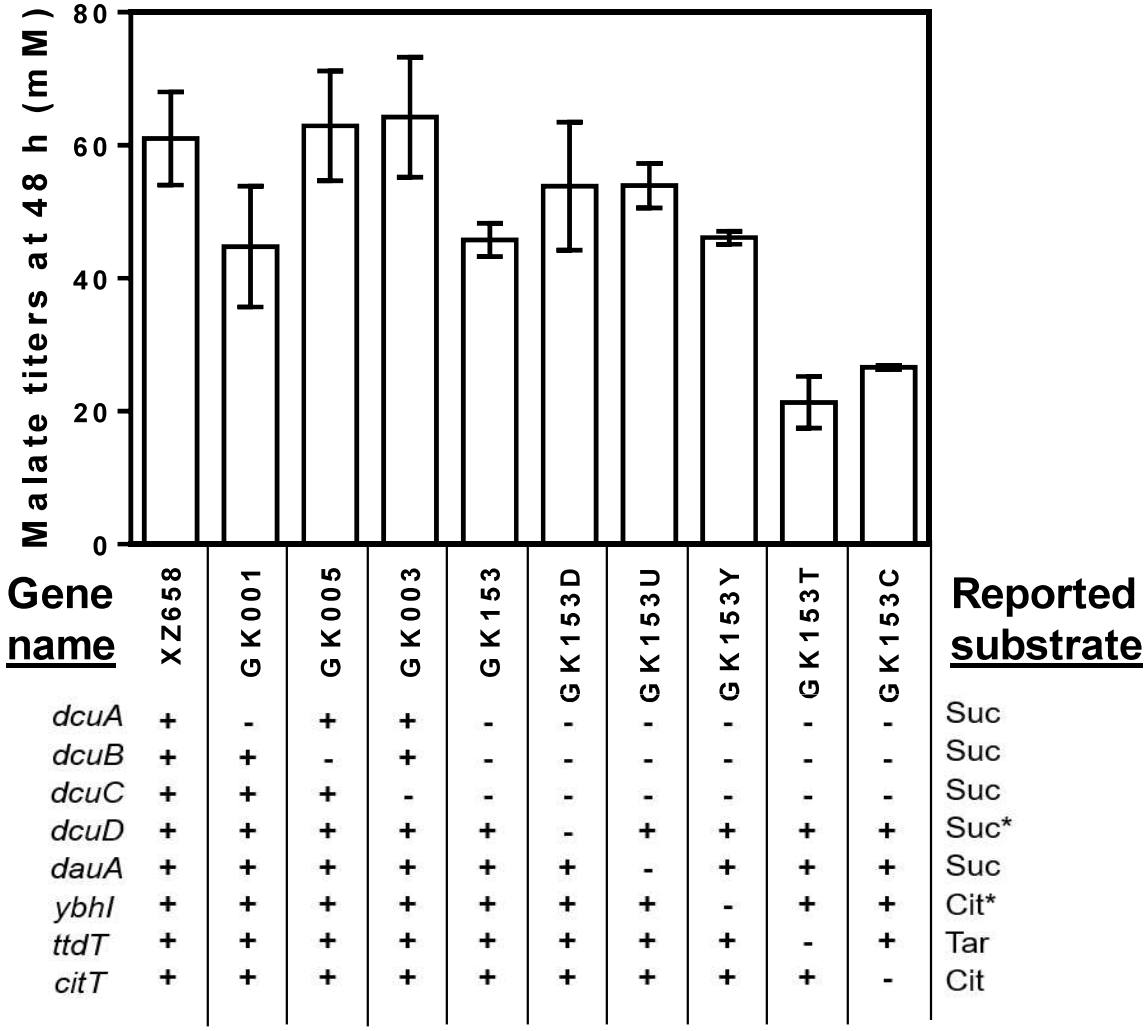


Figure 5.3. Extracellular malate titer of strains fermented in NBS/5% glucose measured at 48 h with gene knockouts corresponding to indicated transporters. Abbreviations: Suc (succinate), Cit (citrate), and Tar (tartrate). *Transport activity inferred from transporter family member from TCDB.

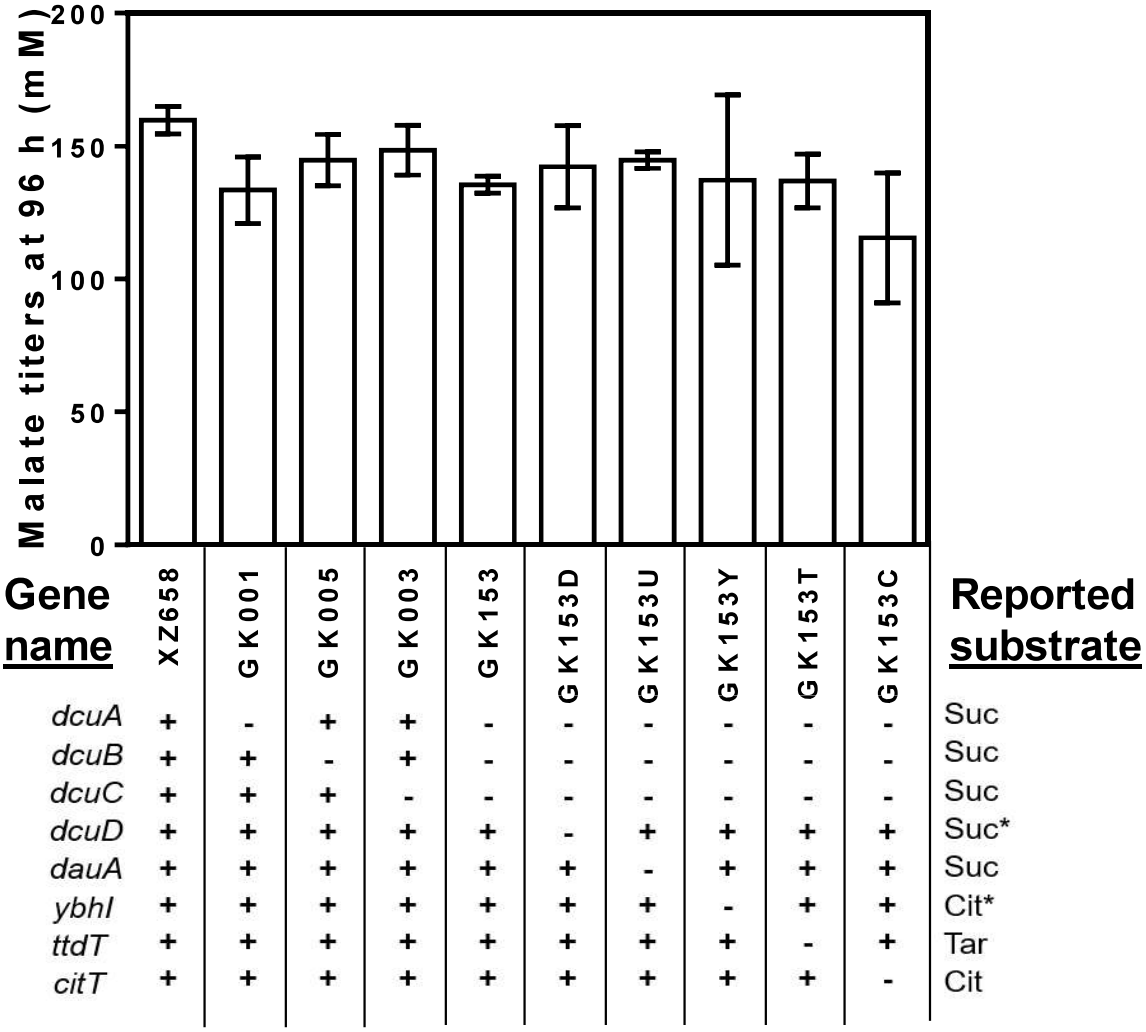


Figure 5.4. Extracellular malate titer of strains fermented in NBS/5% glucose measured at 96 h with gene knockouts corresponding to indicated transporters. Abbreviations: Suc (succinate), Cit (citrate), and Tar (tartrate). *Transport activity inferred from transporter family member from TCDB.

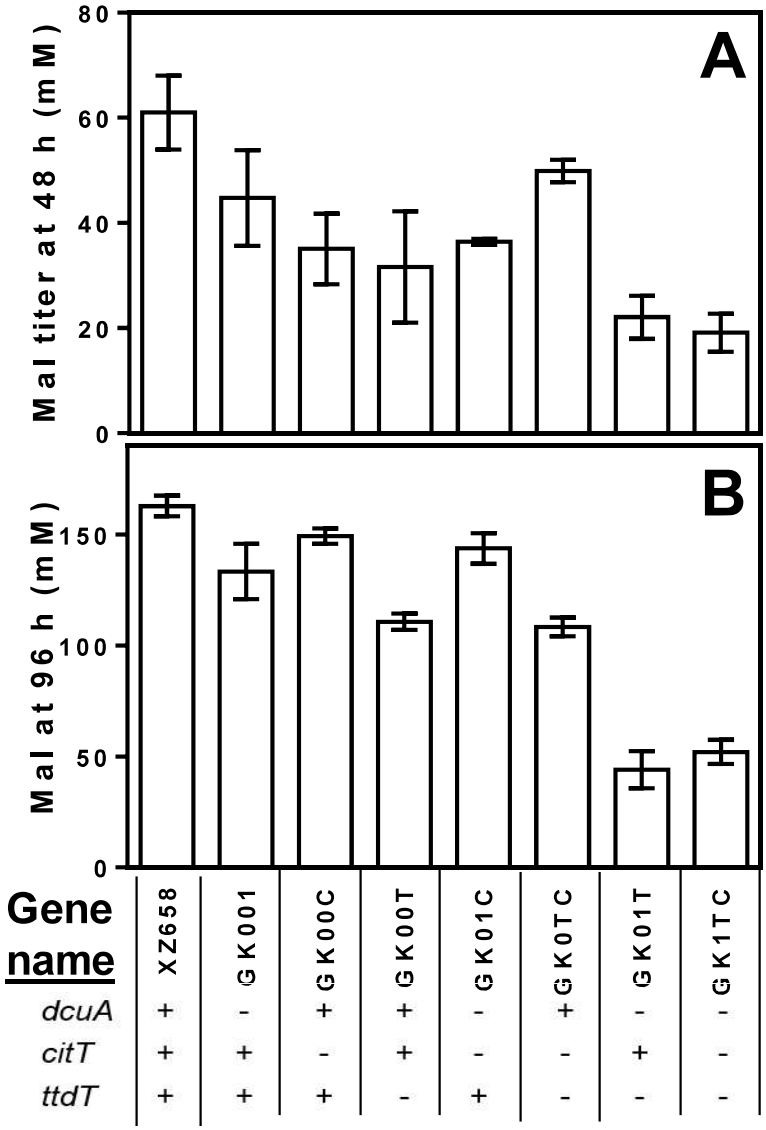


Figure 5,5. Extracellular malate titer of strains fermented in NBS medium supplemented with 5% (w/v) glucose measured at A) 48 h and B) 96 h with gene knockouts corresponding to indicated transporters. Abbreviations: Mal (malate).

