OPTIMIZING PENTOSE SUGAR UTILIZATION IN *ESCHERICHIA COLI* FOR THE PRODUCTION OF BIOFUELS

BY

KHUSHNUMA KOITA

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemical Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Associate Professor Christopher Rao, Chair
Professor Huimin Zhao
Professor Daniel Pack
Associate Professor Kaustubh Bhalerao

Abstract

The hydrolysis of biomass yields a sugar mixture consisting mainly of glucose, arabinose and xylose. Effective metabolism of all sugars in biomass by a microorganism is regarded as essential for commercial biofuel production. However, two of the major challenges that we are currently faced with are the transport of sugars into the microorganism and the co-utilization of these sugars once they are in the cell.

In order to engineer simultaneous multiple sugar utilization in *Escherichia coli*, a better understanding of the pentose sugar pathways is required. In this work, we have investigated the transport of sugars and the regulation of the sugar metabolic pathways within *E. coli* to engineer a strain most efficient in producing biofuels. While extensive research has been carried out to examine the transport mechanisms of sugars into the cell, this research shows that in addition to transporters that pump sugars into the cell, a number of proteins that pump sugars out of the cell are also expressed by *E. coli*. Using genetic approaches, we have demonstrated that by either deleting or overexpressing these efflux transporters we can respectively increase or decrease the uptake of pentose sugars, namely arabinose and xylose, which are abundantly present in the hemicellulose of biomass.

In addition to examining transport mechanisms, this work has also focused on studying and controlling the metabolism of the pentose sugars. By using a novel targeted approach, we can utilize constitutive promoters and chromosomal integration to control the expression of certain metabolic genes, while relieving repression effects. This enables us to regulate the metabolism of pentose sugars such as xylose that are utilized by the cell less efficiently and allows for simultaneous metabolism of sugars, hence leading to a more efficient biofuel production process.

For my parents Abde and Tasneem Koita

Acknowledgements

This project and thesis would have not been possible without the support of many people. For both technical and personal guidance, I must thank my adviser, Chris Rao for correctly guiding me and focusing my interest in the right projects. Without his support, advice and help, these projects would have never been completed successfully. Additionally, I would also like to thank my committee members Dr. Huimin Zhao, Dr. Dan Pack, Dr. Kaustubh Bhalerao and Dr. Nathan Price for being a part of my committee and offering me advice on my projects.

I would like to thank various members of the Rao lab, both past and present. I thank Lon Chubiz for educating me in experimental and microbiology techniques and providing me with valuable advice in conducting experiments. Also for his advice and teaching, I would like to thank Supreet Saini; a special thanks for constructing pPROBE *venus*, which I have used extensively in my experiments. In addition, I should also thank Tasha Desai, who trained me and made me familiar with the workings of the lab when I first joined the group. Also, thank you to Kang Wu for helping me out in the daily hurdles that we face in lab, to Yuki Kimura for being a supportive and fun classmate, and to my undergraduate student, Divya Reddy for her hard work on my projects. A big thanks to all the newbies in our lab as well: Santosh Koirala, Ahmet Badur, Kori Dunn, Shuyan Zhang (Julia), Payman Tohidifar, and Jiewen Zhou. They all have made lab a fun and welcoming environment and it has been a pleasure to work (and play) with all of them. A special thanks to Angel Rivera for his infinite post doctor-ly wisdom! Each one of them has been extremely encouraging and understanding, which has eased the pressure on many difficult days.

Of course, I cannot thank my parents enough for all they have done. From emotional support, to financial support, to constant encouragement, they have been there every step of the way. Thanks to my little brother, Zain Koita, for his sibling love and unwavering confidence in my

capabilities. I would also like to thank all my friends, particularly Moulik Ranka and Caroline Milne. I am deeply grateful to Moulik who was always there to lend me an ear and keep me from pulling out my “stress pony-tails”. Caroline is without a doubt my “Carol-pedia”; I would have been lost without her advice and her company for chocolate, cheese and cookie dough indulgence. A warm thanks to my fitness instructor Lesa Scharnett, and my fellow kickboxers who shared their positive energy and enthusiasm every evening and kept me focused and motivated throughout the day. Without my friends and family, I would not have accomplished what I have or become the person I am today. I hope I have made them proud.

Finally, I would like to thank our collaborator George Ordal and his lab members, especially George Glekas for his advice and encouragement regarding the HPLC. Thanks to the Energy Biosciences Institute for supporting this work.

TABLE OF CONTENTS

CHAPTER 1: Introduction and motivation

1.1 Motivation

Economic and geopolitical factors such as high oil prices, environmental concerns, and supply instability have increased the need to develop renewable, cleaner energy sources. Biofuels are an attractive sustainable transportation fuel because they can be produced from renewable sources and can be readily incorporated into the current transportation infrastructure. Many also have the advantages of low toxicity, higher vapor pressure and a net reduction in the levels of atmospheric carbon dioxide [1]. A goal set forth in “The United States Bioenergy Vision and Sustainable Feedstock Supply” by the Biomass R&D Technical Advisory Committee is that, by year 2030, 30% of current U.S. petroleum-based fuels will be replaced by biofuels [2]. However, there are several factors limiting large scale, efficient processes for biofuels production. Among the many challenges, there exists the need for optimizing cellular systems of microbial hosts. In particular, three of the key issues that we are concerned with are (1) the inability for microorganisms to efficiently and simultaneously ferment all sugars found in biomass feedstocks, (2) tolerate stresses caused by alcohol accumulation and the presence of inhibitors, and (3) process hardiness. The work presented here primarily aims to tackle the issue of inefficient co-utilization of sugars in biomass by microbes; a solution to this challenge will present a large stride towards a large-scale economical biofuel generation.

1.2 Overview of the production of biofuels

The production of biofuels mainly consists of the following steps: pretreatment and hydrolysis of the biomass, fermentation of the sugars from the biomass by microbes, and finally distillation and other separation processes to retrieve the end products. Improvement and optimization at each step of the process is required to make the overall large-scale production of biofuels more efficient and economical.

The types of biofuels from microbes include bioethanol, biodiesel, biobutanol and methane/biogas. Biodiesel and bioethanol are the only fuels presently being produced on a large industrial scale, and bioethanol fermentation is by far the largest scale microbial process [3]. Biobutanol is another promising fuel that is being developed. While ethanol has the advantages of having a higher octane rating and lower toxicity, and requires lower energy for production (due to its lower boiling point), butanol has the benefits of a higher energy content, lower corrosivity, and lower hygroscopicity [4]. However, it is also important to note that biobutanol is produced at a much lower titer, and is not immediately compatible with the current fuel production infrastructure. Aside from butanol, other higher alcohols, alkanes, and various types of oils are possible biochemically derived biofuels (Figure 1.1). It is not clear yet which one(s) will be the ideal biofuel, and the answer to this question would depend on additional factors, such as the type of biomass available, particular climatic conditions, composition of engine emissions, and combustion performance. Therefore, even though it may be beneficial to develop a better biofuel alternative to ethanol as a long-term solution, short-term goals that aim to optimize bioethanol production are also desirable. The work described here is in the context of ethanol production, but much of the research analysis and results can be extended towards producing other biofuels of interest in the future.

Biomass represents the most abundant renewable carbon source available for the production of bioenergy. The US has the potential to produce nearly 1.3 billion dry tons of biomass annually corresponding to a displacement of as much as one-third of current transportation fuel demands [5]. To produce biofuels, biomass is first pretreated to solubilize lignin and to break open crystalline cellulose. Effective pretreatments must improve enzymatic hydrolysis, minimize carbohydrate losses, and prevent formation of by-products that might inhibit subsequent hydrolysis and fermentation steps. Different types of pretreatments include physical, chemical, and biological processes. The pretreatment step is then followed by a hydrolysis step in which

enzymes break open polymer chains and convert cellulose and hemicellulose into fermentable sugars such as glucose and xylose. The product of these two steps is then ready for fermentation by microorganisms that metabolize sugars into biofuels. The process is summarized in Figure 1.2.

At present, the commercial production of biofuels in the U.S. is largely based on the use of starch in grains such as maize. This is a mature commercial technology with well understood economics and the U.S. is now the world's largest producer of starch based ethanol as a 10% additive to gasoline (known as E-10). However, this is not a viable long-term option as increasing demand for corn based ethanol would negatively impact the food and feed industries and result in higher corn prices. In addition, the requirements for arable land would be significant and impractical. Under current conditions, corn based ethanol can only displace a small proportion of gasoline (~10%). It is therefore necessary to pursue other long term ethanol feedstocks to replace corn.

During the fermentation process, two microorganisms commonly used to produce ethanol are *Saccharomyces cervisiae* (yeast) and *Zymomonas mobilis* [6, 7, 8]. Both yeast and *Z. mobilis* produce ethanol and carbon dioxide as principal fermentation products. Yeast is predominantly used in the industry to covert starch to ethanol because of its genetic tractability and widespread industrial use [9]. However, it does not produce comparable ethanol yields to *Z. mobilis* [7, 10]. *Z. mobilis* is an obligately ethanologenic bacterium and hence is an appealing candidate for ethanol production. The challenge, however, is that both microorganisms are unable to utilize multiple sugars. While they readily consume hexoses such as glucose, they cannot natively metabolize pentose sugars such as arabinose and xylose. Therefore, heterologous complementation or significant genetic modification is required.

The current energy productivity (ratio of the energy obtained from ethanol to energy expended in its production) is 1.3-1.6, which is not optimal [11]. Additionally, the current choice of feedstock and overall ethanol production process is not a long-term solution to the growing energy demands. Research in the various fields related to the biofuel production process is essential so that we can overcome the limitations and challenges of the current situation.

1.3 Lignocellulosic biomass

An ideal biomass resource should be high yielding, have low production costs, be readily available, and have consistent desirable chemical concentrations. The feasibility of an energy crop depends largely on its production costs, cost of converting the biomass to usable energy, and cost of competing fuels. Lignocellulosic materials (e.g. wood, grass, forestry waste, agricultural residues, and municipal solid wastes) are most advantageous because they are non-food feedstocks, abundant, and cheap, which make them an attractive substrate for biofuel production. Lignocellulose is a three-dimensional polymeric composite formed by plants as structural material. It is composed of three main fractions: cellulose (~45% of dry weight), hemicellulose (~30% of dry weight), and lignin (~25% of dry weight). Table 1.1 shows the specific composition of potential lignocellulosic biomass resources [5]. Lignin, polyphenolic in structure, is the largest non-carbohydrate fraction of lignocellulose and therefore cannot be utilized in the fermentation process. Cellulose is a high molecular weight linear polymer of β-1,4-linked D-glucose units and often appears as a highly crystalline material. The basic repeating unit is the dissacharide cellobiose. The secondary and tertiary conformation of cellulose, as well as its close association with lignin, hemicellulose, starch, protein and mineral elements, makes cellulose a hydrolysis-resistant molecule. Cellulose can be hydrolyzed chemically by diluted or concentrated acid, or enzymatically (typically by cellulases). By cellulose saccharification, glucose can be obtained and then fermented to produce biofuels. Hemicelluloses are branched polysaccharides consisting of a mixture of hexoses (D-galactose, L-galactose, D-mannose, L-

rhamnose, L-fucose), pentoses (D-xylose, L-arabinose), and uronic acids (D-glucuronic acid). Hemicellulose is more easily hydrolyzed than cellulose, and the exact composition of hemicellulose depends on the source of the raw material, but it is most abundant in xylan (xylose polymer) [12]. Complete hydrolysis of xylan involves three main enzymes: endo-β-1-4-xylanase which primarily targets the internal β-1-4 bonds between xylose units, exoxylanase that releases xylobiose units, and β-xylosidase that releases xylose from xylobiose and short chain xylooligosachharides [13].

A potential lignocellulosic feedstock is switchgrass. Switchgrass shows promise due to its high productivity, suitability for marginal land quality, low water and nutritional requirements, environmental benefits, and flexibility for multipurpose uses. According to Morrow *et al.* [14], a mature bioenergy crop production, in which conversion of both pentoses and hexoses is assumed, would yield 330-380 L of ethanol per Mg of dry switchgrass. The theoretical ethanol production potential from highly adapted switchgrass varieties is between 5000 and 6000 L/ha. In comparison, the theoretical ethanol yield from corn starch is only about 4000 L/ha [5].

1.4 Sugar utilization by microbes

The ideal microorganism for biofuel production should possess high substrate utilization and processing capacities, fast and deregulated pathways for sugar transport, good tolerance to inhibitors and product, and high metabolic fluxes and should produce a single fermentation product [15].

This project is mainly concerned with efficient substrate utilization. As mentioned before, the hydrolysis of hemicellulose yields a sugar mixture consisting mainly of glucose, arabinose and xylose. Effective metabolism of all sugars in biomass is regarded as essential for commercial ethanol production. A significant amount of research has been devoted to using metabolic engineering to construct bacterial and yeast strains which feature traits most efficient in ethanol

production using hemicellulose sugars. Three main emerging microbial platforms are: *Saccharomyces cerevisiae, Zynomonas mobilis* and *Escherichia coli. Z. mobilis* and yeast ferments hexose sugars such as glucose, but cannot natively metabolize the pentose sugars: xylose and arabinose. Unlike yeast or *Z. mobilis*, *E. coli* can natively metabolize all three sugars [16-18]. Metabolic engineering has enabled the addition of pathways for pentose sugar utilization in several of these microorganisms. Examples include: *S. cerevisiae*, *Zymobacter palmae* [19], *Z. mobilis*, and the recently engineered derivative of *Pseudomonas putida* S12 containing *E. coli* genes for xylose metabolism [20]. Furthermore, a great deal of work is focused on the isolation and characterization of *Pichia stipitis*, *Clostridium thermocellum* and *Clostridium phytofermentans* [21-23].

In the case of yeast, strains capable of xylose and arabinose catabolism have been engineered [24-27], but continued optimization of these strains is required. Two types of pentose pathways have been constructed: the oxidoreductase pathway and the isomerase pathway. Both xylose and arabinose can be metabolized through each of these pathways, although arabinose assimilation involves additional steps in both cases [28]. All four possible pathway variants have been previously constructed [28-31], and all feed into native yeast metabolism via D-xylulose or D-xylulose-5-phosphate. Once converted to xylulose-5-phosphate, these sugars are further metabolized through the native pentose phosphate pathway. Ethanol titers of 36 g/L from switchgrass have been achieved [32]. Even though arabinose and xylose can be catabolized by these engineered yeast strains, reported growth rates remain suboptimal for economical production of biofuels [15].

Z. mobilis can natively consume only hexose sugars such as glucose, fructose, and sucrose. However, metabolic engineering has enabled this organism to consume pentose sugars as well. Studies on recombinant strains created at NREL (The National Renewable Energy Laboratories) have involved the construction of integrant xylose-utilizing strains and additionally

an integrant xylose/arabinose-utilizing strain designated AX101. These strains are able to utilize pentose sugars efficiently and produce high yields of ethanol [33]. *Z. mobilis* has several advantages over yeast. Considerable faster rates of sugar uptake and ethanol production, better ethanol tolerance, simpler growth conditions, and higher reported productivities (120-200 g/L hr in continuous processes with cell recycle compared to only 30-40 g/L hr for yeast), make this organism a promising microbe for biofuel production. Another big advantage of this organism is that it is not susceptible to catabolite repression effects like in the case of *E. coli*. Therefore, once pentose sugar utilization is optimized in this organism, challenges related to repression by glucose do not need to be overcome. Engineered *Z. mobilis* strains are reported to produce ethanol at concentrations above 60 g/L within 48 hours (in a 65 g/L glucose and 65 g/L xylose mixture) [33]. Some concerns, however, are that *Z. mobilis* may be less robust than yeast and more susceptible to contamination in large-scale processes. Also, there is a lack of experience with large-scale *Z. mobilis* fermentation in the ethanol industry, which makes the switch from yeast to *Z. mobilis* challenging.

Several studies have explored the issue of using mixed cultures (mixture of several isolated or engineered strains) for more efficient ethanol conversion [34-35]. These studies have shown that mixed cultures can result in 2-12% higher ethanol yields than that of a single microorganism. For example, recent studies have shown that a mixed culture of substrate-selective engineered *E. coli* strains (one glucose-deficient and one xylose-deficient) consumed the sugars in 15% less time than the individual strains when allowed to ferment in a mixture of glucose and xylose. Despite these positive findings, fermentations with a single organism continue to be researched. The difficulties associated with the stability of mixed cultures and the ease of bioprocessing of a single host system are the key reasons why there is continuing interest in finding a single organism solution to the biofuel problem.

While it is possible to engineer microorganisms like *Saccharomyces cerevisiae* and *Z. mobilis* so that they are able to ferment pentose sugars by introducing genes from other organisms containing the requisite pathways, similar sugar utilization efficiencies to that in *E.* coli have not yet been achieved. Other advantages to using *E. coli* are that much is known about the sugar metabolism pathway and transport in this organism, and the bacterium has a robust genetic system. Unfortunately, sugar co-utilization in *E. coli* is hindered by a number of factors, including catabolite repression, sugar transport limitations and intracellular sugar concentration [36-38]. The bacterium will always consume just one sugar while repressing the metabolism of all others [39-41]. In particular, *E.* coli will first consume glucose, then arabinose, and then finally xylose. To overcome this characteristic, several genetic changes have been made and more efficient strains of *E.coli* have been engineered. An important improvement to wild-type *E. coli* has been to introduce the genes *pdc (*pyruvate decarboxylase) and *adhB* (alcohol dehydrogenase II) from *Z. mobilis* to maximize ethanol production during mixed acid fermentation. Additionally, Yomano *et al.* showed that the deletion of methylglyoxal synthase gene increased sugar co-matabolism in ethanol producing *E. coli* [42]. However, a lag in xylose metabolism was observed and they found that injection of small amounts of air was required to relieve this lag. Strains constructed by Trinh *et al*. also achieved high ethanol yields and their minimal cell approach and elementary mode (EM) analysis eliminated the effects of catabolite repression in glucose-xylose sugar mixtures [43]. The reported ethanol yield for their strain, TCS083/pL0I297, is 0.49 ± 0.01 (g ethanol/g sugar). This is very close to the theoretical yield of 0.51g/g and higher than previously reported ethanol yields for engineered *E. coli* (K011 and FBR5/pL0I297) [44-45]. However, the specific growth rate on glucose was reduced in this minimal strain. When grown at a high specific growth rate, pronounced catabolite repression was observed and preferential glucose consumption comparable to that of the wild-type strain was seen under aerobic conditions.

Clearly, there is room for further studies and different approaches. *E. coli* serves as a model organism to test various approaches because of its native ability to degrade a mixture of sugars and because it can be easily engineered according to rational strain design with well-established molecular techniques. The final goal is to be able to genetically engineer *E. coli* to efficiently and robustly utilize multiple sugars to maximize substrate utilization and optimize ethanol production during fermentation.

1.5 Sugar metabolic pathway in *E.coli* and catabolite repression

The *E. coli* metabolic pathway of the three sugars: glucose, arabinose and xylose, is illustrated in Figure 1.3. During the uptake of glucose, one unit of ATP is used to convert the glucose to D-glucose-6-phosphate. The transport of glucose into the cell occurs through the phosphotransferase system (PTS). D-glucose-6-phosphate is then converted to fructose-6-phosphate and finally to pyruvate through the glycolysis pathway. ATP is consumed during glycolysis, but four units are also produced; this results in a positive net energy change of two molecules of ATP [46]. In the case of the pentose sugars, arabinose and xylose are both converted to D-xylulose-5-phosphate before they enter the pentose phosphate pathway and are converted to fructose-6-phosphate. From this point in the pathway, they go through glycolysis to produce pyruvate. Finally, under anaerobic conditions, the pyruvate is converted to mixed acid fermentation products such as lactate, acetate, and most importantly, biofuels like ethanol.

The regulatory system that prioritizes the sequential metabolism of sugars in *E. coli* is carbon catabolite repression, which involves cyclic AMP, cyclic AMP-binding protein (CRP), enzymes of the phophotransferase system, and other components [38]. Through this system, glucose effectively blocks the expression of sugar specific transporters and essential enzymes required for the metabolism of alternative sugars.

The transport of glucose into the cell occurs through the PTS using the glucose-specific PTS permease. The phosphoenolpyruvate-dependent glucose PTS is responsible for the transport and the phosphorylation of the sugar in a two-step reaction: (1) the phosphoenolpyruvate (PEP), catalyzed by the enzyme EI, uses the protein HPr as an intermediate phosphoryl donor; (2) HPr then donates the phosphoryl group to a histidine residue in the EII domains of various substrate specific transporters, and from there to the carbohydrate during its uptake through the membrane domain as shown in Figure 1.4 [38].

Regulation of carbon catabolite repression in *E. coli* is brought about the modulation of the phosphorylation state of $EIIA^{Glc}$ (the EIIA domain of the glucose transporter). $EIIA^{Glc}$ is preferentially dephosphorylated when the cells are growing rapidly in the presence of easily metabolizable sugars, such as glucose. Regulation by phosphorylated $EIIA^{Glc}$ occurs by activation of the membrane bound enzyme adenylate cyclase, which is capable of synthesizing cyclic AMP (cAMP). Once cAMP has been produced, it binds its receptor protein CRP, and the cAMP-CRP complex is then able to activate the promoters of several catabolic genes and operons. In summary, when glucose is present, the PTS system tightly limits the cAMP-CRP complex formation and hence hinders the metabolic pathways of other sugars (e.g. pentose sugars), and prevents their co-utilization in the cell.

1.6 Pentose sugars

The two pentose sugars abundantly present in hemicellulose of biomass are arabinose and xylose. Figure 1.3 illustrates the metabolic pathways of these sugars in *E. coli* before they enter the pentose phosphate pathway, and highlights the metabolic genes and transporters for each sugar.

The genes responsible for arabinose uptake into the cell are *araE and araFGH* [47-49]. Both transporters are induced by arabinose. AraE is a low-affinity transporter, which means that it is

responsible for transporting arabinose when the sugar concentration is high outside the cell. It uses proton motive force as its energy source. On the other hand, AraFGH is a high affinity transporter, responsible for transport when the sugar concentration is low outside the cell [50]. The *araF* gene encodes the arabinose-binding protein (ABP), the *araG* gene encodes a probable ATPase subunit, and the *araH* gene encodes an integral membrane subunit [51]. It forms an ATP binding cassette, using energy from ATP as its energy source. A schematic of these transporters is shown in Figure 1.4. The enzymes required for the metabolism of arabinose are coded by the *araBAD* operon. These encode for L-arabinose isomerase (*araA*), L-ribulokinase (*araB*), and L-ribulose-5-phosphate-4-epimerase (*araD*). The synthesis of these proteins is controlled by the cAMP binding protein (i.e. the catabolite gene activator protein, CRP) and both positively and negatively controlled by the product of the *araC* gene. The regulatory region for the *araC* gene is adjacent to the regulatory region for the *araBAD* operon and their respective promoters are transcribed in opposite directions. The *araC* protein has three regulatory functions: 1) positive control of *araBAD*, 2) negative control of *araBAD*, and 3) negative control of *araC* [52]. In the presence of arabinose, AraC positively activates the transcription of the *araBAD* operon and the *araE and araF* genes. Both the *araBAD* operon and the *araC* gene are positively regulated via the binding of the cAMP-CRP complex to a single site. Therefore positive regulation of the *araBAD* operon requires the presence of both arabinose and CRP. In the absence of arabinose AraC acts as an inhibitor and represses the transcription of the metabolic genes *araA*, *araB* and *araD*, hence inhibiting the metabolism of arabinose.

Analogous to the arabinose transporters are the xylose transporters, *xylE* (low affinity transporter) and *xylFGH* (high affinity transporter) [53-54]. *xylF* encodes for the periplasmic xylose-binding protein, *xylG* encodes for ATPase, and *xylH* encodes for permease. The operon responsible for xylose metabolism is *xylAB*. The *xylA* gene encodes for D-xylose isomerase,

and the *xylB* gene encodes for xylulokinase. The regulator that positively controls the transcription of this operon is *xylR*. [55]. XylR, along with CRP, binds cooperatively to activate transcription of operons involved in transport and catabolism of D-xylose. Gene induction occurs when the physiological inducer, D-xylose, binds to XylR and when cellular cyclic AMP levels are high. It has also been found that arabinose represses the *xyl* operons and previous work has shown that AraC is responsible for the repression caused by arabinose on the expression of the xylose metabolic genes [56].

The effects of catabolite repression and the role that they play in sugar utilization have been extensively studied. However, the mechanisms for the regulation of the sugar hierarchy are not fully understood and optimal ways to bypass this hierarchy have not yet been achieved. Three hierarchical systems need to be overcome in *E. coli* to achieve maximum and simultaneous utilization of all the sugars present in biomass*:* the glucose-arabinose hierarchy, glucose-xylose, hierarchy, and the arabinose-xylose hierarchy. In this work we aimed to increase pentose sugar metabolism in two ways: 1) hypothesize that sugar transport and intracellular sugar concentration are key limiting factors in efficient sugar utilization and study ways in which we can reduce these limitations, and 2) analyze and understand sugar hierarchy and devise techniques to overcome it. The first part of this thesis primarily focuses on transport of the pentose sugars and aims to understand the role that efflux transporters play on pentose sugar metabolism. The second part of this thesis focuses on the hierarchy that exists between glucose and xylose and describes the work done to engineer a strain capable of co-utilizing these two sugars simultaneously.

1.7 Figures and tables

Table 1.1. Detailed compositional analysis of potential lignocellulosic biomass [5].

	Minimum	Maximum	Mean	SD
	% of dry matter			
Corn Stover				
Cellulose[b]	31.3	41	37.5	2.8
Structural glucan	33.8	41	37.5	2.2
Hemicellulose	20	34.4	26.1	4.8
Xylan	19.8	25.8	21.7	2.1
Arabinan	1.7	6.1	2.7	1.6
Galactan	0.7	3	1.6	1
Mannan	0.3	1.8	0.6	1.1
Total lignin	15.8	23.1	18.9	2.6
Acid soluble lignin	1.9	3.6	2.9	0.9
Acid insoluble lignin	13.6	19.8	16.4	3.1
Acid detergent lignin	3.1	5	4.1	1.3
Crude protein	3.5	8.7	4.7	2.2
Ash	4.2	7.5	6.3	1.2
Soil	-	-	1.3	-
Wheat Straw				
Cellulose	31.5	48.6	37.6	5.7
Structural glucan	31.5	32.6	32.1	4.5
Hemicellulose	22.6	38.8	28.8	5.7
Xylan	19.2	19.7	19.5	0.3
Arabinan	2.4	3.2	2.8	0.6
Galactan	0.8	1.5	1.1	0.5
Mannan	0.3	0.9	0.6	0.4
Total lignin	5.3	19	14.5	6.2
Acid soluble lignin	-	-	2.5	-
Acid insoluble lignin	5.3	16.5	10.9	7.9
Acid detergent lignin	7.6	11.2	9.2	1.6
Crude protein	1.9	5.7	3.8	1.9
Ash	1.4	10.2	6.4	3.4
Soil	-	-	-	-
Switchgrass				
Cellulose	31.4	45	37.3	4.4
Structural glucan	31.4	38	34.2	2.1
Hemicellulose	22	35.1	28.5	3.5
Xylan	20.2	24	22.8	1
Arabinan	2.7	3.8	3.1	0.5
Galactan	0.7	1.9	1.4	0.5
Mannan	0.3	0.4	0.3	0
Total lignin	17.7	22	19.1	1.7
Acid soluble lignin	3.3	3.7	3.5	0.3
Acid insoluble lignin	15.8	16.5	16.2	0.5
Acid detergent lignin	4	12	6.4	2.7
Crude protein	1.6	3.8	3.1	0.7
Ash	4.4	8.5	5.9	1
Soil	-	-	-	-

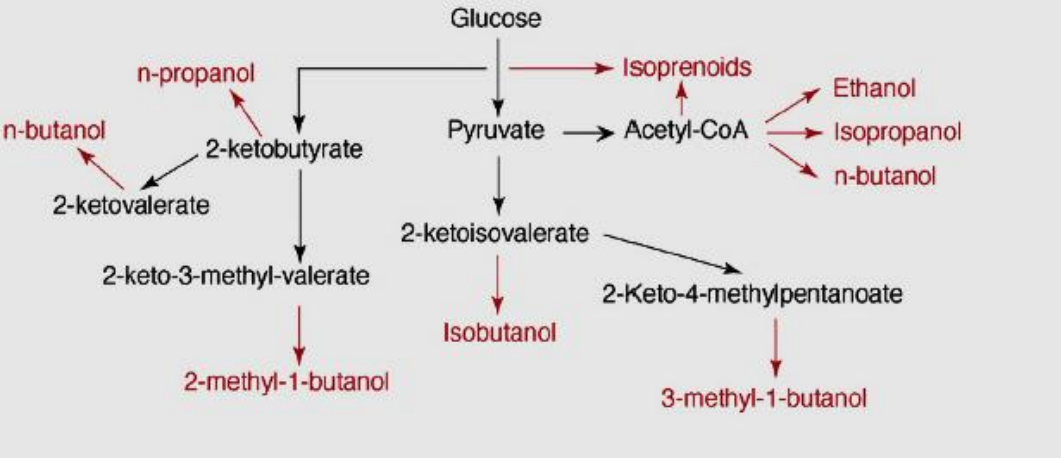

Figure 1.1. Schematic representation of engineered metabolic pathways to produce candidate biofuels.

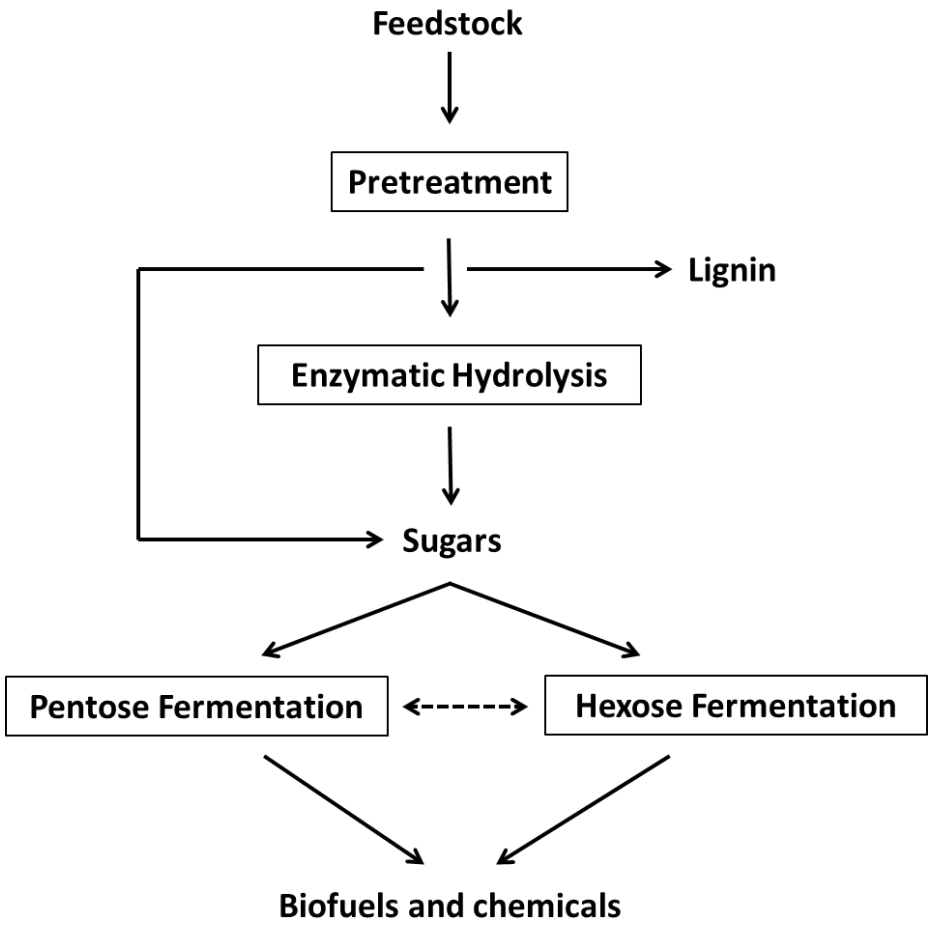

Figure 1.2. Overview of the production of biofuels.

