Copyright 2014 Dawn T. Eriksen
COMBINATORIAL BIOSYNTHETIC PATHWAY ENGINEERING
FOR MICROBIAL PRODUCTION OF BIOFUELS

BY
DAWN T. ERIKSEN

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, 2014

Urbana, Illinois

Doctoral Committee:
Professor Huimin Zhao, Chair
Assistant Professor Brendan Harley
Assistant Professor Charles Schroeder
Professor Isaac Cann
Abstract

To compete in a market dominated by fossil fuels, biofuels must be economically competitive and also offer the variety of molecules and compounds which are currently derived from fossil fuels. This thesis offers potential strategies for biofuels to be both economically competitive and diverse. Effective and economical production of biofuels comes from the optimization of the biosynthetic pathway. We investigated and developed new combinatorial strategies for the optimization of the cellobiose utilization pathway, a pathway which is important in biofuel production. One strategy focused on optimizing enzyme combinations by creating a library of homologous proteins from the pathway. A second strategy investigated engineering all of the proteins in the pathway simultaneously. The improved pathway was assessed based on specific growth rate on cellobiose, with the final mutant exhibiting a 42% increase over the wild-type pathway. Metabolite analysis of the engineered pathway presented a 54% increase in cellobiose consumption (1.68 to 2.82 g cellobiose/(L·h)) and a 74% increase in ethanol productivity (0.59 to 1.03 g ethanol/(L·h)).

The second half of the thesis was focused on creating a biofuel molecule with more diverse applications than the commonly used bioethanol biofuel. A new pathway for biodiesel production was investigated, using a heterologous fatty acid synthesis pathway, which would provide a completely orthologous route for biodiesel production. In this strategy, the endogenous fatty acid flux would not be redirected from cellular metabolism. Through heterologous expression of a Type-I fatty acid synthase, the total production of fatty acid ethyl esters was increased 6.3-fold, from 1670 µg FAEE/ g CDW to 10,498 µg FAEE/ g CDW.

The final work in the thesis surveyed three potential high-throughput screening methods to subject the biodiesel production pathway to the optimization strategies developed earlier in the thesis.
Acknowledgements

Initial acknowledgements go to my parents. To my mother, thank you for helping me through graduate school by reminding me to take it “one step at a time”. To my father, thank you for all those home science experiments when I was young and your subscription to Science News.

Thank you to my research advisor, Professor Huimin Zhao. Throughout the disappointments, he has been steadfast, solid and steady as a rock. His guidance allowed me to conduct research in areas which were exceptionally exciting--a component which has kept me motivated.

I also appreciate the people who trained me in the Zhao Lab: Dr. Ryan Sullivan, Dr. Jing Du, Dr. Nikhil Nair. Thanks to my mentors, Dr. Byoung-jin Kim, Dr. Yongbo Yuan, and Dr. George Schmitz. Thanks to my colleagues in the lab, my EBI members: Sijin Li, Jiazhang Lian, Ran Chao, Dr. Xueyang Feng, Dr. Dan Coursolle, and Sam Hamedi Rad. With a special thank you to Ran and Jiazhang for helpful discussions and Sam for finishing the FAEE work for publication.

And of course, the rest of the Zhao Group members, thank you. A special thanks goes to Ryan Cobb, Zhanar Abil, and Carl Denard from RAL for all of their help and support. And I’m so grateful to Rachel Waldemar, Zhanar Abil, and Kori Dunn and for the GNs throughout the years.

I’m also appreciative of my mentors who contributed to my education before coming to UIUC: Professor Lianhong Sun, Dr. Daniel Sayut, and Dr. Pavan Kambam from UMass Amherst, also Professor Kristala Prather-Jones and Dr. Collin Martin from MIT, and Professor Christopher Snow and Professor Frances Arnold from CalTech. Without their help I would not have become a fellow of the NSF-GRFP, financial support I would also like to acknowledge.

Last, but certainly far from least, thank you to Daniel Stapleton, my significant other. For all the times he would drive to Urbana from Chicago, even though I still had to work. For all he times he would drop me off at lab saying: “Go do your thing, I’ll read baseball news.” Thank you.
# Table of Contents

**Chapter 1: Introduction** ...................................................................................................................... 1  
1.1 Sustainability ........................................................................................................................................ 1  
1.2 Engineering Microorganisms for Sustainable Chemical Production ............................................. 8  
1.3 Scope of Thesis .................................................................................................................................... 31  
1.4 Conclusions ......................................................................................................................................... 35  
1.5 References .......................................................................................................................................... 35  
1.6 Figures ................................................................................................................................................. 41  

**Chapter 2: Combinatorial Library of Homologous Proteins for the Cellobiose Utilization Pathway** ................................................................................................................. 48  
2.1 Background ......................................................................................................................................... 48  
2.2 Results ................................................................................................................................................ 55  
2.3 Discussion and Conclusion ................................................................................................................. 58  
2.4 Materials and Methods ....................................................................................................................... 59  
2.5 References .......................................................................................................................................... 64  
2.6 Tables .................................................................................................................................................. 67  
2.7 Figures ................................................................................................................................................ 68  

**Chapter 3: Developing a Directed Evolution Strategy for the Optimization of Heterologous Metabolic Pathways by Simultaneously Engineering Multiple Proteins** ........................................... 76  
3.1 Background ......................................................................................................................................... 76  
3.2 Results ................................................................................................................................................ 79  
3.3 Discussion .......................................................................................................................................... 83  
3.4 Conclusion ......................................................................................................................................... 87  
3.5 Materials and Methods ....................................................................................................................... 87  
3.6 References .......................................................................................................................................... 96  
3.7 Tables .................................................................................................................................................. 99  
3.8 Figures ................................................................................................................................................. 102
Chapter 4: Engineering a Novel Pathway for Advanced Biofuel Production... 114

4.1 Background ........................................................................................................................................ 114
4.2 Results ............................................................................................................................................... 117
4.3 Discussion .......................................................................................................................................... 123
4.4 Conclusion ......................................................................................................................................... 128
4.5 Materials and Methods .................................................................................................................... 128
4.6 References ......................................................................................................................................... 135
4.7 Tables ................................................................................................................................................ 138
4.8 Figures .............................................................................................................................................. 140

Chapter 5: Developing a High-throughput Screen for FAEE production in S. cerevisiae .... 155

5.1 Background ........................................................................................................................................ 155
5.2 Results ............................................................................................................................................... 159
5.3 Discussion .......................................................................................................................................... 161
5.4 Future Work ....................................................................................................................................... 164
5.5 Materials and Methods ..................................................................................................................... 165
5.6 References ......................................................................................................................................... 170
5.7 Figures .............................................................................................................................................. 173
Chapter 1: Introduction

1.1 Sustainability

The dwindling of natural resources is causing industries and governments to reconsider the design principles of many industrial processes. Sustainability has now become a major component in all sectors of industrial, governmental, and even societal agencies. Though the definition of sustainable development has been argued over the years, a standard definition has emerged as “... a process of change in which the exploitation of resources, the direction of investments, the orientation of technological development, and institutional change are all in harmony and enhance both current and future potential to meet human needs and aspirations. ... (It is) meeting the needs of the present without compromising the ability of future generations to meet their own needs” (1).

1.1.1 Sustainable Development

Sustainable development has been cited as the most difficult challenge that humanity has ever faced (1). To achieve sustainability, many fundamental and well-established processes at local, regional, and global levels must be reevaluated. Industrial sustainability will require a global vision that holistically considers economic, social, and environmental sustainability concerning more than just short-term profit margins, focusing instead on long-term investments. The most considerable challenge in this endeavor is that the sustainable product must be commercially

---

1 Parts of this chapter have been published as:


competitive with its non-sustainable counterpart, if the government does offer a subsidy or mandate. It must be cheaper, more durable, less toxic, and easily recyclable. These sustainable products need to be derived from renewable resources and contribute minimally to the net generation of greenhouse gases (1). These lofty aspirations may sound extreme, but the technologies to attain these goals are emerging.

Many of these nascent technology fields are focusing on creating products derived from renewable resources such as trees and agricultural crops. An interesting discussion involves the life cycle of commodity chemicals such as dyes, solvents and synthetic fibers. At the beginning of the 20th century, many of these chemicals were made from renewable sources like trees and agricultural crops. However, by the late 1960s, many of these bio-based chemical products had been replaced by petroleum derivatives, a non-renewable source (2). In recent decades, liquid petroleum consumption has increased dramatically for both commodity chemicals and energy. Currently, $24 billion worth of hydrocarbon feedstocks are used annually in the chemical industry (1). In recent years, there have been efforts considering ways to return to renewable sources for these chemicals; however the process for sustainable manufacturing of these commodity chemicals is very different from the early 20th century.

Though the chemical industry relies heavily on petroleum-based chemicals, the most significant use of petroleum is for energy. In 2011, the global consumption of petroleum reached 87.42 million barrels each day (3). Furthermore, even though this number is startlingly high, experts predict the energy demand to grow by more than 50% by 2025 (2). The company British Petroleum (BP) has mapped energy consumption since 1986 and the trend illustrates an energy
demand increase at an alarming rate (Figure 1.1) (4). Due to this escalating global demand for more energy, it is imperative to identify sufficient resources to meet this demand. Recent efforts in tar sands, shale oil, shale gas, and natural gas have eased the current supply deficiencies. However, concern with large green-house has emission with the conversion of these resources and also the significant detrimental effects these sources have on the environment have spurred more sustainable strategies. This thesis focuses on potential alternatives for renewable energies, with a focus on microbial production of biofuels.

1.1.2 Alternative and Renewable Energies

There are many avenues for renewable energies originating from natural and renewable resources: sunlight, wind, rain, tides, biomass, and geothermal heat. Currently about 16% of global energy consumption comes from renewables: 10% is produced from traditional biomass (mainly for heating), 3.4% from hydroelectricity, and the final 3% is derived from new renewable technologies, such as biofuels, wind, and solar (5).

To identify a strategy which will be the most influential on achieving sustainability, the largest consumption of energy must be considered. A breakdown of energy consumption with the United States yields the largest fraction of energy consumption to be hydroelectric power at 40%, transportation fuels at 30%, industrial energy at 20%, and residential at 10% (Figure 1.2) (6). When considering the largest amount of energy consumption, hydroelectric power is already derived from renewable sources, thus the next most significant energy consumption is transportation fuels. These fuels are used for cars, planes, and trains. Therefore, to make the greatest impact in sustainable development, major focus must be on the attainment of renewable transportation fuels.
In the transportation industry, there has been some success with electric cars (7), and though there are other potential sustainable technologies relying on fuel cells and hydrogen, these strategies will probably not play a substantial role in heavy duty vehicles, shipping, and aviation because of their limited energy density (8). Moreover, the existing transportation fuel infrastructure relies on the burning of a carbon-based molecule. Thus, the easiest renewable resource for transportation fuels would be biomass which could produce a carbon-based molecule for combustion within the existing car engine. To this end, there has been some success with biofuels: in 2011 the equivalent of 10,000 barrels per day oil was produced as biofuel (4).

1.1.3 Biomass and Strategies for Biomass Conversion

Biomass is a carbon-based renewable energy source derived from garbage, plant, and/or waste. Significant research for biofuels focuses on plant biomass. Cellulose is the most abundant component of plant biomass, typically in the range of approximately 35 to 50% of plant dry weight (9). Cellulose is comprised of long polysaccharide chains of β-linked D-glucose molecules (10) (Figure 1.3). The cellulose fibers are often embedded in a matrix of other structural biopolymers, primarily hemicelluloses and lignin, comprising of 20 to 35% and 5 to 30% of plant dry weight, respectively (9). The hemicellulose is a highly branched chain composed of cross-linked hexose sugars such as D-glucose and also the pentose sugars D-xylose and L-arabinose (10). The lignin consists of phenols and chloroform alcohols which are generally considered unusable. These biopolymers are the dominant structural feature of biomass (Figure 1.4) and the degradation of which limits the rate and extent of biomass utilization (9).
The conversion of biomass into usable fuels can be completed either through thermochemical routes, including gasification and pyrolysis, or through biochemical routes, such as microorganism fermentation. Studies have shown that both processes are economically comparable for the production of biofuels (11).

1.1.3.1 Thermochemical Routes

One thermochemical process is gasification: a transformation of the biomass raw material into a gaseous mixture through a series of chemical reactions in a controlled amount of air and high temperature (12). The gaseous product is an energy rich mixture of H\textsubscript{2}, CO, CO\textsubscript{2}, CH\textsubscript{4}, C\textsubscript{2}H\textsubscript{4}, and other impurities such as nitrogen, sulfur, alkali compounds, and tars (13). Different gases or air used in the gasification chemical reaction can yield different products. Specific gasification of the biomass can be used to convert the biomass into a hot synthesis gas, ‘syn-gas’ (14). Syn-gas can then be converted to a variety of fuels including hydrogen by a water gas-shift reaction, a standard reaction where CO reacts with water to form CO\textsubscript{2} and H\textsubscript{2}. Methanol is also produced through methanol synthesis: after the water gas-shift reaction occurs, the CO reacts with water to form CO\textsubscript{2} and H\textsubscript{2}, and then there is a hydrogenation of CO\textsubscript{2} to methanol. Alkanes are produced from the biomass by Fischer-Tropsch synthesis using Co-, Fe-, or Ru-based catalysts, the overall chemical reaction shown below (12).

\[
\text{CO} + 2\text{H}_2 \rightarrow (1/n)\text{C}_n\text{H}_n + \text{H}_2\text{O}
\]

Another thermochemical process, pyrolysis can produce solid, liquid, and gaseous products from biomass (15). Pyrolysis is differentiated from gasification because the reactions are completed in the absence of oxygen. There has been significant success in chemical routes to produce transportation fuels from biomass (16). Some of these successes include a pilot plant built by
Honeywell for industrial scale biofuels, which is a 50 million gallon facility to produce green gasoline from biomass.

1.1.3.2 Biochemical Routes

Though thermochemical routes have been successful, the processes require expensive catalysts and high temperature and pressures. Thus, a more sustainable route is biochemical, i.e. microbial fermentation. This process uses lower temperatures, atmospheric pressure, and near-neutral pH. Biotechnology had been declared a key technology for a sustainable chemical industry (1), thus it has been a focus as a technology for sustainable production of commodity chemicals and biofuels.

1.1.3.2.1 Pretreatment

In order for the microorganisms to ferment the biomass, the complex biopolymer structure must first be broken down into simple sugars which the microorganisms can utilize. This degradation of the biomass, called pretreatment, has been cited as one of the most expensive processes for the biochemical conversion to biofuels (17). The lignin must first be removed from the biomass mixture and the remaining components are treated through a variety of different methods. The cellulose and hemicellulose can be broken down into individual sugars through either steam explosion, mechanical milling, high temperature treatment, and acid/base treatment to release individual sugar molecules (18). These methods are energy intensive and employ harsh solvents. Therefore, in efforts to make biochemical routes sustainable and economically competitive with petroleum products, steam explosion and mechanical milling are not good methods. Thus, the preferred treatment is enzymatic hydrolysis.
An *in vitro* enzyme cocktail is used to digest the hemicellulose and cellulose molecules. The cellulose molecule is broken down by cellulases, such as endoglucanases and cellbiohydrolases. These enzymes cleave the long polysaccharide chain into disaccharide cellobiose molecules. Subsequently, β-glucosidases hydrolyzes the β1→4 bond of the cellobiose, releasing two D-glucose molecules (9). The hemicellulose fraction consisting of the linked pentose sugars are cleaved by hemicellulases into single molecules of D-xylose and L-arabinose (19). The final hydrolyzate contains a mixture of pentose and hexose sugars such as D-glucose, D-xylose, and L-arabinose. In this enzymatic pre-treatment process, not only are significant amounts of enzymes required to completely break the biomass down into the simple sugars as D-glucose, D-xylose, but the enzymes can be inhibited by the intermediate cellobiose (9,20).

### 1.1.3.2.2 Biochemical Conversion

The use of biological systems, such as microbes, is attractive for sustainable production of transportation fuels and other chemicals due to the low-energy, low temperature, and low pressure systems associated with the process. Additionally, no heavy metals, organic solvents, and strong acids/bases are used, in comparison with traditional chemical synthesis, thus making the biochemical route sustainable and environmentally-friendly (21). Additionally, for many processes, such as butanol and ethanol production, the microorganism has evolved over thousands of years to optimally produce that chemical. Thus biochemical routes for biofuel production hold significant promise.
Different organisms have been used for biofuel production, including *Saccharomyces cerevisiae*, *Escherichia coli*, *Zymomonas mobilis*, and strains of *Clostridium*. Both *S. cerevisiae* and *E. coli* are genetically tractable and desired for most work in microbial catalysis. *S. cerevisiae* is a common chassis based on the fact that the microorganism is very industrially robust: tolerant to low pH, resistant to phages, and has a fast doubling time. Additionally, industrial-scale fermentation of this microorganism has been achieved through the beer/wine industry. *S. cerevisiae* is most often used in current industrial-scale bioethanol production and is a goal for many future industrial processes. However, most of the advanced biofuel studies are conducted in *E. coli* because this organism is easier to work with.

1.2 Engineering Microorganisms for Sustainable Chemical Production

A major thrust in sustainability is to use microorganisms to produce commodity and specialty chemicals. These microorganisms can use renewable resources as a substrate to grow and components produced by the microorganisms will be non-toxic to the environment and to humans. Additionally, these microbes operate at low temperatures and pressures compared to traditional chemical catalysis, thus reducing the amount of energy needed to produce the chemicals. These advantages make microbial production of chemicals a significant area of research in the endeavor to achieve sustainability.

The broad field of metabolic engineering has been established to engineer metabolic pathways within a microorganism to produce commodity and specialty chemicals. Enzymatic pathways within the microorganisms maintain the vitality of all living organisms. These biochemical routes have been exploited to produce numerous commodities since early civilization, examples
including beer, wine, and cheese. With the advance of biotechnology, various genetic tools became available for construction and manipulation of pathways to efficiently convert renewable feedstock to value-added compounds such as specialty chemicals, pharmaceuticals, and biofuels (21). Microbial production of these compounds is usually enabled by over-expressing endogenous or heterologous enzymes of the corresponding pathways. However, over-expression of pathway enzymes alone can be insufficient for optimal metabolite production because of an imbalanced flux through the pathway (21,22). A typical symptom of flux imbalance is the accumulation of unwanted and even toxic intermediates (23,24), which can be detrimental to the productivity of desired compounds. There is seldom a straightforward strategy to resolve the accumulation because enzymes within the pathway are not independent; instead the enzymes are intertwined and cross-regulated with each other and many intricate metabolic networks.

1.2.1 Design and Construction of Pathways

1.2.1.1 Computational Programs for Pathway Design

The first major challenge in engineering microbes for chemical production is designing the enzymatic pathway to produce the target chemical. Identifying a pathway may require sorting through thousands of possible enzymes and reactions wherein all the parameters have different substrate preferences and kinetic features. Computational algorithms have proven vital in pathway design and have the ability to discover de novo pathways by combining enzymes from various sources to produce non-natural products.

Several databases containing enzymes and enzymatic reactions have been developed: Kyoto Encyclopedia of Genes and Genomes (KEGG) (25), BRaunschwig ENzyme DAtabase
(BRENDA) (26), MetaCyc (27), the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD) (28,29), Retro-Biosynthesis Tool (ReBiT) (30), and the Universal Protein Resource (Uniprot) (31). With these databases, it is possible to identify a staggering number of different pathways for synthesis of the same target compound. Thus a major challenge in computational design is to rank candidate pathways to predict which pathway will have the highest yield of the desired compound.

1.2.1.2 Pathway Construction

1.2.1.2.1 Homologous DNA Recombination Techniques

After identifying the enzymes needed to catalyze the reactions, the next challenge is assembling the genes encoding the enzymes together. The three main DNA assembly strategies are based on homologous recombination and DNA repair mechanisms: DNA assembler, SLIC, and Gibson assembly. In the DNA assembler strategy, the endogenous in vivo homologous recombination mechanism in yeast is used to create large pathways in a simple, one-step manner (32-34). The DNA fragments to be assembled are PCR-amplified by oligos designed with homologous regions to the 5’ and 3’ neighboring DNA sequences within the pathway. The linear DNA fragments are co-transformed with the linear plasmid backbone into Saccharomyces cerevisiae and the homologous regions are recognized by the endogenous homologous recombination machinery for repair. These fragments are then combined into a complete pathway.

Mimicking the in vivo homologous recombination mechanisms, in vitro assembly has been accomplished by Sequence and Ligation Independent Cloning (SLIC), a two-step DNA assembly method (35). The DNA sequence of interest is PCR amplified with oligos containing a 30 base
pair (bp) sequence homologous to the 5’ and 3’ flanking DNA sequences. The linearized host vector and the insert DNA fragment are separately treated with T4 DNA polymerase in the absence of deoxynucleotide triphosphates (dNTPs). This enables the exonuclease activity of the T4 polymerase to chew back 3’ terminal end, generating a 5’ overhang which is homologous to the vector/insert. The second step involves addition of RecA and ATP, which can recombine the DNA fragments together into a plasmid, with four nicks in the DNA. Then upon transformation into *E. coli* for DNA amplification, the nicks are repaired by the *in vivo* DNA repair system.

The Gibson assembly method (36) is similar to SLIC, but exploits a specific exonuclease to chew back the 5’ end to generate single stranded complementary overhangs and ligases are incorporated in the reaction mix to seal the DNA nicks. The DNA fragments are PCR amplified with about 15-30 bp of homologous DNA regions to the 5’ and 3’ adjacent DNA sequences. In a single reaction, both the vector and the insert are subjected to T5 exonuclease that chews back the 5’ ends of the DNA fragments, and then the polymerase and ligase combine the homologous ends of fragments to a single circular DNA molecule.

### 1.2.1.2.2 Restriction Digestion and Ligation Techniques

Another family of advanced DNA assembly techniques have been developed based on Type IIS endonucleases, such as *BsaI*. These endonucleases cleave the DNA outside of their recognition sites, resulting in 5’ or 3’ DNA overhangs of nearly any user-defined nucleotide sequence (37,38). This strategy is more advanced than traditional restriction digestion-ligation method because it allows more flexibility in insertion location than cloning into the multiple-cloning site (MCS) on a plasmid. Use of Type IIS endonucleases through the Golden Gate assembly method
is a one-step reaction, which combines restriction digestion and ligation. This method has a high fragment assembly efficiency and proven to be effective in creating gene libraries (37). A continuing area of research with this technique is investigating a more modular approach for pathway and pathway library construction (39,40).

The need for modularity in gene and pathway cloning is becoming more significant with recent focuses on high-throughput DNA assembly and automation. One of the most established strategies for standardized assembly is the BioBrick system (41-46). The BioBrick and BglBrick standards (such as vectors, promoters, RBS) rely on isocaudamer pairs of restriction enzymes to generate compatible cohesive ends, and upon ligation, result in a scar sequence that cannot be cleaved by either of the original restriction digests. DNA fragments flanked with these recognition sequences can be used for modular assembly of a pathway by iterative digestions and ligations. The ePathBricks have been recently developed for BioBricks and are biological components for the fast assembly of full pathways (47). The vectors developed in this system can be used to modify the multi-gene pathway regulatory elements such as promoters, operators, RBSs, and terminators. Modification of these components is significant for pathway optimization.

1.2.2 Traditional Pathway Optimization Strategies

These advanced DNA assembly technologies have made it relatively easy to quickly construct large pathways of enzymes which could produce complex chemicals. However, pathway engineering is more than just recruiting various enzymes and stringing them together (48). In early heterologous gene expression studies, it was discovered that there was significant growth inhibition in cells over-expressing superfluous genes. This inhibition was attributed to the
competition of protein synthesis machinery for essential proteins required for cell growth versus the production of the over-expressed proteins, and the demand for nucleotides in the replication of the DNA itself. The phenomenon became known as the metabolic burden and reduction of the metabolic burden has become essential in the pursuit to identify an optimal metabolic pathway (49-52). In addition to minimizing the metabolic burden, to find an optimal metabolic pathway, the enzymes within the pathway must be balanced. Often one or two enzymes in the pathway will not express well or have low activity. This causes a bottleneck or a build-up of intermediates in the pathway, thereby reducing the overall titer. There have been many examples of strategies to minimize the metabolic burden and to find a balanced pathway, ranging from transcriptional and translational engineering, to protein engineering for higher and more specific activity, to combinatorial pathway libraries.

1.2.2.1 Transcriptional and Translational Engineering for Pathway Optimization

Transcriptional and translational engineering is based on changing the quantity of proteins in the system. Often, a plasmid-based system is used to express the proteins in either an operon or individual promoter basis. There are disadvantages to the traditional plasmid-based expression system. Plasmid instability has shown to be a major problem. Due to the metabolic burden on the plasmid-bearing organism, it has a selective disadvantage compared to a plasmid-free cell. Without selection pressure, an entire culture could be dominated by the plasmid-free cell. Integration into the chromosome leads to long-term stability and circumvents the loss of productivity by allele segregation due to plasmid use. Tyo et al. (53) developed a technique called chemical inducible chromosomal evolution (CiChE) to integrate the heterologous genes into the chromosome at multiple sites, with up to 40 consecutive copies.
In transcription, studies have shown that different organisms have a bias toward preferred codons and if a gene from an organism with a certain codon bias is heterologously expressed in a host with a different codon bias, the translation of the recombinant protein will be suboptimal. This can be overcome by codon optimization, wherein the codon sequence of the target gene is altered to match that of the codon bias of the host. Codon optimization is commercially available from DNA synthesis companies and has proven to increase protein expression in multiple studies (54-57).