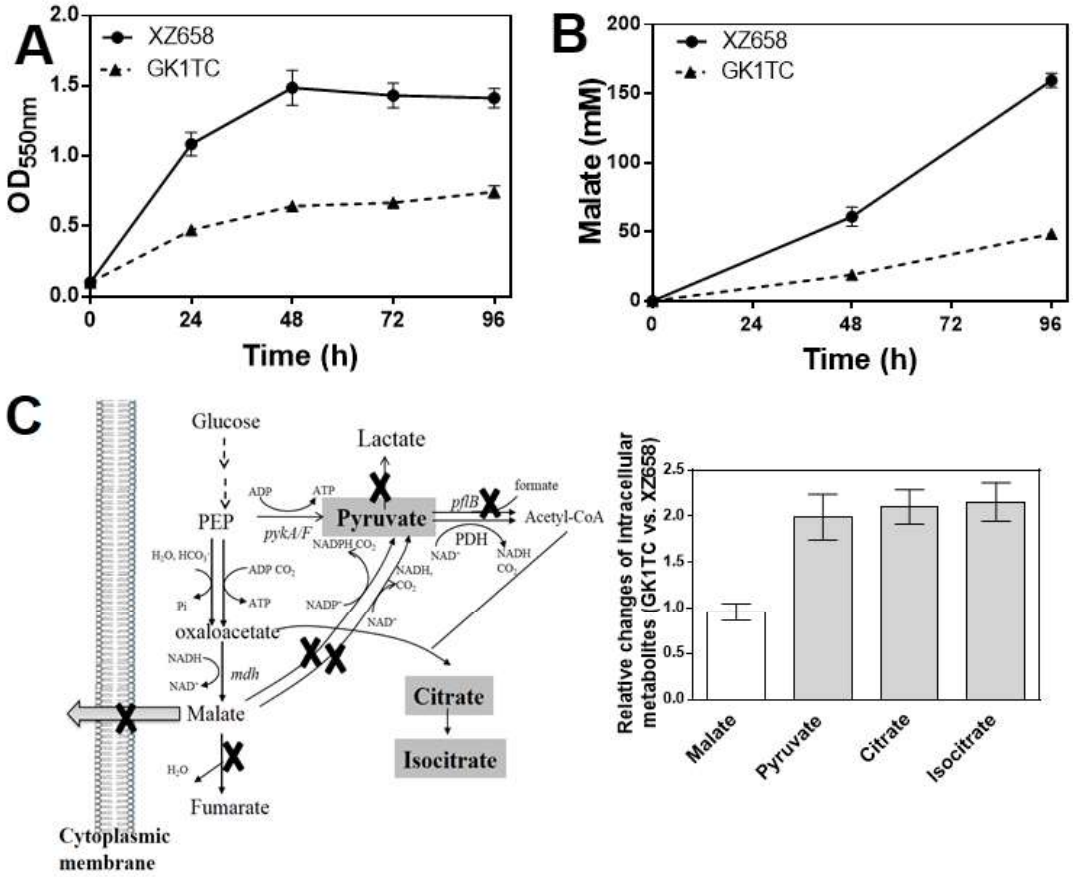


Figure 5,6. Extracellular malate concentrations and significantly changed intracellular metabolites of strains XZ658, and GK1TC (XZ658 Δ*dcuA* Δ*citT* Δ*ttdT*). Cells were fermented in NBS medium supplemented with 5% glucose. A) Cell optical density (OD550nm), B) L-malate concentrations were determined. C) The relative change of intracellular metabolites between GK1TC and XZ658 was determined by GC-MS for cell extractions. Error bars represent SEM with at least three independent experiments

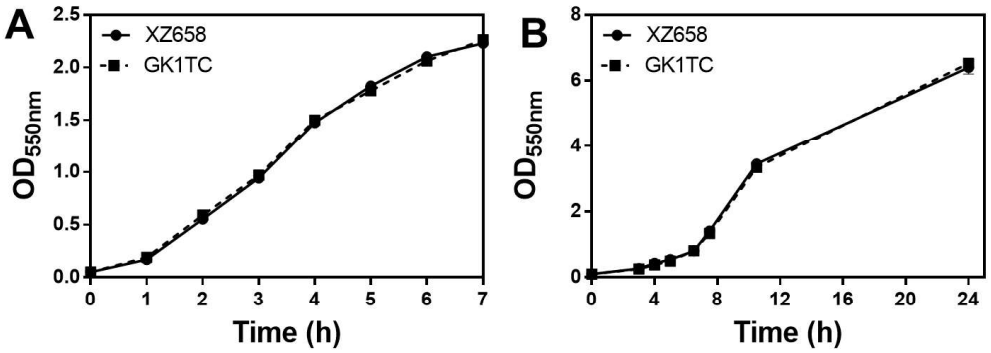


Figure 5,7. Growth of the strains XZ658, and GK1TC (XZ658 Δ*dcuA* Δ*citT* Δ*ttdT*) under non-fermentative conditions. A) Shake flasks at 200 rpm in LB media under aerobic conditions, B) Shake flasks at 200 rpm in NBS media supplemented with 2% glucose under aerobic conditions.

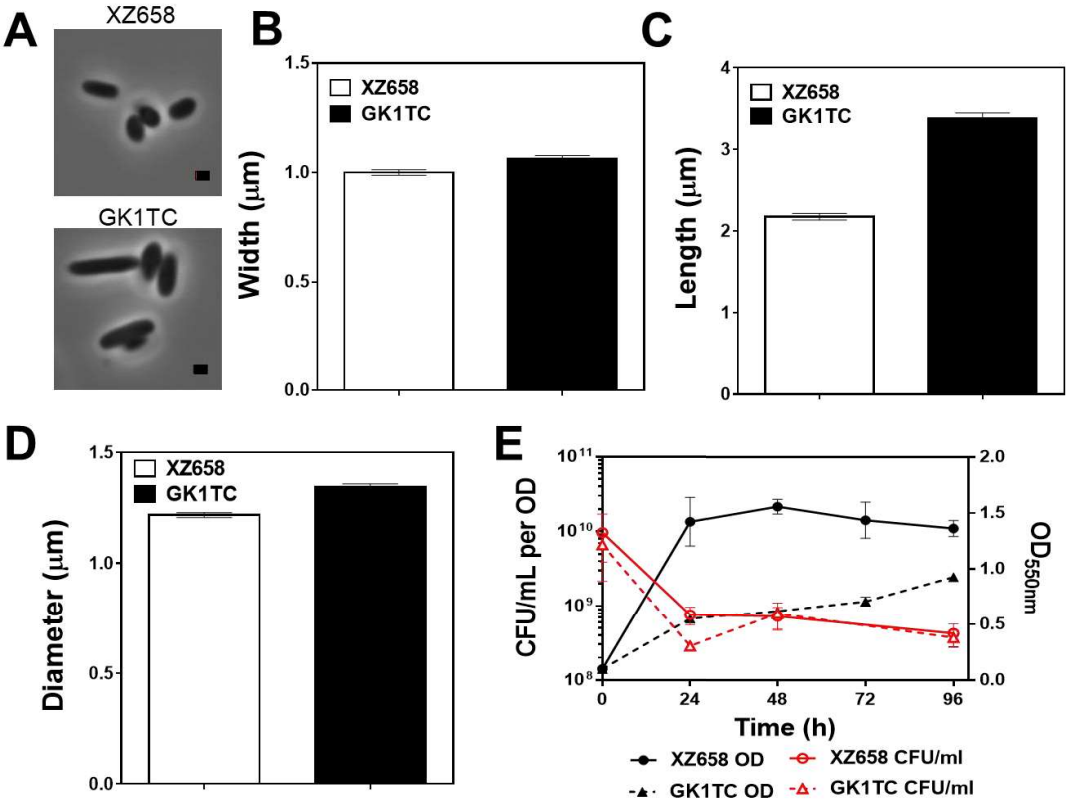


Figure 5,8. The morphological and viability changes of strains XZ658, and GK1TC (XZ658 Δ*dcuA* Δ*citT* Δ*ttdT*) under fermentative conditions. A) micrographs of XZ658 (top) and GK1TC (bottom). B) The cell width, C) length and D) diameter was estimated by counting >100 cells on micrograph images and diameter readout from a cell counter. D) Viability of XZ658 and GK1TC was measured by counting colony forming units (CFU) per mL on LB agar plate with a serial dilution. The different cell numbers per OD were normalized using a cell counter.

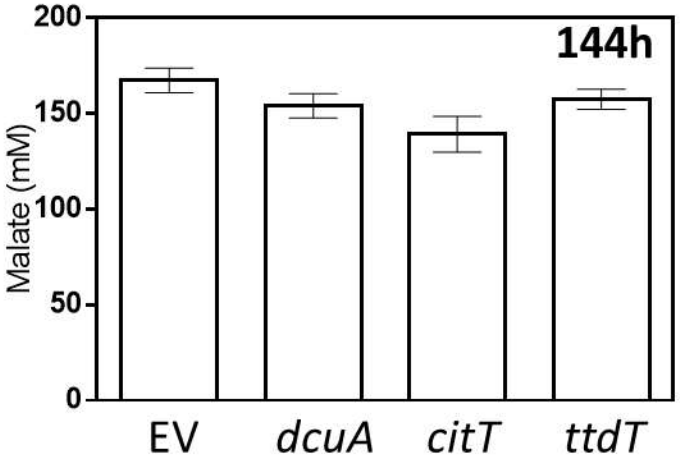


Figure 5,9. Malate titer at 144 h of XZ658 fermented in NBS supplemented with 5% glucose and 10 μM IPTG overexpression either *dcuA, citT* or *ttdT*

References

1. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460:894-898.

2. Luo ML, Leenay RT, Beisel CL. 2016. Current and future prospects for CRISPR-based tools in bacteria. Biotechnol Bioeng 113:930-43.

3. Chae TU, Choi SY, Kim JW, Ko YS, Lee SY. 2017. Recent advances in systems metabolic engineering tools and strategies. Curr Opin Biotechnol 47:67-82.

4. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. 2012. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol 8:536-46.

5. Lee KH, Park JH, Kim TY, Kim HU, Lee SY. 2007. Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol 3:149.

6. Feher T, Planson AG, Carbonell P, Fernandez-Castane A, Grigoras I, Dariy E, Perret A, Faulon JL. 2014. Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. Biotechnol J 9:1446-57.

7. Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, Keasling JD, Hadi MZ, Mukhopadhyay A. 2011. Engineering microbial biofuel tolerance and export using efflux pumps. Mol Syst Biol 7:487.

8. Foo JL, Leong SS. 2013. Directed evolution of an E. coli inner membrane transporter for improved efflux of biofuel molecules. Biotechnol Biofuels 6:81.

9. Fisher MA, Boyarskiy S, Yamada MR, Kong N, Bauer S, Tullman-Ercek D. 2014. Enhancing tolerance to short-chain alcohols by engineering the *Escherichia coli* AcrB efflux pump to secrete the non-native substrate n-butanol. ACS Synth Biol 3:30-40.

10. Kind S, Kreye S, Wittmann C. 2011. Metabolic engineering of cellular transport for overproduction of the platform chemical 1,5-diaminopentane in Corynebacterium glutamicum. Metab Eng 13:617-27.

11. Ginesy M, Belotserkovsky J, Enman J, Isaksson L, Rova U. 2015. Metabolic engineering of *Escherichia coli* for enhanced arginine biosynthesis. Microb Cell Fact 14:29.

12. Yamada S, Awano N, Inubushi K, Maeda E, Nakamori S, Nishino K, Yamaguchi A, Takagi H. 2006. Effect of drug transporter genes on cysteine export and overproduction in *Escherichia coli*. Appl Environ Microbiol 72:4735-42.

13. Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JM, van Dijken JP, Pronk JT, van Maris AJ. 2008. Malic acid production by Saccharomyces cerevisiae: engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. Appl Environ Microbiol 74:2766-77.

14. Brown SH, Bashkirova L, Berka R, Chandler T, Doty T, McCall K, McCulloch M, McFarland S, Thompson S, Yaver D, Berry A. 2013. Metabolic engineering of Aspergillus oryzae

NRRL 3488 for increased production of L-malic acid. Appl Microbiol Biotechnol 97:8903-12.

15. Chen J, Zhu X, Tan Z, Xu H, Tang J, Xiao D, Zhang X. 2014. Activating C4-dicarboxylate transporters DcuB and DcuC for improving succinate production. Appl Microbiol Biotechnol 98:2197-205.

16. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, Jensen NH, Kuijer MB, Matos RC, Tran TB, Whaley R, Glennon RA, Hert J, Thomas KL, Edwards DD, Shoichet BK, Roth BL. 2009. Predicting new molecular targets for known drugs. Nature 462:175-81.

17. Sliwoski G, Kothiwale S, Meiler J, Lowe EW, Jr. 2014. Computational methods in drug discovery. Pharmacol Rev 66:334-95.

18. Pertusi DA, Moura ME, Jeffryes JG, Prabhu S, Walters Biggs B, Tyo KEJ. 2017. Predicting novel substrates for enzymes with minimal experimental effort with active learning. Metab Eng 44:171-181.

19. Faulon JL, Collins MJ, Carr RD. 2004. The signature molecular descriptor. 4. Canonizing molecules using extended valence sequences. J Chem Inf Comput Sci 44:427-36.

20. Churchwell CJ, Rintoul MD, Martin S, Visco DP, Jr., Kotu A, Larson RS, Sillerud LO, Brown DC, Faulon JL. 2004. The signature molecular descriptor. 3. Inverse-quantitative structure-activity relationship of ICAM-1 inhibitory peptides. J Mol Graph Model 22:263-73.

21. Weis D, Visco D. 2010. Computer-aided molecular design using the Signature molecular descriptor: Application to solvent selection. Comput Chem Eng 34:1018-1029.

22. Brown WM, Martin S, Rintoul MD, Faulon JL. 2006. Designing novel polymers with targeted properties using the signature molecular descriptor. J Chem Inf Model 46:826-35.