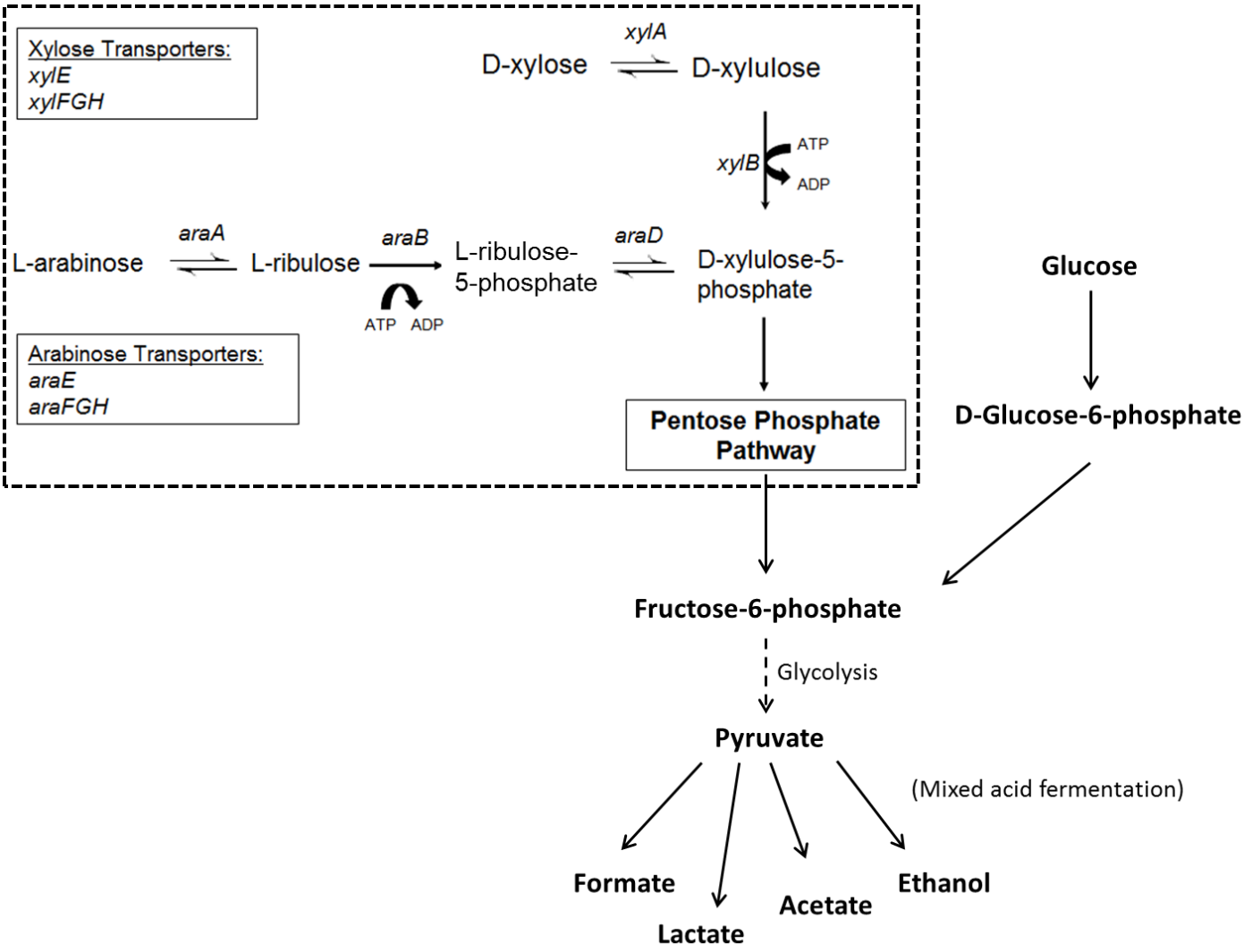

Figure 1.3. Sugar metabolic pathway in *E. coli*.

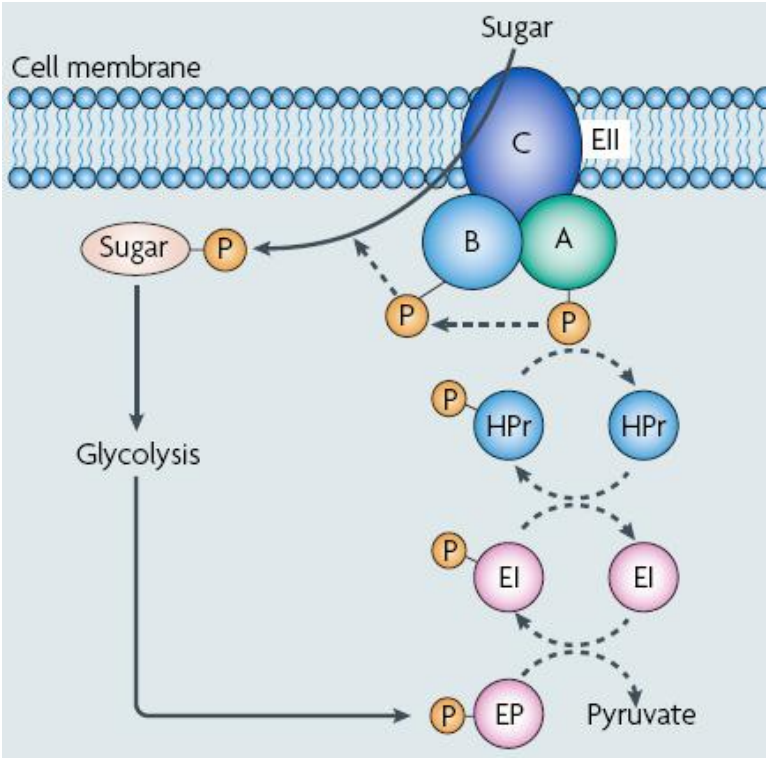

Figure 1.4. The phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS).

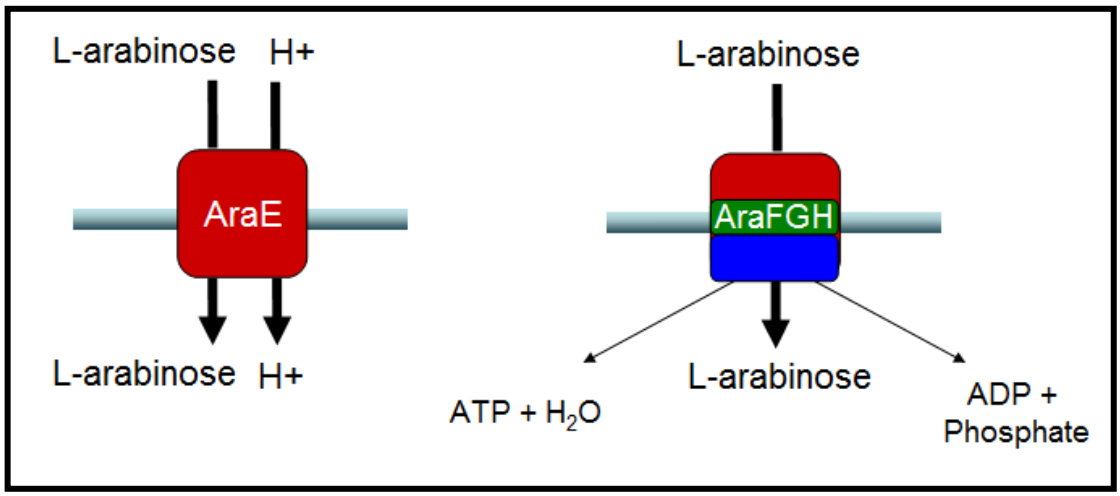

Figure 1.5. Schematic of arabinose transporters.

CHAPTER 2: Identification and analysis of the putative pentose sugar efflux transporters in *Escherichia coli*

2.1 Introduction and background

One of the factors mentioned in our hypothesis that limits metabolism of the pentose sugars in *E. coli* is transport. The metabolic genes of pentose sugars such as arabinose and xylose are controlled by genes that are induced by high sugar concentrations. Hence, the transport mechanism of the pentose sugars through the cell and the factors that contribute to the increased or decreased levels of intracellular sugar concentration needs to be understood in detail. This helps us gain knowledge of how sugar metabolism in the cell can be better regulated.

Escherichia coli expresses proteins that not only transport sugars into cells but also a number that transport them out of the cell. In contrast to the influx transporters described in the previous chapter, which are responsible for sugar uptake, there is also another class of transporters that has been identified, which is responsible for sugar efflux. The canonical example is the SET family of transporters: SetA, SetB, and SetC [57-58]. Efflux transporters are membrane transporter proteins that pump substances out of the cell. There is a vast family of efflux transporters, mainly classified based on the energy source required. These include: ATP-binding cassette superfamily (ABC), small multidrug resistance family (SMR), resistance-nodulation-cell division superfamily (RND), multidrug and toxic compound extrusion family (MATE), and major facilitator superfamily (MFS) [59]. Of interest is the MFS class of efflux transporters. These transporters use a proton symporter as an energy source to pump the substance of interest out of the cell. In *E. coli,* 64 such genes can be identified on the basis of sequence similarity, and the exact number could be somewhat higher [60].

Efflux transporters have been extensively studied in the literature in the context of multi-drug resistance (MDR). These transporters are responsible for pumping out substances that are harmful to the cell. In *E. coli,* efflux systems like AcrAB and EmrAB pump out toxins such as tetracycline and chloramphenicol out of the cell. These transporters are essential for drug resistance in bacteria, and hence a lot of research has been dedicated to studying them [61-68]. However, there is almost no literature that studies them in the context of sugar efflux. Little is known about sugar efflux transporters or necessarily why *E. coli* has them. One possible mechanism is to relieve sugar-phosphate stress. In particular, Sun and Vanderpool demonstrated that SetA participates in the glucose-phosphate stress response [69]. However, they found that SetA does not efflux alpha-methylglucoside, the nonmetabolizable sugar analog used to elicit the response. Another proposal is that they function as safety valves, preventing excess sugar from accumulating within the cell. However, this hypothesis has not yet been tested.

The only genes that have been characterized in the context of sugar efflux are *ydeA*, *setA*, *setB, setC*. Bost *et al.* [70] and Carolé S *et al.* [71] characterized *ydeA* as an arabinose efflux transporter in *E. coli*. Experiments suggest that when *ydeA* is overexpressed, intracellular sugar accumulation of arabinose is lowered. In addition, inactivation of *ydeA* decreases the efflux of arabinose which is indicated by higher intracellular arabinose concentration, thereby affecting the expression of AraC-regulated genes [72]. Both of these results help identify *ydeA* as a successful sugar efflux transporter. The sugar efflux transporter (SET) genes: *setA, setB and setC* are characterized by Jia Yeu Liu, Paul F. Miller *et al* [58]. *setA* and *setB* were found to be responsible for glucose and lactose efflux. In addition, they have broad specificity for various glucosides or galactosides. *setC* was not an efflux transporter of glucose or lactose, but no further experiments were carried out to determine the substrates of *setC*.

In this work, we investigated L-arabinose and D-xylose efflux in *E. coli*. Sequence analysis identified 31 candidate sugar efflux transporters in *E. coli* based on homology to known sugar efflux transporters (Table 2.1). This list was prepared by identifying all the genes that belonged to the same COG (Clusters of Orthologous Groups of proteins) as *ydeA*. The COG's are identified by NCBI by comparing sequences encoded in complete genomes, representing major phylogenetic lineages. Each COG consists of individual proteins or groups of paralogs from at least 3 lineages and thus corresponds to an ancient conserved domain. All the genes in *E .coli* within COG2814, the cluster that included *ydeA* and the *SET* genes, were tested. Using genetic approaches, we tested whether these candidate efflux transporters target arabinose and xylose. We were able to identify multiple putative arabinose efflux transporters but interestingly none for xylose.

2.2 Materials and methods

Media and growth conditions. Luria-Bertani liquid and solid medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) was used for routine bacterial culture and genetic manipulation. All experiments were performed in tryptone broth (10g/L tryptone, 5g/L NaCl) at 37°C. Antibiotics were used at the following concentrations: ampicillin at 100 µg/mL, chloramphenicol at 20 µg/mL, and kanamycin at 40 µg/mL. Inducer isopropyl-β-D-galactopyranoside (IPTG) was used at a concentration of 2 mM unless otherwise specified. All experiments involving the growth of cells carrying the helper plasmid pKD46 were performed at 30 ^{o}C. Loss of pKD46 was achieved by growth at 42 ^{o}C under nonselective conditions on LB agar. The removal of the antibiotic cassette from the FLP recombinant target (FRT)-chloramphenicol/kanamycin-FRT insert was obtained by the transformation of the helper plasmid pCP20 into the respective strain and selection on ampicillin at 30 ^{o}C. Loss of pCP20 was obtained by growth at 42 ^{o}C under nonselective conditions on LB agar. Primers were purchased from IDT, Inc. Enzymes were purchased from New England Biolabs and Fermentas.

Strain and Plasmid Construction. Bacterial strains and plasmids used in this study are described in Tables 2.2 and 2.3 respectively. All strains are isogenic derivatives of *Escherichia coli* K-12 strain MG1655. All cloning steps were performed in *E. coli* strain DH5α (*phi-80d lac*Δm15 *enda1 recA1 hsdR17 supE44 thh*-1 *gyrA96 relA*Δ*lacU169*). Targeted gene deletions and subsequent marker removal were made using the λRed recombinase method of Datsenko and Wanner [73]. The generalized transducing phage P1*vir* was used in all genetic crosses according to standard methods [74-75].

The plasmids pKD3 and pKD4 were used as templates to generate scarred FRT mutants as previously described. The primers used to create the deletions of each efflux transporter gene are displayed in Table 2.4. Mutations were checked by PCR using primers that bound outside the deleted region. These check primers are shown in Table 2.5 and were used in combination with test primers c1 (TTATACGCAAGGCGACAAGG) and c2 (GATCTTCCGTCACAGGTAGG) for *cat,* and k1 (CAGTCATAGCCGAATAGCCT), k2 (CGGTGCCCTGAATGAACTGC), for *kan*. Prior to the removal of the antibiotic resistance marker, the constructs resulting from this procedure were moved into a clean wild-type background by P1*vir* transduction.

The plasmid pPROBE *venus* was constructed by digesting the plasmid pQE80L-*venus* by EcoRI and NheI and cloning the fragment into pPROBE-*gfp*[tagless] digested by EcoRI and NheI. This replaced the *gfp*[tagless] by the fast-folding *yfp* variant Venus in pPROBE [75]. Venus transcriptional fusions were made by amplifying the promoter of interest and then cloning these PCR fragments into the multiple cloning site of pPROBE *venus*. The *araBAD* promoter (region 67884-68377) was amplified using primers TD088f (sequence: 5'- GGA AAG GTA CCC ATT CCC AGC GGT CG) and TD088r (sequence: 5'- GAC TAG AAT TCG CCA AAA TCG AGG CC). The *xylA* promoter (region 3726575-3727102) was amplified using primers TD065f (sequence: 5'- GGA AAG GTA CCT CGA TCT TTT TGC CA) and TD065r (5'- GAC TAG AAT TCG CGA TCG AGC TGG TC). The promoters of the potential efflux genes of interest were

cloned using the primers listed in Table 2.6 and the region amplified is noted. The PCR fragments were then digested with KpnI and EcoRI (sequences underlined) and cloned into the multiple cloning site of the pPROBE *venus* vector. The resulting transcriptional fusions were transformed into MG1655 wild-type *E. coli* strains and the mutated strains containing the deletions of interest. The primers used to check the mutations for all the pPROBE *venus* transcriptional fusions are KK010f (sequence: 5'- TAA ACT GCC AGG AAT TGG GG) and SS123r (sequence: 5'- GGG TCT AGA TTA TTT GTA TAG TTC ATC CAT GCC).

Plasmids overexpressing the genes of interest were constructed by cloning the respective gene into the multiple cloning site of pTrc99A under the control of a strong IPTG inducible promoter, P_{trc} [76]. The primers used to clone the specific genes are shown in Table 2.7, with the enzyme restriction sites underlined. The check primers used are KK027f (sequence: 5'- GAC AGC TTA TCA TCG ACT GC) and KK027r (sequence: 5'- CTG GCA GTT CCC TAC TCT CG).

Fluorescence assays. End-point measurements of the fluorescent reporter system were made using a Tecan Safire 2 microplate reader. 3 mL cultures were grown overnight in tryptone media at 37°C and then subcultured 1:30 after which 0.45mL was transferred to a single well of a polypropylene, 2.2 ml, deep, square, 96-well microtiter plate (VWR; 82006-448). Cultures were then grown at 37°C while shaking at 600 rpm on a microplate shaker (VWR). After an optical density of 0.05 was reached, the cells were induced with varying concentrations of the sugar of interest (either arabinose or xylose). Extracellular sugar concentrations ranged from 1 mM to 10 mM. In addition, strains containing the expression plasmids for the genes of interest were also induced with IPTG. Final volumes in each well were adjusted to 0.5 mL for all cultures. When the cells reached an OD of 0.4-0.5, 100 µL of the culture was transferred from the deep-well plates to black, clear-bottomed Costar 96-well microtiter plates, and the fluorescence (excitation/emission λ, 515/530 nm, bandwidth 5nm, gain 92, and OD at 600nm (OD_{600}) were measured. Fluorescence measurements are reported as the relative fluorescence normalized to

the optical density of the sample to correct for differences in cell density. All experiments were conducted in triplicates and average values with one standard deviation error bars are reported.

2.3 Results

Effect of deleting candidate sugar efflux transporters on arabinose metabolism. We identified 31 candidate sugar efflux transporters in *E. coli* based on homology to the known sugar efflux transporters: *ydeA*, *setA*, and *setB*. These homologs included a number of multidrug efflux transporters, both known and putative, along with five transporters involved in the uptake of various metabolites (*shiA*, *kgtP*, *galP*, *nanT*, and *proP*).

If any of these genes encodes an efflux transporter of arabinose, then deleting the gene should increase the intracellular sugar concentration of arabinose. As an indirect measure of intracellular arabinose concentrations, we employed transcriptional fusions of *araBAD* promoter to Venus, a fast-folding variant of the yellow fluorescent protein [77]. The average activity of the *araBAD* promoter as determined using fluorescence is proportional to extracellular arabinose concentrations and presumably relative intracellular arabinose concentrations as well based on the known mechanism for regulation [78]. To test the effect of deleting these genes, we grew the cells in tryptone broth at varying concentrations of arabinose and then harvested the cells at late-exponential phase to determine relative *araBAD* promoter activities. The resulting data were fit to a Michaelis-Menten dose-response curve with the governing parameters, V_{max} and K_m, reported (Table 2.8).

Deletion of the following eight genes was found to increase intracellular arabinose concentrations as compared to the wild-type control: *setC, cmr, ynfM, mdtD, yfcJ, yhhS, emrD* and *ydhC*. Deleting the other genes had no effect, with the dose-response curves indistinguishable from the wild-type control. A representative dose-response curve for the *Δcmr* mutant is shown in Figure 2.1. These results suggest that these eight genes encode arabinose

efflux transporters. Among them, only SetC has previously been implicated in sugar efflux based on its homology to SetA and SetB, two transporters known to efflux lactose [58]. Cmr, also known as MdfA, and EmrD are known multidrug efflux transporters. Over-expression of Cmr has also been shown to limit the uptake of isopropyl-β-D-thiogalactopyranoside (IPTG), an effect attributed to efflux [79]. YnfM, MdtD, YfcJ, and YdhC are uncharacterized MFS transporters. YhhS is also an uncharacterized MFS transporter that has been implicated glyphosate resistance [80].

Interestingly, we did not observe any effect with the *ydeA* deletion strain, though we did observe an effect when overexpressing it as discussed below. As previously mentioned, this gene is known to encode an arabinose efflux transporter and Carolé and coworkers have shown that disrupting the *ydeA* gene increases intracellular arabinose concentrations. The lack of agreement between their results and ours may be due to the fact that they employed a strain lacking the arabinose metabolic genes, which dramatically increases intracellular arabinose concentrations as shown below, whereas our strain background was wild-type MG1655.

Effect of over-expressing candidate sugar efflux transporters on arabinose metabolism. We also over-expressed these candidate sugar efflux transporters. If these genes encode arabinose efflux transporters, then over-expressing them should reduce intracellular arabinose concentrations. To express these genes, they were cloned on high-copy plasmids under the control of the strong, IPTG-inducible *trc* promoter. Once again, transcriptional fusions to the *araBAD* promoter were employed as indirect measures of intracellular arabinose and the results were recorded in terms of governing parameters for a Michaelis-Menten dose-response curve (Table 2.8).

Over-expression of six genes was found to reduce intracellular arabinose concentrations: *setC*, *ydeA*, *mhpt*, *ybdA*, *ydeE*, and *kgtP*. A representative dose-response curve for *ybdA* is given in

Figure 2.2. Among the five, only *setC* and *ydeA* have been implicated in sugar efflux as discussed above. YbdA, also known as EntS, has been implicated in enterobacterin transport [26] and is also involved in the resistance to multiple chemical stresses [81]. YdeE is an uncharacterized MFS transporter involved in dipeptide transport and resistance [82]. KgtP, interestingly, is known to be α-ketoglutarate transporter [83].

When comparing these results with our deletion results, only *setC* gave a consistent result where loss of the gene increases intracellular arabinose and overexpression decreases it. Over-expression of *yfcJ*, *yhhS*, *emrD*, and *ydhC* had no effect on intracellular arabinose concentrations. Three genes that gave a deletion phenotype were toxic to the cell when over-expressed: *cmr*, *ynfM*, and *mdtD*.

Overall, 12 genes were toxic to *E. coli* when over-expressed. Of the 31 genes, the set of transporters that resulted in cell death when over expressed no matter what IPTG concentration was used included: *ycaD*, *ynfM*, *mdtD*, *bcr*, *ygcS* and *yjhB*. Interestingly, the set of strains overexpressing these genes caused growth defects irrespective of whether the substrate was arabinose or xylose. To test that there is no other factor in the strain itself that is causing cell death the experiments were repeated with no IPTG added. In accordance with intuitive results, the curves imitate the control curve and the optical density of the cells reaches normal levels (0.4-0.6). These results are shown in Figures 2.3 and 2.4.

To limit toxicity, the over-expression experiments were performed at two concentrations of the IPTG inducer. Both *setC* and *ydeE* were toxic to cell when expressed using the higher IPTG concentration; only at the lower concentration did we observe an effect. Conversely, we only observed an effect with *ydeA* and *kgtP* when they were induced at higher IPTG concentration. Only *ybdA* and *mhpT* decrease intracellular arabinose concentrations at both induction levels.

Xylose is not a target for any of the candidate efflux transporters. The candidate efflux transporters were also tested to determine if they have an effect on xylose metabolism. This is important so that we can find out whether all pentose sugars are substrates for the efflux transporters, or if the efflux transporters are specific to arabinose only (and possibly other sugars not tested in this work), but not xylose. Also, if we can find a set of transporters that efflux both arabinose and xylose, it would significantly reduce the amount of genetic engineering required to reach our goal. A similar approach was employed as before except that we used transcriptional fusions of the *xylA* promoter to Venus. The *xylA* promoter is positively regulated by XylR in response to xylose [55] and can equivalently be used to determine relative intracellular xylose concentration. Otherwise, the experiments were performed in an identical manner to the arabinose experiments. Despite the similarities between xylose and arabinose, we did not observe any change in intracellular xylose concentrations as compared to the wild-type control when any of candidate genes were deleted or overexpressed (Figures 2.5 and 2.6).

Arabinose is a substrate for most efflux transporters. The arabinose metabolic intermediate, ribulose-5-phosphate, is toxic to *E. coli* [84]. This may provide one possible explanation as to why arabinose may be effluxed from cells. It also suggests that arabinose itself may not be the target for efflux but rather an intermediate of arabinose metabolism. To test this hypothesis, we blocked different steps in arabinose metabolism by deleting the cognate gene and then determined how the mutation would affect our results concerning arabinose efflux. If arabinose is not the target for these transporters and the downstream metabolite is instead, then we expect that blocking metabolism will reduce the effect of these transporters. We tested three mutants: Δ*araA*, Δ*araB*, and Δ*araBAD*. We did not test a Δ*araD* mutant as arabinose is toxic in this strain. All strains lacking in *araC* suffered growth degradation when arabinose was used as a substrate for the growth cultures. AraC positively regulates the transcription of the arabinose of the four operons in the arabinose regulon in the presence of

arabinose. Therefore, in the absence of *araC*, the metabolic genes are not transcribed and the cells are not able to consume the substrate arabinose and hence suffer growth defects.

Blocking any step in arabinose metabolism significantly increased the concentration of intracellular arabinose relative to the wild-type control. This is expected, as arabinose is no longer being metabolized in any of these strains. However, we did not observe any further increase when *setC, mhpT, cmr, ynfM, mdtD, yfcJ, yhhS,* and *emrD* were deleted as well (Figure 2.7). While these results suggest that these transporters may target ribulose-5-phosphate, a simpler explanation is that arabinose concentrations are already high in these cells and eliminating efflux does not further increase arabinose concentration.

The effect of overexpressing of *setC, ydeE,* and *ydeA* was most pronounced in a Δ*araA* mutant, indicating that the three target arabinose (Figure 3.8). The data for *mhpT* and *ybdA* was more equivocal though again it indicates both target arabinose. Interestingly, we only observed a reduction in relative arabinose concentrations only when *kgtP* was overexpressed in a Δ*araB* mutant or the wild type. These results suggest that ribulose may be the target for KgtP.

Many arabinose efflux transporters are TolC dependent. In general, many efflux transporters are TolC dependent. Martin and Rosner previously found that the *mar*/*sox*/*rob* regulon was activated in a *tolC* mutant [85]. They proposed that this activation may be due to the accumulation of intracellular metabolites due to the loss of efflux. Chubiz and Rao subsequently demonstrated that this activation could partially be attributed to the accumulation of benzoic acid intermediates in the chorismate superpathway [86]. Based on these results, we tested whether any of the identified arabinose efflux transporters were TolC dependent.

In an otherwise wild-type background, deleting *tolC* was found to increase intracellular arabinose concentrations (Figure 2.9A). The magnitude of this increase was similar to that observed with the efflux transporter knockouts. When both sets of mutants were combined, we

observed that the intracellular arabinose concentrations went back to wild-type levels in all cases except with EmrD. These results do not provide any useful information about the link between the efflux transporters and TolC as deletion of both sets of genes indicated no intracellular change in arabinose concentration.

We also tested the effect of overexpressing efflux transporters in a Δ*tolC* mutant (Figure 2.9B). When the efflux transporters were over-expressed in the Δ*tolC* mutant, arabinose concentrations were reduced in all cases except with KgtP and YbdA. These results suggest that these two transporters are linked to TolC. We also found that the effect of over-expressing SetC was somewhat reduced in a Δ*tolC* mutant. Interestingly, we did not observe any increase in arabinose concentrations when *tolC* was deleted. This effect is entirely due to pTrc99A (data not shown). This plasmid was present in both control strains used for results shown in Figure 2.9B. Why the empty plasmid has an effect is not known.

Expression of efflux transporters is induced by arabinose. The last question explored was whether arabinose induced the expression of these transporters. To answer this question, we fused the promoters for these transporters to Venus and then tested whether they were activated by arabinose. Of the thirteen transporters, we were able to construct functional promoter fusions to all but four. Our transcriptional fusions to the *mhpT*, *yfcJ*, and *mdtD* promoters were not functional despite repeated attempts. In addition, the fusions to the *setC* and *ydhC* promoters were very weak and not included in our results as a consequence.

Of the eight remaining functional promoters, only the *ydeA* and *yhhS* promoters were activated at significant levels (Figure 2.10). The remaining six promoters were also activated by arabinose though the effect is less than two fold. How these promoters, in particular for the *ydeA* and *yhhS* genes, are being activated by arabinose is not known. Interesting, *ydeA* is strongly induced by

arabinose yet deleting this gene does not affect intracellular arabinose concentrations. Only when over-expressed do we observe an effect.

2.4 Discussion

We identified 31 candidate sugar efflux transporters in *E. coli* based on sequence homology to known sugar efflux transporters and then tested to see whether they target arabinose and xylose. The results of this study are summarized in Table 2.9. Thirteen putative arabinose efflux transporters were identified based on their ability to alter intracellular arabinose concentration. Interestingly, no xylose transporters were identified. Eight of the thirteen arabinose efflux transporters were found to increase intracellular arabinose concentrations when deleted and six were found to decrease concentrations when overexpressed. Only one transporter, SetC, was found to yield reciprocal results when deleted or overexpressed. Among the thirteen putative transporters, only YdeA has previously been shown to be a sugar efflux transporter. A sequence homology comparison of each of the identified efflux transporters to the *ydeA* gene was carried out using LALIGN (Table 2.10). A few of the transporters have local homology in multiple regions but interestingly most of the homology occurs in the region of ~750-900 base pairs into the *ydeA* gene. Two of the efflux transporters, *ydhC* and *ydeE*, did not show significant homology at a particular location, but some homology was still noted at the 750-900 base pairs region.

We cannot definitely say whether these putative transporters in fact efflux arabinose as we did not directly measure arabinose transport. All that we can say with certainty is that these 13 transporters inhibit the accumulation of arabinose with efflux being a likely but not the only possible mechanism. Alternate mechanisms include the inhibition of the arabinose uptake and the efflux of other compounds that affect arabinose metabolism or possibly stimulate other efflux transporters.

A number of the identified efflux transporters have previously been shown to transport other compounds both in and out of the cell. Cmr, for example, is known to efflux a number of chemically unrelated compounds [87], so it is not entirely implausible that it effluxes arabinose as well. This is consistent with multidrug efflux transporters in general, which are known to have broad substrate specificities. What is surprising is that transporters such as KgtP, a known α-ketoglutarate transporter [83], also inhibit arabinose uptake and perhaps efflux it as well. Clearly further research is required to identify the exact mechanisms for how these transporters affect arabinose uptake and efflux.

Why would *E. coli* want to efflux arabinose from the cell? One possibility is to limit the accumulation of toxic sugar phosphates, such as the arabinose metabolic intermediate L-ribulose-5-phosphate, within the cell. If the flux of arabinose is greater than the pentose phosphate pathway can accommodate, then L-ribulose-5-phosphate may accumulate within the cell due to an effective roadblock within the pathway. Under such a scenario, arabinose efflux transporters would provide a relief valve for the cell. A second possibility is to limit the accumulation of methylglyoxal, another toxic compound to *E. coli* [88]. Methylglyoxal synthase is thought to provide a relief value in glycolysis, preventing the buildup of dihydroxyacetone phosphate when inorganic phosphate is limiting. If the flux of arabinose is too high, it may overwhelm the glycolytic branch of the pathway leading to a buildup of dihydroxyacetone phosphate. Consistent with this model, the addition of cAMP to cells grown on arabinose or xylose, which increases the uptake of these two sugars, produces excess methylglyoxal, arresting cell growth [89]. Interestingly, the effect is less severe with arabinose, consistent with our result showing a lack of xylose efflux transporters. In addition, the overexpression of methylglyoxal synthase is lethal to cells grown on arabinose and xylose but not on glucose. Unlike arabinose and xylose dissimilation, glycolysis is linked to glucose transport through the phosphenolpyruvate transferase system [90]. In the case of arabinose and xylose dissimilation,

transport and metabolism are uncoupled. Arabinose efflux may provide a relief value to accommodate this uncoupling. Why there are not similar relief values for xylose is not known.

Other outstanding questions concern the regulation of these transporters. Arabinose was found to induce the expression of many efflux transporters, where the effects were most pronounced for the YdeA and YhhS transporters. We and others have previously speculated that the multidrug antibiotic resistance (MDR) mechanisms in bacteria evolved in part as a mechanism to prevent the buildup of toxic metabolites. In fact, we found that 2,3-dihydroxybenzoate, an intermediate in enterobactin biosynthesis, directly binds to and inactivates MarR, the key regulator of the *marRAB* operon involved in antibiotic resistance. Whether arabinose and other sugars activate the MDR regulators is not known though it is an intriguing hypothesis nonetheless.