Grouping the genes of interest for the pathway into a single polycistronic mRNA can be beneficial to optimize the enzyme expression. This one-step gene expression allows for simplified induction processes and improved regulatory mechanisms. However, the design of the basic operon can greatly affect expression. Smolke et al. (58) showed that altering the mRNA stability by engineering the intergenic regions would increase protein production and thus overall titer. By introducing RNA hairpin turns into the 3’ and 5’ intergenic region of the gene of interest, the mRNA was less susceptible to mRNA degradation.

Promoters can be mutagenized to achieve precise strength and regulation of protein expression. These promoters of varying strengths have been applied to exhibit a broad range of genetic control, which can be selected to construct an optimal metabolic pathway. Alper et al. (59) used a library of mutant promoters with varying strengths to regulate the expression of the dxs gene, which led to the improvement of the volumetric productivity of lycopene. It was found that
optimal lycopene titer did not involve the strongest promoter, an example of the balancing act needed to optimize the metabolic burden and protein activity.

1.2.2 Directed Evolution and Rational Design of Proteins for Pathway Optimization

Protein engineering can be used to modify and optimize pathways. The previously discussed strategies for optimization rely on balancing protein expression, but these strategies cannot overcome innate inefficiencies in enzymes themselves. Thus, engineering the proteins in the pathway can sometimes be the only way to improve the flux through the pathway.

As one of the most powerful tools in protein engineering, directed evolution mimics the Darwinian evolutionary process in a test tube and involves iterative rounds of creating genetic diversity followed by selection or screening (60-62) (Figure 1.5). The most common genetic diversity creation methods include error-prone PCR, DNA shuffling, chemical mutagenesis, use of a mutator strain, and saturation mutagenesis. To identify improved mutants from this genetic diversity, a myriad of screening/selection methods have been developed such as colorimetric assays, colony size-based growth assays, and fluorescence activated cell sorting (FACS). A major advantage of directed evolution is that no prior knowledge of the enzyme structure or mechanism is required to improve enzymatic properties. Another advantage is the ability to mutate the entire enzyme, thus identifying residues distant to the active site that can affect activity through allosteric interactions. However, a major disadvantage of directed evolution is the impossibility of exploring all sequence diversity, even with the most powerful screening or selection method. Additionally, it can be difficult and time consuming to develop a high throughput screening/selection method for a target enzyme property (Figure 1.5a).
Rational design is a knowledge-driven process which uses *a priori* information about enzymes such as sequence or structure. This knowledge is used to make specific, targeted amino acid mutations which are predicted to affect enzymatic properties vital for the efficiency of the desired reaction (Figure 1.5b). This strategy can be valued more than directed evolution because it limits the onerous task of screening the large libraries from directed evolution. In a sequence-based knowledge approach, researchers pursue systematic comparisons of homologous protein sequences to identify possible residues that could alter protein activity. When the three-dimensional crystal structure of the target enzyme or a homologous enzyme is available, a more direct structure-function relationship study between residues can be investigated. Critical residues that are hypothesized to affect the desired function are targeted for mutagenesis. By visualizing the active site, rational design can be used to mutate the large residues to smaller, hydrophobic residues, thus enlarging the active site and allowing a larger substrate to bind. Various computational tools have been developed to compare the homologous sequences and structural databases to create a mutability map for a target protein (63-65) (Figure 1.5b).

Distinctions between rational design and directed evolution are becoming less clear, as researchers commonly combine these techniques (Figure 1.5c). Though strategies to conjoin the methods vary, one strategy involves a two-step process of directed evolution and then identifying hotspots for rational design, or vice versa if used in *de novo* protein design. Another strategy to combine these techniques is semi-rational design. The power of semi-rational design lies in the reduction of the library size to be screened and an augmentation of the success rate for identifying positive hits (66). Semi-rational design can target specific residues for saturation mutagenesis or target and mutagenize a specific domain that is suspected to have critical
function. These intelligent libraries rely on the ability to identify key beneficial mutations through critical structure-function relationships and knowledge of mutational effects on protein folding and activity. This design process harnesses the power of the directed evolution but reduces the onerous screening. Techniques to develop these smart libraries have been diverse and efficient (61,62).

1.2.2.2.1 Engineering Enzyme Specificity or Selectivity

Both rational design and directed evolution have successfully been used to engineer enzyme specificity or selectivity to improve a pathway performance. Often, the promiscuity of enzymes tends to correlate with low catalytic activity and undesired side-reactions. This yields an inefficient pathway with build-up of intermediates and by-products, which can be toxic to the cell. By redesigning proteins for increased specificity or selectivity toward the desired substrate, less substrate is required for a high titer and fewer by-products are formed. This strategy has proven to be one of the most successful strategies to optimize a pathway through protein design. One particular example of reducing side-reactions of enzymes is the production of specialty chemical triacetic acid lactone (TAL) (67). The fatty acid synthase (FAS) is a bifunctional enzyme, but was rationally designed through sequence homology and structure-function relationships to inactivate the keto reductase domain. This eliminated the enzyme’s ability to utilize NADPH, making one function inactive. As a result, the carbon flux was shifted exclusively to the desired product TAL. Another example did not completely eliminate secondary activity, but increased selectivity towards the desired reaction. Nair and Zhao (68) engineered a xylose reductase with preferred selectivity for xylose over arabinose. Screening the error-prone PCR library revealed mutations clustering around the (β/α)₈-barrel, which led to
speculation that this region was involved in substrate recognition. Targeting this secondary structure, the site-saturation mutagenesis led to a final mutant with an increased selectivity from 2.4- to 16.5- fold preference for D-xylose.

Zhang and coworkers (69) engineered a promiscuous KivD for preferred substrate selectivity, which expanded E. coli metabolism to produce unnatural C5 to C8 alcohols, used for biofuels. The KivD specificity constant $k_{cat}/K_m$ was engineered through rational design to 40-fold higher preference for 2-keto-4-methylhexanoate than its cognate substrate, 2-ketoisovalerate. This was a 10-fold increase over the wild-type selectivity. The α-keto acid binding pocket was enlarged by mutating proximal residues to smaller, hydrophobic amino acids, allowing binding of the larger substrate. In this same report, LeuA was rationally designed for increased activity for the substrate of interest, 2-keto-3-methylhexanoate. The crystal structure suggested that steric hindrance of the methyl group in the 2-keto-3-methylhexanoate could be relieved by mutating proximal residues. The mutation, S139G, was identified to increase the $k_{cat}$ by 7-fold towards the substrate of interest. The combination of the mutated KivD and LeuA produced several non-natural alcohols, which were not produced in the wild-type pathway.

Just as significant as the substrate, the cofactor NAD(P)$^+$ and NAD(P)H usage can be engineered to increase the overall efficiency of the pathway. Properly designed proteins can reduce the competition for the cofactors between cell metabolism and enzymes in the pathway. An internal cofactor regeneration strategy was incorporated into the xylose utilization pathway to reduce the imbalance between the xylose reductase (XR) and xylitol dehydrogenase (XDH) for biofuel production. In one study, the XDH enzyme was engineered for switched cofactor specificity to
NADP\(^+\), which would complement the XR cofactor preference of NADPH (70,71), creating an internal cofactor regeneration mechanism. In another study, the XR was altered to have a preference for NADH, to complement the XDH of NAD\(^+\), which led to a 40-fold increase in ethanol productivity over the wild-type enzyme (72).

1.2.2.2 Improving Protein Activity

Though improving the substrate/cofactor specificity or selectivity of enzymes can increase flux, there are many other ways to engineer enzymes to increase the overall performance of a pathway. More active enzymes can increase the performance of pathways, and the higher enzymatic activity can be achieved by improving the catalytic efficiency, protein solubility, and stability. Additionally, the pathway can be optimized by increasing the activity of the transport system that transfers the extra-cellular substrates into the cell.

Leonard and coworkers (73) engineered two enzymes from the diterpenoid biosynthetic pathway, geranylgeranyl disphosphate synthase (GGPPS) and levopimaradiene synthase (LPS) for levopimaradiene production. Saturation mutagenesis was applied to possible critical residues of the LPS while error-prone PCR was utilized to engineer the GGPPS. The most improved GGPPS (S239C/G295D) and LPS (M593I/Y700F) were cloned into the full pathway and when combined with metabolic engineering strategies, resulted in a 2,600-fold increase in levopimaradiene production. The mutations in the GGPPS were hypothesized to affect the GGPPS catalysis by improving the binding efficiency of the magnesium ions needed for substrate anchoring.
Increasing the overall enzyme activity through traditional directed evolution was eloquently illustrated by Atsumi and Liao in a study focusing on 1-propanol and 1-butanol synthesis in *E. coli* (74). A combination of error-prone PCR and DNA shuffling through six rounds of directed evolution identified a top candidate, called CimaA3.7. The $k_{cat}/K_m$ of CimaA3.7 for acetyl-CoA was 6.7 times greater than the wild-type enzyme at 30 °C. The enzyme activity was screened at varying temperatures and CimA3.7 showed higher activity than wild-type at all temperatures tested. Once re-inserted into the pathway, this enzyme yielded a 9.2-fold improvement in 1-propanol production and a 21.9-fold improvement in 1-butanol production.

Transporters can be the greatest limiting factor in a pathway. However, engineering these proteins has been an underutilized strategy for improving the performance of pathways. Lack of three-dimensional crystal structure has limited the engineering of these proteins to error-prone and directed evolution. Young and coworkers (75) engineered xylose transporters with an improved growth rate of up to 70% on xylose as the sole carbon source. In this engineering effort, the transporters with increased affinity towards xylose were also discovered to have a lower $V_{max}$.

1.2.3 Pathway Libraries for Pathway Optimization

Pathway library screening strategies, as compared to traditional pathway engineering strategies, can be more efficient in the search for an optimized pathway. Traditional strategies optimize individual components of the pathway one at a time to increase flux to the desired product (71,73,76), as previously discussed. However, pathway libraries screening and combinatorial strategies can tune multiple components of the pathway simultaneously, which increases the likelihood of obtaining an optimized flux via balanced gene expression and protein activity.
within the pathway. These combinatorial mutagenesis strategies allow for a comprehensive exploration of the potential diversity of a target pathway, which could discover unexpected synergistic effects and identify the global optimum across the pathway amongst several mediocre, local optimums (77,78). Additionally, a pathway library strategy is advantageous because one does not need a priori information about pathway bottlenecks, as all components of the pathway are simultaneously tuned.

1.2.3.1 Pathway libraries constructed through traditional DNA Assembly Methods

One of the first examples of pathway libraries was demonstrated in controlling gene expression within an operon. Smolke et al. (58) showed that altering the mRNA stability by engineering the intergenic regions would increase protein production and thus overall titer. This idea of independently controlling gene expression within an operon was further explored by Pfleger et al. (24). By creating a library of tunable intergenic regions (TIGRs), the expression of several genes within an operon was simultaneously tuned for optimal expression. The TIGRs contained control elements that include mRNA secondary structures, RNase cleavage sites, and ribosomal binding site (RBS) sequestering sequences. The library of mutated intergenic regions was constructed and applied to the mevalonate pathway. The final mutant product concentration was increased sevenfold compared to the original wild-type pathway.

1.2.3.2 Pathway Libraries Constructed Through Advanced DNA Assembly Methods

The Gibson assembly method was applied to generate a large combinatorial library of promoters and enzymes of a heterologous acetate utilization pathway in E. coli, comprised of an acetate
kinase (*ackA*) and a phosphotransacetylase (*pta*) (79). This combinatorial library was based on three promoter sequences with assorted strengths and four orthologous variants of both genes, generating 144 possible unique combinations of the promoters and genes. The *ackA* and *pta* gene variants were chosen from ClustalW protein alignment scores from the organisms *Methanosarcina acetivorans* (mac), *Clostridium phytofermentans* (cph), *Ruminococcus gnavus* (rgn), and *Pelobacter carbinolicus* (pca). Each gene cassette was synthesized with a RBS, a terminator, and a unique 40-bp DNA linker sequence consisting of homologous DNA directly upstream and downstream of the gene at the terminal ends of the cassette (Figure 1.6). The same 40-bp linker sequence was also cloned into the DNA cassette of the promoters: *recA* (high strength), *ssb* (medium strength), and *lacI* (low strength). To force the gene assembly into a defined promoter-gene order, a unique linker sequence was used for gene 1 and another unique sequence used for gene 2.

The library was assembled via a Gibson assembly reaction, with equimolar amounts of each DNA fragment, then transformed into an acetate utilization knockout *E. coli* strain, and plated on potassium acetate as a carbon source for selection. The total library size was approximately $10^4$, affording 70-fold coverage of the 144 possible combinations. Investigation of the assembly efficiency showed that over 80% of the selected clones harbored a correctly assembled pathway. Sequencing data accounted for 8/16 possible gene pair combinations and 7/9 possible promoter pair combinations, and 14 unique promoter-gene combinations were present out of a theoretical 144. A bias was noted towards a specific combination of genes from the organisms cph and pca, even though each gene fragment was assembled in equal combinations. This bias could have been the result of an assembly bias or it could be the result of a screening bias, as the library was
screened on acetate and these genes could be the most efficient for acetate utilization in *E. coli*. Though pathway optimization was not explicitly demonstrated, the efficient and diverse assembly of genes and their regulatory elements into a large pathway library was established.

Simultaneously modifying all the regulatory elements of the pathway has also been accomplished through ePathBrick from the BioBrick standards assembly platform (47). In this strategy, a more concise library was generated, with each combination constructed individually by the ePathBrick vectors. With this system, Xu and coworkers demonstrated a modular engineering approach for significant titer improvement of a multi-gene fatty acid metabolic pathway by fine tuning gene expression through plasmid copy number and RBS engineering (80). The *E. coli* fatty acid biosynthetic pathway was apportioned and overexpressed into three modules. The first module, the GLY module, contained genes from the glycolysis pathway, the second module, ACA, was the acetyl-CoA activation module, and the final module, FAS, was composed of the fatty acid biosynthetic pathway genes. These modules were successfully expressed on compatible ePathBrick vectors with varying plasmid copy numbers. The total fatty acid production was optimized by overexpressing each module on high, medium, or low copy number plasmids. The highest fatty acid production occurred only when the FAS module was highly expressed and the GLY module was under moderate gene expression, while no correlation was found for ACA expression. The balanced gene expression pathway produced a 4-fold increase in fatty acid titer compared to the lowest-producing pathway.

This pathway library illustrated that the supply of malonyl-CoA (GLY module) and consumption of malonyl-ACP (FAS module) were very important for pathway optimization. Further
improvement strategies focused on translation initiation rates of these modules because RBS engineering has proven to be successful in pathway optimization (81). A library of three different strength RBSs were chosen from the MIT Registry of Biological Standard Parts and applied to the GLY and FAS modules. Maximum fatty acid production was observed under strong RBS of the FAS module and moderate strength RBS of the GLY. The RBS engineering improved fatty acid production by an additional 1.3-fold compared to the pathway with the optimal plasmid copy number.

A randomized BioBrick strategy has also been developed, which combines the power of Gibson assembly and the modularity of the BioBrick standards (82). In this method, all promoters, ribosome binding sites, and transcriptional terminators were randomized within the pathway. Each component was cloned with 18-28 bp linkers of homologous DNA regions to the 5’ and 3’ DNA. Three promoters, three RBSs, and three terminators were simultaneously randomized for the three gene pathway for the lycopene biosynthetic pathway (crtE, crtB, and crtI), generating a library of nearly 20,000 unique clones. After assembly, the library was screened on plates for the orange-colored lycopene product. Of the red-orange colored colonies, 12 were selected and DNA sequencing analysis demonstrated that 7/8 randomized pathways were distinct and four pathways had gene deletions. The study cautions the metabolic burden placed on the cells during the library screening which could have caused the mutations.

A modified version of the DNA assembler method was used by Wingler and Cornish to establish a reiterative recombination method for the in vivo assembly of multi-gene pathway libraries into the chromosome (83). This method involved a 5-day multi-step process for chromosomal
integration of several genes one-by-one. The strategy utilized a pair of alternating orthogonal endonucleases and selectable markers. Homologous recombination and gap repair were used to construct a plasmid containing the gene of interest, marker, and endonuclease, which were recombined into an acceptor strain. This acceptor strain carried a predefined target locus for integration into the chromosome. Galactose-induced expression of the endonuclease initiated the integration. The endonuclease cleaved the double stranded DNA, triggering the homologous recombination and leading to integration of the gene of interest and the auxotrophic marker into the chromosome. The strains were then selected for the new auxotrophic marker and cured against excess donor plasmid. Proof of concept for pathway integration and mock library assembly was demonstrated using the lycopene biosynthetic pathway. Three rounds of induction, curing, and selection were needed to assemble the three pathway genes: \textit{crtE}, \textit{crtB}, and \textit{crtI}. The resulting recombinants from the pathway integration were screened on plates through the orange-colored phenotype of the lycopene metabolite and 99% exhibited the correct assembly. Wingler and Cornish also investigated if this method could be used to create large (>10^4) pathway libraries. The library contained different combinations of the genes in the chromosome, constructed by integration of the plasmid with a library of variant genes at each integration step. To investigate the diversity of the library, the mock library contained various ratios of \textit{crtB} and \textit{crtI} alleles that contained either nonsense or silent mutations, which could produce interrupted or working pathways. The diversity could be judged based on the actual and theoretical percentages of working pathways versus interrupted pathways, visualized based on the color of the colonies on the plate. Each library had the expected percentage of working pathways, indicating a non-biased library assembly into the chromosome.
An example of plasmid-based pathway libraries was constructed by the DNA assembler method and focused on a combinatorial library of different promoter strength for all the genes within the library (84). As a proof of concept in pathway library generation, the xylose and cellobiose utilization pathways for ethanol production were optimized. The fungal xylose utilization pathway consists of three heterologous genes, xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulose kinase (XKS), each expressed under control of PDC1 promoter, ENO2 promoter, and TEF1 promoter, respectively. The cellobiose utilization pathway consists of the β-glucosidase (BGL) and cellodextrin transporter (CDT) under the ENO2 and PDC1 promoters, respectively. The promoters PDC1, ENO2, and TEF1 were mutagenized through nucleotide analog-based error-prone PCR to induce a very high mutation rate and produce promoters of various strengths. After mutagenesis, mutants for each promoter were characterized through green fluorescence protein expression and ten promoters of defined strengths were selected for library construction. These ten promoters in each position of the library resulted in a theoretical library size of $10^2$ and $10^3$ for the cellobiose and xylose utilization libraries, respectively. Each mutant promoter was cloned into a helper plasmid which contained 400 bp sequences homologous to the 5’ DNA region (Figure 1.6). The mutant promoter/gene expression cassettes were co-transformed into a yeast strain with a total library size of $10^5$. To confirm the diversity of the library, over 40 individual colonies from each library were screened from an antibiotic selection marker for plasmid-pathway assembly and not based on sugar utilization. Each colony from this plasmid-marker selection exhibited a unique growth curve on its respective carbon source, which was indicative of a diverse library.
To screen for an optimal pathway with increased sugar utilization, improved pathway phenotype was assessed on the basis of its ability to consume the sugar for growth when that sugar is the sole carbon source available. This could be visualized in a high-throughput manner through colony size on agar plates, wherein larger colony sizes were suggestive of faster sugar utilization and improved growth. In the xylose utilization pathway, a very efficient mutant pathway was identified in a single step. This pathway conferred a xylose consumption rate of 0.73 g/L·h, comparable to some of the fastest xylose consumption rates from strains which had been subjected to multiple generations of optimization strategies. The strain harboring the wild-type pathway did not produce any ethanol, while the mutant pathway conferred an ethanol productivity of 0.17 g/L·h. In the cellobiose utilization strategy, the strain harboring the optimized pathway yielded a 5.4-fold improved cellobiose utilization rate and a 5.3-fold increase in ethanol productivity in a single optimization step.

1.2.4 Engineering Biochemical Pathways for Sustainable Biofuels

As previously mentioned, hexose (D-glucose) is a main constituent of biomass; however pentose sugars such as D-xylose and L-arabinose from the hemicellulose fraction can contribute to a large component of the biomass. An efficient lignocellulosic biofuel production process will need to utilize all sugars from the biomass. However, *S. cerevisiae* and *E. coli* cannot utilize the pentose sugars. Thus, a significant amount of engineering has occurred to design these microorganisms to efficiently utilize these sugars. Bio-ethanol has emerged as a first-generation biofuel and has been the focus of many biofuel processes. However, recent research thrusts have investigated more efficient fuels, as bioethanol has low heating properties and is incompatible with current engines. Advanced biofuels such as butanol, biodiesel, alkanes, and hydrocarbons have become a
major research area. Though many studies recently focused on producing advanced biofuels, either through *E. coli* or *S. cerevisiae*, no study has linked pentose utilization to advanced biofuel production.

### 1.2.4.1 Pentose Sugar Utilization and Bioethanol Production

Optimizing the pentose sugar utilization pathway has two main research areas: engineering transporters to transport the sugars into the cell without incurring a diauxic shift and optimizing the enzymes which convert the sugars to ethanol.

The diauxic shift describes a phenomenon wherein a microorganism will preferentially utilize hexose sugars in a mixture of pentose and hexose sugars. This sequential sugar utilization greatly decreases overall efficiency of the process (10,85-87). Many metabolic engineering techniques have been used to try to overcome this phenomenon. It has been shown that using pentose-specific transporters can reduce the diauxic shift, because the sugars do not enter the cell by the same transporters as the glucose molecules, allowing for simultaneous transport. More than just discovering pentose-specific transporters, it is also possible to engineer the transporters for a higher activity (75,88).

Another strategy for reducing the diauxic shift is to utilize cellobiose as a carbon source. This strategy also relies on using individual transporters for the pentose and hexose sugars. But a very efficient cellobiose transporter was found to be specific for cellohexo- and celloxose (89). Previous work has reduced the diauxic shift by co-expression of the cellohexo and celloxose transporter and a β-glucosidase with the xylose utilization genes (90,91).
One of the most successful strategies for xylose utilization is the heterologous expression of the fungal xylose utilizing pathway. The xylose is reduced to xylitol by a xylose reductase (XR). Then the xylitol is oxidized to D-xylulose by a xylitol dehydrogenase (XDH). Finally, a xylulose kinase (XKS) converts the D-xylulose to D-xylulose-5-P, which is then shuttled to the pentose phosphate pathway for ethanol production (Figure 1.7). The arabinose pathway is similar: the arabinose is reduced to L-arabitol by the XR, a L-arabitol 4-dehydrogenase (LAD) reduces the arabitol to L-xylulose, which is then converted to D-xylulose by a L-xylulose reductase (LXR).

Though the xylose pathway has shown to be efficient (84,92,93), the arabinose pathway is not. The imbalance of the cofactor usage between the enzymes is suspected to be the greatest limiting factor (94). Strategies to overcome this limitation were previously discussed in the chapter, involving engineering enzymes for altered cofactor specificity. Thus internal cofactor regeneration mechanisms were established. Significant work has looked at balancing the promoters of the xylose utilization enzymes. These studies have shown that a high XR expression is important for an efficient pathway (84).

1.2.4.2 Advanced Biofuels

Advanced biofuels, such as butanol, biodiesel, alkanes/alkenes, and hydrocarbons have more similar properties to petroleum-based fuels than ethanol, and are termed ‘drop in’ fuels because they can readily be used by current engines in cars and planes. Recent work has focused on the Ehrlich pathway or the 2-keto-acid pathway for production of butanol and isobutanol. Intermediates of the amino acid pathway, keto acids, are decarboxylated into aldehydes and then
reduced to the products. Isobutanol is perhaps one of the products produced by microbes which are closest to industrial use, with a high energy content and extensive branching (higher octane number). The octane number is a measure of a fuel’s resistance to knocking in spark ignition engines (95). The 2-keto-acid pathway was introduced into *E. coli*, the promiscuous decarboxylase and aldehyde dehydrogenase were engineered for increased substrate specificity in order for the final production of the non-natural carbon chains (69).

Coupling microbial catalysis and chemical catalysis has been shown to be successful with isoprenoid based biofuels, such as farnesane (96) and bisabolane (97). The intermediates farnesene and bisabolene are produced microbially and then chemically converted to farnesane and bisabolane, which are currently being investigated for as alternative biosynthetic diesel. Plants are known to produce terpenes, thus, plant isoprenoid synthases were screened for efficient conversion of farnesyl diphosphate from the mevalonate pathway into bisabolene (97). The cell membrane mainly consists of fatty acids. Certain organisms often overproduce fatty acids which are used for energy storage. The long hydrocarbon fatty acyl chain is energy rich, and can be used as a biofuel. Fatty acids are biosynthesized naturally by either a large multienzyme system the fatty acid synthase or small sub-unit proteins. The multienzyme system uses malonyl-CoA as a building block. Fatty acyl chains are elongated on acyl carrier proteins (ACPs) through repeated cycles of decarboxylative condensation, β-keto reduction, dehydration and enol reduction (95). Thioesterases have been used to release the long-chain fatty acids to produce free fatty acid ethyl ester (FAEEs) (98). Combined with the deletion of genes in the fatty-acid degradation pathway (β-oxidation), free fatty acids have been produced at high titers (1.2 g/L) in *E. coli* (99).
Alkane biosynthesis has recently been a major research focus. Derived from fatty acids, alkanes are a predominant component of diesel. Alkane biosynthetic genes were recently discovered in cyanobacteria (99). The acyl-ACP reductase (AAR) was used to reduce the acyl-ACP to aldehydes which were then converted into alkanes by an aldehyde decarboxylase (ADC). A combination of expression of these enzymes in E.coli produced alkanes from C_{13}-C_{17} with final yield of 300 mg/L. A figure summarizing some of the advanced biofuel pathways is available as Figure 1.8. A more detailed discussion of advanced biofuels is available in Chapter 4.

