The Dissertation Committee for John James Blazeck Certifies that this is the approved version of the following dissertation:

Harnessing *Yarrowia lipolytica's* Potential as a Lipid and Alkane Production Platform

Committee:

Hal S. Alper, Supervisor

Lydia Contreras

Andrew Ellington

George Georgiou

Jennifer Maynard

Harnessing *Yarrowia lipolytica's* Potential as a Lipid and Alkane Production Platform

by

John James Blazeck, B.S.Ch.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2013

Dedication

To Kristina and Rowan
I'll eat you up – I love you so.

Acknowledgements

I would like to first thank my advisor, Dr. Hal Alper. Hal's support and advice have been major reasons for the success of the studies described herein, and I thank him for all he has done for me. He has compassionately supported me through injury and illness. His excitement and determination have allowed us to develop and test more ideas and hypotheses than I would have thought possible during five years time. His knowledge has guided me and has given me every chance to succeed. And his sense of humor has made my entire graduate school experience more enjoyable. Hal has also paid for me to attend and present my work at multiple conferences, and his recent recommendation has helped grant me the opportunity to present at a Distinguished Young Scholar's seminar series.

I would also like to thank the members of my thesis committee, who have all given me thoughtful advice and guidance. First, I want to thank George Georgiou for offering me the opportunity to work in his laboratory as a postdoctoral scholar. I am very excited to join his lab soon. I also want to thank Lydia Contreras, Andrew Ellington, and Jennifer Maynard for their support and for all of the discussions that we have had about my research and my future. These opportunities to talk have been invaluable to me.

Many other people have helped my research succeed or made my experience here at UT more enjoyable. I want to thank all the members of the Alper Laboratory that I have had the pleasure to work alongside and collaborate with, especially the following:

Leqian Liu for helping me tame *Y. lipolytica* into something we could actually work with – there were weeks, months, and even years where it didn't seem like we'd ever get our yeast to do what it should be doing; Eric Young for all of the late night jam sessions and discussions; Andrew Hill for help with the lipid accumulation work; Amanda Lanza and Nathan Crook for help with the GCN5p work; Heidi Redden and Joseph Cheng for early help with experiments; and the rest of the Alper Lab for their input and help throughout my thesis – including Kate Curran, Sun-Mi Li, John Leavitt, Joe Abatemarco, and Jie Sun. I have also had the pleasure of mentoring and instructing numerous undergraduate assistants. I want to deeply thank Jarrett Miller, Annie Pan, Rishi Garg, Ben Reed, Mariam Jamoussi, Rachelle Gerstner, Clinton Holden, and Vaibhav Argawala for their hard work and dedication towards completing many difficult experiments. Their help is a major reason for the success of all of my projects. Finally, I want to thank other UT members that helped contribute towards the success of my experiments, including Mauro Lorenzo, Richard Salinas, and Jon Gengler.

I have been fortunate to develop close friendships while in Austin, and I want to thank Derek Hernandez, Ramiro Palma, Alex Voice, Ryan Nagao, and Avi Murthay for all of the good times we've had. I also want to thank all the members of my incoming Ph.D. class, too numerous to list, whose company I have enjoyed and whose friendship I will miss. My favorite hobby during graduate school has been playing intramural sports, and our chemical engineering graduate teams have been fortunate enough to win six championships across three leagues. I want to thank all the members of our flag football team (Touchdown! M.P.), basketball teams (JT-Tip Drill and Occupy Ball Street), and

softball team (Suns Out Guns Out) that I played with over four seasons. In order of number of games we've played together, I want to thank Derek Hernandez, Ryan Nagao, James Knight, Justin Harris, Doug French, Joe Dekker, Alex Voice, Chet Steinhagen, D.J. McDaniel, Richard Pattison, Bill Liechty, Tom Murphy, David Sanders, Ramiro Palma, David Kyrscio, Jason Cantor, and Marty Gran.

Most importantly, I want to thank my family for their love and support. I want to thank my loving parents, James and Alice, for giving me everything I ever needed to succeed, for giving me the best childhood I could imagine, and for being great grandparents. I want to thank my younger sister, Elizabeth, for always working harder and being more determined than anyone I have ever met, and giving me someone to look up to in the process. I want to thank my older brother, Joseph, for being my best friend for as long as I can remember. I'm sure you've been annoyed by me sometimes, but I've never been by you. I want to thank my son, Rowan, for giving me joy every second I see him. I will always love everything about you. And finally, I want to thank Kristina Ramirez for making the past four years the best of my life. I love you, and thank you for supporting me, for laughing with me, and for caring for me. I am lucky to have met you, and luckier to have kept you.

Harnessing *Yarrowia lipolytica's* Potential as a Lipid and Alkane Production Platform

John James Blazeck, Ph.D.

The University of Texas at Austin, 2013

Supervisor: Hal S. Alper

Engineering cellular phenotype can enable the *in vivo* synthesis of renewable fuels, industrial precursors, and pharmaceuticals. Achieving economic viability requires the use of a cellular platform that generates high titers independent of fermentation condition, through either native or imported biosynthetic metabolism. While lacking fully developed genetic tools, the oleaginous yeast *Yarrowia lipolytica* has the native capacity to produce large titers of lipids and citric acid cycle intermediates. However, unlocking this biosynthetic capacity requires complete rewiring of native metabolism. To this end, this work focuses on the development and engineering of the yeast *Y. lipolytica* to rewire native metabolism and enable the production of lipids, alkanes, and itaconic acid.

Precise control of gene expression is a requisite to enable metabolic and pathway engineering applications for any host organism. However, *Y. lipolytica* lacks promoter elements strong enough to manipulate intracellular metabolism. Thus, we utilized a hybrid promoter engineering approach to produce libraries of high-expressing, tunable promoters, seven-fold stronger than promoters previously characterized in *Y. lipolytica*

[1,2]. We successfully applied this approach to *Saccharomyces cerevisiae*, expanding transcriptional capacity of the strongest constitutive to highlight our hybrid approach as a generalizable method to increase expression capacity in eukaryotic organisms [3].

We utilized our novel *Y. lipolytica* hybrid promoters to drive intracellular metabolism towards lipid production and to overexpress heterologous enzymes that enable alkane and itaconic acid production. Specifically, we implemented a global rewiring of *Y. lipolytica's* native metabolism to increase lipogenesis more than sixty fold to 25.3g/L (the highest lipid production ever reported) and generated cells nearly 90% lipid content. We further expressed a lipoxygenase enzyme to catalyze the novel microbial production of the short-chain *n*-alkane, pentane. Finally, we exploited *Y. lipolytica's* capacity to accumulate citric acid cycle intermediates by expressing a heterologous cis-aconitic acid decarboxylase enzyme to produce itaconic acid. Increasing substrate availability through media optimization and genomic engineering increased pentane and itaconic acid production threefold and eightfold, respectively [4]. Collectively, these studies have facilitated the utilization of *Y. lipolytica* as an industrially relevant microbial platform, and represent a generic approach towards enabling biosynthetic control in microbial hosts will ill-defined gene expression technology.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction and Background

1.1 Metabolic Engineering of Microbial hosts

Metabolic engineering utilizes the expression of recombinant DNA to divert metabolic flux towards a desired product. In this manner, metabolic engineering has enabled large-scale, cell-based production of pharmaceuticals, fuels, and chemicals from renewable resources. Traditionally, model microbial hosts such as *Escherichia coli* or *Saccharomyces cerevisiae* have been utilized for many metabolic engineering applications. However, recent widespread advances in recombinant technologies have enabled basic engineering of a much wider variety of microorganisms. These organisms often have unique biosynthetic capacities not found in standard *E. coli* and *S. cerevisiae* hosts. Genetic tools have been developed for organisms able to accumulate organic acids [5], amino acids [6], butanol [7], and many other valuable compounds. However, these genetic tools are often limited, especially in their ability to enable precise control of metabolic pathways. In particular, the non-conventional yeast, *Yarrowia lipolytica*, has the innate ability to produce large quantities of lipid content [8] and citric acid [9], but has rudimentary gene expression technologies.

Manipulating metabolic flux requires the capacity to precisely control recombinant gene expression level, a capacity that is often non-existent or severely limited in non-conventional microbial hosts. Thus, a question is often raised before undertaking metabolic engineering endeavors – will it be easier to maximize yields in a non-conventional host with innate, beneficial metabolic abilities, or in a conventional *E.*

coli or *S. cerevisiae* host [10]? This dissertation demonstrates a reproducible methodology towards harnessing biosynthetic capacity in non-conventional hosts, by first developing a generalizable promoter engineering approach to permit precise gene expression control (Chapters 2,3, and 4), and then by using this gene expression control to manipulate *Y. lipolytica's* metabolism for alkane (Chapter 5), lipid (Chapter 6), and itaconic acid production (Chapter 7). Chapter 8 provides a brief review of our major findings and recommendations for future studies, and Chapter 9 contains comprehensive materials and methods descriptions.

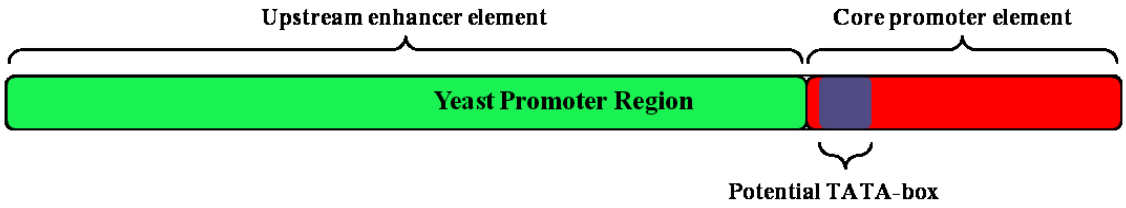
1.2 Controlling Gene Expression With Promoter Engineering

Intracellular metabolic flux is regulated by a series of distinct yet interwoven levels of regulatory control—occurring at the transcriptional, translational, and protein levels [11]. Achieving flux control on the metabolomic level requires a concurrent capacity to directly manipulate gene expression. Gene expression is controlled at its most fundamental level by promoter elements that drive transcription. Hence, metabolic engineering applications have long relied on endogenous promoter discovery and use to control gene expression. However, these promoters are limited in that they do not permit tunable transcriptional control (without differential regulation), and they do not maximize the transcription levels achievable within an organism. To address this issue, the field of promoter engineering attempts to modulate promoter transcriptional capacity by mutating, enhancing, or otherwise altering promoter DNA sequence. In doing so, promoter engineering can help generate the dynamic range necessary to enable fine-tuned gene expression to maximize production formation for metabolic engineering

applications. What follows is a brief overview of eukaryotic promoter structure and function and of previous promoter engineering attempts to enable transcriptional control. As yeast microbial hosts are utilized for all experiments performed in this dissertation, yeast promoters are specifically highlighted.

1.2.1 Overview of eukaryotic/yeast promoter structure and function

Figure 1.1: Overview of promoter structure



A schematic of yeast promoter architecture. Yeast promoters have very few conserved motifs. Instead, it is more applicable to view them as modular enhancer and core elements.

Eukaryotic transcription initiation is complex; requiring DNA sequence-specific transcription factors to bind within a promoter element and interact with transcriptional coactivators that help localize the basic transcriptional machinery [12]. Routinely, eukaryotic promoters are thought to contain two distinct regions: (**1**) a core element and (**2**) an upstream enhancer element. Transcriptional direction and start site are determined by the core promoter element while the upstream enhancer element helps determine transcriptional frequency, or promoter strength. The core promoter is the minimal promoter region required to allow formation of the pre-initiation complex (PIC) and to initiate transcription [12-14], and is typically a succinct stretch of less than 80 nucleotides extending roughly 35 bp upstream or downstream from the transcriptional start site (TSS)

[14,15]. In yeast, little is known about potential core promoter consensus motifs and only ~20% of genes contain a TATA-box (**Figure 1.1**) [16]. As the core promoter's function is to enable basal transcription, a suitable TSS and the capacity to recruit PIC components are equally essential.

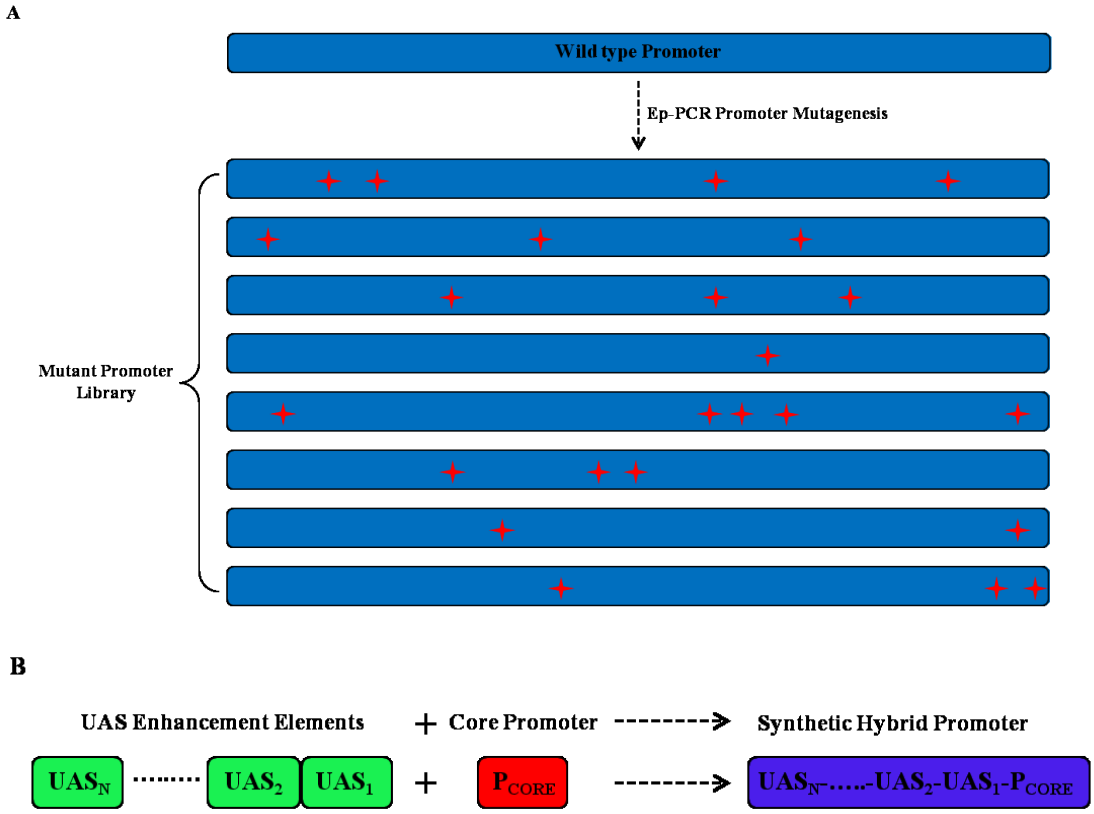
Eukaryotic upstream enhancer elements localize trans-acting regulatory elements (transcription factors) as a means of controlling transcriptional frequency or imparting regulation to a core promoter. Within the enhancer element, concise and specific DNA sequences serve as transcription factor binding sites or "docking points" for transcriptional activators or repressors [17-22]. Promoter-bound transcription factors interact with one another locally and with the basal transcriptional machinery to establish promoter regulation and promoter strength, i.e., frequency of PIC formation and subsequent transcription initiation [23-25]. DNA regions prone to increase transcriptional frequency of a core promoter are commonly referred to as upstream activating sequences (UAS). Similarly, an upstream repressive sequence localizes transcription factors that reduce transcription rate [12]. Promoter regulation (induction or repression dependent on environmental conditions) is also a result of transcription factor mediated interactions in the enhancer element [26-29]. Promoter engineering techniques alter both core and enhancer elements to modulate overall promoter expression capacity.

1.2.2 Promoter engineering strategies

Promoter engineering has become an enabling technology for metabolic engineering purposes by facilitating the creation of novel gene expression control elements. The basis for much of this engineering is related to promoter architecture.

Visualizing eukaryotic promoters as a core-enhancer fusion facilitates engineering hybrid promoters, in which modular enhancer and core combinations can determine promoter regulation and transcriptional capacity. At the more basic level, both enhancer and core elements possess specific transcription factor binding sites that determine overall promoter function. Randomized promoter mutagenesis through error-prone PCR alters transcription factor binding sites thus altering promoter strength. Since binding site mutations are far more likely to reduce transcription factor interactions, these random mutagenesis approaches often produce promoter variants with lower strengths than the template sequence. In fact, the majority of promoter engineering methods, while successful at producing libraries with large ranges in gene expression, have been unable to drastically increase promoter strength.

Figure 1.2: Promoter engineering strategies



(**A**) Ep-PCR introduces random mutations (depicted as red stars) into a wildtype promoter element that alter promoter sequence, adapted from [30]. Large-scale characterization of mutated promoters facilitates isolation of a promoter library that retains endogenous regulation but spans large expression ranges. (**B**) Hybrid promoter engineering utilizes tandem upstream activations sequences to modulate core promoter expression to construct synthetic hybrid promoters with novel strength or regulation, adapted from [3].

1.2.2.1 Error-prone PCR

Error-prone PCR (ep-PCR) introduces random mutations into a DNA sequence of choice (**Figure 1.2a**). When applied to an entire promoter region, mutations occur throughout the consensus and spacer regions and lead to disparate function. This approach is guaranteed to yield novel promoter variants (with sufficient library sizes) and proper selection techniques allow the isolation of promoters driving a wide variation in gene expression. For example, mutagenesis of the bacteriophage-derived P_L-λ

constitutive *E. coli* promoter yielded a library of engineered promoters of varying strengths spanning a 196-fold range with identical regulation [30]. Ep-PCR of the strong constitutive *S. cerevisiae TEF1* promoter generated a similarly diversified promoter library spanning a 15-fold range [30,31]. Finally, random mutagenesis of the oxygen-responsive *S. cerevisiae DAN1* promoter enabled isolation of two mutants induced under less-stringent anaerobiosis than the wild-type promoter [32].

1.2.2.2 Hybrid promoter engineering

A hybrid promoter engineering approach entails the assembly of enhancer element-core promoter fusions to rationally enhance basal core transcriptional capacity or enable novel promoter regulation (**Figure 1.2b**). Basic hybrid promoter work in *E. coli* led to the formation of many commonly utilized promoters, including the *tac* promoter (a fusion derived from the *trp* and *lac* promoters) [33], and the $rhaP_{BAD}$, a tightly regulated arabinose and rhamnose promoter fusion [34]. In yeast, hybrid promoters have traditionally been utilized to dissect promoter function and regulation [33,35-40]. In this light, essential DNA sequences are identified in part due to the modularity of hybrid promoter core and enhancer elements. Upstream enhancer elements contain transcription factor binding sites that enable native regulation and expression activation or repression to be maintained independent of core promoter region. Minimal regions of these DNA regions can identify specific regions essential for transcriptional activation or induction control (i.e., upstream activation sequences). Constructing hybrid promoters composed of tandem repeating UAS elements can theoretically greatly increase core promoter expression capacity, as each additional UAS increases overall hybrid promoter strength.

Chapters 2 – 4 of this dissertation are devoted towards the development of the hybrid promoter engineering approach as a unique tool to increase promoter expression capacity, especially in nonconventional organisms that lack the capacity for strong, tunable gene expression.

1.3 *YARROWIA LIPOLYTICA* AS A METABOLIC ENGINEERING HOST

Yarrowia lipolytica is a widely studied dimorphic yeast with many intriguing metabolic capabilities [41-44], and is a unique host for biochemical production and heterologous protein excretion due to its ability to secrete native and heterologous proteins at high levels [45-47]. Lipase, leucine amino peptidase, hydroperoxide lyase, glucoamylase, alpha-amylase, galactanase, polygalacturonase, cellulase, xylanase, laccase, prochymosin, urokinase-type plasminogen activator, anti-RAS single chain variable fragment, and many other fungal, plant, and mammalian proteins have been successfully expressed and secreted from *Y. lipolytica* [45,47-53]. *Y. lipolytica's* genome consists of six chromosomes, totaling 20.5 million base pairs in length with an average GC content of 49% [54]. *Y. lipolytica* is also very uncommon in its ability utilize hydrophobic and waste carbon sources [55-57]. In particular, *Y. lipolytica* is well-known for its ability to grow robustly on pure and crude glycerol. *Y. lipolytica* is also an industrial producer of the TCA cycle intermediate, citric acid [58], and this tendency to produce citric acid is directly linked to its ability to accumulate lipids [59].

Yarrowia lipolytica is an oleaginous organism, meaning that it can accumulate high levels of lipid content, exceeding 40-70% dry cell weight when cultivated in fatty

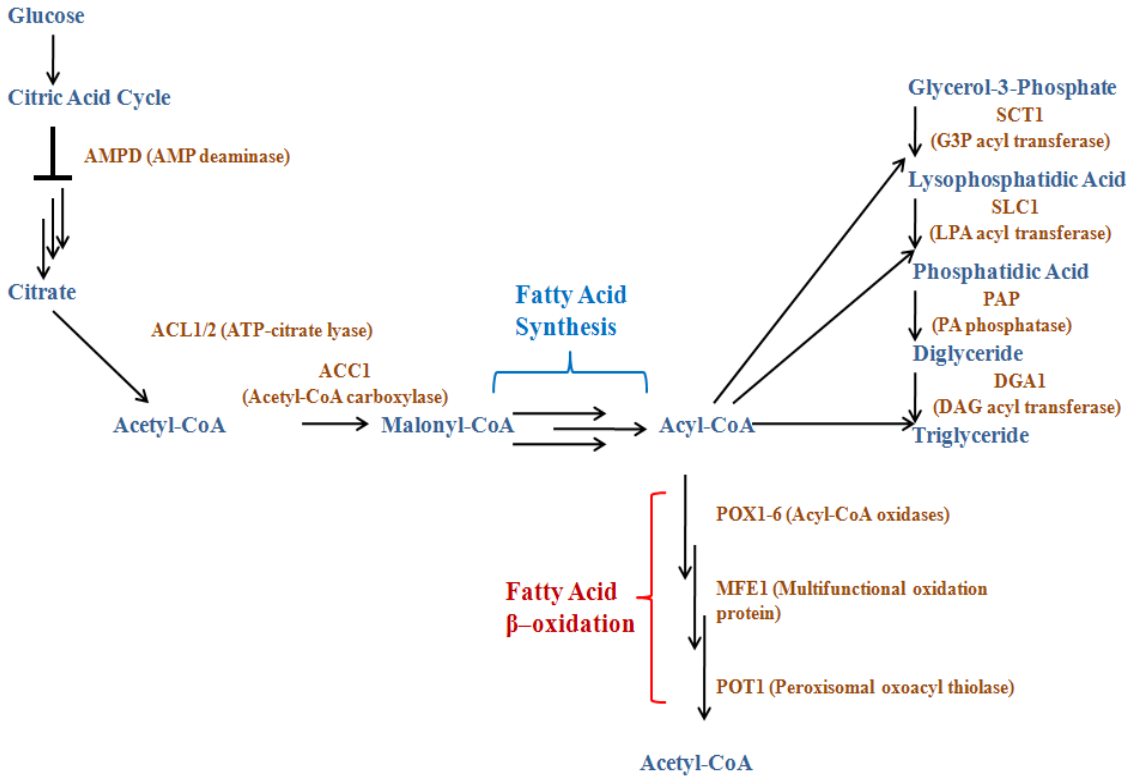
acid containing media [60-65]. The lipid accumulation phenotype is induced by nitrogen starvation conditions that slow cell growth and production. However, the availability of *Y. lipolytica*'s genome sequence [66] along with basic genetic tools such as transformation methods [67,68], gene knockouts [69], and both episomal [70-73] and integrative expression cassettes [74-76] largely enables metabolic engineering approaches. As the only oleaginous organism with developed genetic tools, *Y. lipolytica* represents an ideal host in which to optimize metabolism to enable lipid accumulation across multiple environmental conditions and regardless of carbon source. *In vivo* conversion of lipid content could permit production of oleo-chemicals such as fatty acid methyl esters and alkanes [77-80]. In fact, *Y. lipolytica's* lipid reserves have proven amenable to further *in vivo* manipulation, making the implementation of previously characterized alkane production pathways more promising [77,79,81,82]. Additionally, *Y. lipolytica's* citric acid pool can be exploited for the production of non-native organic acids, namely itaconic acid [83]. However, as stated above, assuming this control of cellular metabolism requires precise control of key gene expression, an area in which *Y.* genetic toolbox is severely underdeveloped [56].

1.4 Fatty Acid and Lipid Metabolism in *Y. lipolytica*

Lipid accumulation in *Y. lipolytica* can be induced by nitrogen starvation and has been associated with the activity of four enzymes: AMP Deaminase (AMPDp), ATP-Citrate Lyase (ACLp), Malic Enzyme (MAEp) and Acetyl-CoA Carboxylase (ACCp) [59,61] (**Figure 1.3**). AMPDp is activated in oleaginous organisms during nitrogen starvation and cleaves AMP into inosine 5'-monophosphate and NH_4^+ to replenish

intracellular nitrogen levels; Low levels of AMP inhibit the isocitrate dehydrogenase enzyme in the citric acid cycle, causing a buildup of isocitrate. Aconitase reversible interconverts isocitric acid and citric acid, but equilibrium concentrations favor high citric acid levels. This citric acid is shuttled out of the mitochondria by the citrate/malate shunt[59,84,85], where ACLp cleaves it into oxaloacetate and acetyl-CoA. Finally, ACCp is activated by high cellular citrate levels [59] and carboxylates acetyl-CoA into malonyl-CoA fatty acid building blocks [86]. Fatty acid synthesis is further encouraged by a MEAp-mediated increase in NADPH levels [59].

Figure 1.3: Lipid metabolism in *Yarrowia lipolytica*



The mechanism for lipid accumulation is diagrammed summarily. In brief, AMPD activity inhibits citric acid cycle under nitrogen starvation condition, causing a buildup of citric acid. Citric acid is cleaved by ACL to form acetyl-CoA, which is further carboxylated by ACC to produce malonyl-CoA fatty acid building blocks. Fatty acids are incorporated in lipids following the Kennedy Pathway in which three acyl-CoA molecules are bound to a single glycerol molecule. Fatty acids are degraded peroxisomally through the four step β-oxidation pathway to replenish acetyl-CoA pools.

Fatty acids can be directly stored in intracellular lipid bodies, though they are typically incorporated into triacylglycerides (lipids) before storage [63]. Lipid synthesis follows the Kennedy Pathway to fuse three fatty acids to a glycerol-3-phosphate (G3P) backbone [87]. The ultimate step of lipid synthesis is catalyzed by the DGA1 or DGA2 acyl-CoA:diacylglycerol acyltransferases [59,88]. G3P backbone is synthesized from dihydroxyacetone phosphate (DHAP) by the cytosolic, NAD^+-dependent glycerol-3-phosphate dehydrogenase (GPD1) and recycled into glycolysis by the mitochondrial, FAD^+-dependent glycerol-3-phosphate dehydrogenase isoform (GUT2) [61]. *Y. lipolytica* also efficiently incorporates extracellular fatty acids, oils and triglycerides by hydrolysis of the hydrophobic substance, transport into the cell, and assembly into triglycerides [59]. A small fraction of free fatty acids are shunted into production of steryl esters, by acylation of a sterol [88]. Lipids, free fatty acids, and steryl-esters accumulate in lipid bodies, where they are stored as energy sources [63].

Lipid hydrolysis mobilizes free fatty acids for peroxisomal degradation through the four step β–oxidation cycle [89] - oxidation by one of six acyl-CoA oxidases (POX1-6), hydration and dehydrogenation by the multifunctional enzyme (MFE1), and thiolysis by a 3-ketoacyl-CoA-thiolase (POT1 or PAT1) [59]. In comparison, *S. cerevisiae* contains only one POX homolog, as it is not known for its ability to accumulate hydrophobic compounds. The PEX10p transcription factor has been implicated in peroxisomal biogenesis and Δ*pex10* mutants display increased lipid content [4,90,91].

Genomic modifications to *Y. lipolytica's* fatty acid, lipid, and central carbon metabolism have shown promise towards increasing lipid accumulation capacity.

Deletion of the six POX genes increased *ex novo* incorporation of oleic acid in *Y. lipolytica*, while deletion of the single MFE1 gene had a similar effect [61,63]. Increasing G3P backbone levels by combining GUT2p deletion and GPD1p overexpression in these β–oxidation deficient backgrounds further increased *ex novo* lipid accumulation to 65-75% lipid content [61]. Overexpression of DGA1p increased *de novo* lipid accumulation fourfold over control levels to 33.8% lipid content, and co-overexpression of ACC1p further increased lipid accumulation slightly to a final yield of 41% lipid content [92]. To date, no study has attempted to completely redirect metabolic flux towards lipid accumulation, by simultaneously manipulating fatty acid, lipid and central metabolism. Furthermore, no study has attempted to utilize *Y. lipolytica* lipid pools for the production of value-added hydrophobic chemicals, like alkanes. Chapters 5 and 6 of this dissertation are devoted to enabling *Y. lipolytica* as an alkane and lipid production platform. Finally, Chapter 7 utilizes *Y. lipolytica's* fatty acid precursor pool to produce itaconic acid.

Chapter 2: Hybrid Promoter Engineering in *Yarrowia lipolytica*

2.1 Chapter Summary

In poorly characterized or nonconventional hosts, a common limitation preventing metabolic engineering applications is a lack of strong, tunable promoter elements to efficiently control gene expression. In particular, *Yarrowia lipolytica* is an intriguing oleaginous organism with developed recombinant DNA technology that is limited in its applications by an inability to manipulate and overexpress enzymatic pathways. Here, we have utilized a hybrid promoter approach to produce libraries of high-expressing, tunable promoters in *Y. lipolytica*. These synthetic promoters are comprised of two modular components—the enhancer element and the core promoter element. By exploiting this basic promoter architecture, we have overcome native expression limitations and provided a strategy for both increasing native promoter capacity and producing libraries for tunable gene expression. In doing so, we have created the strongest promoters ever reported in *Y. lipolytica*. These promoters exhibited a range of more than 400 fold in terms of mRNA levels, and up to eight-fold higher fluorescence levels compared with typically used endogenous promoters. These promoter libraries represent an enabling technology for metabolic engineering applications in *Y. lipolytica*, and we have illustrated that tandem copies of upstream activating regions can serve as synthetic transcriptional amplifiers that may be generically used to increase the expression level of promoters. Thus, we have opened the possibility that the hybrid promoter engineering approach is a generalizable approach applicable to other organisms.

2.2 INTRODUCTION

Developing and establishing a comprehensive suite of promoter elements in organisms with poorly defined genetic tools is essential for enabling metabolic and pathway engineering applications. To this end, the oleaginous yeast *Yarrowia lipolytica* has received attention as a potential biofuels producing host, yet lacks genetic tools for tunable and high level gene expression. In other organisms, prior attempts of promoter engineering relied on modifying the expression range of endogenous promoters through point mutations. As an example, error-prone PCR of the native *Saccharomyces cerevisiae* TEF promoter yielded a library of mutant promoters with a nearly seventeen-fold range in relative expression strength [31]. A similar approach in *E. coli* led to a nearly eighteen-fold range in relative expression strength based on the heterologous Pl-lambda promoter [30]. An alternative strategy aimed at tuning intergenic regions in heterologous *Escherichia coli* operons enabled a seven-fold improvement in pathway throughput and was used to prevent the accumulation of a toxic intermediate [93]. While successful, these methods and applications are quite limited by: (1) their reliance on a strong, well-defined starting core promoter and (2) the tendency of alterations to result in expression levels lower than these baselines. These limitations present a challenge for creating a dynamic range of promoters with similar regulation in a novel organism without highly characterized strong promoter elements.

Beyond point mutations, the generation of hybrid promoters has been previously successful in significantly augmenting promoter architecture and function [33,94,95]. Most of these approaches attempt to create synthetic hybrid promoters by fusing an upstream

activating sequence (UAS) to a separate (often minimal) core promoter region. The result is a functional UAS-core promoter chimera. In this construct, the UAS regulatory site enhances gene expression by localizing trans-acting regulatory elements (transcription factors). As a result, this approach raises the possibility to uniquely engineer promoters as two independent, synthetic parts—activating regions and core regions. We developed a generalizable approach of hybrid promoter engineering for the construction of both very high-strength promoters and tunable promoter libraries in the non-conventional, oleaginous yeast host *Y. lipolytica*.

Y. lipolytica is a unique host for biochemical production and heterologous protein excretion [47,62] that largely enable metabolic engineering approaches [54,69,72,74]. However, it lacks the defined promoter elements to enable high strength enzyme overexpressions [56]. One of the strongest promoters in *Y. lipolytica*, the XPR2 promoter has complex requirements for induction that hinders its industrial applications [53]. Nevertheless, this promoter has been functionally analyzed to reveal a 105 basepair distal UAS fragment named UAS1B [96,97]. Previously, between one to four tandem UAS1B copies were fused to a core minimal LEU2 promoter to create four increasingly strong hybrid chimera promoters, named hp1d through hp4d [53]. As a result, the hp4d promoter has become a commonly used tool for heterologous protein expression in *Y. lipolytica*. A further re-analysis of these four promoters revealed a linear increase of promoter strength as a function of number of tandem UAS1B elements which strongly raises the possibility of further improvements.

2.3 RESULTS AND DISCUSSION

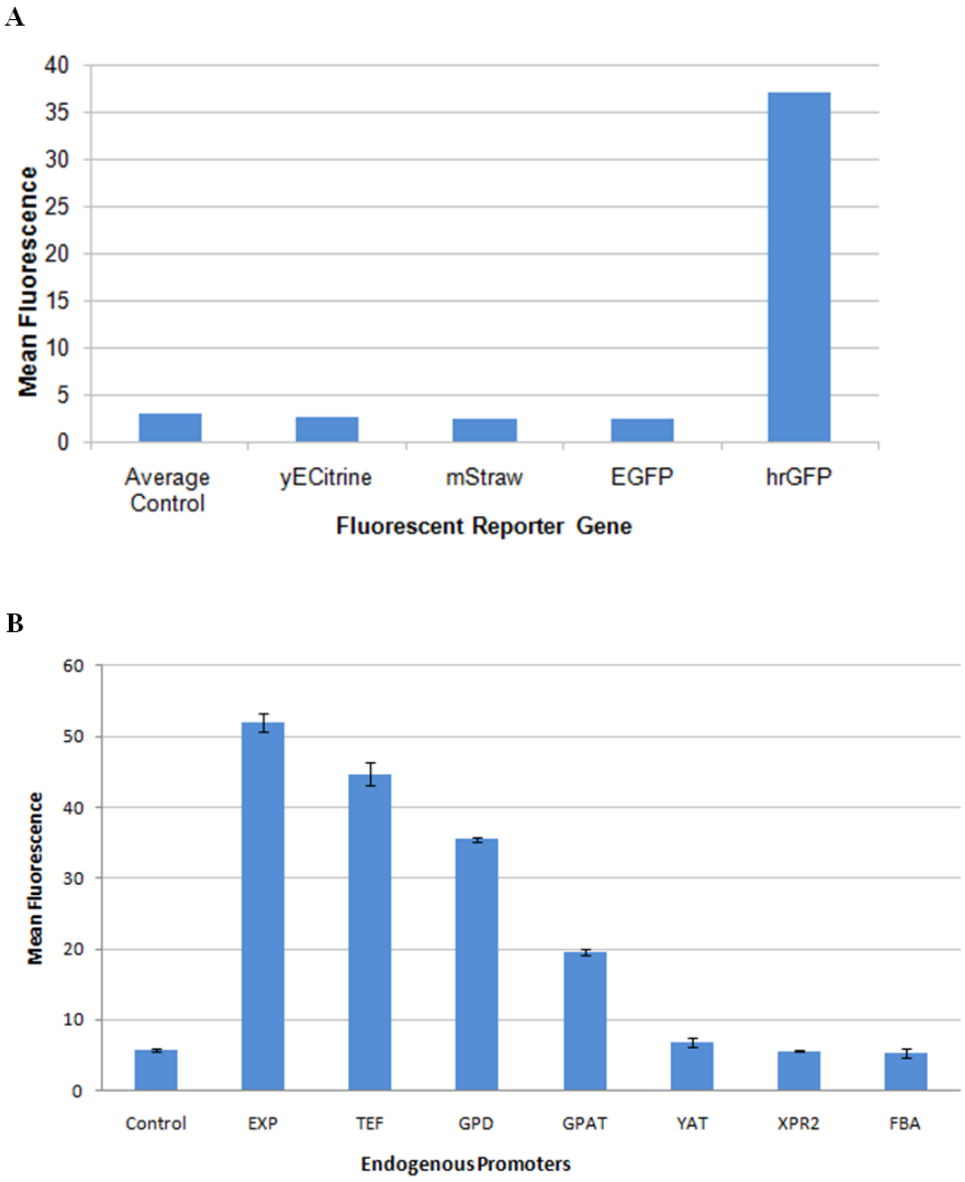
2.3.1 Characterization of endogenous promoters at the single-cell level

Prior studies of promoter strength in *Y. lipolytica* have relied on assaying whole cultures for protein expression level (using reporters such as β-galactosidase) [45,53]. It is commonly known that these methods can mask potential bimodal "on/off" distributions within bacterial populations. Thus, to avoid this complication, we sought to utilize a fluorescence-based assay using the *Y. lipolytica* plasmid pSl16-Cen1-1(227) [98]. All results generated in this study, except where indicated, employed derivatives of this replicative, ARS-CEN based plasmid. Since codon biases are known to limit translation in *Y. lipolytica* [99] and no fluorescent reporter protein has been previously used in *Y. lipolytica* to gauge promoter strength, we initially evaluated several available fluorescent reporter proteins. The TEF promoter was coupled to four different fluorescence genes including yECitrine, EGFP, hrGFP and mStrawberry and flow cytometry was performed to determine reporter functionality. Of these variants, only hrGFP imparted detectable fluorescence (**Figure 2.1a**). This gene, optimized for expression in mammalian cells, has the highest Codon Adaptation Index for *Y. lipolytica* of the four fluorescence genes, [100] indicating the closest compatibility with codon usage frequencies for this organism.

As a result, the hrGFP reporter gene was used to evaluate the promoter strength of seven previously identified endogenous *Y. lipolytica* promoters—TEF, EXP, FBA, GPAT, GPD, YAT, and XPR2 (**Table 2.1**, **Figure 2.1b)** [45,101,102]. Based on this analysis, the relative ordering of promoters strengths is EXP > TEF > GPD > GPAT > YAT > XPR2 > FBA. The low mean fluorescence values of even the strongest of these native

promoters, EXP and TEF, highlight that even strong endogenous promoters in *Y. lipolytica* may be too low for metabolic engineering purposes. When each of these promoters was used in a plasmid-based construct, a bimodal fluorescence distribution was seen. These results strongly suggest an "on/off" switch mechanism resulting from the low copy CEN-based plasmid used in the study. It is interesting to note that this bimodal distribution was absent when using an integrated expression cassette instead. As a result, this work highlights an important consideration for using plasmids in a *Y. lipolytica* system. Based on prior reports to further improve the expression strength, a consensus four nucleotide "caca" sequence 5' of the +1 ATG codon was included [99,103-105]; however, this inclusion was found to be detrimental to expression levels in most cases (data not shown). The differential regulation patterns and small dynamic range of these endogenous promoters require a novel approach to enable metabolic engineering applications in this organism.

Figure 2.1: A fluorescence based assay for endogenous promoter characterization



(**A**) The relative fluorescence levels of yECitrine, mStraw, EGFP and hrGFP driven by the TEF promoter in *Yarrowia lipolytica* were tested with flow cytometry and mean fluorescence from different fluorescent protein were presented for comparison. These results indicate that hrGFP was the only functional fluorescent protein for this system. (**B**) Endogenous promoters including EXP, TEF, GPD, GPAT, YAT, XPR2, and FBA were used to drive the expression of hrGFP. The mean fluorescence data was collected with flow cytometry analysis for comparison at a timepoint of 48 hours in minimal medium. Error bars represent standard deviation from biological replicates. The control for these experiments was *Y. lipolytica* strain Po1f transformed with blank, replicative pMCSCen1 plasmid. The EXP and TEF promoters were identified as the strongest among this tested set.

Table 2.1: Elements utilized in initial hybrid promoter development

Promoter element name	Open reading frame regulated	YALI number	Basepair range	Reference number
EXP1	Export protein	YALI0C12034p	-999 to -1	[102]
GPAT	Glycerol-3-phosphate dehydrogenase	YALI0C00209p	-1130 to -1	[102]
GPD	Glyceraldehyde-3-phosphate dehydrogenase	YALI0C06369p	-931 to -1	[102]
TEF	Translation elongation factor EF-1α	YALI0C09141p	-406 to -1	[102]
YAT1	Ammonium transporter	YALI0E27203p	-775 to -1	[102]
FBA	Fructose-bisphosphate aldolase	YALI0E26004p	-830 to +171	[102]
XPR2	Alkaline extracellular protease	YALI0E26719p	-947 to -1	[53]
TEF(136)	Translation elongation factor EF-1α	YALI0C09141p	-136 to -1	This study
TEF(203)	Translation elongation factor EF-1α	YALI0C09141p	-203 to -1	This study
TEF(272)	Translation elongation factor EF-1α	YALI0C09141p	-272 to -1	This study
TEF(504)	Translation elongation factor EF-1α	YALI0C09141p	-504 to -1	This study
TEF(604)	Translation elongation factor EF-1α	YALI0C09141p	-604 to -1	This study
TEF(804)	Translation elongation factor EF-1α	YALI0C09141p	-804 to -1	This study
TEF(1004)	Translation elongation factor EF-1α	YALI0C09141p	-1004 to -1	This study
Leum	β-isopropylmalate dehydrogenase	YALI0C00407p	-92 to +25	[53]
UAS1B	Alkaline extracellular protease	YALI0E26719p	-805 to -701	[53]

The elements used in this study are listed with their names, open reading frame regulated, YALI accession number and basepair range.

2.3.2 Creating and characterizing a hybrid promoter series using the UAS1B element and minimal leucine core promoter

To bypass the limitations of endogenous promoters in *Y. lipolytica*, we evaluated the generalizable nature of hybrid promoters comprised of fused upstream activating sequences with a minimal core promoter. In prior work, a nearly perfect positive linear correlation was previously detected between the number of tandem UAS1B sequences and promoter outputs in the hp1d to hp4d promoter series [53]. This observation led to our

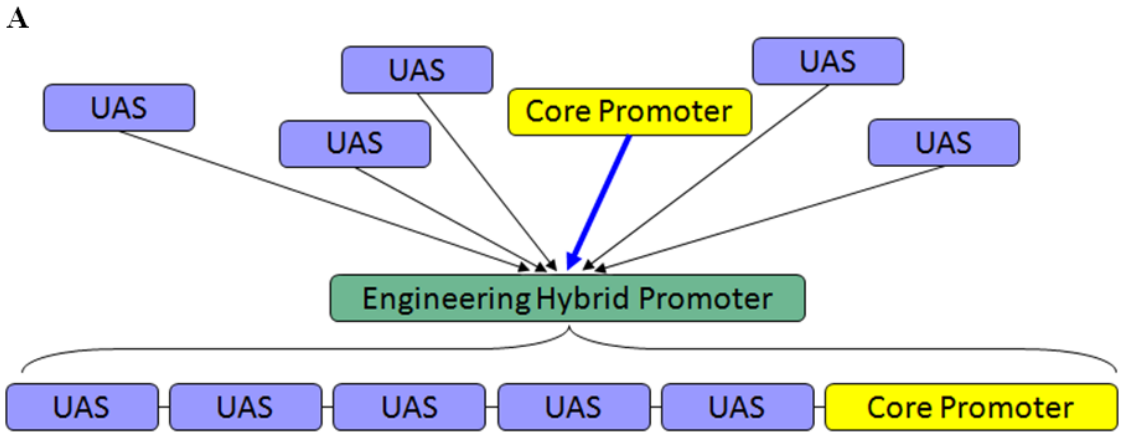
hypothesis that these minimal core promoters and potentially other *Y. lipolytica* full-length, native promoters are enhancer limited. To address this limitation, we created an expanded series of hybrid promoters by fusing between one and thirty two tandem UAS1B enhancer sequences to the leucine minimal promoter (Leum) to form promoters $UAS1B_1$-Leum through $UAS1B_{32}$-Leum (**Figure 2.2a**). The newly constructed UAS1B-Leum promoter series was tested with hrGFP based flow cytometry analysis after ~9 doublings to help insure promoter stability. The fluorescence data displayed several domains for correlation between output fluorescence and the number of upstream UAS elements. Initially, an exponential increase in fluorescence was seen as UAS1B sequence count increased from one to eight. This trend became linear through nineteen tandem repeats (a total of 1995 bp of upstream activating sequences upstream of the core promoter). Finally, the output fluorescence seemed to be saturated through 32 tandem UAS1B repeats (**Figure 2.2b**). This data strongly conformed to a Hill Cooperative Binding model (correlation coefficient of 0.95) and exhibited a high Hill Constant (3.889) which indicates a strong amount of binding cooperativity of the enhancer elements (**Figure 2.2b**). Specifically, this data was fit to the equation:

$$\text{Mean FL} = \text{min mean FL} + \text{max mean FL} - \text{min mean FL}) \times \; . \times \frac{\#\text{ of UAS}^{\text{Hill Coefficient}}}{c^{\text{Hill Coefficient}} + \#\text{ of UAS}^{\text{Hill Coefficient}}}$$

with the resulting coefficients of a= 0.794, Hill Coefficient= 3.889, and c= 10.146. As with the prior tests using endogenous promoters, these plasmid-based expression cassettes showed a bimodal distribution of fluorescence levels that was independent of the promoter being evaluated. However, when these cassettes were integrated into the

genome, these promoters gave a singular, high expression peak which indicates that the bimodal nature was conceivably due to an episomal expression issue related to this organism. These integrative expression cassettes generated roughly two-fold higher fluorescence values than their corresponding replicative-based cassettes, and the lack of bimodal distribution was seen for all promoters tested and was not dependent on UAS1B copy number (data not shown).