We conclude by noting that sugar efflux may be a novel target for metabolic engineering. Clearly, one way to increase metabolic flux for the production of biofuels and other valuable chemicals is to remove sugar efflux. As previously mentioned, the present work was motivated by looking at novel targets for optimizing pentose metabolism in *E. coli*. Whether sugar efflux transporters provide a viable target for metabolic engineering is not known, but it does provide a totally unexplored area that clearly merits further investigation.

2.5 Figures and tables

Table 2.1. List of candidate efflux transporter genes.

Gene	Annotation
setA	broad specificity sugar efflux system
setB	lactose/glucose efflux system
setC	predicted sugar efflux system
ydeA	sugar efflux transporter
mhpT	putative 3-hydroxyphenylpropionic acid transporter
yajR	putative transporter
ybdA (entS)	predicted transporter
cmr (mdfA)	multidrug efflux system protein
ycaD	putative MFS family transporter
yceL	orf, hypothetical protein
ydeE	predicted transporter
ynfM	predicted transporter
ydhP	predicted transporter
mdtG (yceE)	predicted drug efflux system
yebQ	putative transporter
shiA	shikimate transporter
mdtD (yegB)	multidrug efflux system protein
bcr	bicyclomycin/multidrug efflux system
yfcJ	predicted transporter
kgtP	alpha-ketoglutarate transporter
ygcS	putative transporter
galP	D-galactose transporter
nanT	sialic acid transporter
yhhS	putative transporter

Table 2.1 (cont.)

Gene	Annotation
nepI (yicM)	predicted transporter
emrD	2-module integral membrane pump; multidrug resistance
mdtL (yidY)	multidrug efflux system protein
proP	proline/glycine betaine transporter
yjhB	putative transporter
yjiO (mdtM)	multidrug efflux system protein
ydhC	predicted transporter

Table 2.2. Bacterial strains used in this work.

Strain	Genotype or relevant characteristics[a]	Source or reference[b,c]
MG1655	F^- λ^- *ilvG rph-1*	CGSC #7740
DH5α	*phi-80d lac*Δm15 *enda1 recA1 hsdR17 supE44 thh*-1 *gyrA96 relA*Δ*lacU169*	New England Biolabs
CR400	Δ*araA*::*kan*	56
CR401	Δ*araB*::*kan*	56
CR404	Δ*araBAD*::*kan*	56
CR701	Δ*tolC*::*cat*	86
CR702	Δ*tolC*::FRT	86
CR1100	Δ*setA*:: FRT *kan* FRT (77621-78799)	
CR1101	Δ*setB*:: FRT *kan* FRT (2261885-2263066)	
CR1102	Δ*setC*:: FRT *kan* FRT (3834976-3836160)	
CR1103	Δ*ydeA*:: FRT *kan* FRT (1615052-1616242)	
CR1104	Δ*mhpT*:: FRT *kan* FRT (374683-375894)	
CR1105	Δ*yajR*:: FRT *kan* FRT (444526-445890)	
CR1106	Δ*ybdA*::FRT *kan* FRT (621523-622773)	
CR1107	Δ*cmr*::FRT *kan* FRT (882896-884128)	
CR1108	Δ*ycaD*::FRT *kan* FRT (945094-946242)	
CR1109	Δ*yceL*::FRT *kan* FRT (1123341-1124549)	
CR1110	Δ*ydeE*::FRT *kan* FRT (1619356-1620543)	
CR1111	Δ*ynfM*::FRT *kan* FRT (1667723-1668976)	
CR1112	Δ*ydhP*::FRT *kan* FRT (1734145-1735314)	
CR1113	Δ*mdtG*::FRT *kan* FRT (1113487-1114713)	
CR1114	Δ*yebQ*::FRT *kan* FRT (1908300-1909673)	
CR1115	Δ*shiA*::FRT *kan* FRT (2051667-2052983)	
CR1116	Δ*mdtD*::FRT *kan* FRT (2159488-2160903)	
CR1117	Δ*bcr*::FRT *kan* FRT (2276592-2277782)	

Table 2.2 (cont.)

Strain	Genotype or relevant characteristics[a]	Source or reference[b,c]
CR1118	Δ*yfcJ*::FRT *kan* FRT (2436964-2438142)	
CR1119	Δ*kgtP*::FRT *kan* FRT (2722470-2723768)	
CR1120	Δ*ygcS*::FRT *kan* FRT (2894555-2895892)	
CR1121	Δ*galP*::FRT *kan* FRT (3086306-3087700)	
CR1122	Δ*nanT*::FRT *kan* FRT (3369106-3370596)	
CR1123	Δ*yhhS*::FRT *kan* FRT (3608539-3609756)	
CR1124	Δ*nepI*::FRT *kan* FRT (3838572-3839762)	
CR1125	Δ*emrD*::FRT *kan* FRT (3851945-3853129)	
CR1126	Δ*mdtL*::FRT *kan* FRT (3889638-3890813)	
CR1127	Δ*proP*::FRT *kan* FRT (4328525-4330027)	
CR1128	Δ*yjhB*::FRT *kan* FRT (4502081-4503298)	
CR1129	Δ*yjiO*::FRT *kan* FRT (4565310-4566542)	
CR1130	Δ*ydhC*::FRT *kan* FRT (1737935-1739146)	
CR1131	Δ*setA*::FRT	
CR1132	Δ*setB*::FRT	
CR1133	Δ*setC*::FRT	
CR1134	Δ*ydeA*::FRT	
CR1135	Δ*mhpT*::FRT	
CR1136	Δ*yajR*::FRT	
CR1137	Δ*ybdA*::FRT	
CR1138	Δ*cmr*::FRT	
CR1139	Δ*ycaD*::FRT	
CR1140	Δ*yceL*::FRT	
CR1141	Δ*ydeE*::FRT	
CR1142	Δ*ynfM*::FRT	
CR1143	Δ*ydhP*::FRT	

Table 2.2 (cont.)

Strain	Genotype or relevant characteristics[a]	Source or reference[b,c]
CR1144	Δ*mdtG*::FRT	
CR1145	Δ*yebQ*::FRT	
CR1146	Δ*shiA*::FRT	
CR1147	Δ*mdtD*::FRT	
CR1148	Δ*bcr*::FRT	
CR1149	Δ*yfcJ*::FRT	
CR1150	Δ*kgtP*::FRT	
CR1151	Δ*ygcS*::FRT	
CR1152	Δ*galP*::FRT	
CR1153	Δ*nanT*::FRT	
CR1154	Δ*yhhS*::FRT	
CR1155	Δ*nepI*::FRT	
CR1156	Δ*emrD*::FRT	
CR1157	Δ*mdtL*::FRT	
CR1158	Δ*proP*::FRT	
CR1159	Δ*yjhB*::FRT	
CR1160	Δ*yjiO*::FRT	
CR1161	Δ*ydhC*::FRT	
CR1162	Δ*setC*::FRT Δ*araA*::FRT	
CR1163	Δ*cmr*::FRT Δ*araA*::FRT	
CR1164	Δ*ynfM*::FRT Δ*araA*::FRT	
CR1165	Δ*ydhP*::FRT Δ*araA*::FRT	
CR1166	Δ*mdtD*::FRT Δ*araA*::FRT	
CR1167	Δ*yfcJ*::FRT Δ*araA*::FRT	
CR1168	Δ*yhhS*::FRT Δ*araA*::FRT	

Table 2.2 (cont.)

Strain	Genotype or relevant characteristics[a]	Source or reference[b,c]
CR1169	Δ*emrD*::FRT Δ*araA*::FRT	
CR1170	Δ*ydhC*::FRT Δ*araA*::FRT	
CR1171	Δ*setC*::FRT Δ*araB*::FRT	
CR1172	Δ*cmr*::FRT Δ*araB*::FRT	
CR1173	Δ*ynfM*::FRT Δ*araB*::FRT	
CR1174	Δ*mdtD*::FRT Δ*araB*::FRT	
CR1175	Δ*yfcJ*::FRT Δ*araB*::FRT	
CR1176	Δ*yhhS*::FRT Δ*araB*::FRT	
CR1177	Δ*emrD*::FRT Δ*araB*::FRT	
CR1178	Δ*ydhC*::FRT Δ*araB*::FRT	
CR1179	Δ*setC*::FRT Δ*araBAD*::FRT	
CR1180	Δ*cmr*::FRT Δ*araBAD*::FRT	
CR1181	Δ*ynfM*::FRT Δ*araBAD*::FRT	
CR1182	Δ*mdtD*::FRT Δ*araBAD*::FRT	
CR1183	Δ*yfcJ*::FRT Δ*araBAD*::FRT	
CR1184	Δ*yhhS*::FRT Δ*araBAD*::FRT	
CR1185	Δ*emrD*::FRT Δ*araBAD*::FRT	
CR1186	Δ*ydhC*::FRT Δ*araBAD*::FRT	

Table 2.3. Plasmids used in this work.

Plasmids	Genotype or relevant characteristics[a]	Source or reference[b,c]
pKD46	*bla* P_{BAD} *gam bet exo* pSC101 *ori(ts)*	73
pCP20	*bla cat cI*857 λP_R'-*flp* pSC101 *ori(ts)*	91
pKD3	*bla FRT cm FRT oriR6K*	73
pKD4	*bla FRT kan FRT oriR6K*	73
pPROBE *venus*	*kan venus ori* p15a	
pTrc99A	*amp* P_{trc} *ori* (pBR322) lacI^q	76
P_{araB}-*venus*	*kan* P_{araB}-*venus ori* p15a	
P_{xylA}-*venus*	*kan* P_{xylA}-*venus ori* p15a	
pKK001 (pSetA)	*amp* P_{trc} *setA ori* (pBR322) lacI^q	
pKK002 (pSetB)	*amp* P_{trc} *setB ori* (pBR322) lacI^q	
pKK003 (pSetC)	*amp* P_{trc} *setC ori* (pBR322) lacI^q	
pKK004 (pYdeA)	*amp* P_{trc} *setA ori* (pBR322) lacI^q	
pKK005 (pMhpT)	*amp* P_{trc} *mhpT ori* (pBR322) lacI^q	
pKK006 (pYajR)	*amp* P_{trc} *yajR ori* (pBR322) lacI^q	
pKK007 (pYbdA)	*amp* P_{trc} *ybdA ori* (pBR322) lacI^q	
pKK008 (pCmr)	*amp* P_{trc} *cmr ori* (pBR322) lacI^q	
pKK009 (pYcaD)	*amp* P_{trc} *ycaD ori* (pBR322) lacI^q	
pKK010 (pYceL)	*amp* P_{trc} *yceL ori* (pBR322) lacI^q	
pKK011 (pYdeE)	*amp* P_{trc} *ydeE ori* (pBR322) lacI^q	
pKK012 (pYnfM)	*amp* P_{trc} *ynfM ori* (pBR322) lacI^q	
pKK013 (pYdhP)	*amp* P_{trc} *ydhP ori* (pBR322) lacI^q	
pKK014 (pMdtG)	*amp* P_{trc} *mdtG ori* (pBR322) lacI^q	
pKK015 (pYebQ)	*amp* P_{trc} *yebQ ori* (pBR322) lacI^q	
pKK016 (pShiA)	*amp* P_{trc} *shiA ori* (pBR322) lacI^q	

Table 2.3 (cont.)

Plasmids	Genotype or relevant characteristics[a]	Source or reference[b,c]
pKK017 (pMdtD)	*amp* P_{trc} *mdtD ori* (pBR322) $lacI^q$	
pKK018 (pBcr)	*amp* P_{trc} *bcr ori* (pBR322) $lacI^q$	
pKK019 (pYfcJ)	*amp* P_{trc} *yfcJ ori* (pBR322) $lacI^q$	
pKK020 (pKgtP)	*amp* P_{trc} *kgtP ori* (pBR322) $lacI^q$	
pKK021 (pYgcS)	*amp* P_{trc} *ygcS ori* (pBR322) $lacI^q$	
pKK022 (pGalP)	*amp* P_{trc} *galP ori* (pBR322) $lacI^q$	
pKK023 (pNanT)	*amp* P_{trc} *nanT ori* (pBR322) $lacI^q$	
pKK024 (pYhhS)	*amp* P_{trc} *yhhS ori* (pBR322) $lacI^q$	
pKK025 (pNepI)	*amp* P_{trc} *nepI ori* (pBR322) $lacI^q$	
pKK026 (pEmrD)	*amp* P_{trc} *emrD ori* (pBR322) $lacI^q$	
pKK027 (pMdtL)	*amp* P_{trc} *mdtL ori* (pBR322) $lacI^q$	
pKK028 (pProp)	*amp* P_{trc} *proP ori* (pBR322) $lacI^q$	
pKK029 (pYjhB)	*amp* P_{trc} *yjhB ori* (pBR322) $lacI^q$	
pKK031 (pYjiO)	*amp* P_{trc} *yjiO ori* (pBR322) $lacI^q$	
pKK032 (pYdhC)	*amp* P_{trc} *ydhC ori* (pBR322) $lacI^q$	
pKK033 (P_{setC}-*venus*)	*kan* P_{setC}-*venus ori* p15a	
pKK034 (P_{ydeA}-*venus*)	*kan* P_{ydeA}-*venus ori* p15a	
pKK035 (P_{mhpT}-*venus*)	*kan* P_{mhpT}-*venus ori* p15a	
pKK036 (P_{ybdA}-*venus*)	*kan* P_{ybdA}-*venus ori* p15a	
pKK037 (P_{ydeE}-*venus*)	*kan* P_{ydeE}-*venus ori* p15a	
pKK038 (P_{kgtP}-*venus*)	*kan* P_{kgtP}-*venus ori* p15a	
pKK039 (P_{cmr}-*venus*)	*kan* P_{cmr}-*venus ori* p15a	
pKK040 (P_{ynfM}-*venus*)	*kan* P_{ynfM}-*venus ori* p15a	
pKK041 (P_{mdtD}-*venus*)	*kan* P_{mdtD}-*venus ori* p15a	

Table 2.3 (cont.)

Plasmids	Genotype or relevant characteristics[a]	Source or reference[b,c]
pKK042 (P_{yfcJ}-*venus*)	*kan* P_{yfcJ}-*venus ori* p15a	
pKK043 (P_{yhhS}-*venus*)	*kan* P_{yhhS}-*venus ori* p15a	
pKK044 (P_{emrD}-*venus*)	*kan* P_{emrD}-*venus ori* p15a	
pKK045 (P_{ydhC}-*venus*)	*kan* P_{ydhC}-*venus ori* p15a	

a. All strains are isogenic derivatives of *E. coli* K-12 strain MG1655.
b. All strains and plasmids are from this work unless otherwise noted.
c. *E. coli* Genetic Stock Center, CGSC, Yale University.

Table 2.4. Primers used for gene deletions

Efflux Transporter Gene	Primer	Sequence
setA	KK016f	ATG ATC TGG ATA ATG ACG ATG GCT CGC CGT ATG AAC GGT GGT GTA GGC TGG AGC TGC TTC
	KK016r	TCA AAC GTC TTT AAC CTT TGC GGT TAA AAA TAA TGC GAC ACA TAT GAA TAT CCT CCT TAG
setC	KK018f	ATG CAA AAA ACG GCT ACC ACT CCA TCA AAA ATA CTT GAT CGT GTA GGC TGG AGC TGC TTC
	KK018r	CTA AAT ATC TTT AAT AAA CAG CAG GCA AAT CAT CGC AAT ACA TAT GAA TAT CCT CCT TAG
ydeA	TD139f	TCA TTT CAT CAC TTT TTC GCA ACT CAC CCG ATA ATC TGT TGT GTA GGC TGG AGC TGC TTC
	TD139r	ACT ATT GCG TCT GTT CTT CGA GTG CAC TGG CCA GCG GCG CA TAT GAA TAT CCT CCT TAG
mhpT	KK030f	ATG TCG ACT CGT ACC CCT TCA TCA TCT TCA TCC CGC CTG AGT GTA GGC TGG AGC TGC TTC
	KK030r	TCA GGC ATC GGC GCA CGG CTG TAT TCG TGA TCT CCG GCT CCA TAT GAA TAT CCT CCT TAG
yajR	KK032f	ATG AAC GAT TAT AAA ATG ACG CCA GGT GAG AGG CGC GCG AGT GTA GGC TGG AGC TGC TTC
	KK032r	TTA TGC CTG ACG AAT TGC CTG TTC TAT CTC AAA GCG ATT CCA TAT GAA TAT CCT CCT TAG
ybdA	KK034f	ATG AAT AAA CAA TCC TGG CTG CTT AAC CTC AGC CTG TTG AGT GTA GGC TGG AGC TGC TTC
	KK034r	TTA ACT GTC GGA CGC TGT CAC CTG CGG CGG CGT CTG GCG ACA TAT GAA TAT CCT CCT TAG
cmr	KK036f	ATG CAA AAT AAA TTA GCT TCC GGT GCC AGG CTT GGA CGT CGT GTA GGC TGG AGC TGC TTC
	KK036r	TTA CCC TTC GTG AGA ATT TCC CAT CTG TTT ATC TTT TAA ACA TAT GAA TAT CCT CCT TAG
ycaD	KK038f	ATG TCC ACG TAT ACC CAG CCT GTC ATG CTT TTG CTG TCT GGT GTA GGC TGG AGC TGC TTC
	KK038r	TTA CAC GTG AGC AAC GGG TTT CGG CGT ATG ACC GGC GTT GCA TAT GAA TAT CCT CCT TAG
yceL	KK053f	ATG TCC CGC GTG TCG CAG GCG AGG AAC CTG GGT AAA TAT TGT GTA GGC TGG AGC TGC TTC
	KK053r	TCA GGC GTC GCG TTC AAG CAA ACG ACG CGC GGC GCG TTT CCA TAT GAA TAT CCT CCT TAG
ydeE	KK055f	ATG AAC TTA TCC CTA CGA CGC TCT ACC AGC GCC CTT CTT GGT GTA GGC TGG AGC TGC TTC
	KK055r	TCA ACA AAG CGC GGG CTG CCC CCA CGG TCT TGC TCG AAT CCA TAT GAA TAT CCT CCT TAG
ynfM	KK057f	GTG AGC CGT ACT ACA ACT GTT GAT GGC GCT CCG GCA AGC GGT GTA GGC TGG AGC TGC TTC
	KK057r	TCA GGC GTG CAG ACG ACG ATG CAA ACG CGT CCC GAC CAG CCA TAT GAA TAT CCT CCT TAG
ydhP	KK059f	ATG AAA ATT AAC TAT CCG TTG CTG GCG CTG GCG ATT GGC GGT GTA GGC TGG AGC TGC TTC
	KK059r	TTA GCT GTT AGC AAC GCA AAC TGT TTC AGG TTG TTT TCT GCA TAT GAA TAT CCT CCT TAG

Table 2.4 (cont.)

mdtG	KK061f	ATG TCA CCC TGT GAA AAT GAC ACC CCT ATA AAC TGG AAA CGT GTA GGC TGG AGC TGC TTC
	KK061r	TCA GTT CGA TAC CTG GGG TAT TCG ACG ACG ACG TAG ACT GCA TAT GAA TAT CCT CCT TAG
yebQ	KK063f	ATG CCA AAA GTT CAG GCC GAC GGC CTG CCA TTG CCC CAG CGT GTA GGC TGG AGC TGC TTC
	KK063r	TTA TGC CCT GGA TCG TGG CTG AGT GAT ACG TAA ACC ACT GCA TAT GAA TAT CCT CCT TAG
shiA	KK065f	ATG GAC TCC ACG CTC ATC TCC ACT CGT CCC GAT GAA GGG AGT GTA GGC TGG AGC TGC TTC
	KK065r	TCA AGC GCG TTG ACT GTC TTT CAT CAA CAA AGC GGT CAT TCA TAT GAA TAT CCT CCT TAG
mdtD	KK067f	ATG ACA GAT CTT CCC GAC AGC ACC CGT TGG CAA TTG TGG AGT GTA GGC TGG AGC TGC TTC
	KK067r	TCA TTG CGC GCT CCT TTT TCG CCG CGA AAT AGC TAC ATT TCA TAT GAA TAT CCT CCT TAG
setB	KK069f	ATG CAT AAC TCC CCC GCA GTC TCC AGC GCG AAA TCG TTT GGT GTA GGC TGG AGC TGC TTC
	KK069r	TTA AAC ATC TTT AAT CCG CAG TAA GCA AAA CAG AGT GGC GCA TAT GAA TAT CCT CCT TAG
bcr	KK071f	GTG ACC ACC CGA CAG CAT TCG TCG TTT GCT ATT GTT TTT AGT GTA GGC TGG AGC TGC TTC
	KK071r	TCA CCG TTT TTT CGG CCG ACT GGC GTA CAG ACA GAA GAG ACA TAT GAA TAT CCT CCT TAG
yfcJ	KK073f	ATG ACT GCT GTA AGC CAA ACC GAA ACA CGA TCT TCT GCC AGT GTA GGC TGG AGC TGC TTC
	KK073r	TTA ACC CCG ACG AAA TGA CAG TAT CGT GAC AAT AAT ACC CCA TAT GAA TAT CCT CCT TAG
kgtP	KK075f	ATG GCT GAA AGT ACT GTA ACG GCA GAC AGC AAA CTG ACA AGT GTA GGC TGG AGC TGC TTC
	KK075r	CTA AAG ACG CAT CCC CTT CCC TTT GCG ATG TAG CAT CAA ACA TAT GAA TAT CCT CCT TAG
ygcS	KK077f	ATG AAC ACT TCA CCG GTG CGA ATG GAT GAT TTA CCG CTT AGT GTA GGC TGG AGC TGC TTC
	KK077r	TTA AAC GCT AAC AGA ATG TTC ATT CGC ACC TCC TAC ATTT CA TAT GAA TAT CCT CCT TAG
galP	KK081f	ATG CCT GAC GCT AAA AAA CAG GGG CGG TCA AAC AAG GCA AGT GTA GGC TGG AGC TGC TTC
	KK081r	TTA ATC GTG AGC GCC TAT TTC GCG CAG TTT ACG ACC TTT CCA TAT GAA TAT CCT CCT TAG
nanT	KK083f	ATG AGT ACT ACA ACC CAG AAT ATC CCG TGG TAT CGC CAT CGT GTA GGC TGG AGC TGC TTC
	KK083r	TTA ACT TTT GGT TTT GAC TAA ATC GTT TTT GGC GCT GCC ACA TAT GAA TAT CCT CCT TAG
yhhS	KK085f	ATG CCC GAA CCC GTA GCC GAA CCC GCG CTA AAC GGA TTG CGT GTA GGC TGG AGC TGC TTC
	KK085r	TTA AGA TGA TGA GGC GGC CTC AGG GAC GTG TTC CGG AGG CCA TAT GAA TAT CCT CCT TAG
nepI	KK087f	ATG AGT GAA TTT ATT GCC GAA AAC CGC GGC GCG GAT GCC AGT GTA GGC TGG AGC TGC TTC
	KK087r	TCA GGA TTT CTT CAT TTT CAC CTT TGC AGT AAC CAA CAA TCA TAT GAA TAT CCT CCT TAG

Table 2.4 (cont.)

emrD	KK089f	ATG AAA AGG CAA AGA AAC GTC AAT TTG TTA TTG ATG TTG GGT GTA GGC TGG AGC TGC TTC
	KK089r	TTA AAC GGG CTG CCC CTG ATG CGA CAT CCG CGT CGC CAG CCA TAT GAA TAT CCT CCT TAG
mdtL	KK091f	ATG TCC CGC TTT TTG ATT TGT AGT TTT GCC CTG GTT TTA CGT GTA GGC TGG AGC TGC TTC
	KK091r	TCA AGC GTG GTG ATG GAT TTC TTC ATG AGC GGC AAC GGG GCA TAT GAA TAT CCT CCT TAG
proP	KK093f	ATG CTG AAA AGG AAA AAA GTA AAA CCG ATT ACC CTT CGT GGT GTA GGC TGG AGC TGC TTC
	KK093r	TTA TTC ATC AAT TCG CGG ATG TTG CTG CAC CAG GCG GGT ACA TAT GAA TAT CCT CCT TAG
yjhB	KK095f	ATG GCA ACA GCA TGG TAT AAA CAA GTT AAT CCA CCA CAA CGT GTA GGC TGG AGC TGC TTC
	KK095r	TCA TTT AGC CAC GGA TAG TTT ATA AAT TTT ACC TGG AAT ACA TAT GAA TAT CCT CCT TAG
yjiO	KK097f	ATG CCA CGT TTT TTT ACC CGC CAT GCC GCC ACG CTG TTT TGT GTA GGC TGG AGC TGC TTC
	KK097r	TCA CTG CTC CTC CAC TAG CTC GGC TGC CTG ATG CTG GCG CCA TAT GAA TAT CCT CCT TAG
ydhC	KK099f	ATG CAA CCT GGG AAA AGA TTT TTA GTC TGG CTG GCG GGT TGT GTA GGC TGG AGC TGC TTC
	KK099r	TCA GTG TGA TTC GCT ATG AGC GAC TTC GGC ATT GCC ATG ACA TAT GAA TAT CCT CCT TAG

Table 2.5. Knockout check primers

Efflux Transporter Gene	Primer	Sequence
setA	KK017f	CGC CAG CAG ATT ATA CCT GC
	KK017r	GCT TTC CGC TAT CGT CCA CG
setC	KK019f	TTC GCC CGG TGA TTC AGA AG
	KK019r	TGG CTC TGT TTG CGC CAC GC
ydeA	TD140f	AAT TAT CAA ACA TAT GGC GC
	TD140r	TGT GCT CGC TTC CCG GCC GG
mhpT	KK031f	CAC CTG CAC TCG CAA CGA GG
	KK031r	TCG GCG CAC GGC TGT ATT CG
yajR	KK033f	GCA GCA GGC AAT TCG CCG CG
	KK033r	CAC TGC GGC CAG CAT TGC G C
ybdA	KK035f	ATG TGC AGC CTA AGA ATA GG
	KK035r	CGC CTA AAA TCC ACA TCG GC
cmr	KK037f	CTG CGC GAT TAT TAT TGG CG
	KK037r	TGC AGC ATT CCC ATC GCG GC
ycaD	KK039f	TTG GCG GTG GCA TGA TGC GC
	KK039r	CTT TCT CGC AAG CCC ATG CC
yceL	KK054f	AAC TCT CTA TCA AAG CGA CG
	KK054r	GGC AGT AAG GGC AGT GAT CG
ydeE	KK056f	TAG TCG CCG GTT ACG ACG GC
	KK056r	GGT GTG TTT CCT CGC TTA CC
ynfM	KK058f	ACA AAG TAA CGC AGA TGA CG
	KK058r	CTG CGG GAA TTG GCT GGC GC
ydhP	KK060f	TCT CCT TCC TGT TTG TTA GC
	KK060r	CGA TGT AGG TCG GAT AAG GC
mdtG	KK062f	AGC TCT CTG GAT TGC GCC CC
	KK062r	TTG CGT ACA TGG TCA TCA CG
yebQ	KK064f	TTT CGC ACC AGG GTG GTC GC
	KK064r	CAC CCG CCG CTG GAT AAA CG
shiA	KK066f	GTC GAG GTA GTC TCA TAA GG
	KK066r	TCA TGT GTG TCT CCT GTT CC
mdtD	KK068f	AGC TCC TTA CGC TGT ATA CC
	KK068r	AAA ACT GAT ACG CAC CGC CC
setB	KK070f	GGC ATT TTG ACT ATT CGT GG
	KK070r	CGA CCG AAA GTC TGT GCA GG
bcr	KK072f	CGC TGG ATG CTG ATT TAG CC
	KK072r	GGG GAA TGA TAG ATT TGT GG
yfcJ	KK074f	AAG ACG TAC GCG TTA AAT CC
	KK074r	TTC CAC CGA TGA TGG ACA CG
kgtP	KK076f	CCT GCT AAC TAA ATT CAT GG
	KK076r	GGG ATA GCC GCC CCG AAG GG
ygcS	KK078f	TAA TGT GTC AGG CAG CGT TCC
	KK078r	CGA GTA AAC TTC GCA GTA GC

Table 2.5 (cont.)

galP	KK082f	GAT AAA TGT TAG TGT AAG CG
	KK082r	GAT GGC GAT AGG GAG ACG GG
nanT	KK084f	CCC TCG CTC GCC CCT ACC GG
	KK084r	ATC AGG CCA CCG TTA GCA GC
yhhS	KK086f	GTA GCG GGA GTC ACG CCA GC
	KK086r	TAA TCA ACT GCT GAC CAT GC
nepI	KK088f	AGC AAC ACC GTA AAA GTT GC
	KK088r	TTG GGA AGC GTT GGG TAA GG
emrD	KK090f	ATT CCT GGC GTA TAT CTG GC
	KK090r	TTA TTC TTA CTC CGC ACC GC
mdtL	KK092f	CTT GCC TCT CAG GAA ATC AC
	KK092r	GAC TTA CTC ACC GCC GAA GG
proP	KK094f	TTG CAT CCT GCA ATT CCC GC
	KK094r	CCT GGA GGA GAG TAT GCG CG
yjhB	KK096f	GTA CTG GCG TAT TTG AAG AC
	KK096r	AAA CGA GCT AAT TCA GCG CC
yjiO	KK098f	CAT TTG CAG TCT CGT TGC CC
	KK098r	TAA GCA GTA AAG AGA GCG CC
ydhC	KK100f	TGA CCA CAT AGT CTG CCT GC
	KK100r	TCG AAA TTA GTA TAG GCC GC

Table 2.6. Primers used to amplify efflux transporter promoters.

Efflux Transporter Gene	Primer	Promoter Sequence	Region Amplified
setC	KK165f	ATC GGT ACC ATA TAT CAA ATC AAC TTG TT	3834726-3834975
	KK165r	ATC GAA TTC TCT TCT GAA TCA CCG GGC GA	
ydeA	KK166f	ATC GGT ACC AAC CGT GAT GTT ACC GAC TC	1614902-1615051
	KK166r	ATC GAA TTC AAC AGA TTA TCG GGT GAG TT	
mhpT	KK167fii	ATC GGT ACC GAA CGT GCA TTT TGT ATG CC	374119-374682
	KK167r	ATC GAA TTC AAG ATG AAC CTC GTT GCG AG	
ybdA	KK168fi	ATC GGT ACC AAC TTC CAT GAT AAT GAA AT	621416-621522
	KK168r	ATC GAA TTC TAC AAT GCC TTG CCA TCA AA	
cmr	KK171f	ATC GGT ACC GGC ATAG GTTA AAT AAA ACT	882746-882895
	KK171r	ATC GAA TTC GCA ATT TCT TCG CCA ATA AT	
ydeE	KK169f	ATC GGT ACC CCG CCT CTG AAT TTC ATC TT	1619166-1619355
	KK169r	ATC GAA TTC TTG TCT GTC CGG CAG TGC GT	
ynfM	KK172f	ATC GGT ACC AAC ATC TTA TTT GAG ATT AT	1667623-1667722
	KK172r	ATC GAA TTC TTG AAA TCC TTG CTA AAT AT	
mdtD	KK173f	ATC GGT ACC TGT CTG CTG CGT TTT CGC CC	2159248-2159487
	KK173r	ATC GAA TTC TTA CTC GGT TAC CGT TTG TT	
kgtP	KK170f	ATC GGT ACC CTG CGC GTC TTA TAC TCC CA	2722220-2722469
	KK170r	ATC GAA TTC GCC ATT ATG TCT CCT GCC GT	
yfcJ	KK174f	ATC GGT ACC AGT TCC AGG CAA ACT TCC CG	2436824-2436963
	KK174r	ATC GAA TTC GCA TTA CTC CAG AAT GCA GC	
yhhS	KK175f	ATC GGT ACC ATT AAG TGT AGC GGG AGT CA	3608432-3608588
	KK175r	ATC GAA TTC AAA TTC AGG CGC AAT CCG TT	
emrD	KK176f	ATC GGT ACC GGG AAA CCG CTC CCC TTT TA	3851695-3851944
	KK176r	ATC GAA TTC TAT CAC GGA TGC TTT TAT AA	
ydhC	KK177f	ATC GGT ACC AAA TTT TTG AAA CCA GTC AT	1737835-1737934
	KK177r	ATC GAA TTC GTA TAT TCT CTA AAT CTC TC	

Table 2.7. Primers used to amplify efflux transporter genes.