1.3 Scope of Thesis
There are two major components of this thesis: pathway optimization and pathway construction. This thesis is explores new pathway optimization methods through combinatorial protein engineering strategies on the pathway scale. Different schemes for protein pathway libraries are described in Chapters 2 and 3. Chapter 4 describes the design and construction of a heterologous pathway for FAEE production, establishing the pathway using a various combination of traditional pathway optimization strategies to establish the pathway. Chapter 5 once again considers protein pathway libraries in the investigation of a high throughput screening strategy. It is expected that once a high-throughput screen is established, the combinatorial pathway libraries strategies established in Chapters 2 and 3 can be applied to the FAEE pathway.

1.3.1 Chapter 2 Project Overview
This chapter introduces and discusses the importance of cellobiose utilization in S. cerevisiae for biofuel production. Cellobiose is a sugar released from biomass pretreatment and efficient cellobiose utilization by S. cerevisiae could help to make the biofuel production process more
economical. The original wild-type utilization pathway was discovered by our collaborators at UC Berkeley, consisting of a cellodextrin transporter (CDT-1) and a β-glucosidase (GH1-1), both heterologously expressed in *S. cerevisiae* from the filamentous fungi *Neurospora crassa* (89-91). Though the pathway was functional and *S. cerevisiae* was able to utilize cellobiose, it was not very efficient. The strain harboring the wild-type pathway took nearly 50 hours to consume all the cellobiose and unwanted reactions were creating detrimental by-products. We investigated a combinatorial pathway library strategy to consider many different homologous cellodextrin transporters and β-glucosidases. Due to difficulty in cloning and low protein activity, the library consisted of seventeen β-glucosidase enzymes and four cellodextrin transporters. After screening the library based on specific growth rate on cellobiose and ethanol productivity, no statistically significant combination of proteins were found to be better than the original wild-type pathway.

### 1.3.2 Chapter 3 Project Overview

Chapter 3 resumed the work of Chapter 2, in the endeavor to optimize cellobiose utilization in *S. cerevisiae*. We chose to continue working with the original wild-type pathway (CDT-1 and GH1-1) due to the abundance of data available on this pathway. Our next approach for optimizing the pathway was based on protein engineering. Often protein engineering strategies will engineer one enzyme at a time, but the strategy we developed allowed for all proteins in the pathway to be engineered simultaneously. In this work, we chose to conduct error-prone PCR to introduce random mutations into the proteins of the pathway, screening for the desired phenotype, and then repeating the mutagenesis in a directed evolution manner. After two rounds of directed evolution, we identified a 64% increase in ethanol productivity and a 49% increase in cellobiose consumption. The β-glucosidase protein which conferred the improved phenotype was
shown to have two mutations, one from each round of directed evolution: H23L and L173H, respectively. The cellobextrin transporter which conferred the improved phenotype was also shown to have two amino acid mutations, one from each round of directed evolution, D433G and C82S, respectively. We characterized the mutations to rationalize how these mutations could have conferred an increase in the pathway phenotype. *In vitro* protein assays of the β-glucosidase suggested an increased specificity towards cellobiose as a substrate. It was found in the homology structural model of the β-glucosidase, that the L173H mutation introduced a new hydrogen bond directly to the cellobiose molecule, which has been shown to increase substrate specificity (100-102). Analysis of the cellobextrin transporter suggested an improved overall activity, mainly due to mutation C82S. A sequence-mapping software predicted this mutation was located in the first transmembrane helix.

1.3.3 Chapter 4 Project Overview

Chapter 4 changed directions from Chapter 2 and 3 slightly, focusing on the construction of a heterologous pathway, which will later be optimized through the strategies of the previous chapters. The pathway of interest was for the heterologous production of FAEEs, biodiesel. Many reports in FAEE production in yeast have significantly low titers. It is expected that this low titer is due to the endogenous regulation of the cell, limiting the energy intensive fatty acid synthesis (103). We hypothesized that we could overcome the regulation by overexpressing a heterologous fatty acid synthase, which would not be subject to endogenous regulation and would provide a completely orthologous fatty acid synthesis route, to decrease the re-direction of endogenous fatty acids from cellular metabolism. The pathway consists of the wax ester synthetase/acyl-coenzyme A:diacylglycerol acyltransferase (WSDGAT or WS) enzyme, which catalyzes the transesterification of a fatty acid with ethanol to produce the FAEE and a
heterologously expressed fatty acid synthase (FAS). Many traditional pathway engineering strategies were employed in establishing this pathway: varied copy number plasmids, combinations of different plasmids, use of different heterologous enzymes, and employment of various promoters for strongest expression. This project illustrates the arduous endeavors employed in traditional pathway engineering. After testing several promoters, a strain expressing the \textit{ws} gene and the \textit{fas} gene under control of the HXT5 promoter (DE04) resulted in a 6.3-fold increase in total FAEE production compared to a strain expressing only the \textit{ws} gene. Independent biological clones for this project were shown to have great variation in the FAEE titer, however the strain DE04 reproducibly produced significantly increased titer of FAEEs.

\textit{1.3.4 Chapter 5 Project Overview}

After establishing the desired pathway and confirming heterologous expression of the genes in Chapter 4, Chapter 5 revisits the pathway-scale library strategies of Chapter 2 and 3. However, a high-throughput screening method must first be developed. This chapter investigated three different strategies for a high throughput screening strategy to optimize either fatty acid production or FAEE production. The first strategy was a growth-based strategy which was applied in a \textit{S. cerevisiae} strain with the endogenous fatty acid synthesis knocked-out. This allowed for a growth-based screen relying on the complementation of the heterologous FAS to provide FA for growth. The second strategy investigated the use of Nile Red stain, which is a widely-known and commonly-used quantification method for fatty acids. A final screening strategy involved the promoter POX1, which is shown to be sensitive to fatty acid concentration. The promoter controlled GFP expression and through fluorescence studies, the fatty acid concentration could be monitored. Out of the three strategies, it is suggested to use the growth
curve strategy in the FAS knockout-strain. However, further work must be repeated in the Nile Red Assay.

1.4 Conclusions
The global energy demand is rising at an alarming rate. To meet these escalating demands, alternative energies that promote sustainable development must be identified. One of the most exciting alternative energies is biofuels. Engineering microorganisms to produce these biofuels is promising. Though there are many possible ways to engineer these microorganisms, pathway engineering and optimizing the flux through the pathway has been very successful. Protein engineering can be particularly helpful in optimizing the flux and this thesis focuses on the development of new protein based combinatorial strategies to optimize a pathway. A novel biofuel pathway was also constructed and preliminary work in designing a screening method was accomplished, for the application of the protein library strategies to optimize the newly constructed pathway.

1.5 References


---

1.6 Figures
**Figure 1.1:** Energy consumption by region as predicted by British Petroleum in 2012 (4). Global energy demand is increasing significantly. Alternative energy sources must be found to meet these increasing demands.

**Figure 1.2:** Breakdown of energy resources and consumption in the United States in 2011. Each number on the line indicates the percentage of that component (6). Transportation fuels utilize the greatest percentage of petroleum and one of the largest sectors of energy consumption.
**Figure 1.3:** Cellulose is a long polysaccharide chain of glucose monomers. This is the largest constituent of biomass and must be broken down into simple sugars (such as a glucose monomer) for efficient biomass conversion to biofuels.

**Figure 1.4:** Illustration of the many components of biomass (104). Cellulose makes up the largest component, with pentose sugars in the hemicellulose fraction.
Figure 1.5: A schematic of main protein engineering strategies consisting of directed evolution, rational design, and a combined approach. The combined method uses both directed evolution and rational design, though the methodology can vary, one example is using directed evolution which can identify hotspots and then rational design proximal residues to those hot-spots.
**Figure 1.6:** Preparation of DNA fragments for large library generation. Each unique design represents a unique promoter or gene. Varied strength promoters, orthologous genes, or mutated pathway components generate diversity. If the DNA is assembled with homology regions, upstream and downstream of the DNA fragment of interest, the pathway can assemble properly into many different combinations. Each strategy has incorporated different lengths of homology, which can contribute to the efficiency of correct assembly. These DNA fragments are then subjected to the desired DNA assembly reaction with the linearized vector and transformed into the host.
Figure 1.7: Suggested pathway for pentose sugar utilization in yeast, based on the fungal pathway. These pathways were studied extensively by the Zhao Research Group.
Figure 1.8: Biosynthetic pathways for different advanced biofuels (105).
Chapter 2: Combinatorial Library of Homologous Proteins for the Cellobiose Utilization Pathway

2.1 Background

2.1.1 Cellobiose Utilization

The saccharification of lignocellulosic biomass into fermentable sugars is recognized as one of the most expensive operations in biofuel process economics (1,2). Hydrolysis, an enzymatic process for biomass pretreatment, requires an exorbitant amount of enzymes, to release the simple fermentable sugars. The first step in the process requires endoglucanases (EC 3.2.1.4) and exoglucanases (EC3.2.1.74) to collectively break down the cellulose into cellobiose, a β-1,4-glucose disaccharide. Then, β-glucosidases (EC 3.2.1.21) hydrolyze the cellobiose to glucose, which is subsequently fermented by normal cellular metabolism (3). In this process, not only is a significant amount of enzyme required to completely break the biomass down into the simple sugar glucose, but there is also another problem: an inhibiting build-up of the intermediate cellobiose (1,2). Therefore, to make this process more efficient and to reduce the requirement of these in vitro hydrolytic enzymes, recent research has investigated the option of utilizing cellobiose itself as a fermentable sugar. Cellobiose utilization would be advantageous to the biofuel production process by reducing the in vitro enzyme load and inhibition, thus reducing the price of pretreatment, lowering the overall cost of biofuel production (1,2).

---

2 Parts of this chapter were previously published in:
More than just improving the pretreatment process, cellobiose utilization can also be used to overcome a major challenge in biofuel production: the diauxic shift. The diauxic shift is characterized by a lag in growth when the organism’s metabolism shifts from its preferential carbon source hexose sugars to pentose sugars. The major sugars released from biomass include both pentose and hexose sugars and in an efficient process the microorganisms will utilize both sugars simultaneously (4,5). *S. cerevisiae* exhibits a very significant diauxic shift in a mixture of xylose and glucose, preferentially utilizing the glucose molecules first before slowly consuming the xylose. Engineering *S. cerevisiae* to co-utilize pentose sugars and hexose sugars has been a major research thrust and has been an extremely difficult challenge (5-9). It was hypothesized that by engineering *S. cerevisiae* to utilize cellobiose, efficient co-fermentation of cellobiose and xylose could be achieved (10-12).

Three main strategies for heterologous cellobiose utilization have been developed. The first strategy involves cell-surface display of extracellular β-glucosidases (13-17) (Figure 2.1a). In this strategy, the cellobiose is hydrolyzed extracellularly by β-glucosidases docked to the surface of the yeast, then the cell transports and metabolizes the glucose molecules. The hydrolytic pathway is a second strategy, which involves heterologous expression of a cellodextrin transporter and an intracellular β-glucosidase (11,12). After being shuttled into the cell by the transporter, the cellobiose is hydrolyzed by the β-glucosidase into two glucose molecules, which are then metabolized by the cell (Figure 2.1b). A third strategy, the phosphorolytic pathway, relies on heterologous expression of a cellobextrin transporter and a cellobiose phosphorylase (18-20) (Figure 2.1c). The cellobiose is transported into the cell and the intracellular phosphorylase cleaves the disaccharide with an inorganic phosphate, producing a glucose
molecule and α-glucose-1-phosphate, which is quickly metabolized. Though each strategy has its own advantages, no method is as efficient as glucose consumption, which is considered one of the most rapid sugar utilizations. Further engineering is required to improve the cellobiose utilization rates to rival glucose consumption.

Studies in our laboratory have focused on the hydrolytic cellobiose utilization pathway. Though the current pathway is functional, it is not efficient (12). We endeavored to optimize the pathway productivity and sugar utilization. Chapter 1 gave an in-depth description of both traditional and pathway library optimization strategies. Promoter-based optimization has been extremely successful in both of these strategies, even for the cellobiose utilization pathway (21,22). However, engineering the promoters will not optimize the intrinsic inefficiencies of the enzymes themselves. The detrimental effects of side reactions, cofactor specificities, or substrate specificities cannot be improved through promoter engineering. As such, the enzymes of the cellobiose utilization pathway were shown to be inefficient (12), the β-glucosidase exhibited transglycosylation activity and the cellodextrin transporter would transport the cellodextrins outside the cell. Therefore, it is important to consider optimizing the pathway based on the enzymes within the pathway. It is possible to consider a synthetic biology approach to use different orthologs of homologous enzymes that can confer varying activities and characteristics for pathway optimization. A combinatorial library search strategy with these enzymes could be advantageous (23).

2.1.2 Previous Research in Enzyme Combinatorial Library Strategies

The fungal xylose utilization pathway has been shown to be especially sensitive to cofactor imbalances and unbalanced enzyme expression (24-27) and was selected as a good target
pathway to optimize through this strategy. Various combinations of enzymatic properties such as catalytic efficiency, cofactor specificity, stability, and substrate specificity could be explored, allowing for a comprehensive search of the entire potential pathway optimums to identify the most improved optimum.

The homologous genes used in the library were identified through a nucleotide BLAST search of the xylose utilizing enzymes including xylose reductase (XR), xylitol dehydrogenase (XDH), and xylulose kinase (XKS) from the organism Scheffersomyces stipitis. Twenty XR genes, twenty-two XDH genes, and nineteen XKS genes were cloned, which resulted in a total theoretical library size of 8,360 possible unique combinations of the corresponding enzymes. Each enzyme was characterized to show varied activities and cofactor dependencies. The homologous gene sequences were cloned into helper plasmid expression cassettes, containing promoters, terminators, and at least a 400 bp region homologous to the 5’ and 3’ DNA regions of the pathway at the termini of the expression cassette (Figure 2.2). The linear expression cassettes were transformed into a yeast strain, which were then recombined by the homologous recombination mechanism to create a diverse library of pathways containing a unique combination of the enzymes. To confirm the diversity of the library, over 40 individual colonies from each library were screened from an antibiotic selection marker for plasmid-pathway assembly, not based on sugar utilization. Each colony from this plasmid-marker selection exhibited a unique growth curve on xylose as a sole carbon source, which was indicative of a diverse library. Sequencing results of random colonies showed all the genes were recognizable with no major mutations or hybrids and there was no significant bias in the library towards a certain gene. This study improved the xylose utilization in S. cerevisiae and allowed for a
comparison study of fast and slow xylose utilizing pathways to identify important enzyme activity ratios and cofactor specificities of xylose utilization (23).

2.1.3 Strategy Applied to the Cellobiose Utilization Pathway

Similar to the study completed on the xylose utilization pathway, we assembled a library of genes from filamentous fungi encoding for enzymes which are homologous to the enzymes of the cellobiose utilization pathway. The wild-type heterologous pathway consists of a β-glucosidase from *Neurospora crassa* [GenBank Accession number XM_951090] and celldextrin transporter [GenBank Accession number XM_958708], also from *N. crassa* (11,12). Through this combinatorial strategy, we were able to explore various combinations of enzymes in the pathway with different enzymatic properties, such as catalytic activity and substrate specificity (Figure 2.3).

2.1.4 Enzymes of the Pathway

2.1.4.1 β-Glucosidase

β-Glucosidase enzymes catalyze the transfer of glycosyl groups between oxygen nucleophiles on different types of substrates and are classified into two glycosidase hydrolase families: GH1 and GH3 (28). Both of these families were represented in the library. GH1 enzymes are members of the 4/7 super family, consisting of a common (β/α)_{8}-barrel (29). The active site is located on β-strands 4 and 7, within a deep cleft formed by the connection loops at the C-terminal end of the β-sheets of the TIM barrel (30). The conserved catalytic residues of the GH1 family enzymes involve the catalytic acid/base Glu 166 and the nucleophile Glu 377, which cleave the glycosidic bond via the common double displacement mechanism (31). This mechanism acts on β-1,4-linked glucose derivatives and the enzyme is promiscuous, being active towards a range of
substrates with varying affinities (30). The GH3 family of the β-glucosidases has a two-domain structure, a (β/α)₈-barrel followed by a β/α-sandwich, comprising a 6-stranded β-sheet sandwiched between three α-helices on either side. The active site is situated between the (β/α)₈-and the (β/α)₆-sandwhich domains, each of which contributes one catalytic carboxylate residue (32). GH3 enzymes are suspected to be the best at hydrolyzing cellobiose (33). Thus our library consisted of both GH1 and GH3 β-glucosidase enzymes.

A strong reduction in hydrolytic activity has repeatedly been documented for substrate concentrations above the $K_M$ of the enzyme (28). The enzyme is very sensitive to product inhibition, which results in transglycosylation activity. Transglycosylation will yield cellodextrins rather than glucose molecules. During hydrolysis, the cellobiose is cleaved, yielding a glucose leaving group and a covalent glycosyl–enzyme complex. During the second step of the hydrolysis, the intermediate undergoes a nucleophilic attack by a water molecule. However, in transglycosylation, the nucleophile that attacks the covalent glycosyl-enzyme complex is the hydroxyl group of a cellobiose or glucose molecule, which forms a complex saccharide molecule instead of the glucose monomer.

Transglycosylation is considered a major problem in the saccharification of lignocellulosic biomass and was also observed in the original studies of the $S. cerevisiae$ cellobiose utilization with the accumulation of glucose and other cellodextrins (2,11,12,28,33). Different β-glucosidase enzymes have varied sensitivities to product inhibition, with some enzymes having extremely high tolerance (34). Therefore, it was hypothesized that conducting a combinatorial
search of homologous enzymes might result in a cellobiose utilization pathway with protein characteristics more conducive to optimal sugar utilization and ethanol productivity.

2.1.4.2 Cellodextrin Transporters

Biological membrane transporter proteins are divided into two groups: channels (facilitators) and carriers (symporters). Facilitators function as selective pores that open in response to a chemical stimulus, allowing the movement of a solute down the gradient. Symporters use an energy intensive process to translocate one or more substrates across the phospholipid membrane against a concentration gradient (35).

Structural characterization of these proteins is problematic, because membrane proteins are notoriously challenging to crystallize. One reason for the difficulty is the amphipathic nature of the protein surface, with a hydrophobic area in contact with membrane phospholipids on both sides of the membrane (36). However some three-dimensional molecular crystal structures of transporters have been reported and among the crystalized structures, major similarities are observed (35,37). The transporters all have α-helical structures of the membrane-spanning domains, with about 10-12 transmembrane helixes, and some helices have irregular shapes with kinks and bends (35).

In the original cellobiose utilization pathway (10-12), two cellodextrin transporters were characterized from *N. crassa*. The first transporter, cellodextrin transporter 1 (CDT1), is a symporter and the second, cellodextrin transporter 2 (CDT2), is a facilitator. Comparatively, CDT1 has a higher cellobiose uptake activity ($V_{max}$) but requires ATP for transport, while CDT2
has little energy requirements but has a much lower activity than CDT1 (10). Though the facilitator would theoretically be more efficient in anaerobic conditions, we chose to use homologs of both transporters in the library to incorporate a diverse library which would allow for a more comprehensive search for an optimal pathway.

2.2 Results

2.2.1 Gene Cloning

Homologous genes were identified using the bioinformatics tool NCBI BLAST. We investigated putative β-glucosidase genes encoded by the cDNA from a myriad of filamentous fungi. Over forty β-glucosidase genes were successfully amplified from the cDNA, cloned into the helper plasmid, and transformed into *S. cerevisiae* for activity analysis. Upon enzyme activity screening, it was found that only a few of the β-glucosidase enzymes had activity. Subsequent sequencing identified introns to be present in all the inactive genes. We removed the introns through rigorous gDNA decontamination and over-lap PCR construction as necessary. All genes were sequenced to confirm intron removal and proper construction, however only a few new enzymes exhibited activity. A final total of seventeen β-glucosidases were shown to have activity with *p*-nitrophenyl β-D-glucopyranoside (*p*-NPG), a common synthetic substrate for β-glucosidase (30,31). This substrate was used instead of the substrate of interest, cellobiose, due to its ease of use in high-throughput assays. It was noted that in the enzymes we had cloned, the family GH 1 has significantly higher activity than GH 3 for *p*-NPG (Figure 2.4).

Cloning cellobextrin transporters also proved to be difficult. Initial cloning characterized the transporter clones in a high copy plasmid. However, it was found that upon sub-cloning into a
single-copy plasmid (required for controlled library creation and screening), that many strains harboring the transporters were unable to grow on cellobiose (Figure 2.5). It was hypothesized that the activity of the transporters was so low that the high expression of the multi-copy plasmid was necessary to transport a sufficient amount of cellobiose needed for growth. Only three transporters (CDT-1 and CDT-2 and CDT-3) had activity levels which allowed for strains expressing the heterologous transporter in a single copy plasmid to grow on cellobiose.

At this junction, BP researchers discovered two new cellodextrin transporters, named AN1 and CG2, codon-optimized from *Aspergillus nidulans* and *Chaetomium globosum*, respectively. These proteins were discovered after a search for enzymes homologous to CDT-1 and CDT-1 in filamentous fungi. Reports from EBI (38) indicated a slower growth rate on cellobiose compared to CDT-1, but an increased ethanol production (Figure 2.6). These transporters were cloned into the single copy plasmid system and strains harboring these plasmids were able to grow on cellobiose as a sole carbon source. A list of all cloned β-glucosidases and cellodextrin transporters used in the library are available in Table 2.1.

2.2.2 Library Assembly

The combinatorial enzyme library was constructed using DNA Assembler, as previously described (21,23). The homologous gene sequences were cloned into helper plasmid expression cassettes, containing promoters, terminators, and at least a 400 bp region homologous to the 5’ and 3’ DNA regions of the pathway at the termini of the expression cassette. These expression cassettes were amplified from the helper plasmids and then the linear DNA fragments were combined in equimolar concentrations to create the DNA library. This library was then transformed into an industrial strain of *S. cerevisiae* and through the *in vivo* homologous
recombination mechanism, the linear DNA fragments were automatically assembled into a diverse pathway library. The library comprised of the seventeen β-glucosidase clones from GH1 and GH3 families, identified above, and the five active celloextrin transporters. These parameters yielded a theoretical library size of eighty-five possible unique combinations of the enzyme pathway.

The diversity of the library was confirmed by randomly selecting ten colonies from YPD plates containing a selection marker for correct plasmid assembly (not based on cellobiose growth) and then characterizing the diverse growth on cellobiose (Figure 2.7). The total library size screened on cellobiose plates was $10^4$. Over 1,000 colonies were screened for improved specific growth rate on cellobiose, visualized by large colonies on the agarose plate (21,23). Thirty of the largest colonies were selected for further characterization and fermentation screening.

The strains harboring these pathways were compared to a strain containing the original, wild-type pathway previously established (11,12). The wild-type pathway consists of β-glucosidase $gh1-1$ [GenBank Accession number XM_951090] and celloextrin transporter $cdt-1$ [GenBank Accession number XM_958708]. Two top candidates were identified to have a better or comparable phenotype to the strain containing the wild-type pathway. The plasmids from these selected strains were isolated and retransformed into fresh yeast to ensure that the improved phenotype was a result of the pathway and not genomic modification. After transformation, it was shown that these pathways did not have significantly improved activity compared to the wild-type GH1-1 and CDT-1 pathway (Figure 2.8). Additionally, no improvements in transglycosylation activity and transportation of celloextrins to outside the cell.
were observed. Thus, the search for more efficient enzymes for the cellobiose utilization pathway did not identify enzymes with improved characteristics. However, the pathways were a unique combination of the β-glucosidase and cellodextrin transporters for a novel cellobiose utilization with similar characteristics of the wild-type pathway.

2.3 Discussion and Conclusion

Though no significantly improved pathway was discovered, some observations about the enzymes of the best performing pathways can be made. The screen utilized for this library was based on fast growth rate on cellobiose. All of the pathways which conferred the highest growth rate on cellobiose contained the CDT-1 transporter, suggesting that this transporter conferred an activity which is needed for fast growth. Screening the library for a different property, such as ethanol production, might identify a strain consisting of another cellodextrin transporter which may confer a slower overall growth rate on cellobiose, but exhibit an increased ethanol productivity.

All of the fastest growing pathways contained the CDT-1 transporter, suggesting that this is the transporter which confers the fastest growth rate. However there was more diversity in the β-glucosidase enzymes screened from the library. Pathway 1 consisted of the β-glucosidase from *Aspergillus niger* [GenBank Accession number XM_001390218] in the GH1 family. Pathway 23 consisted of the β-glucosidase from *Podospora anserina* [GenBank Accession number XM_001908385], also in the GH1 family. These two enzymatic pathways performed as well as the wild-type pathway containing the β-glucosidase, GH1-1, from *N. crassa* [GenBank Accession number XM_951090]. β-glucosidases from *A. niger* are well-known for their high activity and use in hydrolytic processes. To the best of our knowledge, *P. anserina*
β-glucosidases are not largely known for high activity and are not frequently used in hydrolytic processes. This study suggests that a more rigorous investigation of *P. anserina* for β-glucosidases could lead to the discovery of important enzymes that could be useful in improving hydrolytic processes. All of the fast-growing strains on cellobiose contained pathways with β-glucosidases from the GH1 family. Though it was believed that enzymes from the GH3 family are best at hydrolyzing cellobiose (33), this study suggests that to achieve fast growth rates on cellobiose, the GH1 β-glucosidase family may be more useful than β-glucosidases from the GH3 family.