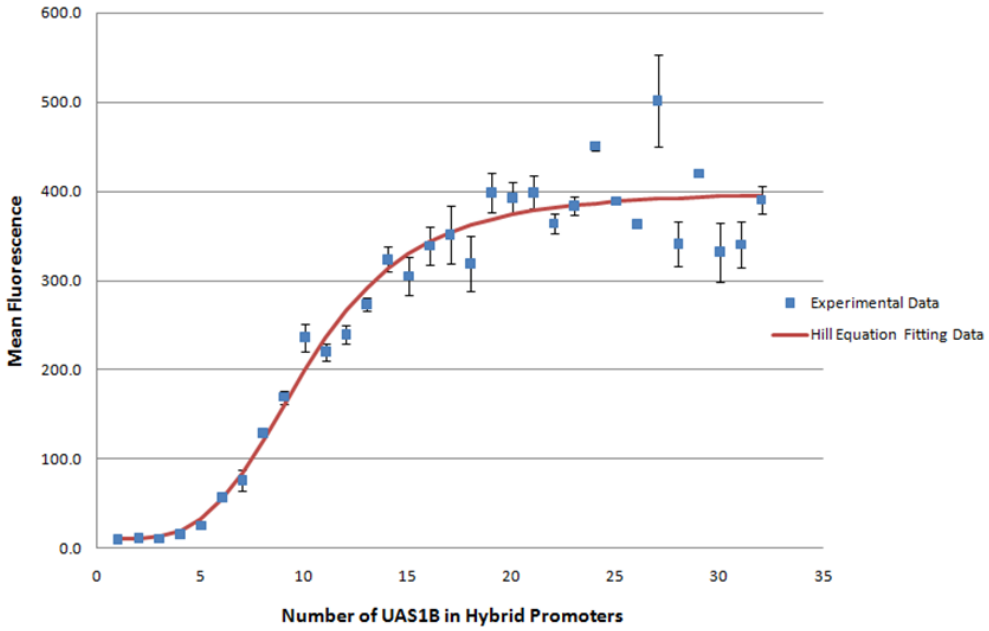
Figure 2.2: Developing and characterizing a UAS1B-Leum hybrid promoter set.



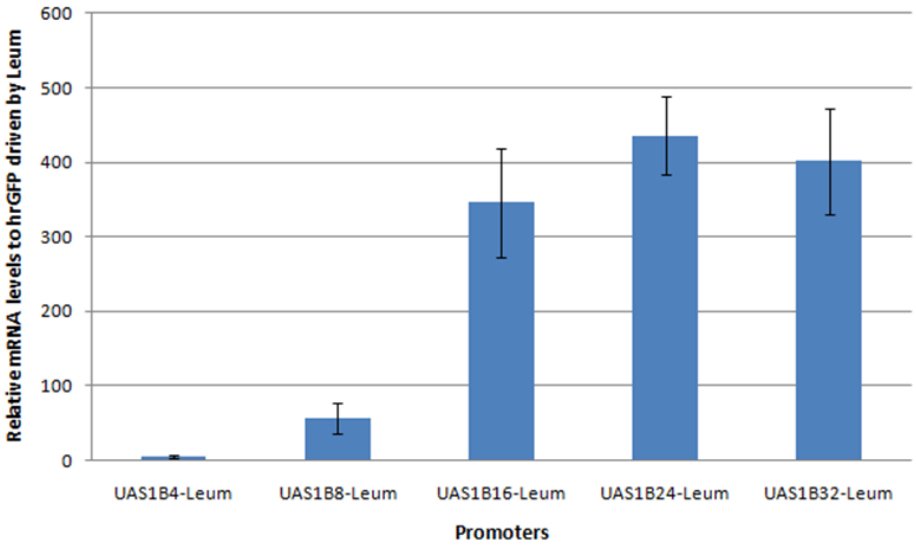
(**A**) A schematic diagram for UAS-enabled hybrid promoter engineering illustrates that these promoters were created from two distinct parts—the core promoter and upstream activation sequence (UAS) elements. By inserting tandem copies of UAS upstream to the core promoter, fine-tuned promoter sets with the strongest expression levels seen in *Y. lipolytica* were created.

Figure 2.2 continued

B

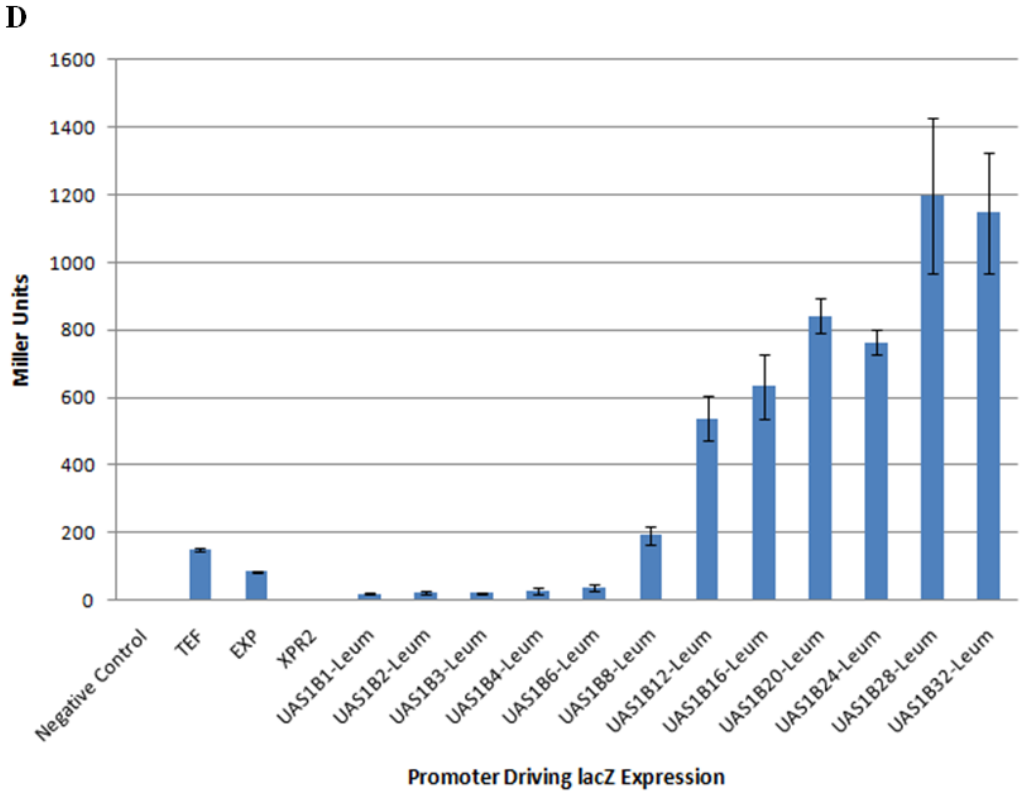


C



(**B**) The tandem insertion of 1 to 32 copies of UAS1B sequences upstream of a minimal LEU promoter enabled construction of the $UAS1B_1$-Leum to $UAS1B_{32}$-Leum promoters. Error bars represent standard deviation from biological triplicates. The relative promoter strengths of UAS1B-Leum promoters were fit to a Hill equation, resulting in a= 0.794, Hill Coefficient= 3.889, c= 10.146 and r^2= 0.9495 using Polymath Software (Willimantic, CT). (**C**) Transcriptional profiling of select promoter constructs was used to calculate mRNA levels relative to the minimal Leum promoter. Expression profiles matched fluorescence data. Error bars represent standard deviation from technical triplicates.

Figure 2.2 continued



(**D**) Promoter driven expression of a β-galactosidase gene yielded similar results to hrGFP. Lysed cultures were pelleted down prior to absorbance measurements. Error bars represent standard deviation from biological triplicates.

2.3.3 Transcriptional analysis of the UAS1B-leum hybrid promoter series

A transcriptional analysis was performed to confirm that the observed effect in fluorescence was indeed manifested at the transcriptional level. To do so, qRT-PCR analysis was employed using the hrGFP mRNA of select promoter constructs (Leum, $UAS1B_4$-Leum, $UAS1B_8$-Leum, $UAS1B_{16}$-Leum, $UAS1B_{24}$-Leum, and $UAS1B_{32}$-Leum) (**Figure 2.2c**). Expression values were normalized to the mRNA level seen with only the minimal leucine promoter used to drive hrGFP. Indeed, the increase in mean fluorescence levels was strongly correlated with the increase in relative mRNA levels.

The relative mRNA levels increased and likewise plateaued for constructs with a high number of UAS1B repeats. Moreover, these results demonstrate an extraordinary range of promoter strength in this series with more than a 400 fold dynamic range of transcript level between the minimal promoter and strongest promoters in this set.

2.3.4 Utility and stability of the UAS1B-leum hybrid promoter series

To ensure that the observed effect was independent of reporter gene, we sought to provide a further characterization of the UAS1B-Leum promoter series with a separate reporter gene, the β-galactosidase gene encoded by *E. coli* lacZ. Select promoter constructs (including the endogenous TEF, EXP, and XPR2 promoters as well as hybrid UAS1B-leum constructs containing 1, 2, 3, 4, 6, 8, 12, 16, 20, 24, 28, and 32 UAS1B copies) were used to construct expression cassettes with lacZ in place of hrGFP. β-galactosidase assays were performed as described previously [106,107] with a maximum value of 1198 miller units generated by the $UAS1B_{28}$-leum construct (**Figure 2.2d**). Promoter strength increased with increasing UAS1B copy number up through 32 UAS1B repeats unlike in the HrGFP assay. However, the β-galactosidase assay results matched well with the hrGFP analysis and the data showed a strong positive statistical correlation (r^2=0.85). The strongest UAS1B-Leum hybrid promoter exhibited a more than eight fold increase in promoter strength in terms of Miller units compared to the strong endogenous promoters tested in this study. These results demonstrate that the UAS1B-leum hybrid promoter series developed here is a generic tool for obtaining tuned gene expression in *Y. lipolytica*. It should be noted that the expression of the $UAS1B_4$-Leum reported here were substantially lower than previously reported. These comparisons are difficult due to

the slight differences in restriction sites used to create these hybrid promoters, the use of rich media in prior reports, and the use of replicative plasmids here compared with integrated plasmids in other studies. Even with these difference and discrepancies, this work still presents up to a fourfold increase in performance compared with the best reported endogenous promoters or previously constructed hybrid promoters [53,108]. This illustrates that multiple tandem repeats of the UAS1B enhancer element activate transcription to levels far stronger than those previously described and that this enhancer activation can occur at regions more than 2000 nucleotides upstream of the start codon (for $UAS1B_{16\text{-}32}$-Leum).

These hybrid promoters rely on a high number of tandem repeats, thus, genetic stability was evaluated. To accomplish this, selected promoter constructs (with 12 or 16 repeated UAS1B elements) were tested on the basis of sequence fidelity after non-selective serial subculturing. These strains were subcultured for a total of 36 generations. After this process, cells were harvested and plasmids were isolated and sequenced to assess gene construct stability. In total, 20 separate plasmids containing the $UAS1B_{12}$-Leum and 20 containing the $UAS1B_{16}$-Leum promoters were evaluated. 17 out of 20 $UAS1B_{12}$-Leum and 20 out of 20 $UAS1B_{16}$-Leum promoters were positively sequence and restriction enzyme digest confirmed after 36 doublings. Only 3 out of 20 $UAS1B_{12}$-Leum promoters were unstable through 36 generations as they were truncated down to $UAS1B_3$-Leum. Thus, these promoters are suitably stable in *Y. lipolytica* for long-term expression and use. Collectively, this data suggests that the expression output from hybrid promoters can be altered by changing the number of activating regions with a

given core promoter region. Next, we sought to address the ability to alter the core promote region of this construct.

2.3.5 Generalizing the hybrid promoter approach by switching the core promoter region

As discussed above, a hybrid promoter has two potential independent elements—activating regions and core promoter region. The data above suggests that tandem UAS elements may serve as movable, synthetic expression amplifiers for a given promoter. Next, we sought to test the hypothesis that even native promoters in *Y. lipolytica* are enhancer limited and can be strengthened by adding additional UAS elements. To do so, we constructed new hybrid promoters containing either eight tandem UAS1B sequences ($UAS1B_8$) or sixteen tandem UAS1B sequences ($UAS1B_{16}$) inserted 5' upstream of a series of different TEF-based core promoters. Specifically, we amplified eight different regions of the TEF promoter spanning 136 bp and 1004 bp upstream of the ATG starting site from PO1f [53] genomic DNA (**Table 2.1**). Included in this set is the consensus 404 bp TEF promoter for *Y. lipolytica* as well as lengthened and truncated versions of this promoter. These eight core TEF promoters and their corresponding $UAS1B_8$ and $UAS1B_{16}$ hybrid promoters were tested and compared with the Leum, $UAS1B_8$-Leum and $UAS1B_{16}$-Leum constructs.

This new series of hybrid promoters was assayed via hrGFP fluorescence by flow cytometry. In the absence of UAS elements, it can be seen that the fluorescence value decreases for truncated promoters below the consensus TEF size as expected. Moreover, the full length (and larger) TEF promoters have more strength than the minimal leucine

promoter (**Figure 2.3a**). The $UAS1B_8$ and $UAS1B_{16}$ enhancer fragments in isolation do not confer any promoter activity (data not shown). When these enhancer fragments were fused with the TEF promoter elements, a substantial increase in the net promoter strength was seen regardless of the TEF variant utilized. The enhancement provided by $UAS1B_8$ was roughly half the value obtained by using $UAS1B_{16}$. Moreover, these enhancements were seen for both more minimal and full length TEF promoter elements, even with the existence of naturally occurring UAS elements in the consensus and longer TEF promoters. Thus, this data suggests shows that even strong endogenous promoters like TEF are enhancer limited in *Y. lipolytica* and their expression capacity can be increased through additional UAS elements. Furthermore, we demonstrated that the ability of a UAS1B element to amplify expression is independent of the core promoter element.

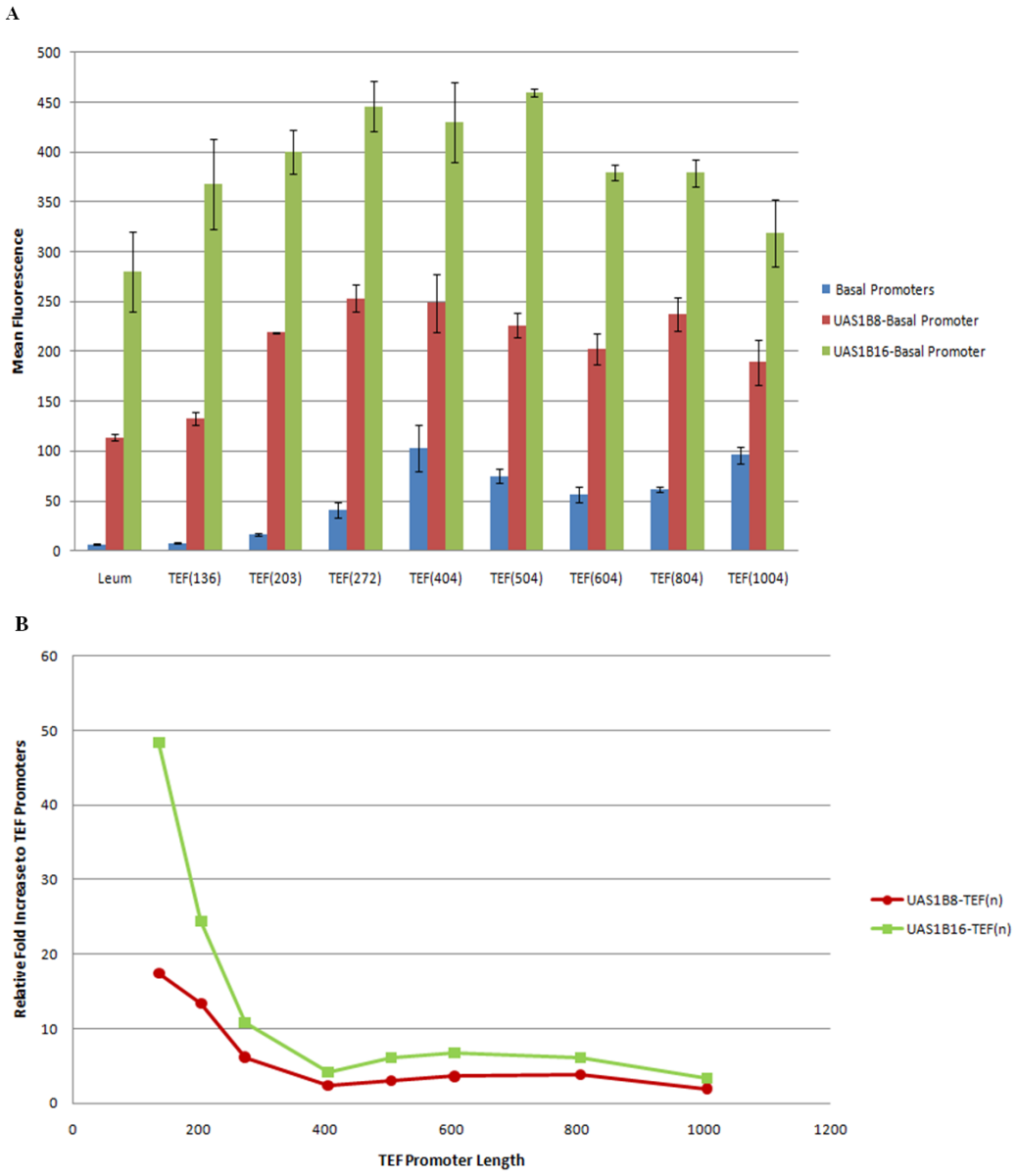
The amplification of expression imparted by the $UAS1B_8$ and $UAS1B_{16}$ enhancer elements was not even for all promoters. The fold increase of the constructs relative to the UAS-free TEF core promoters is plotted in (**Figure 2.3b**). In general, the largest improvement is obtained when the UAS elements are placed most closely 5' of a core promoter region. This fold improvement trend is rationalized in terms of a mechanism of proximity and localization of enhancer elements. However, it is interesting to note that the total promoter size (with 16 activating sequences and a core region) totals upwards of three thousand basepairs for many of these promoters—a region that is quite large for typical yeast constructs. In addition, many of these larger constructs were the best performing promoters. Specifically, nearly all of these promoters proved to be stronger than the corresponding $UAS1B_8$-Leum and $UAS1B_{16}$-Leum promoters, demonstrating

the fitness of the TEF-based core promoter regions for strong hybrid promoter engineering. Thus, both the choice of tandem enhancer element and core promoter element contribute to the collective strength of hybrid promoters. In this regard, this work demonstrates that tandem UAS elements serve to synthetically amplify the expression level imparted by the core promoter element chosen. In this regard, transcription factor binding is posited to serve as a major rate limiting step for transcription at promoter sites with addition of upstream activating sites alleviating this limitation. Thus, it is possible for us to control transcriptional activity by indirectly modulating transcription factor localization or affinity through the choice of enhancer and core elements in a hybrid promoter. This observation raises the possibility of rationally designing hybrid promoters with specified expression strength. The drastic increase in expression levels by both of the $UAS1B_8$ and $UAS1B_{16}$ elements across this series demonstrates that these genetic elements are portable, modular components that can generically alleviate native enhancer-limitation without disrupting endogenous regulation. The modularity of the UAS1B insert and the strength of the UAS1B-Leum and UAS1B-TEF series advocate the use of hybrid promoter engineering as generic approach towards building stronger, fine-tuned promoter libraries with interchangeable, modular components.

A promoter-less plasmid was utilized as a negative control for the majority of these experiments. True controls would consist of 105 basepair DNA segments that lack promoter function fused in tandem, to mimic each $UAS1B_n$-Core promoter construct.

Unfortunately, lack of an acceptable characterized DNA sequence and time constraints prevented construction of these controls.

Figure 2.3: Expanding the hybrid promoter approach by altering core promoter element.



(**A**) The characterization of relative promoter strengths for the core TEF promoters (blue), $UAS1B_8$-TEF promoters (red) and $UAS1B_{16}$-TEF promoters (green) using flow cytometry. This data was compared with the $UAS1B_8$-Leum and $UAS1B_{16}$-Leum promoters. (**B**) The tuning ability of $UAS1B_8$ and $UAS1B_{16}$ decreases as a function of core promoter length.

2.4 SUMMARY AND CONCLUSIONS

Collectively, these results give credence to the theory that *Y. lipolytica* promoters are enhancer-limited, and we have shown that this limitation can be effectively overcome through hybrid promoter engineering. We observed that both the choice of tandem enhancer element and core promoter element contribute to the collective strength of hybrid promoters, and that tandem UAS elements help bypass the enhancer limited nature of promoters by serving as transcriptional amplifiers. The generic approach of hybrid promoter engineering described here is an important, generic synthetic biology tool enabling the construction of high-level and fine-tuned promoters with interchangeable promoter parts. This approach is one of the first to rationally amplify the expression output of a given promoter element. By utilizing this approach, we have expanded the metabolic engineering toolbox in *Y. lipolytica* and developed several novel promoter series - $UAS1B_{1\text{-}32}$-Leum, $UAS1B_{8}$-TEF, and $UAS1B_{16}$-TEF. In doing so, we created the strongest characterized promoters in *Y. lipolytica* and the first ever reported capacity for tunable gene expression in this organism.

Heterologous protein expression requires strong promoters to obtain high protein expression level, while metabolic pathway engineering necessitates strictly controlled, fine-tuned promoters set to optimize pathways. The generic hybrid promoter approach described here accomplished both of these tasks. Moreover, this work establishes synthetic hybrid promoter construction as an approach to expand and enhance the strength of engineered promoters in a manner not accessible through traditional promoter engineering applications.

Chapter 3: *De Novo* Hybrid Promoter Development

3.1 Chapter Summary

Hybrid promoter construction entails two main tasks, first isolating and characterizing core promoter and upstream activating sequence (UAS) elements, and then utilizing these parts to build functional, high strength promoters. The majority of hybrid promoter engineering work - done with the intent to maximize and control gene expression - has been performed in *Y. lipolytica* and have all utilized the same upstream activating sequence, UAS1B, as the sole UAS element. Therefore, the full flexibility of hybrid promoter generation has been limited by reliance on well-established upstream activating regions that are unlikely to be characterized in nonconventional hosts. To circumvent this limitation, we describe here the methodology for the isolation, characterization, and incorporation of novel UAS elements into strong, tunable hybrid promoter libraries in *Y. lipolytica*.

As a test case, the native TEF promoter in *Yarrowia lipolytica* was examined to identify putative UAS elements that serve as modular synthetic transcriptional activators. Resulting synthetic promoters containing a core promoter region activated by between one and twelve tandem repeats of the newly isolated, 230 nucleotide UAS_{TEF}#2 element showed promoter strengths 3 to 4.5-fold times the native TEF promoter. Further analysis through transcription factor binding site abrogation revealed the GCR1p binding site to be necessary for complete UAS_{TEF}#2 function. These various promoters were tested for function in a variety of carbon sources. Finally, by combining disparate UAS elements

(in this case, UAS_{TEF} and UAS1B), we developed a high-strength promoter with for *Y. lipolytica* with an expression level of nearly seven-fold higher than that of the strong, constitutive TEF promoter. Thus, the general strategy described here enables the efficient, *de novo* construction of synthetic hybrid promoters to both increase native expression capacity and to produce libraries for tunable gene expression. Without the prerequisite of a previously characterized UAS element, the hybrid promoter engineering approach is even more useful for nonconventional hosts.

3.2 INTRODUCTION

Synthetic hybrid promoters consist of a core promoter region fused to a single upstream activating sequence (UAS) or multiple tandem UAS repeats that augment and modulate promoter function [2,33,53,94,95], and hybrid promoter engineering has been demonstrated to be the only effective approach to increase expression capacity in an organism [109]. All prior hybrid promoter engineering efforts designed to increase overall promoter strength have been performed in the oleaginous yeast *Y. lipolytica* and have utilized the same 105 basepair upstream activating sequence, named UAS1B, as the sole UAS element. UAS1B was originally isolated through a functional dissection of the strong, tightly regulated XPR2 native promoter and showed promising constitutive transcriptional activation abilities [53,96,110]. Previously, we showed that multiple tandem repeats of this UAS element adjacent to a core minimal LEU2 promoter or core TEF-based promoters resulted in significant transcriptional activation [2,53].

However, fully characterized UAS regions are not likely to be found in nonconventional organisms. Thus, we sought to develop a *de novo* approach, encapsulating UAS isolation and hybrid promoter construction, to enable the application of the hybrid approach into any nonconventional host.

3.3 Results and Discussion

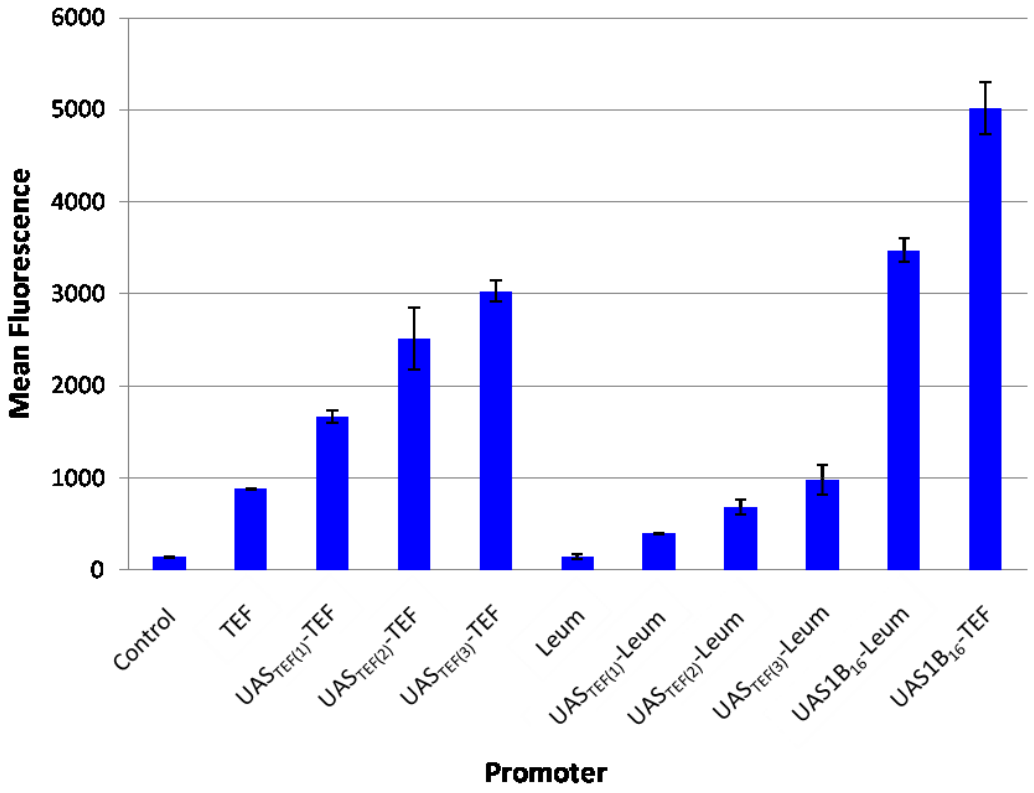
3.3.1 Isolation of a *Y. lipolytica* TEF promoter region UAS

Prior analysis of the 406 basepair full length TEF promoter revealed that the 136 basepair proximal to the ATG start codon, dubbed TEF(136), was unable to drive the expression of an hrGFP reporter gene [2]. However, when this core promoter fragment was used in conjunction with a known UAS element, resulting TEF(136)-based hybrid promoters generated high fluorescence levels. These results established TEF(136) as a functional, low-strength core promoter that had been apparently stripped of its native UAS elements [2]. Analysis of the TEF promoter revealed that a drastic change in GC content coincided with the end of all putative TATA elements at 149 basepairs upstream of the ATG start codon. Consequently, we postulated that a strong upstream activating sequence responsible for the majority of the TEF promoter's transcriptional activation ability must be located further upstream in the remaining 257 basepairs. We tested this hypothesis by constructing two distinct series of hybrid promoters in which between one and three tandem repeats of a putative 257 basepair TEF upstream activating sequence, dubbed UAS_{TEF}, were fused to a core promoter. Hybrid promoter series $UAS_{TEF(n)}$-TEF

employed the 406 basepair native TEF promoter (TEF) as a core promoter, while series $UAS_{TEF(n)}$-Leum employed a minimal core LEU2 promoter (Leum). Flow cytometry analysis of the $UAS_{TEF(n)}$-TEF and $UAS_{TEF(n)}$-Leum hybrid promoters revealed strong expression enhancement by tandem UAS_{TEF} elements **(Figure 3.1)**. Therefore, we validated the existence of a strong UAS element upstream of the minimal core TEF(136) sequence, and demonstrated its capacity as a synthetic amplifier when fused to either the minimal leucine promoter or the full native TEF promoter. The enhancement of fluorescence levels were strongly linearly correlated to the number of UAS_{TEF} copies (r^2 = .9899 for TEF core and r^2 = .9983 for Leum core; r^2 values calculated using a linear regression in Microsoft Excel), demonstrating the modularity and functional additivity of this novel UAS. This linear relationship is in contrast to the Hill-curve dynamics seen with the UAS1B enhancement [2]. This difference suggests that UAS elements have distinct transfer functions of activity. Net promoter strength strongly depended on core promoter choice, as hybrid promoters employing the TEF core were more than three-fold stronger than Leum-based counterparts. The maximum expression amplification imparted to each core promoter differed as well. In this vain, three UAS_{TEF} tandem repeats generated a 6.45-fold amplification of expression when fused to the Leum core promoter (compared to basal Leum), while only a 3.44-fold increase emerged in the corresponding TEF-based constructs. These results echo our previous findings that amplification levels were highest with more minimal core promoter constructs. Moreover, both UAS enhancer element and core promoter elements contribute to overall

hybrid promoter strength in a logical manner, raising the possibility of rationally designing synthetic hybrid promoters.

Figure 3.1: Functional testing of the novel UAS_{TEF} element



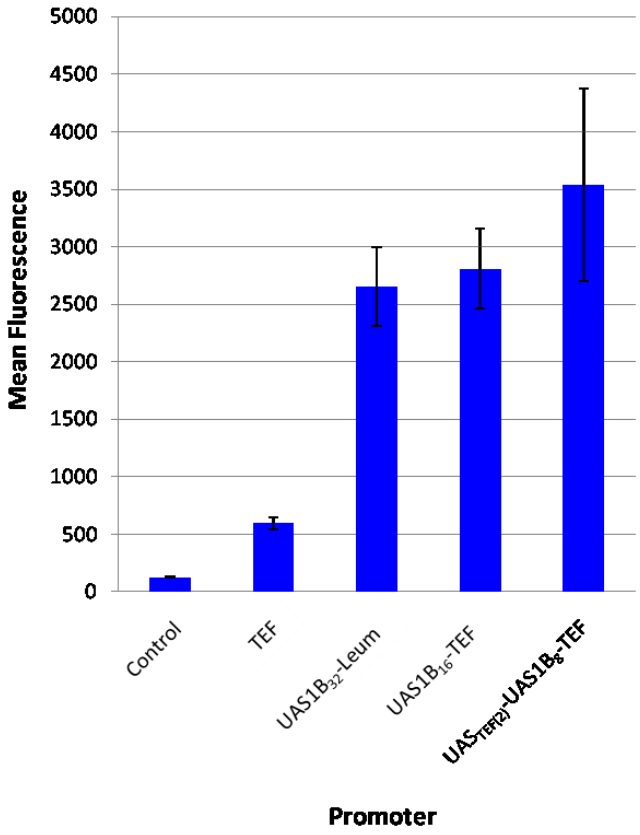
Relative fluorescence values indicate that the UAS_{TEF} element functions as a strong modular upstream activating sequence, linearly increasing expression capacity of both the TEF native promoter (r^2 = .9899) and the Leum minimal promoter (r^2 = .9983). Three tandem repeats of the UAS_{TEF} element elevated protein expression to more than three-fold base levels for the TEF promoter and six-fold for the Leum promoter. Final expression capacity of the $UAS_{TEF(3)}$-TEF promoter approached $UAS1B_{16}$-TEF and $UAS1B_{16}$-Leum levels, two of the strongest previously characterized hybrid promoter in *Y. lipolytica*. Error bars represent the standard deviation of measurements between biological triplicates

3.3.2 Investigating the effect of incorporating disparate UAS sequences

We next investigated the impact of combining disparate UAS elements. The UAS_{TEF} element proximal to the core promoter in promoter $UAS_{TEF(3)}$-TEF was replaced by an $UAS1B_8$ fragment [2] to form hybrid promoter $UAS_{TEF(2)}$-$UAS1B_8$-TEF. The mean

fluorescence levels generated by $UAS_{TEF(2)}$-$UAS1B_8$-TEF are at least as high as the strongest promoters yet described in *Y. lipolytica* [2], advocating future work assessing potential benefits of incorporating multiple, distinct UAS elements into hybrid promoters **(Figure 3.2)**.

Figure 3.2: Combining the novel UAS_{TEF} with the UAS1B element



Relative fluorescence values indicate that the $UAS_{TEF(2)}$-$UAS1B_8$-TEF promoter surpasses in strength the previously characterized $UAS1B_{16}$-TEF and $UAS1B_{32}$-Leum promoters, the strongest promoters constructed in *Y. lipolytica*. Error bars represent the standard deviation of measurements between biological triplicates. P-value comparing $UAS_{TEF(2)}$-$UAS1B_8$-TEF to $UAS1B_{32}$-Leum = 0.17.

As described above, all prior hybrid promoter attempts in *Y. lipolytica* have been limited to only one UAS element. But by combining disparate UAS elements, we obtained surprisingly high levels of gene expression in a much shorter total promoter size

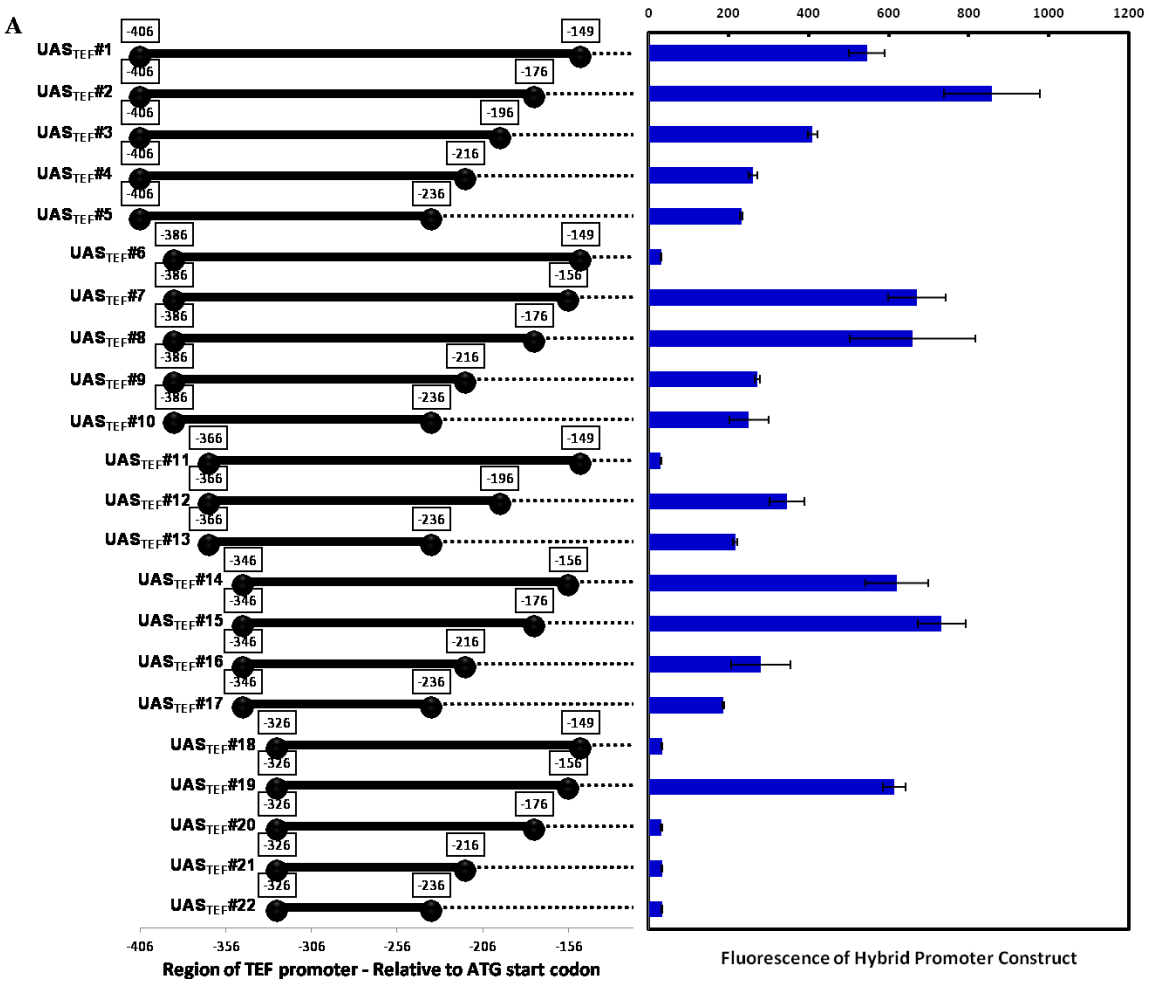
(1850bp vs. up to 3700bp). We hypothesize that the distinct UAS elements localize different transcription activating factors with unexpected cooperative effects and thus more efficiently enhance transcription. Prior evidence has demonstrated that the combination of GCR1 and RAP1 regulatory sequences constitute one of the strongest activating sequences known in *S. cerevisiae* [111]. The UAS1B sequence was initially identified through a deletion analysis of the XPR2 promoter and was shown to be necessary for XPR2 promoter function [110]. The UAS1B sequence contains a TUF/RAP1 transcription factor binding site that by itself rescues native promoter activity when inserted to replace a UAS1B deletion [110]. Our transcription factor deletion analysis (described below) confirmed the importance of a consensus GCR1 binding site found in the UAS_{TEF} sequence towards transcriptional activation. Hence, the unison of GCR1 and RAP1 transcription factor binding sites within the same promoter could drastically increase transcriptional capacity, and it is likely that the localization of these regulatory sequences serves as the mechanism behind the synergy seen when combining the disparate UAS_{TEF} and $UAS1B_8$ elements to form hybrid promoters. These types of synergies will form the basis of developing predictive models for designing *de novo* promoters.

3.3.3 Functional dissection analysis of the UAS_{TEF} element through promoter truncation and transcription factor binding site removal

With the 257 bp UAS_{TEF} established as a modular, synthetic transcriptional amplifier, we sought to dissect its genetic sequence in search of a more minimal, compact UAS element. Twenty-two unique but overlapping fragments spanning the UAS_{TEF}

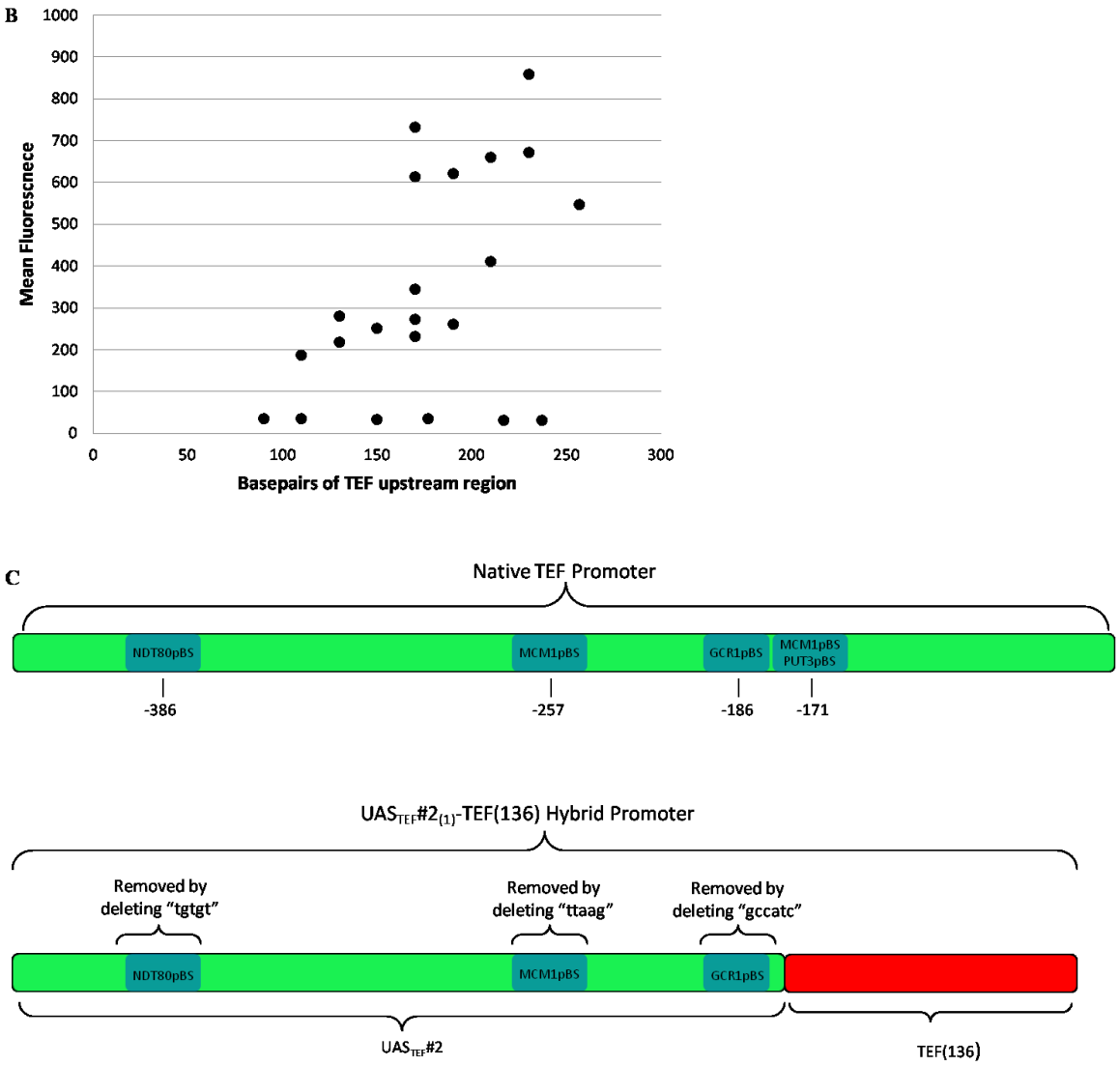
sequence were PCR amplified to create a UAS_{TEF}-based truncation library **(Figure 3.3a)**. Putative UAS library elements were individually fused to the minimal TEF(136) core promoter to form promoters UAS_{TEF}#1-TEF(136) through UAS_{TEF}#22-TEF(136). The minimal TEF(136) core promoter was selected for hybrid promoter construction to allow for the highest sensitivity in measuring UAS activity. The newly constructed hybrid promoter series was tested with hrGFP based flow cytometry analysis. The fluorescence data reveal that the majority of truncated UAS_{TEF} fragments continue to retain at least modest UAS activity **(Figure 3.3a,b)**. This result suggests that nearly all of these fragments can serve to create functional, synthetic hybrid promoters. However, hybrid promoter strength showed a pronounced correlation to UAS_{TEF} fragment length, providing evidence that the majority of the original 257 bp UAS_{TEF} sequence is required for full activity **(Figure 3.3b)**. UAS_{TEF}#2 displayed the highest activating capacity of the UAS_{TEF} truncations tested, elevating the strength of the TEF(136) core promoter to levels 1.5-fold times the native TEF promoter **(Figure 3.3a)**. Thus, we selected UAS_{TEF}#2 for further hybrid promoter library construction and for a more rigorous examination.

Figure 3.3: Dissection of the UAS_{TEF} element.