23. Carbonell P, Faulon JL. 2010. Molecular signatures-based prediction of enzyme promiscuity. Bioinformatics 26:2012-9.

24. Werpy T, Petersen G. 2004. Top value added chemicals from biomass. U.S. Department of Energy, Washington, DC. http://www1.eere.energy.gov/biomass/pdfs/35523.pdf.

25. Zhang X, Wang X, Shanmugam KT, Ingram LO. 2011. L-malate production by metabolically engineered *Escherichia coli*. Appl Environ Microbiol 77:427-34.

26. Deng Y, Mao Y, Zhang X. 2016. Metabolic engineering of a laboratory-evolved *Thermobifida fusca* muC strain for malic acid production on cellulose and minimal treated lignocellulosic biomass. Biotechnol Prog 32:14-20.

27. Dong X, Chen X, Qian Y, Wang Y, Wang L, Qiao W, Liu L. 2017. Metabolic engineering of *Escherichia coli* W3110 to produce L-malate. Biotechnol Bioeng 114:656-664.

28. Janausch IG, Zientz E, Tran QH, Kroger A, Unden G. 2002. C4-dicarboxylate carriers and sensors in bacteria. Biochim Biophys Acta 1553:39-56.

29. Six S, Andrews SC, Unden G, Guest JR. 1994. *Escherichia coli* possesses two homologous anaerobic C4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the aerobic dicarboxylate transport system (Dct). J Bacteriol 176:6470-8.

30. Engel P, Kramer R, Unden G. 1994. Transport of C4-dicarboxylates by anaerobically grown *Escherichia coli*. Energetics and mechanism of exchange, uptake and efflux. Eur J Biochem 222:605-14.

31. Zientz E, Janausch IG, Six S, Unden G. 1999. Functioning of DcuC as the C4-dicarboxylate carrier during glucose fermentation by *Escherichia coli*. J Bacteriol 181:3716-20.

32. Saier MH, Jr., Reddy VS, Tsu BV, Ahmed MS, Li C, Moreno-Hagelsieb G. 2016. The Transporter Classification Database (TCDB): recent advances. Nucleic Acids Res 44:D372-9.

33. Davies SJ, Golby P, Omrani D, Broad SA, Harrington VL, Guest JR, Kelly DJ, Andrews SC. 1999. Inactivation and regulation of the aerobic C(4)-dicarboxylate transport (*dctA*) gene of *Escherichia coli*. J Bacteriol 181:5624-35.

34. Sa-Pessoa J, Paiva S, Ribas D, Silva IJ, Viegas SC, Arraiano CM, Casal M. 2013. SatP (YaaH), a succinate-acetate transporter protein in *Escherichia coli*. Biochem J 454:585-95.

35. Engel P, Kramer R, Unden G. 1992. Anaerobic fumarate transport in *Escherichia coli* by an fnr-dependent dicarboxylate uptake system which is different from the aerobic dicarboxylate uptake system. J Bacteriol 174:5533-9.

36. Golby P, Kelly DJ, Guest JR, Andrews SC. 1998. Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli*. J Bacteriol 180:6586-96.

37. Pos KM, Dimroth P, Bott M. 1998. The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. J Bacteriol 180:4160-5.

38. Kim OB, Unden G. 2007. The L-tartrate/succinate antiporter TtdT (YgjE) of L-tartrate fermentation in *Escherichia coli*. J Bacteriol 189:1597-603.

39. Menzlaff E, Flugge UI. 1993. Purification and functional reconstitution of the 2-oxoglutarate/malate translocator from spinach chloroplasts. Biochim Biophys Acta 1147:13-8.

40. Weber A, Menzlaff E, Arbinger B, Gutensohn M, Eckerskorn C, Flugge UI. 1995. The 2-oxoglutarate/malate translocator of chloroplast envelope membranes: molecular cloning of a transporter containing a 12-helix motif and expression of the functional protein in yeast cells. Biochemistry 34:2621-7.

41. Jantama K, Zhang X, Moore JC, Shanmugam KT, Svoronos SA, Ingram LO. 2008. Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. Biotechnol Bioeng 101:881-93.

42. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-5.

43. Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, Wang X. 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc Natl Acad Sci U S A 114:7349-7354.

44. Quan J, Tian J. 2014. Circular polymerase extension cloning. Methods Mol Biol 1116:103-17.

45. Carbonell P, Carlsson L, Faulon JL. 2013. Stereo signature molecular descriptor. J Chem Inf Model 53:887-97.

46. Faulon JL, Misra M, Martin S, Sale K, Sapra R. 2008. Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor. Bioinformatics 24:225-33.

47. Causey TB, Shanmugam KT, Yomano LP, Ingram LO. 2004. Engineering *Escherichia coli* for efficient conversion of glucose to pyruvate. Proc Natl Acad Sci U S A 101:2235-40.

48. Hasegawa S, Uematsu K, Natsuma Y, Suda M, Hiraga K, Jojima T, Inui M, Yukawa H. 2012. Improvement of the redox balance increases L-valine production by Corynebacterium glutamicum under oxygen deprivation conditions. Appl Environ Microbiol 78:865-75.

49. Cherepanov PP, Wackernagel W. 1995. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158:9-14.

50. Amann E, Ochs B, Abel KJ. 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301-15.

51. Kind T, Wohlgemuth G, Lee DY, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. 2009. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. Anal. Chem. 81 (24), 10038-10048.

52. Agilent Fiehn GC/MS Metabolomics RTL User Guide. 2008. Agilent Fiehn GC/MS Metabolomics RTL User Guide [Online],

CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS

Conclusions

A commercially successful lignocellulose conversion requires cost-effectiveness at all steps: liberating sugars, converting sugars into target products at high titer, yield and productivity, and downstream product recovery. There have been a lot of developments in pretreatment technologies and metabolic engineering strategies for sugar co-utilization and microbial robustness in the past few decades (1). In spite of so many research efforts, challenges such as sugar utilization, degradation recalcitrance, product toxicity, and lignocellulosic toxicity still remain at different degrees, thereby limiting further commercial developments.

Metabolic engineering strategies targeting membrane transport mechanisms for enhancing lignocellulose conversion have been scarce in literature, however multiple potential transporter bottlenecks exist. First, an abundant sugar in lignocellulosic biomass, xylose, is inefficiently transported in most bacterial biocatalysts. This results in a decrease in the energy budget which can be detrimental under microaerobic production conditions (2-6). Second, inhibitors such as furfural in lignocellulose hydrolysate impose cellular toxicity (1, 5, 7). Although cellular efflux has been identified as an important way that many bacteria survive other cytotoxic molecules, such as antibiotics, lignocellulose inhibitor efflux systems have not been identified (8). Third, engineering efflux of target products has shown to increase production metrics, however identification of efflux transporters is notoriously difficult. My work contributed to this field by investigating potential transporter mediated solutions for enhancing lignocellulosic bioconversion. By rationally attempting to probe 1) routes for efficient xylose uptake 2) inhibitor efflux solutions and 3) methods for identifying product exporters, this work has helped lead to the following conclusions.

Energy can be conserved by engineering native and heterologous sugar transporter systems

Although ATP yield has been shown to be decreased under xylose fermentation conditions due to inefficient transport mechanisms, very few robust solutions have been implemented to improve

this (2). The main solution has been to inactivate XylFGH and/or allow adaptive laboratory evolution to naturally edit the genome to enhance ATP conservation (3, 4). Although this previously identified a mutation in *galP* that enhanced xylose to succinate conversion, the authors previously attributed this to enhanced xylose transport affinity by GalP (9). **In Chapter 2**, I have shown that this conclusion is incorrect, and that the true mechanism enhancing xylose to succinate conversion in KJ122 is repression of GalP. I have further shown that inactivation of GalP likely enhances the ATP yield by modulating expression of the anaplerotic enzymes (Ppc and Pck) and shifts carbon flux towards Pck to enhance ATP conservation upon conversion of PEP to OAA. I demonstrate that the transcriptomic changes upon inactivation of GalP have implications not only for xylose to succinate conversion, but also for xylose catabolism in general. This further suggests that utilization of GalP in PTS^- strains for lignocellulose bioconversion may be a non-ideal method to enhance glucose utilization, and suggests that heterologous transporters without unknown regulatory mechanisms may be more ideal. However, this does not necessarily fix the issue of ABC catalyzed ATP utilization during xylose import. **In Chapter 3**, I attempt to provide a robust solution for this issue by engineering a sugar facilitator, Glf, to create an energy efficient xylose uptake system. I rationally created variants of a sugar facilitator, Glf, that are resistant to glucose inhibition to circumvent issues of substrate inhibition of uniporters in industrially relevant sugar mixtures. I further show that many of these mutant transporters maintain wildtype transport efficiency for glucose and xylose *in vivo*, but have shifted sugar preference to prefer xylose over glucose. The causative mutations for relieving glucose repression appear to be diverse, but suggest that there may be multiple ways to release glucose inhibition and that these mechanisms may be widely conserved with other MFS sugar porters. These transporters may serve as an engineering tool to enhance ATP conservation during glucose-xylose fermentation without sacrificing production metrics associated with glucose repression.

SMR transporters can function for furfural efflux to enhance furan aldehyde tolerance

Pretreatment of lignocellulose is known to generate multiple inhibitory compounds including furan aldehydes, aromatics, and organic acids (1, 7). The inhibitory effects of furfural are known to

be particularly detrimental to biocatalysts however due to the ability to potentiate the toxicity of other chemicals (10). Although multiple cytosolic enzymes have been engineered to enhance furfural tolerance by targeting toxicity mechanisms, no work yet has found a transporter to efflux this chemical from the cytoplasm (1). However, furfural is a relatively hydrophobic compound and MDR transporters are known to efflux a broad range of lipophilic compounds, suggesting these transporters may be prospective candidates (8, 11). **In Chapter 4**, I focused on screening a library of MDR transporters to discover potential furfural efflux proteins. From this work I found that overexpression two SMRs (*sugE, mdtJI*), and one lactate permease (*lldP*) enhance furfural tolerance under xylose fermentation conditions. I further demonstrated that MdtJI can be used to enhance xylose to ethanol conversion in the presence of furfural under relevant production conditions. Furthermore, I show that this enhanced tolerance appears to be due to an increase in the viable cell population during furfural stress, likely due to decreased intracellular stress. This decrease in intracellular stress is mediated by functional MdtJI transport activity and appears to be likely due to MdtJI catalyzed furfural efflux. This work ascribes a novel function to SMR transporters, and suggests that directed evolution of these transporters may be useful for enhancing tolerance to furan aldehydes or other inhibitors.

Promiscuous transport systems can be effectively identified using a cheminformatics approach

As stated previously, identification of transporter efflux systems is difficult using many currently available selection methods. Incorporation of these transporters into metabolic engineering regimes has been shown to enhance target product titer, productivity and yield, making advancement of these methods an important target for enhancing lignocellulose bioproduction (12). One particularly striking example has been the incorporation of efflux transporters into fungi to enhance L-malate production (13). *E. coli* has been engineered for L-malate production, however no known malate exporters have been discovered in this organism (14). To elucidate the native malate efflux system of *E. coli,* **in Chapter 5** I used a cheminformatics approach to identify transporters of chemicals similar to L-malate to investigate if these transporters had promiscuous

activity for the target molecule. Using this approach, I identified that citrate and tartrate are the two most similar known di-/tricarboxylic acids that are transported in *E. coli* to L-malate, and inactivation of these transporters decreases malate production in XZ658. This work identified that DcuA, CitT, and TtdT have a role in malate efflux, and demonstrated that malate efflux does not appear to be the current production bottleneck for XZ658. This makes sense, as work in my **Appendix A** demonstrated that the current bottleneck for production of L-malate appears to be relief of allosteric inhibition by NADH in citrate synthase, which results in a 90% increase in titer. However, this work demonstrated that using chemical similarity methods to prospect for target molecule transporters based on the concept of transporter promiscuity is a viable approach to rationally identify efflux systems.

Future directions

My thesis investigated multiple membrane transport associated challenges in converting lignocellulose to valuable products. The field of transporter engineering in general is still relatively nascent and a lot of fundamental transporter:substrate interactions have yet to be characterized. That being said, multiple intriguing questions have yet to be addressed in the future.