2.4 Materials and Methods

2.4.1 Strains, Media, and Culture Conditions

The industrial *Saccharomyces cerevisiae* strain Still Spirits (Classic) Turbo Distiller’s Yeast was purchased from Homebrew Heaven (Everett, WA). Yeast strains were cultivated in YP media (1% yeast extract, 2% peptone) with 2% glucose (YPD) or 2-8% cellobiose (YPC). YPC with 8% cellobiose used in fermentation analysis while 2% cellobiose was used for plate screening. *S. cerevisiae* strains were cultured at 30°C with orbital shaking at 250 rpm for aerobic growth or 100 rpm for oxygen limited conditions. As needed, 200 µg/mL G418 (KSE Scientific, Durham, NC) supplemented YPD for plasmid selection. *Escherichia coli* DH5α (Cell Media Facility, University of Illinois at Urbana-Champaign, Urbana, IL) was used for recombinant DNA manipulations. *E. coli* strains were cultured in Luria broth (LB) (Fischer Scientific, Pittsburgh, PA) at 37 °C and 250 rpm, supplemented with 50 µg/mL ampicillin. Yeast and bacterial strains were stored in 15% glycerol at -80 °C. All chemicals were purchased from Sigma Aldrich or Fisher Scientific.
The filamentous fungi were obtained from ARS Culture Collection (NRRL) in Peoria, IL. Each strain was cultivated in YP media (1% yeast extract, 2% peptone) with 2% glucose (YPD) for growth with 250 rpm orbital shaking. To amplify the genes of interest, the strains were grown on YP with 2% cellobiose agarose plates or in liquid media and the tissue was extracted for mRNA preparation.

2.4.2 Gene Cloning

The total RNA from the sample, which was presumably expressing the gene of interest, was isolated from fresh samples using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA) following manufacturer’s instructions. Isolated RNA samples were cleaned using TURBO DNA-free Kit (AMBION INC, Austin, TX) based on the Rigorous gDNA Decontamination step. Finally, the mRNA was then reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche, Mannheim, Germany) with the oligo-dT primer following the manufacturer’s instructions.

2.4.3 Plasmid and Strain Construction

Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). All cloning work was performed through the DNA assembler method, mediated by yeast homologous recombination (39). Yeast plasmids were isolated using Zymoprep Yeast Plasmid miniprep II kit (Zymo Research, Irvine, CA), then transformed into *E.coli* for isolation of high purity DNA. *E.coli* plasmids were isolated using Qiagen Spin Plasmid Mini-Prep Kit (Qiagen, Valencia, CA). PCR fragments were purified by QIAQuick Gel Purification Kit (Qiagen, Valencia, CA).
The β-glucosidase genes were expressed using a PYK promoter and an ADH terminator. The cellobiose transporters were expressed using a TEF promoter and a PGK terminator as previously constructed (12). To facilitate the creation of a library of cellobiose utilizing pathways, helper plasmids were constructed. A helper plasmid containing the PYK promoter and the ADH terminator followed by the TEF promoter with a unique restriction digest site BamHI between the PYK promoter and ADH terminator was used for cloning the β-glucosidase genes. A helper plasmid containing an ADH terminator followed by a TEF promoter and the PCK terminator, with a unique restriction digest site NotI between the TEF promoter and the PCK terminator was used for cloning the celldextrin transporters. The helper plasmids were linearized using the respective restriction digest enzymes and assembled with the PCR-amplified genes.

2.4.4 Library Creation

Linear DNA fragments were amplified from the helper plasmid which contained the gene, and the promoter/terminator of the gene of interest, including a 500 bp homologous region to the neighboring DNA as shown in Figure 2.2. The library was created according to the previously described method (21,23). The total library size obtained was $10^4$. Ten colonies were randomly selected from the YPD+G418 plate and their plasmids were isolated and sequenced to confirm the diversity of the library (Figure 2.7).

2.4.5 Library Screening

The library was screened on YPC plates for large colony selection (21,23). After confirmation of the top 5 clones with the fastest growth rate, the plasmids were isolated and retransformed. This retransformation ensured that no adaptation occurred in the screening process and the increased
activity is derived from mutations on the plasmid. After retransformation, the cells were seeded overnight in YPD+G418, washed twice with sterile water, and then inoculated into 50 mL YPC in 250 mL flask at 100 rpm and 30 °C. The cultures were sampled and analyzed for growth and also for metabolite production as described previously (21,23).

2.4.6 Fermentation Analysis

Seed cultures were inoculated from freshly streaked frozen stocks. YPD+G418 seed cultures were used for fermentation tests. YPC seed cultures were used for final mutant pathway analysis. The seed cultures were grown at 30 °C and 250 rpm overnight, then washed with sterile water twice before inoculating 50 mL YPC in 250 mL un-baffled flasks to an initial OD of 0.2. The cultures were incubated at 30 °C and 100 rpm for oxygen limited growth. The cultures were sampled and analyzed for growth and metabolite production.

2.4.7 β-Glucosidase Enzyme Activity Assays

Classic strains harboring the β-glucosidase genes were grown to mid-exponential phase and washed three times with potassium phosphate buffer (pH 7). A final cell mass equivalent to an OD of 20 was harvested. Cell free extracts were prepared through YPER Extraction Reagent (Thermo Scientific, Rockport, IL). 125 µL of YPER was used to lyse the cells for 20 minutes at 25 °C, vigorously shaken at 700 rpm in a thermo-mixer (Eppendorf, Germany). After lysing, the cell membrane was pelleted for 10 minutes with 15,000 rpm at 4 °C. The protein concentration was determined via the BCA protein assay kit (Pierce, Rockford IL), following the standard manufacturer protocol.
The lysate was tested for $p$-NPG activity by quantifying the absorbance change of 1 mM $p$-NPG in 100 mM potassium phosphate buffer at pH 7. The colorimetric change was monitored at 405 nm in a 96-well Biotech Synergy 2 plate-reader (Winooski, VT) for 30 minutes at 30°C. The change in absorbance was converted to units (1 U = 1 µmol of $p$-NP/min) using Beer-Lambert law. The extinction coefficient was determined by a standard curve using $p$-NP at pH 7.
2.5 References


### 2.6 Tables

Table 2.1: List of fungal gene names and ID

<table>
<thead>
<tr>
<th>Abbreviated Gene Name</th>
<th>Gene Name</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cellodextrin Transporters</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AN2</td>
<td><em>Aspergillus nidulans</em></td>
<td>ANID_08347.1</td>
</tr>
<tr>
<td>CG1</td>
<td><em>Chaetomium globosum</em></td>
<td>P_001220469.1</td>
</tr>
<tr>
<td>CDT1</td>
<td><em>Neurospora crassa</em></td>
<td>XM_958708.1</td>
</tr>
<tr>
<td>CDT2</td>
<td><em>Neurospora crassa</em></td>
<td>XM_959259.1</td>
</tr>
<tr>
<td>CDT3</td>
<td><em>Neurospora crassa</em></td>
<td>XM_958780.1</td>
</tr>
<tr>
<td><strong>β-glucosidases</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ncrassa1</td>
<td><em>Neurospora crassa</em></td>
<td>XM_951011</td>
</tr>
<tr>
<td>Nhaem2</td>
<td><em>Nectria haematococca</em></td>
<td>XM_003051016</td>
</tr>
<tr>
<td>Tstipitus3</td>
<td><em>Talaromyces stipitatus</em></td>
<td>XM_002478144</td>
</tr>
<tr>
<td>Anidulands5</td>
<td><em>Aspergillus nidulans</em></td>
<td>XM_659164</td>
</tr>
<tr>
<td>Aoryzae6</td>
<td><em>Aspergillus oryzae</em></td>
<td>XM_001816779</td>
</tr>
<tr>
<td>Nhaem8</td>
<td><em>Nectria haematococca</em></td>
<td>XM_003051795</td>
</tr>
<tr>
<td>Tstipitis9</td>
<td><em>Talaromyces stipitatus</em></td>
<td>XM_002479153</td>
</tr>
<tr>
<td>Aflavus10</td>
<td><em>Aspergillus flavus</em></td>
<td>XM_002374789</td>
</tr>
<tr>
<td>Aniger11</td>
<td><em>Aspergillus niger</em></td>
<td>XM_001390218</td>
</tr>
<tr>
<td>Ncrassa18</td>
<td><em>Neurospora crassa</em></td>
<td>XM_951090.1</td>
</tr>
<tr>
<td>Acu19</td>
<td><em>Aspergillus aculeatus</em></td>
<td>JN121997.1</td>
</tr>
<tr>
<td>Aniger21</td>
<td><em>Aspergillus niger</em></td>
<td>DQ220304</td>
</tr>
<tr>
<td>Aniger22</td>
<td><em>Aspergillus niger</em></td>
<td>XM_001212225</td>
</tr>
<tr>
<td>Aterreus23</td>
<td><em>Aspergillus terreus</em></td>
<td>XM_001209807</td>
</tr>
<tr>
<td>Aterreus24</td>
<td><em>Aspergillus terreus</em></td>
<td>XM_001212225</td>
</tr>
<tr>
<td>Aoryzae32</td>
<td><em>Aspergillus oryzae</em></td>
<td>XM_001816940</td>
</tr>
<tr>
<td>Aoryzae33</td>
<td><em>Aspergillus oryzae</em></td>
<td>XM_001819003</td>
</tr>
<tr>
<td>Tstipitus37</td>
<td><em>Talaromyces stipitatus</em></td>
<td>XM_002479606</td>
</tr>
<tr>
<td>Tstipitus38</td>
<td><em>Talaromyces stipitatus</em></td>
<td>XM_002480435</td>
</tr>
<tr>
<td>Aniger50</td>
<td><em>Aspergillus niger</em></td>
<td>EU233788</td>
</tr>
<tr>
<td>Panserina51</td>
<td><em>Podospora anserina</em></td>
<td>XM_001907664</td>
</tr>
<tr>
<td>Zrouxii52</td>
<td><em>Zygosaccharomyces rouxii</em></td>
<td>XP_002497082.1</td>
</tr>
<tr>
<td>Mgrisea63</td>
<td><em>Magnaporthe grisea</em></td>
<td>XM_364315.2</td>
</tr>
<tr>
<td>Ncrassa67</td>
<td><em>Neurospora crassa</em></td>
<td>XM_956275.2</td>
</tr>
<tr>
<td>Aterreus68</td>
<td><em>Aspergillus terreus</em></td>
<td>XM_001210773.1</td>
</tr>
<tr>
<td>Panserina 70</td>
<td><em>Podospora anserina</em></td>
<td>XM_001908385.1</td>
</tr>
<tr>
<td>Aterreus72</td>
<td><em>Aspergillus terreus</em></td>
<td>XM_001210773.1</td>
</tr>
</tbody>
</table>
2.7 Figures

Figure 2.1: Illustration of the potential cellobiose utilization pathways which have been developed. (a) cell-surface display of extracellular β-glucosidases, where the cellobiose is hydrolyzed extracellularly by β-glucosidases docked to the surface of the yeast and then the cell transports and metabolizes the glucose molecules. (b) The hydrolytic pathway involves heterologous expression of a cellodextrin transporter and an intracellular β-glucosidase, the cellobiose is transported into the cell, and the intracellular cellobiose is hydrolyzed by the β-glucosidase into two glucose molecules, which are then metabolized by the cell. (c) The phosphorolytic pathway relies on heterologous expression of a cellodextrin transporter to bring cellobiose into the cell where an intracellular phosphorylase cleaves the disaccharide with an inorganic phosphate, producing a glucose molecule and α-glucose-1-phosphate.
**Figure 2.2** Preparation of DNA fragments for large library generation in a three gene pathway.

Each unique design represents a unique gene. The DNA is assembled into helper plasmids, as shown on the left. Each helper plasmid has homology regions, upstream and downstream of the DNA fragment of interest. With this homology region, the pathway can assemble properly into many different combinations. This strategy requires 400 bp length of homology, which can contribute to the efficiency of correct assembly.
**Figure 2.3:** Workflow for the homologous enzyme library. (1) Amplify homologous genes from cDNA of filamentous fungi. (2) Clone the genes into helper plasmids, designed with promoter and terminator for expression, and homologous DNA regions consisting of upstream and down-stream DNA sequences used for recombination. (3+4) Using DNA Assembler, the library can be generated through *in vivo* homologous recombination. (5) Plate library on cellobiose plates and screen for largest colonies. (6+7) Quantitative analysis through flask fermentation.
Figure 2.4: Characterizing all the different β-glucosidases enzymes for library through *in vitro* whole cell lysate assays. (a) β-glucosidases activity of the GH1 family with *p*-NPG. (b) β-glucosidases activity of the GH3 family with *p*-NPG. The GH1 family had significantly higher overall activity towards *p*-NPG.
Figure 2.5: Growth of strains harboring high (pRS 424) or low (pRS414) copy plasmids of the cellodextrin transporter in YPC media. High and low copy number plasmids translate into either high or low expression levels of the gene. This growth curve indicates that strains harboring the same transporter can only grow on cellobiose when the transporter is highly expressed. Strains harboring the single copy plasmids did not grow on cellobiose as the sole carbon source. A single copy plasmid is required for controlled library screening, therefore these transporters cannot be used in the library.
Figure 2.6: BP researchers discovered two cellodextrin transporters AN2 and CG1 from *Aspergillus nidulans* and *Chaetomium globosum*, respectively. CBT1 and CBT are CDT1 and CDT2 respectively. (a) Growth rate of strains harboring transporters on cellobiose as sole carbon source. (b) Cellobiose utilization of strains harboring transporters. (c) Ethanol production of strains harboring transporters. These transporters were used in the library.
Figure 2.7: Confirmation of the diversity of the library. Assembled pathways are selected based on antibiotic resistance (G418) and are tested for the diversity based on growth on cellobiose. Ten randomly selected clones have a wide range of growth on cellobiose, demonstrating the library has been assembled with a wide variety of potential enzymes with varying degrees of efficiency.
Figure 2.8: Characterization of the retransformed strains harboring top candidate pathways from the homologous enzyme library. (a) Cellobiose consumption. (b) Ethanol production. No significant increase in cellobiose consumption or ethanol production was identified. However, the CDT-1 was in all the top-performing pathways, but two β-glucosidases performed as well as GH1-1, Pathway 1 consisted of β-glucosidase from *Aspergillus niger*, Pathway 23 consisted of the β-glucosidase from *Podospora anserina*, both from family GH-1.
Chapter 3: Developing a Directed Evolution Strategy for the Optimization of Heterologous Metabolic Pathways by Simultaneously Engineering Multiple Proteins

3

3.1 Background

3.1.1 Directed Evolution

Directed evolution is an iterative multi-step optimization strategy and is established as a very powerful technique in synthetic biology for optimizing protein activity (1,2). This thesis chapter expands the application of directed evolution to include pathway-scale protein engineering. Chapter 2 of this thesis focused on optimizing the cellobiose utilization pathway by searching for enzymes which did not confer inefficient activities, such as transglycosylation. No homologous enzymes resulted in a more improved pathway than the original wild-type pathway. Therefore we chose to engineer the original wild-type enzymes for an improved activity. The novelty in this protein engineering project was the expansion to pathway-scale, by engineering multiple proteins simultaneously. Application of directed evolution on the pathway scale is a powerful strategy because no a priori information about pathway bottlenecks or structural information about the pathway enzymes is needed. Another advantage of a pathway-scale protein engineering stems from the major driving force of pathway library strategies - the comprehensive search for a maximum, global pathway optimum rather than a mediocre, local optimum.

---

3 Parts of this chapter were previously published in:


3.1.2 Directed Pathway Evolution Strategies to Improve Cellobiose Utilization

Yuan and Zhao applied directed evolution to mutant promoter pathway libraries of the hydrolytic cellobiose utilization pathway (3). The promoters \textit{PDC1}, \textit{ENO2}, and \textit{TEF1} were mutagenized through nucleotide analog-based error-prone PCR, which induced a very high mutation rate. An average mutation rate of 12-16 nucleotide substitutions per kilobase for the \textit{PDC1} and \textit{ENO2} promoter was obtained. The pathway genes were not mutagenized and these non-mutated DNA fragments were co-transformed with the error-prone promoter library, assembled into a diverse pathway library by the homologous recombination mechanism of \textit{S. cerevisiae}. The total size of the library was $10^4$. The pathway phenotype improvement was assessed by fast sugar utilization, visualized by large colonies on agar plates.

The first round of directed evolution identified a strain with a 5.7-fold increase in cellobiose consumption rate and a 5.5-fold increase in ethanol productivity. Further rounds of evolution yielded incremental subsequent increases (Figure 3.1). After characterizing the mutant promoters, it was found that the expression level ratios had significantly changed. While the parent BGL:CDT relative expression ratio was 13.8:1, the first round of mutagenesis altered the ratio to 2.5:1. This significant increase in relative cellobextrin transporter expression suggested that balancing of the protein expression levels was a bottleneck.

Directed evolution was very successful in the promoter engineering optimization; however, as previously mentioned in Chapter 2, engineering the promoters will not optimize the intrinsic inefficiencies of the enzymes themselves. Though balancing and optimizing the protein expression can greatly improve the pathway, the detrimental effects of side reactions, cofactor
specificities, or substrate specificities cannot be improved through this strategy. Therefore, it is important to consider optimizing the pathway based on the enzymes within the pathway. Chapter 2 of this thesis investigated different protein combinations of the cellobiose utilization pathway in order to find enzymes with reduced unwanted activities. The original cellobiose utilization pathway (CDT1 and GH1-1) was found to confer one of the highest productivities from the library of homologous proteins. However, the proteins in the pathway were still inefficient. Since nature did not provide a more efficient set of enzymes, the next strategy was to engineer the enzymes in the laboratory for an improved activity.

3.1.3 Directed Pathway Evolution Strategies to Improve Protein Activity

Current methods for protein engineering often involve isolating the protein from the context of the pathway and engineering it independently. In this study, we have investigated a complementary strategy for protein engineering wherein we simultaneously modified multiple proteins in the context of the pathway (Figure 3.2). This strategy combines the strength of pathway-scale libraries, which can comprehensively search for a global maximum optimum, with protein engineering. This combination could allow the identification of concurrent mutations in multiple proteins, which could synergistically increase the desired activity. This approach can be widely applicable to any pathway where a high-throughput screening/selection method for the desired phenotype is available. The implementation of this directed evolution strategy will further enhance researcher’s abilities to optimize pathways, introducing a new step in metabolic engineering involving pathway-driven approaches for protein engineering. In this chapter, we describe the first example of simultaneously engineering multiple proteins, β-glucosidase (gh1-1) [GenBank Accession number XM_951090] and cellodextrin transporter (cdt-1) [GenBank Accession number XM_958708] in Saccharomyces cerevisiae for biofuel
production. Through this directed pathway evolution strategy, key mutations were found in both of the proteins, which synergistically improved the overall cellobiose utilization by 49% and ethanol productivity by 64%. These mutations were directly linked to improved activity or altered substrate specificity of the proteins. Though improving the cellobiose utilization for biofuel production is of interest, these results also augment recent research in β-glucosidase substrate specificity (4-7) and continued research in transporter engineering (8-12).

3.2 Results

3.2.1 Library Construction and Screening

We created the library of gh1-1 and cdt-1 genes by independently mutagenizing each gene through error-prone PCR. The genes were then co-assembled through in vivo homologous recombination in S. cerevisiae (13), into the single copy plasmid pRS-KanMX, under control of the PYK1 and TEF1 constitutive promoters, respectively. The intended mutation rate was one amino acid substitution per protein on average. Considering the combination of both genes, the theoretical total number of mutant pathways was 9.9×10^6 (14). Due to limitations of the transformation efficiency into the industrial yeast host, the actual screened library size was ~10^4 mutant pathways. The specific growth rate of the strain on cellobiose was used as an assessment of pathway improvement, which allowed for screening to be based on colony size from agar plates (15,16). Large colonies from the library were visibly distinct from the colonies harboring the wild-type pathway (Figure 3.3). The large colonies were picked for quantitative analysis and the strains exhibiting increased specific growth rates on cellobiose were chosen (Figure 3.4). These plasmids were extracted and retransformed into fresh yeast cells to confirm that the increased specific growth rate resulted from the mutated pathway, not from genome
We successfully screened through two rounds of directed evolution, resulting in mutant pathways, R1 and R2, which conferred a higher specific growth rate on cellobiose when compared to the wild-type pathway (Figure 3.5). A third round of mutagenesis was performed but screening of the library did not result in the identification of a strain with an improved specific growth rate.

3.2.2 Characterization of the Mutant Cellulose Utilizing Pathways

The first mutant pathway from round one of the directed evolution, identified as R1, involved a L173H mutation in the β-glucosidase and a D433G mutation in the cellodextrin transporter. The second round of mutagenesis resulted in pathway R2, with additional mutations: H23L in the β-glucosidase and C82S in the cellodextrin transporter. R1 and R2 pathways conferred an increased specific growth rate on cellobiose, improved cellobiose utilization, and higher ethanol productivity compared to the strains harboring the wild-type cellobiose pathway in oxygen-limited conditions (Figure 3.5; Table 3.1). Oxygen-limited conditions were conducted at 100 rpm orbital shaking in un-baffled flasks, which reduced aeration to the culture (17). There was a 47% increase in specific growth rate in the final mutant strain (0.102 ± 0.002 h⁻¹) compared to the wild-type strain (0.0694 ± 0.003 h⁻¹). For the metabolite analysis, end-point values at the time required to consume more than 95% of total sugar was used to determine rate, yield, and productivity. The strains harboring the engineered pathways presented a 49% increase in cellobiose consumption and a 64% increase in ethanol productivity (Table 3.1). Extracellular glucose production (Figure 3.6), an undesired byproduct, was produced only when the cellobiose had been almost completely consumed and then was quickly consumed itself. The final strain
containing mutant pathway R2 consumed 80 g/L cellobiose in a little less than 30 hours with an ethanol productivity of 1.00 ± 0.03 g/(L·h).

To exclude the possibility that the improved phenotype was a result of improved gene expression, we performed quantitative PCR analysis of the mutant and wild-type strains. We also observed the specific protein abundance by monitoring the fluorescence per cell with green fluorescence protein (GFP)-tagging of the proteins and through Western Blot analysis. There was no statistically significant difference in mRNA levels or protein abundance between the wild-type and mutant proteins, indicating comparable gene expression levels throughout the strains (Figure 3.7 and 3.8).

To determine which mutations conferred the increased specific growth rate and ethanol productivity, the individual and combined mutants from each round were analyzed (Figure 3.9). The most significant mutants for the phenotype improvements of the strains were shown to be the L173H mutation in the β-glucosidase and the C82S mutation in the cellodextrin transporter. Specifically, the L173H mutation was attributed to increased specific growth rate, while the C82S mutation increased both the growth rate and ethanol productivity.

3.2.3 Activity Assays

3.2.3.1 β-Glucosidase Assay

The enzymes were assayed in vitro to better identify how the mutations affected the activity of the individual enzymes. We assayed the activity towards cellobiose by enzymatically quantifying the glucose produced by a timed reaction of β-glucosidase with cellobiose. The L173H mutant
increased the specific activity towards cellobiose by nearly 90% compared to the wild-type. The single H23L mutant had only slightly higher specific activity than the wild-type enzyme. The β-glucosidase combined mutant L173H and H23L increased cellobiose hydrolytic activity by over 150% compared to the wild-type enzyme, illustrating additive effects from the combined mutations (Table 3.2). Further tests involved p-nitrophenyl β-D-glucopyranoside (p-NPG), a synthetic substrate commonly used for determining β-glucosidase activity (4,6). Compared to the wild-type, the L173H mutant decreased hydrolytic activity towards p-NPG by 34%. The single mutant H23L protein had activity slightly higher than wild-type. The combination of the L173H and H23L mutations increased the activity to be slightly higher than the single L173H mutation.

3.2.3.2 Cellodextrin Assay

The specific activity of the cellodextrin transporter was measured through a modified oil-stop assay based on the rate of radiolabeled sugar uptake (18,19). The single mutation D433G conferred an 18% increase in cellodextrin transporter activity. When compared to wild-type, the single mutant protein containing C82S increased the transporter activity by 48%. The combination of D433G and C82S mutations increased the protein activity over wild-type by 60% (Table 3.3), demonstrating a synergistic affect to improve the overall pathway productivity.

3.2.4 Structure/Sequence-Function Relationships

A homology model was constructed for the β-glucosidase and the mutations were investigated for possible structure-function relationships. The H23L mutation is on the periphery of the enzyme, far from the active site (Figure 3.10). The homology model predicted that the L173H mutation was located within the active site, in a domain which has previously been identified as
the substrate entrance region (5). The amino acid L173 did not have any predicted interactions with cellobiose (Figure 3.11a). When the residue was mutated to H173, the model predicted a direct hydrogen bond to the hydroxyl group of the C1 atom of the cellobiose (Figure 3.11b).

Since no crystal structure is available for the cellodextrin transporter, to visualize and hypothesize how these mutations affected the protein, a sequence-based analysis was performed with the aid of HMMTOP software (20,21). This software is used to predict transmembrane helix domains, helix tail domains, and extra/intracellular loop structures (Figure 3.12). Based on the predicted structure, the cellodextrin transporter is comprised of 12 transmembrane helices with one large inside loop and one large outside loop. The D433G mutation is predicted to be located on the large outside loop and the C82S mutation is projected to be in the first transmembrane helix.