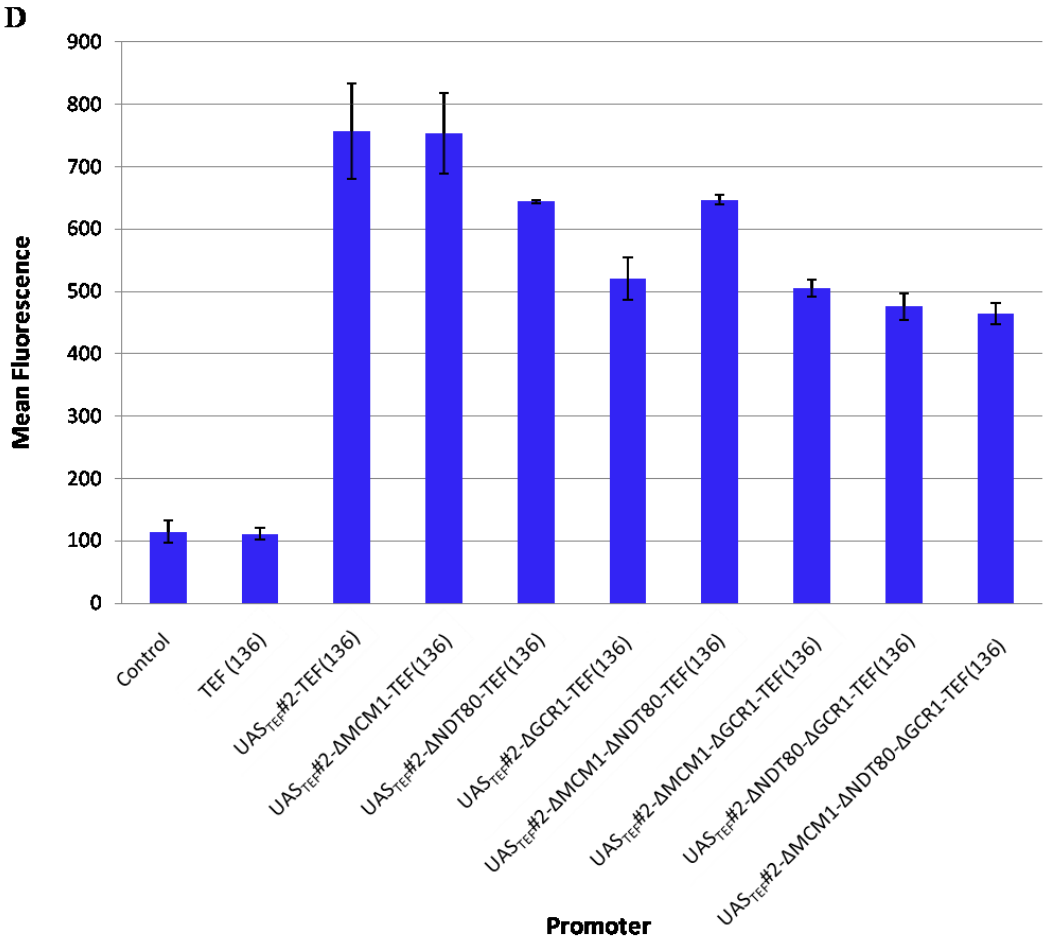
(A) Twenty-two overlapping fragments spanning the UAS_{TEF} sequence were inserted upstream of a TEF(136)-hrGFP expression cassette and tested for transcriptional amplification activity. Relative fluorescence values are shown. Error bars represent the standard deviation of measurements between biological triplicates.

Figure 3.3 continued



(B) Relative fluorescence values are plotted as a function of putative UAS length, with a noticeable correlation between decreasing putative UAS strength and decreasing UAS length. **(C)** A simplified schematic picture is provided detailing the location of specific consensus transcription factor binding sites within the UAS_{TEF}#2 element and the TEF native promoter. Error bars represent the standard deviation of measurements between biological triplicates.

Figure 3.3 continued



(D) Three putative transcription factor binding sites (NDT80p, MCM1p, and GCR1p) were removed individually and combinatorially from the UAS$_{TEF}$#2 element within a UAS$_{TEF}$#2$_{(1)}$-TEF(136)-hrGFP expression cassette. Relative fluorescence values are shown. Error bars represent the standard deviation of measurements between biological triplicates.

The majority of the UAS$_{TEF}$#2 element is necessary for full UAS activity, precluding a more thorough characterization through truncation analyses. We characterized UAS$_{TEF}$#2 through the systematic removal of three putative consensus transcription factor binding sites predicted within this sequence. In our truncation analysis, we observed that removing twenty seven basepair at the 5' end of UAS$_{TEF}$ (-406

to -386 deletion) tended to decrease UAS strength, but further truncation had little effect. We also observed that the removal of twenty seven basepairs from the 3' end of the original UAS_{TEF} (-149 to -176 deletion) often increased UAS strength, but further 3' truncation (-176 to -196/-216/-236 deletion) always decreased UAS strength **(Figure 3.3a,c)**. Thus, we concluded the existence of an upstream repressive element (URS) between basepairs -149 to -176 and enhancing UAS elements in the remaining truncated regions (-406 to -386 and -176 to -196/-216/-236 deletion).

An analysis of the native TEF and the UAS_{TEF}#2-TEF(136) promoters at the sequence level in the yeast promoter database (SCPD) [21] revealed that consensus binding sites for yeast transcription factors MCM1p and PUT3p are lost in the twenty seven basepair truncation. Further sequence analysis of the UAS_{TEF}#2 element also highlighted consensus binding sites for transcription factors NDT80p at position -386, for GCR1p at position -186, and for a separate MCM1p at position -257 **(Figure 3.3c)**. We hypothesized that the GCR1p and NDT80p binding sites grant UAS activity, while this second native MCM1p binding site reduces UAS activity. To test this hypothesis, we employed site-directed mutagenesis to delete the NDT80p, GCR1p, and MCM1p binding sites from the UAS_{TEF}#2-TEF(136) promoter **(Figure 3.3c)**, and tested and compared binding site deletion mutants with hrGFP based flow cytometry. This deletion analysis of putative revealed that MCM1p binding sites have no effect on TEF promoter activity **(Figure 3.3d)**, implicating the PUT3p binding site as a repressive element within the 27bp 3' fragment of UAS_{TEF}. As expected, removal of the NDT80p and GCR1p bindings sites reduced UAS_{TEF}#2-TEF(136) promoter strength, by 15 and 31% respectively

(Figure 3.3d). Combinatorial deletion of NDT80p and GCR1p binding sites only slightly further decreased promoter strength, revealing the GCR1p binding site as the predominate site responsible for UAS_{TEF}#2 function. Thus, the GCR1p binding site motif is essential for UAS capability, while URS activity is most likely conferred by the PUT3p binding site. By combining a truncation analysis with targeted binding site mutagenesis, we were able to propose transcription factors localized by the TEF_{UAS} region that both activate and repress transcription and attain a greater understanding of *Y. lipolytica* native promoters. Knowledge of these specific transcription factors allowed us to explain unexpected interactions resulting from combining disparate UAS elements in a single hybrid promoter.

3.3.4 Constructing hybrid promoters with the novel UAS

Above, we have demonstrated how UAS elements can be deduced and mechanistically studied through promoter truncation and mutagenesis analyses. To complete the generic hybrid promoter construction process, we utilized one, two, three, four, five, six, nine or twelve tandem repeats of the newly isolated UAS_{TEF}#2 sequence to enhance expression of the minimal TEF(136) core promoter. These new UAS_{TEF}#2 based promoters were tested and compared to the native TEF promoter, amongst other controls, via hrGFP fluorescence by flow cytometry. Multiple copies of UAS_{TEF}#2 increased hybrid promoter strength for up to six tandem repeats, at which point expression enhancement from added UAS_{TEF}#2 elements had been saturated **(Figure 3.4a)**. Ultimately, UAS_{TEF}#2-TEF(136) hybrids yielded expression levels 3.5-fold higher than the native, full length TEF promoter starting point **(Figure 3.4a)**. Expression

cassettes containing tandem UAS_{TEF}#2 copies without the TEF(136) promoter generated no expression above background levels. To further demonstrate the utility of the UAS_{TEF}#2-TEF(136) series, we utilized the β-galactosidase gene as alternate reporter protein. Once again, UAS_{TEF}#2-mediated enhancement increased promoter activity, producing a final expression enhancement to nearly twice that of the native TEF using this marker **(Figure 3.4b)**. Thus, UAS elements isolated from native promoters can be utilized to increase available promoter strength within an organism.

Figure 3.4: Functional testing of the novel UAS_{TEF}#2 element to complete de novo hybrid promoter construction

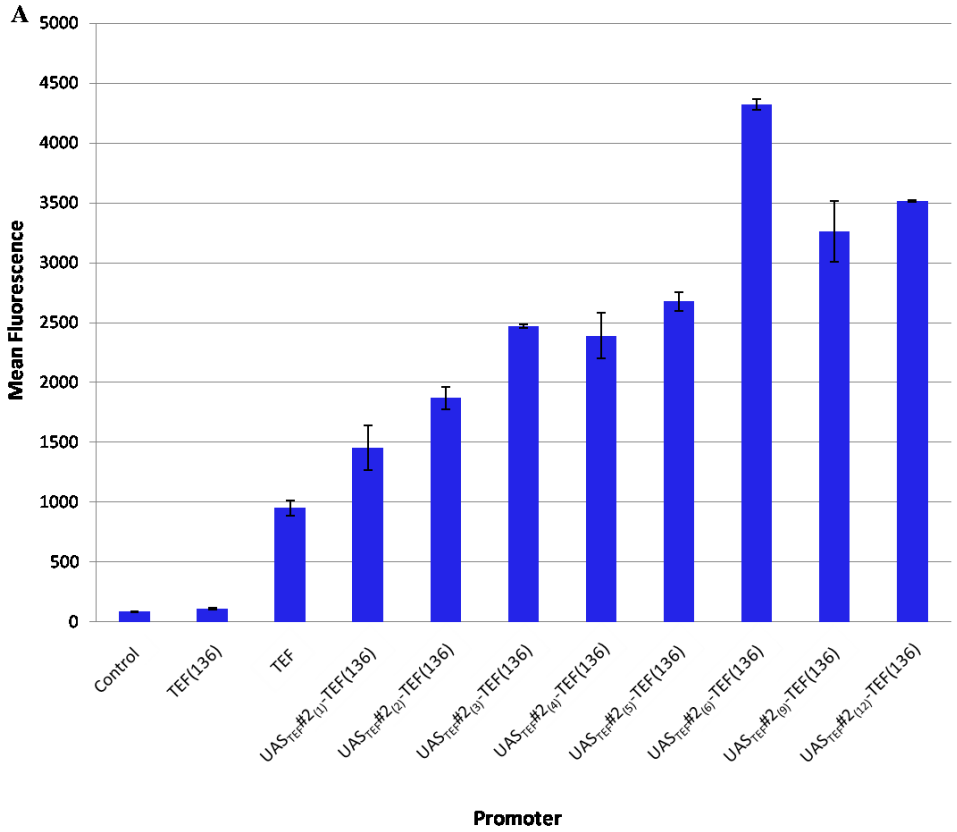
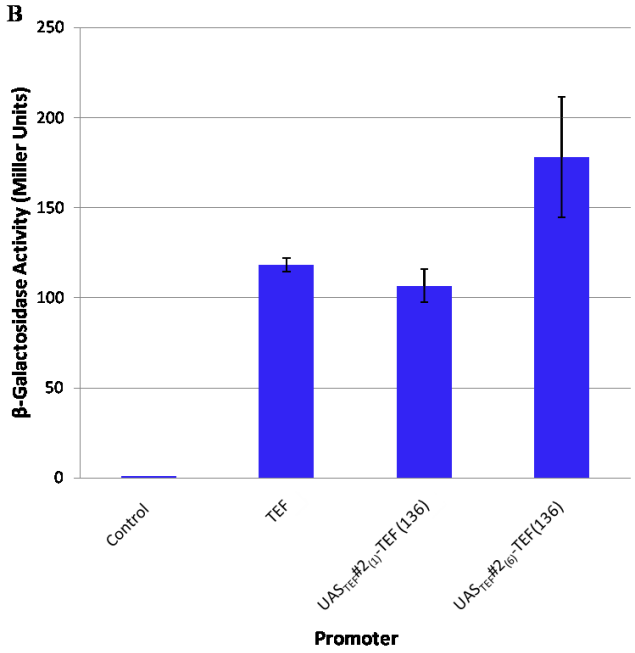


Figure 3.4 continued



(A) Relative fluorescence values indicate that the UAS_{TEF}#2 element functions as an upstream activating sequence, increasing expression capacity of the TEF(136) promoter to levels 3.5-fold higher than the native TEF promoter. Final UAS_{TEF}#2-based promoter strength remained below UAS_{TEF}-based and previously characterized promoters, $UAS1B_{32}$-Leum and $UAS1B_{16}$-TEF. **(B)** Hybrid promoters were tested with a β-galactosidase reporter gene, yielding similar results to the hrGFP assay. Error bars represent the standard deviation of measurements between biological triplicates. P-value comparing UAS_{TEF}#2$_{(6)}$-TEF(136) to UAS_{TEF}#2$_{(1)}$-TEF(136) = 0.06.

3.3.5 Kinetic analysis of hybrid promoters and effect of media formulation

Previous analysis of the UAS1B element (the only other known upstream activating sequence in *Y. lipolytica*) revealed expression profiles dependent on both time course and media formulation [47,53]. Thus, we sought to further characterize our novel UAS_{TEF}-based promoters and compare them to several previously constructed hybrid promoters utilizing a thorough time course kinetic analysis, and analyzing the effects of alternate carbon sources on promoter activity.

We analyzed the UAS_{TEF}#2$_{(n)}$-TEF(136) series, several high expression UAS_{TEF}-truncation hybrids, the novel $UAS_{TEF(2)}$-$UAS1B_8$-TEF promoter, and our previously

constructed strong hybrids promoters UAS1B$_8$-TEF(406), UAS1B$_{16}$-TEF(406), UAS1B$_{16}$-Leum, and UAS1B$_{32}$-Leum [2] via hrGFP flow cytometry after one, two, three, and four days of growth to discern effects of cell phase on promoter activity. UAS$_{TEF}$#2-TEF(136) promoters demonstrated fairly constitutive activity, peaking in expression levels after two days before decreasing **(Figure 3.5a)**. Interestingly, the UAS$_{TEF}$#2$_{(6)}$-TEF(136) and UAS$_{TEF}$#2$_{(12)}$-TEF(136) promoter attain nearly identical peak expression level, but the UAS$_{TEF}$#2$_{(12)}$-TEF(136) promoter retains full activity much longer. As the UAS$_{TEF}$#2$_{(12)}$-TEF(136) promoter retained full activity much longer than other library members, increasing UAS$_{TEF}$#2 copy number may serve to delay the onset of reduced promoter activity. Hybrid promoters based on the UAS$_{TEF}$ truncations displayed similar constitutive expression profiles, with the UAS$_{TEF}$#2-TEF(136) promoting the highest expression levels **(Figure 3.5b)**. While stronger than the UAS$_{TEF}$-based hybrids, the UAS1B$_{16}$-Leum and UAS1B$_{32}$-Leum promoters were not highly activated until the third day of growth, confirming prior observation that the UAS1B element is most active in early stationary phase [53,96]. The addition of a TEF$_{UAS}$ element in the form of a TEF core promoter (UAS1B$_8$-TEF and UAS1B$_{16}$-TEF) shortened this lag period and facilitated quicker promoter activation as the UAS1B$_8$-TEF(406) and UAS1B$_{16}$-TEF(406) promoters displayed very high fluorescence levels and were fully expressed after only two days. The UAS$_{TEF(2)}$-UAS1B$_8$-TEF promoter reached maximal expression capacity quicker than any other tested promoter, further demonstrating the benefit of incorporating multiple, distinct UAS elements into the same hybrid promoter **(Figure 3.5c)**.

Figure 3.5: Kinetic analysis of hybrid promoters

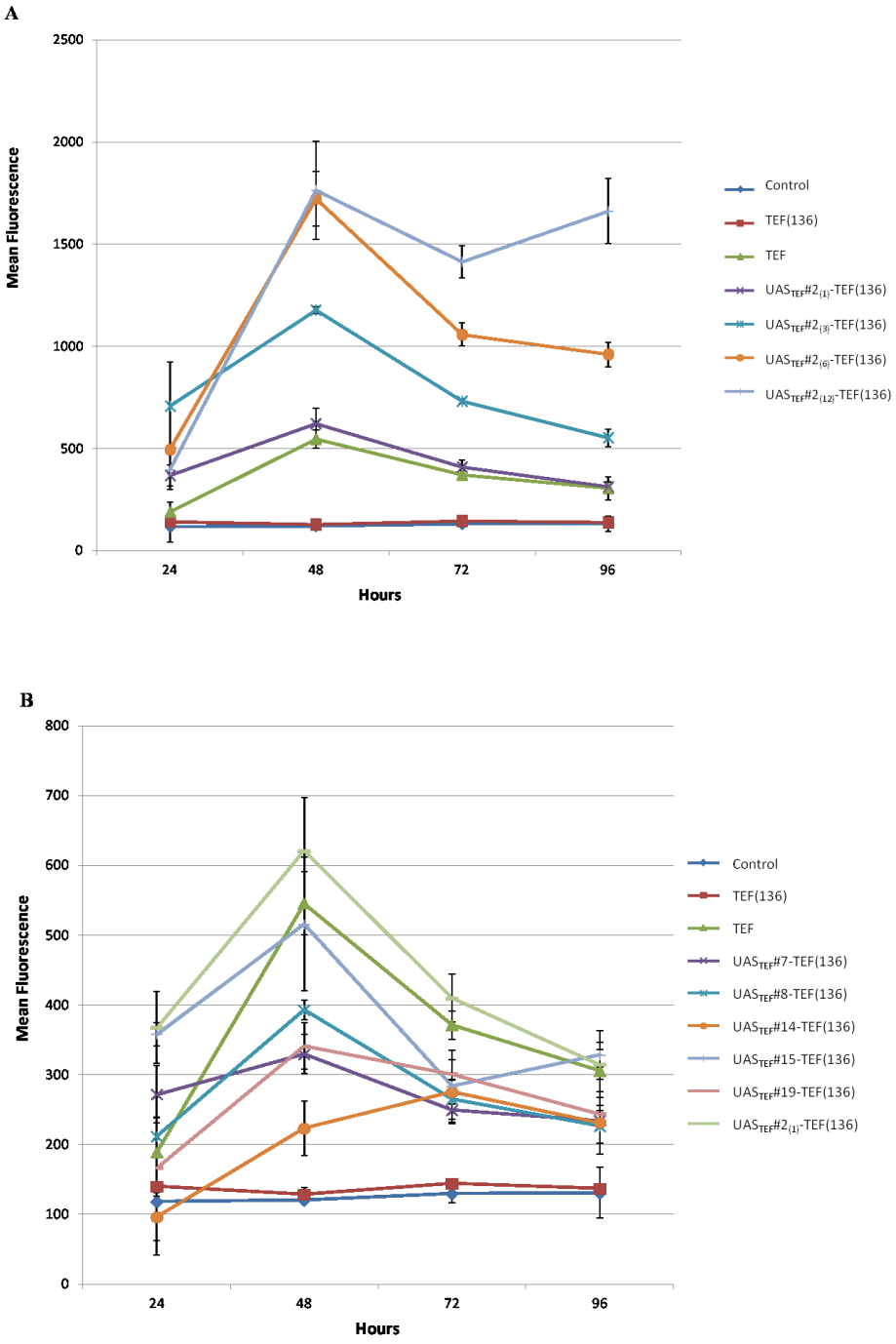
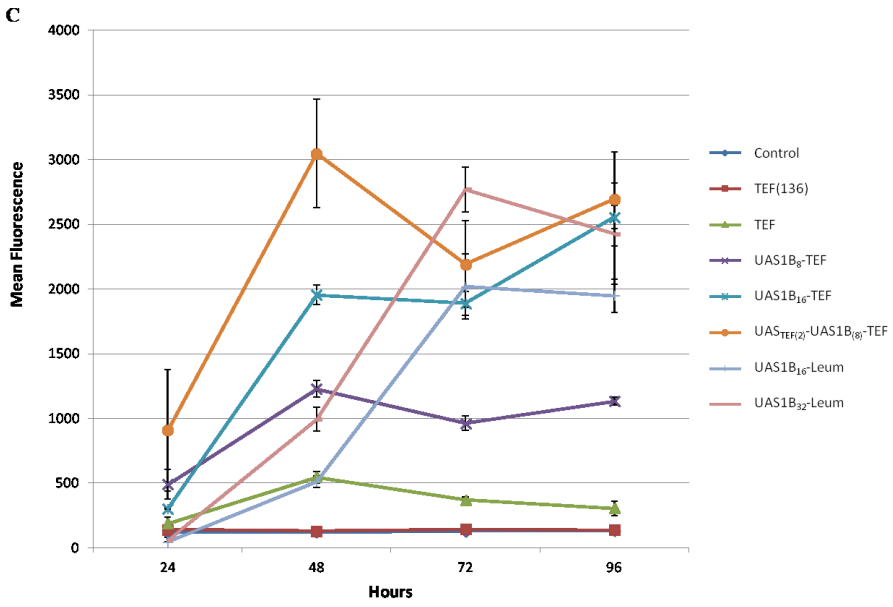


Figure 3.5 continued



(A) Relative fluorescence values indicate that tandem copies of UAS_{TEF}#2 showed progressive increased in transcription capacity. **(B)** Various UAS_{TEF} truncation hybrid promoters (from those described in **Figure 3a**) displayed similar expression profiles, with the UAS_{TEF}#2-TEF(136) promoting the highest expression levels. **(C)** $UAS1B_{16}$-Leum and $UAS1B_{32}$-Leum promoters were not highly activated until the third day of growth. $UAS1B_8$-TEF(406) and $UAS1B_{16}$-TEF(406) promoters displayed very high fluorescence levels and were fully expressed after only two days, and the $UAS_{TEF(2)}$-$UAS1B_8$-TEF promoter exhibited the high expression capacity quicker than any other tested promoter. Error bars represent the standard deviation of measurements between biological triplicates.

To analyze the effects of media on promoter expression capacity, we tested many of the hybrid constructs for protein expression when grown in media utilizing glucose, sucrose, glycerol, or oleic acid as the sole carbon source. UAS_{TEF}#2-TEF(136) promoters were minimally impacted by carbon source, with sucrose enabling the highest expression, and glycerol the lowest expression capacity **(Figure 3.6a)**. Thus, UAS_{TEF}#2-based hybrid promoters retained the constitutive expression properties found in the native TEF promoter and exhibited consistent expression levels regardless of media formulation or cell phase. In contrast, promoters containing UAS1B elements were very strongly dependent on media composition, with highest expression on sucrose or oleic acid

(Figure 3.6b). Thus, hybrid promoters constructed with the newly isolated TEF UAS were more constitutive than previous *Y. lipolytica* hybrid promoters.

Sucrose-activated UAS1B hybrid promoters yielded surprisingly high expression values compared to glucose, presenting it as an alternative carbon source for protein expression. However, maximum cell density was significantly decreased in sucrose-grown cultures. As sucrose is a disaccharide comprised of glucose and fructose, fructose-grown cultures were also analyzed for protein expression capability, but generated less expression than glucose-grown cultures. Thus, further research is necessary to determine the mechanism for such high heterologous protein expression in sucrose-grown cultures.

Figure 3.6: Effect of media formulation on hybrid promoter expression

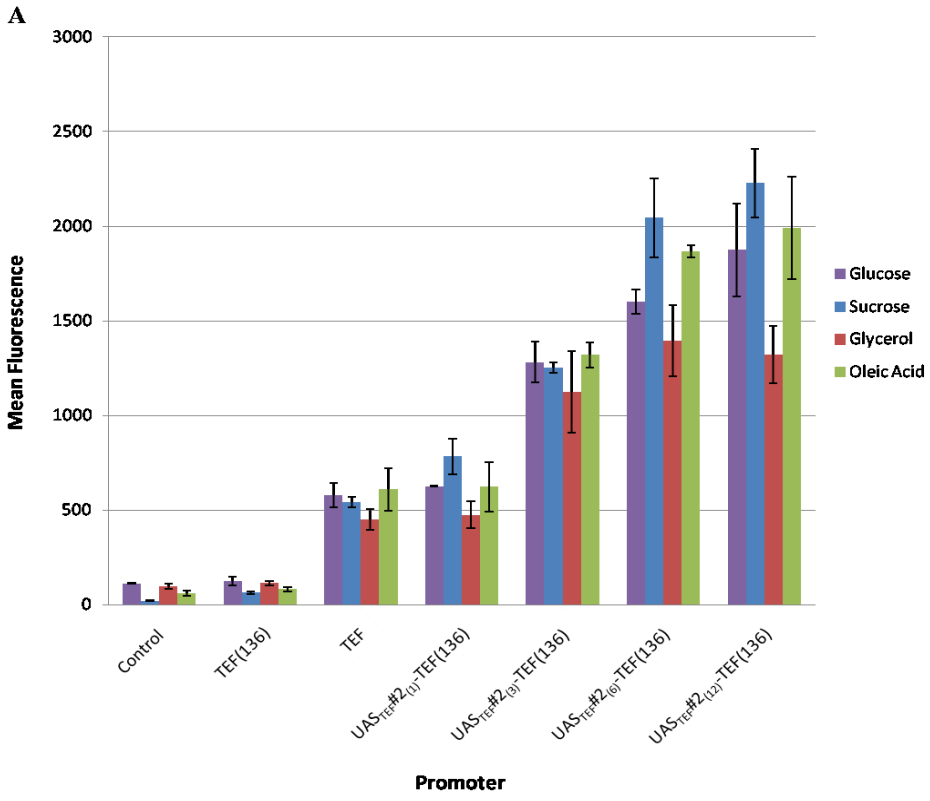
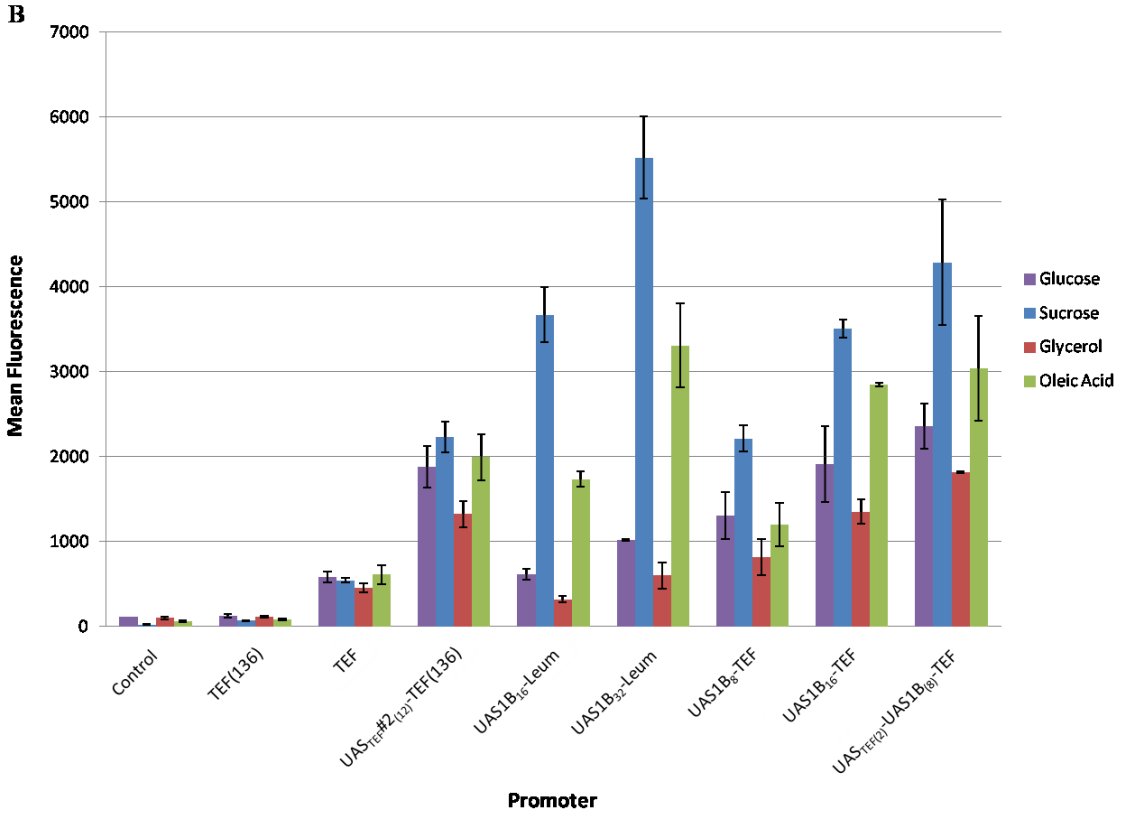


Figure 3.6 continued



(A) Hybrid constructs were tested in medium containing glucose, sucrose, glycerol, or oleic acid as the sole carbon source. UAS$_{TEF}$#2-TEF(136) promoters did not significantly vary with carbon source. **(B)** UAS1B promoters were further activated when grown on sucrose or oleic acid, and had lower levels when grown on glycerol. Error bars represent the standard deviation of measurements between biological triplicates.

3.4 Summary and Conclusion

Collectively, these results illustrate a detailed example of hybrid promoter library construction to increase promoter strength in an organism. The basic architecture of hybrid promoters informs the need for both upstream activation sequences and core promoter regions. Dissection analyses of native promoters can mediate the efficient isolation of novel UAS elements, and UAS elements isolated from native promoters can be utilized to increase available promoter strength within an organism. Tandem UAS

elements help bypass the enhancer limited nature of promoters by serving as transcriptional amplifiers, and fusion of UAS repeats to native promoters elevates their basal strength, and therefore cell-wide transcriptional capacity. We have demonstrated that this amplification is generic for all promoters (regardless of length) and thus not restricted to minimal promoters. Nevertheless, minimal promoters can be isolated through a general truncation and UAS replacement analysis. Minimal promoters increase the ultimate dynamic range of the hybrid promoter approach and allow for fine-tuned gene expression starting at a lower level. Exploitation of both native and minimal promoters under the control of tandem UAS elements permits an otherwise unattainable range of gene expression. Moreover, combining unrelated UAS sequences offers tantalizing potential for ever higher levels of gene expression and controllable regulation. In conclusion, the generic approach for hybrid promoter engineering advanced here is an important synthetic biology method enabling the construction of high-level and fine-tuned promoters, and it represents a unique method to enable high-level overexpression in organisms with ill-defined promoter elements.

Chapter 4: Applying the Hybrid Approach to *Saccharomyces cerevisiae*

4.1 Chapter Summary

Hybrid promoter engineering can enable enhancement of promoter strength in *Y. lipolytica* [1,2]. We sought to establish this approach as a versatile metabolic engineering tool to increase cellular expression capacity by applying it to another host organism. Thus, we applied the synthetic hybrid promoter approach to the model yeast, *Saccharomyces cerevisiae*, for the creation of strong promoter libraries. We demonstrate the utility of this approach with three main case studies. First, we establish a dynamic range of constitutive promoters and in doing so expand transcriptional capacity of the strongest constitutive yeast promoter, P_{GPD}, by 2.5-fold in terms of mRNA levels. Second, we demonstrate the capacity to impart synthetic regulation through adding galactose activation and removing glucose repression. Third, we establish a collection of galactose-inducible hybrid promoters that span a nearly 50-fold dynamic range of galactose-induced expression levels and increase the transcriptional capacity of the Gal1 promoter by 15% (P-value = 0.027). These results demonstrate that promoters in *S. cerevisiae*, and potentially all eukaryotes, are enhancer limited and a synthetic hybrid promoter approach can expand, enhance, and control promoter activity. These results firmly establish hybrid promoter engineering as a novel tool to enhance promoter strength in microbial hosts.

4.2 INTRODUCTION

Well-characterized promoters are essential for pathway engineering and synthetic biology efforts in the model eukaryotic yeast *Saccharomyces cerevisiae*. Specifically, two main types of promoter elements are required: (1) a series of constitutive promoters exhibiting a dynamic range of expression capacities and (2) well-controlled inducible systems with defined expression outputs. This capacity has largely been amassed through the isolation and characterization of numerous native promoters [26,38,112-121]. More recently, synthetic promoter libraries have been developed for fine-tuned transcriptional control in constitutive [30,31] and inducible manners [27,32,122-124]. However, these techniques are limited by their propensity to generate promoters weaker than the starting promoter sequence. In all of these cases, transcriptional capacity remains ultimately bounded by the transcriptional potency of the strongest endogenous promoters: in yeast, the constitutive glyceraldehyde-3-phosphate dehydrogenase promoter (P_{GPD}) and the galactose-inducible P_{GAL} promoter [125]. We seek to remove these limitations in *S. cerevisiae* for both constitutive and inducible promoters by with the promoter approach that we have recently demonstrated in the yeast *Yarrowia lipolytica* [1,2].

Many native promoters have been well characterized in yeast. Constitutive promoters offer fairly constant gene expression levels at the single-cell level without the need for specific inducers or media formulations [125]. Several well-established and commonly used variants, in order of strength, include the strongest constitutive promoter in yeast, P_{GPD} (also known as P_{TDH3}) [125], the promoter for translation elongation factor, P_{TEF} [117], and the fairly weak partial promoter region of iso-1-cytochrome C, P_{CYC} [126,127].

Inducible promoters offer a complementary method for controlling gene expression typically through the use of small molecule inducers. The most widely employed inducible promoters for pathway engineering and protein expression applications are the GAL promoters [113,125,128-130]. Galactose inducible promoters, including P_{GAL}, collectively exhibit tight control with 1000-fold induction ratios between galactose-induced and glucose-repressed expression [131]. To avoid strict on/off states with this promoter, genomic knockouts enabled a galactose dose dependent induction response of the P_{GAL} promoter [132]. Collectively, these core promoters are among the most commonly used in yeast.

Our hybrid promoter engineering approach entails combining core promoters with UAS elements to enable fine-tuned control and amplification of gene expression. Several upstream promoter sequences have previously been identified as transcriptional enhancing elements in *S. cerevisiae*, including a 240 basepair sequence 5' of the mitotic cyclin CLB2 gene [133], a 275 bp sequence 5' of the mitochondrial citrate synthase CIT1 gene [134], and a galactose-controlled 309 bp sequence 5' of the GAL1 gene, [29] referred to henceforth as UAS_{CLB}, UAS_{CIT}, and UAS_{GAL}, respectively (**Table 4.1**). These elements serve as the starting point for this study.

As described above, the field is limited in techniques to both amplify promoter activity and modulate the degree of inducibility of promoters. These limitations ultimately prevent the development of comprehensive libraries of both constitutive and inducible promoters. Based on our previous efforts in *Y. lipolytica* [2], we hypothesize that the expression capacities of even P_{GPD} and P_{GAL} can be enhanced. Hybrid promoters

have been previously used as a promoter dissection tool in *S. cerevisiae* [33,35-40]. Here, we advance the synthetic hybrid promoter approach to the yeast *S. cerevisiae* by creating several synthetic promoter libraries to increase overall cellular promoter strength.

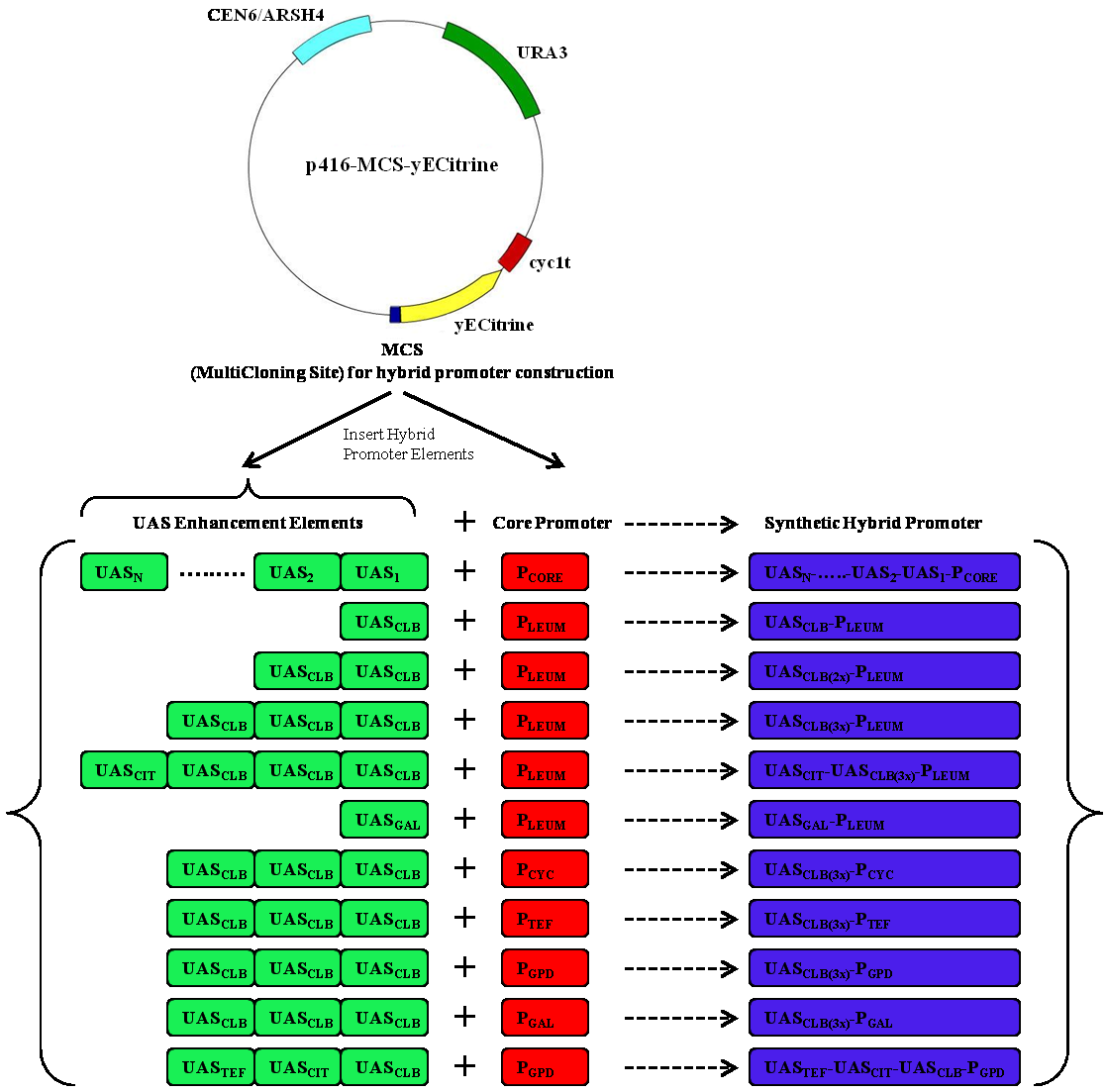
4.3 Results and Discussion

4.3.1 Creating hybrid promoters in S. cerevisiae

As a first proof-of-concept demonstration of synthetic hybrid promoter engineering in *S. cerevisiae*, several of the identified upstream activating sequences were fused to a minimal core promoter echoing our previous trials in *Y. lipolytica* [2]. To do so, a 125 nucleotide truncation of the LEU2 promoter, called P_{LEUM}, was selected as a starting point due to its low constitutive expression levels and lack of observed catabolite-repression [37,38]. While increasing tandem UAS copy number to upwards of twenty was successful in *Y. lipolytica* [2], the highly efficient native homologous recombination machinery makes similarly-sized libraries in *S. cerevisiae* impossible. As a result, a series of P_{LEUM}-based hybrid promoters were created by fusing between one and three UAS_{CLB} or UAS_{CIT} elements to the minimal core promoter (**Figure 4.1 and Table 4.1**). To assess possible synergies resulting from combining distinct UAS elements, a UAS_{CLB} element was added to the $UAS_{CIT(3x)}$-Leum hybrid promoter, creating a promoter contain four UAS elements. Likewise, a UAS_{CIT} was added to the $UAS_{CLB(3x)}$-Leum hybrid promoter construct. This collection of synthetic UAS-Leum hybrid promoters was tested for yECitrine fluorescence via flow cytometry analysis

(**Figure 4.2a**). A linear increase in fluorescence was seen with increasing numbers of UAS elements. Moreover, both promoter constructs containing four UAS elements linked to the P_{LEUM} achieved promoter strengths comparable to the highest endogenous yeast constitutive promoter, P_{GPD}.

Figure 4.1: Construction of *S. cerevisiae* hybrid promoters



A simplified schematic is provided detailing the construction of hybrid promoter-based fluorescence cassettes within the p416-MCS-yECitrine plasmid backbone. A table of hybrid constructs includes a generalizable hybrid naming pattern (top row) along with key hybrid promoters discussed in this paper highlighted.

Table 4.1: List of upstream activation sequences used in *S. cerevisiae*.

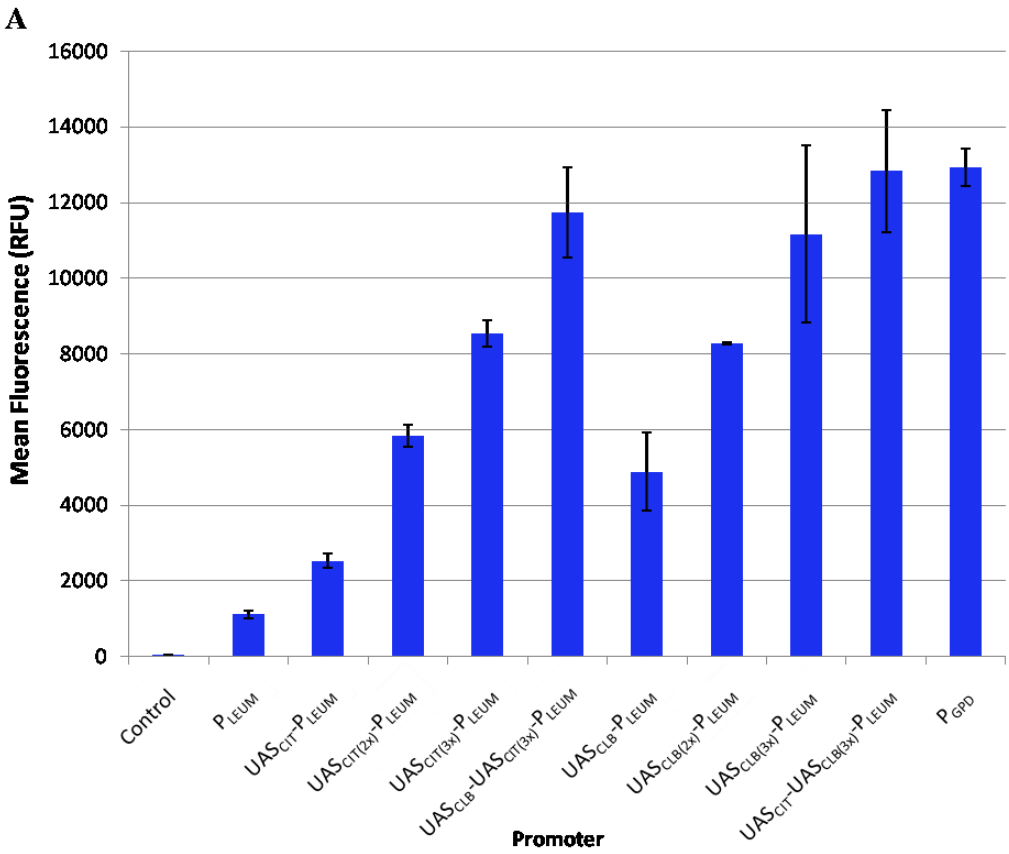
Upstream Activating Sequence	Gene (s) Activated	Inclusive Basepairs (5' of Start Codon)	Length (Basepairs)	Citation
UAS_{CLB}	*CLB2*	-867 to -627	240	[133]
UAS_{CIT}	*CIT1*	-548 to -273	275	[134]
UAS_{GAL}	*GAL1/ GAL10*	-457 to -148 (of GAL1)	309	[29]
UAS_{TEF}	*TEF1*	-401 to -198	203	This work

The UAS elements used in this study are listed with their names, open reading frame regulated, basepair range, size, and reference.