In Chapter 2, I demonstrate that GalP has unknown interactions that appear to affect transcriptional regulation of multiple gene clusters. However, the nature of these interactions remain unknown and may be indicative of an unknown functional domain in sugar porters from the MFS. Many of the upregulated gene clusters include operons that were upregulated by the presence of cAMP, such as the *ara, xyl, gal,* and *rbs* operons. Another protein CstA, a carbon starvation regulator upregulated by cAMP and CRP, is also upregulated ~8 fold in the *galP* inactivation mutant. Thus it may be interesting to investigate the intracellular cAMP levels of both KJ122 and AG055, as this may be a signaling molecule affected by *galP* inactivation. Similarly, it might be interesting to investigate if this mechanism is transport dependent or if it is independent of this function. The mutations found thus far that enhance xylose utilization are in either a glycine residue (G236) that is predicted to be located on the cytoplasmic side between two intracellular helical (ICH) domains or an alanine (A392) in approximately the middle of TM11. By disrupting the

conserved proton coupling aspartate residue between XylE and GalP (GalP D32), we could presumably decipher whether this phenomena is dependent on transport mediated by GalP.

In Chapter 3, I created Glf variants resistant to glucose that appear to have at least two separate mechanisms for relieving glucose inhibition: 1) by disrupting coordination of the 6-hydroxymethyl group of glucose and 2) by altering unknown interactions in TM12. One interesting future direction may be to combine mutations in the best performing mutants of each mechanism, namely mutant GK3 and GK8. This may aid in further reducing glucose inhibition of xylose and further enhance the utility of these variants as xylose uniporters. I also show that variant Glf proteins work for enhancing transport preference towards xylose import, but I do not show an effective demonstration of how increased ATP conservation from transport can benefit bioproduction. Although I attempted to express variant Glf proteins in AG055, this only resulted in a possible slight increase in yield and titer (Figure 3.9). This may be due to the presence of XylFGH still on the chromosome that may be capable of masking the beneficial effects. The growth of AG055 expressing Glf variant GK3 is also decreased compared to an empty vector or wildtype Glf (Figure 3.9). This may be indicative that catabolite repression is still partially present in this strain and *xylR** integration is needed to elucidate the functional benefit of these transporters. Additionally, it may be difficult to prove the efficacy of these transporters in a mono-culture since the presence of glucose should enhance the ATP supply of the biocatalyst. However, our lab has previously developed orthogonal sugar specialist strains that can be co-cultured to effectively ferment glucose-xylose mixtures. Utilization of these transporters in a xylose fermenting specialist may be able to improve the fermentation capabilities of these strains, as we have previously seen that a large excess of xylose specialists is needed to be able to compete for resources against the glucose specialist during co-sugar fermentation. This may be due in part to the energy burden that the xylose specialist faces in addition to inhibition of native xylose uptake systems (e.g. XylE) by glucose. Thus, inactivation of XylFGH and incorporation of a Glf variant into a xylose specialist may prove effective to increase co-culture fermentation robustness.

Although I focused on engineering Glf in Chapter 3 as an ideal xylose transporter, this is not the only potential route to create an energy efficient xylose import mechanism. Another interesting possibility would be to create a PTS transport system for xylose. Since the PTS couples dephosphorylation of PEP to concomitant uptake and phosphorylation of hexose sugars, this could potentially be exploited to catabolize pentose sugars while only utilizing one ATP-equivalent for pentose uptake. However, pentose sugars like xylose and arabinose often go through an isomerization step prior to phosphorylation (15). In order for a system like this to potentially exist, isomerization would need to happen post-phosphorylation of the pentose sugar to integrate the molecule into normal pentose metabolism. There are reports of arabinose-5-phosphate (16) and ribose-5-phosphate isomerases (17), although there is no evidence of these being used to create an alternative mechanism for utilizing xylose/arabinose through the Pentose Phosphate Pathway as described. If utilization of this pathway could be demonstrated, it would be interesting to see if PTS transporters which have been shown to have activity for pentose sugars (such as *gatC*) could be utilized for xylose/arabinose uptake and catabolism (18). Mutations in the glucose PTS transporter, *ptsG,* have also been shown to confer activity for pentose sugars, such as ribose, making these ideal templates to potentially further tune PTS transporter specificity to pentose transport (19).

In Chapter 4, I show that a transporter library derived from the ASKA collection of *E. coli* ORFs is effective in isolating furfural resistance transporters. However, the ASKA library is known to have an artificial sequence at both the N- and C-terminus that may potentially affect transporter activity. In my own work, I have noticed that this does occasionally affect transport activity. For instance, I attempted to complement glucose uptake (PTS^-) and alanine efflux (Δ*alaE*) deficient strains using constructs from within this library at one point, however complementation was unsuccessful. Sub-cloning the exact same sequence into a new plasmid while omitting the artificial N- and C-terminal sequences however successfully complemented this phenotype. This demonstrates that some transporters within this library may lead to the generation of false-positives, suggesting that this library is only sub-optimal. A more ideal library would omit these sequences, as well as expand into

other families of endogenous and heterologous MDR transporters. For instance, the Drug/Metabolite Transporter (DMT) superfamily includes the SMR transporters, but also 26 other families of transporters. Some of these include probable or verified drug exporters that look similar to SMRs but with a variable number of transmembrane spanning regions. This includes transporters such as the Bacterial/Archael Transporters (BAT) and other variations of the SMR family (SMR2, SMR3). Likewise, the library could benefit from the inclusion of other verified MDR transporters from other organisms, such as the SMR transporter Mmr from *Mycobacterium tuberculosis* or YkkCD from *Bacillus subtilis* (11). Additionally, this work could be expanded to attempt to characterize the residues in SMR transporters that control specificity. Ideally, using a small MDR transporter like a SMR would be easier to saturate with protein engineering techniques to potentially enhance and broaden the efflux activity of these transporters. These efforts would allow a more comprehensive screening of transporters for desired compounds and ensure that only the best template is used for further enzyme engineering efforts.

References

1. Nieves LM, Panyon LA, Wang X. 2015. Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. Front Bioeng Biotechnol 3:17.

2. Gonzalez JE, Long CP, Antoniewicz MR. 2017. Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by (13)C metabolic flux analysis. Metab Eng 39:9-18.

3. Utrilla J, Licona-Cassani C, Marcellin E, Gosset G, Nielsen LK, Martinez A. 2012. Engineering and adaptive evolution of *Escherichia coli* for D-lactate fermentation reveals GatC as a xylose transporter. Metab Eng 14:469-76.

4. Khunnonkwao P, Jantama SS, Kanchanatawee S, Jantama K. 2018. Re-engineering *Escherichia coli* KJ122 to enhance the utilization of xylose and xylose/glucose mixture for efficient succinate production in mineral salt medium. Appl Microbiol Biotechnol 102:127-141.

5. Wang X, Yomano LP, Lee JY, York SW, Zheng H, Mullinnix MT, Shanmugam KT, Ingram LO. 2013. Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 110:4021-6.

6. Hasona A, Kim Y, Healy FG, Ingram LO, Shanmugam KT. 2004. Pyruvate formate lyase and acetate kinase are essential for anaerobic growth of *Escherichia coli* on xylose. J Bacteriol 186:7593-600.

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

8. Higgins CF. 2007. Multiple molecular mechanisms for multidrug resistance transporters. Nature 446:749-57.

9. Sawisit A, Jantama K, Zheng H, Yomano LP, York SW, Shanmugam KT, Ingram LO. 2015. Mutation in galP improved fermentation of mixed sugars to succinate using engineered *Escherichia coli* AS1600a and AM1 mineral salts medium. Bioresour Technol 193:433-41.

10. Zaldivar J, Martinez A, Ingram LO. 1999. Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng 65:24-33.

11. Bay DC, Rommens KL, Turner RJ. 2008. Small multidrug resistance proteins: a multidrug transporter family that continues to grow. Biochim Biophys Acta 1778:1814-38.

12. Jones CM, Hernandez Lozada NJ, Pfleger BF. 2015. Efflux systems in bacteria and their metabolic engineering applications. Appl Microbiol Biotechnol 99:9381-93.

13. Brown SH, Bashkirova L, Berka R, Chandler T, Doty T, McCall K, McCulloch M, McFarland S, Thompson S, Yaver D, Berry A. 2013. Metabolic engineering of Aspergillus oryzae NRRL 3488 for increased production of L-malic acid. Appl Microbiol Biotechnol 97:8903-12.

14. Zhang X, Wang X, Shanmugam KT, Ingram LO. 2011. L-malate production by metabolically engineered *Escherichia coli*. Appl Environ Microbiol 77:427-34.

15. Flores A, Kurgan G, Wang X. 2016. Engineering bacterial sugar catabolism and tolerance toward lignocellulose conversion. *In* Gosset G (ed), Engineering of Microorganisms for the

Production of Chemicals and Biofuels from Renewable Resources doi:10.1007/978-3-319-51729-2. Springer Nature.

16. Meredith T, Woodward R. 2005. Identification of GutQ from *Escherichia coli* as a D-ararbinose-5-phosphate isomerase. Journal of Bacteriology. 187:6936-42.

17. Roos A, Mariano S, Kowalinski E, Salmon L, Mowbray S. 2008. D-ribose-5-phosphate isomerase B from Escherichia coli is also a functional D-allose-6-phosphate isomerase, while the Mycobacterium tuberculosis enzyme is not. J Mol Biol. 382:667-79.

18. Utrilla J, Licona-Cassani C, Marcellin E, Gosset G, Nielsen LK, Martinez A. 2012. Engineering and adaptive evolution of Escherichia coli for D-lactate fermentation reveals GatC as a xylose transporter. Metab Eng 14:469-76.

19. Oh H, Park Y, Park C. 1999. A mutated PtsG, the glucose transporter, allows uptake of D-ribose. J Biol Chem. 274:14006-14011

APPENDIX A

RELEASING ALLOSTERIC REGULATION OF CITRATE SYNTHASE ENHANCES L-MALATE PRODUCTION IN *ESCHERICHIA COLI*

APPENDIX

RELEASING ALLOSTERIC REGULATION OF CITRATE SYNTHASE ENHANCES L-MALATE PRODUCTION IN *ESCHERICHIA COLI*

Abstract

Malic acid is a C4-dicarboxylic acid currently produced from petroleum with applications in food industries and it has the potential to be produced as a value-added platform chemical from biomass instead of petroleum. *Escherichia coli* strain XZ658 was previously developed for L-malic acid production using the reductive branch of the tricarboxylic acid (TCA) cycle. Here, we identified the allosteric regulation of citrate synthase as the main metabolic constraint for malate production in *E. coli* and deregulation of this allosteric control significantly increased product titer, yield and productivity. By inactivating lactate export systems to further reduce side product lactate, we have achieved unprecedented yield (1.2 g malate g^{-1} glucose) for malate production, further demonstrating the carbon-fixation capacity of the anaplerotic enzyme for fermentative production.

Importance

Diverse industrial chemicals can be produced from renewable feedstocks by metabolically engineered microbes. However, low production metrics of engineered microbes often limit commercialization. Allosteric regulation of key enzymes is a primary mechanism to control cellular metabolic processes, thereby representing potential regulatory constraints for fermentative microbial production, especially for nonnative fermentation products. Our results suggest releasing allosteric regulation of citrate synthase in *E. coli* as an important approach to optimize fermentative production of chemicals from the reductive branch of the TCA cycle.

Introduction

Microbial conversion of renewable feedstocks, such as lignocellulosic biomass, into fuels and chemicals through fermentation-based manufacturing processes is a desirable alternative to petrochemicals (1-3). The top 12 value-added chemicals derived from biomass with the potential to replace petrochemicals were identified by the Department of Energy (4). 1, 4-C4 dicarboxylates such as succinate, fumarate, L-malate and L-aspartate are among these identified top commodity chemicals with a multibillion dollar market value (4-6). Currently, the major routes to produce these

compounds are still based on the petrochemical platform (5, 6). Although there are many examples in scientific literature regarding metabolic engineering of microbes to produce many valuable renewable chemicals including 1, 4-C4 dicarboxylates, the low production performance (titer, yield and productivity) remains a significant barrier for large-scale production in a cost-effective manner (2, 7, 8).

1, 4-C4 dicarboxylates such as succinate and L-malate are central metabolites without significant cellular toxicity even at high concentrations. Microbial production of these compounds has a high theoretical yield due to the carbon-fixation step of the anaplerotic enzyme (Ppc or Pck) during their biosynthesis as illustrated in Fig. 1. For example, the maximal theoretical yields for succinate and malate are 1.12 and 1.49 g per g glucose, respectively (9-11). A wide range of microorganisms including *Aspergillus flavus* (12), *Aspergillus oryzae* (13, 14), *Zygosaccharomyces rouxii* (15), *Saccharomyces cerevisiae* (16-18), *Aureobasidium pullulans* (19), *Rhizopus delemar* (20, 21), *Thermobifida fusca* (22), and *E. coli* (11, 23) were developed to produce malate using aerobic or microaerobic processes. Although high titers (100 g liter^{-1} or higher) have been achieved after production process optimization in fungal hosts such as *A. oryzae* (14) and *A. pullulans* (19), the yields remain much lower than the theoretical maximum, with the highest reported value being only 69% of theoretical maximum (14). Thus the carbon fixation capacity of the anaplerotic enzyme is not fully exploited yet in these microorganisms.