Efflux Transporter Gene	Primer	Sequence
setA	KK040f	ATC GGT ACC GCC GCG CTA AAA AGG GAA CG
	KK040r	ATC GGA TCC GCT TTC CGC TAT CGT CCA CG
setC	KK026f	ATC GGT ACC ATG CAA AAA ACG GCT ACC ACT CC
	KK026r	ATC GGA TCC CCC AGA CTG GTG AGC TAA ATA TC
ydeA	KK013f	ATC GGT ACC ATT AAA CCG TGA TGT TAC CG
	KK014r	ATC GGA TCC TGG GCC TTT CAA CTA TTG CG
mhpT	KK048f	ATC GGT ACC CTT ATG TCG ACT CGT ACC CC
	KK048r	ATC GGA TCC AAG GCA CGT CAG GCA TCG GC
yajR	KK049f	ATC GGT ACC AAC AGA GGT GGT AAT GAA CG
	KK049r	ATC GGA TCC TTA TGC CTG ACG AAT TGC C
ybdA	KK050f	ATC GGT ACC GCA TGT GCA GCC TAA GAA TAG G
	KK050r	ATC GGA TCC TTA ACT GTC GGA CGC TGT CAC C
cmr	KK143f	ATC GAA TTC TTG GCG AAG AAA TTG CAT GC
	KK143r	ATC GGT ACC TTA CCC TTC GTG AGA ATT TCC
ycaD	KK052f	ATC GGT ACC ATG TCC ACG TAT ACC CAG CC
	KK052r	ATC GGA TCC ATT TAC ACG TGA GCA ACG GG
yceL	KK119f	ATC GGT ACC ATG TCC CGC GTG TCG CAG GC
	KK119r	ATC GGA TCC ATC AGG CGT CGC GTT CAA GC
ydeE	KK120f	ATC GGT ACC ATG AAC TTA TCC CTA CGA CG
	KK120r	ATC GGA TCC TCA ACA AAG CGC GGG CTG CC
ynfM	KK121f	ATC GGT ACC CAA GGA TTT CAA GTG AGC CG
	KK121r	ATC GGA TCC TTT CAG GCG TGC AGA CGA CG
ydhP	KK122f	ATC GGT ACC GCA TGA AAA TTA ACT ATC CG
	KK122r	ATC GGA TCC TAA TTA GCT GTT AGC AAC GC
mdtG	KK118f	ATC GGT ACC AGC GGA TTG CTA TGT CAC CC
	KK118r	ATC GGA TCC AAT CAG TTC GAT ACC TGG GG
yebQ	KK123f	ATC GGT ACC TAT GCC AAA AGT TCA GGC CG
	KK123r	ATC GGA TCC TTT ATG CCC TGG ATC GTG GC
shiA	KK124f	ATC GGT ACC ATG GAC TCC ACG CTC ATC TCC
	KK124r	ATC GGA TCC TCG CCA GGC TAT CAA GCG CG
mdtD	KK125f	ATC GGT ACC GTA AAT GAC AGA TCT TCC CG
	KK125r	ATC GGA TCC TCA TTG CGC GCT CCT TTT TCG
setB	KK126f	ATC GGT ACC AGC ATG CAT AAC TCC CCC GC
	KK126r	ATC GGA TCC CTT AAA CAT CTT TAA TCC GC
bcr	KK146f	ATC GAA TTC GTG ACC ACC CGA CAG CAT TCG
	KK146r	ATC GGT ACC AGA TCA CCG TTT TTT CGG CCG
yfcJ	KK103f	ATC GGT ACC ATG ACT GCT GTA AGC CAA ACC
	KK103r	ATC GGA TCC CTG GTT CGT TAA CCC CGA CG
kgtP	KK104f	ATC GGT ACC ATG GCT GAA AGT ACT GTA ACG
	KK104r	ATC GGA TCC CTA AAG ACG CAT CCC CTT CCC
ygcS	KK105f	ATC GGT ACC ATG AAC ACT TCA CCG GTG CG
	KK105r	ATC GGA TCC TCC GGT TCG GCT TTA AAC GC

Table 2.7 (cont.)

galP	KK144f	ATC GAG CTC ATG CCT GAC GCT AAA AAA CAG G
	KK144r	ATC TCT AGA TTA ATC GTG AGC GCC TAT TTC GC
nanT	KK108f	ATC GGT ACC TCA CAA TGA GTA CTA CAA CC
	KK108r	ATC GGA TCC GCA ACA GGA TTA ACT TTT GG
yhhS	KK109f	ATC GGT ACC CAT GCC CGA ACC CGT AGC CG
	KK109r	ATC GGA TCC TTT AAG ATG ATG AGG CGG CC
nepI	KK110f	ATC GGT ACC CAT GAG TGA ATT TAT TGC CG
	KK110r	ATC GGA TCC TCA GGA TTT CTT CAT TTT CAC C
emrD	KK145f	ATC GAA TTC ATC CGT GAT AAT GAA AAG GC
	KK145r	ATC GGT ACC AAC GGG CTG CCC CTG ATG CG
mdtL	KK112f	ATC GGT ACC CGC TTT AGC CCA TGT CCC GC
	KK112r	ATC GGA TCC GAG ATC AAG CGT GGT GAT GG
proP	KK113f	ATC GGT ACC GGA AAG CTA TGC TGA AAA GG
	KK113r	ATC GGA TCC GCT TAT TCA TCA ATT CGC GG
yjhB	KK114f	ATC GGT ACC CAA ATA TGG CAA CAG CAT GG
	KK114r	ATC GGA TCC CTC CAA TCA TTT AGC CAC GG
yjiO	KK115f	ATC GGT ACC ATG CCA CGT TTT TTT ACC CGC
	KK115r	ATC GGA TCC CAC TGC TCC TCC ACT AGC TCG
ydhC	KK116f	ATC GGT ACC GAA TAT ACA TGC AAC CTG GG
	KK116r	ATC GGA TCC TCA GTG TGA TTC GCT ATG AGC

Table 2.8. K_M and V_{max} values*.

Efflux transporter deletions			Efflux transporter overexpressions				
Gene	**K_M**	**V_{max} x 10^4**	**Gene**	2 mM IPTG		0.2 mM IPTG	
				K_M	**V_{max} x 10^4**	**K_M**	**V_{max} x 10^4**
Control	3.17	6.88	Control	2.73	11.1	2.84	11.9
setC	3.51	12.7	*setC*	Growth defect		5.27	6.49
cmr	3.94	13.2	*ydeE*	Growth defect		4.10	7.59
ynfM	3.77	12.3	*ydeA*	3.11	4.53	No effect	
mdtD	4.17	13.8	*kgtP*	2.75	7.65	No effect	
yfcJ	3.04	9.96	*mhpT*	0.345	7.08	1.35	7.47
yhhS	4.96	14.8	*ybdA*	0.933	6.84	1.33	7.93
emrD	3.61	12.6					
ydhC	3.82	12.4					

*K_M values are reported in mM

V_{max} values are reported as Fluorescence/OD_{600}

Table 2.9. Summary of results.

Efflux transporter	Deletion effect?	Overexpression effect?	Effect on arabinose mutants	TolC dependent?
setC	yes	yes	*ΔaraA, ΔaraB, ΔaraBAD*	no
cmr	yes	no	not observed	undetermined
ynfM	yes	no	not observed	undetermined
mdtD	yes	no	not observed	undetermined
yfcJ	yes	no	not observed	undetermined
yhhS	yes	no	not observed	undetermined
emrD	yes	no	not observed	undetermined
ydhC	yes	no	not observed	undetermined
ydeE	no	yes	*ΔaraA, ΔaraB, ΔaraBAD*	yes
ydeA	no	yes	*ΔaraA, ΔaraB, ΔaraBAD*	yes
kgtP	no	yes	*ΔaraB*	yes
mhpT	no	yes	*ΔaraB, ΔaraBAD*	no
ybdA	no	yes	*ΔaraB, ΔaraBAD*	yes

Table 2.10. Sequence homology of identified putative efflux transporters to *ydeA*.

Efflux transporter	Homology to *ydeA*[1,2]
setC	59.2% identity in 76 nt overlap (66-141:34-108) 65.2% identity in 69 nt overlap (232-298:238-301) 58.4% identity in 125 nt overlap (735-852:1020-1143)
cmr	59.2% identity in 130 nt overlap (196-310:724-853) 76.7% identity in 30 nt overlap (358-387:169-198) 56.1% identity in 107 nt overlap (348-451:198-304)
ynfM	63.6% identity in 55 nt overlap (803-856:679-733)
mdtD	64.8% identity in 54 nt overlap (33-84:43-96) 57.8% identity in 109 nt overlap (290-398:800-903) 56.5% identity in 154 nt overlap (507-660:820-963)
yfcJ	54.5% identity in 202 nt overlap (693-886:111-307)
yhhS	61.1% identity in 131 nt overlap (718-839:307-432) 63.5% identity in 52 nt overlap (836-887:941-992)
emrD	68.6% identity in 51 nt overlap (844-892:919-967)
ydhC	No local region homology
ydeE	No local region homology
kgtP	59.5% identity in 84 nt overlap (184-266:468-548) 62.5% identity in 88 nt overlap (223-303:460-547) 63.9% identity in 61 nt overlap (691-750:793-853)
mhpT	59.7% identity in 77 nt overlap (933-1006:39-115)
ybdA	54.2% identity in 144 nt overlap (997-1136:679-821)

1. Sequences were aligned and analyzed using William Pearson's LALIGN program. The lalign program implements the algorithm of Huang and Miller, published in Adv. Appl. Math. (1991) 12:337-357. Default settings of this program were used.
2. The regions are reported as: sequence region of *ydeA*: sequence region of efflux transporter

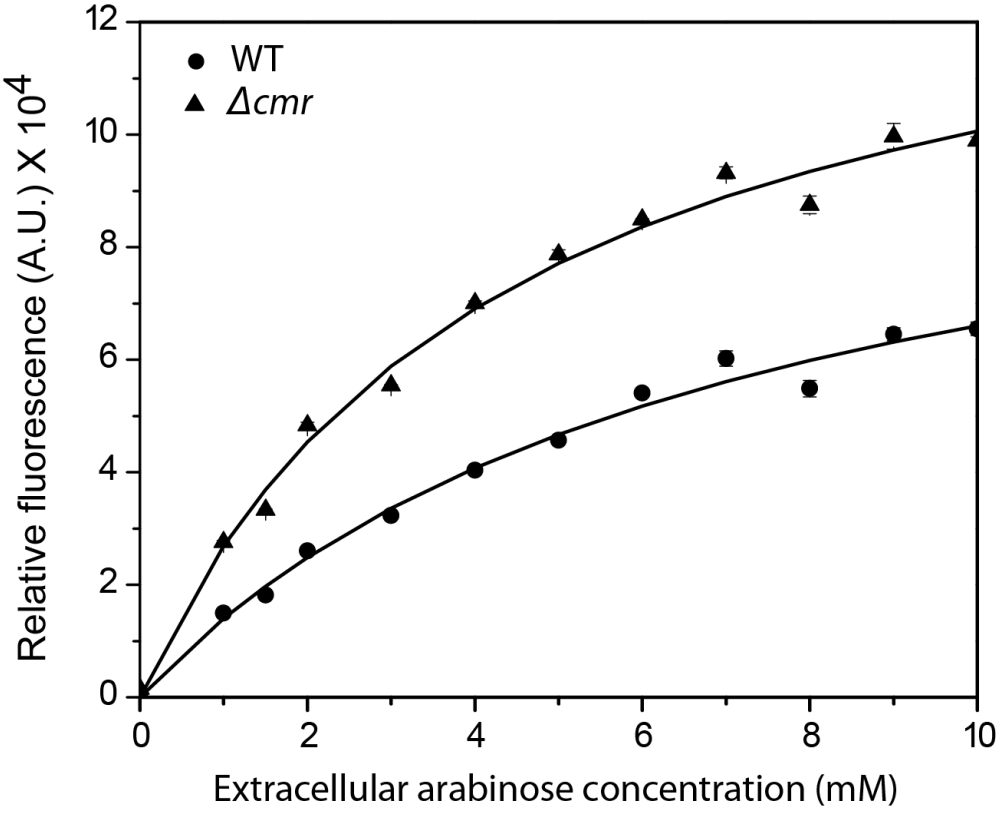

Figure 2.1. Dose-response curve for the *Δcmr* mutant. P_{araB}-*venus* in WT and *Δcmr* mutant strains were measured for fluorescence and the results are reported as Fluorescence/OD_{600}. Strains were induced with arabinose concentrations varying from 0 mM to 10 mM.

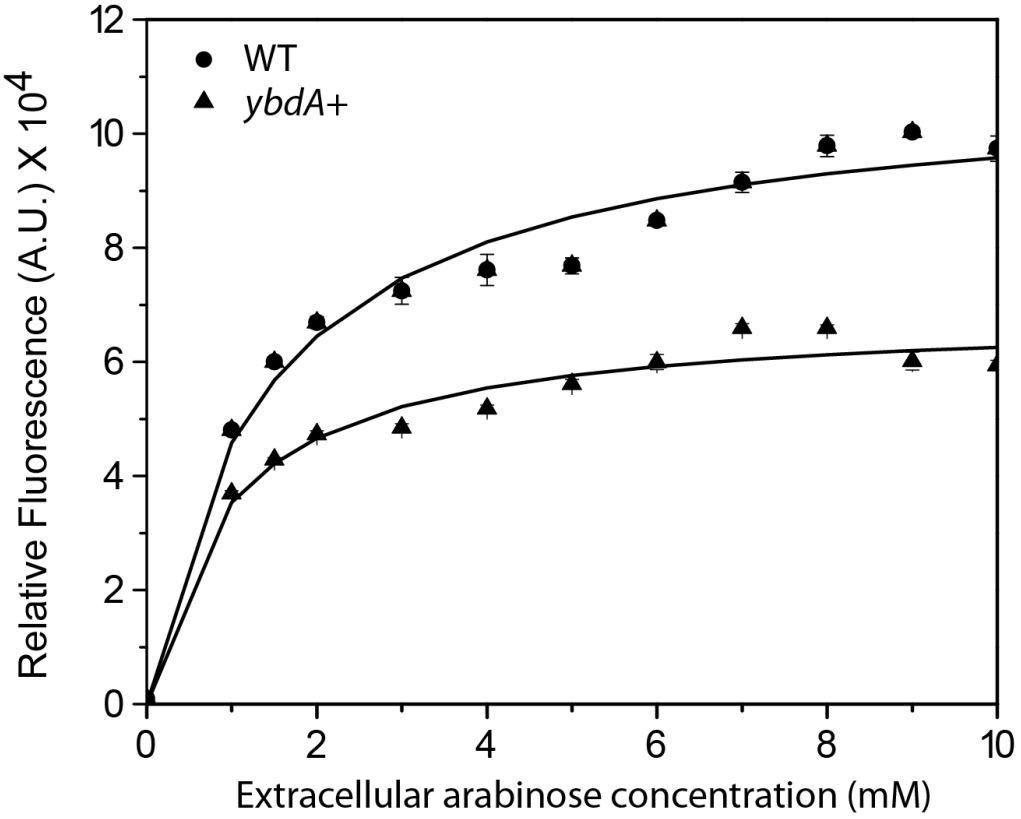

Figure 2.2. Dose-response curve for *ybdA*+. P_{araB}-*venus* in WT and *ybdA* overexpressing strains were measured for fluorescence and the results are reported as Fluorescence/OD_{600}. Strains were induced with arabinose concentrations varying from 0 mM to 10 mM. The *ybdA* overexpressing strain was induced with 2 mM IPTG.

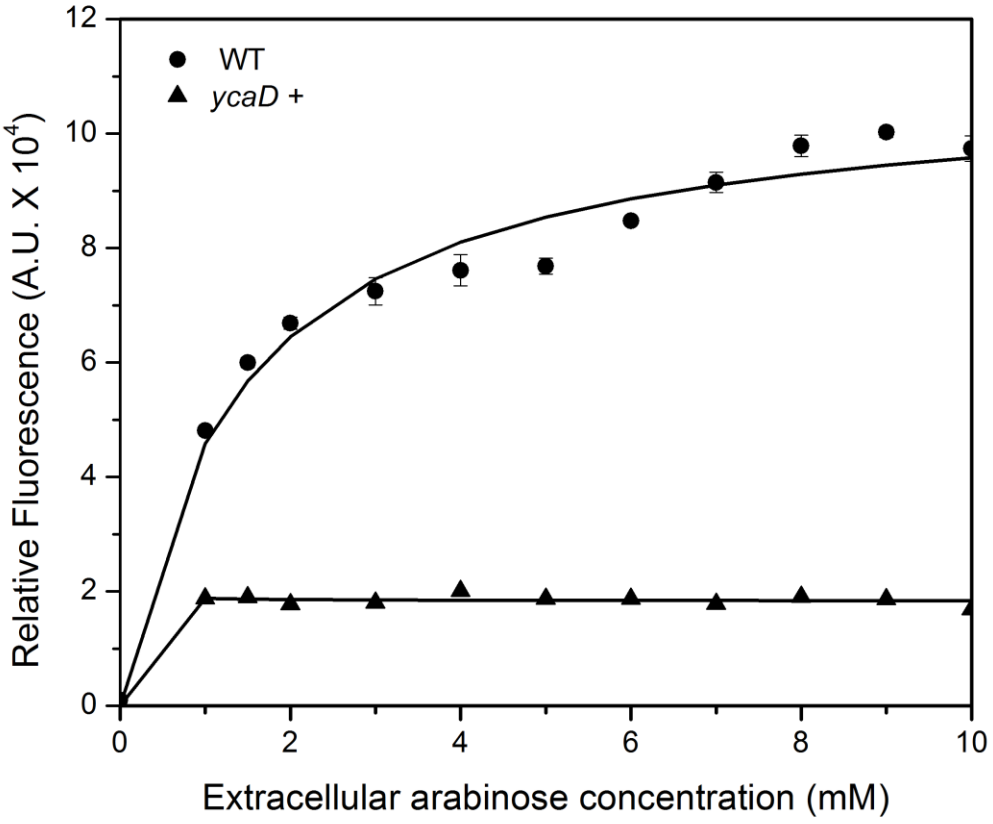

Figure 2.3. **Dose-response curve for *ycaD*+ induced with 2 mM IPTG.** P_{araB}-*venus* in WT and *ycaD* overexpressing strains were measured for fluorescence and the results are reported as Fluorescence/OD_{600}. Strains were induced with arabinose concentrations varying from 0 mM to 10 mM.

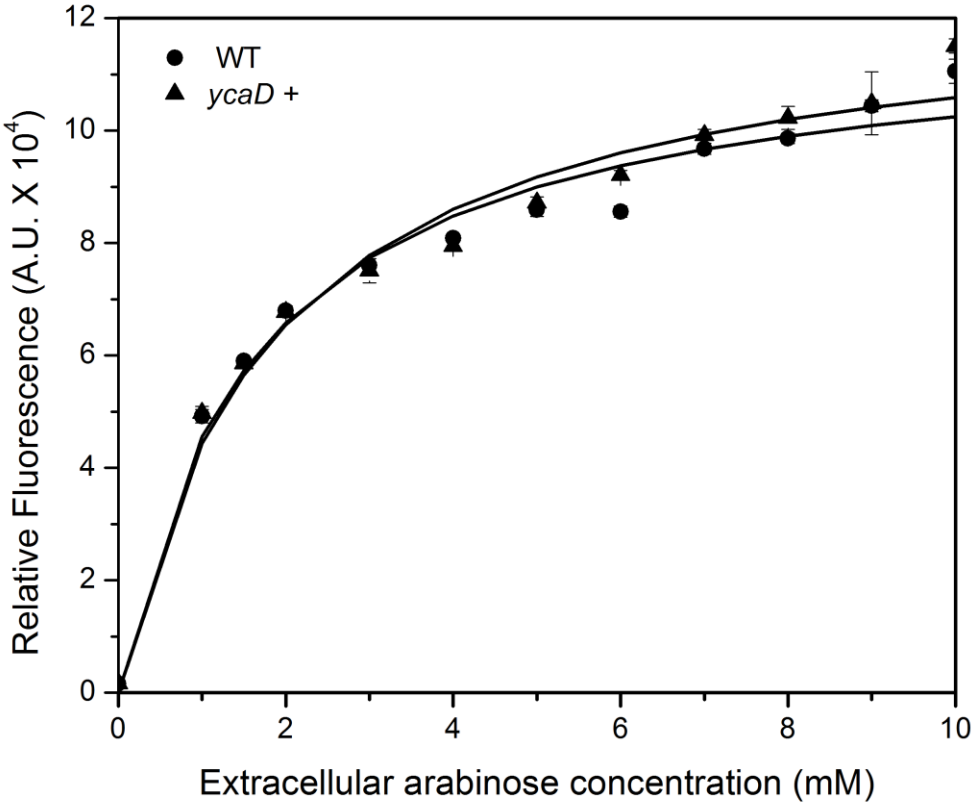

Figure 2.4. **Dose-response curve for *ycaD*+ with no IPTG induction.** P_{araB}-*venus* in WT and *ycaD* overexpressing strains were measured for fluorescence and the results are reported as Fluorescence/OD_{600}. Strains were induced with arabinose concentrations varying from 0 mM to 10 mM.

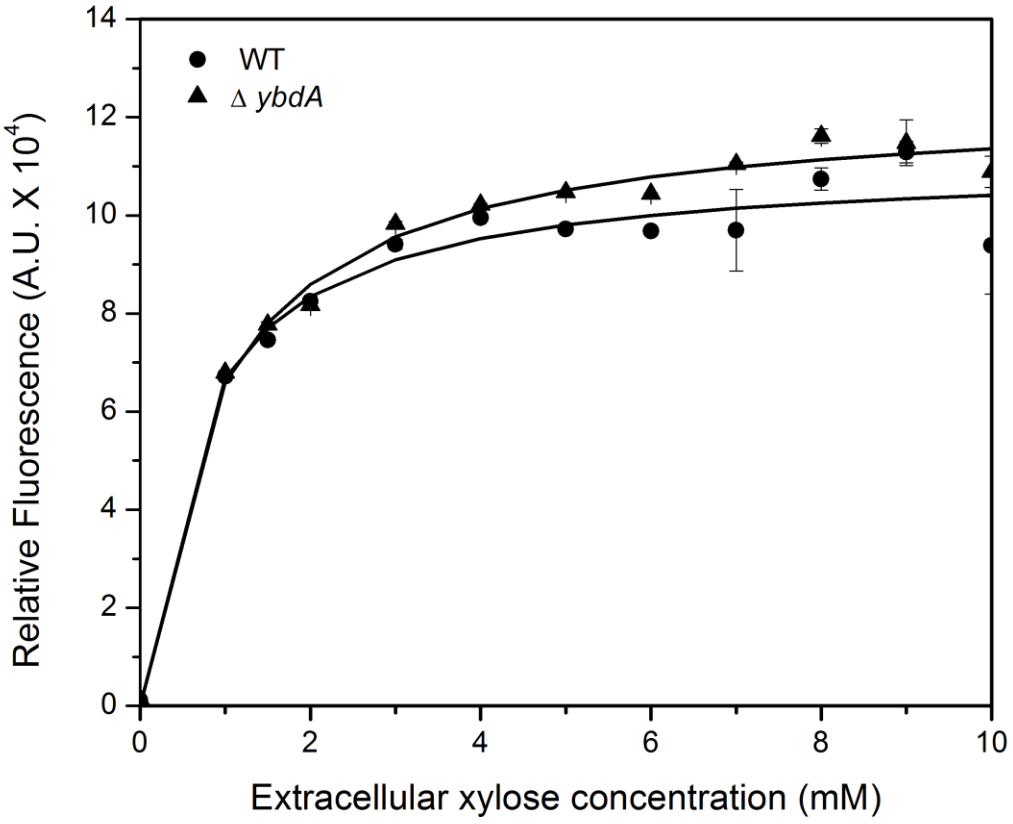

Figure 2.5. Dose-response xylose curve for the *ΔybdA* mutant. P_{xylA}-*venus* in WT and *ΔybdA* mutant strains were measured for fluorescence and the results are reported as Fluorescence/OD_{600}. Strains were induced with xylose concentrations varying from 0 mM to 10 mM.

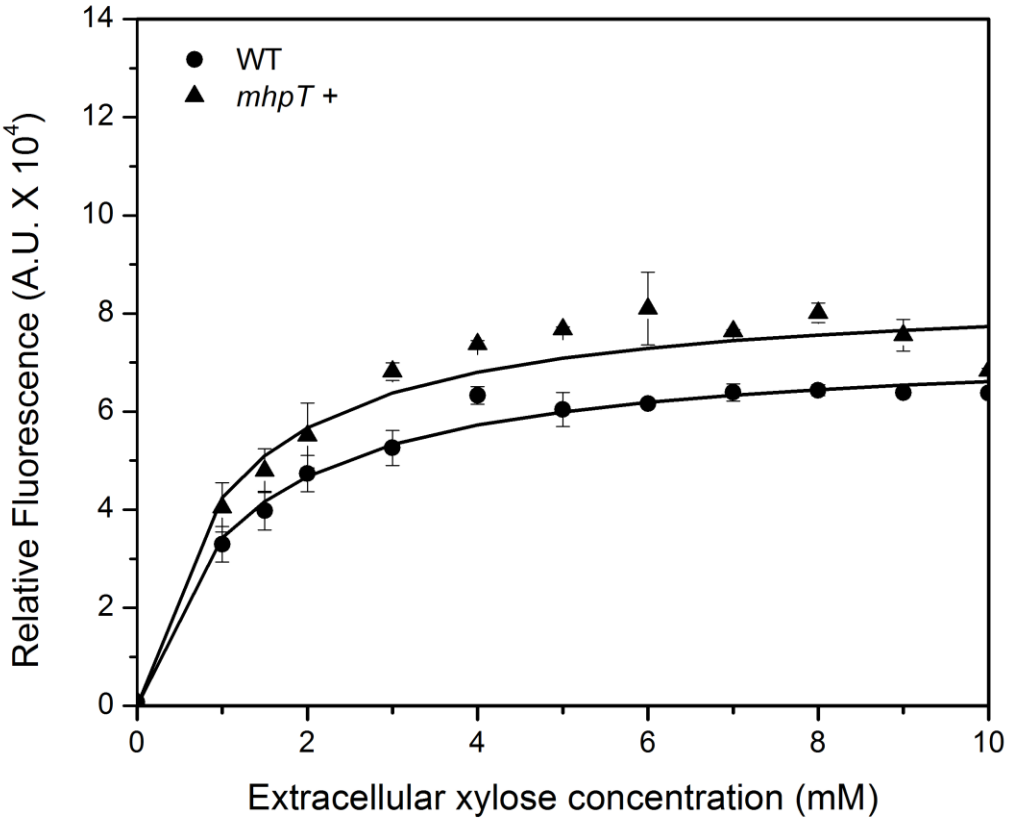

Figure 2.6. Dose-response xylose curve for *mhpT*+. P_{xylA}-*venus* in WT and *mhpT* overexpressing strains were measured for fluorescence and the results are reported as Fluorescence/OD_{600}. Strains were induced with xylose concentrations varying from 0 mM to 10 mM. The *ybdA* overexpressing strain was induced with 2 mM IPTG.

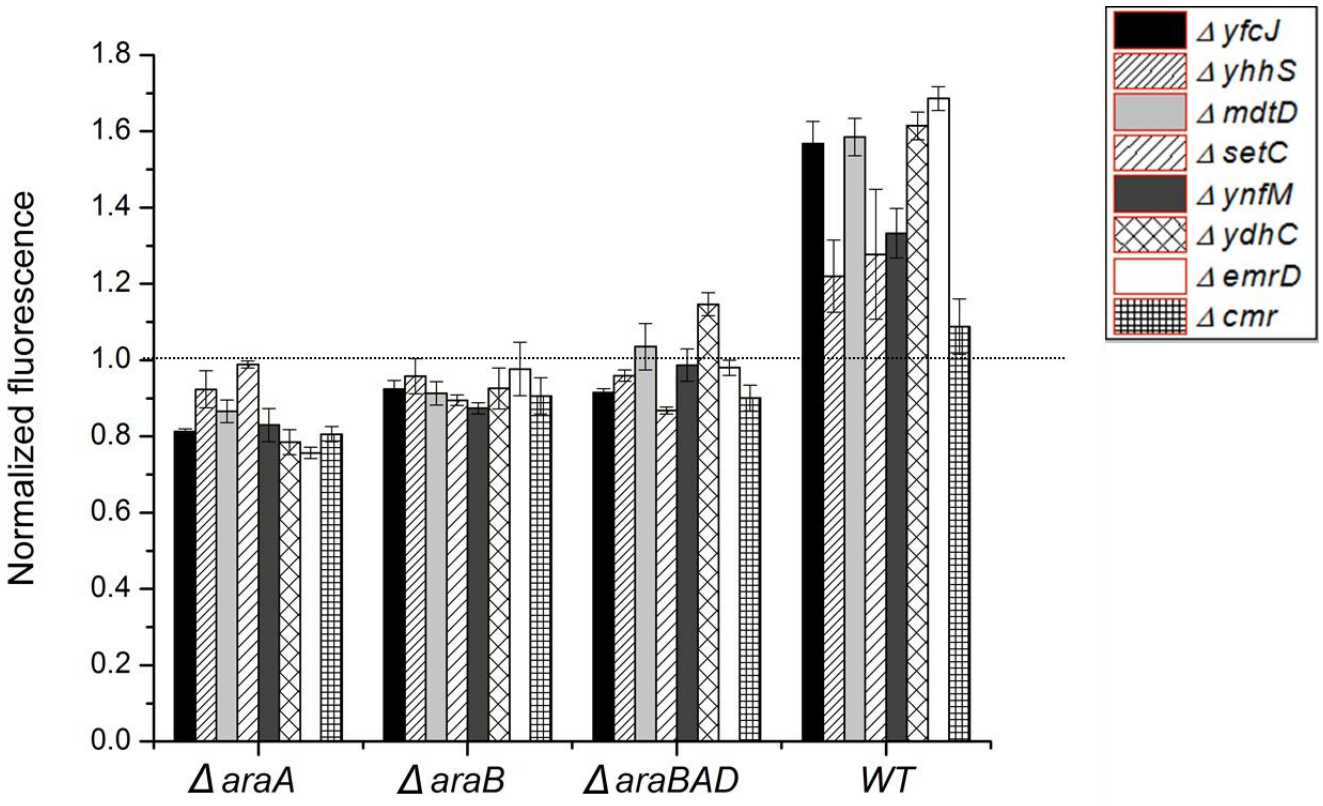

Figure 2.7. Deletion of the efflux transporters in arabinose mutants. The data is normalized to the relative fluorescence values of the control strains (P_{araB}-*venus* in each of the individual arabinose mutants). All of the strains were induced with 5 mM arabinose.