A second promoter library was created utilizing a minimal core promoter derived from the P_{CYC} native promoter, truncated to 158 basepairs in length and dubbed P_{CYC158}. This library exhibited nearly identical characteristics as the P_{LEUM}-based hybrid library, with a nearly linear trend in fluorescence levels culminating in promoter strength akin to P_{GPD} (**Figure 4.2b**). Interestingly, the addition of a third UAS_{CLB} element to $UAS_{CLB(2x)}$-P_{CYC158} did not enhance observed fluorescence levels. In both the P_{LEUM} and P_{CYC158} based hybrid promoters, the UAS_{CLB} element appeared to amplify transcription slightly better than UAS_{CIT} (**Table 4.2**). Nevertheless, these elements functioned together in tandem, raising the possibility that they may serve as movable, synthetic transcriptional enhancers independent of a core promoter choice. As a result, it would be possible to isolate and incorporate novel, comparable UAS elements to enhance transcriptional capacity without introducing homologous DNA regions. To this end, we isolated 203 basepair from the 5' section of the native P_{TEF} from *S. cerevisiae* (created by excluding the AT rich, TATA-box containing region), and fused this element, termed UAS_{TEF} (**Table 4.1**), to the P_{LEUM} promoter and confirmed the capacity of this element to amplify

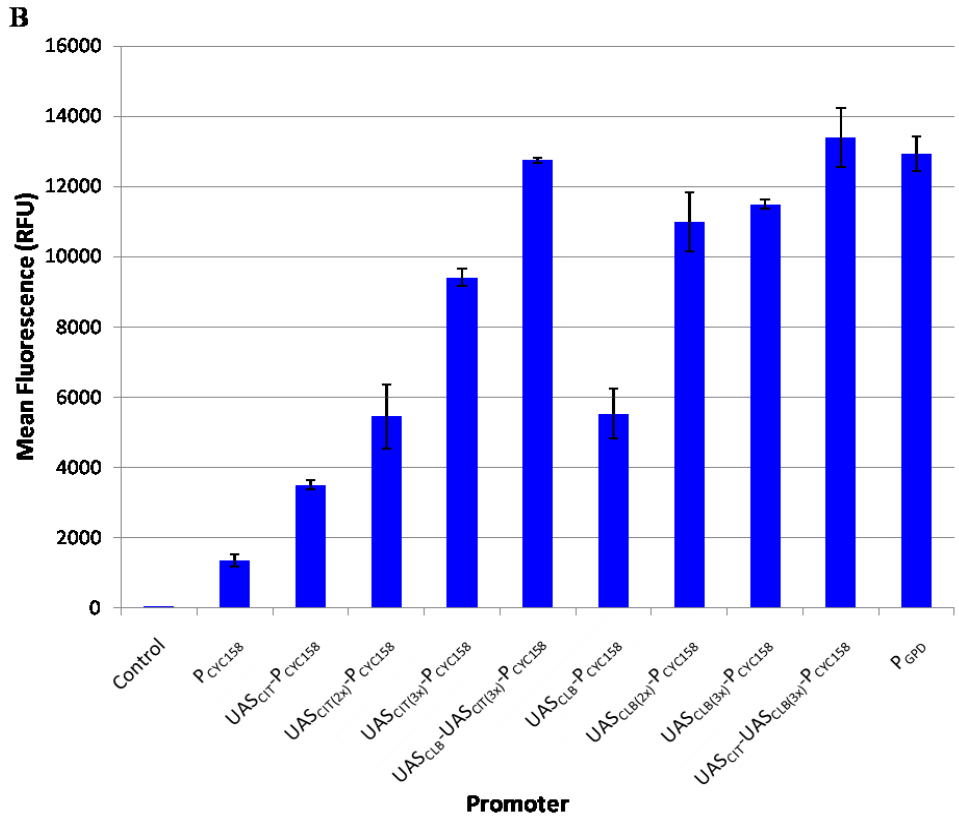
expression (**Figure 4.2c)**. These results demonstrate that UAS_{CLB}, UAS_{CIT}, and UAS_{TEF} elements can act as modular, synthetic amplifiers of transcription.

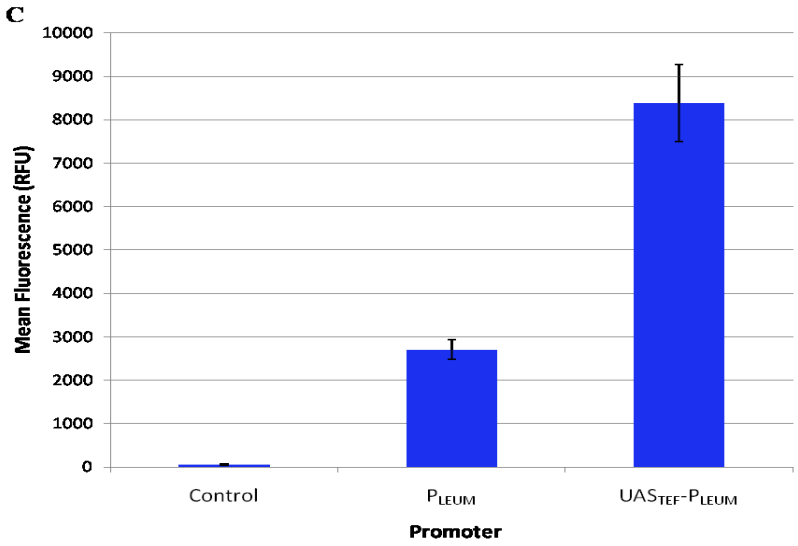
Figure 4.2: Developing UAS-Leum hybrid promoters and isolating a novel UAS element



(**A**) Between one and three copies of UAS_{CLB} or UAS_{CIT} sequences were fused upstream of the minimal P_{LEUM} and resulted in linearly increasing fluorescence values. Further insertion of 1 UAS_{CLB} element upstream of the $UAS_{CIT(3x)}$-P_{LEUM} promoter or one UAS_{CIT} element upstream of the $UAS_{CLB(3x)}$-P_{LEUM} promoter enabled expression levels on par with the strong endogenous P_{GPD} promoter. A more than 10-fold range in fluorescence levels is observed between the P_{LEUM} core promoter and P_{LEUM}-based hybrid promoters. Error bars represent standard deviation from biological triplicates.

Figure 4.2 continued



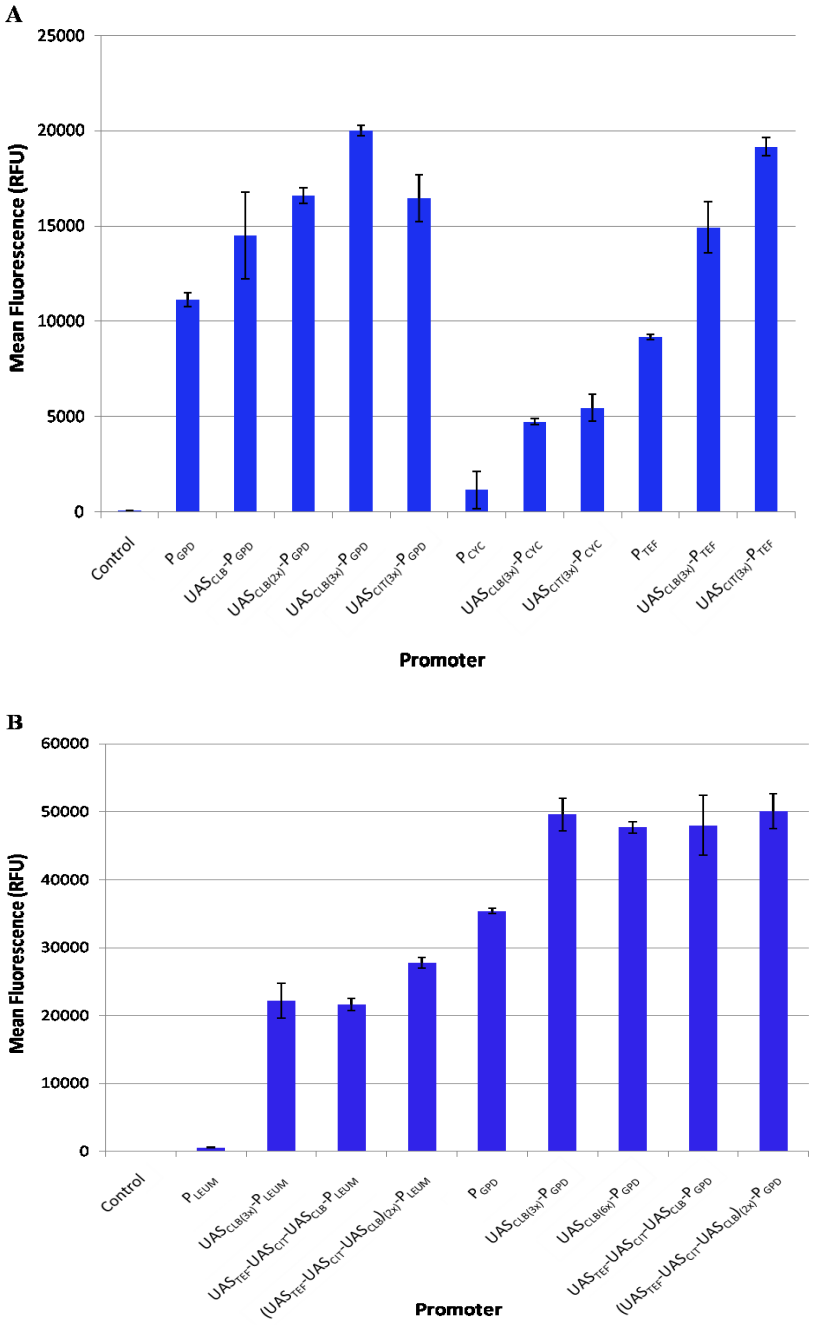


(**B**) Similar results are obtained utilizing P_{CYC158} as a core promoter, including the 10-fold range in fluorescence levels, and final promoter strength comparable to P_{GPD}. (**C**) The insertion of the novel UAS_{TEF} element upstream of the P_{LEUM} core promoter resulted in more than a 3-fold increase in fluorescence levels. The promoter-less plasmid, p416-MCS-yECitrine was included as a negative control (Control). Error bars represent standard deviation from biological triplicates.

4.3.2 Increasing transcriptional capacity of native promoters using a synthetic hybrid approach

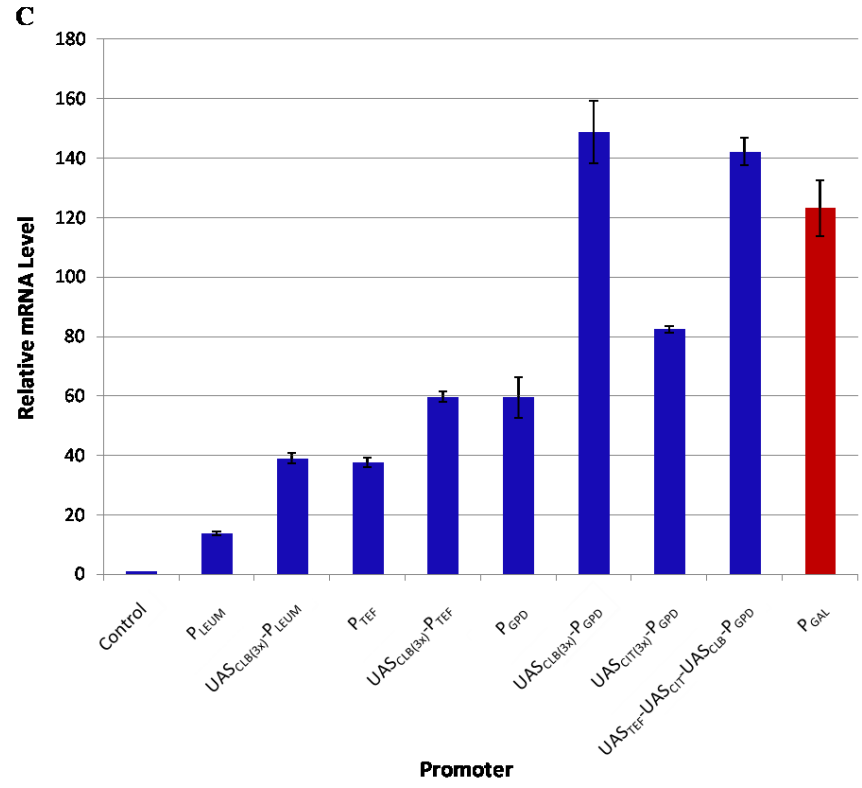
Next, we sought to test the hypothesis that even the strongest native promoters in *S. cerevisiae* are limited by a deficiency of UAS enhancer elements. To do so, we fused three UAS_{CLB} or UAS_{CIT} tandem repeats to the P_{GPD}, P_{CYC}, and P_{TEF} native promoters (**Figure 4.1**), and tested for yECitrine fluorescence (**Figure 4.3a**). Hybrid promoter fluorescence output was roughly doubled compared to basal promoter strength for P_{GPD} and P_{TEF} and more than quadrupled for P_{CYC} (**Table 4.2**). Interestingly, UAS_{CLB} seemed to amplify the P_{GPD} promoter the most whereas UAS_{CIT} elements were more effective with P_{TEF}. Both one and two tandem UAS_{CLB} elements were fused to P_{GPD} to enable a linear promoter range proportional to the number of synthetic UAS elements (**Figure 4.3a**). Finally, in an attempt to further saturate the transcriptional capacity in yeast, we fused six UAS_{CLB} activating sequences to P_{GPD}, despite the inherent instability of this repetitious sequence. As an alternative, we also fused single copies of the UAS_{CLB}, UAS_{CIT}, and UAS_{TEF} elements in tandem to the P_{LEUM} and P_{GPD} core promoters (**Figure 4.1**). This entire UAS_{CLB}-UAS_{CIT}-UAS_{TEF} fragment was also duplicated to construct hybrid promoters comprised of two sets of UAS_{CLB}-UAS_{CIT}-UAS_{TEF} enhancer elements activating the P_{LEUM} and P_{GPD} cores. Flow cytometry characterization of promoter strength (**Figure 4.3b**) depicted an evident trend towards saturation of P_{GPD} through UAS-based hybrid promoters. Moreover, the UAS_{CLB}-UAS_{CIT}-UAS_{TEF}-P_{GPD} hybrid promoter, containing no regions of homology, exhibited fluorescence levels on par with the strongest promoters constructed here.

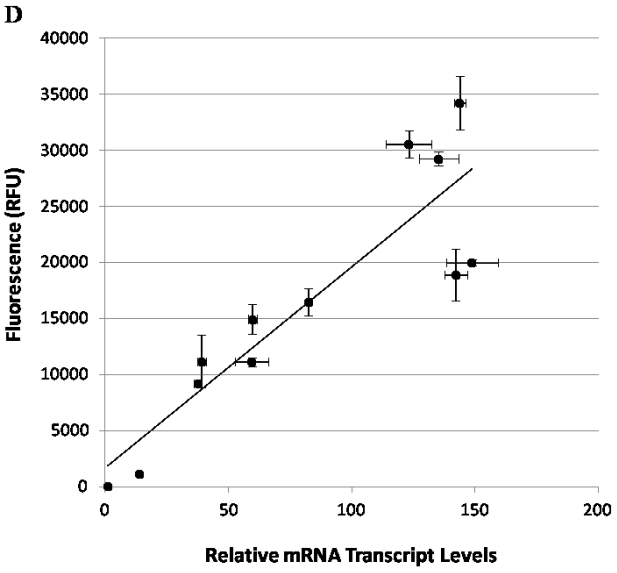
Figure 4.3: Expanding the transcriptional capacity of *S. cerevisiae*



(**A**) Insertion of three tandem repeats of either UAS_{CLB} or UAS_{CIT} upstream of endogenous promoters P_{CYC}, P_{TEF}, or P_{GPD} greatly increases expression capability. Higher and tunable expression is enabled by fusing one to three UAS_{CLB} elements to the P_{GPD}. (**A**) P_{GPD} capacity is enhanced and eventually saturated by three or more UAS_{CLB} elements, or by a UAS_{TEF}-UAS_{CIT}-UAS_{CLB} combination. P_{GPD}-based hybrid promoters represent the strongest characterized promoters in yeast, a 1.4-fold improvement. As a note, an irreversible instrument recalibration at the core facility results in the higher than previously observed fluorescence levels. Error bars represent standard deviation from biological triplicates.

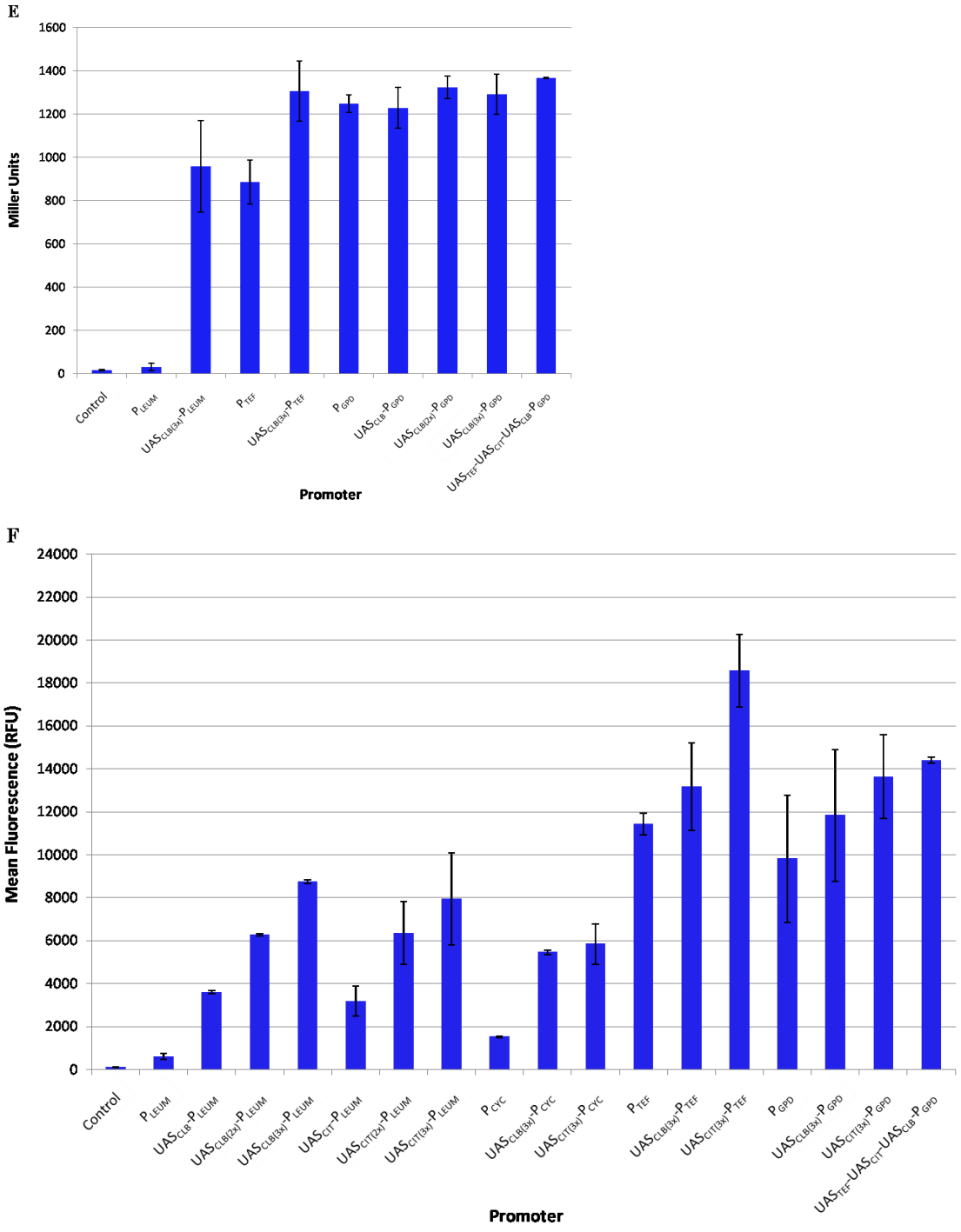
Figure 4.3 continued





(**C**) Transcriptional profiling of select promoter constructs was used to calculate mRNA levels relative to the p416-MCS-yECitrine plasmid. Expression profiles matched well with fluorescence data and the $UAS_{CLB(3x)}$-P_{GPD} and UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-P_{GPD} promoters exhibited a 2.5-fold increase in mRNA levels compared to the endogenous P_{GPD} promoter, and outperformed an induced GAL1 promoter. (**D**) Fluorescence levels correlate linearly with relative mRNA transcript levels with an $R^2 = 0.7955$. Error bars represent standard deviation from technical triplicates.

Figure 4.3 continued



(**E**) Endogenous and hybrid promoters were tested with a β-galactosidase reporter gene. (**F**) Endogenous and hybrid promoters were tested with the yECitrine reporter gene in an industrial Sake yeast to validate that enhancer elements function in different genetic backgrounds. Results agree well with prior tests with the exception of the $UAS_{CLB(3x)}$-P_{TEF} promoter performing as the strongest hybrid promoter. The promoter-less plasmid, p416-MCS-yECitrine was included as a negative control (Control). Error bars represent standard deviation from biological triplicates.

Table 4.2: Fold fluorescence increase upon addition of an enhancer element

	Enhancer Elements					
Core Promoter	UAS_{CLB}	$UAS_{CLB(2x)}$	$UAS_{CLB(3x)}$	UAS_{CIT}	$UAS_{CIT(2x)}$	$UAS_{CIT(3x)}$
P_{LEUM}	4.38	7.43	10.02	2.27	5.25	7.66
P_{CYC158}	4.07	8.09	8.45	2.57	4.01	6.92
P_{CYC}			4.15			4.79
P_{TEF}			1.63			2.09
P_{GPD}	1.30	1.49	1.80			1.48

Fold fluorescence improvement between basal core strength and enhancer-activated hybrid promoter strength is given for various constitutive hybrid promoters created in this study. Ratios correlate to fluorescence levels seen in **Figures 4.2a,b** and **4.3a**. Gray boxes signify potential hybrid promoters not constructed in this work.

A transcriptional analysis was performed to confirm that the impact of the tandem UAS elements was indeed manifested at the transcriptional level. To do so, qRT-PCR analysis was employed using the yECitrine mRNA of select promoter constructs (P_{LEUM}, $UAS_{CLB(3x)}$-P_{LEUM}, P_{TEF}, $UAS_{CLB(3x)}$-P_{TEF}, P_{GPD}, $UAS_{CLB(3x)}$-P_{GPD}, $UAS_{CIT(3x)}$-P_{GPD}, UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-P_{GPD}) (**Figure 4.3c**). Expression values were normalized to the mRNA level seen from a plasmid containing the yECitrine gene and no promoter (labeled control). Indeed, mean fluorescence levels were strongly correlated with relative mRNA levels (**Figure 4.3d**), and a 10 fold dynamic range of transcript level between the minimal promoter and strongest promoters was observed in this set. Moreover, UAS augmentation of P_{GDP} increased its transcriptional capacity 2.5 fold, and P_{GDP}-based hybrid promoters generated more mRNA transcripts than a fully induced GAL1 promoter.

Select promoter constructs, including the endogenous P_{LEUM} and P_{TEF} and their corresponding hybrid constructs (activated by three tandem UAS_{CLB} elements) were used

to construct expression cassettes with lacZ gene in place of yECitrine. The strong P_{GPD} promoter was similarly tested for basal strength and when activated by one, two, or three UAS_{CLB} elements, or the UAS_{CLB}-UAS_{CIT}-UAS_{TEF} fragment. β-galactosidase assays were performed as described previously [2,106] with a maximum value of 1368 miller units generated by the UAS_{CLB}-UAS_{CIT}-UAS_{TEF} enhanced P_{GPD} hybrid construct (**Figure 4.3e**). LacZ expression appeared saturated for promoters stronger that P_{GPD} unlike in the yECitrine assay. However, UAS_{CLB} markedly increased β-galactosidase activity for the P_{LEUM} and P_{TEF} core promoters, demonstrating UAS-mediated expression enhancement to be a generic tool for increasing gene expression in *S. cerevisiae*.

Finally, to evaluate the generalized utility of these synthetic hybrid promoters, seventeen expression cassettes were transformed into the industrial Kyokai 7 diploid Sake yeast and tested for fluorescence via flow cytometry (**Figure 4.3f**). In each of these cases, increasing the number of tandem UAS elements can improve transcriptional activity. Interestingly, the P_{TEF} native promoter generated higher fluorescence values than P_{GPD} in this genetic background. However, P_{TEF} hybrid promoters still exhibited a large increase in strength compared to the basal promoter, validating the hybrid promoter approach as an effective tool across genetic backgrounds.

Collectively, these data establish that even the strongest constitutive promoters in *S. cerevisiae* are enhancer limited and that yeast transcriptional capacity can be increased through additional UAS elements. The magnitude of transcriptional enhancement depends on both UAS element and core promoter, as seen previously in *Y. lipolytica* [2]. Regardless, all core promoters tested could be amplified through hybrid promoter

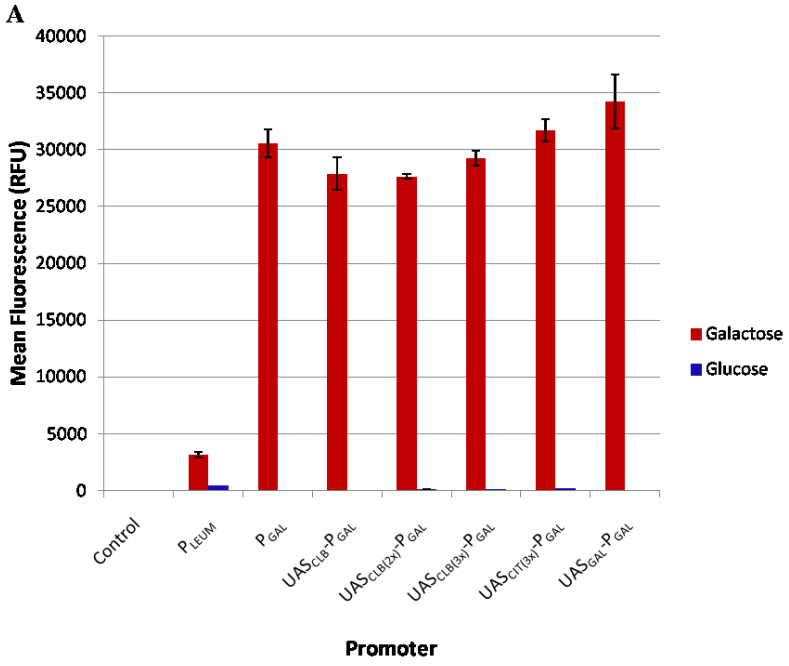
engineering. A hybrid approach amplified the strength of the strongest characterized constitutive promoter, P_{GPD}, by nearly 2.5 fold on the basis of transcript level (**Figure 4.3c)**. These levels are higher than that achieved for the induced P_{GAL} promoter. Moreover, this enhanced P_{GPD} hybrid construct was created using three distinct UAS elements which preclude the potential for homologous recombination-mediated promoter degradation. We postulate that these tandem UAS elements alleviate enhancer limitation through the localization of transcription factors, removing a major rate limiting step for transcription initiation. In this context, it is unsurprising that UAS-mediated enhancement differed between individual core promoters, as synergy (or lack thereof) between specific UAS-core hybrids could promote transcription with higher (or lower) efficacy. In each of these cases, total transcription factor availability is not seemingly a limiting factor as cells containing the strongest hybrid promoters exhibited no growth defects, making it unlikely that these novel, synthetic promoters will deplete or starve cells of transcription factors.

4.3.3 UAS-mediated derepression of the GAL1 promoter and UAS-mediated regulation of constitutive promoters

After demonstrating the utility of the synthetic hybrid promoter approach for controlling constitutive gene expression, we sought to apply our approach to control promoter regulation. By fusing UAS_{CLB} or UAS_{CIT} elements to the P_{GAL} promoter, we removed the strict on/off phenotype observed in native galactose controlled promoters. Subsequent tandem copies of UAS_{CLB} elements stimulated a linear trend derepressing the P_{GAL} promoter under glucose-repressive conditions (**Figure 4.4a,b**). In this manner, the

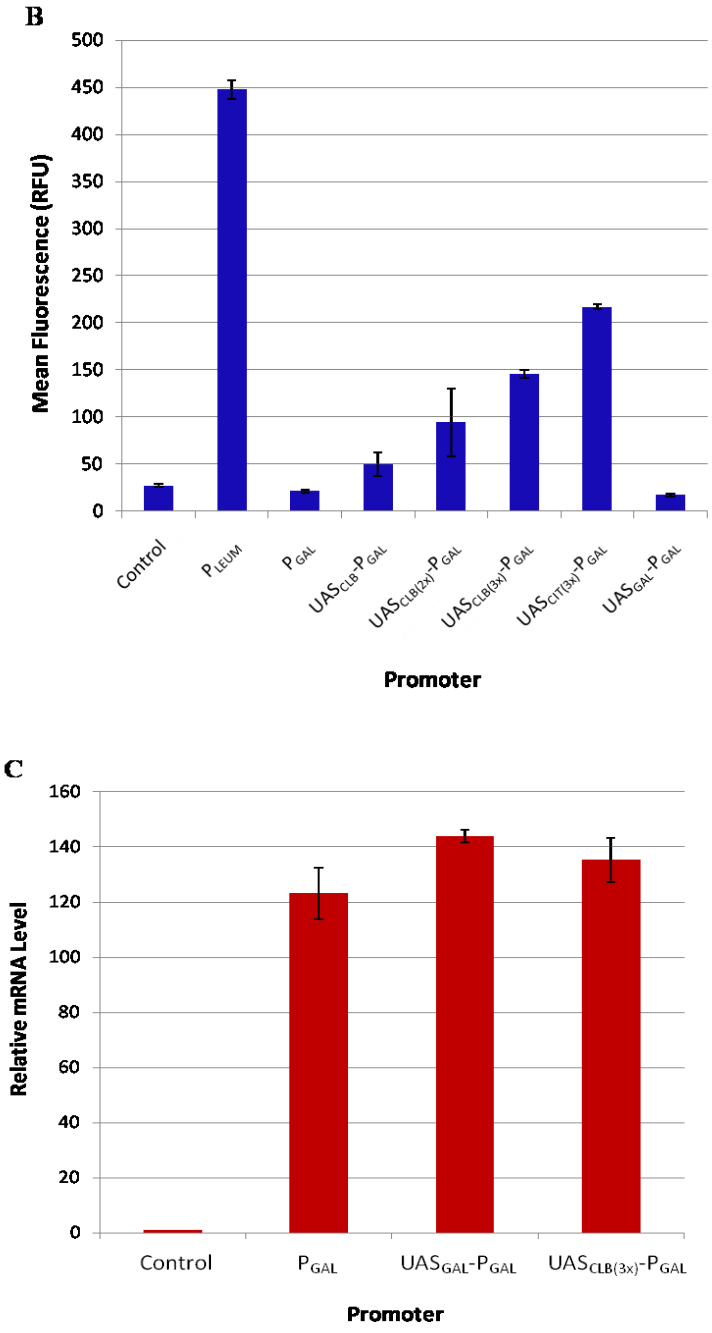
tandem UAS sites served to localize expression enhancing transcription factors upstream of the glucose-repressed GAL1 promoter and effectively created leaky hybrid promoters that allow low expressions levels even when grown on glucose. Alternatively, fusion of the GAL1-derived UAS_{GAL} element to P_{GAL} further repressed gene expression levels by 7% in glucose conditions. Moreover, this promoter exhibited a nearly 15% increase in transcript level compared to the basal P_{GAL} promoter, validating the capacity of UAS elements to synthetically amplify transcriptional capacity (**Figure 4.4c**).

Figure 4.4: Engineering promoter regulation in *S. cerevisiae*



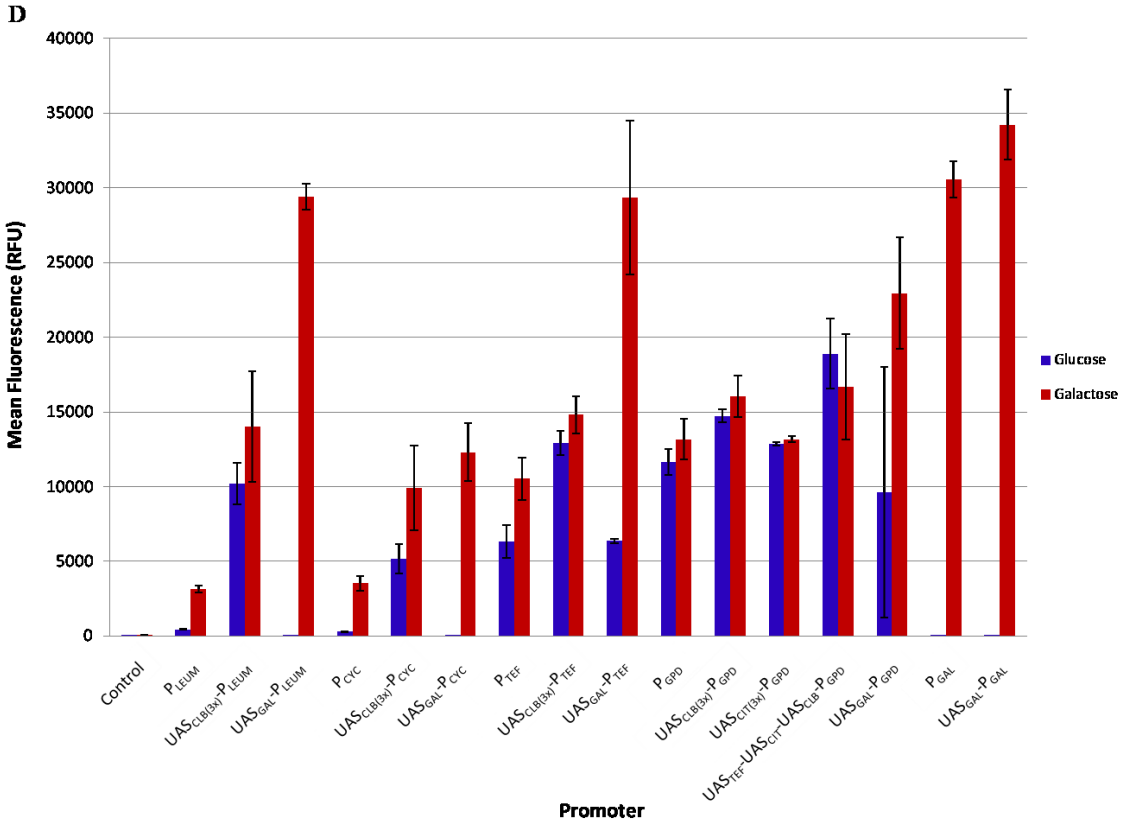
(**A**) Characterization of promoter P_{GAL}-based hybrid promoter strength shown for both galactose-induced and glucose-repressed conditions. Error bars represent standard deviation from biological triplicates. P-value comparing UAS_{GAL}-P_{GAL} to P_{GAL} = 0.054.

Figure 4.4 continued



(**B**) Characterization of promoter P_{GAL}-based hybrid promoter strength shown only for glucose-repressed conditions. Fusing enhancing UAS elements to the P_{GAL} promoter causes leaky expression on glucose, while fusing the UAS_{GAL} element to P_{GAL} further represses expression. (**C**) Transcriptional profiling of select promoter constructs was used to calculate mRNA levels relative to the p416-MCS-yECitrine plasmid. Expression profiles matched well with fluorescence data. The UAS_{GAL}-P_{GAL} and $UAS_{CLB(3x)}$-P_{GAL} hybrid promoters were more highly expressed that the basal P_{GAL}. Error bars represent standard deviation from biological triplicates in (**B**) and technical replicates in (**C**). P-value comparing UAS_{GAL}-P_{GAL} to P_{GAL} = 0.027.

Figure 4.4 continued



(**D**) Galactose-induced and glucose-repressed fluorescence levels are shown for various constitutive hybrid promoters, with high expression similarity across carbon sources. UAS_{GAL} fusions to the endogenous promoters P_{CYC} and P_{LEUM} created hybrids strongly induced galactose and fully repressed by glucose. UAS_{GAL} fusions to the endogenous promoters P_{TEF} and P_{GPD} created hybrids strongly induced galactose and unrepressed by glucose. The promoter-less plasmid, p416-MCS-yECitrine was included as a negative control (Control). Error bars represent standard deviation from biological triplicates.

We next sought to evaluate whether inclusion of a galactose-inducible UAS element can impart synthetic regulation to standard, constitutive core promoters. To do so, the UAS_{GAL} element was fused 5' of various constitutive promoters including P_{TEF}, P_{GPD}, P_{CYC}, and P_{LEUM} and resulting fluorescence was tested under both glucose-repressive and galactose-activating conditions (**Figure 4.4d**). Additionally, several constitutive hybrid promoters were also tested in the same flow cytometry run to enable comparisons in these two conditions (**Figure 4.4d**). As expected, these constitutive hybrid promoters performed similarly across the two carbon sources. In contrast, the

UAS_{GAL} was able to impart galactose-induced activation to core promoters. Interestingly, the UAS_{GAL}-P_{LEUM} hybrid promoter appears quite similar to the native P_{GAL}, potentially owing to relative activation of P_{LEUM} by galactose. By fusing the UAS_{GAL} element to P_{TEF} and P_{GPD} core regions, we created hybrid promoters that were quite active when grown on glucose and further induced by galactose. We hypothesize that transcriptional repressors normally recruited by the UAS_{GAL} element were incapable of stymieing constitutive expression, potentially due to *in vivo* occupation of nearby binding regions by other transcription factors. This effect has been observed in the context of the UAS_{CLB} element, in which binding by the essential transcription factor Reb1p prohibited Gal4p binding and subsequent transcriptional activation [133]. The result is a galactose-inducible promoter without pure on/off functionality.

In general, a synthetic hybrid promoter approach can be used to impart novel promoter traits of regulation. The UAS_{GAL} synthetic element could efficiently introduce novel regulation to constitutive promoters, while glucose-repressed expression from the native GAL1 promoter could be enhanced or reduced by incorporating synthetic UAS elements. Thus, hybrid promoter regulation and potency can be controlled through the choice of enhancer element or core promoter.

4.3.4 Construction of a tunable galactose-inducible promoter library

As a final demonstration of hybrid promoter engineering in *S. cerevisiae*, we sought to create a continuum of galactose-induced expression to change the magnitude of the "on" state for this promoter. Modifications described above instilled galactose control, but gene expression was still limited to discreet points of either very high or very

low expression. Of the galactose-controlled hybrid promoters described above, only the UAS_{GAL}-P_{CYC} fusion exhibited reduced galactose induction. Thus, we sought to control this induction level by constructing two partial dissections of the 5' region of the P_{CYC} promoter (named CU1 and CU2) to reduce UAS_{GAL} activation of a P_{LEUM} core promoter (**Figure 4.5a**). These upstream regions of the CYC1 promoter tempered galactose-induced expression, beginning to fill in gaps in the expression continuum (**Figure 4.5b**).

The remainder of the continuum seen in **Figure 4.5b** was enabled through the construction of P_{LEUM}-based hybrid promoters fused to portions of Gal4p binding sites. Previously, a 54 nucleotide fragment was confirmed as adequate and essential to impart the galactose-induced phenotype associated with the GAL1 promoter [29]. Here, we dissected this region by examining four putative Gal4p binding site DNA motifs (hereby named G4BS1-4) identified through a transcription factor motif search using the Promoter Database of *Saccharomyces cerevisiae* [21]. Putative Gal4p binding sites were fused individually or in combination to the P_{LEUM} core promoter (**Figure 4.5a**), generating a library of synthetic hybrid promoters differing in expression when activated by galactose (**Figure 4.5b**). The capacity of these Gal4p binding sites to mediate effective galactose induction was as follows: G4BS2 < G4BS1 < G4BS3 = G4BS4. Employing single or multiple Gal4p binding sites as hybrid enhancer elements enabled differential levels of expression, and produced a library of promoters that fully sampled the range of galactose induction. Only the inclusion of all Gal4p binding sites permitted full galactose induction, implying that these sites are bound *in vivo* by separate tandem Gal4 proteins. Collectively, this collection of hybrid galactose inducible promoters spans

a relative yECitrine expression range of nearly 50-fold without the need for any genetic changes to the strain. This complete coverage of intermediate expression levels by our hybrid promoter library fully enables fine-tuned inducible expression, and will facilitate pathway engineering and applications in yeast.

Figure 4.5: Engineering precise control of galactose-induced expression

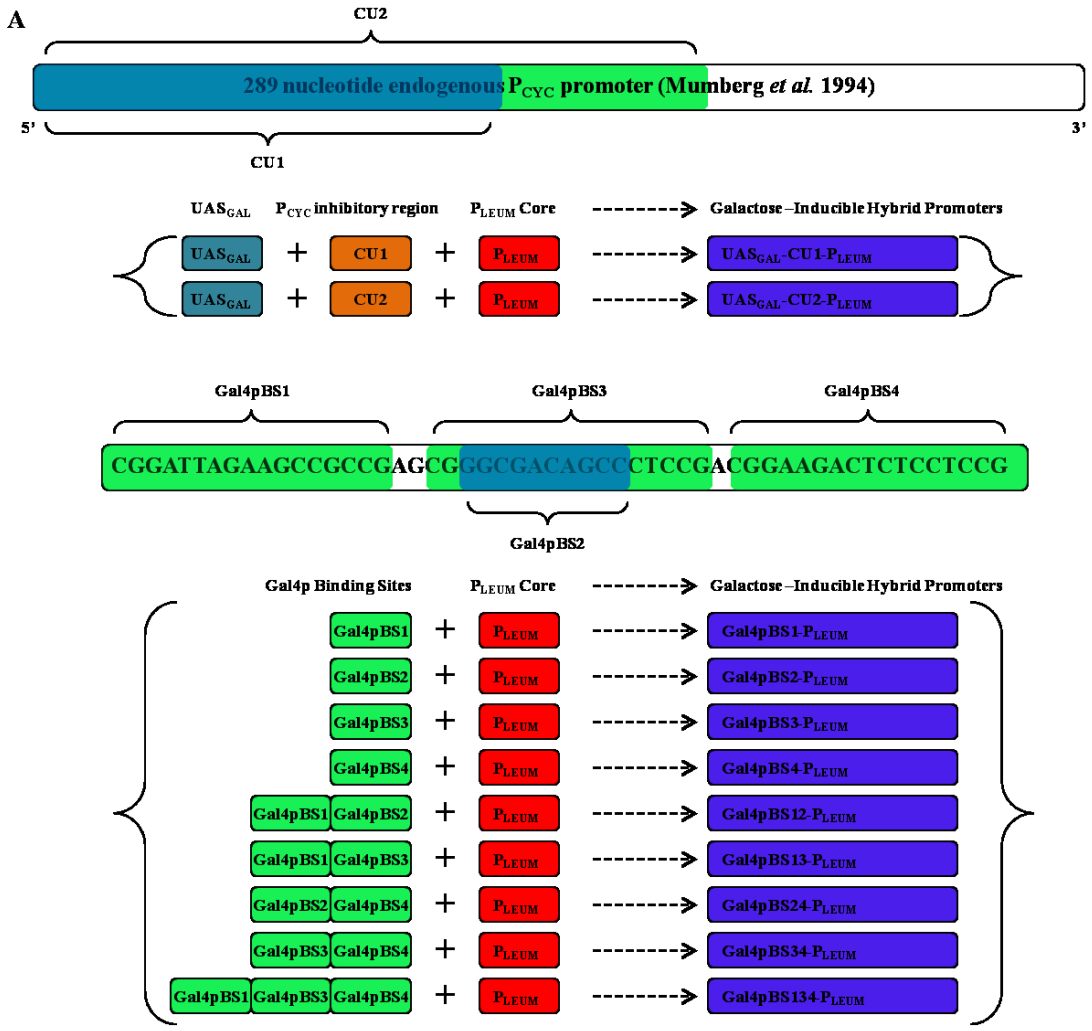
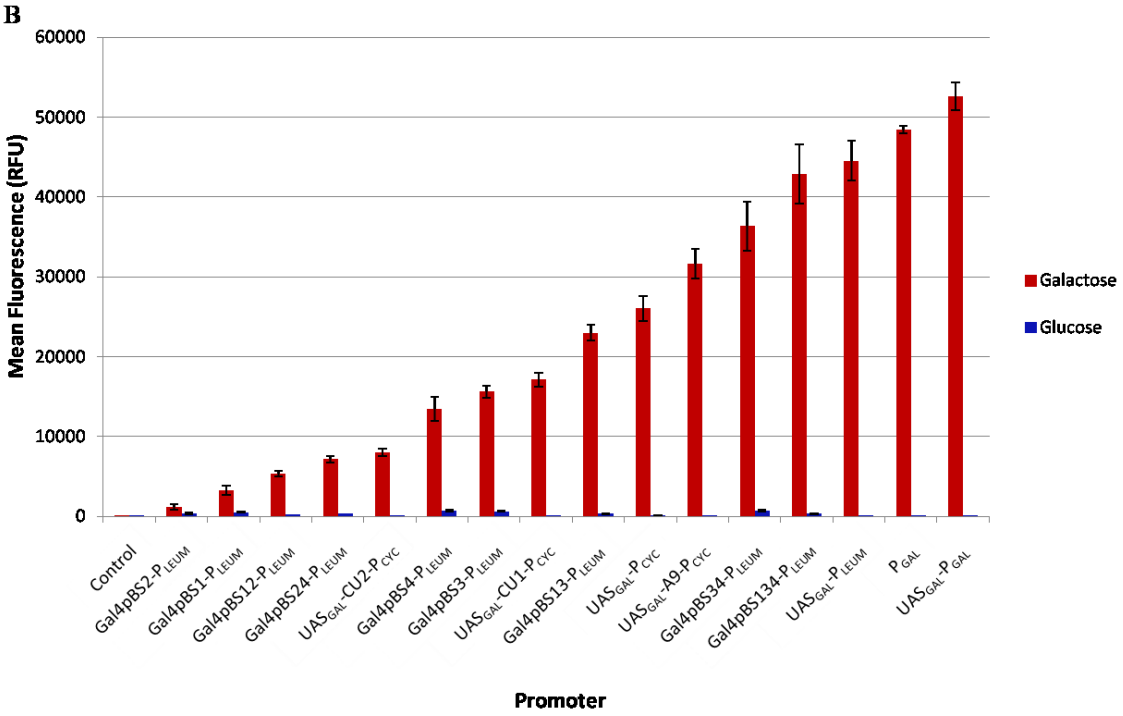


Figure 4.5 continued



(**A**) A simplified schematic picture is provided, demonstrating the construction of hybrid promoters utilizing the CU1 or CU2 regions to inhibit UAS_{GAL} activation of a P_{LEUM} core promoter (Top). The figure also details the dissection of a key 54bp sequence that confers galactose induction (Middle) and shows the construction of Gal4p binding site controlled hybrid promoters (Bottom). (**B**) The characterization of a library of hybrid promoters that enable a 50-fold range of controlled galactose-induced expression is shown. This novel library includes Gal4pBS-P_{LEUM} hybrid promoters, endogenous promoters controlled by the UAS_{GAL} element, and hybrid promoters containing DNA inserted between a UAS_{GAL} element and a P_{LEUM} core to decrease galactose induction. Fluorescence levels observed from cultures grown on galactose are shown in red, and glucose in blue. The promoter-less plasmid, p416-MCS-yECitrine was included as a negative control (Control). Error bars represent standard deviation from biological triplicates.