Homo-malate fermentation is a redox-balanced process with net ATP generation if phosphoenolpyruvate (PEP) carboxykinase (Pck) is used instead of native PEP carboxylase (Ppc) (Figure 7.1) (11, 24). Production of each malate molecule requires the fixation of one CO_2/HCO_3^- meanwhile preserving all glucose carbons, thereby yielding up to 2 moles of malate per mole of glucose (Figure 7.1). We successfully engineered *E. coli* for malate production using a precursor strain with increased PCK activity and multiple deliberate gene deletions to funnel metabolic flux into malate (11). However, the slow growth, low yield of cell mass and low production metrics suggest that intrinsic metabolic constraints limit malate production under oxygen-limiting fermentative conditions.

Here, we hypothesize that accumulated intracellular metabolites and cofactors such as PEP and NADH allosterically inhibit important central metabolic pathways, thereby limiting cell growth and fermentative malate production. We investigated the impact of allosteric regulation associated with malate production, focusing on three main allosteric regulatory sites involved in metabolic pathways of C4 dicarboxylates production: 6-phosphofructokinase (Pfk) (25), lipoamide dehydrogenase (Lpd) (26, 27), and citrate synthase (GltA) (28). Pfk catalyzes the irreversible phosphorylation of fructose 6-phosphate at an early step in glycolysis (29). Lpd is the main component of the pyruvate dehydrogenase complex that catalyzes the decarboxylation of pyruvate to form acetyl-CoA (30). GltA catalyzes the formation of citrate from oxaloacetate and acetyl-CoA in the TCA cycle, which is the first committed step for production of α-ketoglutarate, a precursor important for amino acid biosynthesis (31, 32). In *E. coli*, PEP is a negative allosteric regulator for Pfk (25) and high intracellular levels of NADH strongly inhibits the activities of Lpd (26, 27) and GltA (28). This allosteric inhibition can be mitigated by replacing these native enzymes with heterologous isozymes or mutants that are resistant to inhibition (26, 33-35).

In this study, we identified GltA as a major allosteric regulation site for malate production and mitigation of this allosteric inhibition in a previously engineered malate producer increased malate titer and yield by 90% and 40%, respectively. Further deletions of lactate permeases decreased side product accumulation and led to a yield at 1.2 g g^{-1} (80% of the theoretical maximum).

Results

Overexpression of the mutants resistant to allosteric regulation in a malate-producing *E. coli* strain

The genes encoding *pfkA*, *lpd* and *gltA* variants resistant to allosteric inhibition were cloned into pTrc99A with their native ribosomal binding sites and terminator regions (Table 7.1). The missense mutations of T125S in PfkA and E354K in Lpd were used to release the allosteric regulation of 6-phosphofructokinase and pyruvate dehydrogenase in a previously engineered malate producer XZ658 (11). There are two types of bacterial citrate synthases with distinct subunit arrangements and allosteric regulatory properties: the type I enzymes in Gram-positive bacteria and archaea; type II enzymes in Gram-negative bacteria (31, 32). NADH strongly inhibits the type

II citrate synthase in *E. coli* but has no effect on type I enzymes (28, 31, 33). To release allosteric inhibition of citrate synthase, two different types of variants that are insensitive to NADH inhibition were used: i) the mutant of *E. coli* citrate synthase (GltA R163L) (33) and ii) citrate synthase genes *citZ* from *Bacillus subtilis* and *Bacillus licheniformis* (35). Two concentrations of IPTG (0 and 0.1 mM) were used to test the effect of induced expression of target genes on malate production. Overexpression of *pfkA* and *lpd* mutants had essentially no impact on malate production (Table 7.2). In contrast, induced expression of either the *gltA* mutant or *citZ* from *Bacillus* strains increased malate titer and yield ranging from 40 to 55% and 45 to 60%, respectively. Even without induction, the leaky expression of these variant genes led to increased malate titers and yields (Table 7.2). This increase of malate production is correlated to the enhanced biomass yield and sugar consumption (Table 7.2). Titers of side products, especially the main side product D-lactate, are similar to the control strain with empty vector and more sugar conversion into malate led to an increase in malate yield (~ 1.0 g malate g^{-1} glucose used) (Table 2).

Chromosomal integration of citrate synthase variants to enhance malate production

Plasmid-based systems have several disadvantages in large-scale microbial production, such as genetic instability, performance inconsistency, metabolic burden and the requirement of antibiotics and inducers (40-42). We integrated *citZ* from both *Bacillus* strains into the *frdBC* locus on the chromosome since the *frdBC* operon was already deleted in the malate producing strain XZ658 (11) and the upstream promoter is active under anaerobic fermentation conditions (43, 44). GltA R163L was not used for chromosomal integration due to potential homologous recombination to restore the mutation and cause undefined genetic changes. Integrations of both *Bacillus citZ* increased cell growth, biomass yield and malate production metrics (Table 7.3). The integration of *B. subtilis citZ* yielded the strain JB02 that produced 42 g $liter^{-1}$ with a product yield of 1.1 g g^{-1} glucose by a simple batch fermentation. Compared to the precursor strain XZ658, this genetic modification increased the titer, yield, and productivity by 90%, 40%, and 90%, respectively (Table 7.3). In addition, the replacement of native *gltA* with *B. licheniformis citZ* in XZ658 also showed similar positive effect, suggesting that type II citrate synthase is able to functionally replace native type I enzyme without noticeable growth defectiveness under fermentation conditions (Table 7.3).

Allosteric regulation of citrate synthase in *E. coli* was found to limit cell growth in an ethanol producer (35). It was indicated that the growth inhibition is due to the starvation of α-ketoglutarate, an important precursor for amino acid biosynthesis (35). We similarly hypothesized that NADH accumulation caused by inefficient malate production inhibited citrate synthase, and thereby limited the amount of citrate and α-ketoglutarate that stunts cell growth. To directly test this hypothesis, 30 mM citrate and α-ketoglutarate were supplemented in the medium for XZ658 fermentation, respectively. Both chemicals enhanced cell growth, leading to ~50% increase of malate titer and glucose consumption (Table 7.3). Supplementation of α-ketoglutarate did not increase malate yield because of the increased amount of side products. Addition of citrate had no significant impact for JB01 (XZ658 *frdBC*::*citZ*$_{B.\ licheniformis}$) fermentation, suggesting that *citZ* integration and addition of citrate may share a common working mechanism to improve malate production (Table 7.3).

Inactivation of lactate permeases to decrease lactate accumulation

Although lactate dehydrogenase gene, *ldhA*, was deleted in XZ658, uncharacterized reduction of pyruvate led to lactate accumulation that uses ~30% glucose (11). The genes encoding the enzymes involved in lactate metabolism including Dld (D-lactate dehydrogenase)(45), LldD (encoding L-lactate dehydrogenase)(46), YbiC (putative alpha-keto acid dehydrogenase) (47), YkgEFG (predicted lactate-related catabolism) (48) were deleted separately, but none of these genetic modifications essentially decreased lactate accumulation relative to XZ658 (data not shown).

To direct more carbon flux towards malate production, we inactivated lactate export systems in XZ658 as an alternative approach to prevent lactate accumulation. There are three known transporters potentially involved in lactate export: LldP, GlcA and FocA (49, 50). The *focA* gene was already deleted in the malate producing strain XZ658 (10, 11). Chromosomal inactivation of both *lldp* and *glcA* in JB02 yielded the strain AS55. Interestingly, inactivation of both known lactate transporters did not completely block lactate accumulation, suggesting the presence of an unknown lactate export system in *E. coli*. Lactate titer was reduced by 35% and the malate yield reached 1.2 g g^{-1} (1.6 mol mol^{-1}) in AS55. This yield is higher than previously reported results in *A. oryzae* (14)

and other fungal strains (12, 17-23). This high yield shows the potential of the anaplerotic enzyme to fix carbon and increase total carbon conversion efficiency.

Discussion

The CO_2 fixation step of the anaplerotic enzyme prior to the reductive branch of the TCA cycle (Figure 7.1) has been proposed to have high potential for carbon assimilation in terms of pathway kinetics and overall low energy cost (51). L-malate production takes advantage of this carbon-fixing step to achieve a higher yield (1.49 g g^{-1}; 2.0 mol mol^{-1}) than many other fermentation products, such as ethanol (0.51 g g^{-1}) and lactate (1.0 g g^{-1}). However, metabolic constraints limit malate production and high yields have not been achieved in the previous reports (11, 13, 14, 17-19, 23, 52). In this study, we discovered that the allosteric regulation of citrate synthase inhibited malate fermentative production and cell growth in *E. coli* and introduction of citrate synthase variants insensitive to NADH inhibition significantly improved malate production.

There are four putative metabolic routes to produce malate: 1) oxaloacetate reduction catalyzed by malate dehydrogenase in the reductive branch of TCA cycle; 2) pyruvate carboxylation by malic enzymes; 3) the oxidative branch of TCA cycle; and 4) the glyoxylate shunt with co-production of succinate (Figure 7.1). The first route has the best kinetics and thermodynamic advantages (51, 53). The second route using malic enzyme to fix CO_2 is limited by unfavorable thermodynamics and kinetics (54-56). Route 3 and 4 are not redox-balanced and produce multiple side-products, thereby leading to low product yields. In *E. coli*, the native PEP carboxylation step employs PEP carboxylase in a process that loses the energy contained in PEP with the release of inorganic phosphate, which limits net ATP generation during anaerobic fermentation (11). Some bacteria living in bovine rumen such as *Actinobacillus succinogenes* (57-59) and adapted *E. coli* for succinate fermentative production (24) use an alternative carboxylase, Pck. Use of Pck as the anaplerotic enzyme conserves energy in PEP by generating ATP during the PEP carboxlyation step (Fig. 1). This Pck-based malate production should theoretically maximize carbon yield and achieve a homo-malate fermentative production. Fungal malate-producing strains such as *S. cerevisiae* (17), *A. oryzae* (14) and *A. pullulans* (19) use pyruvate carboxylase to convert pyruvate and bicarbonate into oxaloacetate with the cost of ATP. Therefore, there is no net ATP generated

for the glucose to malate conversion in these fungal strains. Moreover, the production procedures using these fungal strains require oxygen to provide energy for cellular growth, thereby further decreasing malate yield due to carbons lost in side products and biomass. In addition, $CaCO_3$ was used to maintain pH and provide a CO_2 source for fungal malate production (13, 14, 16, 17, 19), which will lead to gypsum accumulation during downstream processes, a potential disposal challenge for organic acid fermentative production (60). The engineered malate-producing *E. coli* strain directly adds carbons from bicarbonate into the malate carbon skeleton without the requirement for $CaCO_3$. If the issues associated with allosteric regulation, enzyme kinetic properties, side-product accumulation and product export are further addressed, this Pck-based malate production pathway has the potential to achieve a yield close to theoretical maximum under oxygen-limiting conditions.

Allosteric regulation of key enzymes is one of the primary mechanisms that control complex cascades of cellular metabolic processes. Identification and deregulation of these biochemical controls have been found to be effective in improving microbial production, especially for nonnative fermentation products such as amino acids and aromatic compounds (61-63). Releasing allosteric regulation of citrate synthase in XZ658 not only increased cell growth but also enhanced malate production metrics on a per cell basis (Table 3). The increased malate production is not from the enhanced glyoxylate shunt pathway because succinate production in the modified strains (JB01, JB02 and JB03) is still negligible (< 1g $liter^{-1}$) compared to the increased malate production (~15-20 g $liter^{-1}$). The released allosteric regulation presumably increases intracellular levels of citrate and α-ketoglutarate that are important for amino acid biosynthesis and other physiological processes. This hypothesis was further supported by the chemical supplementation experiments that exogenous citrate and α-ketoglutarate also improved fermentative cell growth and malate production. A similar observation was made in an ethanologenic *E. coli* KO11 (35). Releasing allosteric regulation of citrate synthase in KO11 improved fermentative cell growth and ethanol production when cells used xylose as the sole carbon source (35). During sugar fermentation to generate different products NADH molecules produced during glycolysis need to be reoxidized by fermentative pathways to achieve redox balance. Inefficient NADH oxidation steps at fermentation

pathways, for example, OAA reduction by malate dehydrogenase (Fig. 1), will lead to the accumulation of NADH, thereby representing a potential bottleneck to different fermentative producing strains.