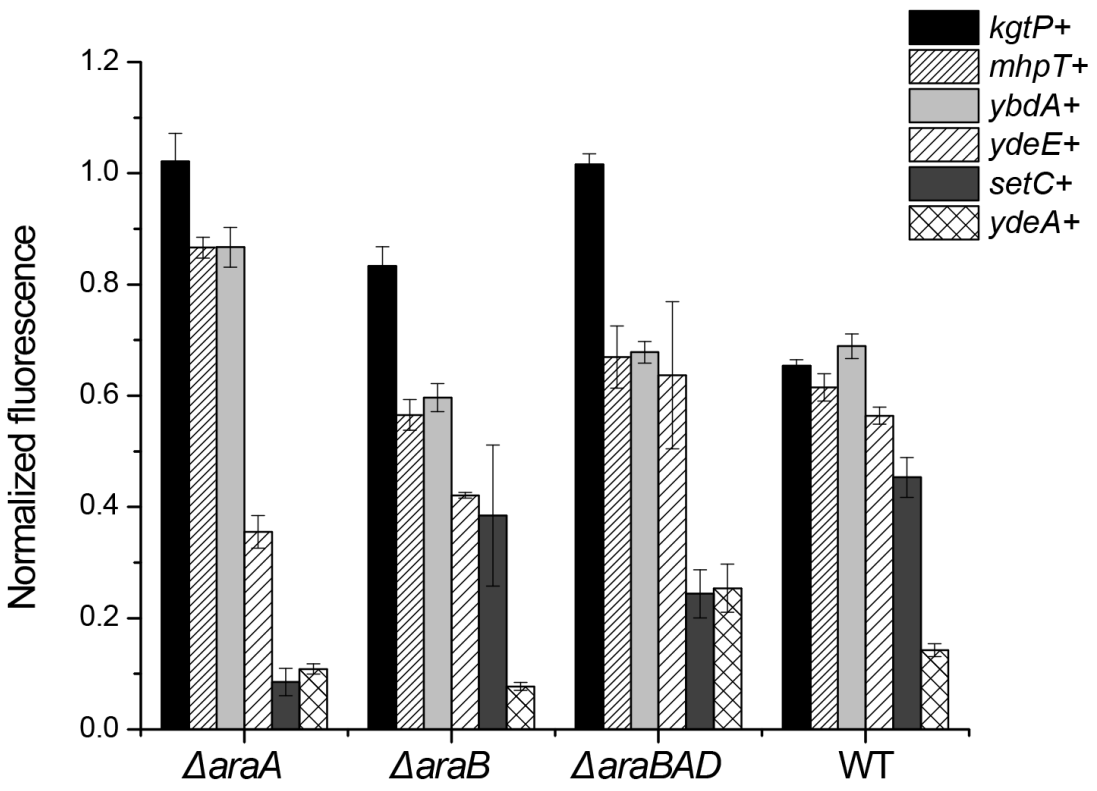

Figure 2.8. Overexpression of the efflux transporters in arabinose mutants. The data is normalized to the relative fluorescence values of the control strains (P_{araB}-*venus* and pTrc99A (empty plasmid) in each of the individual arabinose mutants). All of the strains were induced with 5 mM arabinose. Overexpression of the the efflux transporters was induced with 2 mM IPTG except for the overexpression of *setC* and *ydeE* where 0.2 mM IPTG was used.

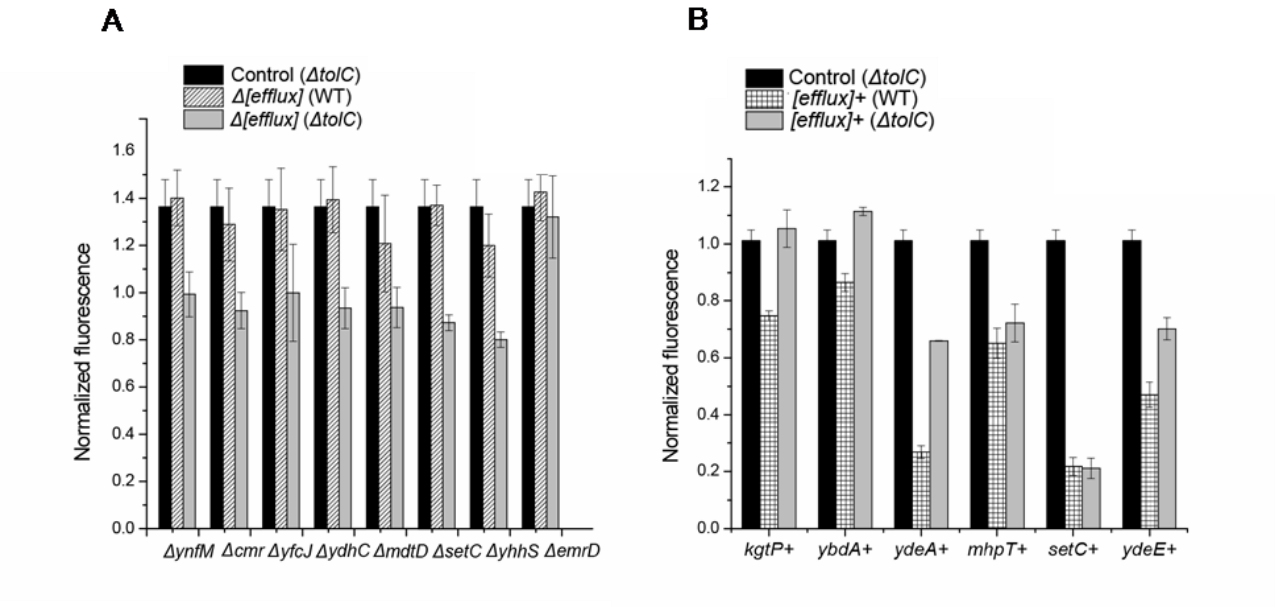

Figure 2.9. Effect of deleting *tolC* in the. **A)** efflux transporter knockout strains and **B)** efflux transporter overexpressing strains. The data is normalized to the relative fluorescence values of the WT control strain (P_{araB}-*venus* and pTrc99A in WT). All of the strains were induced with 5 mM arabinose. Overexpression of the the efflux transporters was induced with 2 mM IPTG except for the overexpression of *setC* and *ydeE* where 0.2 mM IPTG was used.

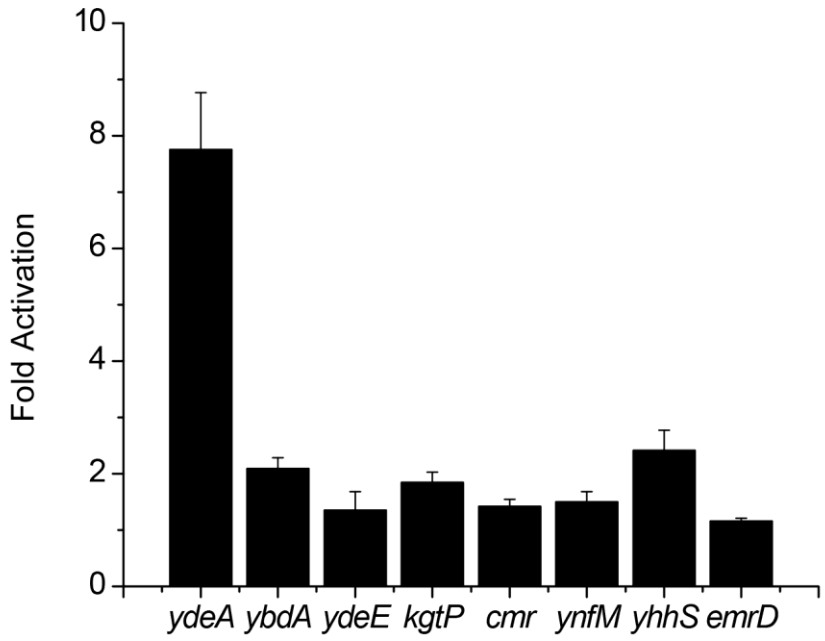

Figure 2.10. Efflux promoter transcriptional fusions. Reporters are in WT background and the data is normalized to the relative fluorescence values when no arabinose is added. All of the strains were induced with 10 mM arabinose.

CHAPTER 3: Targeted rational approach to simultaneous sugar co-utilization in *Escherichia coli*

3.1 Introduction and background

The goal of this project is to engineer a strain that is efficient in co-utilizing multiple sugars simultaneously. As mentioned previously, the three main sugars of concern are glucose, arabinose and xylose. In the case where all three sugars are present in a mixture, *E. coli* will consume glucose first, then arabinose and then finally xylose. During the consumption of each individual sugar, a diauxic shift exists and this lag phase results in inefficient consumption of the sugar mixture. To find a solution to this problem, it is necessary to study the hierarchy that exists between glucose, arabinose, and xylose consumption and research possible ways to eliminate this hierarchy. Previous work has been devoted towards investigating the glucose-arabinose hierarchy as well as the hierarchy that exists between arabinose and xylose [56]. Current findings on the glucose-xylose hierarchy, however, have not been thoroughly studied.

Given the central role of PTS and CRP in glucose-mediated catabolite repression, a significant amount of research has been devoted to exploring the use of *ptsG* (glucose-specific transporter of the PTS) knockouts as a strategy for enabling simultaneous sugar utilization [92]. Mutations in *ptsG* have been investigated and successful results have been achieved. It was previously shown that in *E. coli*, deletion of the *ptsG* gene can remove catabolite repression at the expense of a lower glucose uptake rate [45]. The time required to complete fermentation of sugar mixtures in rich medium has been greatly decreased [93]. However, similar results have not been achieved in mineral salts medium, which is particularly desirable for commercial processes [94]. Mutations that activate CRP can also result in higher co-utilization of sugars. However, because CRP is a global regulator controlling the expression of approximately 200 promoters

[95], its activation leads to a large-scale reprogramming of cellular physiology. The significances of these changes are not known and rather than attempting to analyze every change possible, other approaches to engineering a strain capable of co-utilizing sugars are more desirable.

Our goal is to eliminate repression of the metabolic and transport genes of xylose, without affecting cAMP or CRP. The operon responsible for the metabolism of xylose in *E. coli* is regulated by other genes such as *xylR* and *araC*. We proposed a mechanism by which this operon could be controlled by a constitutively "on" promoter, so that it is not affected by repression of other genes. This allows expression of the xylose metabolic genes even in multiple sugar mixtures where catabolite repression would normally result in repression of the *xylAB* operon and cease xylose metabolism. To do this we first replaced the *xylA* promoter with a *tetRA* element and selected on tetracycline. We then replaced the *tetRA* element with a constitutively "on" promoter and selected on minimal xylose plates (to ensure functionality of the *xylA* operon). A schematic diagram for the proposed promoter replacement is shown in Figure 3.1.

3.2 Materials and methods

Media and growth conditions. Luria-Bertani liquid and solid medium (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) was used for routine bacterial culture and genetic manipulation. All experiments were performed in M9 minimal media (6 g/L $Na_2 HPO_4$, 3 g/L $KH_2 PO_4$, 0.5 g/L NaCl, 1 g/L NH_4 Cl, 1 mL of a 1 M solution of $MgSO_4$. $7H_2O$, 50 µL of 1M solution of $CaCl_2$) at 37**°**C. Antibiotics were used at the following concentrations: ampicillin at 100 µg/mL, chloramphenicol at 20 µg/mL, and kanamycin at 40 µg/mL unless otherwise noted. All experiments involving the growth of cells carrying the helper plasmid pKD46 were performed at 30 ^{o}C. Loss of pKD46 was achieved by growth at 42 ^{o}C under nonselective conditions on LB agar. Sugars were added at 20mM concentrations for overnight cultures and experiments unless otherwise specified. The sugars and antibiotics were sterile filtered before adding them

to the growth flasks. Primers were purchased from IDT, Inc. Enzymes were purchased from New England Biolabs and Fermentas.

Tetracycline counter selection. Chromosomal tetracycline counter-selection was done using the method described previously [96]. Briefly, the method consists of two steps. In the first step, the region of interest to be mutated was first replaced with a *tetRA* element from the transposon Tn*10*. This was done using λ-Red recombination using the plasmid pKD46. The resultant strain is resistant to tetracycline and was selected for on tetracycline plates. In the second step, the tetracycline resistance marker on the plates was replaced by the desired mutation. The strain CR1264 was made by first replacing the *xylA* promoter (region 3727343-3727450) with a *tetRA* element. The *tetRA* element was then moved into a clean wild-type background by P1*vir* transduction. *The tetRA* element was then replaced by the $P_{LtetO-1}$ promoter amplified from the pPROTet.E plasmid. The $P_{LtetO-1}$ promoter was amplified using primers KK148fii (sequence: 5'-GAA TTA CCC AGT TTC ATC ATT CCA TTT TAT TTT GCG AGC GCT TAG CTC CTG AAA ATC TCG) and KK148r (sequence: 5'- CGA GCT GGT CAA AAT AGG CTT GCA TAT TGA ACT CCA TAA TGG TCA GTG CGT CCT GCT GAT), where the underlined sequence is the region on the pPROTet.E plasmid. This was done using λ-Red recombination using the plasmid pKD46 and the colonies were selected for on minimal (M9) xylose plates. Colonies from these plates were then re-streaked on LB plates, first at 42°C and then at 37°C.

Strain and plasmid construction. Bacterial strains and plasmids used in this study are described in Tables 4.1 and 4.2 respectively. All strains are isogenic derivatives of *Escherichia coli* K-12 strain MG1655. All cloning steps were performed in *E. coli* strain DH5α (*phi-80d lac*Δm15 *enda1 recA1 hsdR17 supE44 thh*-1 *gyrA96 relA*Δ*lacU169*). The generalized transducing phage P1*vir* was used in all genetic crosses according to standard methods [19].

The plasmid pAraE (pKK046) was constructed by first amplifying the *araE* gene (region 2978766-2980223) using primers KK141f (sequence: 5'- ATC GAA TTC CTC TCT GCT GGC AGG AAA AA) and KK141r (sequence: 5'- ATC GGT ACC AGC ACA TCC GGC CCG TGA AA). The PCR product was then digested with EcoRI and KpnI (sequences underlined) and cloned into pPROTet.E (Clontech) under the control of a strong promoter, $P_{LTetO-1}$ [97]. The plasmid pXylE (pKK047) was constructed by first amplifying the *xylE* gene (region 4238782-4240296) using primers KK140f (sequence: 5'- ATC GAA TTC AGA ATG GTC TAA GGC AGG TC) and KK140r (sequence: 5'- ATC GGA TCC GGC GCG GCG TGC TGG ACA GG). The PCR product was then digested with EcoRI and BamHI (sequences underlined) and cloned into pPROTet.E. The pPDC AdhB was constructed by Kori Dunn (unpublished) in a sequential process. First, the *pdc* gene from Zymomonas mobilis (ZM4) [98] was amplified using the primer sequences: 5'- ATG CGA ATT CTA TAT GGA GTA AGC AAT GAG TTA TAC TGT CGG TAC C (forward) and 5'- ATG CGG ATC CCT AGA GGA GCT TGT TAA CAG GC (reverse). The PCR product was then digested with EcoRI and BamHI and cloned into pPROTet.E, resulting in plasmid pPDC. The *adhB* gene was amplified from Zymomonas using primers: 5'- ATG CGG ATC C**AG GAG TG**T TAT AGC TAT GGC TTC TTC AAC TTT TTA TAT TCC (forward; bold letters are the added sequence for an optimal ribosome binding site for cloning in *E. coli*) and 5'- ATG CCT GCA GGA AAT TAG AAA GCG CTC AGG (reverse). The PCR product was digested with BamHI and PstI and cloned into plasmid pPDC to yield pPDC AdhB.

CRIM integrations were performed in accordance to the method described by Haldimann and Wanner [99]. In order to carry out the *araE* integration, the region consisiting of the *araE* gene controlled by the $P_{LTetO-1}$ promoter from the pAraE plasmid was first amplified. The primers used for this amplification are KK161f (sequence: 5'- ATC CTG CAG TCG AGT CCC TAT CAG TGA TA) and KK141r. For the integration of *xylE,* the region consisiting of the *xylE* gene controlled by the $P_{LTetO-1}$ promoter from the pXylE plasmid was amplified. The primers used for this

amplification are KK161f and KK140r. The PCR products were then digested with PstI and KpnI and cloned into the multiple cloning sites of pAH143K and pAH153C. The pAH143K and pAH153C plasmids were constructed by Lon Chubiz (unpublished), in which the gentamicin resistant gene in pAH143 and pAH153 CRIM plasmids is replaced by the kanamycin and chloramphenicol resistant genes respectively. The region containing the *pdc* and *adhB* genes in the pPROTet.E plasmid, controlled by the $P_{LTetO-1}$ promoter was amplified from the plasmid using primers KK161f and KK180r (ATC TCT AGA GAA ATT AGA AAG CGC TCA GG). The PCR product was digested with PstI and XbaI and cloned into the multiple cloning sites of pAH143K, yielding plasmid pKK052. The CRIM helper plasmids pAH69 (helper plasmid for CRIM plasmid pAH143K) and pAH123 (helper plasmid for CRIM plasmid pAH153C) were used to complete the integrations in *E.coli* and the integrated plasmids were moved into different mutant strains by P1*vir* transduction.

Kinetic growth experiments. Single colonies were picked from freshly streaked plates and grown overnight in glass test-tubes containing 3 ml of minimal medium (M9) and 20mM glucose. For preliminary kinetic growth measurements, overnight cultures were subcultured to an OD of 0.05 in fresh medium and 150 µL was transferred to 96-well microtiter plates. Plates were sealed with Breath-Easy membranes (Diversified Biotech) to reduce evaporation. The temperature was maintained at 37°C and OD_{600} readings were taken every 20 minutes.

For shake flasks growth experiments, the overnight cultures were inoculated in pre-sterilized 250 mL Erlenmeyer shake flasks containing 50 mL of M9 and 20mM total concentration of sugars. The initial OD_{600} after inoculation was set to 0.02. The temperature and agitation rate were set at 37°C and 500 rpm respectively. The optical densities of the cultures were measured every 30 minutes at a wavelength of 600 nm in 1.5 cm cuvettes using a spectrophotometer.

Analytical measurements. Sugar concentrations were measured using high-performance liquid chromatography (HPLC). Cells were first grown overnight in test tubes with M9 supplemented with glucose (20 mM). The cells were then subcultured 1:30 into fresh M9 supplemented with the appropriate sugars and were grown in test tubes. Periodically, a sample was first filtered using a 0.22 µm pore size filter (Millipore) and then run on a Gilson HPLC equipped with an autosampler, a cation exchange Aminex HPX-87H column (Bio-Rad), and an evaporative light scattering detector (ELSD). The column was operated at a temperature of 65°C. A 10 mM Formic Acid (HPLC grade) solution was used as the mobile phase and run isocratically at a flow rate of 0.6 mL/min. Each sample was run through the column for 15 minutes (retention times for glucose and xylose are 10.12 and 11.1 minutes respectively). The chromatogram peaks were analyzed using the Gilson Safire software. All experiments were conducted in triplicates and average values with one standard deviation error bars are reported.

3.3 Results

Replacing the *xylA* promoter with a constitutive promoter improves co-utilization in a mixture of sugars, but results in poor xylose utilization. Growth curves of wild-type *E. coli* demonstrate a diauxic shift in a mixture of sugars. As mentioned before, our primary focus is on improving xylose metabolism and overcoming the inhibiting effects of catabolite repression in the presence of glucose. In order to gain an understanding of how the two sugars are utilized in the cell, two main analyses were made. First, we produced growth curves in minimal media to observe how the cells grow in 1) individual sugars and 2) mixtures of sugars. Second, we analyzed specifically how each of these sugars was utilized by the cell over time.

The growth curves for wild-type *E. coli* in individual sugars are shown in Figure 3.2 (glucose) and Figure 3.3 (xylose). Figure 3.4 shows the growth of the cells when grown in a mixture of glucose and xylose. The diauxic shift is clearly visible between 9 hrs and 12 hrs. There is a

distinct lag phase during which one sugar (glucose) is completely consumed and the cells are preparing to use the next sugar in line (xylose). Using HPLC analysis, we were also able to measure extracellular sugar concentrations over time. This gives a measure of how the cells are utilizing each sugar during the course of their growth. Cells were supplemented with 2mM of each sugar and allowed to grow for eight hours. Figure 3.5 shows the depletion of the sugars for six hours of growth. What is most evident from this analysis is that in a glucose-xylose mixture, the cells only start to consume xylose when all the glucose is depleted. This is in agreement with the diauxic shift we observe in growth curves. Glucose is consumed to completion within four hours whether it is the sole carbon source or not. The cells take longer to consume xylose (five hours) when it is the sole carbon source.

To compare the performance of our engineered strain (CR1264) with the wild-type strain, we did similar growth and sugar utilization analyses. A micro-plate reader was used for a quick overnight kinetic growth measure. The results for the individual sugars as well as glucose-xylose mixtures are shown in Figures 3.6, 3.7 and 3.8. The engineered strain grows normally on glucose. In a mixture of sugars, however, we can see a distinct difference between the growth of the engineered strain and that of wild-type. While the diauxic shift is completely eliminated in the engineered strain, we observe that the cells grow much slower than the wild-type cells. We can see the reason for this effect more clearly in Figure 3.8. The engineered strain does not grow as well in xylose as the wild-type strain. It should be noted, however, that this strain is able to use xylose as the sole carbon source and cultures reach normal OD_{600} levels when left to grow overnight. The growth rate in xylose, however, was significantly decreased. These results are also confirmed by HPLC analysis (Figure 3.9). We see that there is no xylose depletion even after eight hours of growth.

Adding a transporter to the engineered strain significantly improves co-utilization of sugars. We hypothesized that the reason the engineered strain was not able to efficiently utilize

xylose was due to transport limitations. Even though our strain now had a constitutive promoter that was able to control metabolism of xylose effectively, there were still restrictions on how xylose was getting into the cell. To account for this, we constructed plasmids with the arabinose low affinity transporter *araE* and the xylose low affinity transporter *xylE*. Previous work has suggested that these transporters are not sugar specific [100]. Therefore, in the absence of arabinose, *araE* should be able to transport xylose and hence it is worth testing this transporter along with the native transporter for xylose.

The results showed a significant increase in co-utilization of the two sugars (Figure 3.10 shows the growth curve for the strain with the *araE* transporter and Figure 3.11 shows the growth for the strain with the *xylE* transporter). With the addition of a transporter, the diauxic shift was completely eliminated and the engineered strain was able to grow faster than the wild-type strain. Where the wild-type *E. coli* strain was lagging during consumption of the individual sugars, the engineered strain was growing continuously and faster.

The analysis of the individual sugars was less straightforward. Growth in glucose was comparable to the wild-type cells for both the CR1264::pAraE and CR1264::pXylE strains (Figures 3.12 and 3.13). The growth for the strain with the *araE* transporter was not as fast as the wild-type strain, but this was due to the fact that there was a discrepancy in the OD's of the overnight cultures before subculturing; the OD of the wild-type culture was higher than that of the engineered strain. Growth in xylose of the engineered strain was definitely improved with the addition of the transporters, but not restored to wild-type levels (Figures 3.14 and 3.15). In order to makes sure that the lack of the diauxic shift was not just a delayed diauxic shift, OD measurements were taken well beyond 24 hours and up to 48 hours to confirm that the stationary phase was in fact reached.

HPLC analysis confirmed what we observed with the growth curve experiments when the strains were grown in a glucose-xylose mixture (Figures 3.16 and 3.17). The sugars are simultaneously depleted, which suggests co-utilization of glucose and xylose by these strains. The individual sugars are consumed similarly to the wild-type strains, with glucose being depleted within four hours and xylose within five hours of growth. The discrepancy between these results and the growth curve results suggests that there may be more complicated pathways involved in growth of the cells that do not necessarily reflect sugar consumption directly.

The addition of the *araE* transporter, rather than the *xylE* transporter, proved to be better for growth on xylose. The addition of transporters to the engineered strain significantly increased co-utilization of glucose and xylose. The transporters were also individually integrated into the chromosome of the engineered strain, CR1264, to produce strains CR1265 (*araE* integrated) and CR1266 (*xylE* integrated). The transporters in this case were controlled by the constitutive $P_{LTetO-1}$ promoter and hence their expression was not inhibited by catabolite repression effects. The growth and sugar utilization in these strains was analyzed once again to determine which strain proved to be the most efficient in co-utilization. While both CR1265 and CR1266 showed improved co-utilization, the strain with an *araE* transporter showed better growth on xylose. Strains CR1267 and CR1268 were also tested to confirm that the attachment sites for integration don't affect the results that we see. The results for CR1267 are reported. Figures 3.18, 3.19 and 3.20 show the growth curves for glucose only, xylose only and glucose-xylose sugar mixtures respectively. HPLC analysis confirmed improved co-utilization for this engineered strain (Figure 3.21).

Co-utilization was observed even in anaerobic conditions. To confirm that our strain was not exhibiting favorable properties only in conditions in which oxygen was available, we repeated the experiments in purely anaerobic conditions. Experiments were carried out in

sealed Erlenmeyer shake flasks (with a side spout) that were sparged with nitrogen prior to the start of the experiment. Time-point OD measurements were taken as described before except that the samples were drawn from the side spout using a sterile BD 5mL syringe with a disposable needle (Luer-Lok™ Tip with BD Precision Glide™ Needle 22G x 1 ½ (0.7mm 40mm)) so as not to allow oxygen to enter the sealed flask. As expected, in anaerobic conditions the growth is affected and hence the OD's were all lower than before for the same time period. However, the strain still displayed improved co-utilization properties in a glucose-xylose sugar mixture. It can be seen from Figure 3.22 that the diauxic shift is absent in both oxygen limiting (or micro-aerobic) and anaerobic conditions. Additionally, we can see from Figure 3.23, that the sugar utilization is also similar in anaerobic conditions. Xylose is consumed simultaneously with glucose in a glucose-xylose mixture and is consumed in exactly the same time period as when the strain is grown in micro-aerobic conditions. This is an important result as it verifies that our engineered strain can be grown in an anaerobic environment (which is typical of a fermentation reactor) and still utilize sugars simultaneously.

Adaptation of the engineered strain to xylose did not improve xylose utilization. In an effort to improve the growth of the strain CR1267 on xylose, we applied an evolutionary approach. Cells were subcultured every day for three weeks in fresh minimal media supplemented with 20 mM xylose. The hope was that the cells would eventually adapt in a xylose environment and their growth would be improved on xylose as the sole substrate. After two weeks of adaptation, it was observed that the cultures started to become dense faster than before and that growth on xylose was definitely improved, with final OD's much higher than the strain CR1267.

The challenge, however, was that the characteristics of this strain were not consistent from one day to the next. Growth curve experiments performed on different days resulted in inconsistent results. Results also varied between different colonies from the same plate. The second

problem was that even when the final OD was higher than the wild-type strain, the kinetic growth trend of this strain did not follow the exponential curve. Rather there was a long lag phase during which the cells did not grow at all and then suddenly grew rapidly (Figure 3.24). HPLC analysis does not explain this lag in growth. According to the sugar utilization analysis, the xylose seems to be depleting as fast as in the wild-type strain. The adapted strain would have to be sequenced and the mutations analyzed to determine what exactly is causing the variation in properties and whether xylose utilization in this strain is in fact improved.

3.4 Discussion

We have shown that by replacing the native *xylA* promoter with a constitutive promoter that is not restricted by catabolite repression effects, we can significantly improve co-utilization of glucose and xylose in *E. coli*.

As previously mentioned in the introductory chapter, there are several factors affecting efficient utilization of a mixture of sugars. A combination of factors involving metabolism, transport and regulation is the key to achieving efficient co-utilization. Replacement of the *xylA* promoter with a constitutive promoter attacked the problem of negative regulation of this operon in the presence of glucose and ensured metabolism of xylose. However, this strain's efficiency was limited by transport of xylose into the cell. This was fixed by introducing transporters such as *araE* and *xylE*, under the control of a strong promoter, into the cell. While this solution significantly improved co-utilization, growth on xylose as the sole carbon source was still not restored. This suggests that there are other factors that come into play whenever changes are introduced into a metabolic pathway. The exact responses to these changes are unknown but the effect is noticed when the strain is grown on xylose as the only substrate. When the strains are grown in a mixture of glucose and xylose, however, the properties of the engineered strain meet our goals and more efficient utilization of the sugar mixture takes place. We do not know

for a fact whether these positive changes are because catabolite repression effects are overcome as this has not been directly tested. It is, however, our hypothesis that the replacement of the *xylA* promoter with a constitutively “on” promoter, which is unaffected by catabolite repression responses of the cell, helps to express the metabolic genes of xylose even in the presence of glucose.

The addition of the *araE* transporter helped the engineered strain grow on xylose better than the addition of the *xylE* transporter. However, HPLC analysis of these two strains showed that the sugars were consumed in exactly the same way; xylose did not seem to be utilized faster in the CR1264::pAraE strain. The reason for this discrepancy is not known.

Another important question to ask is whether there are detrimental effects to the bacteria with the introduction of these metabolic changes. For example, the increase in metabolism of sugars leads to a build-up of sugar phosphates and hence a stronger drive towards the methylglyoxal pathway. The methylglyoxal pathway diverts carbon from the lower branch of glycolysis to produce methylglyoxal. This diversion helps bacterial cells maintain the rate of carbon flux by temporarily relieving the stress caused by phosphorylated intermediates (such as sugar phosphates) and allows the cells to grow for a limited time until phosphate balance is restored. However, in the scenario that the flux balance is continually not maintained by the cell, the production of methylglyoxal continues to dangerous levels, causing cell death. If this is the case, it is important to investigate whether causing mutations to the cell to improve co-utilization also causes toxicity.

The success of this approach opens doors to more possibilities in engineering a more efficient strain for co-utilizing sugars. The methodology can be used to solve the problem of the glucose-arabinose sequential utilization as well. A similar replacement of the *araB* promoter with a constitutive promoter could lead to better co-utilization of glucose and arabinose and eventually

all three sugars prominently present in hemicellulose. Furthermore, this engineering approach can also be extended to improve the transport of pentose sugars into the cell.

The final construction of our engineered strain that showed the most efficiency in co-utilization of glucose and xylose is CR1267. To summarize, this strain has the native *xylA* promoter replaced with a constitutive promoter ($P_{LtetO-1}$) and an araE transporter that is controlled by a strong promoter $P_{LtetO-1}$ integrated at the *ϕ80* attachment site of the *E. coli* chromosome.

3.5 Figures and tables

Table 3.1. Bacterial strains used in this work.

Strain	Genotype or relevant characteristics[a]	Source or reference[b,c]
MG1655	$F^- \lambda^-$ *ilvG rph-1*	CGSC #7740
DH5α	*phi-80d lac*Δm15 *enda1 recA1 hsdR17 supE44 thh*-1 *gyrA96 relA*Δ*lacU169*	New England Biolabs
CR1263	ΔP_{xylA}::*tetRA*	
CR1264	ΔP_{xylA}::$P_{LtetO-1}$	
CR1265	ΔP_{xylA}::$P_{LtetO-1}$ *attHK022*::pKK048	
CR1266	ΔP_{xylA}::$P_{LtetO-1}$ *attHK022*::pKK049	
CR1267	ΔP_{xylA}::$P_{LtetO-1}$ *att ϕ80*::pKK050	
CR1268	ΔP_{xylA}::$P_{LtetO-1}$ *att ϕ80*::pKK051	
CR1269	ΔP_{xylA}::$P_{LtetO-1}$ *att ϕ80*::pKK050 *attHK022*::pKK052	

Table 3.2. Plasmids used in this work.

Plasmids	Genotype or relevant characteristics[a]	Source or reference[b,c]
pKD46	*bla* P_{BAD} *gam bet exo* pSC101 *ori(ts)*	
pAH143K	*kan MCS tL3 rgnB attHK022 oriR6K*	Lon Chubiz
pAH153C	*cm MCS tL3 rgnB attϕ80 oriR6K*	Lon Chubiz
pAH69	*bla* Int_{HK022} *oriR6K*	
pAH123	*bla Int ϕ80 oriR6K*	
pPDC AdhB	*cm* $P_{LtetO-1}$ *pdc adhB ori* ColE1	Kori Dunn
pPROTet.E	*cm* $P_{LtetO-1}$ *ori* ColE1	Stratagene
pKK046 (pAraE)	*cm* $P_{LtetO-1}$ *araE ori* ColE1	
pKK047 (pXylE)	*cm* $P_{LtetO-1}$ *xylE ori* ColE1	
pKK048	*kan* $P_{LtetO-1}$ *araE tL3 rgnB attHK022 oriR6K*	
pKK049	*kan* $P_{LtetO-1}$ *xylE tL3 rgnB attHK022 oriR6K*	
pKK050	*cm* $P_{LtetO-1}$ *araE tL3 rgnB attϕ80 oriR6K*	
pKK051	*cm* $P_{LtetO-1}$ *xylE tL3 rgnB attϕ80 oriR6K*	
pKK052	*kan* $P_{LtetO-1}$ *pdc adhB tL3 rgnB attHK022 oriR6K*	

a. All strains are isogenic derivatives of *E. coli* K-12 strain MG1655.
b. All strains and plasmids are from this work unless otherwise noted.
c. *E. coli* Genetic Stock Center, CGSC, Yale University.