4.4 Summary and Conclusions

Here, we have advanced a generic synthetic hybrid promoter engineering approach into *S. cerevisiae*, and in the process demonstrated that native yeast promoters are enhancer limited. Obtaining high heterologous protein expression levels necessitates strong promoters, while pathway engineering applications demand precise control of gene expression levels to optimize pathway throughput. By augmenting native promoters

with modular transcriptional amplifiers, we have produced the strongest promoters yet characterized in yeast and have also created a range of expression capacities. We further enabled fine-tuned galactose-induced expression by fusing Gal4p binding sites to a minimal core promoter. Given the results of this chapter and the prior two chapters constructing and studying *Y. lipolytica* hybrid promoters [1,2], it is conceivable that transcriptional capacity is enhancer limited in all eukaryotes, and the hybrid promoter approach could facilitate removal of this limitation. In particular, the hybrid promoter approach is an ideal methodology to enable strong expression in organisms lacking well characterized, high strength promoters.

Chapter 5: Heterologous Production of Pentane in *Yarrowia lipolytica*

5.1 CHAPTER SUMMARY

The complete biosynthetic replacement of petroleum transportation fuels requires metabolic pathways capable of producing alkanes. Microbial production of long-chain alkanes has recently been described, but current biofuel technology lacks a complementary pathway to produce short-chain *n*-alkanes [77]. *Yarrowia lipolytica* is an ideal organism for oleo-chemical production due to its capacity to accumulate hydrophobic carbon content. Having developed high strength-promoters designed to control metabolic flux in *Y. lipolytica,* we attempted to redirect lipid synthesis flux towards short-chain alkane production. Here, we report and characterize a proof-of-concept pathway that enables production of the C_5 *n*-alkane, pentane, in *Y. lipolytica*. This pathway utilizes a soybean lipoxygenase enzyme to cleave linoleic acid to pentane and a tridecadienoic acid byproduct. Initial expression of the soybean lipoxygenase enzyme within a *Yarrowia lipolytica* host yielded 1.56 mg/L pentane. Efforts to improve pentane yield by increasing substrate availability and strongly overexpressing the lipoxygenase enzyme successfully increased pentane production three-fold to 4.98 mg/L. This represents the first-ever microbial production of pentane and demonstrates that short chain *n*-alkane synthesis is conceivable in microbial hosts. In this regard, we demonstrate the potential pliability of *Y. lipolytica* towards the biosynthetic production of value-added molecules from its generous fatty acid reserves.

5.2 INTRODUCTION

The complete replacement of petroleum-derived liquid transportation fuels can be achieved in one of two means: by using an alternative, drop-in fuel molecule (such as an alcohol) or by recreating the major constituents of gasoline in a renewable manner. Gasoline and jet fuel are complex, variable mixtures of hydrocarbons and other additives that contain large quantities of short and long-chain n-alkanes [135-140]. To this end, biosynthetic hydrocarbon production in engineered cellular hosts has received great attention recently. In *Escherichia coli* and *Saccharomyces cerevisiae*, heterologous pathway expression has enabled the production of isoprenoids [141-144], but their abundance of carbon-carbon double bonds results in only the alkene fraction of fuels. Biosynthetic production of long chain n-alkanes has been achieved through the reduction and decarbonylation of fatty acids to their aliphatic backbones [77,80,145,146]. In particular, this pathway enabled synthesis of tridecane, pentadecene, pentadecane, and heptadecane in *E. coli* [77]. Despite these advances, current biofuel technology is limited by a lack of pathways to produce short-chain *n*-alkanes. Here, we present a proof-of-concept pathway demonstrating that it is possible to produce such short alkanes via a lipoxygenase-mediated reaction in *Yarrowia lipolytica*. Specifically, we demonstrate the first microbial production of pentane by importing a lipoxygenase-based pathway found in soybeans.

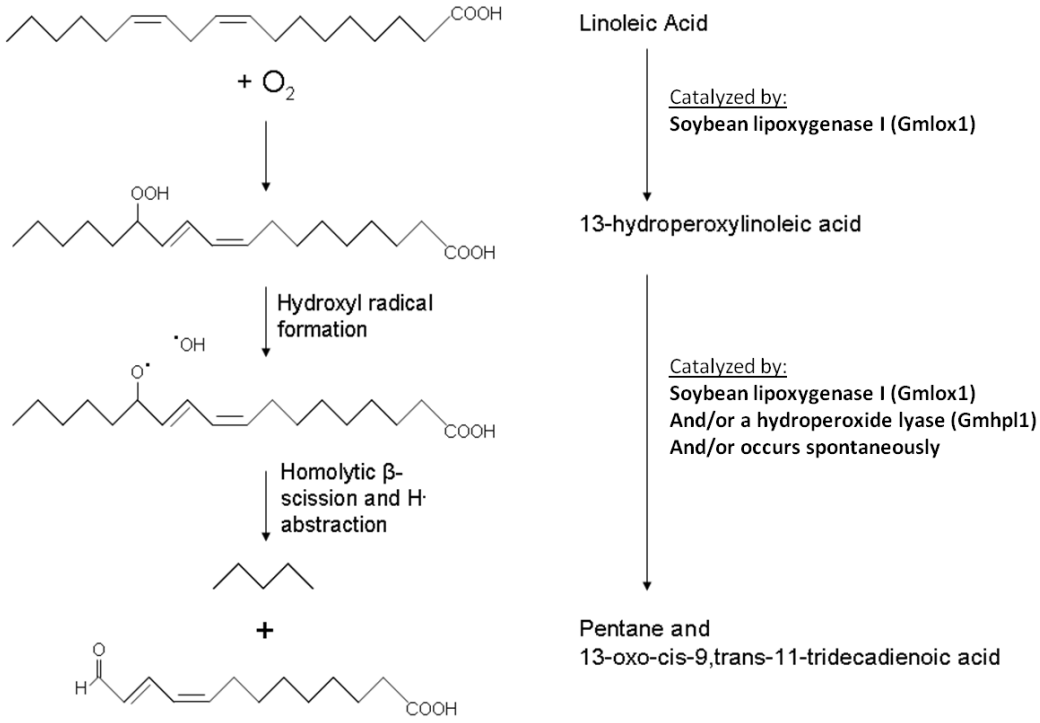
Lipoxygenase enzymes convert polyunsaturated fatty acids into an unsaturated fatty acid hydroperoxide by adding a molecular oxygen to a (Z,Z)-1,4-pentadiene structural unit located within the fatty acid [147,148]. Pentane production by means of this pathway has been previously demonstrated in both soybeans and peanuts [79,82]. In

particular, soybean (*Glycine max*) lipoxygenases I and II and peanut (*Arachis hypogaea*) lipoxygenase convert linoleic acid (C18:2) into a 13-hydroperoxy linoleic acid intermediate (13-HPOD). Further catalysis via a homolytic-β-scission reaction and hydride abstraction converts 13-HPOD into pentane and 13-oxo-cis-9-trans-11-tridecadienoic acid, or into *n*-hexanal and 12-oxo-cis-9-dodecenoic acid [82,149-151]. Since no lipoxygenase pathway for short-chain alkane synthesis has been synthetically imported into a microbial system, there is discrepancy as to the enzymes required for this pathway. In particular, while it is clear that lipoxygenase enzymes mediate at least the first step to pentane (**Figure 5.1**), hydroperoxide lyase enzymes have been implicated in n-hexanal formation from 13-HPOD degradation [149,152-154]. Thus, it is unclear if the conversion of 13-HPOD to pentane product occurs spontaneously, is mediated by the lipoxygenase, or requires a hydroperoxide lyase enzyme. *In vitro* analysis has demonstrated that soybean lipoxygenase I [155], henceforth referred to as Gmlox1, is highly active under a broad range of conditions, can produce pentane from linoleic acid substrate, and does not produce n-hexanal - making it promising for short-chain alkane production [150,156].

Biosynthetic pentane production requires a host system capable of producing linoleic acid substrate. Thus, the oleaginous yeast *Yarrowia lipolytica* was chosen for this study based on reports that linoleic acid can account for up to 47% of total lipid extract [63]. Gene deletion analyses of β-oxidation enzymes and other key fatty acid enzymes in *Y. lipolytica* have established mechanisms to further increase lipid accumulation [63,88,157-159]. *Y. lipolytica* possesses a developed genetic toolbox [8,69,72,75], and

our work generating hybrid promoters has significantly increased *Y. lipolytica's* transcriptional capacity. These promoters enable high-level genetic overexpressions necessary for heterologous enzymes such as the lipoxygenase utilized here [1,2,160].

Figure 5.1: Lipoxygenase-mediated conversion of linoleic acid to pentane



Soybean lipoxygenase I, Gmlox1, catalyzes the addition of molecular oxygen to the *cis* double bond within linoleic acid to form 13-hydroperoxylinoleic acid. Scission of this intermediate to pentane and a byproduct can proceed either spontaneously or through further Gmlox1 catalysis, and evidence suggests that a hydroperoxide lyase enzyme may assist this reaction.

5.3 Results and Discussion

5.3.1 Enabling the novel heterologous production of a short-chain alkane

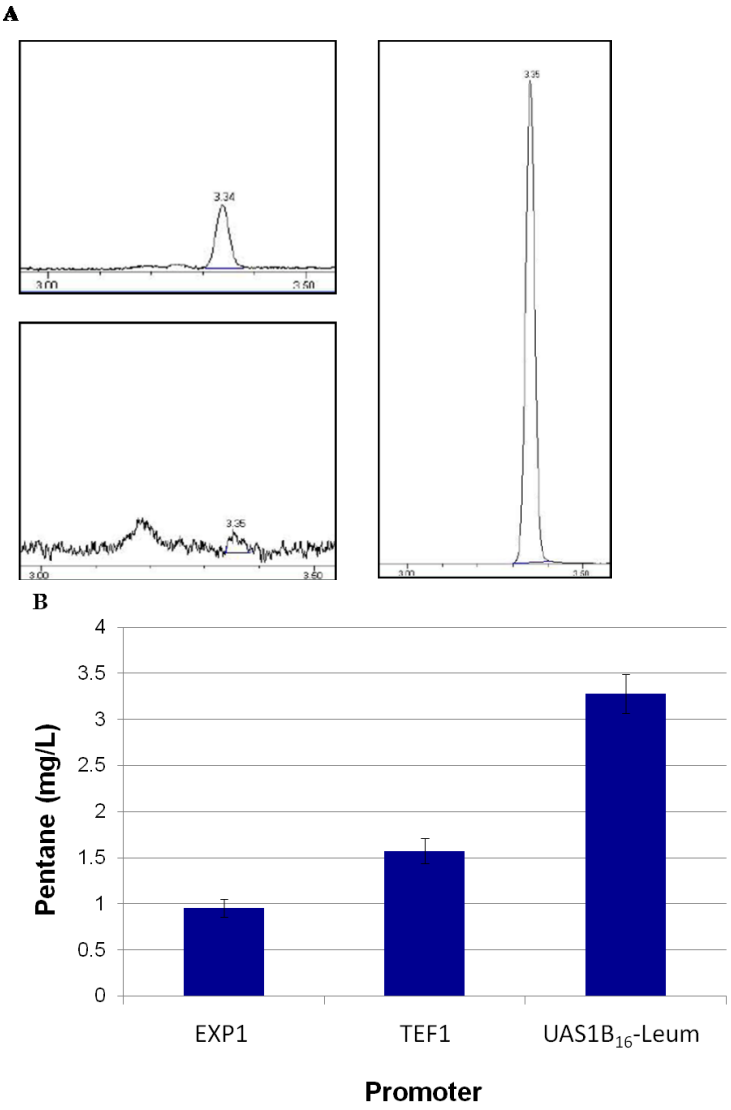
To enable synthesis of the short-chain alkane pentane in *Y. lipolytica*, *Glycine max* lipoxygenase isozyme 1, Gmlox1, was codon optimized and inserted into plasmid-

borne expression cassettes driven by the well-characterized TEF, EXP, and UAS1B$_{16}$-Leum promoters [2]. We had previously found that the strength of these promoters varied such that TEF < EXP << UAS1B$_{16}$-Leum. *Y. lipolytica* strains harboring episomal Gmlox1 expression cassettes were cultured in standard minimal media for eight days and assayed for pentane production by sampling the headspace and injecting into GC. In each of these cases, heterologous overexpression of the Gmlox1 enzyme enabled the novel microbial production of a short-chain alkane, pentane, as determined by GC (**Figure 5.2a**). The highest level expression of Gmlox1, by the UAS1B$_{16}$-Leum promoter, resulted in 3.28mg/L pentane. Yet expression of Gmlox1 by the far weaker TEF promoter (by nearly seven fold) yielded a value of nearly half, 1.56mg/L (**Figure 5.2b**). Based on this experiment, it was evident that transcription of Gmlox1 was not the only rate limiting step for pentane production in an unoptimized, episomal expression system. Therefore, we determined that the UAS1B$_{16}$-Leum promoter sufficiently overexpressed the heterologous Gmlox1 enzyme, translational issues being outside the scope of this study. More importantly, these results demonstrated that the lipoxygenase enzyme alone was necessary and sufficient to produce pentane.

Our initial experimentation utilized an eight-day culturing period to permit *Y. lipolytica* ample time to accumulate lipid content and linoleic acid substrate for the lipoxygenase enzyme. To confirm this eight-day culturing period maximized pentane output, we performed a time course of pentane production with a *Y. lipolytica* strain harboring the UAS1B$_{16}$-Leum-Gmlox1 expression cassette. Pentane production was analyzed after two, four, six, eight, and ten days of growth, with a maximum yield of

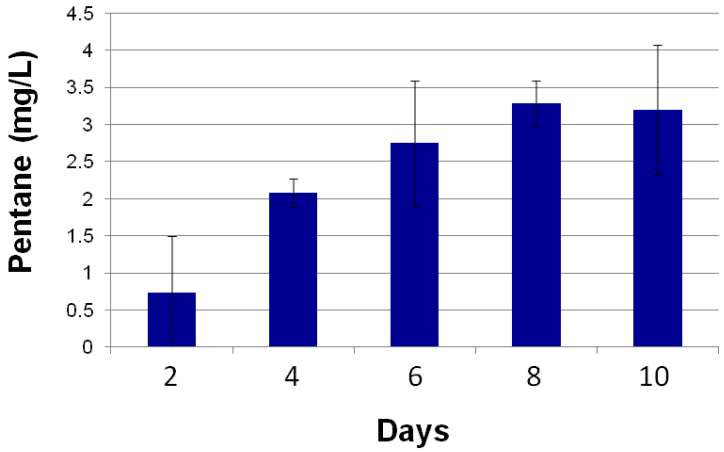
3.28mg/L after eight days, confirming that the eight-day growth period maximizes pentane production (**Figure 5.3**).

Figure 5.2: Enabling short chain *n*-alkane production in *Y. lipolytica*



(**A**) Top left - A chromatogram illustrating the peak at 3.34 minutes representing pentane. Bottom left – Negative Control - A chromatogram illustrating the lack of the peak at 3.34 minutes in a strain not expressing the Gmlox1 gene. Right – Positive Control - A chromatogram illustrating the peak at 3.34 minutes in a culture supplemented with 100 mg/L pentane. (**b**) A codon optimized soybean lipoxygenase gene (Gmlox1) was expressed in *Y. lipolytica* by three different promoters, the endogenous TEF and EXP promoters, and the high strength UAS1B$_{16}$-Leum synthetic hybrid promoter. This Gmlox1 expression enabled the first reported microbial production of a short chain *n*-alkane. Expression of Gmlox1 by the UAS1B$_{16}$-Leum promoter produced 3.28mg/L pentane. Error bars represent standard deviation from biological triplicates.

Figure 5.3: Time-course of pentane production



To determine the cultivation time required to maximize pentane yield, pentane production was analyzed from strains harboring an episomal UAS1B$_{16}$-Leum-Gmlox1 expression cassette after two, four, six, eight, and ten days of growth. An eight day growth period yielded the highest titer of 3.28mg/L pentane, with an insignificant reduction after ten days. Error bars represent standard deviation from biological triplicates.

5.3.2 Increasing linoleic acid substrate availability by supporting lipid accumulation

To increase linoleic acid content in our host cells, we utilized two distinct attempts to increase overall fatty acid and lipid accumulation. First, we altered the carbon and nitrogen supply in our media to shunt carbon flux directly into lipid and fatty acid synthesis. Secondly, we reduced intracellular fatty acid and lipid catabolism with two genetic knockouts.

5.3.2.1 Optimizing media formulation

The impetus for lipid accumulation in *Y. lipolytica* is nitrogen starvation that triggers the shuttling of carbon directly into fatty acid synthesis [161,162]. Depending on the ratio of carbon flux versus nitrogen flux into the cell, *Y. lipolytica* can alter its metabolic state to degrade or accumulate intracellular lipid content [62,89]. Increasing carbon to

nitrogen ratio in the growth media shifts carbon metabolism towards lipid accumulation until a threshold is reached, after which citric acid is accumulated [9,59,163]. Optimization of media for fatty acid accumulation is dependent on culturing condition (batch, fed-batch, continuous) and strain genotype. Here we chose to select two conditions for comparison. First, we used the standard, minimal media formulation containing 20 g glucose and ~5 g ammonium sulfate per liter. Second, we selected a 160g glucose and 0.2g ammonium sulfate per liter formulation, dubbed High C:N media, to induce lipid accumulation. Cultivation in High C:N media increased lipid % dry cell weight (DCW) by 67% for *Y. lipolytica* PO1f, from 9.5% lipid DCW to 15.9% lipid DCW (**Figure 5.4a**). This increase in lipid accumulation was not accompanied by an increase in organic production, as the concentration of citric acid in the supernatant decreased from 109.1 mg/L in minimal media to 82.3 mg/L in High C:N media. Media formulation altered the overall fatty acid profile for PO1f slightly, and relative linoleic acid substrate accumulation decreased from 1.39% to 1.12% (**Table 5.1**). However, total linoleic acid content per cell increased when utilizing High C:N media as a result of higher overall lipid weight per cell.

Table 5.1: Fatty acid profile analysis

	Fatty acid accumulation (%)				
Strain (Media)	**C16:0**	**C16:1**	**C18:0**	**C18:1**	**C18:2**
PO1f (minimal)	18.54%	7.66%	18.90%	35.00%	1.39%
PO1f (High C:N)	18.14%	5.07%	24.14%	41.51%	1.12%
PO1f-Δmfe1 (High C:N)	11.48%	4.02%	21.31%	45.87%	2.11%
PO1f-Δpex10 (High C:N)	15.14%	3.94%	21.06%	39.47%	0.59%

The *Y. lipolytica* strains harboring the three episomal Gmlox1 expression cassettes, driven by the EXP1, TEF1 and UAS1B$_{16}$-Leum promoters, were cultured in High C:N media for eight days and analyzed for pentane production (**Figure 5.4b**). A large increase in pentane production was observed for each construct by switching from standard media to High C:N media, culminating in a 4.19mg/L pentane yield by the highest strength UAS1B$_{16}$-Leum-Gmlox1 cassette (**Figure 5.4b**). Therefore, increasing lipid levels (and thus concomitant increases in fatty acid substrate availability) by altering growth media formulation was an effective measure to increase alkane product formulation. The increase in *n*-pentane yield attributable to the 67% increase in overall lipid content more than compensated for the slight reduction in linoleic acid accumulation. Thus, we concluded that free linoleic acid substrate availability was a major factor limiting alkane production in our system, and could be effectively overcome by encouraging general fatty acid and lipid accumulation.

Figure 5.4: Increasing pentane production by increasing lipid levels

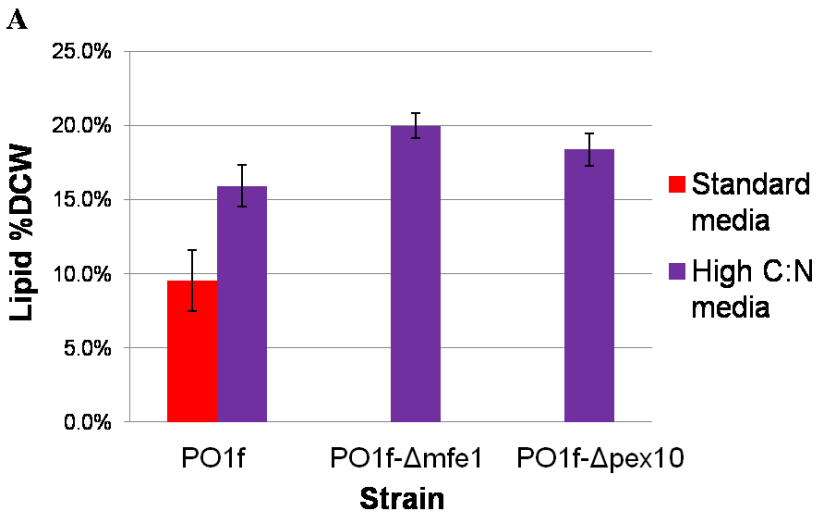
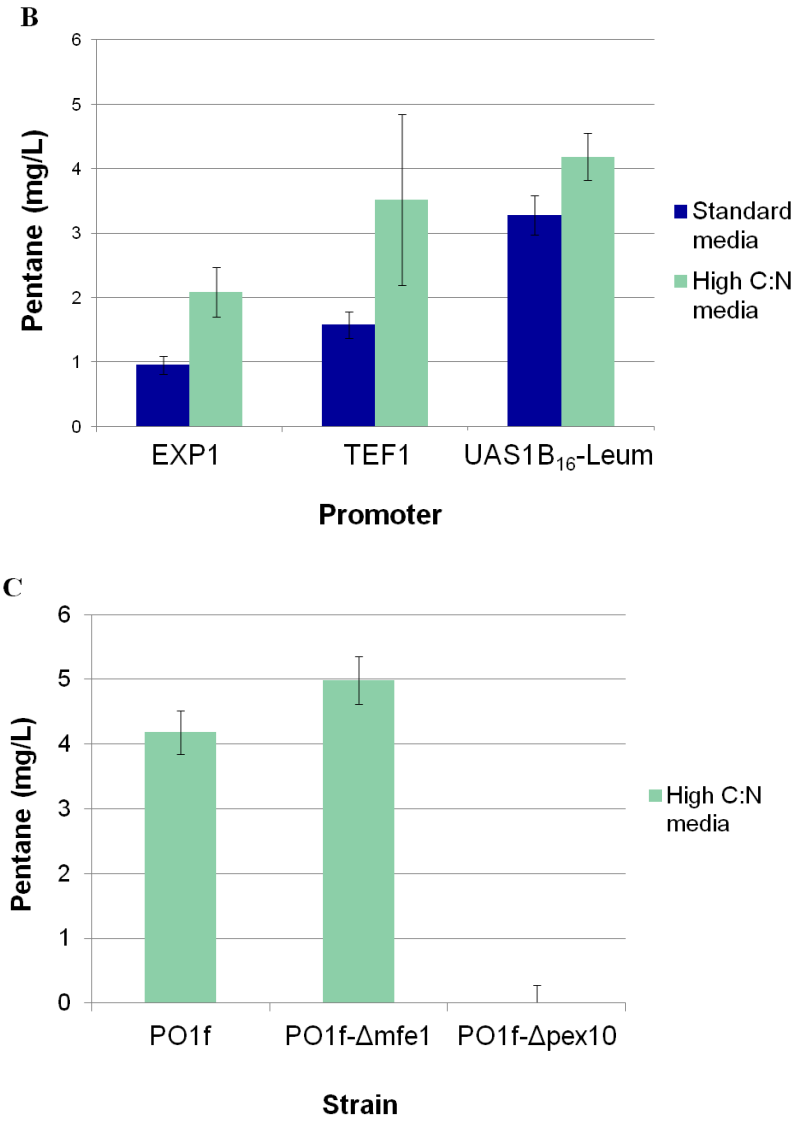


Figure 5.4 continued



(**A**) The percentage of lipid and fatty acid dry cell weight (Lipid %DCW) is shown for *Y. lipolytica* strain PO1f cultivated in standard synthetic minimal media or in High C:N media and for strains PO1f-Δmfe1 and PO1f-Δpex10 cultivated in High C:N media. Cultivation conditions were identical to those employed for pentane production, and all strains harbored plasmid pMCSCen1. Growth on High C:N media boosts lipid %DCW by 67% in strain PO1f, from 9.5% to 15.9%. Deletion of key fatty acid degradation enzymes further increases Lipid %DCW to 20.0% and 18.4% for mfe1 and pex10 gene deletions, respectively. (**B**) *Y. lipolytica* PO1f strains harboring TEF, EXP, and UAS1B_{16}-Leum driven Gmlox1 expression cassettes were cultivated for eight days in standard minimal media or in High C:N media. Cultivation in High C:N media resulted in increased pentane production for all strains, due to a concurrent increase in linoleic acid substrate accumulation. Cultivation of a *Y. lipolytica* strain harboring the UAS1B_{16}-Leum-Gmlox1 expression cassette in High C:N media produced 4.19 mg/L pentane. (**C**) *Y. lipolytica* PO1f Δmfe1 and Δpex10 strains harboring UAS1B_{16}-Leum driven Gmlox1 expression cassettes were cultivated for eight days in High C:N media. The pex10 gene deletion abolished pentane production. Deletion of the mfe1 gene increased yields by 19% to 4.98mg/L pentane, due to a decrease in linoleic acid substrate degradation. Error bars represent standard deviation from biological triplicates.

5.3.2.2 Removing β-oxidation

By altering media composition to shift *Y. lipolytica*'s metabolism towards fatty acid and lipid accumulation, we demonstrated a positive correlation between substrate availability and short-chain *n*-alkane production. Removal of fatty acid β-oxidation capacity in *Y. lipolytica* represents a well-established alternative method to increase fatty acid (including linoleic acid) and lipid accumulation. The bulk of fatty acid degradation occurs inside peroxisomes, specialized organelles prevalent in *Y. lipolytica* [164]. Deletion of the six acyl-CoA oxidase enzymes (POX1-6) that catalyze the first step of fatty acid β-oxidation or of the multifunctional enzyme (MFE1) that catalyzes both the second and third step of fatty acid degradation have been shown to increase lipid accumulation in *Y. lipolytica* [63,158]. Therefore, we sought to evaluate the impact of two gene deletions on pentane production in *Y. lipolytica*. We created single knockouts strains of the MFE1 β-oxidation gene and of the PEX10 gene, a transcription factor necessary for correct peroxisomal biogenesis and morphology [90,91]. We assayed *mfe1* and *pex10* gene deletion strains for lipid accumulation (**Figure 5.4a**) and fatty acid content (**Table 5.1**) when cultivated in High C:N media. The *mfe1* gene deletion increased fatty acid and lipid accumulation by 25% from 15.9% to 20.0% lipid DCW (**Figure 5.4a**), while simultaneously nearly doubling linoleic acid accumulation from 1.12% to 2.11 % (**Table 5.1**). The *pex10* gene deletion also increased lipid accumulation to 18.4% lipid DCW (**Figure 5.4a**), but significantly reduce production of linoleic acid to only 0.59% of total fatty acid content (**Table 5.1**). Once again, organic acid production remained fairly

constant, producing 123.4 and 107.6 mg/L citric acid in the *mfe1* and *pex10* deletion strains, respectively.

We transformed the high strength UAS1B16-Leum-Gmlox1 episomal expression cassette into each knockout mutant, and then cultivated these strains in High C:N media. The *mfe1* gene deletion increased pentane production by 19% to 4.98mg/L, while the *pex10* gene deletion surprisingly completely abolished pentane production (**Figure 5.4c**). It has been demonstrated that Pex10p controls the composition of intracellular fatty acid content, and the *pex10* gene deletion has been associated with decreased linoleic acid content [90]. In this regard, we show that the *pex10* deletion specifically reduced linoleic acid concentration, despite an overall lipid content improvement. This once again points toward linoleic acid availability as the major rate limiting step in this process.

Deletion of MFE1 improved pentane production to our highest yield of 4.98mg/L, and represented the cumulative effect of three combinatorial strategies to increase alkane yield, including optimizing media formulation and removing fatty acid degradation to increase substrate availability, and strongly overexpressing the heterologous Gmlox1 enzyme.

5.3.3 Investigating the necessity of a hydroperoxide lyase enzyme

Finally, we sought to evaluate the necessity of a hydroperoxide lyase enzyme in the synthetic pentane production pathway. The lipoxygenase-mediated cleavage of linoleic acid to pentane and the 13-oxo-cis-9-trans-11-tridecadienoic acid byproduct follows a two-step catalytic pathway (**Figure 5.1)**. Numerous hydroperoxide enzymes have been characterized to cleave long-chain fatty acid hydroperoxides into shorter-chain

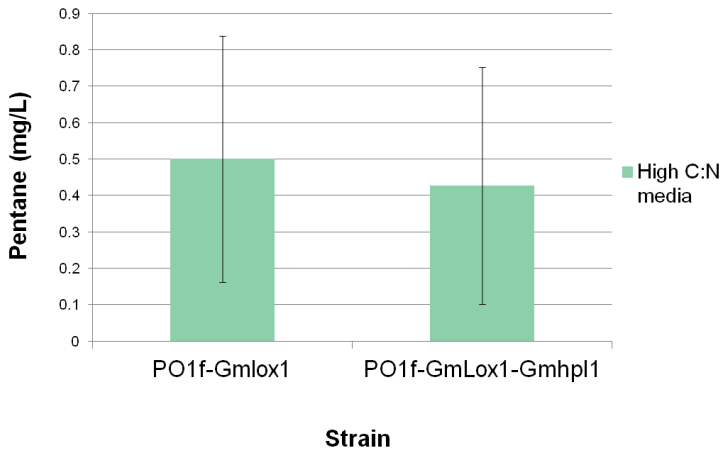
products (predominantly *n*-aldehydes) [152-154]. Hydroperoxide lyase activity has been implicated for n-hexanal production in soybeans, the source of the Gmlox1 lipoxygenase enzyme [165-167], yet the responsible gene for this is unknown. Through bioinformatics analysis, we found that soybeans contain a protein homolog known as CYP74B15 with 63% identity to an *n*-hexanal forming hydroperoxide lyase from green bell pepper [154,168]. We hypothesized that this hydroperoxide lyase homolog could improve pentane production when expressed in tandem with the lipoxygenase enzyme if this second step of the mechanism was indeed catalyzed by a separate enzyme.

The soybean hydroperoxide lyase homolog, dubbed Gmhpl1, was codon optimized and inserted into a high-strength integrative expression cassette. As described above, consecutive integrations of this cassette and a high strength Gmlox1 expression cassette enabled the construction of two stable, markerless strains – for lone Gmlox1 expression or for simultaneous Gmlox1/Gmhpl1 expression. We tested each strain for pentane production after cultivation on High C:N media, but saw no increase in alkane production from Gmhpl1 co-overexpression **(Figure 5.5**). Thus, the secondary hydroperoxide lyase enzyme had little to no impact towards short-chain *n*-alkane production via lipoxygenase catalyzed cleavage of linoleic acid and was inessential for the biochemical reaction mechanism. In the first enzymatic step (**Figure 5.1**), the lipoxygenase enzyme (Gmlox1) catalyzes the addition of molecular oxygen to linoleic acid to form the 13-HPOD intermediate. The second step of the reaction is either spontaneous or catalyzed by the lipoxygenase. This one-enzyme heterologous pathway is

comparably simpler than pathways for higher-chain alkane production previously described [77,169].

It is interesting to note that we observed a reduction in pentane production when integrating the Gmlox1 expression cassette (comparing **Figures 5.4c and 5.5**), despite that integration should result in a rough doubling of protein expression due to a documented "half on/half off" tendency of *Y. lipolytica* episomal expression cassettes [2]. We hypothesize that the strongly expressed leucine marker present on our base plasmid was influencing and increasing lipid accumulation in these plasmid-borne strains [170]. As we demonstrated with media modification and gene deletions, an increase in lipid accumulation could result in a concurrent increase in linoleic fatty acid substrate concentration and downstream pentane production.

Figure 5.5: Concurrent hydroperoxide lyase expression does not improve pentane production.



Y. lipolytica PO1f Gmlox1 and Gmlox1-Gmhpl1 strains harboring only the Gmlox1 ($UAS1B_{16}$-Leum driven) or both the Gmlox1 and the Gmhpl1 integrative expression cassettes, respectively, were cultivated for eight days in High C:N media. Co-expression of the Gmhpl1 hydroperoxide lyase enzyme did not increase pentane yield. Error bars represent standard deviation from biological triplicates.

5.4 SUMMARY AND CONCLUSIONS

The metabolic engineering of *Y. lipolytica* for short-chain *n*-alkane production is an important step towards enabling microbial replacement of petroleum products. We successfully diverted endogenous fatty acids towards the production of a bio-value added product by expressing the soybean lipoxygenase I enzyme. In this regard, we have expanded the chemical palette of cells through the incorporation of a novel heterologous metabolic pathway [11,171]. Our highest pentane titer must be improved, likely by stimulating fatty acid accumulation and lipogenesis to encourage alkane yields. Strain engineering and media optimization increased cellular lipid content more than twofold, stimulating a concurrent increase in pentane production. Thus, we demonstrated that *Y. lipolytica* metabolism can be manipulated to stimulate lipogenesis and that fatty acid pools can be diverted to more valuable metabolites such as alkane biofuels. This example serves to demonstrate that short-chain alkane synthesis is conceivable in cellular systems and can be improved through further metabolic engineering and protein engineering efforts. In the next chapter, we focus solely on stimulating lipogenesis to enable high fatty acid and lipid titers.

Chapter 6: Harnessing Lipogenesis in *Yarrowia lipolytica*

6.1 Chapter Summary

Microbial biosynthesis of fuels (such as ethanol and biodiesel) and industrial chemical precursors provides a renewable means to reduce dependence on petroleum feedstock [172-176]. Economic feasibility of these processes hinges upon harnessing either native or imported biosynthetic metabolism to achieve high titers and yields in a feedstock-independent manner [10]. To this end, *Yarrowia lipolytica* provides an ideal platform for oleochemical synthesis due to genetic tractability, broad substrate specificity, and native capacity for lipogenesis [8,59]. This chapter describes the largest engineering effort ever attempted to rewire an oleaginous organism, in which we couple combinatorial multiplexing of lipogenesis targets with phenotypic induction, creating a strain with unsurpassed lipogenesis capability. Specifically, tri-level control of *Y. lipolytica* metabolism resulted in saturated cells containing upwards of 90% lipid content and titers as high as 25.3 g/L lipids, representing a more than 60-fold improvement over parental strain and conditions. We further demonstrate that these lipids can be easily converted into FAMEs suitable for biodiesel. Through this rewiring effort, we uncover several unique, previously unreported facets of lipogenesis, including: (1) that lipogenesis is dependent on absolute environmental carbon content, (2) that lipid accumulation phenotypes are dependent on leucine biosynthetic capacity, and (3) that rare odd-chain fatty acids pathways are naturally activated by high lipogenesis. Finally, the high titers

and carbon-source independent nature of this lipogenesis in *Y. lipolytica* highlights the potential of this organism as a platform for efficient oleochemical and fuels production.

6.2 INTRODUCTION

Bio-based production of oils and lipids provides a unique platform for the sustainable production of biodiesel and other important oleochemicals [77,176]. Most efforts for developing such a platform involve either rewiring *E. coli* or cultivating cyanobacteria. These attempts suffer from low titers (less than 3g/L) and variable lipid content (ranging between 10 and 87%), with the highest levels typically occurring in non-tractable, slow growing hosts cultivated in oil containing media (*ex novo* lipid incorporation instead of *de novo* lipogenesis) [174,176-180]. As an alternative, several groups have explored oleaginous organisms such as the fungus *Y. lipolytica*, but total oil content and titers are still limited [61,63,92]. Yet the genetic tractability of this organism [2,54,69,72,74,181] coupled with its modest, innate *de novo* lipogenesis (~10-15% lipid content in wildtype [4,63,92]) make *Y. lipolytica* a potential candidate as a platform organism for superior lipid production. Lipogenesis is generally accepted to be induced by nitrogen starvation conditions, in which relative nitrogen content is far less than relative carbon content. It is also assumed that it is the ratio of carbon to nitrogen that is important towards lipid accumulation and that adjusting absolute nutrient content has little effect [59,62,89]. Lipid biosynthesis is primarily initiated by the activity of four enzymes - AMP Deaminase (AMPD), ATP-Citrate Lyase (ACL), Malic Enzyme (MAE) and Acetyl-CoA Carboxylase (ACC) – that cooperatively divert carbon flux from central carbon

metabolism towards fatty acid precursors [59,61]. Previous efforts to increase lipid accumulation have shown promise but have been limited by their breadth of metabolic control and their comprehensiveness of genotypic and phenotypic sampling towards complete redirection of metabolic flux towards lipid accumulation [4,61,63,92].

6.3 RESULTS AND DISCUSSION

6.3.1 Genomic engineering to increase lipogenesis

We performed a combinatorial, multiplexing of targets spanning fatty acid, lipid, and central metabolism through the overexpression of five lipogenesis enzymes in four genomic backgrounds marked by differential fatty acid catabolic capacity (**Table 6.1, Figure 6.1a**). Specifically, AMPDp, ACLp, and MAEp overexpressions were investigated for their potential to increase acetyl-CoA precursors (ACCp overexpression has not been reported to improve lipogenesis and was excluded [92]) and DGA1p and DGA2p (acyl-CoA:diacylglycerol acyltransferases isozymes I and II) were included for their potential in catalyzing the ultimate step in triglyceride synthesis [88]. These overexpression targets were multiplexed with several deletions that served to reduce fatty acid catabolism by reducing one or both of β–oxidation (via *mfe1* deletion) [61,63] and peroxisome biogenesis (via *pex10* deletion) [4,90] (**Table 6.1, Figure 6.1a**). We demonstrated in Chapter 5 that restoration of a complete leucine biosynthetic pathway potentially increased lipid accumulation more than alleviation of uracil auxotrophy in a PO1f base strain [4]. Thus, we included the complementing of leucine and uracil

biosynthetic capacity both singly and in tandem as targets for this multiplexing. Integrated expression cassettes were driven by our high strength synthetic UAS1B$_{16}$-TEF constitutive hybrid promoter, constructed and characterized in Chapter 2 [2]. Collectively, the combinatorial multiplexing of enzyme overexpressions, fatty-acid inhibition knockouts, and auxotrophies resulted in 57 distinct genotypes that were analyzed for lipogenesis capacity compared to the wild-type strain (**Table 6.2**). Initially, a nile red based fluorescence assay coupled with flow cytometry [182] was used to efficiently determine relative lipid content and assess critical genotype synergies. Across the resulting lipogenesis metabolic landscape, we observed a significant range in lipid accumulation that spanned a 74-fold improvement in fluorescence over unmodified *Y. lipolytica* PO1f (**Figure 6.1b**). By using fluorescence microscopy or by physically lysing the cells, it is evident that cells become larger and more saturated with lipid content across the resulting lipogenic continuum in this landscape (**Figure 6.1c,d**).

Table 6.1: List of genomic parts used in Chapter 6

Genomic backgrounds	
Name	**Genotype and *(function of knockout)***
POlf	MatA, leucine$^-$, uracil$^-$, no extracellular proteases
pex10	POlf-Δpex10 *(prevents peroxisome biogenesis)*
mfe1	POlf-Δmfe1 *(prevents β-oxidation)*
pex10 mfe1	POlf-Δpex10 Δmfe1 *(prevents peroxisome biogenesis and β-oxidation)*
Enzymatic overexpressions	
Name	**Function**
AMPD	Inhibits TCA cycle, increasing citric acid level
$ACL_{subunit}1$	Cleaves citric acid to acetyl-CoA
$ACL_{subunit}2$	Cleaves citric acid to acetyl-CoA
MEA	Increases NADPH cofactor supply
$DGA_{isozyme}1$	Catalyzes lipid synthesis step
$DGA_{isozyme}2$	Catalyzes lipid synthesis step
Auxotrophic markers	
Name	**Utilized for expression**
Leucine$^{+/-}$	Episomally and chromosomally
Uracil$^{+/-}$	Chromosomally

AMPD = Adenosine monophosphate deaminase; ACL = ATP-Citrate Lyase; MAE = Malic Enzyme; DGA = acyl-CoA:diacylglycerol acyltransferases. ACL is a heterodimeric protein so only dual overexpressions of the $ACL_{subunit}1$ and $ACL_{subunit}2$ were constructed and tested.