3.3 Discussion

Efficient production of biofuels and specialty chemicals hinges on optimization of the metabolic pathways associated with the desired product. In this endeavor, a variety of successful pathway optimization methods, mainly based on transcriptional engineering, have been developed (16,22-28). Though successful, these methods are unable to overcome inefficiencies based on innate enzyme characteristics such as $V_{\text{max}}$ and $K_{\text{m}}$, therefore protein engineering strategies must also be applied (29). To improve these properties, often the enzymes are removed from the pathway context and engineered in vitro or in vivo for the highest activity. Engineering each enzyme independently can be arduous and there is no guarantee the engineered enzyme will reduce the bottleneck when it is reinstated into the pathway. We are introducing a new approach to combine
pathway and protein engineering, involving the evolution of all proteins within the pathway simultaneously. Through this strategy, no *a priori* information about pathway bottlenecks or enzymes is required for the directed evolution. Mutant proteins of the pathway are not chosen for high activity, but instead interrogated as a whole for a balanced, increased activity and flux which is demonstrated by an improved pathway phenotype. Perhaps the most advantageous reason for a pathway-scale protein engineering strategy is the ability to comprehensively search all the possible global pathway optimums to identify the high optimum, which could be a combination of unexpected synergistic effects (26,30).

Recent advances in large-scale library creation with high efficiency have made the multiplex protein engineering library a possibility (15,16,31-33). Our study illustrates an *in vivo* library creation for the simultaneous engineering of multiple proteins in a pathway through directed evolution in *S. cerevisiae*. To the best of our knowledge, there has been only one example of mutagenesis of multiple proteins through directed evolution, which was illustrated in *Escherichia coli* for arsenate resistance (34). This strategy has not been widely implemented and has not been applied to other systems. Our work optimizes a sugar utilization pathway for increased metabolite production with a simple one-step assembly method, allowing for quick and efficient rounds of evolution. This represents the first example of pathway optimization through multiple protein engineering in *S. cerevisiae*.

The final optimized pathway (R2) increased cellobiose consumption by 49% and ethanol productivity by 64%. Characterization of the strains expressing the mutations which conferred these improved phenotypes illustrated that, the highest specific growth rates on cellobiose were
achieved only when the β-glucosidase L173H mutation is present. In the cellodextrin transporter, the most significant mutation was the C82S mutation. It was shown that the optimized pathway involved an overall increase of protein activity: the β-glucosidase had an increased specificity towards cellobiose and the cellodextrin transporter exhibited a higher overall activity.

3.3.1 β-Glucosidase Characterization

The β-glucosidase from glycoside hydrolase family 1 (GH1) contains the standard (α/β)-barrel structure of the enzyme family. The active site is located in a deep cleft formed by the connection loops at the C-terminus of the β-sheets of the TIM barrel with negatively charged residues surrounding the bottom of the active site (6). The conserved catalytic residues of the GH1 family involve the catalytic acid/base Glu 166 and the nucleophile Glu 377, which cleave the glycosidic bond via the common double displacement mechanism (4). This mechanism acts on β-1,4-linked glucose derivatives; the enzyme is promiscuous, being active towards a range of substrates with varying affinities (6). In this study, the wild-type GH1-1 β-glucosidase enzyme exhibited an innate substrate preference for p-NPG over cellobiose. The mutation L173H shifted the substrate specificity towards cellobiose. Homology modeling suggested the H173 residue could have a direct hydrogen bond with the hydroxyl group of the C1 atom of the cellobiose (Figure 3.11), which was not predicted with the wild-type β-glucosidase. It is possible that this new hydrogen bond could increase the affinity of the enzyme for cellobiose by firmly keeping the substrate in place during hydrolysis, a hypothesis previously established by studies in substrate recognition for the β-glucosidase family (7). There has been significant interest in substrate specificity of the GH1 family, which is suggested to occur via an extensive network of polar interactions with well-ordered water molecules and enzyme-substrate bonding (4-7).
specific study of interest involved, a β-glucosidase from *Trichoderma reesei* [GenBank Accession number AB003110.1] which was recently investigated to identify critical residues within the substrate entrance region (5). Through rational design, substitutions at residue 172 were shown to improve substrate recognition, thermostability, and enzyme activity. This β-glucosidase has a 73% sequence identity to the GH1-1 enzyme studied here. Sequence alignment and structural modeling predicted this residue to be in the same position as the L173 residue from this work, corroborating the importance of this residue position. The H23L mutation is hypothesized to be an overall activity enhancer, with only slightly improved enzyme activity towards both substrates *p*-NPG and cellobiose. H23 is located on the periphery of the enzyme, distant from the active site (Figure 3.10).

### 3.3.2 Cellodextrin Characterization

Engineering sugar transporters has recently been shown to greatly enhance pathway productivity (8,12), a result also illustrated here. In our directed evolution effort, the combined D433G and C82S mutations of the cellodextrin transporter increased the specific activity by 60% compared to the wild-type. The specific activity is directly proportional to the $V_{\text{max}}$ of the transporters, which is the protein property that has been most often enhanced by sugar transporter engineering (8,12). It is probable that this property is the most likely to be improved due to the screening conditions, which were performed at saturating conditions. The cellodextrin transporter model predicted that the D433G mutation is located on a large outside loop. It is possible that this loop could have been causing a steric hindrance for the mass transfer of cellodextrin to the transporter, which could have been alleviated by the altered loop structure associated with this mutation. The C82S mutation, which conferred the highest increase in activity, is predicted to be located in the
first transmembrane helix (Figure 3.12). The location in the helix and the increased activity suggest that this mutation could be directly associated with cellobiose interactions or the protein complex stability. Young et al. investigated directed evolution of xylose transporters and identified mutations which tended to cluster around the first transmembrane helix (8), a similar finding to this study.

3.4 Conclusion

Cellobiose utilization has recently become a significant consideration in economical biofuel production. Efficient utilization could decrease the expensive in vitro enzyme load in biomass pretreatment. We successfully optimized the hydrolytic cellobiose utilization pathway through a new directed evolution strategy: simultaneously engineering multiple proteins within the context of the pathway. The optimized pathway involved a β-glucosidase mutant with an increased specificity towards cellobiose and a cellodextrin transporter mutant with an increased overall activity. By applying directed evolution to the entire pathway, engineered proteins can be found which synergistically improve the phenotype either through a balanced flux of the pathway or through overall protein activity improvement. Simultaneous multi-protein engineering is expected to expand researchers’ abilities to optimize biosynthetic metabolic pathways.

3.5 Materials and Methods

3.5.1 Strains, Media, and Culture Conditions

The industrial Saccharomyces cerevisiae strain Still Spirits (Classic) Turbo Distiller’s Yeast was purchased from Homebrew Heaven (Everett, WA). Yeast strains were cultivated in YP media (1% yeast extract, 2% peptone) with 2% glucose (YPD) or 2-8% cellobiose (YPC). YPC with
8% cellobiose was used in fermentation analysis while 2% cellobiose was used for plate screening. *S. cerevisiae* strains were cultured at 30 °C with orbital shaking at 250 rpm for aerobic growth or 100 rpm for oxygen limited conditions. As needed, 200 µg/mL G418 (KSE Scientific, Durham, NC) supplemented YPD for plasmid selection. *Escherichia coli* DH5α (Cell Media Facility, University of Illinois at Urbana-Champaign, Urbana, IL) was used for recombinant DNA manipulations. *E. coli* strains were cultured in Luria broth (LB) (Fischer Scientific, Pittsburgh, PA) at 37 °C and 250 rpm, supplemented with 50 µg/mL ampicillin. Yeast and bacterial strains were stored in 15% glycerol at -80 °C. All chemicals were purchased from Sigma Aldrich or Fisher Scientific.

### 3.5.2 Plasmid and Strain Construction

Restriction enzymes were purchased from New England Biolabs (Ipswich, MA). All cloning work was performed through yeast homologous recombination mediated by the DNA Assembler method (13). Yeast plasmids were isolated using Zymoprep Yeast Plasmid miniprep II kit (Zymo Research, Irvine, CA), then transformed into *E. coli* for isolation of high purity DNA. *E. coli* plasmids were isolated using Qiagen Spin Plasmid Mini-Prep Kit (Qiagen, Valencia, CA). PCR fragments were purified by QIAQuick Gel Purification Kit (Qiagen, Valencia, CA).

The β-glucosidase gene gh1-1 (GenBank Accession number XM_951090) from *N. crassa* was expressed using a PYK promoter and an ADH terminator. The cellobiose transporter gene cdt-1 (GenBank Accession number XM_958708) from *N. crassa* was expressed using a TEF promoter and a PGK terminator as previously constructed (19). To transfer the pathway to the pRS-KanMX plasmid, primers kanMX-PYKp-F and PGKt-kanMX-R (Table 3.4) were used to
amplify the full cellobiose utilizing pathway. To facilitate the creation of a library of cellobiose utilizing pathways, a helper plasmid was constructed. The helper plasmid contained the PYK promoter and the PGK terminator, separated by a unique restriction enzyme recognition site BamHI for plasmid linearization. The helper plasmid will later be linearized and used as a backbone for the library creation. The PYK promoter was amplified using primers kanMX-PYKp-F and PYKp-BamHI-PGKt-R (Table 3.4). The PGK terminator was amplified using primers PYKp-BamHI-PGKt-F and PGKt-kanMX-R (Table 3.4).

3.5.3 Library Creation

The ADH terminator and TEF promoter were not subjected to random mutagenesis, thus this cassette was amplified using primers middle-BGL-ADHt-F and middle-TEFp-CDT-R (Table 3.4). This fragment was transformed with the error-prone PCR fragments and linear helper plasmid. The gh1-1 and cdt-1 genes were subjected to error prone PCR and error rates of 2-3 bp mutations per 1kb were used for library creation. The library was created according to the previously described method (15,16). The total library size obtained was $10^4$. Ten colonies were randomly selected from the YPD+G418 plate and their plasmids were isolated and sequenced to confirm the diversity of the library.

3.5.4 Library Screening

The library was screened on YPC plates for large colony selection (15,16). After confirmation of the top 5 clones with the fastest growth rate, the plasmids were isolated and retransformed. This retransformation ensured that no adaptation occurred in the screening process and the increased activity is from mutations on the plasmid. After retransformation, the cells were seeded
overnight in YPD+G418, washed twice with sterile water, and then inoculated into 50 mL YPC in 250 mL flask at 100 rpm and 30 °C. The cultures were sampled and analyzed for growth and also for metabolite production as described previously (15,16). After confirmation of improved mutant pathways, the plasmids were isolated and sequenced to identify the mutations within the genes in the pathway. After confirmation of the mutations, that pathway was used as a template for a second round of error-prone PCR. The library creation and screening process was repeated.

3.5.5 Fermentation Analysis

Seed cultures were inoculated from freshly streaked frozen stocks. YPD+G418 seed cultures were used for the single/double mutant fermentation tests. YPC seed cultures were used for final mutant pathway analysis. The seed cultures were grown at 30 °C and 250 rpm overnight, then washed with sterile water twice before inoculating 50 mL YPC in 250 mL un-baffled flasks to an initial OD of 0.2. The cultures were incubated at 30 °C and 100 rpm for oxygen limited growth. The cultures were sampled and analyzed for growth and metabolite production.

3.5.6 Construction of Single Mutants

The single mutant genes were constructed through site-directed mutagenesis and the mega primer PCR method with primers listed in Table 3.4. After PCR amplification, the genes were transformed into the Classic strain along with the ADHt/TEFp cassette into a linearized pRS-KanMX helper plasmid. The final plasmid was purified and sequenced for confirmation.
3.5.7 \( \beta \)-Glucosidase Enzyme Activity Assays

Classic strains harboring an empty vector, the wild-type, and mutant pathways were grown to mid-exponential phase and washed three times with potassium phosphate buffer (pH 7). A final cell mass equivalent to an OD of 20 was harvested. Cell free extracts were prepared through YPER Extraction Reagent (Thermo Scientific, Rockport, IL). 125 µL of YPER was used to lyse the cells for 20 minutes at 25 °C, vigorously shaken at 700 rpm in a thermomixer (Eppendorf, Germany). After lysing, the cell membrane was pelleted for 10 minutes with 15,000 rpm at 4 °C. The protein concentration was determined via the BCA protein assay kit (Pierce, Rockford IL), following the standard manufacturer protocol.

The lysate was tested for \( p \)-NPG activity by 1 mM \( p \)-NPG in 100 mM potassium phosphate buffer at pH 7. The colorimetric change was monitored at 405 nm in a 96-well Biotech Synergy 2 plate-reader (Winooski, VT) for 30 minutes at 30°C. The change in absorbance was converted to units (1 U = 1 µmol of \( p \)-NP/min) using Beer-Lambert law. The extinction coefficient was determined by a standard curve using \( p \)-NP at pH 7. For cellobiose-based enzymatic assay, the linear range of the \( \beta \)-glucosidase was determined. The reactions were carried out at 60 mM cellobiose in 100 mM potassium phosphate buffer at pH 7. After addition of lysate, the reaction was allowed to react for 15, 30, 45, and 60 minutes at 30 °C. The reaction was stopped by boiling at 100 °C for 10 minutes. The samples were then centrifuged at 15,000 rpm for 5 minutes before being stored on ice. The amount of glucose which had been produced in the allotted time frame was then measured using the D-glucose kit (R-Biopharm, Germany). Standard manufacturer’s instructions were followed. One unit (U) is defined as micromoles of glucose.
produced per minute. The specific activity was determined by normalizing the rate to the amount of protein in the assay.

To determine the β-glucosidase protein abundance in each strain, the wild-type and mutant proteins were fused to GFP. There was no statistically significant difference in β-glucosidase abundance of each strain (Figure 3.7 and 3.8). The total protein concentration, as determined through BCA assay, was comparable in each lysate. Therefore, the ratio of β-glucosidase to total protein was assumed constant, and hence the activity was normalized to total protein concentration. One unit (U) is defined as one micromole of glucose produced per minute.

3.5.8 CDT Activity Assay

The cellodextrin transporter was assayed using the oil-stop protocol previously reported (18,35) with few modifications. Cultures of the wild-type and mutant pathways were grown to an OD of 15-20, washed three times with ice-cold assay buffer (30 mM MES-NaOH + 50 mM ethanol), and then normalized to an OD of 20. 50 μL of cells were added to 50 μL of [3H]-cellobiose at 30 °C and layered over 100 μL of silicone oil (Sigma 85419), incubated for 10, 20, 40, and 80 seconds. The cells were then centrifuged through the oil at 15,000 rpm for 30 seconds. After ethanol/dry-ice bath, the cell pellets were solubilized in NaOH overnight. The amount of [3H]-cellobiose present in the cells was then quantified via a liquid scintillation counter. The amount of labeled cellobiose that was taken up by the cell was plotted against the time of reaction. One unit (U) is defined as the micromoles of cellobiose taken up by the cell per min. The dry cell weight of the cells was determined and the rate of the cellobiose uptake was normalized by the gram cell dry weight (gdcw). The rate of the cellobiose uptake was normalized by the transporter
abundance determined through GFP fluorescence measurements and Western Blot analysis (Figure 3.7 and 3.8) and the gram cell dry weight (gcdw).

3.5.9 qPCR Analysis

Samples of cultures grown in the same conditions as protein activity assays, beginning-mid exponential phase, and were used to compare relative mRNA expression levels of the mutant pathways via quantitative PCR (qPCR). The total RNA was isolated from fresh samples using the yeast FastRNA SPIN kit (MP Biomedical Solon, OH) following manufacturer’s instructions. Isolated RNA samples were cleaned using TURBO DNA-free Kit (AMBION INC, Austin, TX) and then reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche, Mannheim, Germany) with the oligo-dT primer following the manufacturer’s instructions. The qPCR was performed with LightCycler 480 SYBR Green Master reagents (Roche) using the Roche Light Cycler® 480 System (Roche, Indianapolis, IN). The relative abundance of the mRNA levels for the target genes was normalized to the asparagine-linked glycosylation 9 gene (alg9) expression to determine expression levels.

3.5.10 GFP Fusion

The β-glucosidase and the cellodextrin transporter genes were C-terminally tagged with GFP via the (Gly-Gly-Gly-Gly-Ser)2 linker using primers described in Table 3.4 and DNA Assembler. Strains harboring the GFP-tagged proteins were grown in the same conditions as the protein activity assays: inoculated to an initial OD of 0.2 and grown at 30°C at 100 rpm orbital shaking until the cells reached OD 15-20. The cells were then diluted in PBS to an OD of about 0.1-0.2 and were analyzed on a flow cytometer (Biosciences LSR II Flow Cytometry Analyzer, BD
Biosciences, San Jose, CA) to quantify GFP fluorescence of each mutant protein in the strain, with excitation wavelengths at 488 nm and emission band pass 550/30.

3.5.11 Western Blot Analysis

The β-glucosidase and the cellodextrin transporter genes were C-terminally tagged with a 3x-flag tag and a c-myc antibody tags respectively, using the primers described in Table 3.4 and DNA Assembler. Strains harboring the tagged proteins were grown in the same conditions as the protein activity assays: inoculated to an initial OD of 0.2 and grown at 30 °C at 100 rpm orbital shaking until the cells reached OD 15-20. Cells were lysed following standard manufacturer’s protocol of the YPER Extraction Reagent (Thermo Scientific, Rockford, IL). The protein concentration was determined via the BCA protein assay kit (Pierce, Rockford IL), following the standard manufacturer protocol. Each sample was normalized by total protein concentration and applied to a 12% agarose- SDS-page gel (Bio-Rad, Hercules, CA) at 100 V for 60 minutes. The gel was then transferred to a PVDF membrane for 60 minutes at 100 V at 4°C. The membrane was washed three times with PBS before being treated with the respective 3x Flag tag or c-myc antibodies from mouse (Thermo Fischer Scientific, Rockford, IL) for two hours are room temperature. The membrane was then washed and treated with goat anti-mouse antibody for two hours at room temperature. Finally, the membrane was washed and stained with Western Blue stabilized substrate (Promega, Madison Wisconsin) until bands were visualized.

3.5.12 Homology Modeling

A homology model of the β-glucosidase enzyme was constructed to identify the structure-function relationships of the mutations discovered. The gene encoding the β-glucosidase from
*Trichoderma reesei* (6) (PDB accession code 3AHY) afforded the highest homology to the *gh1-l* gene from *N. crassa* with 73% sequence identity and was used as a template for homology modeling. The structure model of the β-glucosidase from *N. crassa* was constructed using the modeling program Molecular Operating Environment (Chemical Computing Group, Montreal, Canada). After constructing the homology model, the substrate in the co-crystal structure of *Neotermes koshunensis* β-glucosidase (PDB accession code 3VIK) was docked into the model (5). The model was energy minimized before ligand interactions were investigated. To identify the effects of the mutations, the mutations were introduced to the model and energy minimized again before investigating the ligand interactions.
3.6 References


3.7 Tables

Table 3.1: Parameters for the improved pathways identified through directed evolution. Key parameters for an improved pathway include specific growth rate, cellobiose consumption, and ethanol productivity. Errors are derived from biological duplicates. Fermentations conditions in 8% cellobiose rich media and oxygen limited conditions.

<table>
<thead>
<tr>
<th></th>
<th>Wild-type</th>
<th>R1</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Specific Growth Rate</strong> (h⁻¹)</td>
<td>0.0694 ± 0.003</td>
<td>0.0919 ± 0.006</td>
<td>0.102 ± 0.002</td>
</tr>
<tr>
<td><strong>Cellobiose Consumption</strong> (g cellobiose/(L·h))</td>
<td>1.78 ± 0.06</td>
<td>2.33 ± 0.02</td>
<td>2.65 ± 0.02</td>
</tr>
<tr>
<td><strong>Ethanol Productivity</strong> (g ethanol/(L·h))</td>
<td>0.611 ± 0.02</td>
<td>0.815 ± 0.03</td>
<td>1.00 ± 0.03</td>
</tr>
<tr>
<td><strong>Yield</strong> (g ethanol / g cellobiose)</td>
<td>0.4073 ± 0.02</td>
<td>0.4236 ± 0.0008</td>
<td>0.4363 ± 0.004</td>
</tr>
</tbody>
</table>

*Values determined at time point when >95% cellobiose was consumed for the strain*
Table 3.2: Specific activity measured from the crude lysate of the engineered β-glucosidase. One unit is defined as one µmol/min and the control is the lysate from a strain with no heterologously expressed β-glucosidase. Each mutant was quantified with p-NPG and cellobiose as a substrate to demonstrate the increased substrate specificity towards cellobiose. Errors are derived from the standard deviation of biological triplicates.

<table>
<thead>
<tr>
<th></th>
<th>p-NPG (U/mg protein)</th>
<th>Cellobiose (U/mg protein)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.00729 ± 0.0001</td>
<td>0.0021 ± 0.002</td>
</tr>
<tr>
<td>Wild-type</td>
<td>0.193 ± 0.01</td>
<td>0.0668 ± 0.001</td>
</tr>
<tr>
<td>L173H</td>
<td>0.130 ± 0.01</td>
<td>0.124 ± 0.004</td>
</tr>
<tr>
<td>H23L</td>
<td>0.219 ± 0.004</td>
<td>0.0738 ± 0.006</td>
</tr>
<tr>
<td>L173H + H23L</td>
<td>0.156 ± 0.007</td>
<td>0.170 ± 0.01</td>
</tr>
</tbody>
</table>

Table 3.3: Specific activity measurements from the cellobextrin transporter as determined through radioactive-labeled cellobiose uptake rate assay. One unit is defined as one µmol/min. The control is the lysate from a strain with no heterologously expressed cellobextrin transporter. Errors represent the standard deviation of biological triplicates.

<table>
<thead>
<tr>
<th></th>
<th>Cellobiose Uptake Rate (U/gCDW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>0.006 ± 0.001</td>
</tr>
<tr>
<td>Wild-type</td>
<td>1.77 ± 0.07</td>
</tr>
<tr>
<td>D433G</td>
<td>2.09 ± 0.05</td>
</tr>
<tr>
<td>C82S</td>
<td>2.63 ± 0.2</td>
</tr>
<tr>
<td>D433G + C82S</td>
<td>2.82 ± 0.2</td>
</tr>
<tr>
<td>Table 3.4: Primers used in this study</td>
<td></td>
</tr>
<tr>
<td>--------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>PGKt-kanMX-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CCTCACAAGGAACAAACAGCTGGAGCTCCACCACAGTTGGGCAGGAAGAATACACTATAC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL EP-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTCTCTCTTCTTTATTTCTAAGACACCTTACAATCCTGCTACATGGAAGAATACACTATAC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL EP-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GGATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>CDT EP-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTCTTGTGCTATGAGAGGAGATGCAATCTAATTGATATTTCAATTCAATTAAATCAGATTAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>CDT EP-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GGATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>internal-BGL-ADHt-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTTTTGAACATCAGCAGCCCAAAGG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>internal-TEFp-CDT-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GAATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>B2-H23L-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GGGATTATCTTCCGAGCCCGGGCGCGGC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>B2-H23L-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GGATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>C2-C82S-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CCATTTTGAGGATGAGCCTGGACCCCAAGG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>C2-C82S-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GAATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>qPCR-BGL-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-AACGGCATCAGCAGCCCAAAGG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>qPCR-BGL-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GAATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>qPCR-CDT-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-AATCCCTCTGCCTCTTATTTG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>qPCR-CDT-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-TCTCTGATACGTTACCTACACAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>CDT sequencing-TEFp-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTTTGATACGCTTACCTACACAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>CDT sequencing-PGKt-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CAACGTTGAGTTAAAAAGGAGGCGCCCTATGCTGCTGCTCCTGCTGCTATGAAAGGAGTC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL sequencing-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTTTGATACGCTTACCTACACAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL sequencing-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GAATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>Prs Kamm-PYK-BGL-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CAACGTTGAGTTAAAAAGGAGGCGCCCTATGCTGCTGCTCCTGCTGCTATGAAAGGAGTC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL-Linker+GFP-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CAACGTTGAGTTAAAAAGGAGGCGCCCTATGCTGCTGCTCCTGCTGCTATGAAAGGAGTC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL-Linker+GFP-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CAACGTTGAGTTAAAAAGGAGGCGCCCTATGCTGCTGCTCCTGCTGCTATGAAAGGAGTC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>Linker+GFP-ADHt-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GGGATTATCTTCCGAGCCCGGGCGCGGC-3’</td>
<td></td>
</tr>
<tr>
<td><strong>ADHt-CDT Cas-KanMX-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTTTGATACGCTTACCTACACAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL CDT Cassette-Linker GFP-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTTTGATACGCTTACCTACACAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>BGL CDT Cassette-Linker GFP-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-CTTTTGATACGCTTACCTACACAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>Linker GFP-PGKt-R</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GAATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
<tr>
<td><strong>Linker GFP-PGKt-F</strong></td>
<td></td>
</tr>
<tr>
<td>5’-GAATGGGAAAAGAAGAAGAAAAATGATCTATCATTTGACATTGCCAGAG-3’</td>
<td></td>
</tr>
</tbody>
</table>
3.8 Figures