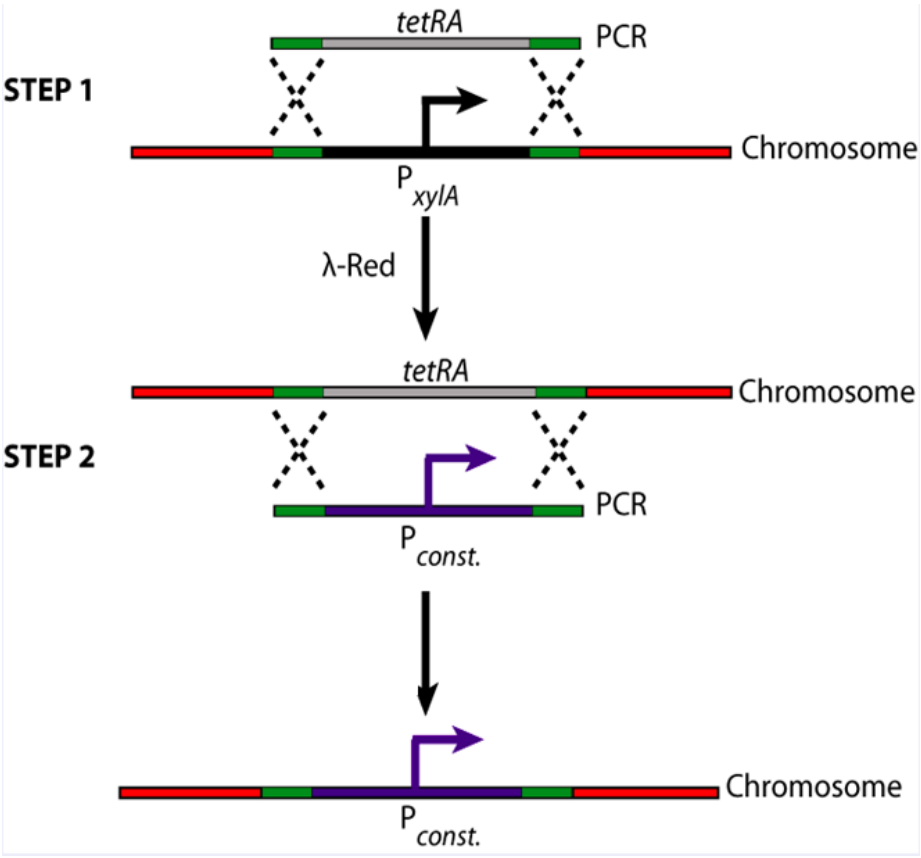

Figure 3.1. Approach for the replacement of the *xylA* promoter with a constitutive promoter.

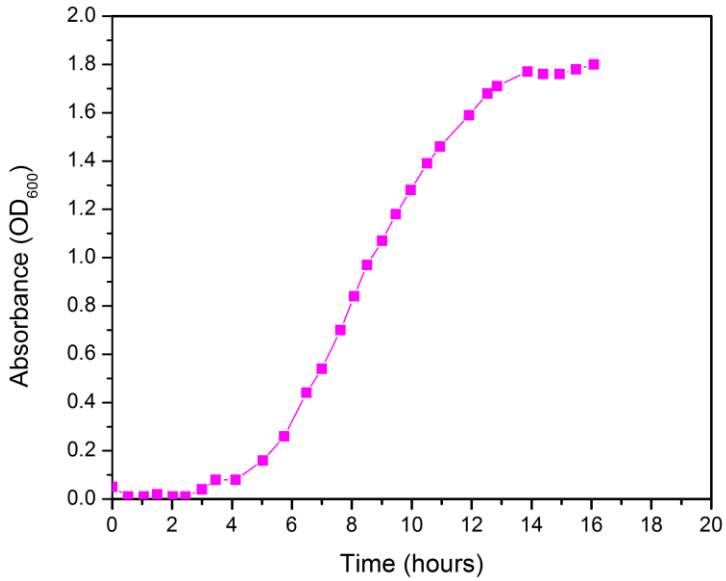

Figure 3.2. Growth curve of wild-type in glucose. Cells were grown in M9 with 20 mM glucose as the sole carbon source.

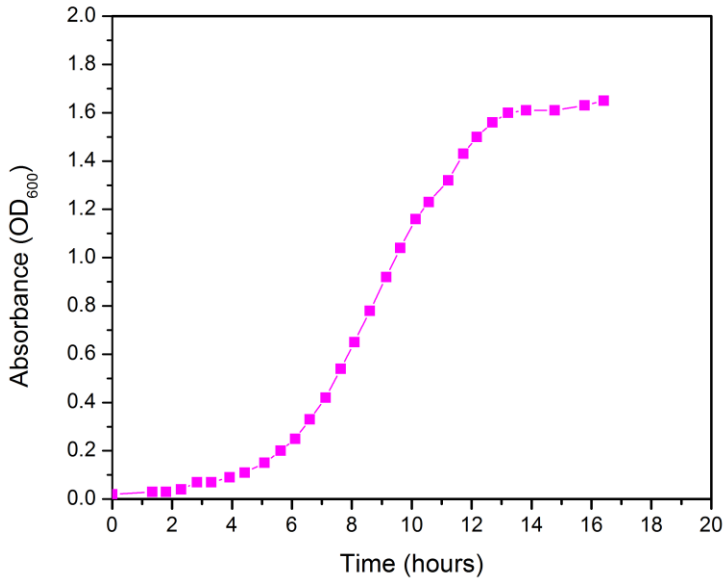

Figure 3.3. Growth curve of wild-type in xylose. Cells were grown in M9 with 20 mM xylose as the sole carbon source.

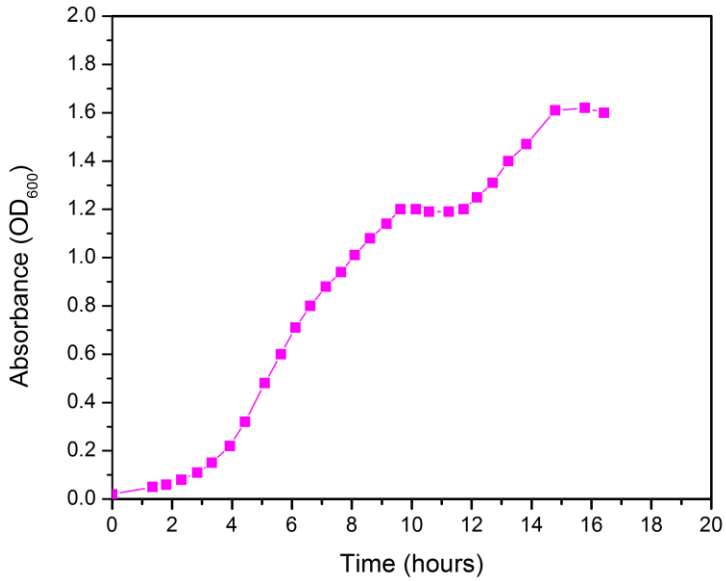

Figure 3.4. Growth curve of wild-type in a mixture of glucose and xylose. Cells were grown in M9 with glucose (10 mM) and xylose (10 mM).

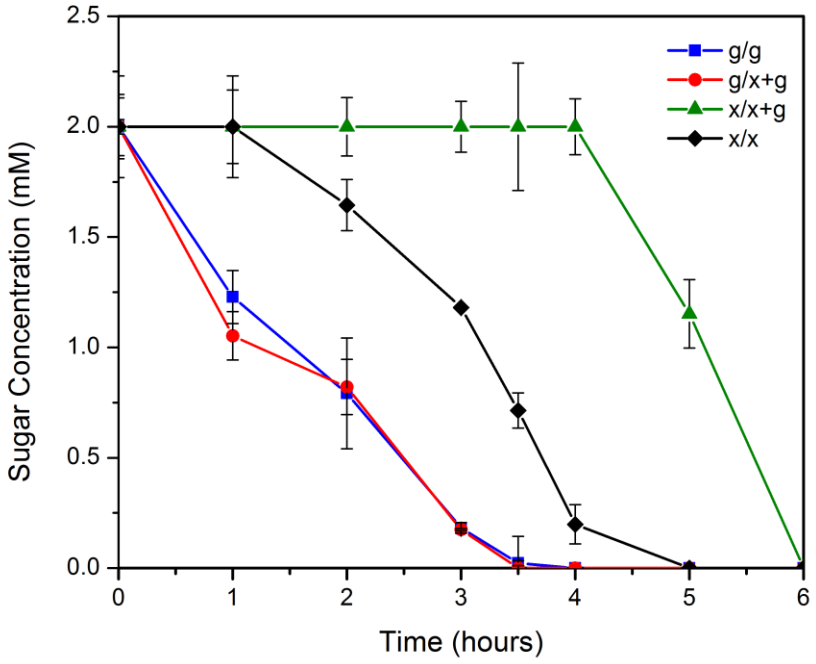

Figure 3.5. Sugar utilization in wild-type (HPLC analysis). Cells were supplemented with 2 mM sugar concentrations and allowed to grow for eight hours.

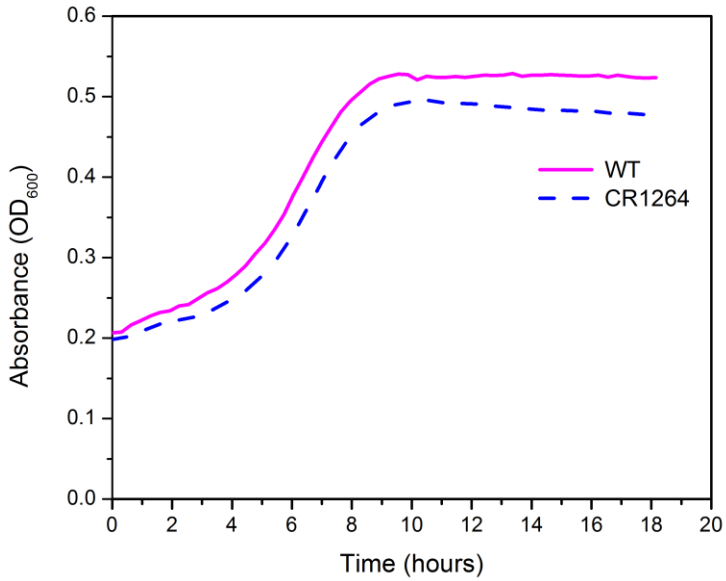

Figure 3.6. Growth curve of the engineered strain (CR1264) and wild-type (WT) in glucose. Cells were grown in M9 with 20 mM glucose as the sole carbon source.

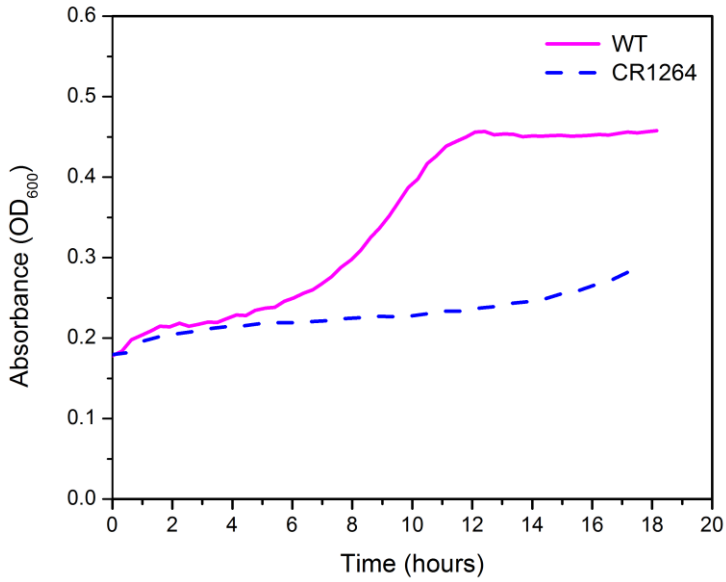

Figure 3.7. Growth curve of the engineered strain (CR1264) in xylose. Cells were grown in M9 with 20 mM xylose as the sole carbon source.

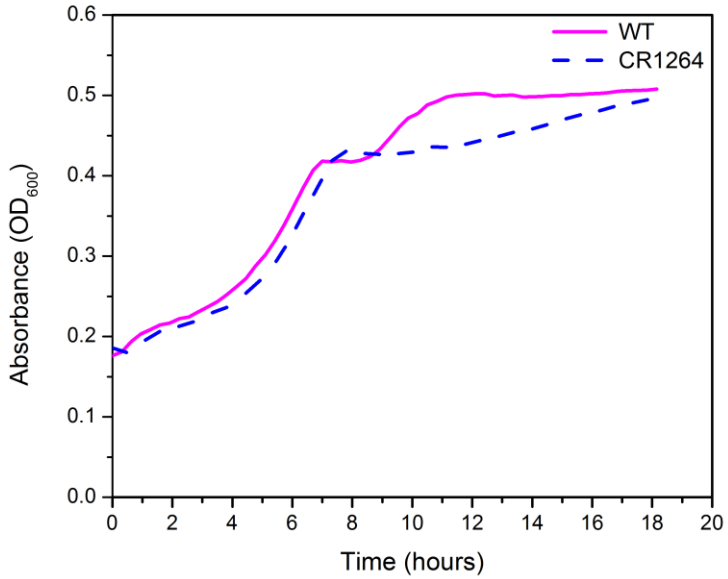

Figure 3.8. Growth curve of the engineered strain (CR1264) in a mixture of glucose and xylose. Cells were grown in M9 with glucose (10 mM) and xylose (10 mM).

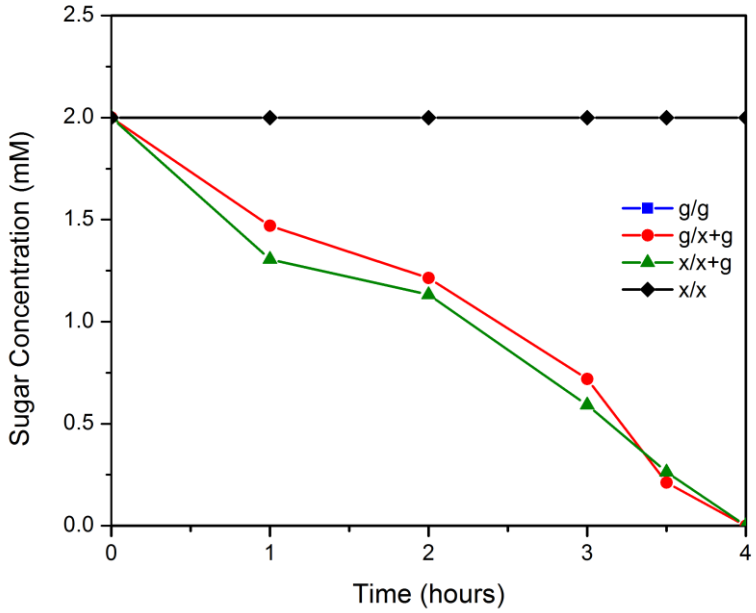

Figure 3.9. Sugar utilization in the engineered strain (CR1264). Cells were supplemented with 2 mM sugar concentrations and allowed to grow for eight hours.

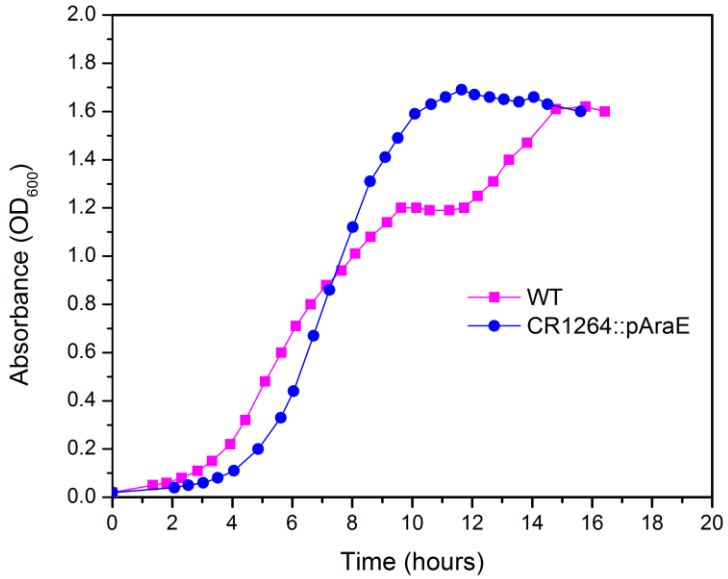

Figure 3.10. Growth curve of the engineered strain with pAraE (glucose + xylose). Cells were grown in M9 with 10 mM glucose and 10 mM xylose.

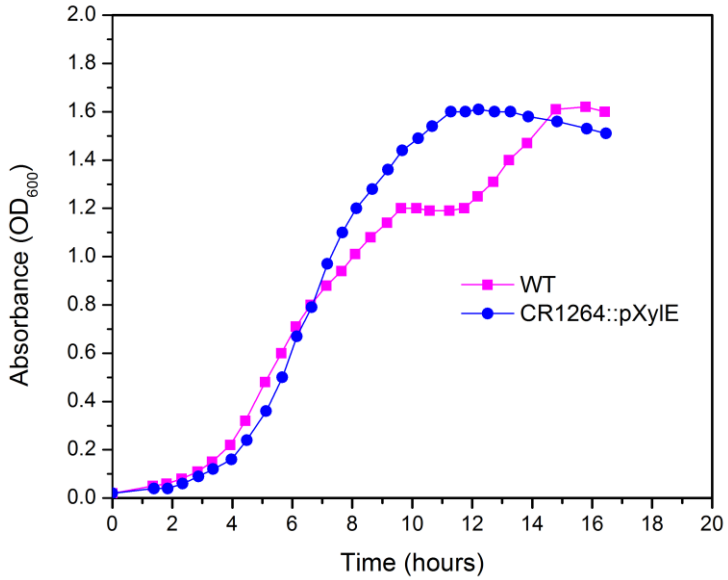

Figure 3.11. Growth curve of the engineered strain with pXylE (glucose + xylose). Cells were grown in M9 with 10 mM glucose and 10 mM xylose.

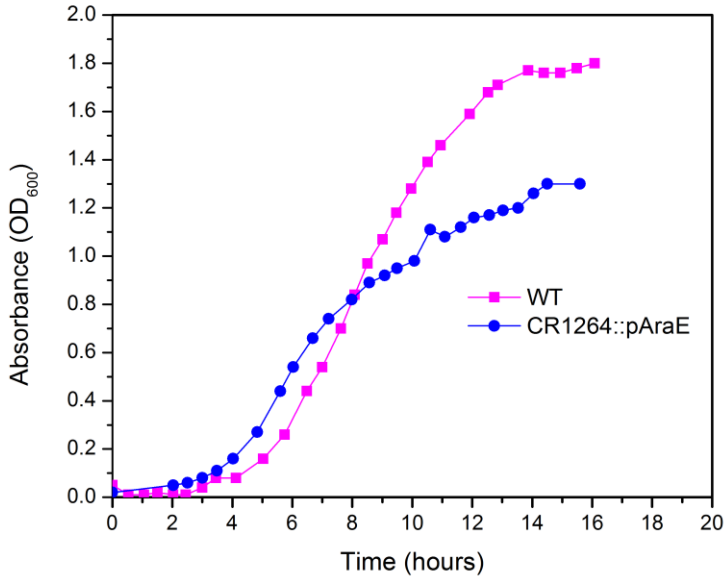

Figure 3.12. Growth curve of the engineered strain with pAraE (glucose). Cells were grown in M9 with 20 mM glucose as the sole carbon source.

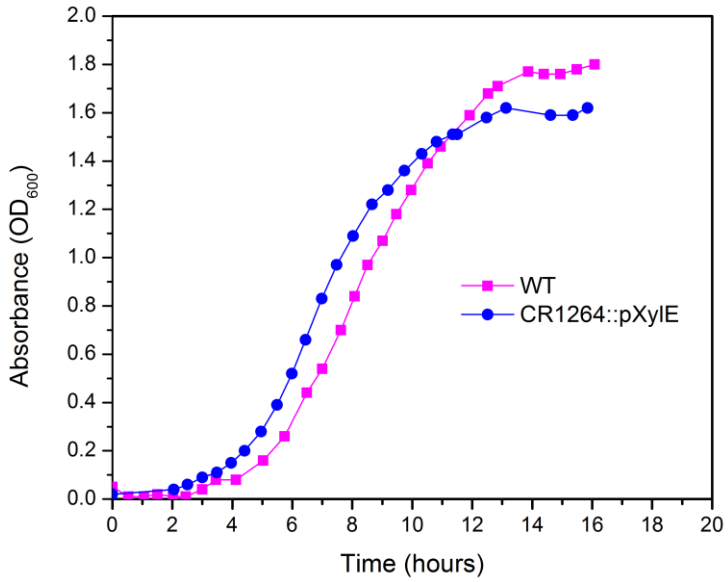

Figure 3.13. Growth curve of the engineered strain with pXylE (glucose). Cells were grown in M9 with 20 mM glucose as the sole carbon source.

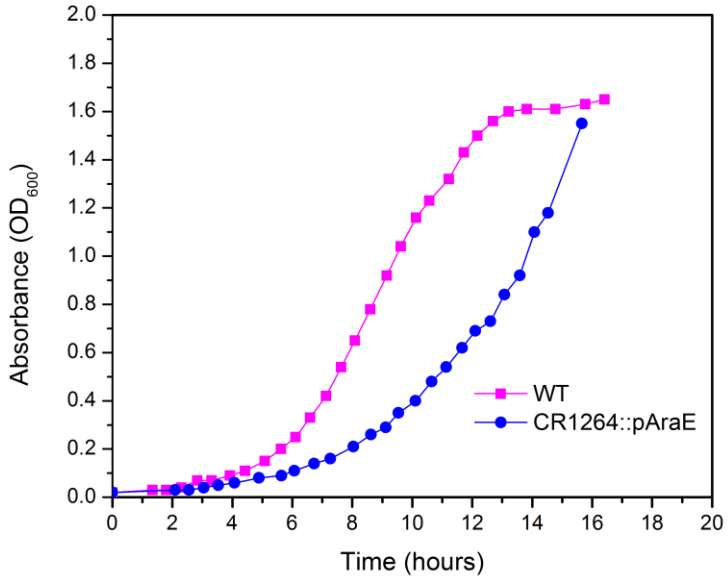

Figure 3.14. Growth curve of the engineered strain with pAraE (xylose). Cells were grown in M9 with 20 mM xylose as the sole carbon source.

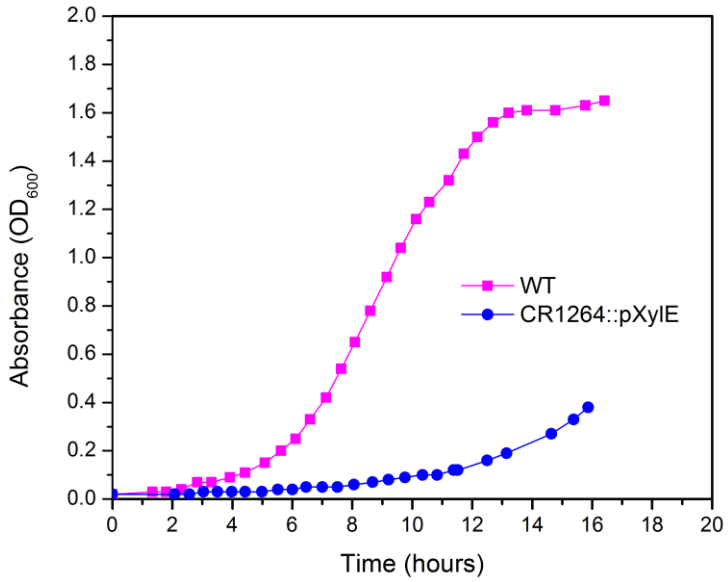

Figure 3.15. Growth curve of the engineered strain with pXylE (xylose). Cells were grown in M9 with 20 mM xylose as the sole carbon source.

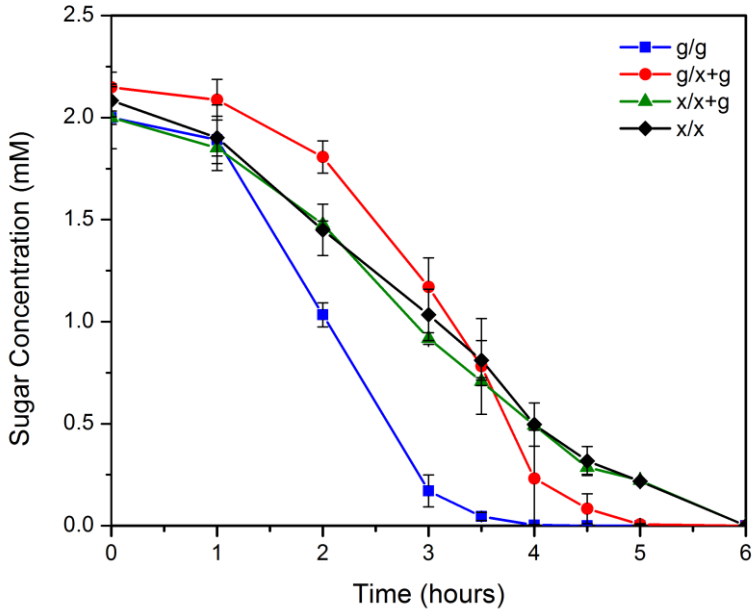

Figure 3.16. Sugar utilization in strain CR1264::pAraE. Cells were supplemented with 2 mM sugar concentrations and allowed to grow for eight hours.

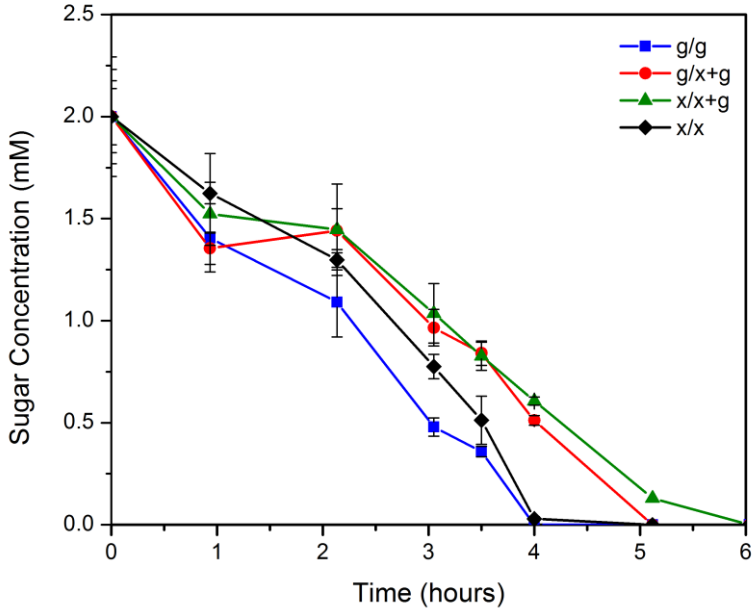

Figure 3.17. Sugar utilization in strain CR1264::pXylE. Cells were supplemented with 2 mM sugar concentrations and allowed to grow for eight hours.

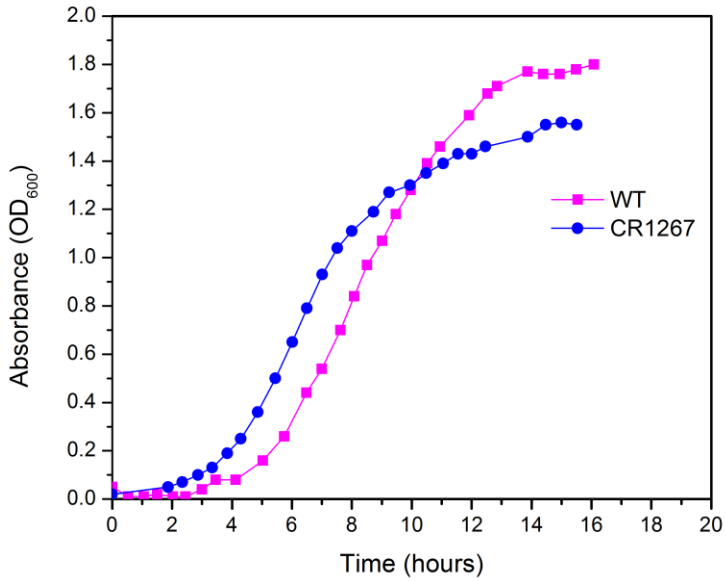

Figure 3.18. Growth curve of the engineered strain with the *araE* transporter integrated (glucose). Cells were grown in M9 with 20 mM glucose as the sole carbon source.

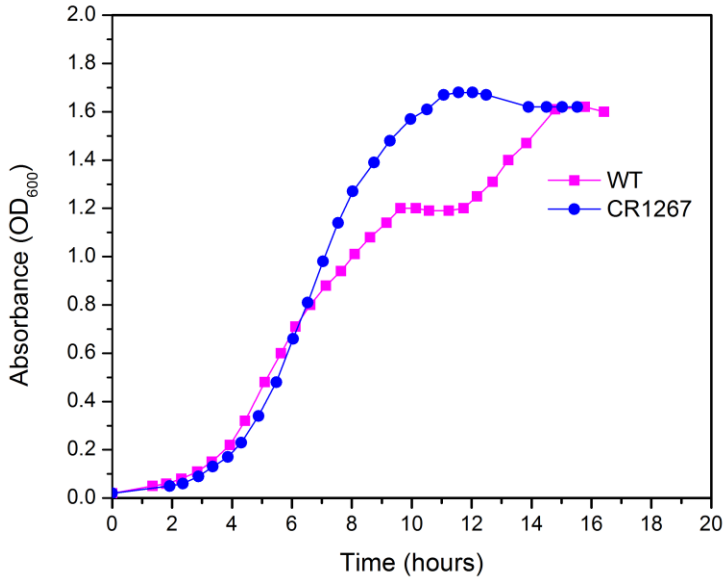

Figure 3.19. Growth curve of the engineered strain with the *araE* transporter integrated (glucose + xylose). Cells were grown in M9 with glucose (10 mM) and xylose (10 mM).

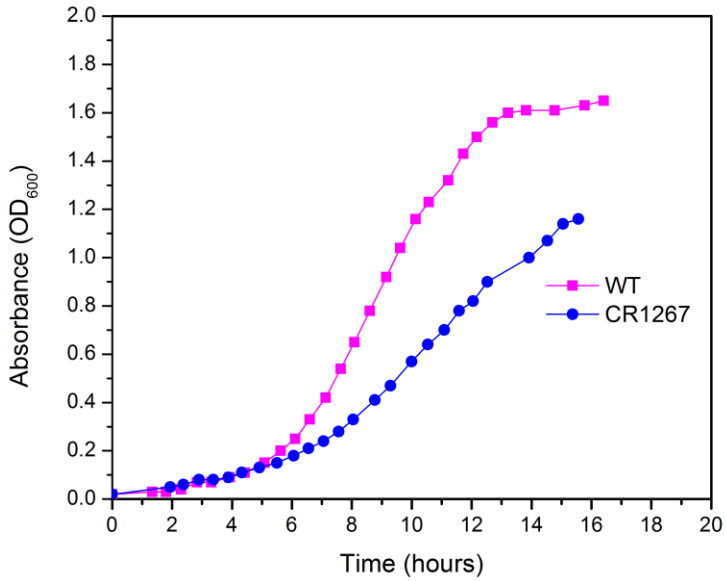

Figure 3.20. Growth curve of the engineered strain with the *araE* transporter integrated (xylose). Cells were grown in M9 with 20 mM xylose as the sole carbon source.