D-lactate is the main side product for our malate-producing strain. Our previous results indicated that lactate is likely derived from pyruvate through an unknown metabolic route (11). Attempts to inactivate putative lactate metabolic enzymes including Dld, LldD, YbiC and YkgEFG did not decrease lactate accumulation. To mitigate the lactate accumulation by targeting its export, two known lactate transporters (GlcA and LldP) were inactivated and lactate concentration was decreased by only 35% in the resulting strain AS55. Surprisingly, the majority of lactate is still able to cross cellular membranes without these two lactate transporters, suggesting the presence of an unknown lactate export system in *E. coli*. If carbon flux from this lactate pathway (~20% sugars) is re-directed to produce malate, the malate yield can be further increased close to ~1.4 g/g (>90% of theoretical maximum). However, our incomplete understanding of microbial product export limits the effectiveness of this strategy to minimize side-product accumulation. The strain AS55 has a lower growth rate and biomass yield, thereby leading to a lower malate titer compared to its precursor strain JB02 (Table 7.3). It is plausible that lactate production facilitates cell growth by oxidizing abundant reducing power generated by glycolysis to maintain redox balance. Further laboratory adaptation may be helpful to improve cell growth and malate titer.

Materials and Methods

Bacterial strains and plasmids

Strains and plasmids used in this study are listed in Table 7.1. XZ658 was kindly provided by Dr. Lonnie Ingram's group at University of Florida (11). Other strains were constructed using either one-step inactivation (36) or two-step integration as previously described (37, 38). To express heterologous or mutant genes resistant to allosteric regulation, the ribosomal binding sites, the open reading frames and terminator regions were cloned into pTrc99A under the control of the isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible *trc* promoter. The construction details such as used restriction sites are described in Table 7.1.

Media and growth conditions

During strain construction, cultures were grown aerobically at 30°C, 37°C, or 39°C in Luria-Bertani broth as needed. NBS (New Brunswick Scientific) mineral salts medium supplemented with 50 g liter^{-1} glucose, 10 mM acetate and 100 mM potassium bicarbonate was used in batch fermentation as previously described (11). Ampicillin (50 mg liter^{-1}), kanamycin (50 mg liter^{-1}), or chloramphenicol (40 mg liter^{-1}) were added as needed.

Genetic methods

Red recombinase technology was employed to perform seamless chromosomal deletion and gene replacement as previously described (37). The DNA fragments used for the first-step integration were obtained by PCR to amplify the *cat-sacB* cassette from plasmid pXW1 and primers provided 50 bp or longer homologous sequences to facilitate homologous recombination. The DNA fragments used for the second-step integration containing homologous regions (~500 bp flanking sequence on each side) were assembled by fusion PCR (39). Point mutations in target genes were also introduced using fusion PCR (39) with the relevant primers listed in Table 7.1. During strain construction, 5% (w/v) arabinose and 10% (w/v) sucrose were used for induction of λ-Red recombinase synthesis and for *sacB* counterselection, respectively. Positive clones from genetic manipulation were verified by colony PCR and Sanger sequencing of the PCR-amplified target regions.

Fermentation

All batch fermentations were performed using NBS (New Brunswick Scientific) mineral salts medium supplemented with 50 g liter^{-1} glucose, 100 mM potassium bicarbonate and 10 mM sodium acetate in small fermentation vessels with 300 ml as the working volume as previously described (11). Pre-inoculum for fermentations was started by transferring fresh colonies from NBS 2% glucose (w/v) plates into a shake flask (100 ml NBS medium, 2% glucose). After 16 h (37ºC, 120 rpm), this culture was diluted into a fermentation vessel containing 300 ml NBS medium (5% glucose, 100 mM potassium bicarbonate, 10 mM acetate) to provide an inoculum of 0.044 g cell dry wt (CDW) liter^{-1} (an optical density value of 0.05) and batch fermentations were performed for 144 h. Batch fermentations of strains with plasmids were operated in a similar manner with the supplementation of 50 mg liter^{-1} ampicillin and 0.1 mM IPTG to induce the expression of tested

genes. All fermentations were maintained at pH 7.0 by the automatic addition of base solution (2.4 M potassium carbonate and 1.2 M potassium hydroxide) as previously described (11).

Analyses

Glucose and organic acids in fermentation broth were measured by high-performance liquid chromatography (HPLC) using an Aminex® HPX-87H column (Bio-Rad) and 4 mM sulfuric acid as the mobile phase (11). Cell dry weight was derived from the measured optical density at 550 nm. Experimental data represent an average of three or more measurements with standard deviations.

Figures

Table 7.1: Strains, plasmids and primers used in this work

	Relevant characteristics	Reference
Strains		
KJ073	*E.coli* ATCC 8739 Δ*ldhA,* Δ*ackA,* Δ*adhE*, Δ*focA-pflB* Δ*mgsA*, Δ*poxB;* adapted and engineered for succinate production	(10)
XZ658	KJ073 Δ*frdBC* Δ*sfcA* Δ*maeB* Δ*fumB* Δ*fumAC*; engineered for L-malate production	(11)
JB01	XZ658 *frdBC*::*citZ*$_{B.\ licheniformis}$	This study
JB02	XZ658 frdBC::citZ$_{B.\ subtilis}$	This study
JB03	XZ658 *gltA*:: citZ$_{B.\ licheniformis}$	This study
JB04	XZ658 *glcA*::citZ$_{B.\ subtilis}$	This study
GX01	XZ658 *dld*::FRT	This study
GX02	XZ658 *lldD*::FRT	This study
GX03	XZ658 *ybiC*::FRT	This study
GX04	XZ658 *ykgEFG*::FRT	This study
AS55	JB02 *glcA*::FRT *lldP*::FRT	This study
Plasmids		
pTrc99A	P*trc bla oriR rrnB lacI*^{q}	(64)
pKD46	*bla*, *γ β exo* (Red recombinase)	(36)
pKD4	*bla,* FRT-kan-FRT	(36)
pPfkA T125S	*E. coli pfkA* ORF with T125S (A to T nucleotide change) point mutation and its 50 bp upstream sequence cloned into between KpnI and XbaI sites in pTrc99A	This study
pLpd E354K	*E. coli lpd* ORF with E354K (G to A nucleotide change) point mutation and its 50 bp upstream sequence cloned into between EcoRI and XbaI sites in pTrc99A	This study
pGltA R163L	*E. coli gltA* ORF with R163L (G to T nucleotide change) point mutation and its 50 bp upstream sequence cloned into between EcoRI and XbaI sites in pTrc99A	This study

pCitZ$_{B.\ licheniformis}$	*B. licheniformis citZ* ORF and its 50 bp upstream sequence cloned into between BamHI and SalI in pTrc99A	This study
PcitZ$_{B.\ subtilis}$	*B. subtilis citZ* ORF and its 50 bp upstream sequence into between BamHI and SalI in pTrc99A	This study
PXW1	The *cat-sacB* cassette with the *sacB* native terminator cloned into a modified vector pLOI4162 (37)	(38)
Primers		
pfkA T125A cloning	5' CGACGGTACCCAAGAAGACTTCCGGCAACAG 5' ATGGTCTAGAGTTGAGGGATTAAAAAGGCGG 5' CCGTGCATCGGTCTGCCGGGCTCTATCGACAA CGACATC 5' CCTTTGATGTCGTTGTCGATAGAGCCCGGCAGA CCGATG	This study
Lpd E354K cloning	5' CGTACGAATTCGAAAGACGACGGGTATGACCG 5' GTGTTCTAGACGACTGGAAAGGTAAATTACAGA CG 5' ATCGCCTATACCAAACCAGAAGTTG 5' CAACTTCTGGTTTGGTATAGGCGAT	This study
gltA R163L cloning	5' CGTACGAATTCGTTCCGGCAGTCTTACGC 5' GTGTTCTAGAGTTCAGCCATATAAAAAGAACCC GC 5' GCCGCGTTCCTCCTGCTGTCGA 5' TCGACAGCAGGAGGAACGCGGC	This study
citZ$_{B.\ subtilis}$ cloning	5' AGAAGTGGGATCCTGGGGGAGAGAAATACTTG C 5' TTAATGGGTCGACTCGGGTTGTTTGGTACGTTT	This study
citZ$_{B.\ licheniformis}$ cloning	5' GAAGTGGGATCCCAACATGCGTTGTTTTTGCTC C 5' TTAATGGGTCGACACATTTGTTTCCCCCGTTTG	This study

frdBC::cat-sacB integration	5'AGCGGATGCAGCCGATAAGGCGGAAGCAGCC AATAAGAAGGAGAAGGCGATCGAGTGTGACGG AAGATCA 5'ATACCGGTTCGTCAGAACGCTTTGGATTTGGA TTGATCATCTCAGGCTCCTTAGCCATTTGCCTG CTTTT 5' TATCTACCGTACGCCGGAAC (used for sequencing) 5' AGCAAATGTGGAGCAAGAGG (used for sequencing)	This Study
frdBC::citZ$_{B. licheniformis}$ *and frdBC::citZ*$_{B. subtilis}$ *integrations*	5'AGCGGATGCAGCCGATAAGGCGGAAGCAGCC AATAAGAAGGAGAAGGCGAATGACAGCGACAC GCGGTC 5'ATACCGGTTCGTCAGAACGCTTTGGATTTGGA TTGATCATCTCAGGCTCCTTAGGCTCTTTCTTCA ATCGGAACGAA	This study
gltA::cat-sacB	5'AAATTTAAGTTCCGGCAGTCTTACGCAATAAG GCGCTAAGGAGACCTTAATCGAGTGTGACGGA AGATCA 5'CCCGCCATATGAACGGCGGGTTAAAATATTTA CAACTTAGCAATCAACCATTAGCCATTTGCCTG CTTTT	This study
gltA::citZ$_{B. licheniformis}$ *and gltA::citZ*$_{B. subtilis}$ *integrations*	5'AAATTTAAGTTCCGGCAGTCTTACGCAATAAG GCGCTAAGGAGACCTTAAATGACAGCGACACG CGGTC 5'CCGCTCTATTAAAGGCGGGTCCGGAAAGTAA ACGGCTTAGCAATCAATCATTAGGCTCTTTCTTC AATCGGAACGAA	This study
*dld::*FRT	5'AGGCGCTATTCTAGTTTGTGATATTTTTTCGCC ACCACAAGGAGTGGAAAGTGTAGGCTGGAGCT GCTTC 5'ACAGAACGATTAAGTGAATTCGGATGGCGATA CTCTGCCATCCGTAATTTCATATGAATATCCTCC TTAGT	This study
*lldD::*FRT	5'ATTACCCGCCTGCCCGGTGAGCATAATGAGC ATTCGAGGGAGAAAAACGCGTGTAGGCTGGAG CTGCTTC 5'CAAGAGACAGCTCCCTGAGGGAGAGGGTTAG GGGGAGGGGGCGCAAACGACATATGAATATCC TCCTTAGT	This study
*ybiC::*FRT	5'GCGTGCAGCATAAAAATACAACAAACACATAA CATAAACAGGAGTTAACCGTGTAGGCTGGAGCT GCTTC	This study

	5’TGATCCTATGTTACCCTTGTTTGCCCGTCCGC CACTGGACGGGCTTTTTTCATATGAATATCCTC CTTAGT	
ykgEFG::FRT	5’GGGGGGGAAACATCGGGCATAAATGGGCATG AAGTAATGGAGTATTAGTTGTGTAGGCTGGAGC TGCTTC 5’TTCAACCTCGTATGCCTTCTTCAGGTTTATGTC CAGACTTCATATCTCTCCATATGAATATCCTCCT TAGT	This study
glcA::FRT	5’TACAACCTTAAGTATATCAAGCATATAAAGATA ATAAGAGACTGAACAATGTGTAGGCTGGAGCTG CTTC 5’ATAATGTAAATGAAGCCGGATGATATTAACGA TCATCCGGCATTATTGATCATATGAATATCCTCC TTAGT	This study
lldP::FRT	5’GTGTGACAAGGAGATGAGCAACAGACTCATTA CACGATGTGCGTGGACTCCAGGAGACCTGCAG TGTAGGCTGGAGCTGCTTC 5’CCGCACACGATCGGCAACCTCGTCTGACAGG CGTCTGGGTAAAACAATCACATATGAATATCCT CCTTAGT	This study