Figure 3.1: Fermentation profiles of the directed evolution rounds for the pathway libraries from Yuan et al. (3). (a,b) Cellobiose consumption and ethanol production of the cellobiose utilization pathway from the promoter-based directed evolution. The square represents the parent pathway with no mutations in the PDC1 and ENO2 promoter. The circles are the first round of error-prone PCR of both promoters. The triangles represent the second round of directed evolution, and the grey triangle represents the final round of mutagenesis.
Figure 3.2: A general scheme for the directed evolution of multiple proteins within a metabolic pathway. Diversity is introduced to the genes through error-prone PCR, DNA shuffling, or other mutagenesis techniques. The genes are then recombined through homologous recombination into the pathway. A non-mutagenized fragment for the interior - promoter and terminator is also amplified and assembled. Screening is accomplished via colony size on cellobiose agarose plates. Large colonies are associated with faster growth and selected for further analysis. To confirm phenotype is a result of the improved pathway and not from adaption, the pathways are retransformed. Improved mutants are selected for a second round of diversification and screening.
Figure 3.3: Selection plates of the error-prone library on cellobiose plates, depicting the large colonies. (a) wild-type pathway with all uniform colonies (b) error-prone PCR library.
Figure 3.4: (a) Specific growth rates of mutants in the top 10% fastest growers in the initial tube screening of the library created by the first round of error-prone PCR. (♦) is the wild-type pathway (♦) depicts the selected clones, numbered on the x-axis. (b) Cellobiose utilization from confirmation of clones with high specific growth rate fermentation carried out in oxygen limited conditions in un-baffled flasks. (■) (♦) denote the wild-type pathway cellobiose consumption, (■) (♦) represent two of the faster cellobiose utilizing pathways.
Figure 3.5: Fermentation profile of the improved cellobiose utilization pathways under oxygen limited conditions: (■) wild-type pathway, (●) R1 pathway, (▲) R2 pathway (a) Growth curve, (b) Cellobiose utilization, (c) Ethanol production. Error bars derived from biological triplicates.
Figure 3.6: Glucose accumulation in oxygen limited fermentation (■) wild-type pathway, (●) R1 pathway, (▲) R2 pathway.
Figure 3.7: Consistent protein expression levels between mutants. Quantitative PCR analysis of relative mRNA levels of the mutations (a) of the gh1-1 and (b) the cdt-1. Relative expression is normalized to the average value of the housekeeping gene. GFP-fusion to the wild-type and mutant proteins was also performed to compare specific protein concentration of the strains (c) β-glucosidase and (d) cellodextrin transporter. Arbitrary fluorescence units (AFU) are normalized to the wild-type average value. Error bars represent the standard deviation of at least three biological triplicates. A two-tailed student t-test confirmed that there were no significant differences in mRNA levels or protein concentration levels when tagged with GFP (p-values all greater than 0.05).
**Figure 3.8:** Consistent protein expression levels between mutants shown through Western blot analysis. Cell lysates have been normalized by total protein concentration. (a) Samples read from left to right, β-glucosidase wild-type, R1, and R2 and (d) Samples read from left to right, cellodextrin transporter wild-type, R1, and R2. No significant difference in protein expression can be discerned through Western Blot analysis.
Figure 3.9: (a) Growth rate and (b) ethanol productivity of the single and double mutants on cellobiose as a sole carbon source. Error bars are derived from biological triplicates. Each bar represents the mutation in the β-glucosidase coupled with that of the cellodextrin transporter. The black bar is the wild-type, the red bar is the L173H, the blue bar represents the double mutant L173H and H23L, and the final grey bar is the single H23L. The dotted line denotes statistically significant values over the wild-type (p < 0.05).
Figure 3.10: Homology model of the *N. crassa* β-glucosidase with the cellobiose substrate docked into the active site. The L23 mutation is notably far from the active site and the cellobiose substrate down in the center.
**Figure 3.11:** Homology model of the *N. crassa* β-glucosidase with the cellobiose substrate docked into the active site. (a) The critical residue in the wild-type enzyme L173 has no predicted interaction with the cellobiose molecule. (b) The mutated residue H173 is predicted to have direct hydrogen bonding with the hydroxyl group on the C1 atom of the cellobiose molecule.
Figure 3.12: Sequence-structure based mapping of the cellodextrin mutations. The HMMTOP software predicted sequence-structure of the transmembrane helix, inner, and outer loops of the cellodextrin transporter in the cell membrane. The top of the graphic depicts outside of the cell, while the bottom. The mutations were overlaid onto the structure prediction.
Chapter 4: Engineering a Novel Pathway for Advanced Biofuel Production

4.1 Background

To compete in a market dominated by fossil fuels, biofuels must be economically competitive and also offer the variety of molecules and compounds which are currently derived from fossil fuels. In the evolution of microbial-based biofuels, bioethanol has been the first major player, with a large market share coming from Brazil and the USA (1). However, as a fuel alternative ethanol is not ideal when compared to petrol, due to its low energy density, high recovery costs, corrosiveness, and high hygroscopy (2-4). Additionally, though bioethanol could be used in car engines, it is likely not to be successful in higher powered vehicles such as planes, buses, and trains (1). The molecule is also curbed in potential market share, due to the limited commodity and special chemicals which can be derived from it. The fossil fuels that biofuels are endeavoring to replace more closely resemble fatty acids.

Fatty acids are only one catalytic step away from petroleum-derived diesel molecules and have become a focus in advanced biofuel production. Derived from fatty acids, fatty acid ethyl esters (FAEEs), a type of biodiesel, are a non-toxic, biodegradable, drop-in fuel. Compared to fossil fuel derived diesel, the use of biodiesel in current engines can reduce emissions due to less carbon monoxide, sulfur, aromatic hydrocarbons, and soot particles produced from the combustion process (4,5). Considering market prospects, FAEEs and longer chain hydrocarbons are already one of the largest renewable fuels in Europe (6) and have the potential to meet commodity/specialty chemical demands (7).
4.1.1 Fatty Acid Synthesis and FAEE Production

Fatty acid biosynthesis is a complicated, iterative process which requires the coordination of many different reactions to elongate the carbon chain. There are various platforms and systems to produce fatty acids in different organisms; however the overall catalytic mechanisms are the same. The first committed step in fatty acid biosynthesis is the carboxylation of acetyl-CoA by acetyl-CoA carboxylase (ACC). This forms malonyl-CoA, which is subsequently converted by a malonyl CoA:ACP transacylase to malonyl-ACP. A series of iterative chain elongation steps that occur in a cyclical fashion then occurs. The growing acyl chain is covalently attached to the acyl-carrier protein (ACP) through a thioester linkage on the terminal sulfhydryl of the phosphopantetheine prosthetic group. The first elongation step is initiated by the Claisen condensation of malonyl-ACP with an acyl-CoA, catalyzed by the first condensing enzyme β-ketoacyl-ACP synthase III. This forms 3-ketoacyl-ACP. The β-ketoacyl-ACP synthases condense malonyl-ACP with acyl-ACP to extend the acyl chain by two carbons. The NADPH-dependent β-ketoacyl-ACP reductase reduces the 3-keto group to a 3-hydroxy intermediate. This intermediate is then dehydrated by the 3-hydroxyacyl-ACP dehydratase to a trans-2-enoyl-ACP. The final step reduces the enoyl chain by a NAD(P)H dependent enoyl-ACP reductase. After each complete cycle of condensation and chain reduction, the growing acyl chain is transferred back to the ketosynthase to initiate the next cycle (8).

Two classes of enzymes catalyze these reactions: Type-I and Type-II, which are categorized based on the enzyme architecture. Type-II fatty acid synthases are comprised of distinct individualized enzymes, completing each condensation and reduction separately. All of these enzymes require independent expression mechanisms or operon structures. Type-I fatty acid
synthase (FAS) is a large multifunctional enzyme, wherein nearly all of the enzymes involved in the fatty acid synthesis are combined into a single protein.

Fatty acids can be converted to FAEEs through an esterification reaction with ethanol, performed by a bacterial wax ester synthase (WS) (5). The initial microdiesel studies were done in E. coli (5). However, S. cerevisiae is more intuitive chassis for the biodiesel production due to its endogenous ethanol supply. Some of the original work on FAEE production in S. cerevisiae investigated five different wax ester synthases in S. cerevisiae and identified the most active enzyme as WS2 from Marinobacter hydrocarbonoclasticus (9). This enzyme was subjected to further enzymatic studies and the residues critical for selectivity to reduce the promiscuity were discovered (10,11).

Host engineering strategies have been successful in increasing FAEE production in S. cerevisiae (9,12-14). However, in these studies, the endogenously produced fatty acids which are used to generate the FAEEs are diverted from growth and cellular metabolism. Synthetic biology offers an orthologous approach to the fatty acid production, by heterologously overexpressing a parallel fatty acid synthesis pathway. This would provide additional fatty acids to supplement growth and to provide fatty acids needed for FAEE synthesis. By supplying a heterologous fatty acid pool, the flux redirected from endogenous cellular metabolism to FAEE production would be reduced. More importantly, it was hypothesized the endogenous fatty acid system could be limiting the flux towards fatty acids due to strong regulation. Therefore, it was expected that the endogenous regulation for fatty acid biosynthesis could be overcome by the heterologous expression of the fatty acid synthesis genes (15-17). In this study, we endeavored to overexpress the bacterial FAS
Brevibacterium ammoniagenes FAS-B (bafas) gene (18). This baFAS enzyme has previously exhibited a high activity in S. cerevisiae (19). The ppt1 gene from Brevibacterium ammoniagenes was co-expressed with the bafas gene, encoding for the phosphopantetheine transferase (PPT1), which activates the FAS. Co-expression of the bafas and ppt1 genes in E. coli produced palmitic acid (18). Fatty acids produced by the baFAS would then be converted to the activated acyl-CoA form by an acyl-CoA ligase (FAA1). In the final step on the pathway, the fatty acid acyl-CoA molecules would be converted to FAEEs by a wax ester synthetase/acyl-coenzyme A:diacylglycerol acyltransferase (WSDGAT), which has been extensively studied for FAEE production (5,9,12,13,20) (Figure 4.1).

4.2 Results

4.2.1 WS Functionality

Initial experiments investigated the activity of two WS/DGAT genes. One WS originated from Marino hydrocarbonoclasticus (ws2) and the other from Acinetobacter baylyi (ab). The ws2 gene was codon optimized for expression in S. cerevisiae by DNA 2.0 (Menlo Park, CA), whereas the ab gene was cloned directly from the native organism using the primers described in Table 4.1. Both genes were cloned under the PGK promoter and PGK terminator into a pRS426 plasmid. Using culture conditions which were shown to produce FAEEs (9), it was discovered the AB enzyme did not produce any significant FAEEs. Considering that this gene was not codon optimized for expression in S. cerevisiae, this result is not surprising. However, the strain harboring the codon optimized gene encoding for the WS2 enzyme did produce FAEEs (Figure 4.2a). The WS2 enzyme was further characterized: the activity towards its native substrate
hexadecanol and ethanol were analyzed (Figure 4.2b). This WS2 has been previously utilized in FAEE studies in yeast (9,12), therefore our continued work focused on this WS2 enzyme.

4.2.2 FAS Functionality

The bafas and ppt1 genes were cloned from plasmid PGM44 (18,21) using the primers from Table 4.1. The bafas gene was cloned under control of the TPI promoter, which was expected to be a strong promoter (22). The ppt1 gene was cloned under control of the TEF1 promoter, also expected to be strong (22). The corresponding bafas and ppt1 expression cassettes were cloned into two different plasmids for use in different strains with varying selection markers: pRS424 and pRS426. The pRS424-FAS&PPT1 plasmid was co-expressed with the PRS426-WS2 plasmid (strain DE17) and improved FAEE production through the baFAS enzyme was not observed (Figure 4.3).

To investigate if the heterologous genes bafas and ppt1 could function similarly to the endogenous fatty acid synthase in S. cerevisiae, the genes were expressed in a BY4741 strain deficient in endogenous fatty acid production, which was previously constructed by graduate student Jiazhang Lian in our laboratory (Table 4.2). The Δfas strain harboring the pRS426-FAS&PPT1 plasmid generated strain DE09. Strain DE09 was able to grow without exogenous fatty acid supplementation, therefore it was concluded that the bafas and ppt1 genes complemented fatty acid synthesis (Figure 4.4). The titer of the heterologous free fatty acids of strain DE09 was then compared to the endogenous free fatty acid pool titer of strain DE13, which is parental BY4741 (Figure 4.5). Strain DE09 yielded a 2.75-fold increase in free palmitic acid (p-value <0.05) compared to strain DE13.

118
4.2.3 Confirmation of a Heterologous FAEE Pathway

The endogenous *fas* knockout strain was also used to confirm if a strain harboring the full biodiesel pathway (FAS-PPT1-FAA1-WS2) could successfully produce FAEEs without endogenous fatty acids. The heterologous pathway enzymes would need to provide fatty acids for both cell growth and also for FAEE synthesis. The plasmid pRS426-FBDP (Table 4.2) was transformed into the knockout strain, generating strain DE10. After a very long recovery phase, the strain did grow and produced detectable FAEEs.

The FAEE production of strain DE10 was compared to the FAEE production of the BY4741 strain, which over-expressed only the *ws2* gene (DE01). Strain DE10 yielded a 7-fold increase in total FAEE production (Figure 4.6) over strain DE01. It is important to note that this increase in FAEE titer could be due to an increase in copy number of the plasmid, based on an increased selection pressure, which would increase *ws2* gene expression. This work demonstrates that the heterologous fatty acid synthase genes *bafas* and *ppt1* can restore fatty acid synthesis for cell growth while simultaneously providing sufficient fatty acids for FAEE production in yeast.

4.2.4 Media Testing

Fatty acid production is a very energy intensive process and it was hypothesized that the system could be nutrient limited. The pRS426-WS2 plasmid was reconstructed with antibiotic based markers (G418) for fermentation in rich medium rather than drop-out synthetic media and tested for FAEE production (strain DE06). However, no FAEEs were detected in the rich media fermentations. It was possible that the selection pressure using the antibiotic marker was not high
enough for adequate expression of the *ws2* gene. The experiment was repeated using strain DE10, which had previously demonstrated high expression of the pathway genes when cultured in synthetic media. However, fermentation in rich (YPD) media produced very small amounts of FAEEs while the same clone cultured in synthetic media (SCD-URA) produced significant FAEEs (Figure 4.7). Strain DE10 was able to grow in YPD media, therefore the plasmid was present and the baFAS enzyme was producing enough fatty acids. The WS2 enzyme must have become inactivated, either through loss of expression or inhibition. The *PGK1* promoter controlling the *ws2* gene has shown to be active in YPD media (23). Therefore, the absence of FAEE production in YPD media has remained unsolved, though future industrial systems will apply a synthetic complex media, more similar to the SCD media. Thus, the SCD media fermentations are recommended for all further analysis.

4.2.5 *HXT5 Promoter*

Expressing the pRS426-FBDP yielded heterologously produced FAEEs in the ∆fas strain (DE10). However, transformation of the pRS426-FBDP plasmid into BY4741 (DE03) did not yield significantly improved FAEE titers. Subsequent qPCR analysis revealed very low expression of the *bafas* gene (Figure 4.8). The *bafas* gene mRNA levels showed a 10-fold decrease in the levels of the *bafas* gene in DE03 compared to DE10. It was determined that the selection pressure was higher in DE10, which improved the expression of the *bafas* gene. A stronger promoter was needed to express the *bafas* gene in the wild-type BY4741 strain. A systematic look at promoters through microarray studies conducted by graduate student Jonathan Ning suggested that the promoter *HXT5* had significantly strong activity, especially in SC media (Figure 4.9).
The *bafas* gene was cloned under control of the strong *HXT5* promoter, generating plasmid pRS426-HFBDP. This plasmid was transformed into BY474, generating strain DE04, and the FAEE production from the strain was compared to the titer of the BY4741 strain harboring only the pRS426-WS2 plasmid (DE01). Strain DE04 resulted in an improved FAEE production compared to FAEEs produced from strain DE01. Palmitic acid ethyl ester production in DE04 was increased by 4.6-fold with a *p*-value of 0.003 using a standard one-tailed, two sample unequal variances t-test. Palmitoleic acid ethyl ester production increased 6.6-fold over DE01 with a *p*-value of 0.047. Stearic acid ethyl ester was also increased by 7.3-fold with a *p*-value of 0.017. Oleic acid ethyl ester production increased by 12-fold compared to DE01 with a *p*-value of 0.009. Total FAEE production was improved 6.3-fold in DE04 compared to DE01 with a *p*-value of 0.007 (Figure 4.10). Strain DE01 produced 1670 µg FAEE/ g CDW compared to the 10,498 µg FAEE/ g CDW, produced by strain DE04. Sample size for the strain DE01 was 9 and sample size of the DE04 was 11. Margin of error was high, a common problem found in the fatty acid ethyl ester quantification (12).

4.2.6 Integrated *ws2* Gene

In efforts to reduce the variability in the data and to simplify the system, the *ws2* expression cassette was integrated into multiple δ-sites of the *S. cerevisiae* chromosome. The integrated *ws2* expression system produced only slightly more FAEEs than the plasmid system, but still with large margins of error (Figure 4.11). The clone named Int7 (DE14) conferred the lowest margin of error and was chosen for further analysis. Interestingly, mRNA levels of the *ws2* gene were extremely high in the integrated version. The integrated clones of DE14 had a 65-fold
improvement in \( ws2 \) gene mRNA levels as compared to the mRNA levels of strains harboring the plasmid version of the \( ws2 \) gene (DE01). The discrepancy in high mRNA level and moderate FAEE production has not been investigated (Figure 4.12)

4.2.7 Integrated Coexpression

Super-copy plasmids with truncated promoters for the uracil selection gene were constructed by graduate student Jiazhang Lian and undergraduate Run Jin. These plasmids have been shown to have extremely high expression abilities due to increased selectivity (24). The \( bafas \) gene under the \( TPI \) promoter and also the \( HXT5 \) promoter was cloned with the \( ppt1 \) expression cassette into the supercopy plasmid. These fatty acid synthase plasmids were expressed in the \( ws2 \) integrated strain, named DE11 and DE12 for the \( TPI \) and \( HXT5 \) promoter respectively (Table 4.2). However, no increase in FAEE production compared to FAEEs produced from DE14 was observed (Figure 4.13).

4.2.8 Cellobiose Utilization for FAEE production

Ultimately, any advanced biofuel production will come from biomass, of which cellobiose is a major constituent, therefore proof-of-concept studies to be able to produce FAEEs from cellobiose are significant. The cellobiose utilization pathway which had been engineered in Chapter 3 was transformed into the DE14 strain. After 200 hours in SC media supplemented with 8% cellobiose as the sole carbon, nearly 2/3 of the cellobiose had been consumed. Extraction and quantification of the FAEEs in the culture yielded FAEE production from cellobiose as the sole carbon source (Figure 4.14).
4.3 Discussion

Fatty acids are energy-dense long chain carboxylic acids used in energy storage for many different organisms. These molecules are the precursors to fatty acid ethyl esters (FAEEs), which are a more effective biofuel than bioethanol and significant efforts have investigated biodiesel production in microbes (4,5,9,12-14,25). Host engineering strategies have been successful in increasing FAEE production in S. cerevisiae. Shi et al. overexpressed the endogenous ACCp1 gene to overproduce acetyl-CoA, a major precursor of fatty acids (9) and more recent work looked into multiple integrations of the ws2 gene into the chromosome (26). Rungphan et al. considered endogenous regulation, by replacing the endogenous promoter of the FAS1 and FAS2 with the TEF1 constitutive promoter (13). In another host engineering strategy, de Jong et al. overexpressed genes in the ethanol degradation system to improve availability of NADPH, a major component in fatty acid synthesis. A second strategy de Jong used was to over-express a phosphoketolase pathway to re-channel the carbon flux through the oxidative section of the pentose phosphate pathway, pushing the production of precursors to fatty acids (12). These strategies have all focused on using the endogenous fatty acid biosynthetic pathway of S. cerevisiae. Fatty acid synthesis genes were over-expressed to produce a larger flux of fatty acids to be shared between basic cell metabolism and growth with the FAEE production.

A complementary strategy to these host engineering methods is the use of a heterologous fatty acid synthetic pathway. This strategy can offer a completely parallel and orthologous system to produce sufficient fatty acids to minimize the flux being redirected from cellular metabolism. Additionally, this strategy can overcome the endogenous regulation mechanisms which have evolved to control the energy-intensive fatty acid biosynthesis. The endogenous FAS system has
evolved to have tight regulation and control, which could be difficult to overcome with only host engineering strategies.

4.3.1 Fatty Acid Synthesis Systems

Here we describe a heterologous FAS production system, through the use of a Type-I FAS from actinobacteria. Type-I FAS systems offer a simplified approach for overproduction of fatty acids, as compared to the Type-II FAS systems, which have been investigated for over production of fatty acids in *E. coli* (25). Type-II FAS systems consist of independently expressed genes which could require some significant engineering and balancing to optimize the pathway in a heterologous host. The Type-I system relies on a single multi-functional enzyme under a single expression system and the iterative chain elongation mechanism involves enzyme-bound intermediates being cycled through the catalytic domains. A single enzyme system is beneficial because gene expression of the individual domains do not need to be balanced and mass transfer between the enzymes is not a concern as the enzymes are linked together in their own natural protein scaffold. Balancing protein expression and mass transfer are two problems which could occur in heterologous expression of Type-II systems and have been shown to be significant in optimizing biosynthetic pathways (27-29). Each specific Type-I FAS will produce different chain length fatty acids, an inherent characteristic of the enzyme (30). Bacterial Type-I fatty acid synthases have shown to be different from other FASs by being primarily bimodal in the chain length products, generally producing a chain length of 16 to 18 and also 24 to 30 carbons.
4.3.2 Heterologous Fatty Acid Ethyl Ester Synthesis

The heterologous pathway in this work involved the Type-I fatty acid synthase (FAS) from *B. ammoniagenes* coupled with independent expression of its respective phosphopantetheine transferase (PPT1). The PPT1 is responsible for activating the acyl carrier protein of the baFAS. Additionally, the free thiol moiety of the PPT1 covalently binds to the elongating fatty acid chain, transferring the chain through the distinct reaction domains of the FAS (31). The heterologously produced fatty acids are processed by an over-expressed fatty acid acyl-CoA ligase from *S. cerevisiae*. The activated fatty acids are then used as a co-substrate with ethanol by the wax synthase to produce FAEEs. Through this strategy, total FAEE production was increased 6.3-fold compared to a strain expressing only the *ws* gene, with a *p*-value of 0.0078 using a standard one-tailed, two sample unequal variances t-test (Figure 4.10). The *p*-value statistical analysis is important to consider based on the variation from biological clones. It is expected that the variation is due to the 2-micron plasmids variations, introducing different copy numbers of plasmids to each clone.

4.3.3 HXT5 Promoter and FAS Expression

Addition of the *HXT5* promoter was also important in this study as the *bafas* gene would not express unless it was under the control of the very strong promoter *HXT5*. Graduate student Jonathan Ning investigated a microarray study to determine strong promoters. Of the eight promoters investigated in his study, the *HXT5* promoter was the strongest in synthetic (SC) drop-out media. Often the *TEF1* is considered a strong promoter, but the *HXT5* exhibited a 3-fold increase in GFP-expression in control studies compared to the *TEF1*. The *HXT5*
transporter protein is known to have a low affinity for glucose and is expressed in nutrient limiting conditions, therefore a strong expression is required for growth in minimal media (32).

It is important to note that there was significant variation in this system. With each biological replicate there was a large error for the FAEE quantification, which is attributed to the variation in the copy number of the plasmid and also in the extensive extraction protocol. One strain (DE04) exhibited the 6.3-fold improvement in the total FAEE production. Initial work with this clone resulted in a total FAEE titer of 10,498 µg/ gCDW, averaged from 11 independent clones. However, after retrieving this clone from the frozen stock, the titer decreased to 2,780 µg/ gCDW, though the overall titer from DE01 also decreased. The fold improvement was still 6-fold over the DE01 strain. To confirm this improvement was coming from the plasmid, the DNA was transformed into fresh BY4741 cells, clones from these new transformants also exhibited increased FAEE production, however none has significantly improved as strain DE04. These fresh transformants, compared to DE01, exhibited a 2- and 5-fold improvement instead. Additionally, some clones exhibited a decrease in FAEE production compared to DE01. Causes of these major variations are currently unknown, but assumed to be due to copy number variations in the plasmid. However, DE04 has been tested many times and has continuously exhibited the significant improvement in the FAEE production, with a final titer of 2,780 µg/ gCDW.

4.3.4 Cellobiose Consumption for FAEE Production

Coming full-circle in this thesis, the cellobiose utilization pathway which was optimized in Chapter 4 was used to produce FAEEs from cellobiose as a sole carbon source. The media and
host conditions were not the same as which were used to optimize the cellobiose utilization pathway, therefore this pathway is not extremely efficient. However, we wanted to simply demonstrate that cellobiose can be consumed as the sole carbon source to produce FAEEs. It is hypothesized that use of cellobiose or xylose as a substrate could improve the availability of NADPH, by shuttling through the pentose phosphate pathway, instead of glycolysis, however no significant controlled experiments were conducted for this analysis in this study.

This work represents the first time a heterologous FAS pathway has been constructed to over-produce fatty acids for the production of biofuels in *S. cerevisiae*. This strategy yielded a 6.3-fold improvement in total FAEEs produced. This improvement is based on increased availability of heterologously produced fatty acids providing an orthologous route for FAEE production, which essentially decoupled the FAEE production from endogenous metabolism. However, it should be noted that even though the fatty acid synthesis is heterologous and orthologous, the fact that the precursor of fatty acid synthesis, acetyl-CoA, is not heterologously produced should not be ignored. Using host engineering strategies to improve production of the acetyl-CoA could increase the productivity of the pathway even more. Additionally, combining this synthetic biology approach with the host engineering strategies previously mentioned could synergistically increase the FAEE titer (9,12,13).