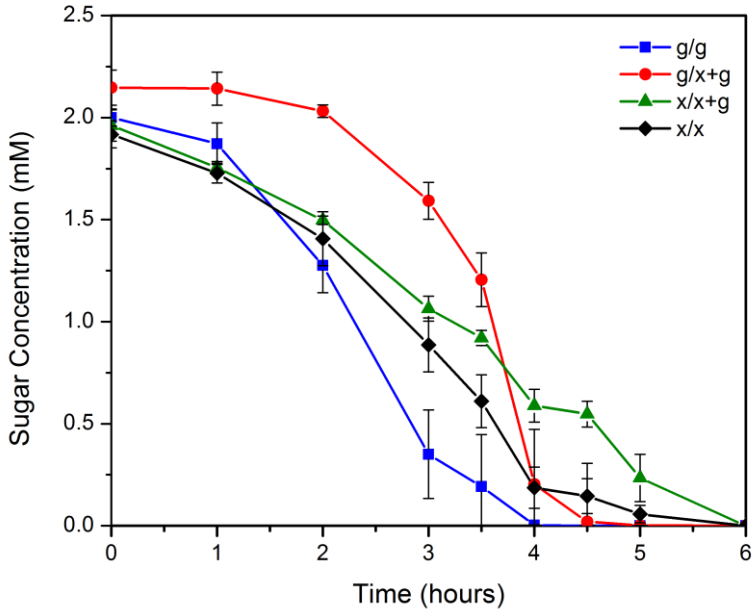

Figure 3.21. Sugar utilization in CR1267. Cells were supplemented with 2 mM sugar concentrations and allowed to grow for eight hours.

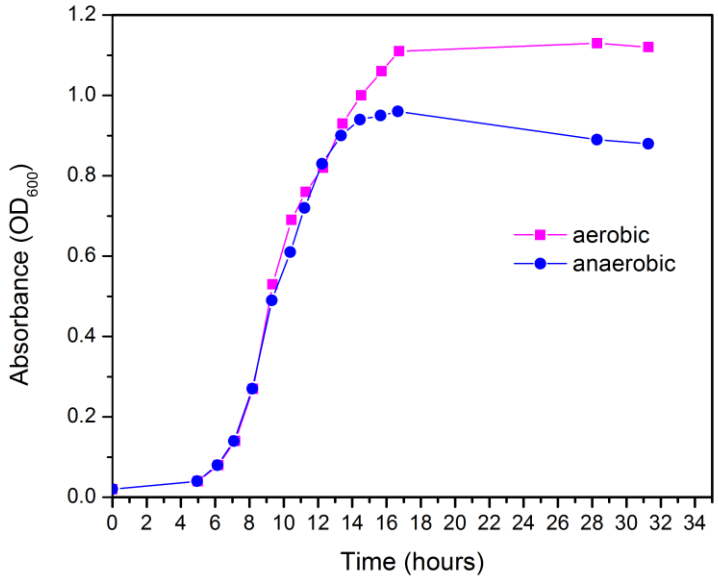

Figure 3.22. Growth curve of the CR1267 strain in micro-aerobic and anaerobic conditions. Cells were grown in M9 with 10 mM glucose and 10 mM xylose.

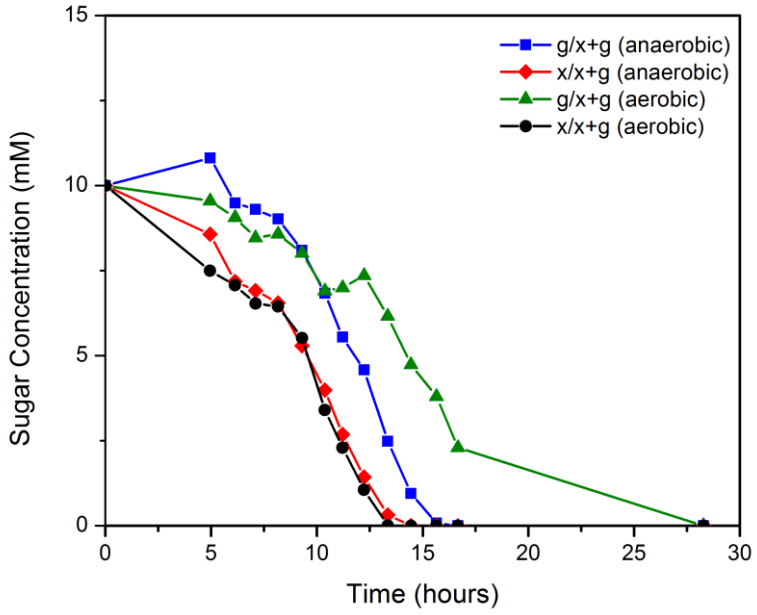

Figure 3.23. Sugar utilization in strain CR1267 under micro-aerobic and anaerobic conditions. Cells were supplemented with 10 mM glucose and 10mM xylose and allowed to grow until stationary phase.

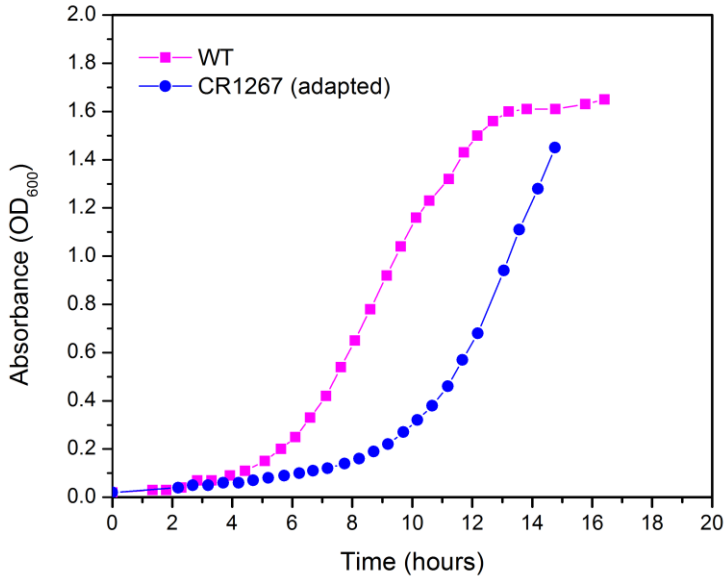

Figure 3.24. Growth curve of the adapted CR1267 strain in xylose. Cells were grown in M9 with 20 mM xylose as the sole carbon source.

CHAPTER 4: Further engineering of strain CR1267 for analysis of ethanol production under anaerobic conditions

4.1 Introduction

The description of a strain that is more efficient than wild-type *E. coli* in co-utilizing glucose and xylose simultaneously was given in the previous chapter. It was concluded that strain CR1267 showed the most co-utilizing efficiency as shown by growth curve experiments and sugar utilization (HPLC) analysis. This strain had the native *xylA* promoter replaced with a constitutive promoter and was supplemented with the *araE* transporter.

In this chapter, we describe how the strain CR1267 was engineered further to optimize it for ethanol production. Two main changes to the engineered strain and the experimental conditions were made. First, in order to maximize ethanol production we introduced the genes *pdc (*pyruvate decarboxylase) and *adhB* (alcohol dehydrogenase II) from *Zymomonas mobilis* [101, 102] to divert the production of organic acids to ethanol as the principal fermentation product in *E. coli* [103]. Secondly, all experiments were carried out in a purely anaerobic environment to emulate fermentation conditions.

With the addition of these changes, we repeated the growth experiments to observe anaerobic growth. Previously, we had monitored the sugar utilization of our engineered strain for eight hours only and started with a sugar concentration of only 2 mM. In these experiments, however, we monitored sugar utilization of the strain throughout its growth period in the 20 mM concentrated sugar media. We also analyzed the ethanol and acetate concentrations in the wild-type cultures during the growth period to quantify ethanol production. This serves as a preliminary step to allow for further analysis for our engineered strain in the future.

4.2 Materials and methods

Media and growth conditions as well as strain and plasmid construction is described in the previous chapter. The strain and plasmids specifically used in this study are CR1267, CR1269, pPDC AdhB, and pKK052 (Tables 3.1 and 3.2).

Anaerobic growth experiments. Single colonies were picked from freshly streaked plates and grown overnight in glass test-tubes containing 3 mL of minimal medium (M9) and 20mM glucose. The cultures were then inoculated in pre-sterilized 250 mL Erlenmeyer shake flasks (with a side spout) containing 150 mL of M9 and 20mM total concentration of sugars. The initial OD_{600} after inoculation was set to 0.02. To maintain anaerobic growth conditions, nitrogen was sparged into the flasks through a 0.22 µm pore size filter for 15 minutes and sealed with a rubber stopper at both openings. The temperature and agitation rate were set at 37°C and 500 rpm respectively. Five milliliters of a culture was withdrawn periodically from the shake flasks' side spouts using a sterile BD 5mL syringe with a disposable needle (Luer-Lok™ Tip with BD Precision Glide™ Needle 22G x 1 ½ (0.7mm 40mm)). 1mL of the withdrawn culture was transferred to a 1.5 cm cuvette and the OD_{600} was measured using a spectrophotometer. The rest of the withdrawn culture was used for ethanol and sugar concentration analysis. All experiments were conducted in triplicates and average values with the standard deviations are reported.

Analytical measurements. To measure sugar concentrations, samples were first filtered using a 0.22 µm pore size filter (Millipore) and diluted 50% with deionized, autoclaved water before running them on the HPLC. The dilution step was necessary because the chromatogram peaks reach a maximum at 10mM sugar concentration due to software limitations. Operating conditions were the same as described in the previous chapter. The chromatogram peaks were analyzed using the Gilson Safire software.

Ethanol and acetate concentrations were determined by GC/MS at the metabolomics center of the University of Illinois at Urbana-Champaign. The culture samples were first filtered using a 0.22 μm pore size filter prior to submission to the metabolomics center. The following protocol is provided by Alexander Ulanov from the metabolomics center. Samples (1 μL) were injected in split mode (5:1) to the GC/MS system consisting of an Agilent 7890 gas chromatograph, an Agilent 5975 mass selective detector and HP 7683B (Agilent Inc, Palo Alto, CA, USA) autosampler. Analysis was done on a 30 m ZB-WAX*plus* column with 0.32 mm I.D. and 0.25 μm film thickness (Phenomenex, Torrance, CA, USA) with an injection port temperature of 200^0C. The interface was set to 200^0C, and the ion source adjusted to 230^0C. The helium carrier gas was set at a constant flow rate of 3 mL/min. The temperature program was as follows: 6 minutes isothermal heating at 55^0C, followed by an oven temperature increase of 20^0C/min to 170^0C for 0 minutes and then 10^0C/min to 220^0C for 2 minutes. The mass spectrometer was operated in positive electron impact mode (EI) at 69.9 eV ionization energy in m/z 25-300 scan range.

The spectra of all chromatogram peaks were evaluated using the HP Chemstation (Agilent, Palo Alto, CA, USA) program. Identification and quantification of samples was performed using the mass spectra obtained from the authentic standards and additionally confirmed with NIST08 and W8N08 libraries (John *Wiley* & Sons, *Inc*., USA).

4.3 Results

Confirming simultaneous co-utilization of the engineered strain in anaerobic conditions. Before the *pdc* and *adhB* genes were introduced in our engineered strain, we wanted to verify that the simultaneous co-utilization of glucose and xylose is still observed for the specific experimental conditions described above. The engineered strain, therefore, was grown under anaerobic conditions and the extracellular sugar concentrations were measured over the course of growth and compared to wild-type strains grown under the same conditions. Figure 4.1 shows

that the engineered strain still consumes the two sugars simultaneously. This is a vast improvement to the wild-type strain, where the depletion of xylose remains stationary until all the glucose in the culture has been utilized. This result confirms that our strain still retains its desired properties under the conditions that will be required to test for ethanol yields.

Ethanol production in wild-type *E. coli*. In order to have a model to compare our results with, we first carried out measurements for the wild-type *E. coli* strain without any mutations or plasmids in anaerobic conditions. The OD growth curves and sugar utilization results are similar to that obtained under aerobic or oxygen limiting conditions as described in Chapter 3 and are therefore not shown here. The end-point ethanol and acetate production for different sugar mixtures are reported in Table 4.1. Ethanol production values cannot be compared with previous studies in the literature because of variations in experimental conditions and fermentation reactor volumes. Hence, it is important to have this reference standard in our study. Figure 4.2 shows how the ethanol production varies over time when the cells are grown in a glucose-xylose mixture. It can be seen from this graph that between 9.5 hrs and 11 hrs, the ethanol concentration is constant, hence indicating that the ethanol production is stagnant. This is in agreement with the diauxic shift observed between 9 hrs and 11 hrs in the growth curve when cells are grown in a glucose-xylose mixture (Figure 3.4).

Ethanol production in wild-type *E. coli* containing the plasmid pPDC AdhB. Following the experiments carried out with wild-type *E. coli*, we proceeded to repeat the experiments with the strain WT::pPDC AdhB (wild-type strain containing the pProtet.E plasmid with the *pdc* and *adhB* genes cloned). With the introduction of the *pdc* and *adhB* genes, the ethanol concentrations would be expected to increase and the acetate production would be decreased. The final optical densities of the cells and sugar utilization, however, would be expected to stay approximately the same. We confirmed that the growth is unaffected when the *pdc* and *adhB* genes are introduced in the wild-type strain and grown on glucose (Figure 4.3). Figures 4.4 and 4.5

illustrate the growth on xylose and a glucose-xylose mixture respectively. The growth on xylose is affected slightly and the cells reach lower optical densities than wild-type cells grown without the *pdc* and *adhB* genes. This result is evident in both figures 4.4 and 4.5. In a glucose-xylose mixture, the cells grow much more slowly after all the glucose has been consumed. The slow xylose consumption rate after the glucose is consumed is probably due to a combination of cellular inhibitory effects resulting from the formation of by-products such as succinic acid, lactic acid, acetic acid, ethanol, and formic acid. Additionally, when cells are grown in xylose as the sole carbon source, the final optical density reached is about 0.2 lower. HPLC analysis of this strain confirms that the strain utilizes glucose and xylose in a sequential manner (Figure 4.6). Note that only when all the glucose has been consumed, does the organism begin to utilize xylose. The beginning of xylose consumption occurs at about 11 hours of growth. This is in strong agreement with the growth curve data, where we see that at the 11^{th} hour, the cells begin to grow exponentially after a lag period of two hours (between 9-11 hrs of growth). Finally, analysis of the ethanol and acetate concentrations in the extracellular environment confirm that the addition of the *pdc* and *adhB* genes into the wild-type strain does in fact drive the production of ethanol as the principal fermentation product in *E. coli* (Table 4.2). The ethanol concentrations are higher than in the wild-type strain (an increase from 0.4 g/L to 2.1 g/L when grown in glucose) and the acetate production is reduced (from 30.9 mM to 13.2 mM when grown in glucose).

Ethanol production in wild-type *E. coli* with the *pdc* and *adhB* genes integrated. To avoid the addition of antibiotics to the fermentation process, we integrated the *pdc* and *adhB* genes into the chromosome under the control of the $P_{LTetO-1}$ promoter. The growth experiment, HPLC analysis and ethanol analysis were repeated with this strain to check for consistency. Figures 4.7, 4.8, and 4.9 illustrate the growth of this strain on glucose, glucose-xylose mixture, and xylose respectively. The growth is exactly the same as the strain WT::pPDC AdhB. The HPLC

results show that the sugar utilization of this strain is also consistent (Figure 4.10). Glucose was consumed within 16 hours of growth and xylose was consumed within 17 hours. In a mixture of glucose and xylose all the glucose was consumed within 10 hours of growth. During the time between 9 hrs and 11 hrs the xylose concentration was constant as no xylose is utilized by the cells. At about the 11^{th} hour of growth, xylose utilization was initiated and xylose was completely consumed after approximately 25 hours. Ethanol analysis showed that the ethanol and acetate production is similar to the strain WT::pPDC AdhB. These results verified that the integration of the ethanol producing genes, *pdc* and *adhB* did not affect the characteristics of the wild-type strain except for improving ethanol production.

Ethanol production in the engineered strain CR1269. Verification of the successful integration of the *pdc* and *adhB* genes into the wild-type strain shows that integration of these genes does not change the characteristics of the strain. This is important so that we can successfully incorporate these ethanol producing genes into our engineered strain without compromising its favorable characteristics. The final step, therefore, was to integrate the *pdc* and *adhB* genes into our engineered strain, CR1267, to produce the final strain CR1269. To test the performance of this strain, it is necessary to simulate fermentation conditions more accurately where anaerobic conditions, temperature, pH, and pressure are more tightly regulated.

4.4 Discussion

The design of the final engineered strain CR1269 is described pictorially in Figure 4.11. To summarize, this strain has the native *xylA* promoter replaced with a constitutive promoter ($P_{LtetO-1}$), an araE transporter that is controlled by a strong promoter $P_{LtetO-1}$ integrated at the *ϕ80* attachment site and the *pdc* and *adhB* genes integrated at the *HK022* site of the *E. coli* chromosome also controlled by the $P_{LtetO-1}$ promoter.

The addition of the *pdc* and *adhB* genes into our strains verified that the ethanol production can be increased in strains of *E. coli* and other mixed fermentation products (such as acetate) can be reduced. Furthermore, we proved that the engineered strain shows co-utilization of sugars even when grown in anaerobic conditions. This is important because it confirms that fermentation conditions do not affect the characteristics of our engineered strain. The results shown in this chapter help to verify that our engineered strain would be able to co-utilize sugars just as well in anaerobic conditions as in micro-aerobic conditions and with the aid of the ethanol producing genes *pdc* and *adhB* further increase in ethanol yields can be achieved.

4.5 Figures and tables

Table 4.1. End-point ethanol and acetate concentrations for wild-type *E. coli*.

Glucose		**Glucose-Xylose mixture**		**Xylose**	
Ethanol (g/L)	Acetate (mM)	Ethanol (g/L)	Acetate (mM)	Ethanol (g/L)	Acetate (mM)
0.4	30.9	0.3	22.9	0.3	24.2

Table 4.2. End-point ethanol and acetate concentrations for the strain WT::pPDC AdhB.

Glucose		**Xylose**	
Ethanol (g/L)	Acetate (mM)	Ethanol (g/L)	Acetate (mM)
2.1 ± 0.2	13.2 ± 2.1	1.9 ± 0.2	13.1 ± 1.9

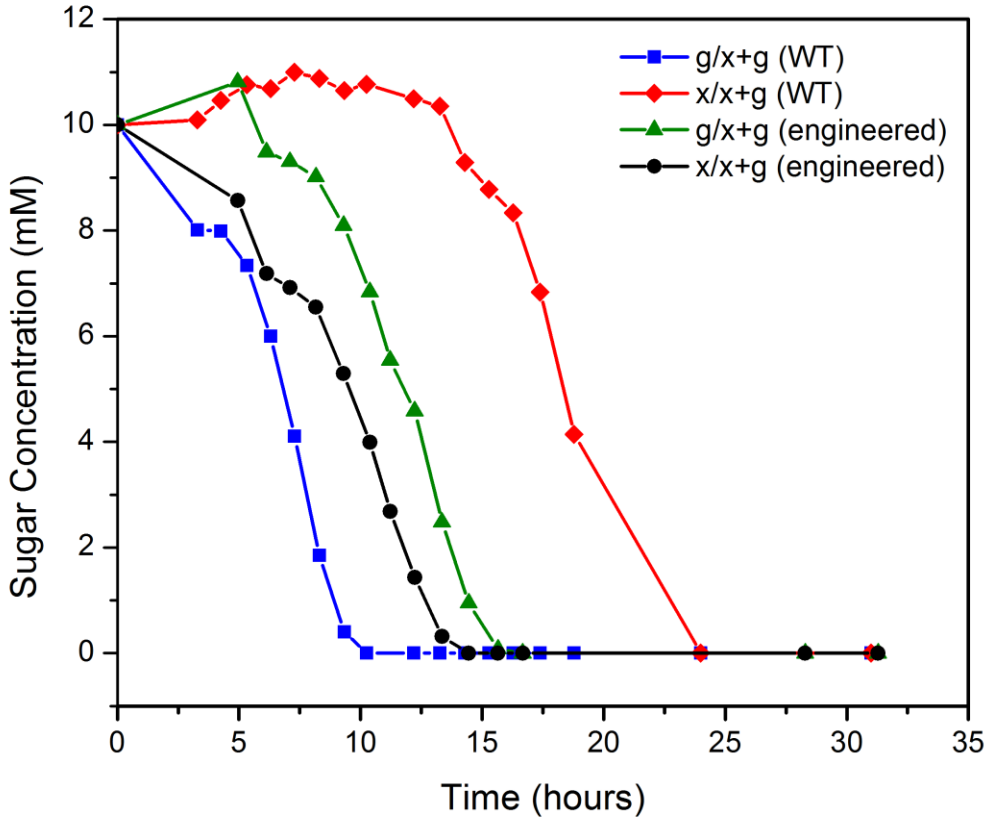

Figure 4.1. Sugar utilization of the engineered strain CR1267 versus the wild-type strain under anaerobic conditions. Cells were supplemented with 10 mM glucose and 10 mM xylose and allowed to grow until stationary phase was reached.

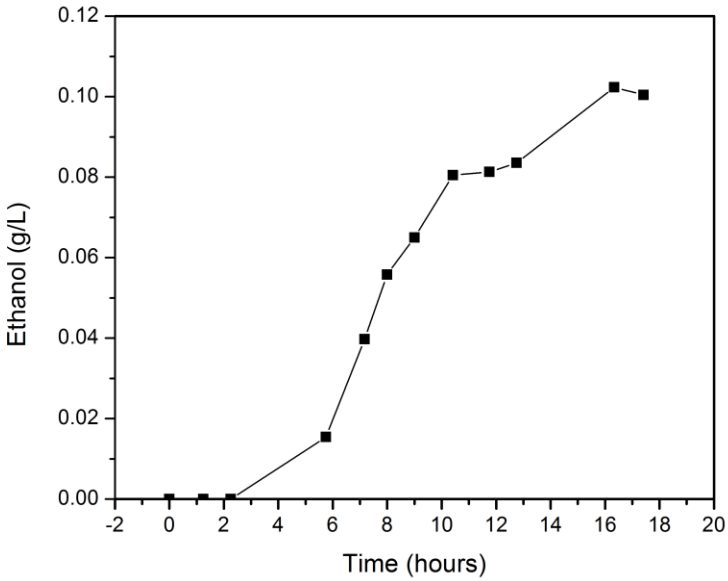

Figure 4.2. Production of ethanol over time in the wild-type strain. Cells were grown in M9 with glucose (10 mM) and xylose (10 mM). Samples were filter sterilized before ethanol was analyzed.

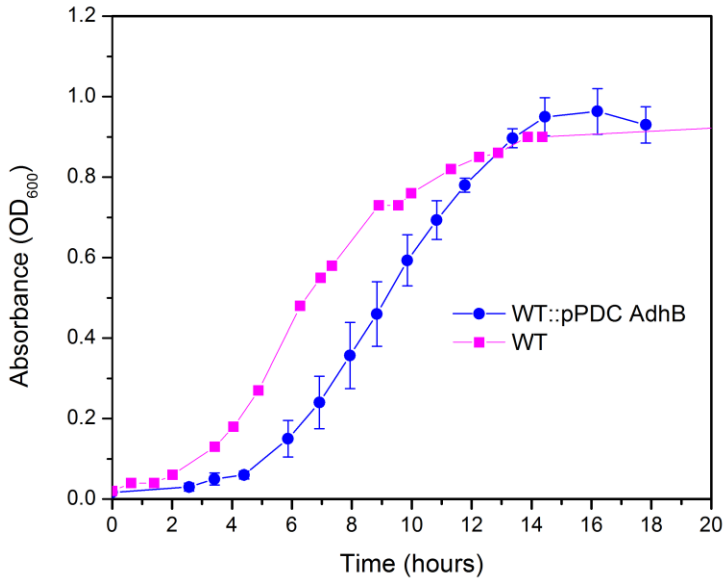

Figure 4.3. Growth curve of WT::pPDC AdhB in glucose. Cells were grown in M9 with 20 mM glucose as the sole carbon source.

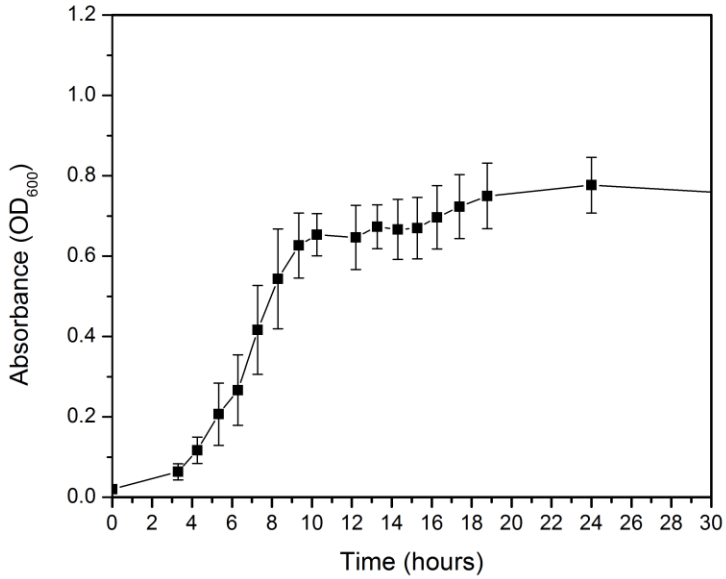

Figure 4.4. Growth curve of strain WT::pPDC AdhB in a mixture of glucose and xylose. Cells were grown in M9 with glucose (10 mM) and xylose (10 mM).

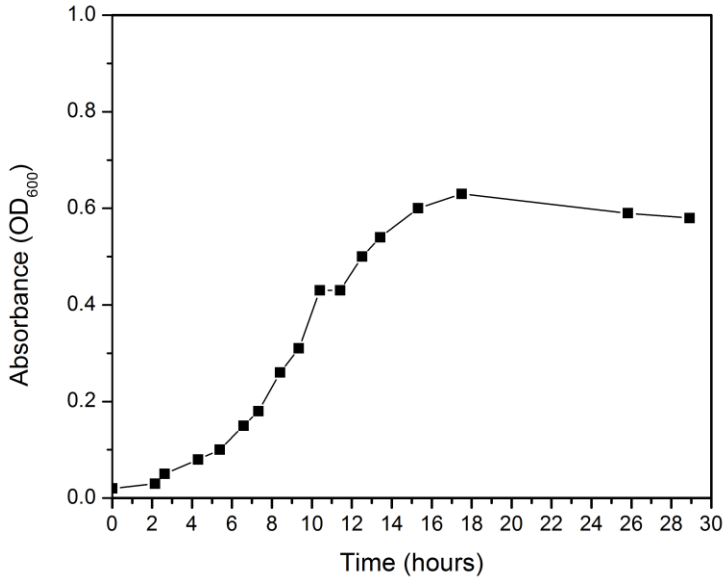

Figure 4.5. Growth curve of strain WT::pPDC AdhB in xylose. Cells were grown in M9 with 20 mM xylose as the sole carbon source.

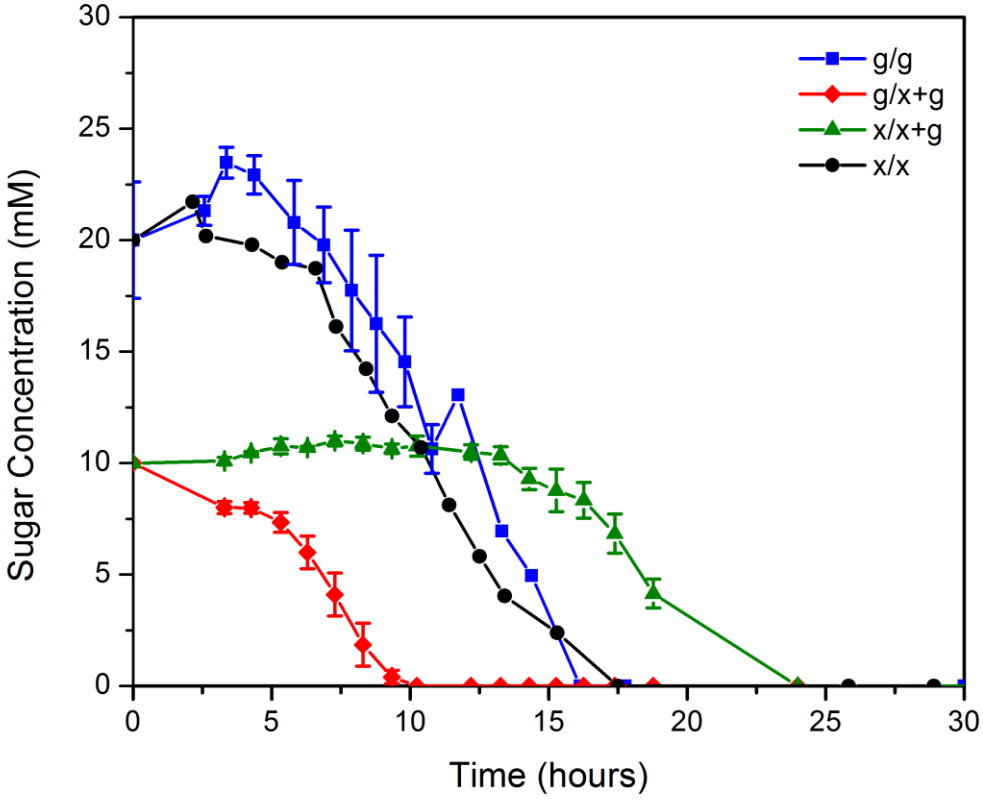

Figure 4.6. Sugar utilization in strain WT::pPDC AdhB. Cells were supplemented with 20 mM sugar concentrations (total) and allowed to grow until stationary phase was reached.

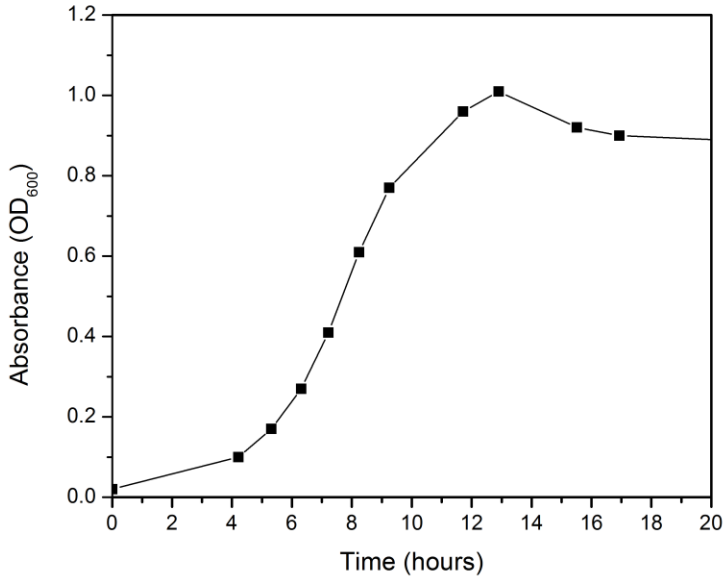

Figure 4.7. Growth curve of strain WT::pKK052 in glucose. Cells were grown in M9 with 20 mM glucose as the sole carbon source.

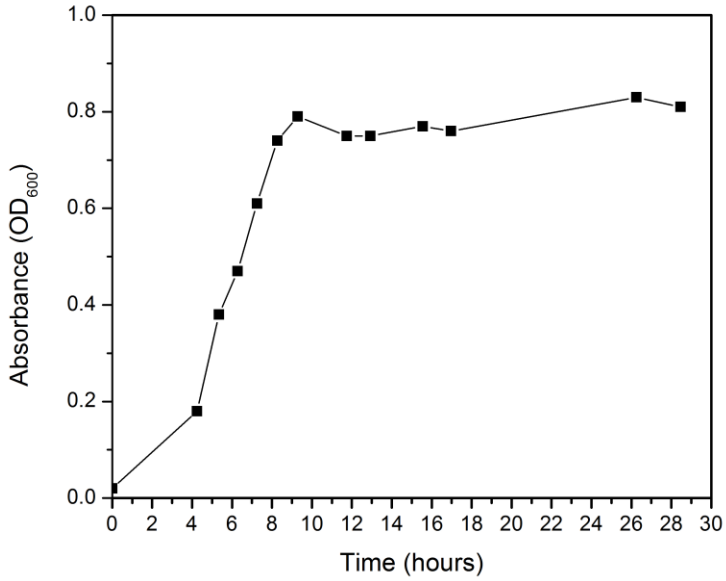

Figure 4.8. Growth curve of strain WT::pKK052 in a mixture of glucose and xylose. Cells were grown in M9 with glucose (10 mM) and xylose (10 mM).