Table 6.2: *Yarrowia lipolytica* strains used and constructed in Chapter 6

Host Strain Name	**Genotype**	**Reference or Source**
***Yarrowia lipolytica base* strains**		
POlf	MatA, leucine-, uracil-, xpr2-322, axp1-2	[53]
pex10	MatA, leucine-, uracil-, xpr2-322, axp1-2, Δmfe1	[4]
mfe1	MatA, leucine-, uracil-, xpr2-322, axp1-2, Δpex10	[4]
pex10 mfe1	MatA, leucine-, uracil-, xpr2-322, axp1-2, Δpex10, Δmfe1	This work
***Yarrowia lipolytica* overexpression strains**		
Polf background		
POlf uracil$^+$	MatA, leucine-, uracil+, xpr2-322, axp1-2	This work
Polf leucine$^+$	MatA, leucine+, uracil-, xpr2-322, axp1-2	This work
Polf leucine$^+$ uracil$^+$	MatA, leucine+, uracil+, xpr2-322, axp1-2	This work
Polf leucine$^+$ Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2	This work
Polf uracil$^+$ AMPD	MatA, leucine-, uracil+, xpr2-322, axp1-2, AMPD	This work

Table 6.2 continued

Polf leucine^{+} AMPD	MatA, leucine+, uracil-, xpr2-322, axp1-2, AMPD	This work
Polf uracil^{+} MEA	MatA, leucine-, uracil+, xpr2-322, axp1-2, MEA	This work
Polf leucine^{+} MEA	MatA, leucine+, uracil-, xpr2-322, axp1-2, MEA	This work
Polf leucine^{+} uracil^{+} AMPD MEA	MatA, leucine+, uracil+, xpr2-322, axp1-2, AMPD, MEA	This work
Polf leucine^{+} DGA1 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, DGA1	This work
Polf leucine^{+} DGA2 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, DGA2	This work
Polf leucine^{+} DGA1	MatA, leucine+, uracil-, xpr2-322, axp1-2, DGA1	This work
***mfe1* background**		
mfe1 uracil^{+}	MatA, leucine-, uracil+, xpr2-322, axp1-2, *mfe1*	This work
mfe1 leucine^{+}	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*	This work
mfe1 leucine^{+} uracil^{+}	MatA, leucine+, uracil+, xpr2-322, axp1-2, *mfe1*	This work
mfe1 uracil^{+} AMPD	MatA, leucine-, uracil+, xpr2-322, axp1-2, *mfe1*, AMPD	This work
mfe1 leucine^{+} AMPD	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*, AMPD	This work
mfe1 uracil^{+} MEA	MatA, leucine-, uracil+, xpr2-322, axp1-2, *mfe1*, MEA	This work
mfe1 leucine^{+} MEA	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*, MEA	This work
mfe1 leucine^{+} uracil^{+} AMPD MEA	MatA, leucine+, uracil+, xpr2-322, axp1-2, *mfe1*, AMPD, MEA	This work
mfe1 leucine^{+} uracil^{+} ACL1 ACL2	MatA, leucine+, uracil+, xpr2-322, axp1-2, *mfe1*, ACL1, ACL2	This work
mfe1 leucine^{+} DGA1 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*, DGA1	This work
mfe1 leucine^{+} DGA2 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*, DGA2	This work
mfe1 leucine^{+} DGA1	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*, DGA1	This work
mfe1 leucine^{+} DGA2	MatA, leucine+, uracil-, xpr2-322, axp1-2, *mfe1*, DGA2	This work

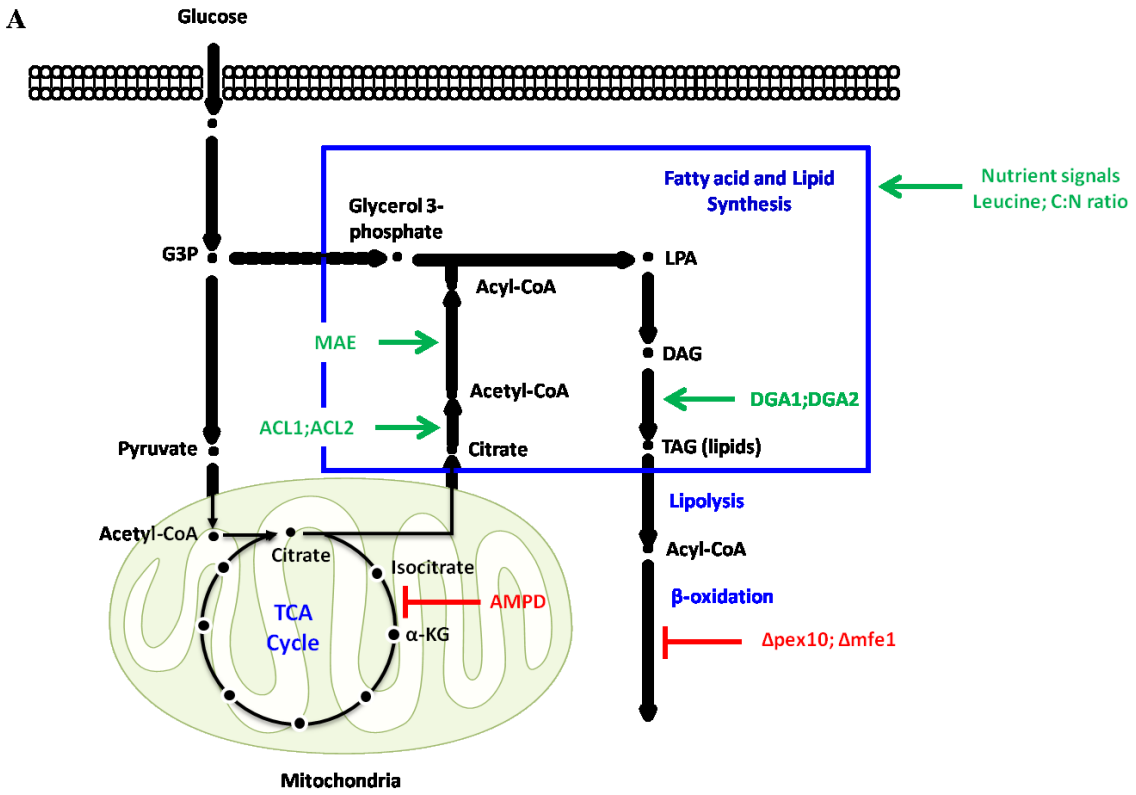
Table 6.2 continued

***pex10* background**		
pex10 uracil$^+$	MatA, leucine-, uracil+, xpr2-322, axp1-2, *pex10*	This work
pex10 leucine$^+$	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10*	This work
pex10 leucine$^+$ uracil$^+$	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10*	This work
pex10 uracil$^+$ AMPD	MatA, leucine-, uracil+, xpr2-322, axp1-2, *pex10*, AMPD	This work
pex10 leucine$^+$ AMPD	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10*, AMPD	This work
pex10 uracil$^+$ MEA	MatA, leucine-, uracil+, xpr2-322, axp1-2, *pex10*, MEA	This work
pex10 leucine$^+$ MEA	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10*, MEA	This work
pex10 leucine$^+$ uracil$^+$ AMPD MEA	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10*, AMPD, MEA	This work
pex10 leucine$^+$ DGA1 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10*, DGA1	This work
pex10 leucine$^+$ DGA2 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10*, DGA2	This work
pex10 leucine$^+$ DGA1	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10*, DGA1	This work
***pex10 mfe1* background**		
pex10 mfe1 uracil$^+$	MatA, leucine-, uracil+, xpr2-322, axp1-2, *pex10*, *mfe1*	This work
pex10 mfe1 leucine$^+$	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10, mfe1*	This work
pex10 mfe1 leucine$^+$ uracil$^+$	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*	This work
pex10 mfe1 uracil$^+$ AMPD	MatA, leucine-, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, AMPD	This work
pex10 mfe1 leucine$^+$ AMPD	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10, mfe1*, AMPD	This work
pex10 mfe1 uracil$^+$ MEA	MatA, leucine-, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, MEA	This work
pex10 mfe1 leucine$^+$ MEA	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10, mfe1*, MEA	This work
pex10 mfe1 leucine$^+$ uracil$^+$ AMPD MEA	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, AMPD, MEA	This work

Table 6.2 continued

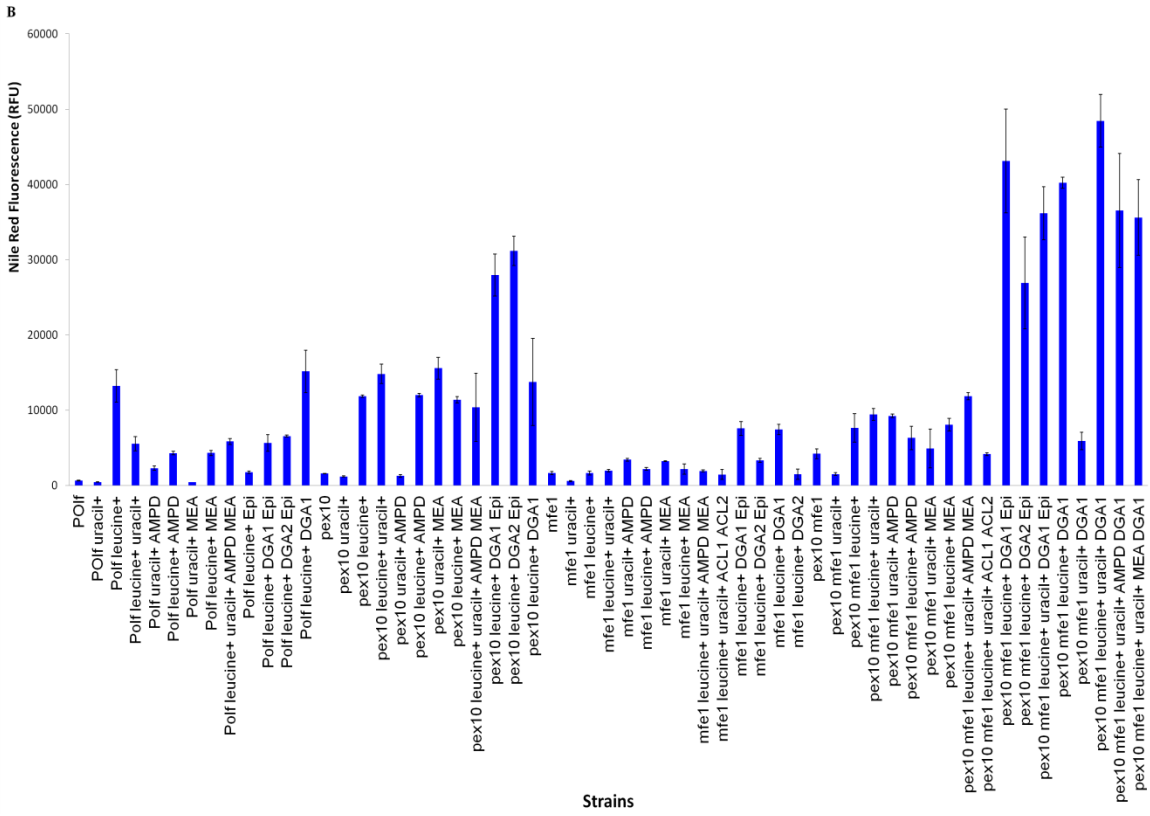
pex10 mfe1 leucine$^+$ uracil$^+$ ACL1 ACL2	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, ACL1, ACL2	This work
pex10 mfe1 leucine$^+$ DGA1 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10, mfe1*, DGA1	This work
pex10 mfe1 leucine$^+$ DGA2 Epi	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10, mfe1*, DGA2	This work
pex10 mfe1 leucine$^+$ uracil$^+$ DGA1 Epi	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, DGA1	This work
pex10 mfe1 leucine$^+$ DGA1	MatA, leucine+, uracil-, xpr2-322, axp1-2, *pex10, mfe1*, DGA1	This work
pex10 mfe1 uracil$^+$ DGA1	MatA, leucine+, uracil-, xpr2-322, axp1-2, pex10, mfe1, DGA1	This work
pex10 mfe1 leucine$^+$ uracil$^+$ DGA1	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, DGA1	This work
pex10 mfe1 leucine$^+$ uracil$^+$ AMPD DGA1	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, AMPD, DGA1	This work
pex10 mfe1 leucine$^+$ uracil$^+$ MEA DGA1	MatA, leucine+, uracil+, xpr2-322, axp1-2, *pex10, mfe1*, MEA, DGA1	This work

Figure 6.1: Combinatorial genomic engineering



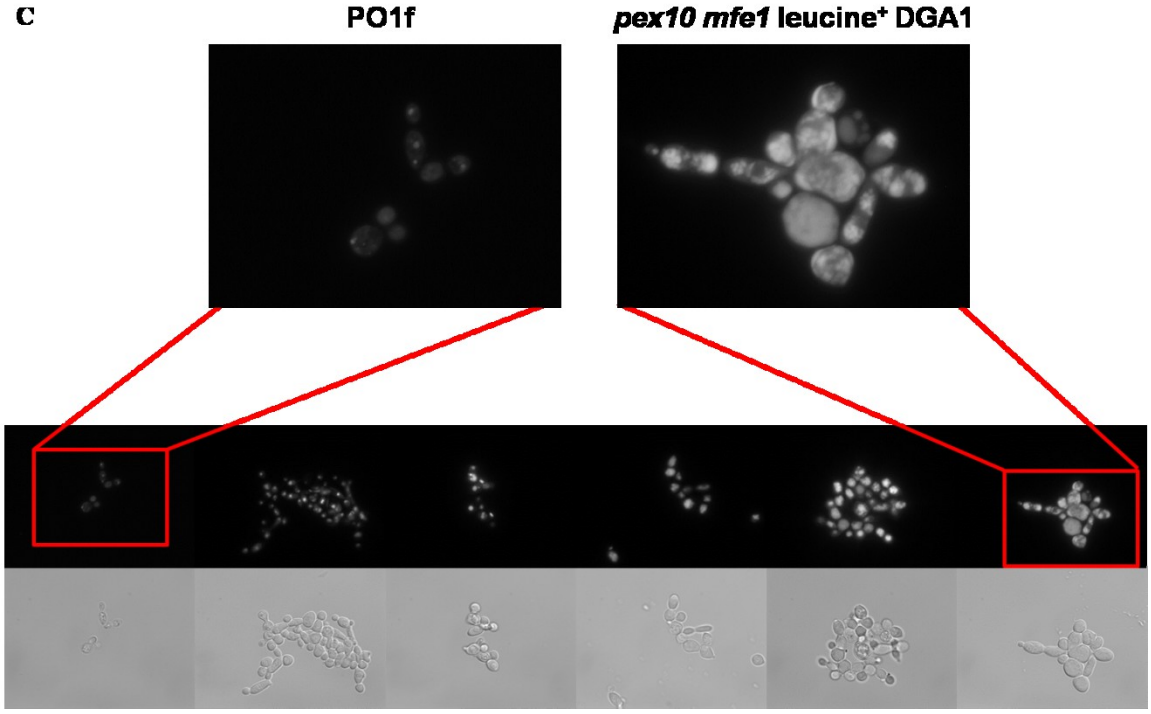
(**A**) A schematic illustrating how rewiring *Y. lipolytica's* metabolism can increase lipid content. Glucose is converted to acetyl-CoA fatty acid substrate through glycolysis and the TCA cycle. The TCA cycle can be inhibit by AMPD activity to increase isocitrate and citrate availability. Citrate is cleaved by ACL to form acetyl-CoA. Fatty acid synthesis produces acyl-CoA chains that join with a glycerol 3-phospate backbone during lipid synthesis. Three acyl-CoA molecules join to a single glycerol 3-phosphate molecule to complete lipid creation, and DGA1 and DGA2 catalyze the final step of lipogenesis. During lipolysis, lipases cleave lipids to reform free acyl-CoAs. Acyl-CoAs are degraded during fatty acid β-oxidation, which can be inhibited through pex10 or mfe1 deletion. Fatty acid and lipid synthesis can be manipulated with nutrient signals, including leucine and carbon availability. AMPD, ACL, MAE, DGA, pex10, and mfe1 are described in **Table 6.1**. G3P = Glyceraldehyde 3-Phosphate; LPA = lysophosphatidic acid; DAG = diacylglceride; TAG = Triacylglceride.

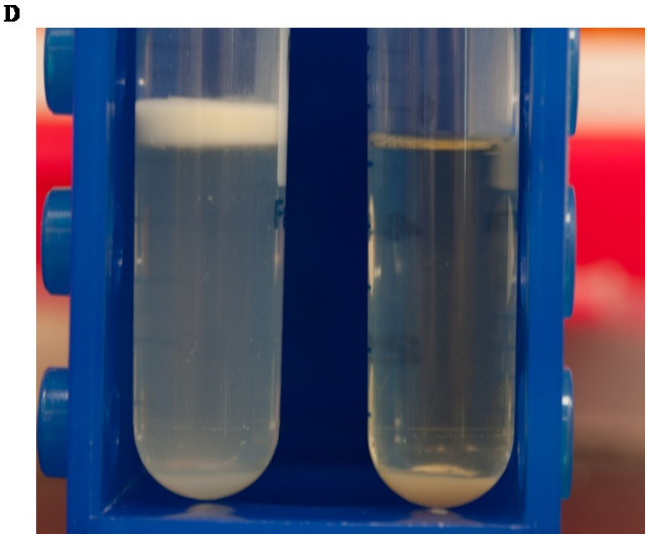
Figure 6.1 continued



(**B**) Nile red fluorescence analysis of lipid levels of the 57 strains constructed in this study. Strain names include strain background (POlf, *pex10*, *mfe1*, *pex10 mfe1*), auxotrophies relieved (leucine$^+$ or uracil$^+$), and enzymes overexpressed. "Epi" denotes episomal overexpressions. The *pex10* and *pex10 mfe1* backgrounds consistently display improved fluorescence levels compared to wildtype PO1f background. DGA1p overexpression and leucine biosynthetic capacity are key effectors of high lipogenesis. The highest levels of fluorescence were observed when staining the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain, the strain selected for bioreactor fermentation. AMPDp and MEAp overexpressions improve fluorescence signals in low lipid production strains, but do not enhance fluorescence in the highest production strain (far right, and second from right). Error bars represent standard deviations of biological triplicates.

Figure 6.1 continued

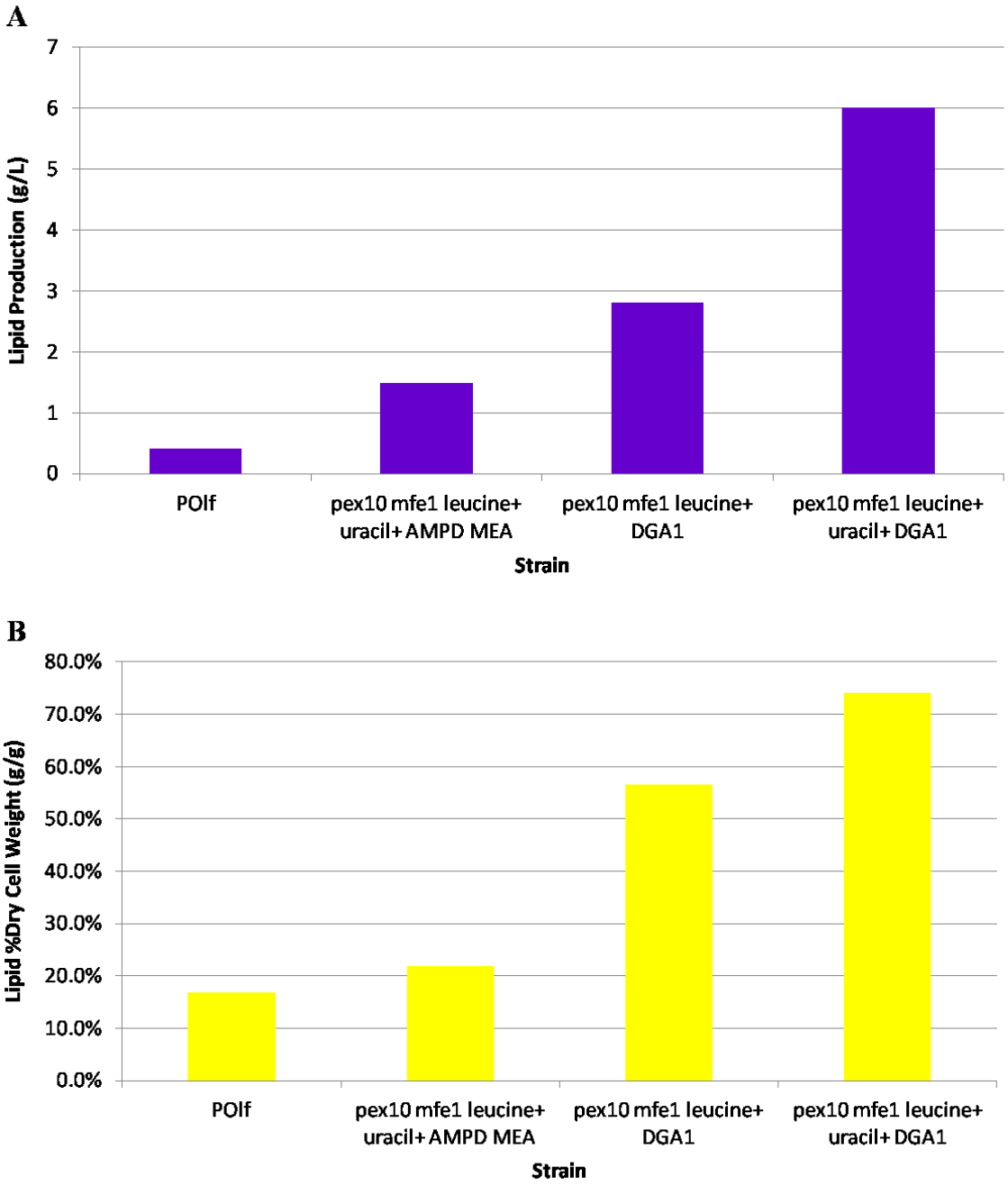




(**C**) Fluorescence light microscopy images of six strains increasing in lipid content (white) from left (unmodified POlf) to right (*pex10 mfe1* leucine$^+$ DGA1). (**D**) Image of lysed *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 (left) and POlf control (right) cells. Lipids remain floating after centrifugation at max speed.

Three dominant genetic targets exhibited cooperativity in the lipogenesis metabolic landscape—*pex10* deletion, DGA1p or DGA2p overexpression, and restoration of a complete leucine biosynthetic pathway (leucine$^+$ genotype) (**Figure 6.1b**). Each of these targets improved nile red based fluorescence by more than threefold, and DGA1p overexpression outperformed that of DGA2p. Overexpression of the MEA and AMPD proteins were positive effectors of lipogenesis, but offered no cooperative advantage when combined the *pex10* leucine$^+$ DGA1 background (**Figure 6.1b**). Deletion of the *mfe1* gene was not seen to alter total nile-red based lipid fluorescence measurements, but its removal reduces fatty acid degradation in carbon starvation conditions [4,61]. Similarly, the uracil$^+$ genotype had minimal effect on fluorescence, but improved growth rate and permitted cultivation in a pure minimal medium composition. Thus, the *pex10, mfe1,* leucine$^+$, uracil$^+$, DGA1p genotype, the strain with the highest lipogenesis potential in terms of fluorescence, was selected as our most advantageous strain. We extracted and measured lipid content to confirm nile red-based flow cytometry assessment of lipid content (**Figures 6.2a,b**). In small scale, test-tube cultivations, the final engineered strain outperformed all others, yielding 6.00 g/L lipids with 74% lipid content, a nearly 15-fold improvement over control (0.41 g/L lipid and 16.8% lipid content).

Figure 6.2: Improvement of lipid production in small-scale cultivation



(A,B) Four strains were assayed for lipid production and for % lipid content (g lipids/g biomass) when cultivated in 80g/L glucose, 6.7g/L Yeast Nitrogen Base w/o amino acids, and 0.79g/L CSM supplement – POlf, *pex10 mfe1* leucine$^+$ uracil$^+$ AMPD MEA, *pex10 mfe1* leucine$^+$ DGA1, and *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1. The *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain yielded the highest lipid production of 6g/L and highest % lipid content (74%), nearly 15-fold and fourfold improvements, respectively, over POlf control (0.41g/L lipids and 16.8% lipid content). Without uracil biosynthetic capacity, the *pex10 mfe1* leucine$^+$ DGA1 strain generated less than 3g/L lipids.

6.3.2 Coupling genotypic engineering with phenotypic induction

We next sought to understand the complex relationship between *de novo* lipid accumulation and nutrient levels. It is generally accepted that lipogenesis is highly dependent on the ratio of available carbon and nitrogen (C:N ratio) and induction requires a nitrogen starvation mechanism [8,85]. However, no definitive, quantitative relationship between genotype and lipogenesis induction has been determined. Thus, we analyzed the effect of nitrogen starvation and carbon availability on lipogenesis for unmodified *Y. lipolytica* PO1f and eleven engineered strains spanning the lipogenesis landscape (**Table 6.3**). Cultivation of these twelve strains in media formulations containing between 10g/L and 160g/L glucose and 0.055g/L and 1.365g/L ammonium revealed that absolute glucose level, rather than generally accepted C:N ratio, is crucial towards inducing lipid synthesis, and high lipid producers realized optimal accumulation in higher glucose media (**Table 6.4, Figure 6.3a-l**). In particular, unmodified *Y. lipolytica* PO1f and other low performing strains were most strongly induced by a lower carbon level (20g/L glucose and 0.273g/L ammonium), but responded poorly at similar C:N ratios with higher glucose and ammonium concentrations (**Figure 6.3a-d**). Moderate lipid accumulators were highly induced at intermediate glucose levels but again responded poorly at similar C:N ratios (**Figure 6.3e-i**). The highest accumulators, including the *pex10 mfe1* leucine$^+$ DGA1p genotype, were induced most intensely by higher levels of carbon and nitrogen (80g/L glucose and 1.365g/L ammonium) (**Figure 6.3j-l**). Thus, the current paradigm asserting the necessity of nitrogen starvation and that similar C:N ratios beget similar induction irrespective of overall carbon and nitrogen levels is incorrect. Instead, a

defined amount of carbon content ultimately controls lipid synthesis, and this favorable carbon level increases in strains capable of superior lipogenesis.

Table 6.3: Twelve strains analyzed to determine dependency of lipid accumulation induction on genotype

PO1f
mfe1 uracil$^+$ leucine$^+$ ACL1 ACL2
pex10 uracil$^+$ AMPD
pex10 mfe1
mfe1 uracil$^+$ leucine$^+$ AMPD MEA
mfe1 uracil$^+$ MEA
pex10 mfe1 uracil$^+$ leucine$^+$ ACL1 ACL2
uracil$^+$ leucine$^+$
leucine$^+$ DGA1
pex10 leucine$^+$ AMPD
pex10 leucine$^+$ MEA
pex10 mfe1 leucine$^+$ DGA1

Table 6.4: Media formulations utilized in lipid accumulation induction-genotype dependency assay.

	Carbon Source	Nitrogen Source
	Glucose	Ammonium Sulfate (Ammonium)
Media Name	(g/L)	(g/L)
$C_{20}N_{0.2}$	20	0.2 (0.055)
$C_{20}N_{1}$	20	1.0 (0.273)
$C_{20}N_{5}$ (YSC)	20	5.0 (1.365)
$C_{40}N_{0.2}$	40	0.2 (0.055)
$C_{40}N_{1}$	40	1.0 (0.273)
$C_{40}N_{5}$	40	5.0 (1.365)
$C_{80}N_{0.2}$	80	0.2 (0.055)
$C_{80}N_{1}$	80	1.0 (0.273)
$C_{80}N_{5}$	80	5.0 (1.365)
$C_{160}N_{0.2}$	160	0.2 (0.055)
$C_{160}N_{1}$	160	1.0 (0.273)
$C_{160}N_{5}$	160	5.0 (1.365)

Figure 6.3: Dependence of lipid induction phenotype on genotype and environmental conditions

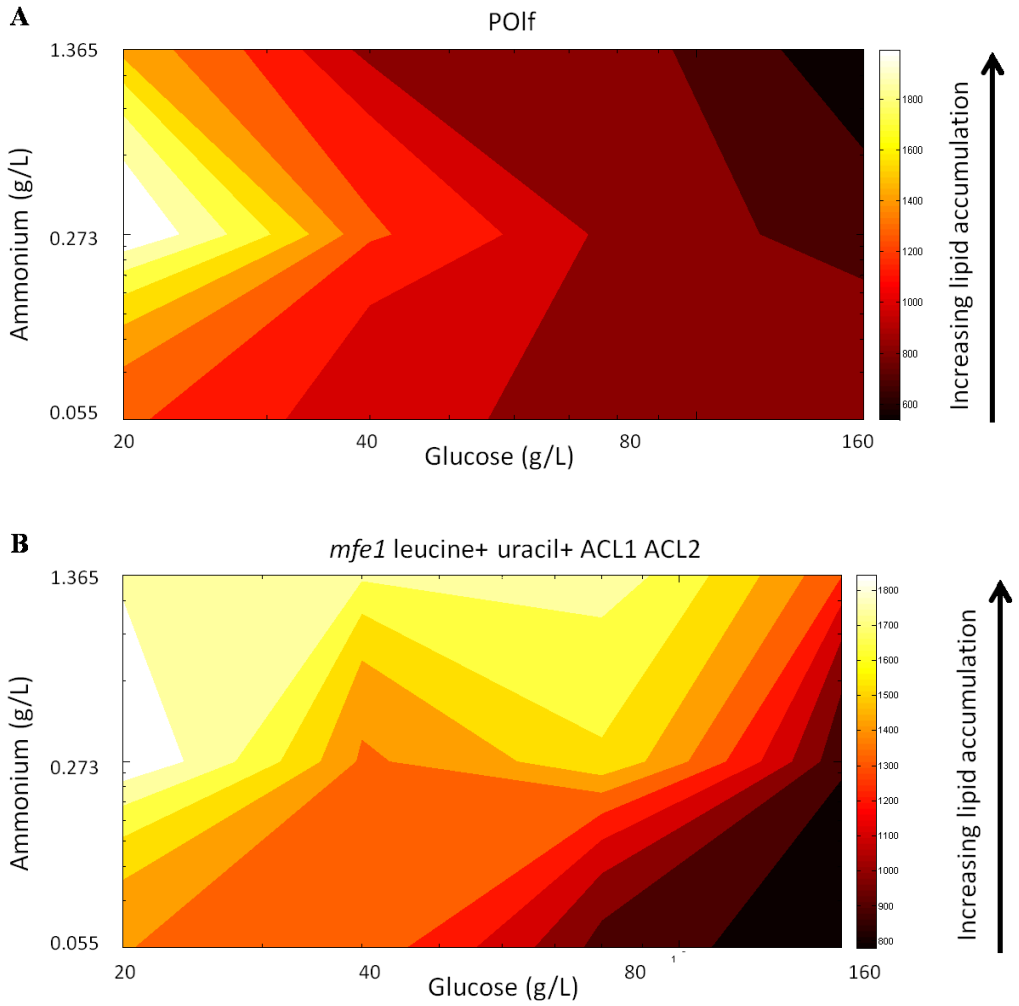


Figure 6.3 continued

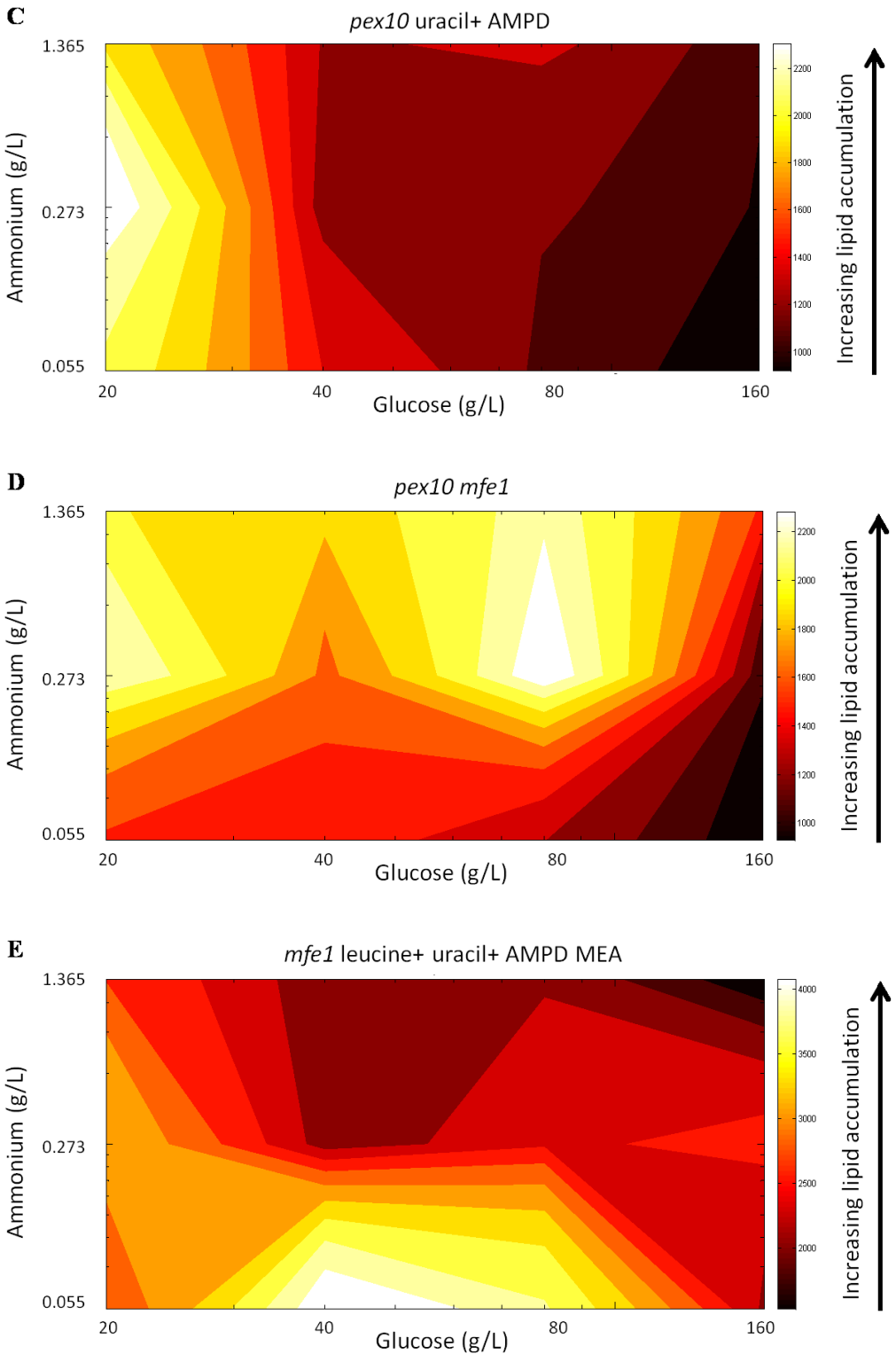


Figure 6.3 continued

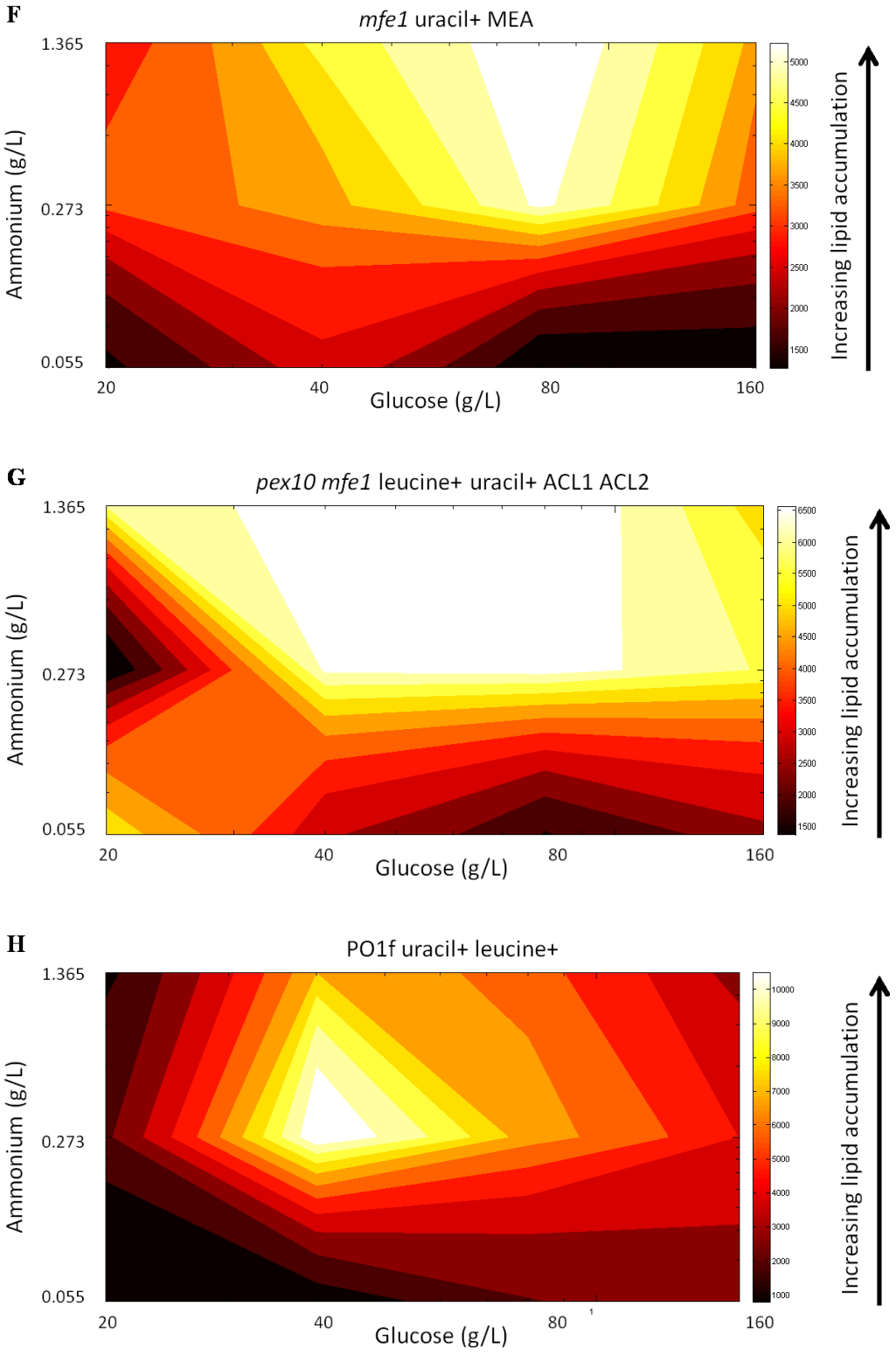


Figure 6.3 continued

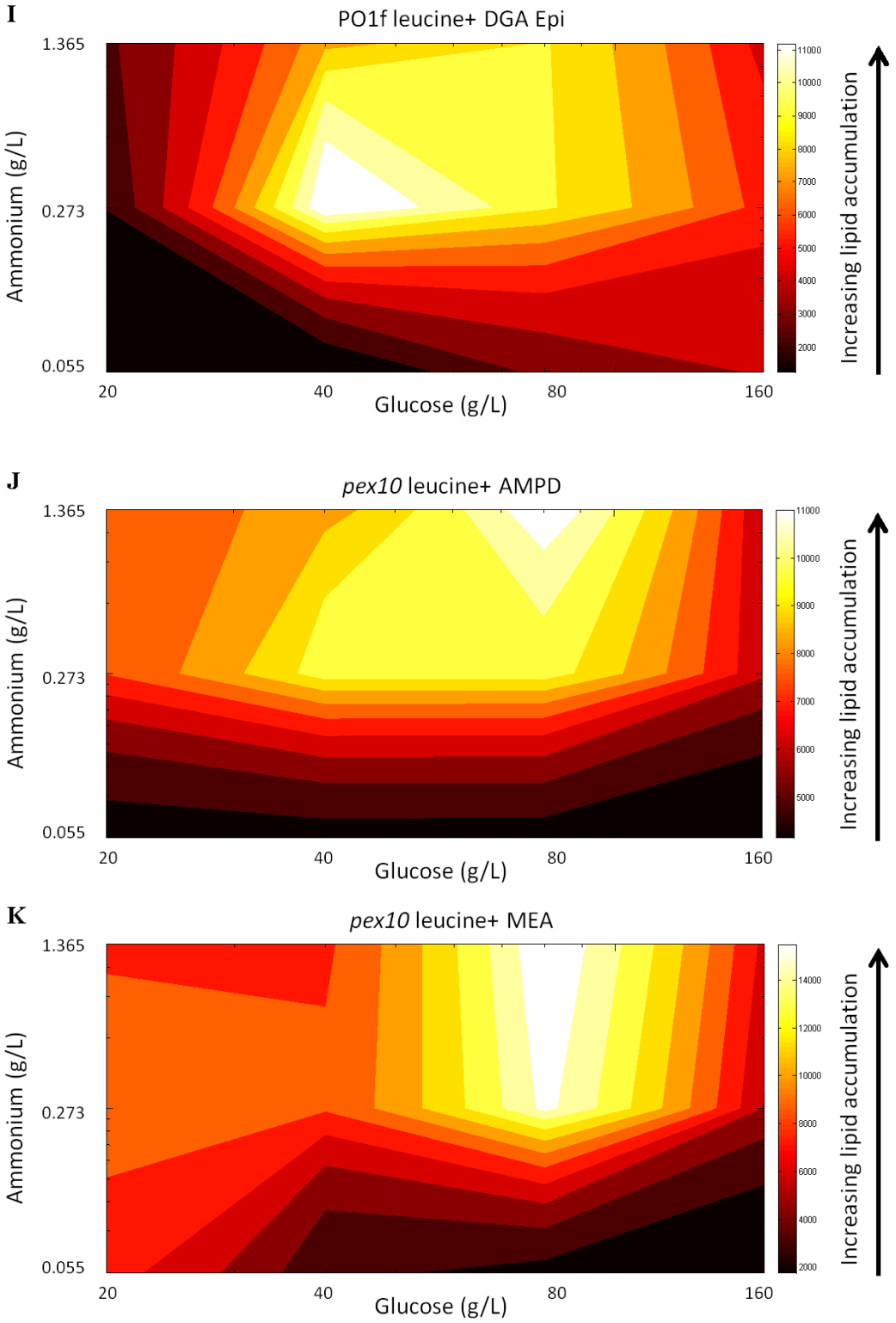
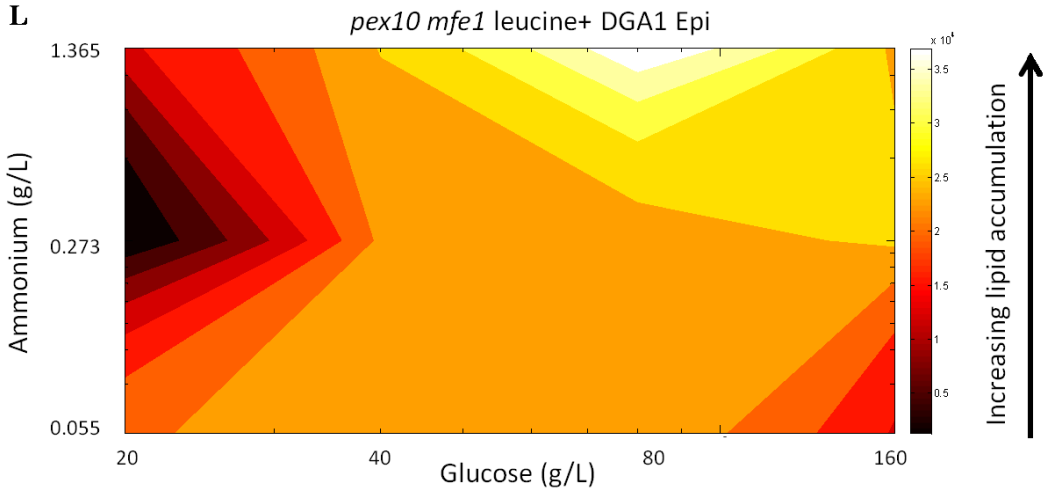


Figure 6.3 continued



Heat maps of nile red stained lipid fluorescence for the twelve strains listed in **Table 6.3** when cultivated in media formulations varying glucose between 20g/L and 160g/L and ammonium between 0.055g/L and 1.365g/L. (**A**) PO1f: Low lipid accumulator; Highest fluorescence values were induced by 20g/L glucose and 0.273g/L ammonium. (**B**) *mfe1* leucine$^+$ uracil$^+$ ACL1 ACL2: Low lipid accumulator; Highest fluorescence values were induced by 20g/L glucose and 0.273g/L ammonium. (**C**) *pex10* uracil$^+$ AMPD: Low lipid accumulator; Highest fluorescence values were induced by 20g/L glucose and 0.273g/L ammonium. (**D**) *pex10 mfe1*: Low lipid accumulator; Highest fluorescence values were induced by 80g/L glucose and 0.273g/L ammonium. (**E**) *mfe1* leucine$^+$ uracil$^+$ AMPD MEA: Medium lipid accumulator; Highest fluorescence values were induced by 40g/L glucose and 0.055g/L ammonium. This represents the only strain of the twelve tested to be induced optimally with 0.055g/L ammonium. (**F**) *mfe1* uracil$^+$ MEA: Medium lipid accumulator; Highest fluorescence values were induced by 80g/L glucose and 1.365g/L ammonium. (**G**) *pex10 mfe1* leucine$^+$ uracil$^+$ ACL1 ACL2: Medium lipid accumulator; Highest fluorescence values were induced by 80g/L glucose and 1.365g/L ammonium. (**H**) PO1f leucine$^+$ uracil$^+$: Medium lipid accumulator; Highest fluorescence values were induced by 40g/L glucose and 0.273g/L ammonium. (**I**) PO1f leucine$^+$ DGA Epi: Medium lipid accumulator; Highest fluorescence values were induced by 40g/L glucose and 0.273g/L ammonium. (**J**) *pex10* leucine$^+$ AMPD: High lipid accumulator; Highest fluorescence values were induced by 80g/L glucose and 1.365g/L ammonium. (**K**) *pex10* leucine$^+$ MEA: High lipid accumulator; Highest fluorescence values were induced by 80g/L glucose and 1.365g/L ammonium. (**L**) *pex10 mfe1* leucine$^+$ DGA1 Epi: High lipid accumulator; Highest fluorescence values were induced by 80g/L glucose and 1.365g/L ammonium.

6.3.3 Optimizing fermentation conditions in a bioreactor

Taken together, these dominant lipogenesis targets and carbon levels (specifically, 80 g/L glucose) enabled an optimization of fermentation conditions for the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain to maximize lipid accumulation in a bioreactor setting. We cultivated this fully prototrophic strain in an inexpensive minimal media formulation consisting of only glucose, ammonium sulfate, and yeast nitrogen base. By additionally controlling pH, temperature, and dissolved oxygen levels, we observed significantly

improved lipid titer to 16.1 g/L with cells containing up to 88% lipid cellular content, the highest reported yield and content to date (**Figure 6.4a-d)**. This represents a 5.4 fold increase over a POlf leucine^{+} uracil^{+} control and 63% of the theoretical stoichiometric yield. Moreover, this engineered strain exhibit a significantly reduced citric acid production rate, and decreased biomass formation (**Figure 6.4c,d**). Increasing overall glucose availability to (160g/L with double ammonium levels to 2.73g/L) increased titer to 25.3 g/L, again the highest reported titer (**Figure 6.4a**). As we have observed previously, maintaining identical C:N ratio while altering glucose and ammonium availability did result in identical lipogenesis. When cultivating the *pex10 mfe1* leucine^{+} uracil^{+} DGA1 strain in the 160g/L glucose - 2.73g/L ammonium media, we saw lipid content reaching only 70% of cell dry weight and reached only 50% of theoretical yield.