Table 7.2. Effects of different plasmids on malate production in XZ658

Plasmids	Maximum Cell mass (g liter^{-1})	Glucose used (g)	Malate yield (g g^{-1}) [c]	Fermentation products (g liter^{-1}) [d]				
				Mal	Lac	Pyr	Ace	For
pTrc99A[a]	0.80 ± 0.10	28 ± 4	0.72	20 ± 3	8.4 ± 1	3.3 ± 0.7	0.86 ± 0.2	1.5 ± 0.1
pTrc99A[b]	0.75 ± 0.09	29 ± 1	0.69	20 ± 2	8.5 ± 0.6	2.0 ± 0.1	0.51 ± 0.2	1.4 ± 0.3
pPfkA T125S[a]	0.84 ± 0.09	29 ± 2	0.76	22 ± 2	10 ± 2	2.2 ± 0.7	0.99 ± 0.1	1.5 ± 0.1
pPfkA T125S[b]	0.70 ± 0.2	31 ± 4	0.74	23 ± 2	9 ± 2	2.0 ± 0.1	0.81 ± 0.1	1.2 ± 0.1
pLpd E354K[a]	0.70 ± 0.2	24 ± 1	0.75	18 ± 2	8.2 ± 0.7	1.8 ± 0.4	0.92 ± 0.2	1.4 ± 0.2
pLpd E354K[b]	0.80 ± 0.2	25 ± 3	0.8	20 ± 2	7 ± 2	2.2 ± 0.1	1.1 ± 0.6	1.4 ± 0.7
pGltA R163L[a]	1.1 ± 0.1	35 ± 3	0.74	26 ± 4	10 ± 3	3.6 ± 0.6	1.2 ± 0.4	2.3 ± 0.2
pGltA R163L[b]	0.88 ± 0.1	29 ± 2	1.0	29 ± 3	7.2 ± 0.4	3 ± 1	1.0 ± 0.4	2.1 ± 0.3
pCitZ$_{B.\ subtilis}$[a]	0.70 ± 0.09	25 ± 2	1.0	26 ± 1	6.1 ± 0.9	2.5 ± 0.1	0.4 ± 0.1	1.4 ± 0.1
pCitZ$_{B.\ subtilis}$[b]	0.92 ± 0.09	27 ± 4	1.0	28 ± 2	8.1 ± 1	2.9 ± 0.6	0.5 ± 0.1	1.4 ± 0.1
PcitZ$_{B.\ licheniformis}$[a]	0.97 ± 0.09	26 ± 5	1.0	26 ± 1	8.6 ± 0.2	3.4 ± 0.8	0.6 ± 0.1	1.6 ± 0.1
PcitZ$_{B.\ licheniformis}$[b]	0.90 ± 0.09	28 ± 3	1.1	31 ± 3	8.7 ± 0.5	3.7 ± 0.4	0.6 ± 0.1	1.4 ± 0.1

[a] Fermentations (n ≥3) without IPTG as controls were carried out in NBS mineral salts medium with 50 g liter^{-1} glucose and 100 mM potassium bicarbonate for 144 hours (37°C, pH 7.0).

[b] Fermentations (n ≥3) with 0.1 mM IPTG added to induce the expression of candidate genes.

[c] Yield was calculated as g malate produced per g glucose consumed.

[d] All data at 144 hours were shown here and SD were included. Abbreviations: Mal, malate; Lac, lactate; Pyr, pyruvate; Ace, acetate; For, formate.

Table 7.3. Effects of chromosomal integrations and chemical supplementations on malate production

Strains	Genotypes or added chemicals	Glucose used (g)	Maximum cell mass (g liter^{-1})	Malate yield (g g^{-1})	Fermentation products (g liter^{-1}) [a, b]				
					Mal	Lac	Pyr	Ace	For
XZ658	Precursor strain	28 ± 1	0.84 ± 0.2	0.78	22 ± 1	7.8 ± 0.5	3.0 ± 0.8	0.5 ± 0.1	1.5 ± 0.3
JB01	XZ658 *frdBC*::citZ$_{B. licheniformis}$	38 ± 4	0.97 ± 0.04	0.95	36 ± 2	8.9 ± 1	3.2 ± 0.9	0.9 ± 0.4	3.0 ± 0.8
JB02	XZ658 *frdBC*::CitZ$_{B. subtilis}$	37 ± 2	0.95 ± 0.04	1.1	42 ± 2	11 ± 1	2.3 ± 0.2	0.5 ± 0.1	1.8 ± 0.2
JB03	XZ658 *gltA*::citZ$_{B. licheniformis}$	35 ± 1	1.2 ± 0.04	1.1	37 ± 3	11 ± 1	2.3 ± 0.2	0.5 ± 0.1	1.8 ± 0.2
AS55	JB02 Δ*glcA* Δ*lldP*	29 ± 1	0.92 ± 0.04	1.2	34 ± 1	5.1 ± 0.5	1.1 ± 0.2	0.9 ± 0.1	2.0 ± 0.3
XZ658[c]	citrate	32 ± 5	1.0 ± 0.09	1	31 ± 2	10 ± 1	3.8 ± 0.7	1.5 ± 0.3	1.8 ± 0.6
XZ658 [c]	α-ketoglutarate	45 ± 1	1.4 ± 0.2	0.73	33 ± 2	12 ± 1	4.8 ± 0.7	0.6 ± 0.3	2.8 ± 0.2
JB01[c]	citrate	40 ± 3	1.1 ± 0.2	0.95	38 ± 1	11 ± 1	4.4 ± 0.4	1.3 ± 0.2	1.4 ± 0.4

[a] Yield was calculated as g malate produced per g glucose consumed.

[b] Fermentations were carried out in a 500 ml fleaker with 300 ml NBS mineral salts medium with 5% glucose, 10 mM acetate, and 100 mM potassium bicarbonate (37°C, pH 7.0, 150 rpm). Acetate was added to improve cell growth. Abbreviations: Mal, malate; Fum, fumarate; Suc, succinate; Pyr, pyruvate; Lac, D-lactate; Ace, acetate.

[c] Fermentation supplemented with 30 mM citrate or α-ketoglutarate.

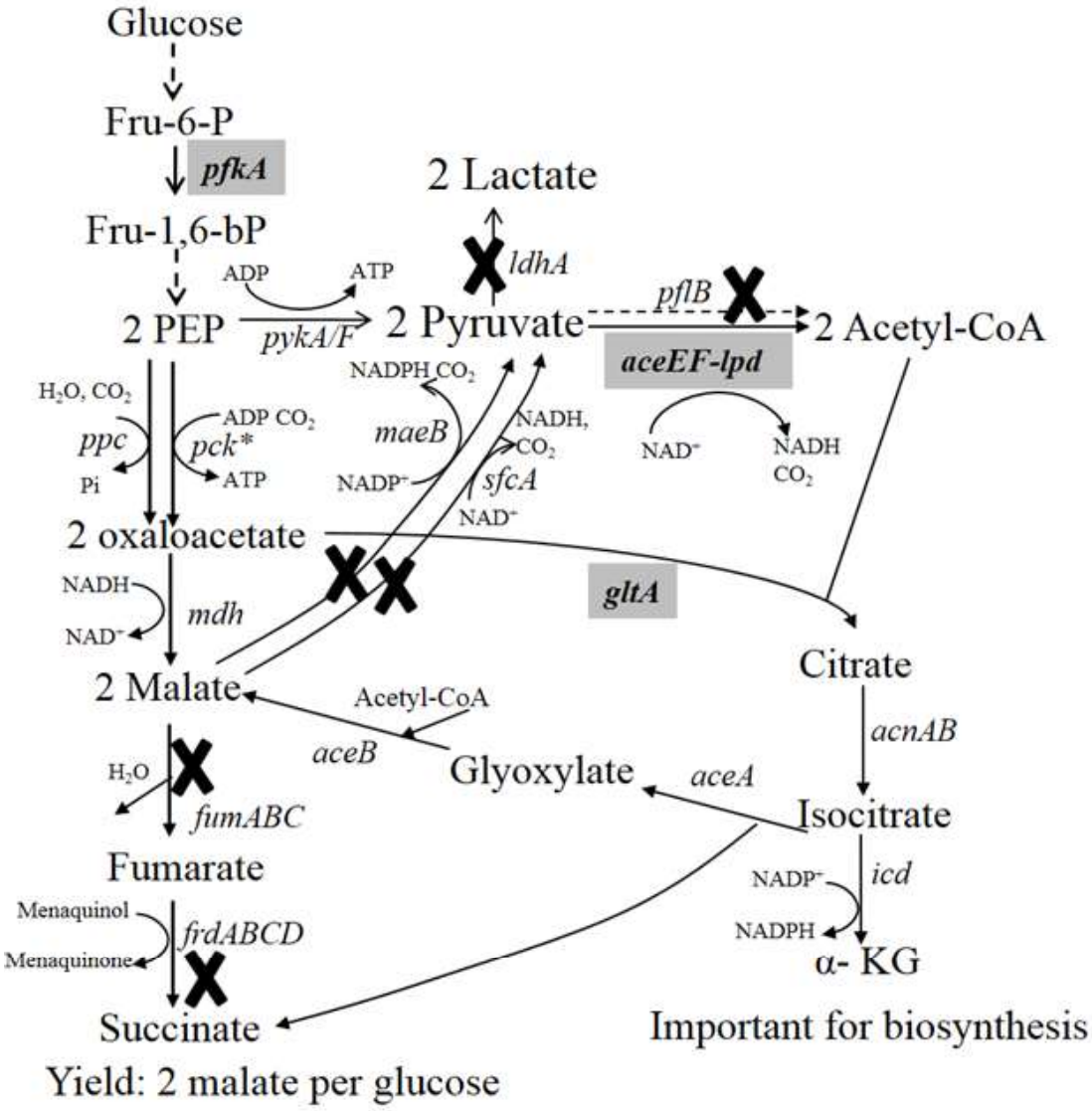


Figure 7.1. Potential metabolic allosteric regulation for malate production. The reductive branch of TCA cycle was engineered to accumulate malate as the sole fermentation product by deleting genes encoding fumarases (*fumB* and *fumAC*), fumarate reductase (*frdABCD*), and malic enzymes (*maeB* and *sfcA*). A redox-balanced malate-producing pathway converts one molecule of glucose into two molecules of malate. Some intermediate steps in glycolysis are not shown for clarity and simplicity. The malate-producing strain contains spontaneous mutations in *pck* and *ptsI* that were important to enhance carbon flow in the reductive branch of TCA cycle (11). Three potential allosteric regulation sites are indicated in gray boxes: 6-phosphofructokinase, pyruvate dehydrogenase, and citrate synthase encoded by *pfkA*, *aceEF-lpd* and *gltA*, respectively. Abbreviations: PEP, phosphoenolpyruvate; α-KG, α-ketoglutarate; acetyl-CoA, acetyl-Coenzyme A.

References:

1. Nieves LM, Panyon LA, Wang X. 2015. Engineering Sugar Utilization and Microbial Tolerance toward Lignocellulose Conversion. Front Bioeng Biotechnol 3:17.

2. Van Dien S. 2013. From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. Current Opinion in Biotechnology 24:1061-1068.

3. Kircher M. 2015. Sustainability of biofuels and renewable chemicals production from biomass. Current Opinion in Chemical Biology 29:26-31.

4. Werpy T, and G. Petersen. 2004. Top value added chemicals from biomass.U.S. Department of Energy, Washington, DC. http://www1.eere.energy.gov/biomass/pdfs/35523.pdf.

5. McKinlay JB, Vieille C, Zeikus JG. 2007. Prospects for a bio-based succinate industry. Appl Microbiol Biotechnol 76:727-40.

6. Yin X, Li J, Shin HD, Du G, Liu L, Chen J. 2015. Metabolic engineering in the biotechnological production of organic acids in the tricarboxylic acid cycle of microorganisms: Advances and prospects. Biotechnol Adv 33:830-41.

7. Huang WD, Zhang YHP. 2011. Analysis of biofuels production from sugar based on three criteria: Thermodynamics, bioenergetics, and product separation. Energy & Environmental Science 4:784-792.

8. Zheng YN, Li LZ, Xian M, Ma YJ, Yang JM, Xu X, He DZ. 2009. Problems with the microbial production of butanol. Journal of Industrial Microbiology & Biotechnology 36:1127-1138.

9. Lin H, Bennett GN, San KY. 2005. Fed-batch culture of a metabolically engineered *Escherichia coli* strain designed for high-level succinate production and yield under aerobic conditions. Biotechnology and Bioengineering 90:775-779.