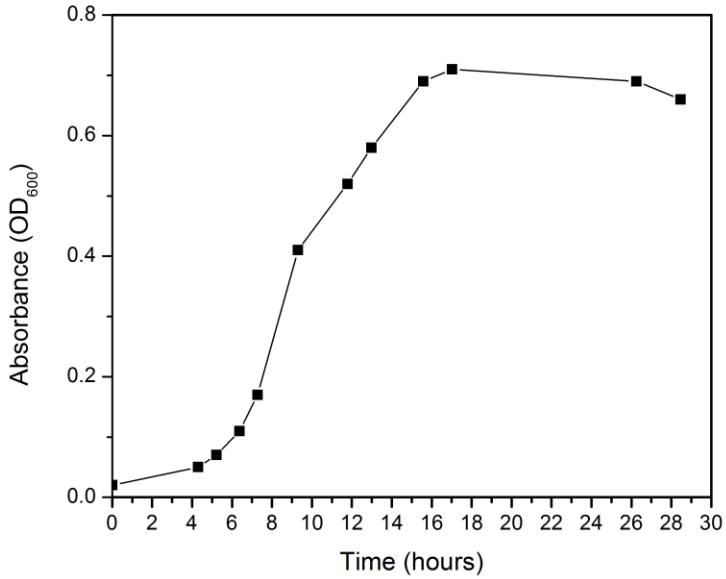

Figure 4.9. Growth curve of strain WT::pKK052 in xylose. Cells were grown in M9 with 20 mM xylose as the sole carbon source.

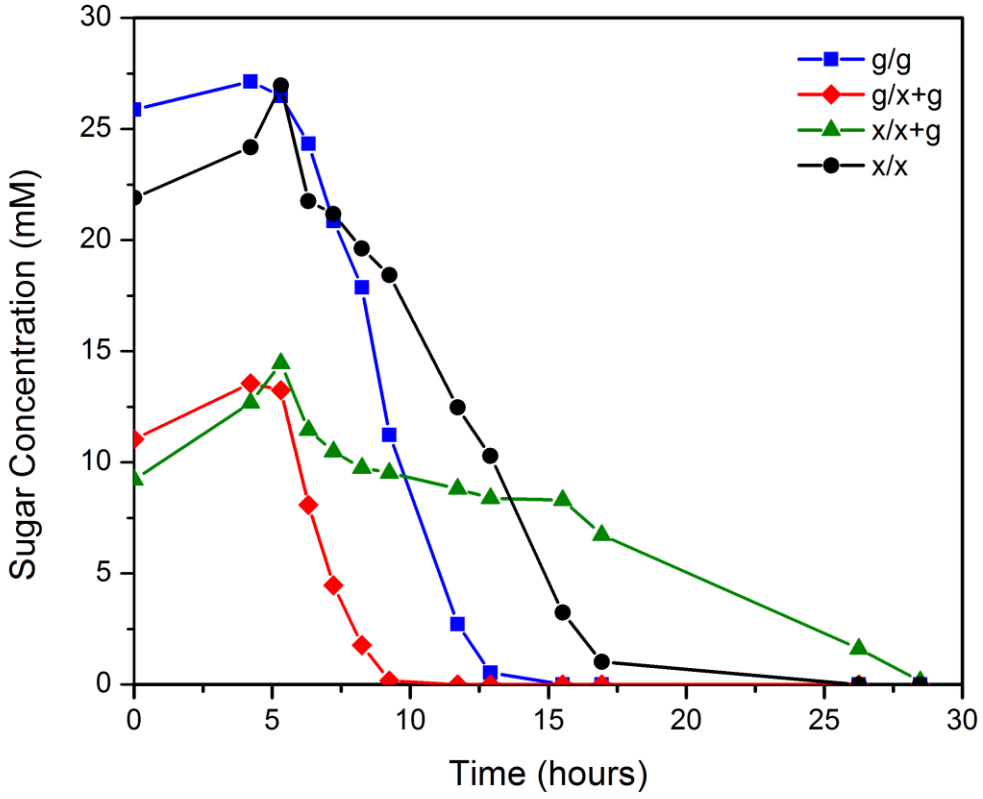

Figure 4.10. Sugar utilization in strain WT::pKK052. Cells were supplemented with 20 mM sugar concentrations (total) and allowed to grow until stationary phase was reached.

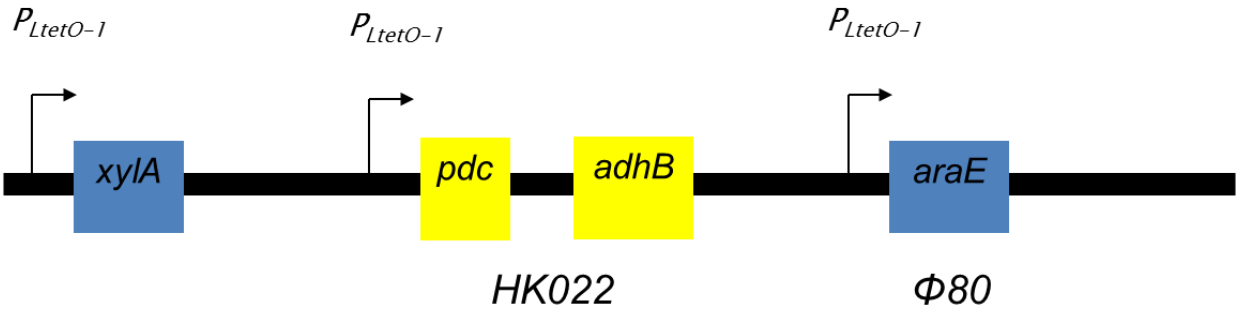

Figure 4.11. Design of the final engineered strain CR1269.

CHAPTER 5: Conclusions and recommendations

5.1 Analysis of efflux transporters: conclusions

We identified thirteen arabinose efflux transporters. Eight of these transporters showed an increase in intracellular arabinose concentrations when deleted. Six of these transporters showed a decrease in intracellular arabinose concentration when over expressed, and overexpression of twelve of these transporters proved to be toxic to the cell. None of these transporters had an effect on xylose metabolism.

We also investigated the possible substrates for the identified efflux transporters. Our results suggest that arabinose is a substrate for most of these transporters and ribulose may be a substrate for KgtP. We also studied the effects of these transporters on *tolC* mutants and found that most of these transporters are *tolC* dependent. Our final transcriptional reporter results also showed that the promoters of these transporters are arabinose inducible.

To summarize, we have identified sugar export as one of the key mechanisms of sugar transport for the cell. We have studied candidate efflux transporters for arabinose and showed that deletion of these transporters may lead to higher metabolism of arabinose in *E. coli.* This study allows us to develop a novel method to regulate the expression of genes involved in the sugar metabolic pathways within the cell.

5.2 Analysis of efflux transporters: recommendations for future work

Our experimental analysis gives an indirect measure of sugar metabolism in the cell. We did not directly measure the intracellular concentration of arabinose. Rather, we only have data to prove that the promoter activity of the gene that is responsible for arabinose breakdown (*araB)* is affected with the presence of these transporters in the cell. Further investigation would be

required to confirm whether intracellular arabinose concentration is affected. Our results suggest that arabinose accumulation is inhibited by the presence of these transporters. However, we have not confirmed that efflux is the only mechanism by which this occurs. Other possible mechanisms may exist, which need to be further investigated and/or ruled out.

It is also important to analyze why the cell would have a mechanism in which sugars are effluxed from the cell. Our hypothesis is that these systems exist to expel harmful sugar analogues (such as ribulose-5-phosphate and methylglyoxal). However, we have not confirmed this with experimental data, or identified the substrates for these efflux transporters.

A step further would be to identify and analyze efflux transporters for xylose. We studied 31 transporters and none of these showed any effect on intracellular xylose concentration. However, the candidate genes that we chose to study were picked based on homology to *ydeA,* which is known to efflux arabinose. These transporters could be substrate specific and hence we do not observe any effects with xylose. For future work, it is more advantageous to study ways in which we can improve xylose metabolism in the cell because xylose is most abundant in hemicellulose and yet least utilized during fermentation. Identifying a possible way to increase intracellular xylose concentrations in the cell and overcoming transport limitations would prove to be favorable for making the biofuel production process more efficient.

5.3 Strain engineering design: conclusions

Through rational strain engineering, we successfully built an *E. coli* strain capable of simultaneously co-utilizing sugars in a mixture of glucose and xylose. This strain was able to grow in a mixture of glucose and xylose without a diauxic shift or lag in growth due to sequential sugar utilization. HPLC results also confirmed that both sugars are utilized at the same time during the growth period of the cell. Our engineered strain retained these favorable

characteristics in aerobic as well as anaerobic conditions and can therefore be used in the industry for large-scale biofuels production.

5.4 Strain engineering design: recommendations for future work

The natural next step would be to test the engineered strain in a fermenter and measure ethanol yields to quantitatively measure its performance. Ethanol yields of 0.49 have previously been achieved, which is very close to the theoretical value of 0.51. In order to compare the performance of our engineered strain to that of others', it would be necessary to test the growth and ethanol production of these strains in similar fermentation conditions.

Another aspect that is not fully understood in our experiments is the changes in kinetic growth of the cells. One of the effects we observed with growth curves of our engineered strain is that the strains grow very poorly on xylose as the sole carbon source. The reason for this is not known. One possible explanation is that the cell is inhibited by mixed acid fermentation products and that the mutations introduced to improve co-utilization may make the cell more susceptible to these inhibitory affects. However, this has not been tested and the metabolic pathways have not been fully analyzed to determine whether the mutations are causing these changes in growth.

In addition, it is also recommended that different constitutive promoters be tested. Varying the strength of the promoter that is controlling the expression of the xylose metabolic genes may have differing effects on how the strain grows on xylose as the sole carbon source. Ribosomal promoters (P_{recA}) are possible options. Other promoters include synthetic promoters, which are also good alternatives as the strength of the promoter can be varied [104].

Finally, this approach of strain design can be applied to other metabolic genes in the xylose pathway and extrapolated to include the arabinose pathway as well. This would serve to overcome catabolite repression effects in the presence of glucose. Since CRP is not affected with this approach of strain engineering, we don't need to compensate for unwanted changes in

cell physiology that may develop with changes that are caused to the CRP regulatory system. This provides us with a novel approach towards strain design and helps to engineer a strain that is capable of metabolizing all the sugars in biomass most efficiently for maximum biofuel yields.

REFERENCES

1. Zaldivar, J., J. Nielsen, and L. Olsson, *Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration.* Appl Microbiol Biotechnol, 2001. **56**(1-2): p. 17-34.

2. DoKyoung Lee, V.N.O., Peter Jeranyama, *Herbaceous Biomass Feedstocks*, in *Sun Grant Initiative*. June 2007, South Dakota State University South Dakota.

3. Antoni, D., V.V. Zverlov, and W.H. Schwarz, *Biofuels from microbes.* Appl Microbiol Biotechnol, 2007. **77**(1): p. 23-35.

4. Fortman, J.L., et al., *Biofuel alternatives to ethanol: pumping the microbial well.* Trends Biotechnol, 2008. **26**(7): p. 375-81.

5. Keshwani, D.R. and J.J. Cheng, *Switchgrass for bioethanol and other value-added applications: a review.* Bioresour Technol, 2009. **100**(4): p. 1515-23.

6. Bothast, R.J., et al., *Fermentation of L-Arabinose, D-Xylose and D-Glucose by Ethanologenic Recombinant Klebsiella-Oxytoca Strain P2.* Biotechnology Letters, 1994. **16**(4): p. 401-406.

7. Dien, B.S., M.A. Cotta, and T.W. Jeffries, *Bacteria engineered for fuel ethanol production: current status.* Applied Microbiology and Biotechnology, 2003. **63**(3): p. 258-266.

8. Jarboe, L.R., et al., *Development of ethanologenic bacteria.* Biofuels, 2007. **108**: p. 237-261.

9. Ho, N.W., Z. Chen, and A.P. Brainard, *Genetically engineered Saccharomyces yeast capable of effective cofermentation of glucose and xylose.* Appl Environ Microbiol, 1998. **64**(5): p. 1852-9.

10. Jeffries, T.W. and Y.S. Jin, *Metabolic engineering for improved fermentation of pentoses by yeasts.* Applied Microbiology and Biotechnology, 2004. **63**(5): p. 495-509.

11. Goettemoeller, J. and A. Goettemoeller, *Sustainable ethanol : biofuels, biorefineries, cellulosic biomass, flex-fuel vehicles, and sustainable farming for energy independence*. 2007, Maryville, Mo.: Prairie Oak Pub. 195 p.

12. Hahn-Hagerdal, B., et al., *Bio-ethanol - the fuel of tomorrow from the residues of today.* Trends in Biotechnology, 2006. **24**(12): p. 549-556.

13. Saha, B.C. and R.J. Bothast, *Pretreatment and enzymatic saccharification of corn fiber.* Appl Biochem Biotechnol, 1999. **76**(2): p. 65-77.

14. Morrow, W.R., W.M. Griffin, and H.S. Matthews, *Modeling switchgrass derived cellulosic ethanol distribution in the United States.* Environmental Science & Technology, 2006. **40**(9): p. 2877-2886.

15. Alper, H. and G. Stephanopoulos, *Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential?* Nature Reviews Microbiology, 2009. **7**(10): p. 715-723.

16. Alterthum, F. and L.O. Ingram, *Efficient Ethanol-Production from Glucose, Lactose, and Xylose by Recombinant Escherichia-Coli.* Applied and Environmental Microbiology, 1989. **55**(8): p. 1943-1948.

17. Dien, B.S., et al., *Conversion of corn milling fibrous co-products into ethanol by recombinant Escherichia coli strains K011 and SL40.* World Journal of Microbiology & Biotechnology, 1997. **13**(6): p. 619-625.

18. Dien, B.S., L.B. Iten, and R.J. Bothast, *Conversion of corn fiber to ethanol by recombinant E-coli strain FBR3.* Journal of Industrial Microbiology & Biotechnology, 1999. **22**(6): p. 575-581.

19. Yanase, H., et al., *Genetic engineering of Zymobacter palmae for production of ethanol from xylose.* Applied and Environmental Microbiology, 2007. **73**(8): p. 2592-2599.

20. Meijnen, J.P., J.H. de Winde, and H.J. Ruijssenaars, *Engineering Pseudomonas putida S12 for efficient utilization of D-xylose and L-arabinose.* Applied and Environmental Microbiology, 2008. **74**(16): p. 5031-5037.

21. Warnick, T.A., B.A. Methe, and S.B. Leschine, *Clostridium phytofermentans sp. nov., a cellulolytic mesophile from forest soil.* Int J Syst Evol Microbiol, 2002. **52**(Pt 4): p. 1155-60.

22. Zhang, Y.H. and L.R. Lynd, *Cellulose utilization by Clostridium thermocellum: bioenergetics and hydrolysis product assimilation.* Proc Natl Acad Sci U S A, 2005. **102**(20): p. 7321-5.

23. Jeffries, T.W., et al., *Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast Pichia stipitis.* Nat Biotechnol, 2007. **25**(3): p. 319-26.

24. Matsushika, A., et al., *Ethanol production from xylose in engineered Saccharomyces cerevisiae strains: current state and perspectives.* Appl Microbiol Biotechnol, 2009. **84**(1): p. 37-53.

25. Hahn-Hagerdal, B., et al., *Metabolic engineering for pentose utilization in Saccharomyces cerevisiae.* Adv Biochem Eng Biotechnol, 2007. **108**: p. 147-77.

26. van Maris, A.J.A., et al., *Development of efficient xylose fermentation in Saccharomyces cereviside: Xylose lsomerase as a key component.* Biofuels, 2007. **108**: p. 179-204.

27. Jeffries, T.W., *Engineering yeasts for xylose metabolism.* Current Opinion in Biotechnology, 2006. **17**(3): p. 320-326.

28. Richard, P., et al., *Production of ethanol from L-arabinose by Saccharomyces cerevisiae containing a fungal L-arabinose pathway.* Fems Yeast Research, 2003. **3**(2): p. 185-189.

29. Kotter, P., et al., *Isolation and Characterization of the Pichia-Stipitis Xylitol Dehydrogenase Gene, Xyl2, and Construction of a Xylose-Utilizing Saccharomyces-Cerevisiae Transformant.* Current Genetics, 1990. **18**(6): p. 493-500.

30. Walfridsson, M., et al., *Ethanolic fermentation of xylose with Saccharomyces cerevisiae harboring the Thermus thermophilus xylA gene, which expresses an active xylose (glucose) isomerase.* Appl Environ Microbiol, 1996. **62**(12): p. 4648-51.

31. Becker, J. and E. Boles, *A modified Saccharomyces cerevisiae strain that consumes L-arabinose and produces ethanol.* Applied and Environmental Microbiology, 2003. **69**(7): p. 4144-4150.

32. Jin, M.J., et al., *Two-step SSCF to convert AFEX-treated switchgrass to ethanol using commercial enzymes and Saccharomyces cerevisiae 424A(LNH-ST).* Bioresource Technology, 2010. **101**(21): p. 8171-8178.

33. Rogers, P.L., et al., *Zymomonas mobilis for fuel ethanol and higher value products.* Biofuels, 2007. **108**: p. 263-288.

34. Szambelan, K., J. Nowak, and Z. Czarnecki, *Use of Zymomonas mobilis and Saccharomyces cerevisiae mixed with Kluyveromyces fragilis for improved ethanol production from Jerusalem artichoke tubers.* Biotechnology Letters, 2004. **26**(10): p. 845-848.

35. Patle, S. and B. Lal, *Ethanol production from hydrolysed agricultural wastes using mixed culture of Zymomonas mobilis and Candida tropicalis.* Biotechnology Letters, 2007. **29**(12): p. 1839-1843.

36. Deutscher, J., *The mechanisms of carbon catabolite repression in bacteria.* Current Opinion in Microbiology, 2008. **11**(2): p. 87-93.

37. Deutscher, J., C. Francke, and P.W. Postma, *How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria.* Microbiology and Molecular Biology Reviews, 2006. **70**(4): p. 939-+.

38. Goerke, B. and J. Stulke, *Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.* Nature Reviews Microbiology, 2008. **6**(8): p. 613-624.

39. Inada, T., K. Kimata, and H.J. Aiba, *Mechanism responsible for glucose-lactose diauxie in Escherichia coli: Challenge to the cAMP model.* Genes to Cells, 1996. **1**(3): p. 293-301.

40. Lawford, H.G. and J.D. Rousseau, *Relative Rates of Sugar Utilization by an Ethanologenic Recombinant Escherichia-Coli Using Mixtures of Glucose, Mannose, and Xylose.* Applied Biochemistry and Biotechnology, 1994. **45-6**: p. 367-381.

41. Kang, H.Y., S. Song, and C. Park, *Priority of pentose utilization at the level of transcription: Arabinose, xylose, and ribose operons.* Molecules and Cells, 1998. **8**(3): p. 318-323.

42. Yomano, L.P., et al., *Deletion of methylglyoxal synthase gene (mgsA) increased sugar co-metabolism in ethanol-producing Escherichia coli.* Biotechnology Letters, 2009. **31**(9): p. 1389-1398.

43. Trinh, C.T., P. Unrean, and F. Srienc, *Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses.* Applied and Environmental Microbiology, 2008. **74**(12): p. 3634-3643.

44. Yomano, L.P., S.W. York, and L.O. Ingram, *Isolation and characterization of ethanol-tolerant mutants of Escherichia coli KO11 for fuel ethanol production.* Journal of Industrial Microbiology & Biotechnology, 1998. **20**(2): p. 132-138.

45. Nichols, N.N., B.S. Dien, and R.J. Bothast, *Use of catabolite repression mutants for fermentation of sugar mixtures to ethanol.* Applied Microbiology and Biotechnology, 2001. **56**(1-2): p. 120-125.

46. Keseler, I.M., et al., *EcoCyc: a comprehensive database resource for Escherichia coli.* Nucleic Acids Research, 2005. **33**: p. D334-D337.

47. Guzman, L.M., et al., *Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter.* Journal of Bacteriology, 1995. **177**(14): p. 4121-30.

48. Hendrickson, W., C. Stoner, and R. Schleif, *Characterization of the Escherichia coli araFGH and araJ promoters.* Journal of Molecular Biology, 1990. **215**(4): p. 497-510.

49. Schleif, R., *Regulation of the L-arabinose operon of Escherichia coli.* Trends Genet, 2000. **16**(12): p. 559-65.

50. Horazdovsky, B.F. and R.W. Hogg, *High-affinity L-arabinose transport operon. Gene product expression and mRNAs.* Journal of Molecular Biology, 1987. **197**(1): p. 27-35.

51. Kehres, D.G. and R.W. Hogg, *Escherichia coli K12 arabinose-binding protein mutants with altered transport properties.* Protein Sci, 1992. **1**(12): p. 1652-60.

52. Lee, N.L., W.O. Gielow, and R.G. Wallace, *Mechanism of araC autoregulation and the domains of two overlapping promoters, Pc and PBAD, in the L-arabinose regulatory region of Escherichia coli.* Proc Natl Acad Sci U S A, 1981. **78**(2): p. 752-6.

53. Sumiya, M., et al., *Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in Escherichia coli.* Receptors Channels, 1995. **3**(2): p. 117-28.

54. Sofia, H.J., et al., *Analysis of the Escherichia coli genome. V. DNA sequence of the region from 76.0 to 81.5 minutes.* Nucleic Acids Research, 1994. **22**(13): p. 2576-86.

55. Song, S.G. and C. Park, *Organization and regulation of the D-xylose operons in Escherichia coli K-12: XylR acts as a transcriptional activator.* Journal of Bacteriology, 1997. **179**(22): p. 7025-7032.

56. Desai, T.A. and C.V. Rao, *Regulation of arabinose and xylose metabolism in Escherichia coli.* Appl Environ Microbiol, 2010. **76**(5): p. 1524-32.

57. Liu, J.Y., et al., *The identification of a new family of sugar efflux pumps in Escherichia coli.* Mol Microbiol, 1999. **31**(6): p. 1845-51.

58. Liu, J.Y., et al., *Functional and biochemical characterization of Escherichia coli sugar efflux transporters.* J Biol Chem, 1999. **274**(33): p. 22977-84.

59. Lee, M.D., et al., *Microbial fermentation-derived inhibitors of efflux-pump-mediated drug resistance.* Farmaco, 2001. **56**(1-2): p. 81-85.

60. Paulsen, I.T., M.K. Sliwinski, and M.H. Saier, *Microbial genome analyses: Global comparisons of transport capabilities based on phylogenies, bioenergetics and substrate specificities.* Journal of Molecular Biology, 1998. **277**(3): p. 573-592.

61. Nikaido, H., *Multidrug Resistance in Bacteria.* Annual Review of Biochemistry, 2009. **78**: p. 119-146.

62. Li, X.Z. and H. Nikaido, *Efflux-Mediated Drug Resistance in Bacteria An Update.* Drugs, 2009. **69**(12): p. 1555-1623.

63. Nishino, K., E. Nikaido, and A. Yamaguchi, *Regulation and physiological function of multidrug efflux pumps in Escherichia coli and Salmonella.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2009. **1794**(5): p. 834-843.

64. Nikaido, H. and Y. Takatsuka, *Mechanisms of RND multidrug efflux pumps.* Biochimica Et Biophysica Acta-Proteins and Proteomics, 2009. **1794**(5): p. 769-781.

65. Nikaido, E., A. Yamaguchi, and K. Nishino, *AcrAB multidrug efflux pump regulation in Salmonella enterica serovar typhimurium by RamA in response to environmental signals.* Journal of Biological Chemistry, 2008. **283**(35): p. 24245-24253.

66. Li, X.Z., L. Zhang, and H. Nikaido, *Efflux pump-mediated intrinsic drug resistance in Mycobacterium smegmatis.* Antimicrobial Agents and Chemotherapy, 2004. **48**(7): p. 2415-2423.

67. Li, X.Z. and H. Nikaido, *Efflux-mediated drug resistance in bacteria.* Drugs, 2004. **64**(2): p. 159-204.

68. Li, X.Z., et al., *Role of Efflux Pump(S) in Intrinsic Resistance of Pseudomonas-Aeruginosa - Active Efflux as a Contributing Factor to Beta-Lactam Resistance.* Antimicrobial Agents and Chemotherapy, 1994. **38**(8): p. 1742-1752.

69. Sun, Y. and C.K. Vanderpool, *Regulation and Function of Escherichia coli Sugar Efflux Transporter A (SetA) during Glucose-Phosphate Stress.* Journal of Bacteriology, 2011. **193**(1): p. 143-153.

70. Bost, S., F. Silva, and D. Belin, *Transcriptional activation of ydeA, which encodes a member of the major facilitator superfamily, interferes with arabinose accumulation and induction of the Escherichia coli arabinose P-BAD promoter.* Journal of Bacteriology, 1999. **181**(7): p. 2185-2191.

71. Carole, S., S. Pichoff, and J.P. Bouche, *Escherichia coli gene ydeA encodes a major facilitator pump which exports L-arabinose and isopropyl-beta-D-thiogalactopyranoside.* Journal of Bacteriology, 1999. **181**(16): p. 5123-5125.

72. Schleif, R. and J.T. Lis, *The regulatory region of the L-arabinose operon: a physical, genetic and physiological study.* Journal of Molecular Biology, 1975. **95**(3): p. 417-31.

73. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6640-6645.

74. JH, M., *A Short Course in Bacterial Genetics: A Laboratory Manual for Escherichia coli and Related Bactera.* 1992, Cold Spring Harbor Laboratory: Cold Spring Harbor, New York.

75. Lindsay, S.E., R.J. Bothast, and L.O. Ingram, *Improved strains of recombinant Escherichia coli for ethanol production from sugar mixtures.* Appl Microbiol Biotechnol, 1995. **43**(1): p. 70-5.

76. Amann, E., B. Ochs, and K.J. Abel, *Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli.* Gene, 1988. **69**(2): p. 301-15.

77. Nagai, T., et al., *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.* Nat Biotechnol, 2002. **20**(1): p. 87-90.

78. Bohn, C. and P. Bouloc, *The Escherichia coli cmlA gene encodes the multidrug efflux pump Cmr/MdfA and is responsible for isopropyl-beta-D-thiogalactopyranoside exclusion and spectinomycin sensitivity.* Journal of Bacteriology, 1998. **180**(22): p. 6072-6075.

79. Staub, J.M., et al., *Bacterial glyphosate resistance conferred by overexpression of an E. coli membrane efflux transporter.* Journal of Industrial Microbiology & Biotechnology, 2012. **39**(4): p. 641-647.

80. Furrer, J.L., et al., *Export of the siderophore enterobactin in Escherichia coli: involvement of a 43 kDa membrane exporter.* Molecular Microbiology, 2002. **44**(5): p. 1225-1234.

81. Han, X.L., et al., *Escherichia coli genes that reduce the lethal effects of stress.* Bmc Microbiology, 2010. **10**.

82. Hayashi, M., et al., *Effect of multidrug-efflux transporter genes on dipeptide resistance and overproduction in Escherichia coli.* Fems Microbiology Letters, 2010. **304**(1): p. 12-19.

83. Seol, W. and A.J. Shatkin, *Escherichia-Coli Kgtp Encodes an Alpha-Ketoglutarate Transporter.* Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(9): p. 3802-3806.

84. Englesberg, E., et al., *L-Arabinose-Sensitive, L-Ribulose 5-Phosphate 4-Epimerase-Deficient Mutants of Escherichia Coli.* Journal of Bacteriology, 1962. **84**(1): p. 137-&.

85. Rosner, J.L. and R.G. Martin, *An Excretory Function for the Escherichia coli Outer Membrane Pore TolC: Upregulation of marA and soxS Transcription and Rob Activity Due to Metabolites Accumulated in tolC Mutants.* Journal of Bacteriology, 2009. **191**(16): p. 5283-5292.

86. Chubiz, L.M. and C.V. Rao, *Aromatic Acid Metabolites of Escherichia coli K-12 Can Induce the marRAB Operon.* Journal of Bacteriology, 2010. **192**(18): p. 4786-4789.

87. Edgar, R. and E. Bibi, *MdfA, an Escherichia coli multidrug resistance protein with an extraordinarily broad spectrum of drug recognition.* Journal of Bacteriology, 1997. **179**(7): p. 2274-2280.

88. Ferguson, G.P., et al., *Methylglyoxal production in bacteria: suicide or survival?* Archives of Microbiology, 1998. **170**(4): p. 209-219.

89. Ackerman, R.S., Cozzarel.Nr, and W. Epstein, *Accumulation of Toxic Concentrations of Methylglyoxal by Wild-Type Escherichia-Coli K-12.* Journal of Bacteriology, 1974. **119**(2): p. 357-362.

90. Postma, P.W., J.W. Lengeler, and G.R. Jacobson, *Phosphoenolpyruvate - Carbohydrate Phosphotransferase Systems of Bacteria.* Microbiological Reviews, 1993. **57**(3): p. 543-594.

91. Cherepanov, P.P. and W. Wackernagel, *Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant.* Gene, 1995. **158**(1): p. 9-14.

92. Gosset, G., *Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate : sugar phosphotransferase system.* Microbial Cell Factories, 2005. **4**.

93. Lindsay, S.E., R.J. Bothast, and L.O. Ingram, *Improved Strains of Recombinant Escherichia-Coli for Ethanol-Production from Sugar Mixtures.* Applied Microbiology and Biotechnology, 1995. **43**(1): p. 70-75.

94. Zhang, J. and R. Greasham, *Chemically defined media for commercial fermentations.* Applied Microbiology and Biotechnology, 1999. **51**(4): p. 407-421.

95. Hollands, K., S.J.W. Busby, and G.S. Lloyd, *New targets for the cyclic AMP receptor protein in the Escherichia coli K-12 genome.* Fems Microbiology Letters, 2007. **274**(1): p. 89-94.

96. Karlinsey, J.E., *lambda-Red genetic engineering in Salmonella enterica serovar Typhimurium.* Methods Enzymol, 2007. **421**: p. 199-209.

97. Lutz, R. and H. Bujard, *Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I-1-I-2 regulatory elements.* Nucleic Acids Research, 1997. **25**(6): p. 1203-1210.

98. Seo, J.S., et al., *The genome sequence of the ethanologenic bacterium Zymomonas mobilis ZM4.* Nat Biotechnol, 2005. **23**(1): p. 63-8.

99. Haldimann, A. and B.L. Wanner, *Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria.* Journal of Bacteriology, 2001. **183**(21): p. 6384-93.

100. Desai, T.A., *Regulation of hexose and pentose metabolism by Escherichia coli*. 2009, University of Illinois at Urbana-Champaign. p. x, 149 leaves, bound.

101. Conway, T., et al., *Promoter and nucleotide sequences of the Zymomonas mobilis pyruvate decarboxylase.* Journal of Bacteriology, 1987. **169**(3): p. 949-54.

102. Conway, T., et al., *Cloning and sequencing of the alcohol dehydrogenase II gene from Zymomonas mobilis.* Journal of Bacteriology, 1987. **169**(6): p. 2591-7.

103. Ingram, L.O., et al., *Genetic engineering of ethanol production in Escherichia coli.* Appl Environ Microbiol, 1987. **53**(10): p. 2420-5.

104. Jensen, P.R. and K. Hammer, *The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters.* Applied and Environmental Microbiology, 1998. **64**(1): p. 82-87.