4.3.5 Future work

Figure 4.15 is a thin layer chromatography plate visualizing the lipids extracted from strains DE01 and DE04. Unidentified lipid spots in strain DE04, harboring the full biodiesel pathway, have a higher concentration than compared to spots on DE01, the strain harboring only the ws2
gene. It is suggested for future work in this pathway to consider these unidentified lipids and knockout these competing routes to redirect the flux towards FAEEs.

4.4 Conclusion
Strategies and solutions to identify future renewable energy sources will be complex and diverse. In order to construct a broad arsenal of possible routes for success, different strategies must be explored and combined for synergistic improvements. We offer here, a new pathway for FAEE production in *S. cerevisiae*. By an orthogonal fatty acid synthesis, the supply of fatty acids for both growth and for FAEE production is increased, therefore not diverting the flux of fatty acids away from cellular metabolism.

4.5 Materials and Methods
4.5.1 Strains and plasmids
*S. cerevisiae* strains used in this work are provided in Table 4.2. All plasmid constructions were accomplished through DNA assembler (33) unless otherwise noted. The *ws2* gene was codon optimized and synthesized by DNA 2.0 (Menlo Park, CA) and assembled under control of the *PGK1* promoter/terminator in a pRS426 plasmid (pRS426-WS2). The *bafas* and *ppt1* genes from *Brevibacterium ammoniagenes* were cloned from pGM44 (18) under control of the *TPI1/TPI1* promoter/terminator and *TEF1/TEF1* promoter/terminator respectively (pRS426-FAS). The *faa1* gene from *S. cerevisiae* was expressed under control of the *PYK1/ADHt* promoter and terminator (pRS425-scFAA1). Due to the size and GC composition of the *bafas* gene, we endeavored to avoid excessive PCR amplification of the gene. For ease of plasmid construction, a helper plasmid was assembled first which included the expression cassettes for the *ppt1*, *ws2*, and *faa1*. 
The pRS426-FAS plasmid was linearized and used as backbone to insert the helper plasmid expression cassettes in 5 kilobase (kb) fragments with 2 kb homologous region overlap. Further work required replacing the *TPI1* promoter of the *bafas* gene with the *HXT5* promoter, which was completed by using restriction digests *PacI* and *RsrII* to separate the *TPIp* and part of the *bafas* gene. The plasmid was reassembled with the *HXT5* promoter and the *bafas* gene fragment. Construction of all plasmids was confirmed via restriction digest maps and partial sequencing.

The SuperCopy plasmid was constructed by graduate student Jiazhang Lian and undergraduate student Run Jin. Truncation of the promoter of the selection marker weakened the cell’s ability to synthesize uracil. This weakened expression forced a high copy number of the plasmid to accommodate for the loss of activity. This plasmid was used to force a high selection pressure and therefore high expression of the *bafas* and *ppt1* gene.

The BY4741 Δfas was constructed by graduate student Jiazhang Lian. The full *fas1* gene was replaced with a KanMX expression cassette, completely disrupting the endogenous fatty acid synthesis.

All transformations followed standard Lithium Acetate transformation protocol for *S. cerevisiae* as described previously (33). However, transformation of DNA into the Δfas strain required a 12 hour incubation period after inoculation, supplemented with exogenous 2% palmitic acid and G418 for selection of the knockout. The transformants were recovered in YPD media with 0.01% palmitic acid for five hours. The cells were then transferred to agar plates or liquid media with and without palmitic acid supplementation.
4.5.2 Media, and Culture Conditions

Seed cultures of strains for FAEE detection were grown in 2% glucose CSM-URA or CSM-LEU (MP Biomedical, Santa Ana California) to stationary phase at 30°C with orbital shaking at 250 rpm and then transferred to 125-mL baffled flasks with 25 mL SC-URA at an initial OD of 0.05 at 30°C with orbital shaking at 100 rpm. As needed, 200 µg/mL G418 (KSE Scientific, Durham, NC) supplemented media for selection. *Escherichia coli* 10-β (New England Biolabs, Ipswich, MA) was used for recombinant DNA manipulations. *E. coli* strains were cultured in Luria broth (LB) (Fischer Scientific, Pittsburgh, PA) at 37°C and 250 rpm, supplemented with 50 µg/mL ampicillin. Yeast and bacterial strains were flask frozen and stored in 15% glycerol at -80°C. All chemicals were purchased from Sigma Aldrich or Fisher Scientific.

4.5.3 Lipid Extraction and Thin Layer Chromatography

FAEE detection was conducted as previously reported (5,9,34). Cells were harvested and washed with sterile water before being applied to a freeze-dryer overnight. Dried cell weight was normalized and then rehydrated with digestion buffer (1.2 M glycerol, 100 mM sodium thioglycolate, 50 mM Tris-SO4, pH 7.5) and 15 Units of zymolase (Zymo Research, Irvine CA). After incubation at 37°C and 250 rpm, cells were extracted twice with 2:1 chloroform methanol mixture at 4°C with continuous vortexing. 25 µg of lauric acid ethyl ester was used as an internal standard. The lipid mixture was applied to a TLC silica gel 60 F254 plate (EMB Millipore, Darmstadt, Germany) with a mobile phase of heptane, 2-propanol, and acetic acid (95:5:1 v/v/v). Lipids were visualized through a 0.05% 2,7-dichlorofluoresceine mixture in ethanol. The FAEEs were isolated by scraping the FAEEs off the TLC plates using a scalpel and then extracted from
the TLC power with a mixture of hexane, methanol, and water (3:2:2). The upper phase was separated and transferred to glass GC-vial inserts for analysis.

4.5.4 Free Fatty Acid Extraction

Quenching was conducted on the cells to quantify the free fatty acid pool. Cells were flash frozen by a methanol-dry ice bath. The mixture was centrifuged at 4000 rpm for 5 minutes and then the cell pellet was separated from the media. Cells pellets were then lysed by boiling at 100°C in ethanol for 10 minutes. Cell membrane was separated by centrifugation and the supernatant was applied to a freeze-dryer overnight or until dry. The lipids were then extracted twice with 2:1 chloroform/methanol using 50 µM myristic acid as an internal standard. After extraction, the fatty acids were allowed to derivatize for 2 hours at room temperature with 200 µL of 2N TMS-diazomethane, 10 µL methanol and 10 µL HCL. The sample was then transferred to a glass GC vial for analysis.

4.5.5 GC-MS Analysis

The fatty acids were separated and quantified through a GC-MS-QP 2010 Plus (Shimadzu, Kyoto Japan). The separation was performed by a DB-WAX GC column with the following dimensions: 30 mm x 0.25 mm internal diameter, 0.25 µm film thickness, from Agilent Technologies (Santa Clara, CA). A 1 µL portion was injected via split-less injection at 250 °C at an initial pressure of 98 kPA and a total flow of 14.1 mL/min helium carrier gas. The chromatograph separation initially occurred at a temperature of 40 °C for 1.2 minutes and then increased at a rate of 30 °C per minute to 220°C and then held for fifteen minutes. The mass
transfer line and ion source were at 250 and 200 °C respectively. The fatty acids were detected with an electron ionization method in scan mode from 50 to 650 m/z.

FAEEs were quantified through the same GC, column, and injection procedure. The chromatograph separation initially occurred at a temperature of 50°C for 1.5 minutes and then increased at a rate of 25°C per minute to 180°C, finally increasing the temperature at 10°C per minute to 250°C which was held for 3 minutes. The mass transfer line and ion source were at 250 and 200 °C respectively. The FAEEs were detected with an electron ionization method in scan mode from 50 to 650 m/z. The identification of both the fatty acids and the FAEEs were achieved by comparison of retentions times and mass spectrum.

4.5.6 qPCR Analysis

The total RNA was isolated from fresh or frozen samples stored at -80°C using the MP Biomedical Fast RNA Prep for Yeast kit (Santa Ana, California) following manufacturer’s instructions. Samples were reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit (Roche, Mannheim, Germany) with the oligo-dT primer following the manufacturer’s instructions. The qPCR was performed with LightCycler 480 SYBR Green Master reagents (Roche) using the Roche Light Cycler® 480 System (Roche, Indianapolis, IN) using gene specific primers (Table 4.1). The relative abundance of the mRNA levels for the target genes was normalized to the asparagine-linked glycosylation 9 gene (alg9) expression to determine expression levels.
4.5.7 Radioactive WS Enzyme Activity Assay

Cells harboring the pRS426-WS2 plasmids were harvested at mid to late exponential phase and washed twice with potassium phosphate containing 2 mM magnesium chloride at pH 7.5, then resuspended with 1 mM DTT in the potassium phosphate buffer. Cells were lysed using MP Biomedical FastPrep 24 with Lysing Matrix B. The total protein concentration was determined via the BCA protein assay kit (Pierce, Rockford, IL) and the standard manufacturer protocol was followed. 3.75 mM of the substrate of interest (ethanol or hexadecanol) were combined with 5 mg/mL BSA in potassium phosphate buffer with 2 mM MgCl2. 1.17 ng of C-14 labeled palmitoyl-CoA labeled was added. The reaction was started with the addition of 100 µg of crude cell lysate and allowed to react for 30 minutes. The reaction was halted by addition of 500 µL of chloroform:methanol. 25 µg of lauric acid ethyl ester was used as an internal standard. The reaction mixture was applied to a TLC silica gel 60 F_{254} plate (EMB Millipore, Darmstadt, Germany) with a mobile phase of heptane, 2-propanol, and acetic acid (95:5:1 v/v/v). Lipids were visualized through a 0.05% 2,7-dichlorofluoresceine mixture in ethanol. The FAEEs were isolated by scraping the FAEEs off the TLC plates using a scalpel. The FAEEs were extracted from the powder by vortexing in hexane, methanol, and water. Finally, the hexane layer was added to 3 mL scintillation fluid and analyzed for radioactivity, quantified by a Beckman Coulter LS 6500 liquid scintillation counter (Brea, CA).

4.5.8 Integration of WS2

Integration of the ws2 gene into S. cerevisiae chromosome was achieved through recombination at the δ sites. The plasmid pITy2 (35) was linearized by digestion with KpnI and SacI to insert the WS2 expression cassette under control of the PGK promoter using Gibson Assembly (New
England Biolabs, Ipswich, MA). For integration, the plasmid harboring the \textit{ws2} expression cassette was linearized with \textit{XhoI} and 5 \( \mu \text{g} \) of linearized DNA was transformed into CEN.PK \textit{S. cerevisiae} strain using the Lithium Acetate method. The transformed cells were plated on 3.5 g/L G418 for selection of the integrated expression cassette. Integration into the chromosome was confirmed by whole-cell PCR amplification of the \textit{ws2} cassette.
4.6 References


## 4.7 Tables

Table 4.1: Selected primers used in this study.

<table>
<thead>
<tr>
<th>Genes</th>
<th>Forward</th>
<th>Reverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAS Primers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPI F</td>
<td>5' - GACGGTGTAAACGACGGCCAGTGAGCGCGCTAATACGATATATCTAGGAACCAGTA</td>
<td>3'</td>
</tr>
<tr>
<td>TPI R</td>
<td>5' - AAACACATACTAAAAACGACTATGGGATCTCTACAAACCACGCCCTGGTAGAAGCG</td>
<td></td>
</tr>
<tr>
<td>FAS F</td>
<td>5' - TACGCGCAAGCCACGGATGAGATTATAATAATATATAAATATTATCCTCTTCTT</td>
<td>3'</td>
</tr>
<tr>
<td>FAS R</td>
<td>5' - CAAGGCGTGTGGCTAAGCGATCTAGCGCTAAATCTAGGGATCCTCTAAGC</td>
<td>3'</td>
</tr>
<tr>
<td>TPIt F</td>
<td>5' - AAATTTCAACTGTAGTTATATAATAGGATTTTCTACTCTTTTTAATTCCAG</td>
<td>3'</td>
</tr>
<tr>
<td>TPIt R</td>
<td>5' - GGTGAGAAGATGTTCTTATTCCAAATTTACATGTTATAATAGGCTC</td>
<td></td>
</tr>
<tr>
<td>TEF F</td>
<td>5' - GTACGCGATACCCAGTGAAAGATTTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>TEF R</td>
<td>5' - CAAACACGATACCCAGTGAAAGATTTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>WSDGAT Primers</td>
<td>5' - GGAAGAAAGCATACACCTGCAAGCTTAGACATTTCTGCTAGTTGGCTTGCAG</td>
<td>3'</td>
</tr>
<tr>
<td>PGKp F</td>
<td>5' - TACAACAAAATGTTGACGCTATACGCCAGATATATAACATCTGGCAAAATAGGCA</td>
<td>3'</td>
</tr>
<tr>
<td>PGKp R - to ab</td>
<td>5' - CAAAGAAGATTTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>ab F</td>
<td>5' - AAACACGATACCCAGTGAAAGATTTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>ab r</td>
<td>5' - GGAAGTAATTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>PGKp RtoWS2</td>
<td>5' - GTATAGTGATATCTCTCTGCTAGTTGGCTTGCAG</td>
<td>3'</td>
</tr>
<tr>
<td>WS2 F</td>
<td>5' - GCGTGCGTACTAGAAAATGTTAATGTGAAATATTGGTATCAGAAATTGCTAGTCTCTAG</td>
<td>3'</td>
</tr>
<tr>
<td>WS2 R</td>
<td>5' - CAAAGAAGATTTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>PGKt F to ab</td>
<td>5' - CAAAGAAGATTTATCTACTTTTTACATTAGTGGATGATG</td>
<td>3'</td>
</tr>
<tr>
<td>PGKt F to WS2</td>
<td>5' - GGTGAGAAGATGTTCTTATTCCAAATTTACATGTTATAATAGGCTC</td>
<td>3'</td>
</tr>
<tr>
<td>PGKt R</td>
<td>5' - CAAGGCGTGTGGCTAAGCGATCTAGCGCTAAATCTAGGGATCCTCTAAGC</td>
<td>3'</td>
</tr>
</tbody>
</table>
### Table 4.2: Strains and plasmids used in this study

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Name</th>
<th>Genes</th>
<th>Promoters/Terminator</th>
<th>Marker</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pRS426- abWS</td>
<td>A. baylyi WS</td>
<td>PGK1/PGKt</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS426- WS2</td>
<td>M. hydrocarbonoclasticus WS2</td>
<td>PGK1/PGKt</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS424-FAS</td>
<td>B. ammoniagenes FAS and PPT1</td>
<td>TPI/TPI , TEF1/TEF</td>
<td>TRP</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS426-FAS</td>
<td>B. ammoniagenes FAS and PPT1</td>
<td>TPI/TPI , TEF1/TEF</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS425-scFAS</td>
<td>S. cerevisiae FAA1</td>
<td>PYK1/ADH1</td>
<td>LEU</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS416-Helper</td>
<td>PPT1, FAA1, WS2</td>
<td>TEF1/TEF, PYK1/ADH1, PGK1/PGK1</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS426-FBDP</td>
<td>baFAS, PPT1, FAA1, WS2</td>
<td>TPI/TPI , TEF1/TEF, PYK1/ADH1, PGK1/PGK1</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS426-HFBDP</td>
<td>baFAS, PPT1, FAA1, WS2</td>
<td>HXT5/TPI , TEF1/TEF, PYK1/ADH1, PGK1/PGK1</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS426-K1 FBDP</td>
<td>KanMX, baFAS, PPT1, FAA1, WS2</td>
<td>TPI/TPI , TEF1/TEF, PYK1/ADH1, PGK1/PGK1</td>
<td>G418</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pRS426- K1 WS2</td>
<td>KanMX, WS2</td>
<td>PGK1/PGK1</td>
<td>G418</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>SuperCopy (SC)</td>
<td></td>
<td></td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>SC-TFAS</td>
<td>baFAS</td>
<td>TPI/TPI</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>SC-HFAS</td>
<td>baFAS</td>
<td>HXT5/TPI</td>
<td>URA</td>
<td>This study</td>
</tr>
<tr>
<td></td>
<td>pTy</td>
<td></td>
<td></td>
<td></td>
<td>(31)</td>
</tr>
<tr>
<td></td>
<td>pTy+WS2</td>
<td>WS2</td>
<td>PGK1/PGKt</td>
<td>G418</td>
<td>This study</td>
</tr>
</tbody>
</table>

### Strains

<table>
<thead>
<tr>
<th>Genetic Background</th>
<th>Plasmid</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>BY4741</td>
<td>pRS426- WS2</td>
<td>DE01</td>
</tr>
<tr>
<td>BY4741</td>
<td>pRS426-FAS</td>
<td>DE02</td>
</tr>
<tr>
<td>BY4741</td>
<td>pRS426-FBDP</td>
<td>DE03</td>
</tr>
<tr>
<td>BY4741</td>
<td>pRS426-HFBDP</td>
<td>DE04</td>
</tr>
<tr>
<td>BY4741</td>
<td>pRS426-K1 FBDP</td>
<td>DE05</td>
</tr>
<tr>
<td>BY4741</td>
<td>pRS426- K1 WS2</td>
<td>DE06</td>
</tr>
<tr>
<td>BY4741</td>
<td>SC-TFAS</td>
<td>DE07</td>
</tr>
<tr>
<td>BY4741</td>
<td>SC-HFAS</td>
<td>DE08</td>
</tr>
<tr>
<td>BY4741</td>
<td>Δfas1</td>
<td>DE09</td>
</tr>
<tr>
<td>BY4741</td>
<td>Δfas1</td>
<td>DE10</td>
</tr>
<tr>
<td>CEN.PK ::WS2</td>
<td>SC-TFAS</td>
<td>DE11</td>
</tr>
<tr>
<td>CEN.PK ::WS2</td>
<td>SC-HFAS</td>
<td>DE12</td>
</tr>
<tr>
<td>BY4741</td>
<td></td>
<td>DE13</td>
</tr>
<tr>
<td>CEN.PK ::WS2</td>
<td></td>
<td>DE14</td>
</tr>
<tr>
<td>CEN.PK</td>
<td>pRS426- WS2</td>
<td>DE15</td>
</tr>
<tr>
<td>CEN.PK</td>
<td>pRS426- WS2 + pRS424</td>
<td>DE16</td>
</tr>
<tr>
<td>CEN.PK</td>
<td>pRS426- WS2 + pRS424-FAS</td>
<td>DE17</td>
</tr>
<tr>
<td>CEN.PK</td>
<td></td>
<td>DE18</td>
</tr>
<tr>
<td>CEN.PK ::WS2</td>
<td></td>
<td>DE19</td>
</tr>
</tbody>
</table>

The strains and plasmids used in this study are listed in Table 4.2. The genetic background and plasmid used for each strain are specified. The marker for each plasmid is also provided.
4.8 Figures

**Figure 4.1:** Description of microbial production of FAEEs. Heterologous pathway used in this study is depicted by red arrows (9). In this study, the heterologous expression of the Type-I FAS is a key component. This heterologous expression provides an orthogonal route for fatty acid synthesis to reduce the flux diverted from cellular metabolism. The WS/DGAT enzyme uses the activated form of fatty acids and ethanol as a co-substrate to form fatty acid ethyl esters.
Figure 4.2: Initial characterization of the WS/DGAT genes cloned. (a) FAEE production from heterologous expression of the wax synthase genes. The gene \textit{ab} is cloned from the bacteria \textit{Acinetobacter baylyi} and the \textit{ws2} gene has been codon optimized for \textit{S. cerevisiae} from \textit{Marino hydrocarbonoclasticus}. (b) Radio-labeled activity assay of the WS2 enzyme testing for preferred substrate.
Figure 4.3: Dual plasmid expression of the *baf* and *ws2* genes in a single strain (DE17) in two different plasmids. No increase in FAEE production was observed with the addition of the FAS enzyme. C16 represents the palmitic acid ethyl ester, C16:1 is palmitoleic acid ethyl ester. Values are averages of two biological replicates.
**Figure 4.4:** Strain DE09, which expressed the *bafas* gene in the ∆*fas* BY4741 strain under the *TPI* promoter, was able to grow without exogenously supplied fatty acids (FA). This illustrated that the *bafas* gene was able to complement the ∆*fas* strain. Therefore, the FAS is functional in yeast and can produce palmitic acid, which can be used for cell growth.
**Figure 4.5:** Free fatty acid pools of the wild-type BY4741 strain (DE13) compared to the free fatty acid pools of the BY4741 ∆fas1 + pRS426-baFAS (DE09). Strain DE09, heterologously expressing the *bafas* gene, was able to produce up to 2-fold increase in free fatty acids compared to strain DE13. Values are averages of biological triplicates with one standard deviation. The asterix symbolizes samples with a *p*-value of less than 0.05. C16 is palmitic acid, C16:1 is palmitoleic acid. C18 is stearic acid, and C18:1 is oleic acid. Values are relative to the internal standard.
Figure 4.6: Comparing the FAEE production of the BY4741 strain expressing only the ws2 gene (DE01) to the Δfas knockout strain complemented with the heterologous FAS and WS2 pathway (DE10). There is a 7-fold increase in total FAEE production in DE10 compared to FAEE production in strain DE01. Each sample was repeated in triplicate and one standard deviation is represented. The asterix symbolizes a $p$-value of less than 0.05 compared to the parental BY4741 strain expressing the ws2 gene. C14 is myristic acid ethyl ester. C16 is palmitic acid ethyl ester and C16:1 is palmitoleic acid ethyl ester. C18 is stearic acid ethyl ester and C16:1 is oleic acid ethyl ester.
Figure 4.7: The BY4741 Δfas1 complemented with the pRS426-FBDP (DE10) in different media selections. A clone which had previously exhibited significant FAEE production in SCD media was separated into YPD and SCD media, using three different biological replicates. Even considering a strong selection pressure to complement the Δfas1 for growth, no significant FAEEs were produced in YPD media. C14 is myristic acid ethyl ester. C16 is palmitic acid ethyl ester and C16:1 is palmitoleic acid ethyl ester. YPD media is not suggested for any future work with the WS2 enzyme.
**Figure 4.8:** mRNA levels of the *bafas* and *ws2* genes in different strains. The strain DE09 (Δfas+pRS426-FAS) exhibited a 10-fold increase in *bafas* mRNA levels compared to the strain DE03 (BY4741+pRS426-FBDP) and DE02 (BY4741+ pRS426-FAS). This experiment demonstrates that the FAS can be expressed in *S. cerevisiae*. The relative abundance of the mRNA levels for the target genes was normalized to the asparagine-linked glycosylation 9 gene (*alg9*) expression to determine expression levels.
Figure 4.9: Fluorescence levels of green fluorescence protein (GFP) under control of eight different promoters in synthetic drop-out media. Data collected by graduation student Jonathan Ning. HXT5 was shown to express extremely well in SC media.
Figure 4.10: Comparison of FAEE production strain harboring the heterologous pathways in wild type BY4741 strain. The FAEE production is increased in strain DE04, when the HXT5-FBDP (FAS+PPT1+FAA1+WS) is heterologously expressed. This strain exhibited a 6.3-fold increase in total production of FAEEs compared to strain DE01. The asterix symbols represent chain lengths of fatty acids which are statistically greater than the pRS426-WS control (p<0.05), the double asterix symbol represents chain lengths of fatty acids which are significantly statistically greater than the pRS425-WS control (p<0.01). Sample sizes ranged from n= 9 for DE01 and n= 11 for DE 04. C14 is myristic acid ethyl ester. C16 is palmitic acid ethyl ester and C16:1 is palmitoleic acid ethyl ester. C18 is stearic acid ethyl ester and C16:1 is oleic acid ethyl ester.
**Figure 4.11:** FAEE production of the WS integrated version compared to the WS plasmid version (DE01). There was no major increase in FAEE production or significant reduction in error in the integrated version as compared to the plasmid-based system. Continued work with the integrated strain used clone Int7 (DE14) due to the lower error margin. C14 is myristic acid ethyl ester. C16 is palmitic acid ethyl ester and C16:1 is palmitoleic acid ethyl ester. C18 is stearic acid ethyl ester and C16:1 is oleic acid ethyl ester.
**Figure 4.12:** Expression levels of the *ws* integrated version (DE14) compared to the plasmid strain (DE01). Though the increase in FAEE production of the integrated strain was marginal, the relative expression level of the *ws2* gene in the DE14 strain was nearly 60-fold greater than that of the mRNA levels of the *ws2* gene in the plasmid version (DE01). The relative abundance of the mRNA levels for the target genes was normalized to the asparagine-linked glycosylation 9 gene (*alg9*) expression to determine expression levels.
Figure 4.13: FAEE production in co-expression of integrated WS strain (DE14) and supercopy FAS. Both promoters $TPI$ (DE11) and $HXT5$ (DE12) were investigated to determine if the $bafas$ gene would express under these highly-selective conditions. After 48 hours, no significant increase in FAEE production was detected though expression of the FAS under the super-copy plasmid. C14 is myristic acid ethyl ester. C16 is palmitic acid ethyl ester and C16:1 is palmitoleic acid ethyl ester. C18 is stearic acid ethyl ester and C16:1 is oleic acid ethyl ester.
**Figure 4.14:** FAEE production from cellobiose as a sole-carbon source. Ultimately, any advanced biofuels will be produced from biomass, of which cellobiose is a major constituent of. Therefore a proof-of-concept study was examined here to identify if FAEEs could be produced from cellobiose as a sole carbon source, using the pathways optimized in Chapter 3 for cellobiose utilization. This pathway was transformed into DE14, a strain with the \( \text{ws} \) gene integrated into the chromosome and grown on SC-8% cellobiose. Samples taken after 200 hour fermentation when nearly all of the cellobiose had been utilized. C14 is myristic acid ethyl ester. C16 is palmitic acid ethyl ester and C16:1 is palmitoleic acid ethyl ester. C18 is stearic acid ethyl ester and C16:1 is oleic acid ethyl ester.
**Figure 4.15:** Thin layer chromatography visualization of the lipids extracted from the cells. The first three lanes are biological replicates of strain DE01, BY4741 harboring the WS2 enzyme. The last three samples are biological replicates of strain DE04, harboring the full biodiesel pathway. The black arrows indicate unidentified lipids which are produced in higher quantities in DE04 than DE01. Future work is suggested to identify these lipids and consider knocking out the key enzymes in their synthesis to redirect flux towards the FAEEs.
Chapter 5: Developing a High-throughput Screen for FAEE production in S. cerevisiae

5.1 Background

Recent advances in DNA assembly methods have provided scientists and engineers extraordinary freedom in constructing large DNA pathways (1). This has given synthetic biologists an unprecedented ability to make vast quantities of biosynthetic systems for nearly any desired purpose, with applications in genetic circuitry, microbial chemical production, or biomedical devices and treatments. Furthermore, high efficiency and modularity of these advanced DNA assembly methods makes in silico design (2) and automated assembly (3) of these constructs possible. Large combinations of DNA parts and libraries can be individually constructed by robotic platforms, further increasing the potential for immense libraries of pathways and biosynthetic systems.