10. Jantama K, Haupt MJ, Svoronos SA, Zhang X, Moore JC, Shanmugam KT, Ingram LO. 2008. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. Biotechnol Bioeng 99:1140-53.

11. Zhang X, Wang X, Shanmugam KT, Ingram LO. 2011. L-malate production by metabolically engineered *Escherichia coli*. Appl Environ Microbiol 77:427-34.

12. Battat E, Peleg Y, Bercovitz A, Rokem JS, Goldberg I. 1991. Optimization of L-Malic Acid Production by Aspergillus-Flavus in a Stirred Fermenter. Biotechnology and Bioengineering 37:1108-1116.

13. Knuf C, Nookaew I, Brown SH, McCulloch M, Berry A, Nielsen J. 2013. Investigation of Malic Acid Production in Aspergillus oryzae under Nitrogen Starvation Conditions. Applied and Environmental Microbiology 79:6050-6058.

14. Brown SH, Bashkirova L, Berka R, Chandler T, Doty T, McCall K, McCulloch M, McFarland S, Thompson S, Yaver D, Berry A. 2013. Metabolic engineering of Aspergillus oryzae NRRL 3488 for increased production of L-malic acid. Applied Microbiology and Biotechnology 97:8903-8912.

15. Taing O, Taing K. 2007. Production of malic and succinic acids by sugar-tolerant yeast Zygosaccharomyces rouxii. European Food Research and Technology 224:343-347.

16. Zelle RM, De Hulster E, Kloezen W, Pronk JT, van Maris AJA. 2010. Key Process Conditions for Production of C-4 Dicarboxylic Acids in Bioreactor Batch Cultures of an Engineered Saccharomyces cerevisiae Strain. Applied and Environmental Microbiology 76:744-750.

17. Zelle RM, de Hulster E, van Winden WA, de Waard P, Dijkema C, Winkler AA, Geertman JMA, van Dijken JP, Pronk JT, van Maris AJA. 2008. Malic acid production by Saccharomyces cerevisiae: Engineering of pyruvate carboxylation, oxaloacetate reduction, and malate export. Applied and Environmental Microbiology 74:2766-2777.

18. Chen XL, Wang YC, Dong XX, Hu GP, Liu LM. 2017. Engineering rTCA pathway and C4-dicarboxylate transporter for L-malic acid production. Applied Microbiology and Biotechnology 101:4041-4052.

19. Zou X, Zhou YP, Yang ST. 2013. Production of polymalic acid and malic acid by Aureobasidium pullulans fermentation and acid hydrolysis. Biotechnology and Bioengineering 110:2105-2113.

20. Wang HT, Yang Y, Wang HL, Sun T, Liu Y, Pan LJ, Li XJ. 2014. Improvement of the Tolerance to Toxic Inhibitor for Rhizopus Delemar Producing Malic Acid from Crop Straw Hydrolyte. Basic & Clinical Pharmacology & Toxicology 115:8-9.

21. Li XJ, Liu Y, Yang Y, Zhang H, Wang HL, Wu Y, Zhang M, Sun T, Cheng JS, Wu XF, Pan LJ, Jiang ST, Wu HW. 2014. High levels of malic acid production by the bioconversion of corn straw hydrolyte using an isolated Rhizopus delemar strain. Biotechnology and Bioprocess Engineering 19:478-492.

22. Deng Y, Mao Y, Zhang XJ. 2016. Metabolic engineering of a laboratory-evolved Thermobifida fusca muC strain for malic acid production on cellulose and minimal treated lignocellulosic biomass. Biotechnology Progress 32:14-20.

23. Dong XX, Chen XL, Qian YY, Wang YC, Wang L, Qiao WH, Liu LM. 2017. Metabolic engineering of *Escherichia coli* W3110 to produce L-malate. Biotechnology and Bioengineering 114:656-664.

24. Zhang X, Jantama K, Moore JC, Jarboe LR, Shanmugam KT, Ingram LO. 2009. Metabolic evolution of energy-conserving pathways for succinate production in *Escherichia coli*. Proc Natl Acad Sci U S A 106:20180-5.

25. Babul J, Robinson JP, Fraenkel DG. 1977. Are Aerobic and Anaerobic Phosphofructokinases of Escherichia-Coli Different. European Journal of Biochemistry 74:533-537.

26. Kim Y, Ingram LO, Shanmugam KT. 2008. Dihydrolipoamide dehydrogenase mutation alters the NADH sensitivity of pyruvate dehydrogenase complex of *Escherichia coli* K-12. J Bacteriol 190:3851-8.

27. Kim Y, Ingram LO, Shanmugam KT. 2007. Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes. Appl Environ Microbiol 73:1766-71.

28. Danson MJ, Harford S, Weitzman PDJ. 1979. Studies on a Mutant Form of Escherichia-Coli Citrate Synthase Desensitized to Allosteric Effectors. European Journal of Biochemistry 101:515-521.

29. Kotlarz D, Garreau H, Buc H. 1975. Regulation of the amount and of the activity of phosphofructokinases and pyruvate kinases in *Escherichia coli*. Biochim Biophys Acta 381:257-68.

30. Sahlman L, Williams CH, Jr. 1989. Lipoamide dehydrogenase from *Escherichia coli*. Steady-state kinetics of the physiological reaction. J Biol Chem 264:8039-45.

31. Weitzman PD. 1967. Allosteric fine control of citrate synthase in *Escherichia coli*. Biochim Biophys Acta 139:526-8.

32. Duckworth HW, Anderson DH, Bell AW, Donald LJ, Chu AL, Brayer GD. 1987. Structural basis for regulation in gram-negative bacterial citrate synthases. Biochem Soc Symp 54:83-92.

33. Stokell DJ, Donald LJ, Maurus R, Nguyen NT, Sadler G, Choudhary K, Hultin PG, Brayer GD, Duckworth HW. 2003. Probing the roles of key residues in the unique regulatory NADH binding site of type II citrate synthase of *Escherichia coli*. J Biol Chem 278:35435-43.

34. Auzat I, Lebras G, Garel JR. 1994. The Cooperativity and Allosteric Inhibition of Escherichia-Coli Phosphofructokinase Depend on the Interaction between Threonine-125 and Atp. Proceedings of the National Academy of Sciences of the United States of America 91:5242-5246.

35. Underwood SA, Buszko ML, Shanmugam KT, Ingram LO. 2002. Flux through citrate synthase limits the growth of ethanologenic *Escherichia coli* KO11 during xylose fermentation. Appl Environ Microbiol 68:1071-81.

36. Datsenko KA, Wanner BL. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640-5.

37. Jantama K, Zhang X, Moore JC, Shanmugam KT, Svoronos SA, Ingram LO. 2008. Eliminating side products and increasing succinate yields in engineered strains of *Escherichia coli* C. Biotechnol Bioeng 101:881-93.

38. Sievert C, Nieves LM, Panyon LA, Loeffler T, Morris C, Cartwright RA, Wang X. 2017. Experimental evolution reveals an effective avenue to release catabolite repression via mutations in XylR. Proc Natl Acad Sci U S A 114:7349-7354.

39. Higuchi R, Krummel B, Saiki RK. 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16:7351-67.

40. Ow DSW, Nissom PM, Philp R, Oh SKW, Yap MGS. 2006. Global transcriptional analysis of metabolic burden due to plasmid maintenance in *Escherichia coli* DH5 alpha during batch fermentation. Enzyme and Microbial Technology 39:391-398.

41. Keasling JD. 2008. Synthetic biology for synthetic chemistry. ACS Chem Biol 3:64-76.

42. Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, Ingram LO. 2010. Metabolic engineering for production of biorenewable fuels and chemicals: contributions of synthetic biology. J Biomed Biotechnol 2010:761042.

43. Ruch FE, Kuritzkes DR, Lin EC. 1979. Use of lac operon fusions to isolate *Escherichia coli* mutants with altered expression of the fumarate reductase system in response to substrate and respiratory controls. Biochem Biophys Res Commun 91:1365-70.

44. Lohmeier E, Hagen DS, Dickie P, Weiner JH. 1981. Cloning and Expression of the Fumarate Reductase Gene of Escherichia-Coli. Canadian Journal of Biochemistry 59:158-164.

45. Matsushita K, Kaback HR. 1986. D-lactate oxidation and generation of the proton electrochemical gradient in membrane vesicles from *Escherichia coli* GR19N and in proteoliposomes reconstituted with purified D-lactate dehydrogenase and cytochrome o oxidase. Biochemistry 25:2321-7.

46. Dong JM, Taylor JS, Latour DJ, Iuchi S, Lin EC. 1993. Three overlapping lct genes involved in L-lactate utilization by *Escherichia coli*. J Bacteriol 175:6671-8.

47. Sevin DC, Fuhrer T, Zamboni N, Sauer U. 2017. Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in *Escherichia coli*. Nat Methods 14:187-194.

48. Pinchuk GE, Rodionov DA, Yang C, Li X, Osterman AL, Derwyn E, Geydebrekht OV, Reed SB, Romine MF, Collart FR, Scott JH, Fredrickson JK, Beliaev AS. 2009. Genomic reconstruction of Shewanella oneidensis MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. Proc Natl Acad Sci U S A 106:2874-9.

49. Nunez MF, Kwon O, Wilson TH, Aguilar J, Baldoma L, Lin EC. 2002. Transport of L-Lactate, D-Lactate, and glycolate by the LldP and GlcA membrane carriers of *Escherichia coli*. Biochem Biophys Res Commun 290:824-9.

50. Wang ZW, Saini M, Lin LJ, Chiang CJ, Chao YP. 2015. Systematic Engineering of *Escherichia coli* for d-Lactate Production from Crude Glycerol. J Agric Food Chem 63:9583-9.

51. Bar-Even A, Noor E, Lewis NE, Milo R. 2010. Design and analysis of synthetic carbon fixation pathways. Proc Natl Acad Sci U S A 107:8889-94.

52. Liu JJ, Xie ZP, Shin HD, Li JH, Du GC, Chen J, Liu L. 2017. Rewiring the reductive tricarboxylic acid pathway and L-malate transport pathway of Aspergillus oryzae for overproduction of L-malate. Journal of Biotechnology 253:1-9.

53. Bar-Even A, Noor E, Milo R. 2012. A survey of carbon fixation pathways through a quantitative lens. Journal of Experimental Botany 63:2325-2342.

54. Ye XT, Honda K, Morimoto Y, Okano K, Ohtake H. 2013. Direct conversion of glucose to malate by synthetic metabolic engineering. Journal of Biotechnology 164:34-40.

55. Stols L, Donnelly MI. 1997. Production of succinic acid through overexpression of NAD(+)-dependent malic enzyme in an *Escherichia coli* mutant. Applied and Environmental Microbiology 63:2695-2701.

56. Bologna FP, Andreo CS, Drincovich MF. 2007. *Escherichia coli* malic enzymes: Two isoforms with substantial differences in kinetic properties, metabolic regulation, and structure. Journal of Bacteriology 189:5937-5946.

57. Kim P, Laivenieks M, Vieille C, Zeikus JG. 2004. Effect of overexpression of Actinobacillus succinogenes phosphoenolpyruvate carboxykinase on succinate production in *Escherichia coli*. Applied and Environmental Microbiology 70:1238-1241.

58. Laivenieks M, Vieille C, Zeikus JG. 1997. Cloning, sequencing, and overexpression of the Anaerobiospirillum succiniciproducens phosphoenolpyruvate carboxykinase (pckA) gene. Applied and Environmental Microbiology 63:2273-2280.

59. Podkovyrov SM, Zeikus JG. 1993. Purification and Characterization of Phosphoenolpyruvate Carboxykinase, a Catabolic Co2-Fixing Enzyme, from Anaerobiospirillum-Succiniciproducens. Journal of General Microbiology 139:223-228.

60. Eggeman T, Verser D. 2005. Recovery of organic acids from fermentation broths. Appl Biochem Biotechnol 121-124:605-18.

61. Frost JW, Draths KM. 1995. Biocatalytic Syntheses of Aromatics from D-Glucose - Renewable Microbial Sources of Aromatic-Compounds. Annual Review of Microbiology 49:557-579.

62. Li K, Mikola MR, Draths KM, Worden RM, Frost JW. 1999. Fed-batch fermentor synthesis of 3-dehydroshikimic acid using recombinant *Escherichia coli*. Biotechnology and Bioengineering 64:61-73.

63. Wendisch VF. 2014. Microbial production of amino acids and derived chemicals: Synthetic biology approaches to strain development. Current Opinion in Biotechnology 30:51-58.

64. Amann E, Ochs B, Abel KJ. 1988. Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene 69:301-15.