Figure 6.4: Overall increases in lipid content and titer from controlled fermentations

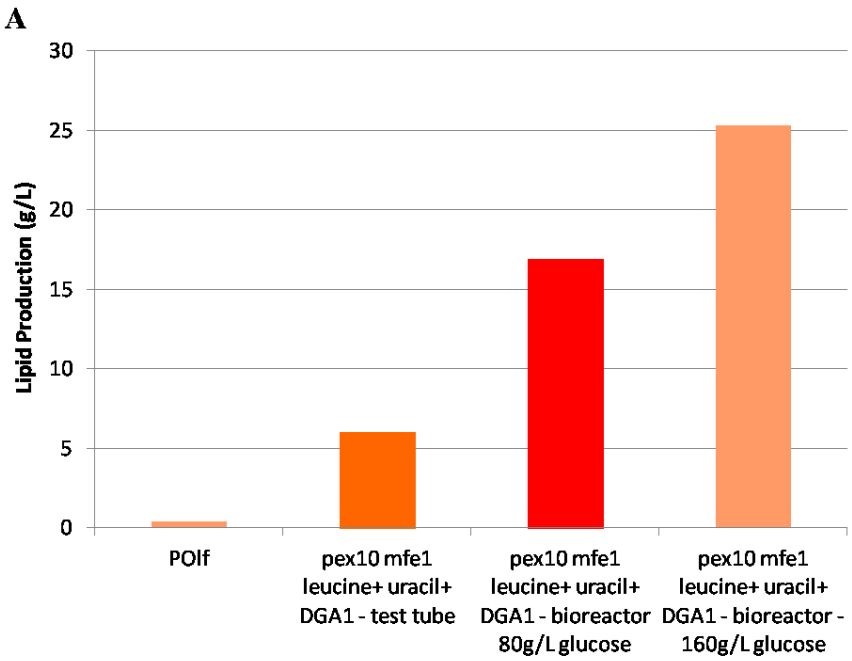
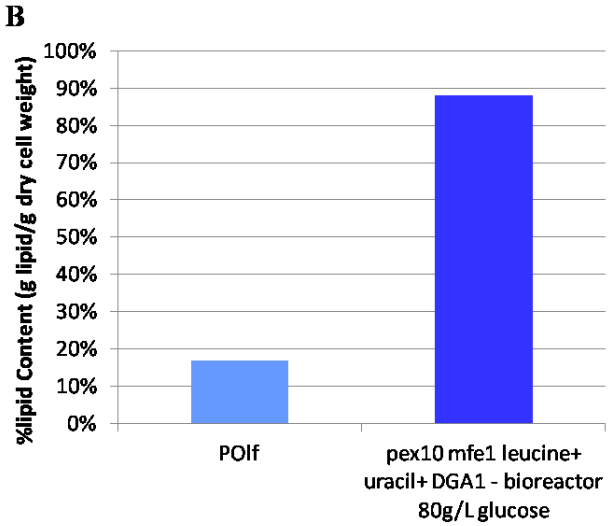
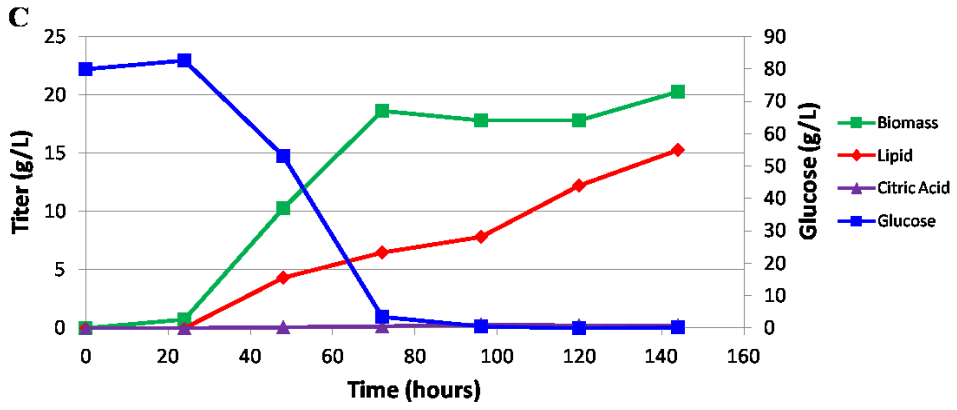
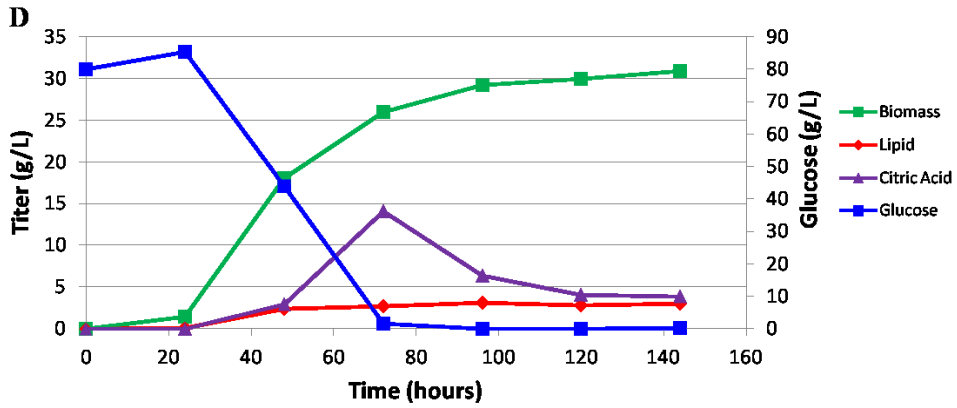


Figure 6.4 continued





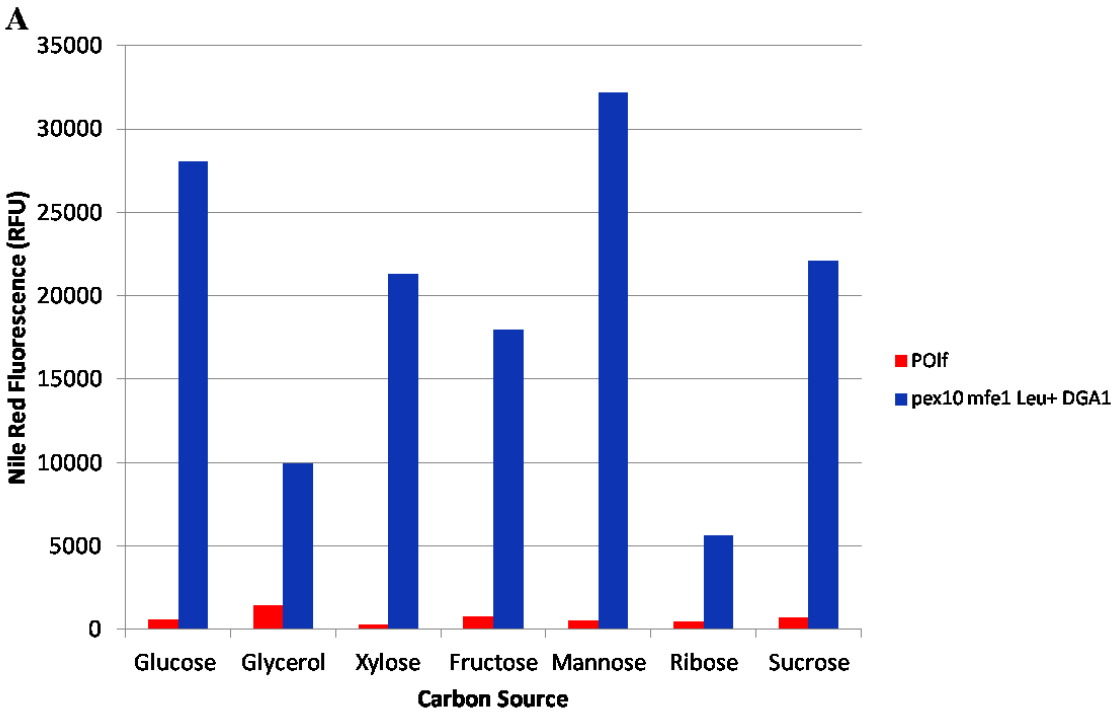


(**A,B**) Increase in lipid titer and lipid content from POlf to *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 at small scale production to *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 fermentation in bioreactor with 80g/L glucose and 1.365 g/L ammonium or 160g/L glucose and 2.73g/L ammonium. (**C,D**) Fermentation profiles for the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 (**C**) and POlf leucine$^+$ uracil$^+$ (**D**) strains cultivated in media containing 80g/L glucose and 1.365 g/L ammonium. The POlf leucine$^+$ uracil$^+$ strain accumulated more than 2g/L citric acid not seen in the engineered strain. Dissolved oxygen levels were held at or above 0.50.

6.3.4 Lipogenesis on alternative carbon sources, analysis of lipid content, and biodiesel synthesis

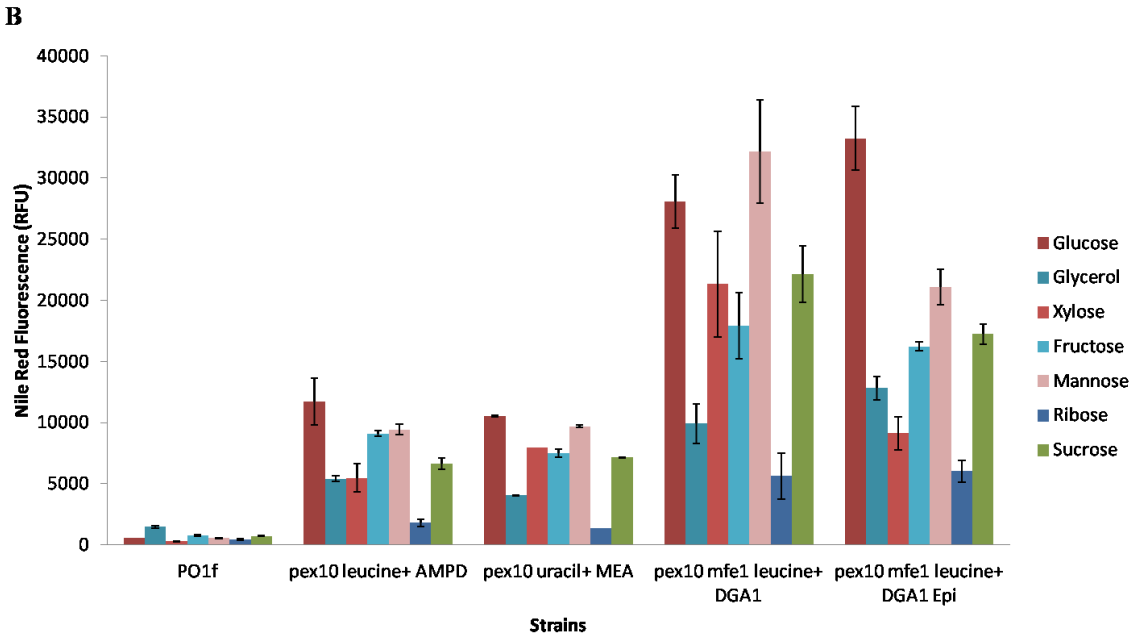
We further tested our *pex10 mfe1* leucine^{+} uracil^{+} DGA1 strain and five other strains on alternative carbon sources to assay for carbon-source independent lipogenesis. We observed that the *pex10 mfe1* leucine^{+} uracil^{+} DGA1 strain exhibited superior production in nearly all carbon sources, establishing these lipogenesis targets as essential in rewiring this organism into an oleochemical platform strain (**Figure 6.5a,b**).

Figure 6.5: Lipogenesis on alternative carbon sources



(**A**) The *pex10 mfe1* leucine^{+} uracil^{+} DGA1 strain effectively generates lipid content when grown on a variety of sugar and carbon sources. The PO1f base strain is included for comparison. Error bars represent standard deviations of biological triplicates

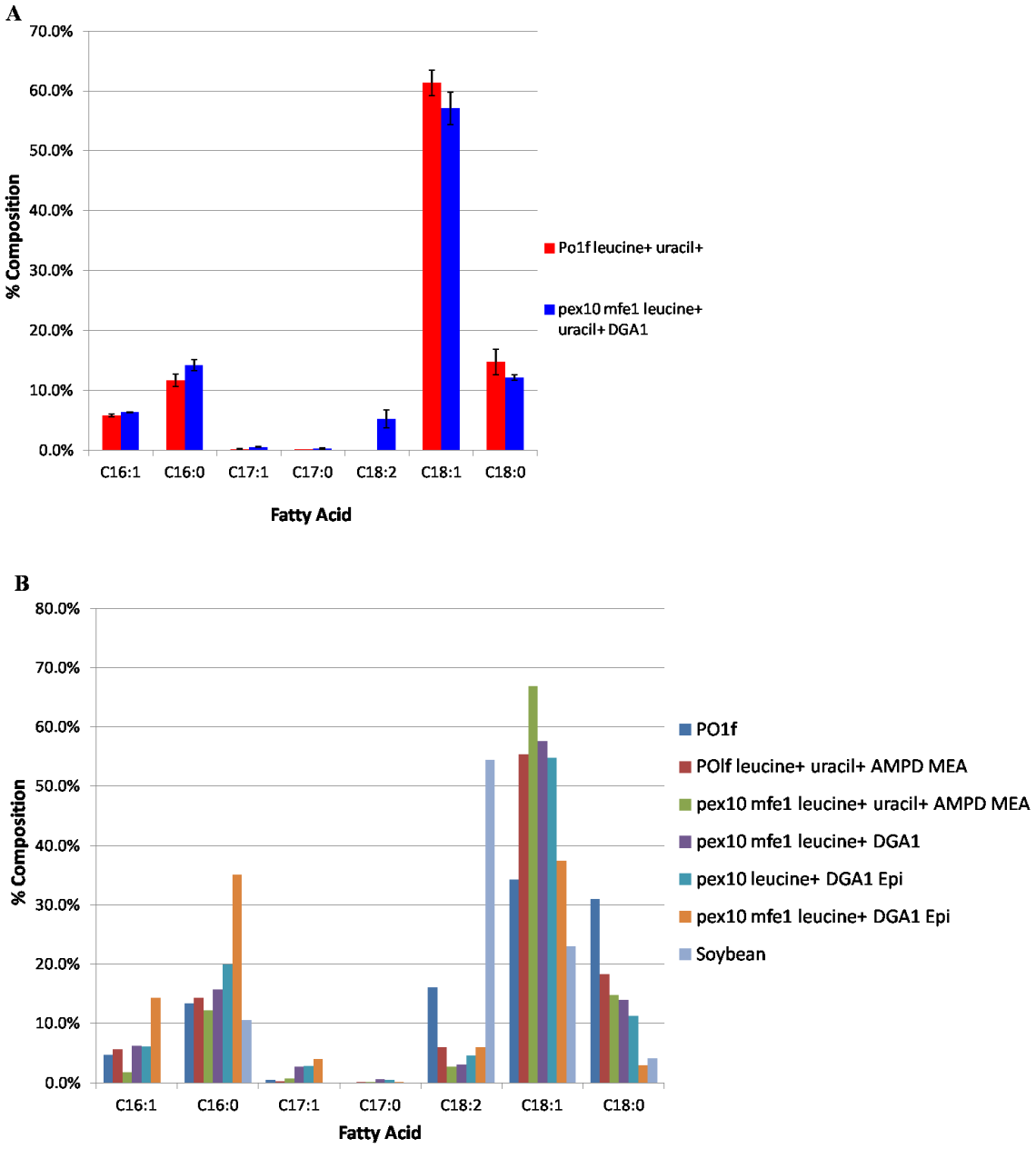
Figure 6.5 continued



(**B**) We analyzed the ability of unmodified POlf, *pex10* uracil^{+} MEA1p, *pex10* leucine^{+} AMPD, *pex10 mfe1* leucine^{+} DGA1 Epi, and *pex10 mfe1* leucine^{+} DGA1 Epi to generate lipids when utilizing glucose, glycerol, xylose, fructose, mannose, ribose, or sucrose as their sole carbon source. *Pex10 mfe1* leucine^{+} DGA1 strains (expressed chromosomally or episomally) generated the highest lipid content in all cases, and all engineered strains demonstrated the capacity to utilize each carbon source for lipid production. Glucose and mannose utilization generated the highest nile red stained lipid fluorescence, while ribose generated the least. Error bars represent standard deviations of biological triplicates

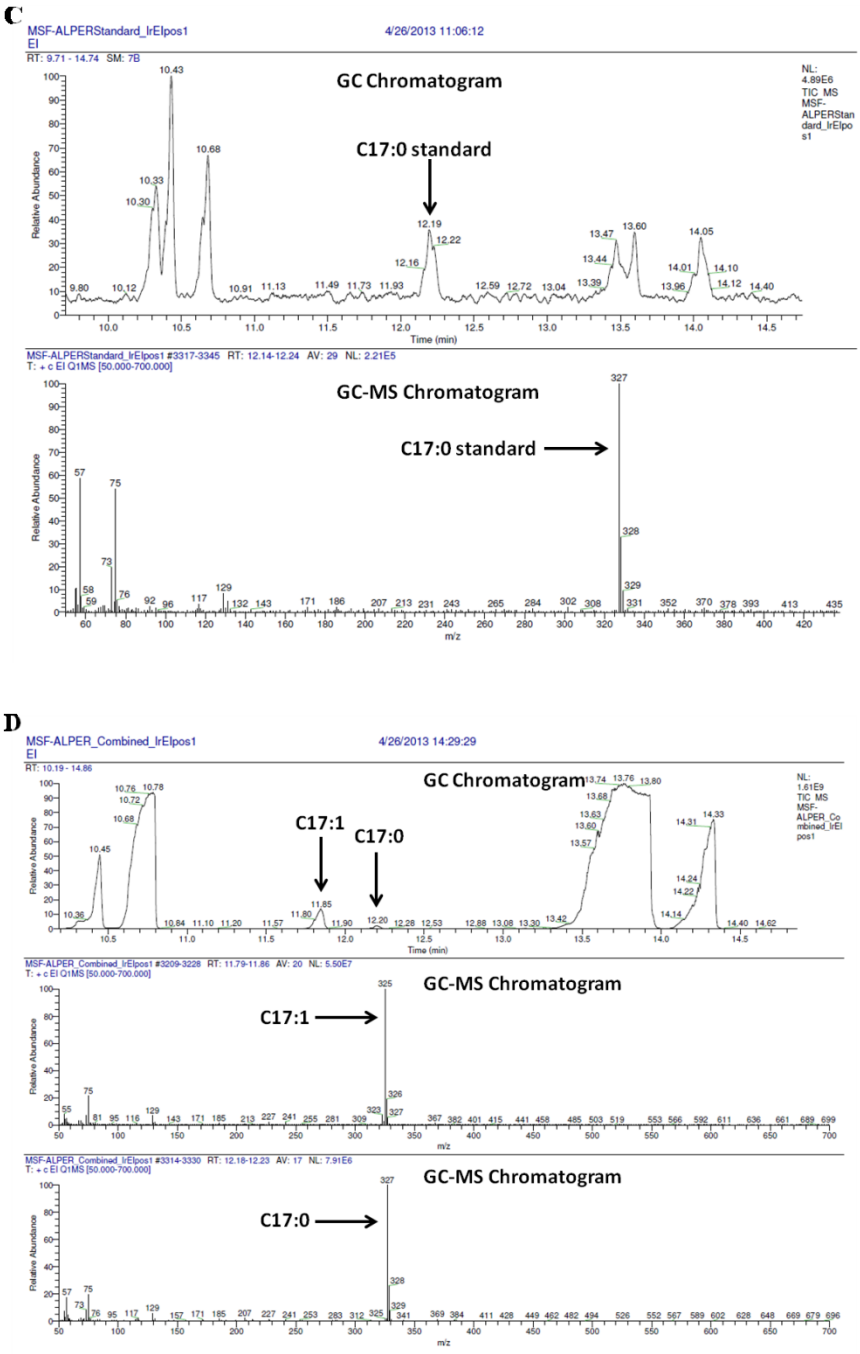
We analyzed lipid content of several strains with GC and GC-MS and saw predominantly C16:0, C16:1, C18:0, C18:1, and C18:2 fatty acids (very similar to a soybean-derived biodiesel profile), making these lipid reserves ideal feedstock for biodiesel synthesis [183]. We also produced 127 mg/L C17 fatty acids, a rare metabolite in cells (**Figures 6.6a-d**). We hypothesize that extremely active lipogenesis enables less favored odd chained synthesis and such high titers permit detection and characterization. Finally, a standard methanol transesterification reaction of bioreactor-extracted lipids completed *de novo* biodiesel production (**Figure 6.6e,f**).

Figure 6.6: Analysis of lipid composition and transesterification to biodiesel



(**A**) We performed GC analysis of lipid extract from strains *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 and POlf leucine$^+$ uracil$^+$ after 6-day bioreactor batch fermentations to determine fatty acid profile. We observed predominantly C16 and C18 fatty acid content, as expected, with a noticeably amount of C17 accumulation. We observed C18:2 (linoleic acid) production in the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 but not in the POlf leucine$^+$ uracil$^+$ control. (**B**) We performed GC and GC-MS analysis of lipid extract from strains POlf, POlf leucine$^+$ uracil$^+$ AMPD MEA, *pex10 mfe1* leucine$^+$ uracil$^+$ AMPD MEA, *pex10 mfe1* leucine$^+$ DGA1, *pex10* leucine$^+$ DGA1 Epi, *pex10 mfe1* leucine$^+$ DGA1 Epi. We observed predominantly C18:1, C18:0, and C16:0 fatty acid content, and noticed a general tendency towards increasing prevalence of C16s, C17s, and desaturated bonds as lipid accumulation capacity increased. Typical soybean fatty acid profile is included for comparison.

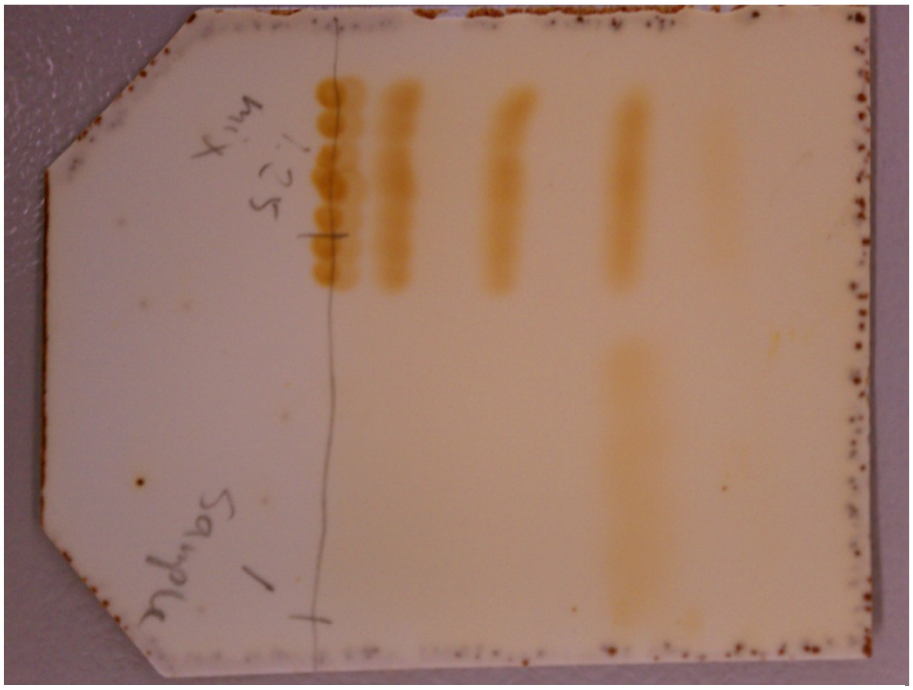
Figure 6.6 continued



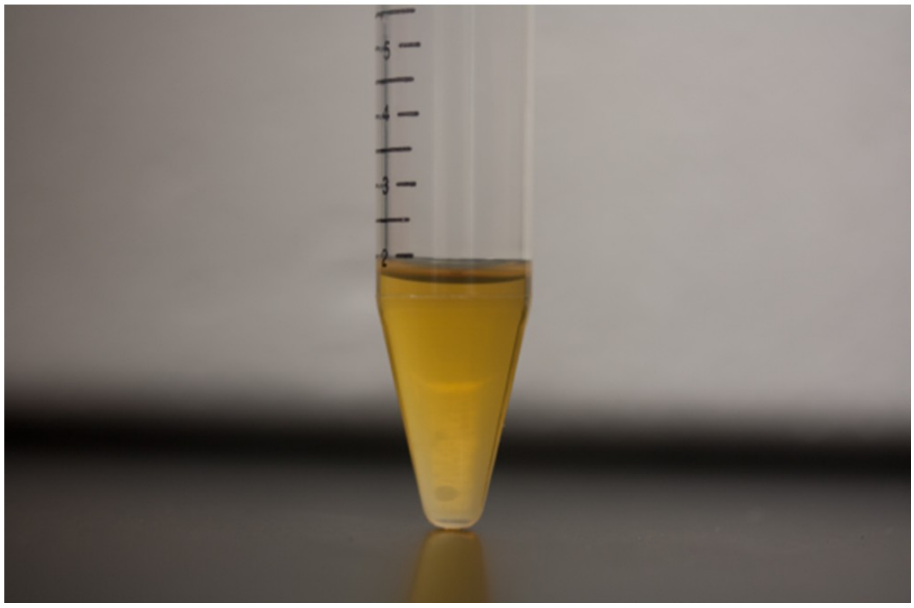
(**C**) The C17:0 hexadecanoic acid standard was analyzed with GC (top panel) and GC-MS (bottom panel) and showed the major expected peak of 327. (**D**) Lipid extract from the *pex10 mfe1* leucine^{+} DGA1 strain was analyzed with GC (top panel) and GC-MS (bottom two panels). Coinciding with the C17:0 standard retention time, C17:0 was present between 12.16-12.23, and analysis with GC-MS and showed the major expected peak of 327 (bottom panel). C17:1 was present on the GC between 11.79 and 11.86, and analysis with GC-MS (middle panel) revealed the expected major peak of 325, which corresponds to C17:0 losing two hydrogen atoms as it is desaturated to C17:1.

Figure 6.6 continued

E



F

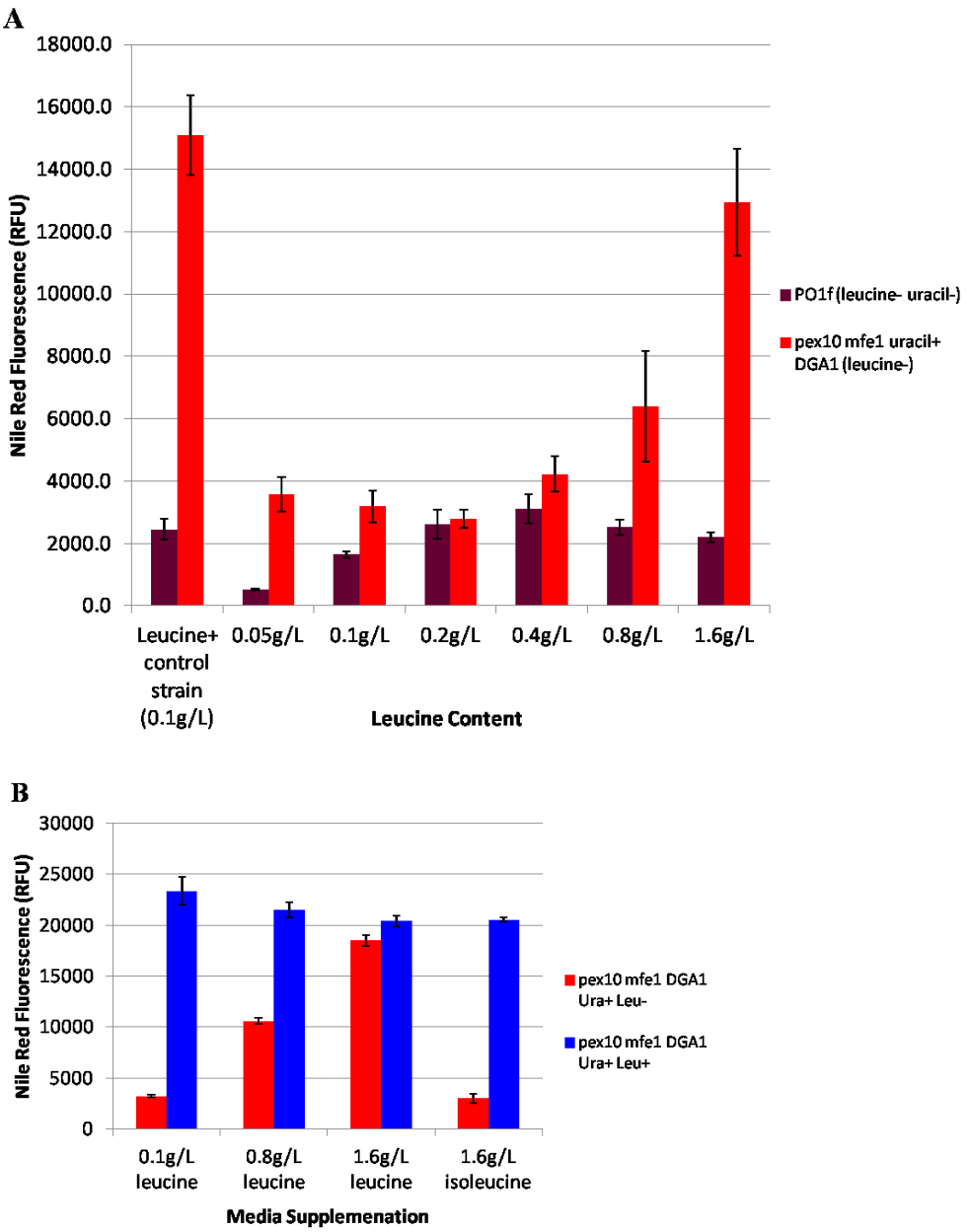


(**E**) We transesterified extracted lipid content from *pex10 mfe1* leucine^{+} uracil^{+} DGA1p cells fermented in a bioreactor with methanol to form FAMEs (biodiesel). We ran these FAMEs (bottom lane) on a TLC plate next to a mixed standard (top lane) containing free fatty acids, lipids, and glycerol and as expected, only FAMEs were present. (**F**) A picture of ~1500 µl sample of biodiesel is shown in a 15mL Falcon tube against a white backdrop.

6.3.5 Understanding lipogenesis in *Y. lipolytica*

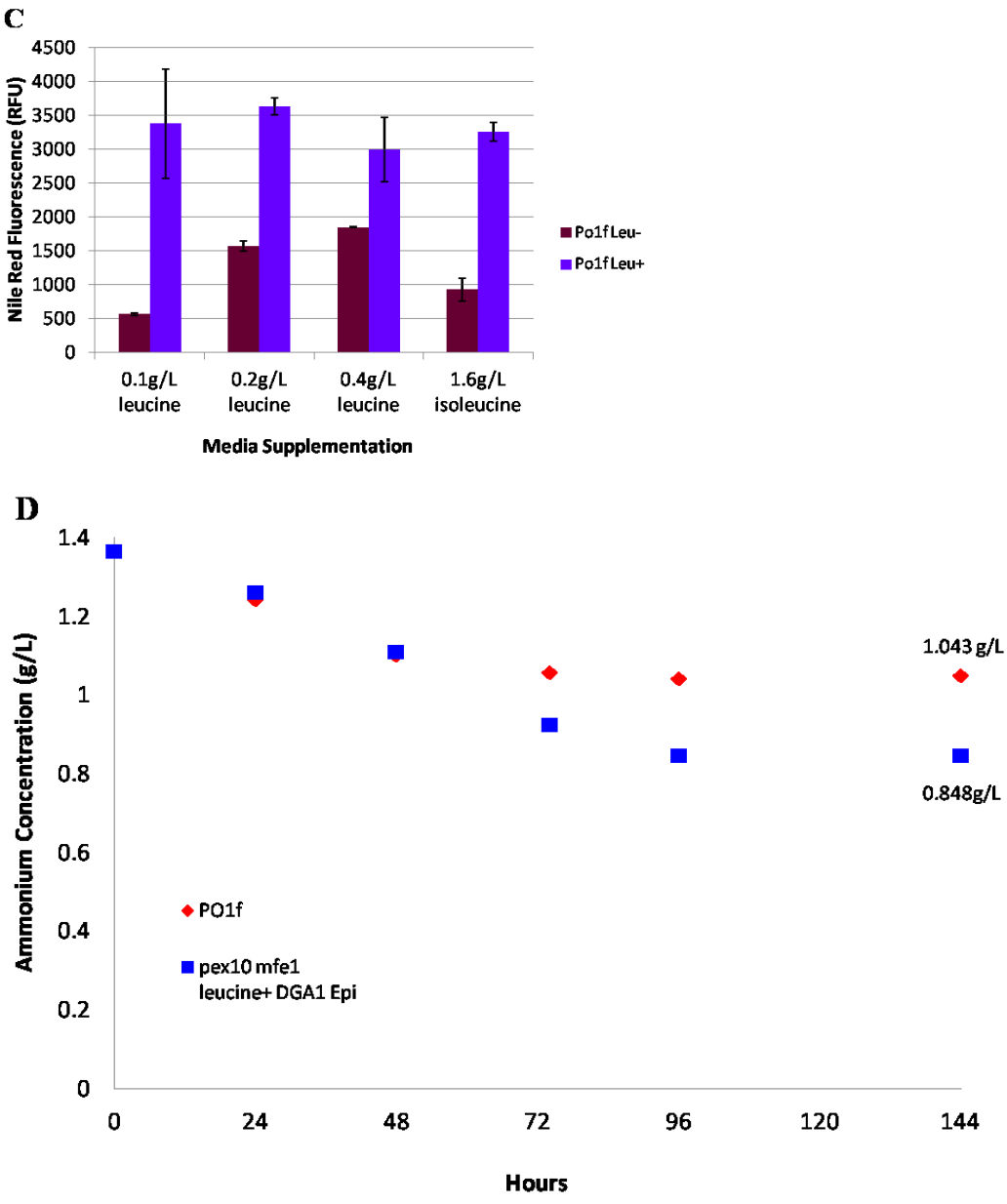
In performing these rewiring efforts on *Y. lipolytica*, we observed and explored several previously unreported facets of lipogenesis. First, we sought to explain the benefit bestowed by restoration of the leucine biosynthetic pathway towards lipogenesis by comparing leucine supplementation to genetic complementation in leucine$^-$ backgrounds. We observed that leucine supplementation mimicked genotypic complementation in leucine$^-$ strains but did not affect leucine$^+$ backgrounds (**Figure 6.7a-c**). Neither background responded to isoleucine supplementation, insuring that leucine-induced lipogenesis is not a result of leucine catabolism (**Figure 6.7b,c**). Furthermore, we determined that both leucine supplementation and genetic complementation can enable a pronounced reduction in steady state nitrogen concentration that correlates with lipogenesis in small scale cultivations (**Figure 6.7d,e**). Thus, we demonstrate that leucine acts as an intracellular trigger to enable lipid accumulation and this lipid accumulation requires a specific amount of nitrogen availability. We further determined that extremely high lipogenesis levels correlate with a metabolomic shift to sustain steady state ammonium level, again revealing the importance of nitrogen availability towards lipogenesis (**Figure 6.8a-c**). Specifically, we saw that ammonium was fully depleted in both the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1p and PO1f leucine$^+$ uracil$^+$ fermentations after three days, but was replenished in the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1p cultivation (**Figure 6.8a,b**). Ammonium replenishment was not solely a result of intracellular protein degradation as cellular protein content decreased in both strains as the fermentation progressed (**Figure 6.8c**).

Figure 6.7: Understanding lipogenesis by exploring the effects of leucine biosynthesis and nitrogen availability



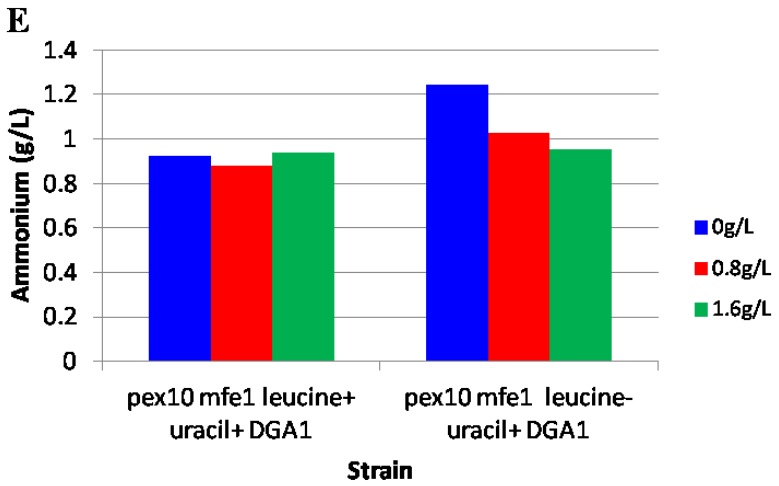
(**A**) We tested if leucine could induce lipid accumulation in POlf leucine$^-$ (dark red) to the same levels as seen in POlf leucine$^+$ (dark red – far left) and in *pex10 mfe1* leucine$^-$ DGA1 (light red) to the same levels as seen in *pex10 mfe1* leucine$^+$ DGA1 (light red – far left). We observed that 0.4g/L leucine supplementation could complement the leucine$^+$ phenotype in the unmodified PO1f leucine$^-$ strain, while 1.6g/L was necessary for the engineered *pex10 mfe1* leucine$^-$ DGA1 strain. (**B**) We tested the ability of leucine and isoleucine supplementation to complement the leucine$^+$ phenotype in the *pex10 mfe1* leucine$^-$ DGA1 background. Isoleucine had no effect on lipid accumulation. Error bars represent standard deviations of biological triplicates.

Figure 6.7 continued



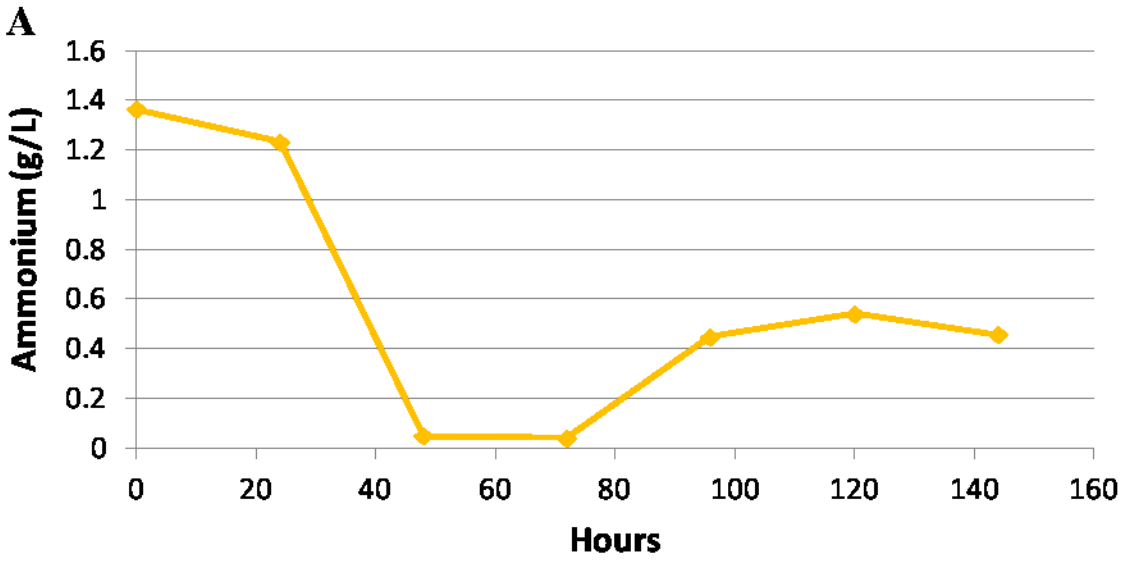
(**C**) We tested the ability of leucine and isoleucine supplementation to complement the leucine^{+} phenotype in the POlf leucine^{-} background. Isoleucine had no effect on lipid accumulation, while we observed that 0.4g/L leucine supplementation could partially complement the leucine^{+} phenotype in the unmodified PO1f leucine^{-} strain. Error bars represent standard deviations of biological triplicates. (**D**) We analyzed steady state ammonium concentration in PO1f and in the *pex10 mfe1* leucine^{+} DGA1 Epi background when cultivated in 30 mL media formulated with 80g/L glucose, 6.7g/L Yeast Nitrogen Base w/o amino acids, and 0.79g/L CSM supplement. High lipid production strain maintains a lower steady state ammonium concentration that POlf.

Figure 6.7 continued



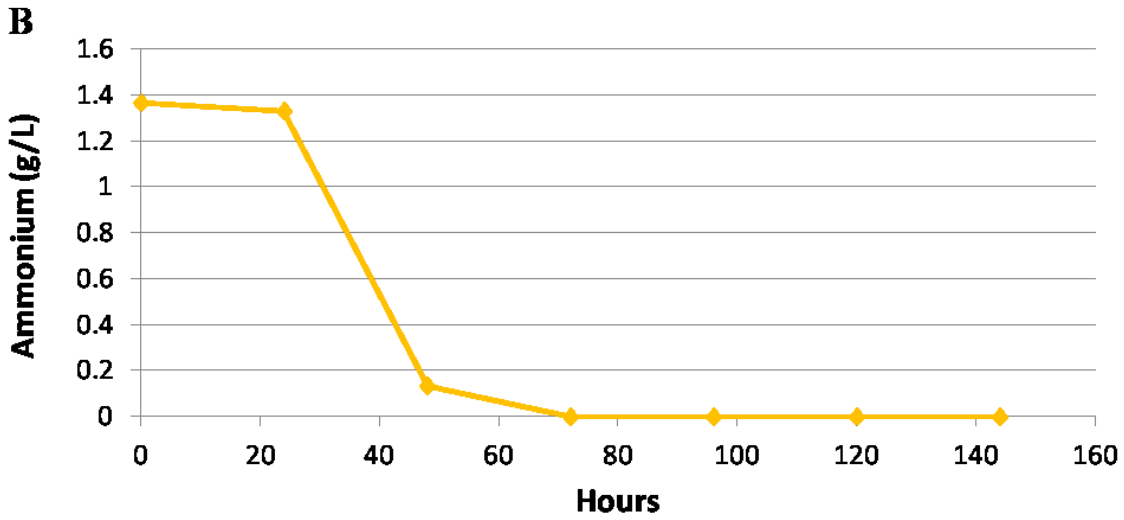
(**E**) We analyzed the effect of leucine supplementation on ammonium concentration in the *pex10 mfe1* leucine$^-$ uracil$^+$ DGA1 and the *pex10 mfe1* leucine$^-$ uracil$^+$ DGA1 background. The *pex10 mfe1* leucine$^-$ uracil$^+$ DGA1 had much higher ammonium levels when not supplemented with leucine. However, supplementation of leucine induced the *pex10 mfe1* leucine$^-$ uracil$^+$ DGA1 strain towards the leucine+ phenotype. Leucine supplementation had no effect on the *pex10 mfe1* leucine$^-$ uracil$^+$ DGA1 strain. Error bars represent standard deviations of biological triplicates

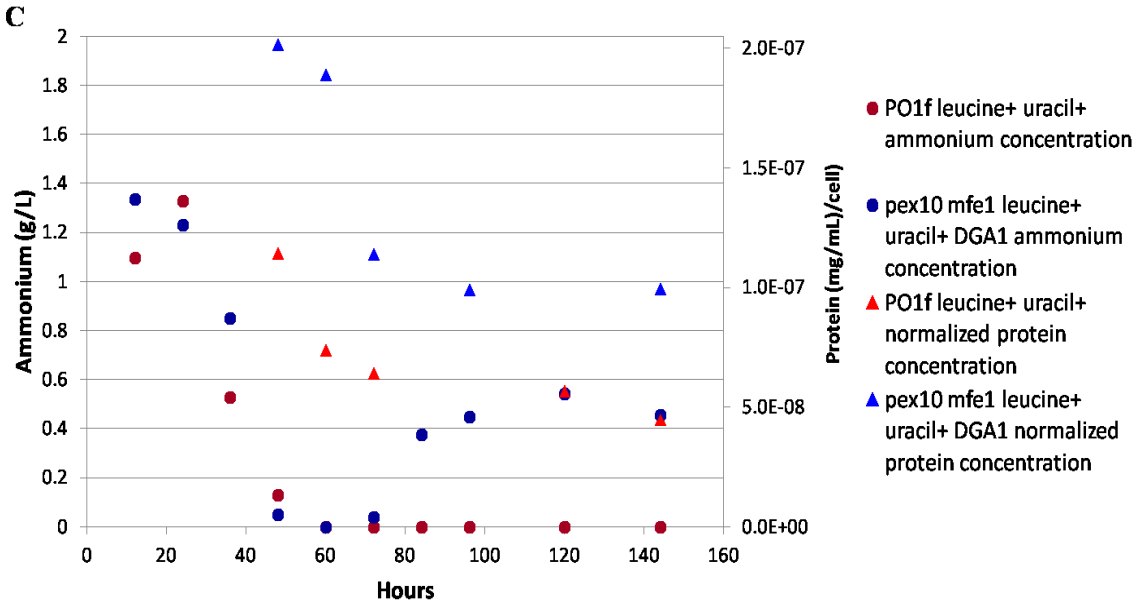
Figure 6.8: Ammonium is required for highly active lipogenesis



(**A**) A time course is shown of ammonium levels in the bioreactor batch fermentation of *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain in 80g/L glucose, 5g/L ammonium, and 1.7g/L YNB (no amino acids, no ammonium sulfate). Ammonium was fully consumed after 72 hours, but surprisingly, ammonium level was replenished to a steady state level of ~0.5g/L.