Despite the rapid progress of DNA assembly technologies, widespread application of large-scale DNA assembly and pathway libraries is currently limited by the ability to effectively and efficiently interrogate the resulting pathways and libraries. Without the ability to easily and economically quantify for the phenotype of interest, these large-scale methods will not be able to fulfill their maximum potential.

Chapters 2 and 3 of this thesis focused on developing pathway library optimization strategies. It is desired to apply these optimization methods to the fatty acid ethyl ester (FAEE) pathway constructed in Chapter 4. However, there is a major hurdle which must first be overcome before applying these methods to the FAEE pathway. Optimization of the cellobiose utilization pathway from Chapter 2 and 3 was readily performed through visual check of large colonies (4). Strains
harboring an improved cellobiose utilization pathway would grow the quickest on media with cellobiose as the sole carbon source. However, such a simple screen is not available for FAEE synthesis in *S. cerevisiae*. This final chapter of the thesis will consider three different strategies to develop a high throughput screen and ultimately apply pathway optimization strategies to FAEE synthesis developed in Chapter 4.

5.1.1 *Transcription Factor based Biosensors*

Biosensors have significant potential in high-throughput screening methods. This has been demonstrated by the large number of transcription factor based biosensors engineered to detect small molecules (5-8). These biosensors can link the small molecule concentration to an easily measurable signal such as fluorescence and cell growth via gene circuits.

Promoters have evolved over millions of years to function in response to certain stimuli, therefore utilizing transcription factors and their cognate promoters seems only sensible for a high-throughput screen. Developing screens based on the response of transcription factors for a detectable, phenotypic response has been well documented (5,8). Biosensors can be used to detect environmental factors such as pH, temperature, light, and nutrient limiting conditions. Though these sensors are significant, the most important type of biosensors for metabolic engineering is sensors which can detect intracellular chemicals and metabolites. Regulatory RNA molecules have been evolved by nature to act as biosensors for intercellular metabolites. For example, *B. subtilis* has evolved a riboflavin or a thiamin-responsive mRNA riboswitch: once the ligand (riboflavin) binds to the RNA, a conformational change is induced in the mRNA, which leads to premature transcription termination (9). It is possible to mine other organism genomes to
identify other regulatory elements which respond to a metabolite of interest. Once a transcription factor and its respective promoter which are sensitive to that metabolite are identified, the DNA parts can be cloned in a system that will trigger a phenotype in the presence of the metabolite.

5.1.2 Growth Based Screen

Perhaps one of the most basic forms of a high-throughput screen is the growth-based screen. In this strategy the metabolite or component being interrogated is fundamental to cell growth. The presence or absence of this compound will affect cell growth, either positively or negatively. This strategy was used by Du et al. and Kim et al. to screen for improved pentose sugar utilization (4,10). The libraries were grown on the pentose sugar as a sole carbon source, strains which harbored an improved sugar utilization pathway would grow faster. Therefore, these improved strains would have a larger colony size on the agar plate. A similar strategy was used by Kambam et al. (11). In this work to engineer the substrate specificity of the RhII, a component in a complex genetic circuit, the RhII specificity was linked to ampicillin resistance. Therefore, a mutated RhII protein with improved substrate specificity towards acyl-ACPs would yield a higher ampicillin resistance. Thus, colonies which appeared first on high concentrations of ampicillin agar plates were associated with the improved protein mutants. Another example is screening in toxicity studies. Genetic diversity can be introduced to the cell through a variety of methods, either through pathway engineering (4,10,12), RAGE (13), MAGE (14), mutator strains (15), and then the strains are applied to a toxic chemical with increasing concentrations (16). The strains which grow the fastest and the best in these toxic situations have improved tolerance.
5.1.3 Dyes and stains

Dyes and stains have also often been used for screening strategies. These strategies were originally developed as visualization methods of structures within the cell. Of particular interest in this project are lipophilic fluorescent dyes, including Sudan black (17), Nile Red (18), Nile blue (19) and BODIPY (20). The Nile Red (9-diethyl amino-5H-benzo[a]phenoxazine-5-one) stain has been used most commonly (21). When this fluorogenic dye interacts with cellular lipids in solution, it emits a fluorescence signal which is significantly higher than background. Additionally, the emission wavelengths change with the hydrophobicity. Polar lipids interacting with the Nile Red will lead to an emission longer than 590 nm, while neutral lipid interaction will emit at 520 and 290 nm (21).

High-throughput applications of Nile Red have also been studied (22,23). Of particular interest, Sitepu et al. investigated the high-throughput use of Nile red in yeast species (23). In this study, the addition of dimethyl sulfoxide (DMSO) was used to improve the diffusion of the Nile red across the cell membrane. Additionally, kinetic studies were implemented, monitoring the change in fluorescent emission over a 20 min period. This allowed for the optimal fluorescence reading to be discovered.

5.1.4 Project Overview

This project was intended to identify a high–throughput screening method which would allow for optimization of the FAEE pathway constructed in Chapter 4. Unfortunately, direct product detection will be extremely difficult as a complicated extraction and separation method is used which depends on a GC-MS analysis. However, during construction of the pathway, expression
of the fatty acid synthase (FAS) gene was very difficult. Therefore, it was expected that if a screen could be based on the product of the FAS gene, it would improve overall activity of the pathway. Therefore, the following assays are based on fatty acid detection.

5.2 Results

5.2.1 POX1- GFP Expression Dynamic Range

For fatty acid detection, a green fluorescence protein (GFP) was put under control of the fatty acid sensitive POX1 promoter. In this system, with the presence of fatty acids, the POX1 would initiate expression of the GFP. This expression cassette was successfully cloned into a p403 integration plasmid. Integration plasmid p403 was required due to the auxotrophic markers of strain BY4741, which was used in this study. The linear plasmid was transformed into BY4741 and plated into SC-Leu. Any resulting colonies were confirmed for integration (Figure 1).

POX1 is inhibited by glucose, therefore the strain was grown in raffinose and glycerol and supplemented with exogenous fatty acids. After six hours, the fluorescence was quantified. Though in this study, no fluorescence was observed, graduate student Tong Si has shown the range to be 0.5-2 g/L oleic acid (Figure 2), with the linear range being between 0 and 1 g/L of exogenously added oleic acid.

5.2.2 Growth Based Assay

Fatty acids are essential for growth. Therefore, a Δfas strain cannot grow without fatty acids and the more fatty acids present in vivo, the faster the strain will be able to grow. Initial high-throughput tests were conducted on the Bioscreen C and exogenously supplied fatty acid
It was shown that increased concentration of exogenous fatty acids did improve cell growth of the ∆fas strain. Further testing considered in vivo fatty acid concentration. Graduate student Sam Hamedi Rad cloned a series of homologous FAS genes from several organisms in the actinobacteria family and also eukaryotic origins: Brevibacterium ammoniagenes, Gordonia terrae, Mycobacterium smegmatis, Saccharomyces cerevisiae, and Yarrowia lipolytica. When these genes were expressed in S. cerevisiae, the resulting proteins produced varied free fatty acid pool concentrations (Figure 3). These FAS genes were then expressed in the ∆fas strain and each strain harboring the different FAS genes were shown to have unique growth patterns in YP and SC Media (Figure 4). The difference in growth rate amongst the strains is attributed to the different free fatty acid pool concentration. Further testing on the system incorporated a high-throughput analysis on the Bioscreen C to test for error ranges (Figure 5).

5.2.3 Nile Red Fluorescence Assay

Parental strain BY4741 was grown overnight to saturation and then inoculated to an OD of 0.2 and grown in the presence of 1% and 2% palmitic acid for six hours in YP media at 250 rpm and 30°C. The cells were then stained with Nile red using the Nile Red Lipid Staining kit from Cayman Chemical (Ann Arbor, Michigan), according to the manufacturer’s protocol. The samples were then read on a Tecan plate reader. However, there was no discernable difference in fluorescence between the controls with no exogenous fatty acid addition and the samples with exogenous fatty acids added (Figure 6).
5.3 Discussion

High-throughput screens are very important for interrogating genetically diverse pathways and strains for the phenotype of interest. Researchers are increasingly realizing that high-throughput screen and selection methods are needed to reach a new phase of synthetic biology and metabolic engineering: large-scale combinatorial library engineering. The ability to construct these large libraries are now well established (1,12), but without a high-throughput screen, these methods will not be very effective.

Here we investigated three different strategies to use as a high-throughput screen for the optimization of a heterologous fatty acid ethyl ester synthesis pathway. This pathway was successfully developed in Chapter 4, however pathway optimization is needed. Therefore, it is desired to apply pathway optimization strategies, the same techniques developed in Chapter 2 and 3. However, there is no high-throughput screen for FAEEs. Detection of these molecules requires an arduous 12-hour multi-step extraction and isolation, which is then applied to a GC-MS (24,25). A high-throughput screen to detect FAEEs directly is not likely, therefore the screens studied in this chapter investigated a precursor of FAEEs and a major component of the pathway established in Chapter 4: fatty acids.

5.3.1 Transcriptional Biosensor

In this study the POX1 promoter is sensitive to acyl-CoA motives, therefore any increase in acyl-CoA molecules could trigger the POX1 expression (26). A disadvantage of this is that the POX1 expression has no specificity in distinguishing between acyl-CoA molecules. However, as this strategy would be used for the optimization of the FAS enzyme, the only variation and
diversity introduced to the system would be based on the FAS activity. Therefore, any increase compared to the wild type enzyme would be directly related to the fatty acids produced by the FAS. Work completed by graduate student Tong Si showed that the POX1 promoter had a linear dynamic range from 0 g/L to 1 g/L of exogenous fatty acids (Figure 2). However, as POX1 is the transcriptional control for β-oxidation, the process in which fatty acids are broken down for energy consumption, the promoter is inhibited by glucose (26). Therefore, for the use of POX1 as a biosensor, the strain must use non-fermentable sugars as a carbon source, such as glycerol. Unfortunately, the ws2 gene and activity of the WS2 enzyme has been shown to be greatly affected by culture conditions (see Chapter 4) and also, it has been proven that different substrates can greatly affect the optimization of the pathway (4). Therefore, if pathway optimization strategies will be applied, future work with this biosensor is not suggested. However, overall host engineering strategies to improve fatty acids could potentially find this screening strategy very helpful.

5.3.2 Growth Based Complementation

By knocking-out the cell’s ability to synthesize its own fatty acids, the cells cannot grow without either exogenous supplementation of fatty acids or through complementation of a heterologous gene. In this study, the efficiency of the heterologously expressed FAS could be assessed based on the growth rate of the strain harboring the heterologously expressed FAS. The free fatty acid pool had been quantified to show each enzyme studied produced a different concentration of fatty acids (Figure 3). This difference in fatty acid concentration was exemplified by different growth rates of strains harboring these different fatty acids (Figure 4). A more high-throughput version of this screen was performed through the use of the BioscreenC (Oy Growth Curves Ab
Ltd, Finland) (Figure 5). In this study, the differences in growth were significant for small changes in fatty acid concentration in YP media. However, the BioscreenC readings using SC media were subject to large margins of error. A moving average was used for analysis. If the screen were conducted in YP, there would be enough sensitivity and dynamic range for a controlled screen. However, if the screen were conducted in SC media, the error in the readings would limit the sensitivity of the assay. Further analysis could be conducted by changing settings on the machine to improve readings. Chapter 4 reported that the biodiesel pathway was not functional in YP media and it has been established that promoter engineering is sensitive to the media screened in (10). Though this strategy is promising, further tests in SC media are suggested. Additional experiments could focus on colony-size on agar plates screening, which could be sensitive enough to identify improved fas gene expression.

5.3.3 Nile Red Staining

Using the lipophilic stain was actually more difficult than anticipated. After six hour incubation with palmitic acid, the BY4741 strain did not react to the Nile Red stain (Figure 6). It is assumed in this study that the difficulty lies in the transient nature of exogenous fatty acids. Though the time-frame was suggested by the manufacturer’s protocol, it was still difficult to discern if the exogenously fatty acids were still present in the cell. Continued work with this screen is suggested, especially using intra-cellular fatty acid concentrations, which could be compared using plasmids and strains with known differences in fatty acid contents.
5.4 Future Work

Final product detection will be difficult to develop a high-throughput screen in this FAEE project because FAEEs require GC-MS analysis for quantification. Though, if the above-mentioned strategies are not stringent enough for further optimization and engineering, another strategy is suggested here. One of the major challenges of Chapter 4 and the development of the biodiesel pathway was expression of the *fas* gene. The qPCR analysis illustrated that in samples with high selection of the *fas* gene, there was significantly higher mRNA levels for the *fas* gene than for samples without the high selection (Figure 7). Therefore, another high-throughput strategy is suggested: to increase mRNA levels and consistency of the *fas* gene expression. It has been shown that mRNA stability and relative levels can affect protein expression and overall titer of proteins expressed by the engineered mRNA (27). Reasons for differential mRNA levels can be complicated. However, promoter strength has shown to be a major factor in mRNA levels (28,29). Promoter engineering, either combinatorial or through directed evolution (10,30), could be monitored through mRNA levels. Increased mRNA levels would be indicative of improved expression. The mRNA could be quantified and monitored through a high-throughput Taq-man RT-PCR (31,32). This method has been shown to be reliable and accurate at low mRNA levels. In 384-well plates, a diverse library of either combinatorial promoters or mutated promoters could be screened to identify a pathway with improved expression of the *fas* gene. It is suggested that this be considered as a future potential screen to improve the *fas* gene expression.
5.5 Materials and Methods

5.5.1 Plasmid Construction and Strains

All plasmid construction was accomplished through DNA assembler (33) unless otherwise noted. The fas and pptI genes from Brevibacterium ammoniagenes were cloned from pGM44 (34) under control of the TPI1/TPI1 promoter/terminator and TEF1/TEF1 promoter/terminator respectively (pRS426-FAS). Due to the size and GC composition of the fas gene, we endeavored to reduce PCR amplification of the gene. The GT, MS, SC, YL fas genes were cloned from Gordonia terrae, Mycobacterium smegmatis, Saccharomyces cerevisiae, and Yarrowia lipolytica respectively, by graduate student Sam Hamedi Rad. Each gene was cloned with a TPI1 promoter/terminator and the PPT1 gene from Brevibacterium ammoniagenes, also under the TEF1/TEF1 promoter/terminator in the pRS 426 plasmid.

The BY4741 Δfas was constructed by graduate student Jiazhang Lian. The full fas1 gene was replaced with a KanMX expression cassette, completely disrupting the endogenous fatty acid synthesis. All transformations followed standard Lithium Acetate transformation protocol for S. cerevisiae as described previously (33). However, transformation in the Δfas strain required a 12 hour incubation period after inoculation, supplemented with exogenous 2% palmitic acid and G418 for selection of the knockout. The transformants were recovered in YPD media with 0.01% palmitic acid for five hours. The cells were then transferred to agar plates or liquid media with and without palmitic acid supplementation. Parental BY4741 was used for Nile Red staining and also the POX1 biosensor.
5.5.2 Culture Conditions

Seed cultures of strains for free fatty acid detection and growth curves were grown in 2% glucose CSM-Ura (MP Biomedical, Santa Ana California) or yeast (1%), peptone (2%), dextrose (2%) (YPD) supplemented with 200 µg/mL G418 (KSE Scientific, Durham, NC) to stationary phase at 30°C with orbital shaking at 250 rpm. Seed cultures were then transferred to 125-mL baffled flasks with 25 mL CSM-URA or YPD supplemented with 200 µg/mL G418 (KSE Scientific, Durham, NC) at an initial OD of 0.05 at 30°C with orbital shaking at 100 rpm.

For Nile Red staining, seed cultures of BY4741 were grown overnight in YPD. The cultures were inoculated to an OD(600nm) of 0.2 in fresh YPD media and varied concentrations of palmitic acid for six hours at 250 rpm and 30°C. Culturing to test for the activity of the POX1 promoters involved growing seed cultures of the POX1 biosensor strain in 2% glucose CSM-His (MP Biomedical, Santa Ana California). The cultures were inoculated to an OD (600 nm) of 0.2 in CSM-His with 2% galactose or 2% glycerol and various concentrations of palmitic acid.

*Escherichia coli* 10-β (New England Biolabs, Ipswich, MA) was used for recombinant DNA manipulations. *E. coli* strains were cultured in Luria broth (LB) (Fischer Scientific, Pittsburgh, PA) at 37°C and 250 rpm, supplemented with 50 µg/mL ampicillin. Yeast and bacterial strains were flash frozen and stored in 15% glycerol at -80 °C. All chemicals were purchased from Sigma Aldrich or Fisher Scientific.

5.5.3 Integration of POX-GFP

Integration of the POX-GFP expression cassette into *S. cerevisiae* chromosome was achieved through recombination into the *his* selection marker of BY4741 strain. The p403 plasmid was
linearized by BamHI and ClaI at the multiple cloning site and the POX-GFP cassette was inserted into the plasmid through Gibson Assembly (New England Biolabs, Ipswich, MA). For integration, this plasmid was linearized in the his3 selection marker through digest with HindIII. 5 µg of linearized plasmid was transformed into the BY4741 strain using the Lithium Acetate method. The transformed cells were plated on SC-his plates and colonies which grew were tested by colony PCR to confirm the integration.

5.5.4 Free Fatty Acid Extraction

Quenching was conducted on the cells to quantify the free fatty acid pool (35). Cells were flash frozen by a methanol-dry ice bath. The mixture was centrifuged at 4000 rpm for 5 minutes and then the cell pellet was separated from the media. Cells pellets were then lysed by boiling at 100°C in ethanol for 10 minutes. Cell membrane was separated by centrifugation and the supernatant was applied to a freeze-dryer overnight or until dry. The lipids were then extracted twice with 2:1 chloroform/methanol using 50 µM myristic acid as an internal standard. After extraction, the fatty acids were allowed to derivatize for 2 hours at room temperature with 200 µL of 2N TMS-diazomethane, 10 µL methanol and 10 µL HCL. The sample was then transferred to a glass GC vial for analysis.

5.5.5 Free fatty acid detection

The fatty acids were separated and quantified through a GC-MS-QP 2010 Plus (Shimadzu, Kyoto Japan) (36). The separation was performed by a DB-WAX GC column with the following dimensions: 30mm x 0.25 mm internal diameter, 0.25 µm film thickness, from Agilent Technologies (Santa Clara, CA). A 1 µL portion was injected via splitless injection at 250 °C at
an initial pressure of 98 kPa and a total flow of 14.1 mL/min helium carrier gas. The chromatography separation initially occurred at a temperature of 40°C for 1.2 minutes and then increased at a rate of 30 °C per minute to 220 °C and then held for fifteen minutes. The mass transfer line and ion source were at 250 and 200°C respectively. The fatty acids were detected with an electron ionization method in scan mode from 50 to 650 m/z.

5.5.6 Nile Red Staining
Strains to be tested through Nile red staining were inoculated to an initial OD of 0.2 and grown at 30 °C at 250 rpm orbital shaking until the desired time-points. At inoculation, 1% and 2% exogenous palmitic acid was added as needed. Cells were normalized to an OD of 20 and were aliquoted into black 96-well UV-plates flat bottom plates (Corning, Corning, NY). Cells were washed twice with Potassium Phosphate Buffer 100mM, as supplied by the Caymen Chemical (Ann Arbor, Michigan) Nile Red Staining Kit. Then a fixative agent was also used to wash the cells per the protocol. Finally, the cells were stained with Nile Red staining solution, per manufacturer’s instructions for 5 minutes. After washing with Potassium Phosphate Buffer, the cells were then visualized on a 96-well plate reader from Tecan (Mannendorf, Germanyh) at an excitation of 485 and emission of 535 nm.

5.5.7 FACS
Integrated strains of the POX1 promoter were inoculated to an initial OD of 0.2 and grown at 30°C at 100 rpm orbital shaking until the cells reached OD 15-20. The cells were then diluted in PBS to an OD of about 0.1-0.2 and were analyzed on a flow cytometer (Biosciences LSR II Flow
Cytometry Analyzer, BD Biosciences, San Jose, CA) to quantify GFP fluorescence, with excitation wavelengths at 488 nm and emission band pass 550/30.
5.6 References


the two type-I fatty acid synthases of *Brevibacterium ammoniagenes*. *Eur J Biochem*, **248**, 481-487.


5.7 Figures

Figure 5.1: Confirmation of correct integration of the POX1-GFP expression cassette into chromosome of BY4741 using gel electrophoresis. Expression cassette is 1.8 kilobase (kb) and the PCR amplification yielded a band at 1.8 kb according to the 1kb Plus TrackIt ladder (Invitrogen). The first three samples are the BY4741 integration with the POX1-GFP, PCR conducted from whole-cells. The final sample is a control, which used the integration plasmid as a template for PCR. This strain was used as a biosensor for increased fatty acid production, as the POX1 promoter is sensitive to acyl-CoA molecules such as fatty acids.
Figure 5.2: Dynamic range of the POX1-GFP fluorescence with respect to exogenous fatty acids. Data collected by graduate student Tong Si. POX1 is a promoter sensitive to fatty acids and was shown to activate expression of a GFP gene in the presence of exogenously added fatty acids. Though this strategy is promising, the POX1 promoter is inhibited by glucose and is only effective in non-fermentable sugar substrates. Ultimately, the FAEE pathway will be produced with glucose as the substrate and optimizing the pathway in another sugar substrate will not transfer to the glucose system (4).
Figure 5.3: Free fatty acid concentration in the $\Delta f_{as}$ BY4741 strain complemented with heterologous FAS genes as measured by GC-MS analysis at 100 hours in YPD media. Sample BA is the fatty acid synthase cloned from *Brevibacterium ammoniagenes*. Sample GT is the fatty acid synthase cloned from *Gordonia terrae*. Sample YL is the fatty acid synthase cloned from *Yarrowia lipolytica*. Data is the result of one biological replicate.
Figure 5.4: Growth curve of strains deficient in endogenous fatty acid synthesis complemented with heterologous FAS expression in (a) YPD media and (b) SC media. The BA, GT, MS, SC, YL FAS genes were cloned from *Brevibacterium ammoniagenes*, *Gordonia terrae*, *Mycobacterium smegmatis*, *Saccharomyces cerevisiae* and *Yarrowia lipolytica* respectively. The growth profile is indicative of fatty acid availability within the cell. Strains harboring more efficient FAS enzymes, which produce increased fatty acids (Figure 3), will grow faster than strains with inefficient FAS enzymes. A growth curve based assay is beneficial to consider for optimization of the *fas* gene either through protein engineering or promoter engineering.
Figure 5.5: Growth curve profiles of strains deficient in endogenous fatty acid synthesis complemented with heterologous FAS expression completed in a high throughput manner on the BioscreenC. Data is the average of twenty samples from one biological replicate for each fas gene. The wt is parental strain BY4741 with no knockout of the fatty acid synthesis. The BA, SC, YL FAS genes were cloned from *Brevibacterium ammoniagenes*, *Saccharomyces cerevisiae*, and *Yarrowia lipolytica* respectively. The margin of error was very low if screening in YP, however use of SC media increased error significantly. Data for the SC media screen is based on a moving average analysis for the profile. Sensitivity becomes a problem if screening in SC media, as promoter engineering could introduce small changes in expression which would not be distinguished by this screen.
**Figure 5.6:** Nile Red staining with exogenously supplied fatty acids with BY4741 strain. Cells were incubated with the fatty acid for six hours, per manufacturer’s protocol and then were stained with Nile Red for five minutes. Samples were run on a plate reader with an excitation at 485 nm and emission at 535 nm. No significant differences in fluorescence reading were found in comparison to the negative. Continued work with this strain is suggested with strain containing *in vivo* differences in fatty acid concentrations.
Figure 5.7: Relative expression or mRNA levels of the *bafas* gene under different selection pressures. mRNA levels of the *bafas* gene in a highly selective system were extremely high, while the *bafas* gene expression using the *HXT5* promoter in the wild-type system had very low mRNA levels. It has been shown that mRNA stability and relative levels can affect protein expression and overall abundance of proteins expressed by the engineered mRNA (27). Increasing the mRNA levels could improve the *bafas* gene expression and overall improve the FAEE production. It is suggested to use mRNA levels as a high-throughput screen to improve the *bafas* gene expression.