Figure 6.8 continued





(**B**) A time course is shown of ammonium levels in the bioreactor batch fermentation of PO1f leucine$^+$ uracil$^+$ strain in 80g/L glucose, 5g/L ammonium, and 1.7g/L YNB (no amino acids, no ammonium sulfate). Ammonium was fully consumed after 72 hours and was not replenished. (**C**) Ammonium concentration (left axis, circles) plummets to zero in both the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 and POlf leucine$^+$ uracil$^+$ fermentations. However, during *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 fermentation, ammonium levels are replenished. We tested if this could be a result of intracellular protein degradation by analyzing protein content (right axis, triangles) before, during, and after this dip in ammonium concentration. Cellular protein concentration decreases in both strains cells, so protein degradation is not the only mechanism used to replenish ammonium levels.

6.4 SUMMARY AND CONCLUSION

In summary, we have conducted the largest metabolic rewiring of an oleaginous organism ever attempted, and we have successfully engineered and enhanced lipid accumulation in *Y. lipolytica* by more than 60-fold. In doing so, we identified several unique features of lipogenesis, determined dominant genotype-phenotype dependencies, demonstrated that lipid accumulation approaching 90% of cell mass is possible, and demonstrated the feasible conversion of these lipids into FAMEs. This will serve as a stepping stone towards creating a robust, efficient, high production platform for ubiquitous conversion of carbon towards value-added oleochemical and biofuels products.

Chapter 7: Itaconic Acid Production in *Y. lipolytica*

7.1 Chapter Summary

During lipogenesis, carbon flux through the citric acid cycle is inhibited resulting in a buildup of citric acid and fatty acid precursors. We efficiently redirected this organic acid pool towards lipid accumulation in Chapter 6, but organic acids themselves (like citric acid and its derivatives) often represent value-added products [184]. Here, we have attempted to harness *Y. lipolytica's* capacity to accumulate citric acid for the production of itaconic acid, a top 12 value-added chemical from biomass according to the United States Department of Energy's 2004 report [184]. Specifically, we utilized our strong UAS1B$_{16}$-TEF promoter to overexpress a heterologous *cis*-aconitic acid decarboxylase enzyme (CAD1) to catalyze the conversion of *cis*-aconitic acid to itaconic acid. We produced 33mg/L itaconic acid with episomal CAD1 expression in an unmodified *Y. lipolytica* PO1f background, and co-overexpression of the native AMP deaminase enzyme (AMPD) increased itaconic acid titers to 141mg/L. Integration of CAD1 enzyme drastically increased titers to 136mg/L and 226mg/L for the unmodified and AMPD overexpression backgrounds, respectively. Utilization of minimal media formulations further increased itaconic acid titer to our highest production of 273mg/L, over eightfold higher than the starting strain. Optimization of carbon and nitrogen availability also showed promise towards increasing itaconic acid production by enabling high-level production of citric acid precursor. Our work represents the first attempt to use *Y. lipolytica's* capacity to accumulate organic acids to produce nonnative chemicals.

7.2 INTRODUCTION

Control of metabolic flux in model hosts can enable efficient production of commodity chemicals. Of particular interest are compounds that serve as renewable alternatives for petroleum derived chemicals. In this regard, itaconic acid is a naturally produced versatile monomer with diverse applications as a copolymer in traditionally petro-derived plastics and rubbers [185-187], and polymerized polyitaconic acid is a functional alternative for polyacrylic acid, a high volume commodity petrochemical [188,189]. In fact, itaconic acid was named a top 12 value-added chemical from biomass by the Department of Energy in 2004 [184]. The expanding market for itaconic acid derived products requires a concurrent decrease in production costs [187,188]. Here, we have undertaken the metabolic engineering of *Yarrowia lipolytica* for itaconic acid production.

Itaconic acid production was first observed in *Aspergillus itaconicus* in 1932 [190], and current industrial production occurs through fermentation by *Aspergillus terreus* [5]. Moderate successes towards rationally engineering *Aspergillus* species for increased itaconic acid yield have been described but remain difficult due to a lack of well-defined genetic tools [5,191]. Media optimization and large-scale mutagenesis approaches have proven far more successful towards increasing itaconic acid yield [192,193]. While high titers of itaconic acid have been realized in *A. terreus* fermentations, *A. terreus* growth is naturally inhibited in media formulated to induce itaconic acid accumulation, and *A. terreus* is negatively impacted by shear stress, preventing cultivation in standard stirred tank bioreactors typically utilized for microorganism fermentations [5,193,194]. Metabolic profiling of *A. terreus* demonstrated that itaconic acid production proceeds through the decarboxylation of the cis-aconitate TCA cycle intermediate [83], and the *cis*-aconitic acid decarboxylase encoding gene (CAD1) was recently isolated and functionally confirmed [195]. Thus, the potential to import CAD1 functionality into a model host organism is an

ideal possibility towards furthering itaconic acid biosynthetic capacity while avoiding difficulties inherent in *A. terreus* cultivations.

Lipid accumulation in *Y. lipolytica* proceeds via accumulation of citric acid cycle intermediates, mainly citric acid and isocitric acid [59]. Citric acid and isocitric acid are interconverted in a two step enzymatic process by the TCA cycle enzyme, *cis*-aconitase. During this interconversion, the *cis*-aconitic acid intermediate is physically released from the *cis*-aconitase enzyme, allowing its use as a substrate for further catalysis [196]. Thus, we have attempted to divert *Y. lipolytica's* citric and isocitric acid reserves towards the production of itaconic acid by expressing the *cis*-aconitic acid decarboxylase enzyme. *Y. lipolytica* has a fully developed genetic toolbox to enable rational metabolic engineering approaches and is becoming a more standard host for biosynthetic production and metabolic engineering [1,2,4]. *Y. lipolytica* offers numerous advantages towards industrial fermentations including robust growth in low pH and lower temperatures, high tolerance to shear stress, lack of phage-contamination, and ease of separation. Here, we seek to enable heterologous production of itaconic acid by importing *cis*-aconitic decarboxylase activity into *Y. lipolytica*.

7.3 Results and Discussion

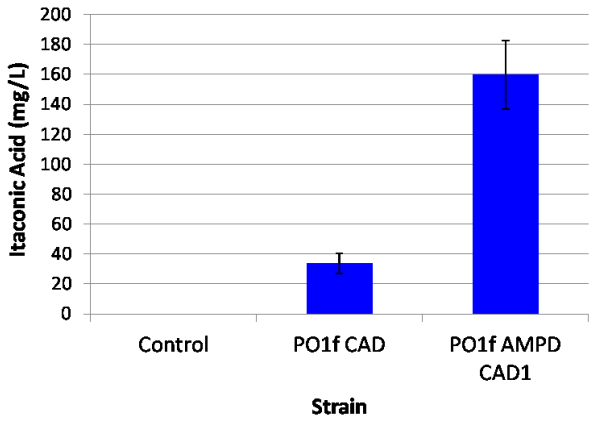
7.3.1 Episomal expression of the CAD1 gene in *Y. lipolytica*

Recent characterization of the *cis*-aconitic acid decarboxylase gene (CAD1) enables its utilization for itaconic acid production in microbial hosts. We inserted the CAD1 gene into a high-strength UAS1B$_{16}$-TEF expression cassette on an episomal plasmid to allow for expression in *Y. lipolytica*, and we observed 33 mg/L itaconic acid titer (**Figure 7.1**). This is the first time that itaconic acid has been produced by *Y.*

lipolytica and illustrates that CAD1 expression can enable itaconic acid production in *Y. lipolytica*. We should note that the CAD1 gene had not been codon optimized from its original codon usage in *Aspergillus terreus*. In fact, use of a CAD1 optimized for *S. cerevisiae* expression resulted in no itaconic acid production in *Y. lipolytica*. This demonstrates once again the importance of codon usage for heterologous protein expression in *Y. lipolytica* (see Chapter 2 - hrGFP expression) [2].

We attempted to increase itaconic acid production by expressing CAD1 (again episomally) in a *Y. lipolytica* strain with the AMP Deaminase (AMPD) enzyme constitutively overexpressed in a UAS1B_{16}-TEF-driven chromosomal expression cassette. Constitutive expression of AMPD should halt the citric acid cycle at the isocitric acid intermediate, effectively increasing *cis*-aconitic acid substrate levels. We observed a nearly fivefold increase in itaconic acid in this AMPD overexpression background strain, to 159 mg/L (**Figure 7.1**). Thus, AMPD overexpression is an effective measure towards increasing itaconic acid production. This increase is likely mediated through inhibition of the TCA cycle to increase organic acid substrate levels.

Figure 7.1: Itaconic acid production in *Y. lipolytica* with episomal CAD1 expression



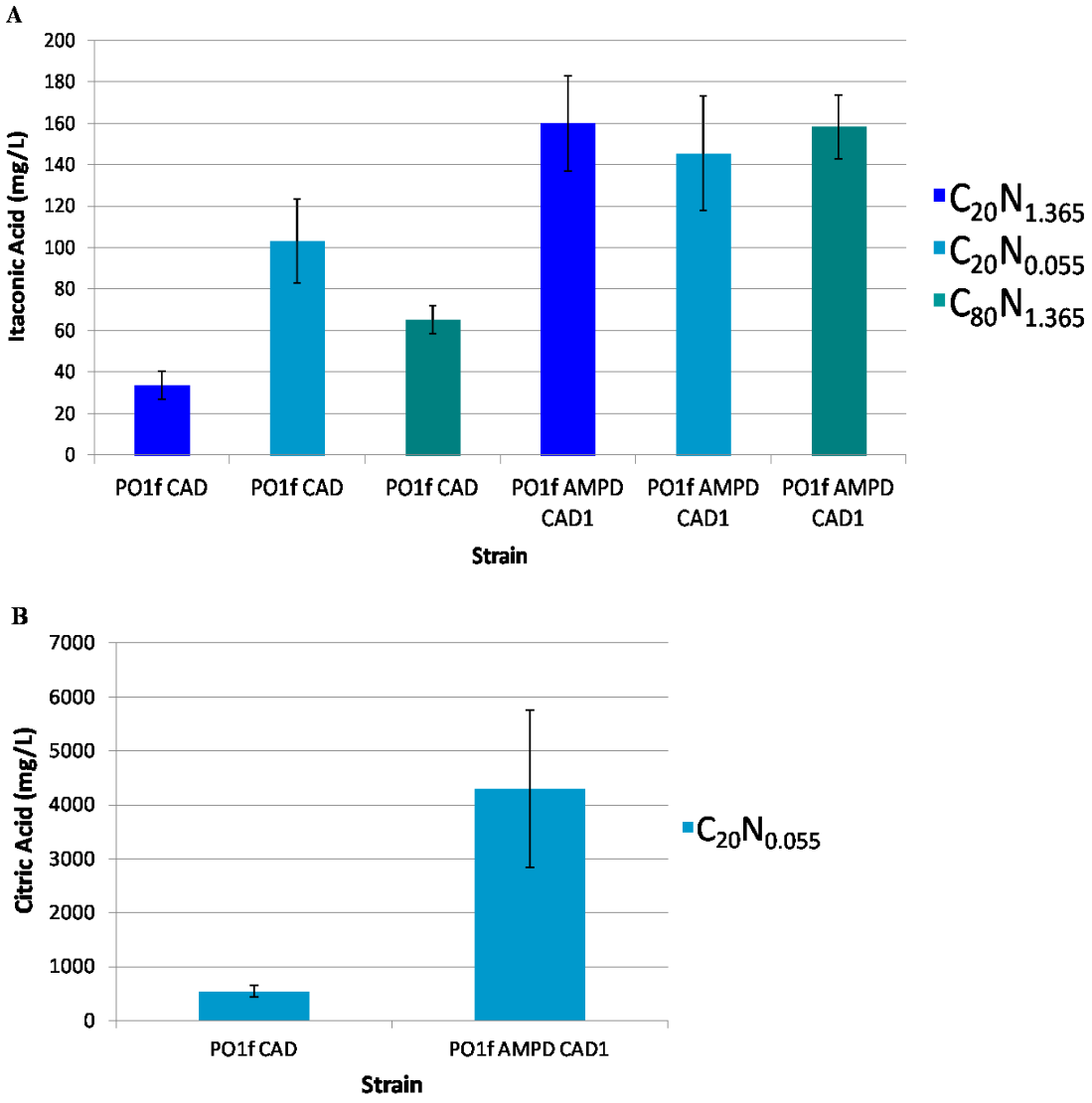
Episomal expression of the CAD1 enzyme, driven by the UAS1B_{16}-TEF promoter, in *Y. lipolytica* strain PO1f enables itaconic acid production at 33mg/L. Similar expression of CAD1 in a *Y. lipolytica* strain engineered to constitutively express AMPD increased itaconic acid production to 159mg/L. Strains were cultivated in $C_{20}N_{1.365}$ media with amino acid supplementation (described below). Error bars represent the standard deviation of biological triplicates.

7.3.2 Optimizing C:N ratio for itaconic acid production

We have shown that *Y. lipolytica's* central carbon metabolism is pliable to manipulation by overexpression of the AMPD enzyme for increased itaconic acid production. In Chapter 6, we thoroughly demonstrated that *Y. lipolytica's* lipid accumulation potential can be drastically manipulated by altering the ratio of carbon (glucose) and nitrogen (ammonium) in media formulations (C:N ratio). Lipid accumulation, C:N ratio, and citric acid accumulation (a metabolic precursor for *cis*-aconitic acid CAD1 substrate) are all linked in *Y. lipolytica* [59,62]. Increasing C:N ratio stimulates lipid accumulation to a certain extent, until further increases shuttle carbon directly into citric acid accumulation. We attempted to increase citric acid accumulation and itaconic acid production by increasing C:N ratio. Specifically, we cultivated unmodified *Y. lipolytica* PO1f and the *Y. lipolytica* AMPD overexpression background (both episomally expressing CAD1) in two new media formulations and assayed for itaconic acid production (**Figure 7.2a**). Our original media formulation contained 20g/L glucose and 1.365g/L ammonium ($C_{20}N_{1.365}$), and these two new formulations contained 20g/L glucose and 0.055g/L ammonium ($C_{20}N_{0.055}$), or 80g/L glucose and 1.365g/L ammonium ($C_{80}N_{1.365}$). All three formulations also contained yeast nitrogen base and CSM-leucine amino acid supplementation. We observed that increasing C:N ratio improved itaconic acid production to more than 100mg/L in the unmodified *Y. lipolytica* background, but had little benefit when the AMPD enzyme was coexpressed, instead resulting in drastic citric acid buildup to more than 4g/L (**Figure 7.2b**). We concluded that both AMPD constitutive expression and high C:N ratios inhibit the TCA cycle to increase organic acid levels. However, when these two methods are employed simultaneously, further metabolic rewiring is necessary to draw flux towards itaconic acid production. Only the $C_{20}N_{0.055}$ media formulation resulted in citric acid

accumulation, demonstrating the nitrogen limitation is more important than C:N ratio towards TCA cycle inhibition in the PO1f and PO1f AMPD backgrounds (**Figure 7.2b**).

Figure 7.2: Altering C:N ration to increase organic acid production

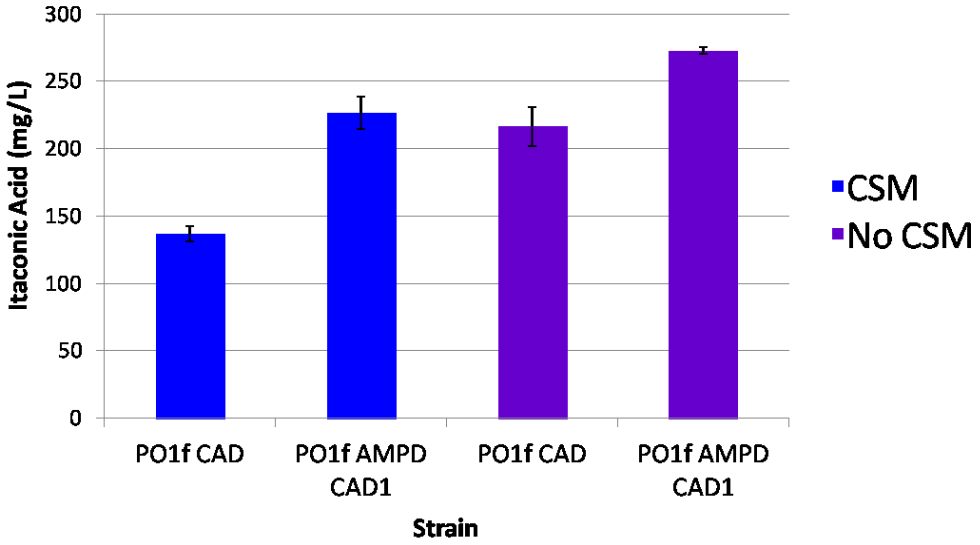


(**A**) Two *Y lipolytica* strains episomally expressing the CAD1 enzyme, unmodified strain PO1f and a PO1f derivative constitutively overexpressing AMPD, were analyzed for itaconic acid production when cultivated in three media formulations, $C_{20}N_{1.365}$, $C_{20}N_{0.055}$, and $C_{80}N_{1.365}$, where C and N represent g/L glucose and g/L ammonium, respectively. Increasing C:N ratio, by decreasing nitrogen level or increasing glucose level, effectively increased itaconic acid production in PO1f. No effect on itaconic acid production was seen in the AMPD expression background. (**B**) Interestingly, the $C_{20}N_{0.055}$ formulation stimulated exceedingly high citric acid production in the AMPD expression background. Other media formulation did not stimulate citric acid accumulation. Error bars represent standard deviations of biological triplicates.

7.3.3 Integration of CAD1 to increase itaconic acid production

We have previously observed increased protein activity using chromosomal expression compared to episomal expression in Chapter 1 [2]. Thus, we integrated the CAD1 gene into the PO1f and PO1f AMPD overexpression backgrounds and assayed for itaconic acid production (**Figure 7.3**). We observed a pronounced increase in itaconic acid production, suggesting that CAD1 expression is a limiting factor in our systems.

Figure 7.3: Chromosomal CAD1 expression increases itaconic acid production



PO1f and POlf AMPD overexpression backgrounds, harboring chromosomal CAD1 expression cassettes, were assayed for itaconic acid production after cultivation in standard $C_{20}N_{1.365}$ media (including CSM amino acid supplementation). Chromosomal CAD1 expression increased itaconic acid titers to 136 and 226mg/L for the PO1f and PO1f AMPD backgrounds, respectively. Cultivation in ($C_{20}N_{1.365}$) media without amino acid supplementation increased itaconic acid production to our highest yield of 272mg/L in the AMPD CAD1 chromosomal expression strain. Error bars represent the standard deviation of biological triplicates.

We have also observed an increase in fatty acid accumulation when cultivating in minimal media (media not containing CSM-based amino acid supplementation), specifically for bioreactor fermentations described in chapter 6. Thus, we assayed our two chromosomal CAD1 expression strains for itaconic acid when cultivated in minimal media ($C_{20}N_{1.365}$ without CSM). The PO1f chromosomal CAD1 expression strain required additional supplementation with 20mg/L uracil due to an auxotrophy, while

uracil auxotrophy had been alleviated in the AMPD overexpression background by insertion of the AMPD expression cassette (**Figure 7.3**). We observed another pronounced increase in itaconic acid production, culminating in 272mg/L produced by the AMPD overexpression background. Thus, reducing amino acid supplementation and increasing C:N ratio are important towards increasing itaconic acid production in *Y. lipolytica*.

7.4 SUMMARY AND CONCLUSION

Y. lipolytica has the capacity to accumulate lipid content and organic acids [9]. These accumulation mechanisms are interrelated, as fatty acid accumulation requires an inhibition and reversal of flux through the TCA cycle to supply acetyl-CoA and malonyl-CoA fatty acid building blocks. In *Y. lipolytica* strains without advanced lipid accumulation capabilities, TCA flux is inhibited but is not drawn towards fatty acid synthesis. Instead, organic acid intermediates are accumulated, predominantly as citric and isocitric acid. We have attempted to utilize these organic acid reserves for the production of itaconic acid, a value-added chemical monomer with diverse applications. Specifically, we enabled the first reported production of itaconic acid in *Y. lipolytica* through the episomal expression of a CAD1 (*cis*-aconitic acid decarboxylase) enzyme. The substrate for the CAD1 enzyme, *cis*-aconitic acid, is formed during a two-step isomerization reaction between citric acid and isocitric acid. Citric acid, isocitric acid, and *cis*-aconitic acid are present during standard cellular respiration. We attempted to increase *cis*-aconitic acid availability through the overexpression of an AMP deaminase enzyme, an enzyme know to inhibit the TCA cycle at the isocitric acid level. AMPD coexpression successfully increased itaconic acid, as did modulation of C:N content, a more thoroughly studied mechanism to inhibit TCA cycle flux. However, combining

these two methods did not increase itaconic acid synthesis, instead resulting in high levels of citric acid accumulation. Future work will be devoted to pulling this citric acid content towards itaconic acid production. We further increased itaconic acid by chromosomally expressing the CAD1 gene, a known method to increase gene expression by avoiding a document "half on/ half off" phenotype observed in centromeric *Y. lipolytica* plasmids [2]. Finally, we utilized a minimal media formulation to increase itaconic acid to our highest titer, 272 mg/L, illustrating once again that amino acid supplementation disrupts efficient diversion of TCA flux towards a desired product, be it fatty acid production (Chapter 6) or itaconic acid production.

Chapter 8: Major Findings and Proposals for Future Work

8.1 Major Findings

The experiments described herein had two major goals. First, develop and generalize the hybrid promoter engineering for widespread application, and second, engineer *Yarrowia lipolytica* to produce short chain alkanes, high lipid titers, and itaconic acid. For our first goal, we successfully employed the hybrid promoter approach to an expression limited system, *Y. lipolytica*, and increased overall promoter strength sevenfold. Our promoter libraries also enabled the first reported capacity for tunable gene expression in this organism, spanning a range of more than 400 fold in terms of mRNA levels. We then demonstrated *de novo* hybrid promoter construction in *Y. lipolytica* and applied the hybrid promoter approach to a second yeast, *S. cerevisiae*. In *S. cerevisiae*, we exploited a large collected of well characterized promoter parts to increase the expression of the strongest yeast constitutive (P_{GDP}) and inducible (P_{GAL}) promoters by 2.5 fold and 15%, respectively. We also engineered promoter regulation and imported synthetic galactose induction to constitutive promoters. Finally, we created a collection of galactose-inducible promoters that span a nearly 50-fold dynamic range of induced expression levels. Our results demonstrated that hybrid promoter engineering can efficiently increase expression capacity in two distinct eukaryotic lineages. Combination of modular UAS and core promoter elements enabled a predictable gene expression and regulation response in resultant hybrid promoters. Thus, we have opened the possibility of creating synthetic designer promoter constructs.

For our second goal, we successfully produced short-chain alkanes, itaconic acid, and high lipid titers in *Y. lipolytica*. Heterologous expression of a soybean lipoxygenase enzyme was necessary and sufficient to enable pentane production. Gene expression, media, and strain background improvements increased yield threefold, but pentane

production was still too low to consider scale up. We had much more success enabling the heterologous production of itaconic acid, up to 272mg/L with modest flux control and media optimization. Much work can be done to improve this titer, described below.

Our greatest success was engineering and enhancing lipid accumulation in *Y. lipolytica*. We fully sampled the lipogenesis metabolic landscape through AMPDp, ACLp, MAEp, DGA1p and DGA2p overexpressions multiplexed with deletions that prevent β–oxidation and peroxisome biogenesis. We determined three dominant genetic targets that exhibited cooperativity in the lipogenesis metabolic landscape—*pex10* deletion, DGA1p or DGA2p overexpression, and restoration of a complete leucine biosynthetic pathway (leucine$^+$ genotype) – to enhance lipid levels nearly 15-fold over control to 6.00 g/L in small scale cultivation. We analyzed the effect of nitrogen starvation and carbon availability on lipogenesis to understand the complex relationship between *de novo* lipid accumulation, strain genotype, and nutrient levels. We realized that the current paradigm stating that - (1) nitrogen starvation is necessary for lipogenesis and (2) similar C:N ratios beget similar induction - is incorrect. Instead, a defined amount of carbon content ultimately controls lipid synthesis, and this favorable carbon level increases in strains capable of superior lipogenesis. Controlled fermentation of the *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain improved lipid titer to up to 25.3 g/L (representing a more than 60-fold improvement over parental strain and conditions) and produced cells almost fully consisting of lipid content (88% lipid dry cell weight). These represent the highest reported lipid yield and content to date. Finally, we demonstrated production of rare C17 fatty acids, and determined that leucine biosynthetic capacity stimulates lipogenesis.

Engineering *Y. lipolytica*'s metabolism for lipid and itaconic acid production entailed simultaneous stimulus of phenotypic-induction (through media optimization).

Our experience maximizing phenotypic output in this manner allowed use to realize one essential point – that genotypic perturbations (even seemingly benign modifications like auxotrophy relief) can drastically redistribute intracellular metabolic flux and should be followed by a re-optimization of media formulation. Specifically, we observed how altering carbon and nitrogen availability produces different phenotypic responses in a strain dependant manner. Although labor intensive, this property increases the pliability of *Y. lipolytica's* central carbon metabolism for redirection towards multiple secondary metabolites.

8.2 Proposals for Future Work

The studies described herein demonstrate how to enable metabolic engineering applications in a non-conventional host. We have described a start to finish protocol to improve native biosynthetic capacity in an organism with ill-defined gene expression tools. It would be interesting to see if this protocol would be effective in other poorly characterized hosts that have basic recombinant DNA technology (and the native capacity to produce an interesting product). For instance, itaconic acid production in *Aspergillus* species [187] or alkane production in algae populations [197] may be improved using our general techniques of promoter engineering, high-strength heterologous and endogenous enzymatic overexpressions, large-scale strain development, and thorough media optimization.

As described above, we developed a generalizable hybrid promoter engineering approach to modulate and increase promoter strength in two fungal species. This method should be applied to other industrially relevant hosts to illustrate its utility across eukaryotic systems. In particular, gene expression is a major limiting factor towards protein and therapeutic production in mammalian cell lines. Utilizing a hybrid approach

is an ideal method to generate high capacity promoters that efficiently drive gene expression regardless of cell line or integration loci. Suitable UAS and strong core promoter elements have been characterized in mammalian systems [198,199] and fluorescent reporter proteins are available [200], enabling immediate mammalian hybrid promoter construction.

Our production of pentane is significant as it represents the first time this molecule has been produced synthetically in an organism. Nevertheless, there is considerable strain engineering that must be done to further improve yields and titers. Specifically, free linoleic acid availability must be increased. Efforts to increase free fatty acids through expression of thioesterases [63,88,158,201] must be used in *Y. lipolytica* to further increase pentane production. An additional consideration toward improving yield is the byproduct formed through the lipoxygenase reaction. In the case of pentane production, 13-HPOD is also converted into 13-oxo-cis-9-trans-11-tridecadienoic acid. To increase overall yield, it will be necessary to catabolically recycle this byproduct back into linoleic acid. Finally, we demonstrated that the soybean lipoxygenase enzyme is sufficient to produce the short-chain alkane pentane. It is conceivable that protein engineering efforts can be used to alter either the specificity of substrate or the product profile from this enzyme. In fact, our lab recently engineered an *E. coli* biosensor to be highly responsive to pentane and hexane [202]. This biosensor can be utilized to isolate mutant lipoxygenases that can produce alternative short-chain alkanes. Finally, we have produced a robust, efficient, high production lipid accumulator, the *Y. lipolytica* PO1f *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain. We should assay this strain for pentane production after inserting a heterologous Gmlox1 expression cassette.

Our *Y. lipolytica* PO1f *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1 strain accumulates the highest lipid levels reported, and we have shown that *Y. lipolytica* is amenable to further

in vivo catalysis of its lipid reserves in Chapter 5 [4]. Overexpression of fatty acid or lipid modifying enzymes can isolate or further alter the carbon backbone. For instance, coexpression of a fatty acyl reductase and an aldehyde decarbonylase could enable long chain alkane production by isolating the carbon acyl backbone [77]. Expression of wax ester synthetases coupled with methanol or ethanol supplementation could enable *in vivo* synthesis of fatty acid methyl esters or fatty acid ethyl esters (biodiesel) [78]. Finally, expression of additional fatty acid elongases and desaturases can enable production of EPA (eicosapentaenoic acid) or DHA (docosahexaenoic acid) omega-3 fatty acids [91,203,204].

We also enabled itaconic acid production in *Y. lipolytica* PO1f and AMPD overexpression strains through heterologous CAD1 (*cis*-aconitic acid decarboxylase) expression. Itaconic acid yield can invariably be improved through effective engineering of central carbon metabolism and control of media formulation. For instance, we observed high level citric acid production in the AMPD overexpression strain when cultivated in nitrogen starvation conditions. Overexpression of the native *cis*-aconitase enzyme can reactivate flux conversion of citric acid to *cis*-aconitic acid substrate.

Nitrogen starvations results in the transportation of citric acid out of the mitochondria via the citrate-malate shunt [62]. Therefore, it may be more beneficial to express a cytoplasm-targeted version of the *cis*-aconitase enzyme. Cytoplasmic detargeting can be accomplished by removal of the mitochondrial localization signal (MLS) found in the first ~40 N-terminal amino acids of mitochondrial-localized enzymes. Exact placement of the MLS can be determined through a simple comparison of *Y. lipolytica* and *S. cerevisiae cis*-aconitase amino acid sequence. Specifically, MLS activity is dependent only on relative hydrophobicity making sequence conservation unnecessary, while *cis*-aconitase sequence is highly conserved. Thus, removal of

nonconserved N-terminal amino acids will enable cytoplasmic detargeting. It may be beneficial to redirect the entire itaconic acid production pathway to the cytoplasm by further retargeting of the citrate synthase enzyme. Alternatively, it may be beneficial to target the CAD1 enzyme to the mitochondria through addition of an N-terminal MLS. In the past, mitochondrial targeting and cytoplasmic detargeting have proven equally beneficial [205,206]. Expression of a unidirectional *cis*-aconitase enzyme could further increase *cis*-aconitic acid substrate levels [207]. Finally, optimization of media formulation and fermentation condition could grant large gains in itaconic acid production, similar to those we have seen for lipogenesis.

Collectively, my research has had two main thrusts (1) development of a novel promoter engineering technique to enable high strength gene expression for metabolic engineering applications and (2) utilization of these promoters to enable short-chain alkane, lipid and itaconic acid production in *Y. lipolytica*. In particular, I was able to employ large-scale strain and media optimization to generate, analyze, and characterize the most efficient lipid producer ever created.

Chapter 9: Materials and Methods

9.1 Common Materials and Methods

9.1.1 Common *E. coli* and yeast growth conditions

E. coli strain DH10B was used for all cloning and plasmid propagation. DH10B was grown at 37°C with constant shaking in Luria-Bertani (LB) Broth (Teknova) supplemented with 50μg/ml ampicillin for plasmid propagation. *Yarrowia lipolytica* strain PO1f (ATCC # MYA-2613), a leucine and uracil auxotroph devoid of any secreted protease activity [53] was used as the starting point for all strain construction *Y. lipolytica* studies. YSC media consisted of 20g/L glucose (Fisher Scientific), 0.79g/L CSM supplement (MP Biomedicals), and 6.7g/L Yeast Nitrogen Base w/o amino acids (Becton, Dickinson, and Company). YSC-URA, YSC-LEU, and YSC-LEU-URA media contained 0.77g/L CSM-Uracil, 0.69g/L CSM-Leucine, or .67g/L CSM-Leucine-Uracil in place of CSM, respectively. YPD media contained 10g/L yeast extract (Fisher Scientific), 20g/L peptone (Fisher Scientific) and 20g/L glucose, and was supplemented with 300μg/ml Hygromycin B (Invitrogen) for *Y. lipolytica* knockout selection.

9.1.2 Cloning and transformation procedures.

All restriction enzymes were purchased from New England Biolabs and all digestions were performed according to standard protocols. PCR reactions were set up with recommended conditions using Phusion high fidelity DNA polymerase (Finnzymes), or LongAmp *Taq* DNA polymerase (New England Biolabs). Ligation reactions were

performed overnight at room temperature using T4 DNA Ligase (Fermentas). Gel extractions were performed using the Fermentas GeneJET extraction kit purchased from Fisher ThermoScientific. Purification of small DNA fragments (<200 bp) generated during plasmid construction were performed using the MERmaid Spin Kit (Qbiogene). *E. coli* minipreps were performed using the Zyppy Plasmid Miniprep Kit (Zymo Research Corporation). *Y. lipolytica* minipreps were performed using Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research Corporation). *E. coli* maxipreps were performed using the Qiagen HiSpeed Plasmid Maxi Kit. Transformation of *E. coli* strains was performed using standard electroporator protocols (Sambrook and Russell, 2001). Large amounts of linearized DNA (>20μg), necessary for *Y. lipolytica* PO1f transformation were cleaned and precipitated using a standard phenol:chloroform extraction followed by an ethanol precipitation.

Genomic DNA (gDNA) was extracted from *Y. lipolytica* or *S. cerevisiae* using the Wizard Genomic DNA Purification kit (Promega). Transformation of *Y. lipolytica* and *S. cerevisiae* with episomal expression plasmids was performed using the Zymogen Frozen EZ Yeast Transformation Kit II (Zymo Research Corporation), with plating on appropriate selection plates. Transformation of *Y. lipolytica* PO1f with linearized cassettes was performed as described previously [4]. Briefly, PO1f was inoculated from glycerol stock directly into 10mL YPD media, grown overnight, and harvested at an OD_{600} between 9 and 15 by centrifugation at 1000 x g for 5 minutes. Cells were washed in 8.0mL TE buffer (10mM Tris, 1mM EDTA, pH = 7.5), spun down, and resuspended in 8.0mL TE buffer. 10^8 cells were dispensed into separate microcentrifuge tubes for each

transformation, spun down, and resuspended in 1.0mL LiOAc buffer (100mM LiOAc, adjusted to pH = 6.0 with 2M acetic acid). Cells were incubated with shaking at 30°C for 60 minutes, spun down, resuspended in 90µL LiOAc buffer, and placed on ice. 1-5µg of linearized DNA was added to each transformation mixture in a total volume of 10µL, followed by 25µL of 50mg/mL boiled salmon sperm DNA (Sigma-Aldrich). Cells were incubated at 30°C for 15 minutes with shaking, before adding 720µL PEG buffer (50% PEG8000, 100mM LiOAc, pH = 6.0) and 45µL 2M Dithiothreitol. Cells were incubated at 30°C with shaking at 225rpm for 60 minutes, heat shocked for 10 minutes in a 39°C water bath, spun down and resuspended in 1mL LiOAc buffer. 200µL of cells were plated on appropriate selection plates.

9.1.3 Citric acid and itaconic acid quantification

A 1-2 mL culture sample was pelleted down for 5 minutes at 3000 x g, and the supernatant was filtered using a 0.2 mm syringe filter (Corning Incorporated). Filtered supernatant was analyzed with a HPLC Ultimate 3000 (Dionex) and a Zorbax SB-Aq column (Agilent Technologies). A 2.0 µL injection volume was used in a mobile phase composed of a 99.5:0.5 ratio of 25 mM potassium phosphate buffer (pH=2.0) to acetonitrile with a flow rate of 1.25 mL/min. The column temperature was maintained at 30°C and UV–Vis absorption was measured at 210 nm. Citric acid and itaconic acid standards (Sigma-Aldrich) were used to detect and quantify organic acid production.

9.2 MATERIALS AND METHODS FOR CHAPTER 2

9.2.1 Calculation of codon adaptation index.

Codon Adaption Indices were calculated for the hrGFP, mStrawberry, EGFP, and yECitrine genes using the CAIcal server [208] and the codon usage table for *Y. lipolytica* available on the Codon Usage Database [209].

9.2.2 Plasmid construction.

9.2.2.1 Construction of endogenous promoter fluorescence cassettes (Figure 9.1a)

A table of primer sequences can be found in **Table 9.1**. All plasmids employed for gene expression in *Y. lipolytica* were centromeric, replicative vectors based off plasmid pSl16-Cen1-1(227) [98], which was initially modified to include a new multicloning site and redubbed pMCSCen1. The cyc1 terminator (cyc1t) was amplified from p416-TEF-yECitrine [30,210] and inserted into pMCSCen1 with an EcoNI/BlpI digestion to form pMCScyc1t. Endogenous promoters as defined previously [102] EXP1 (JB096/97), GPAT (JB094/95), GPD (JB088/89), TEF (JB104/105), YAT1 (JB090/91), XPR2 (JB275/276), and XPR2fus (JB277/276) (**Table 2.1**) were amplified from *Yarrowia lipolytica* PO1f genomic DNA and ligated into pMCScyc1t using XmaI/AscI for FBA and BstBI/AscI for the rest to form pMCS-Promoter serial constructs. Reporter genes including yECitine [30] (JB083/084), mStrawberry (pmstrawberry, Clonetech) (JB153/155), EGFP [211] (JB156/158), hrGFP (pIRES-hrGFP, Agilent) (JB160/161) and lacZ [212] (JB312/313) were amplified using the indicated primers pairs and inserted into

appropriate pMCS-Promoter constructs to form different pMCS-Promoter-Reporter constructs. Additionally, hrGFP and lacZ were inserted into pMCS-cyc1t to form plasmids pMCS-hrGFP and pMCS-lacZ.

9.2.2.2 Constructions of UAS1B$_1$-Leum through UAS1B$_{32}$-Leum expression cassettes (Figure 9.1b)

Primers JB162/163 amplified a 140bp minimal leucine promoter, Leum (**Table 2.1**) from plasmid pMCSCen1 that was inserted into pUC19 with SphI/HindIII to form pUC-Leum. Additionally, Leum was amplified using primers LQ19/20 and inserted into pMCS-HrGFP with BstBI/AscI to form plasmid pLeum-hrGFP.

Template for UAS1B was created by annealing primers JB177/178. Primers JB164/165 amplified a UAS1B oligo that was inserted into pUC-Leum with SalI-HF/SphI-HF to form pUC-UAS1B$_1$-Leum-No5'3'. Additionally, Primers JB174/165 amplified a UAS1B oligo that was inserted into pUC-Leum with SacI/SphI to form pUChp1dins. pUC-1dins contained only EcoRI and SacI sites 5' of the UAS1B to allow for future insertion of four tandem UAS1B sequences.

Primers JB167/168 amplified a UAS1B oligo for BamHI-HF/SalI-HF insertion into pUC-UAS1B$_1$-Leum-No5'3' to form pUC-UAS1B$_2$-Leum-No5'3', while primers JB174/168 were used to create plasmid pUC-2dins from pUC-UAS1B$_1$-Leum-No5'3'. Primers JB169/170 amplified a third UAS1B for XbaI/BamHI-HF insertion into pUC-UAS1B$_2$-Leum-No5'3' to create pUC-UAS1B$_3$-Leum-No5'3', while primers JB174/170 were used to create plasmid pUC-3dins from pUC-UAS1B$_2$-Leum-No5'3'. Primers JB171/172 amplified a fourth UAS1B for XbaI/SacI-HF insertion into pUC-UAS1B$_3$-

Leum-No5'3' to create pUC-UAS1B$_4$-Leum-No5'3', while primers JB173/172 were used to form pUC-4d5'ins from pUC-UAS1B$_3$-Leum-No5'3'. pUC-4d5'ins was edited with a SalI/SpHI mediated UAS1B replacement (from primers JB163/166) to form pUC-4dins.

Plasmid pUC-4dins was digested with EcoRI/SacI to extract a 444bp fragment containing four sequential UAS1Bs that was ligated into digests of pUC-1dins, pUC-2dins, pUC-3dins, and pUC-UAS1B$_4$-Leum-No5'3' to form plasmids pUC-UAS1B$_5$-Leum-No5'3' through pUC-UAS1B$_8$-Leum-No5'3'. An AscI restriction enzyme site was added 3' of leum in these eight plasmid with primers JB251/252 and an BstBI-PstI-KpnI sequence was added 5' to the EcoRI site of the AscI altered plasmids using primers JB249/250 to form plasmids pUC-UAS1B$_1$-Leum through pUC-UAS1B$_8$-Leum. pUC-UAS1B$_8$-Leum-No5'3' was modified to include a 5' (of EcoRI) PstI site and a 3' (of leum) KpnI site using primer pairs JB253/249 and JB254/252 respectively to create plasmid pUC-8dins. pUC-8dins was PstI/KpnI digested to extract a 902bp fragment containing eight UAS1Bs that was inserted into pUC-UAS1B$_1$-Leum through pUC-UAS1B$_8$-Leum to create plasmids pUC-UAS1B$_9$-Leum through pUC-UAS1B$_{16}$-Leum.

Plasmid pUC16dblank was created by annealing together primers JB289/290, and inserting this 74bp oligo into pUC19 with NdeI/HindIII. Plasmid pUC-UAS1B$_8$-leum was digested with KpnI and then SphI-HF, and a 895bp fragment containing "KpnI-EcoRI-UAS1B$_8$-SphI" was extracted and inserted into pUC16dblank to form pUC16d8dins. pUC-UAS1B$_8$-leum was digested with EcoRI-HF, and a 901bp UAS1B$_8$ fragment was extracted and inserted into pUC16d8ins to create pUC16dins. A 1808bp UAS1B$_{16}$ fragment was BstBI/PstI gel extracted from pUC16dins and inserted into

vectors pUC-UAS1B$_1$-Leum through pUC-UAS1B$_{16}$-Leum to form plasmids pUC-UAS1B$_{17}$-Leum through pUC-UAS1B$_{32}$-Leum.

UAS1B$_n$-Leum promoter elements were cut out using BstBI/AscI and inserted 5' of hrGFP and lacZ reporter genes in pMCS-hrGFP or pMCS-lacZ constructs in which the hrGFP and lacZ genes lacked their native ATG start site.

9.2.2.3 Construction of TEF-based promoters and expression cassettes (Figure 9.1c)

The 1004bp region upstream and including the TEF promoter was amplified from PO1f gDNA using primers LQ13 and JB105 and inserted into a pMCS-HrGFP expression cassette in which the hrGFP gene included its native ATG start site using BstBI/AscI to form plasmid pMCS-TEF(1004)-HrGFP. Promoters TEF(804) (LQ12), TEF(604) (LQ10), TEF(504) (LQ9), TEF(272) (LQ16), TEF(203) (LQ15), and TEF(136) (LQ14) were amplified from plasmid pMCS-TEF(1004)-HrGFP using the indicated primer and JB105. These seven promoters replaced TEF(1004) in pMCS-TEF(1004)-HrGFP to form the pMCS-TEF(n)-HrGFP core TEF promoter series (**Table 2.1**).

Plasmid pUC19-8d was formed by the insertion of a BstBI/HindIII digested, gel extracted UAS1B$_8$ fragment from pUC16d8dins in place of UAS1B$_{16}$-Leum in digested pUC-UAS1B$_{16}$-Leum vector. Plasmid pUC19-16d was formed by the insertion of a BstBI/HindIII gel extracted UAS1B$_{16}$ sequence from pUC16dins in place of UAS1B$_{16}$-Leum in digested pUC-UAS1B$_{16}$-Leum vector. The TEF series of promoters was reamplified using primers as follows: TEF(1004) (LQ29/17), TEF(804) (LQ28/17), TEF(604) (LQ26/17), TEF(504) (LQ25/17), TEF (LQ18/17), TEF(272) (LQ32/17), TEF(203) (LQ31/17), and TEF(136) (LQ30/17) and inserted into plasmids pUC19-16d

and pUC19-8d using a HindIII/AscI digest to form pUC-UAS1B$_{8/16}$-TEF(n) vectors. UAS1B$_{8/16}$-TEF(n) promoters were cut out with BstBI/AscI and inserted in place of TEF(136) in the pMCS-TEF(136)-hrGFP vector to form pMCS-UAS1B$_{8/16}$-TEF(n)-hrGFP vectors.

Table 9.1: Primers used in Chapter 2

Primer	**Sequence (5'-3')**
JB085	CGGGATCCCCCCCGGGAATTCGAATTGGCGC
	GCCCCTTAATTAAGGCACGTGCCTAAAAAAGGCGGACCGG
	GCTTAGCTTGTTTAAACAACTGCAGTTTT
JB088	GGTTCGAAGACGCAGTAGGATGTCCTGCA
JB089	TTGGCGCGCCGTTGATGTGTGTTTAATTCAAGAATGA
JB090	GG TTCGAAATAAGTTTGCAAAAAGATCGTATTATAGT
JB091	TTGGCGCGCCTTGTGAATTAGGGTGGTGAGA
JB092	TCCCCCCGGGTAAACAGTGTACGCAGTACTATAGA
JB093	TTGGCGCGCCGGAGAGCTGGGTTAGTTTGT
JB094	GG TTCGAACAACTTTTCTTGTCGACCTGAGAT
JB095	TTGGCGCGCCTTAGCGTGTCGTGTTTTTGTTGT
JB096	GGTTCGAAGGAGTTTGGCGCCCGTTTTT
JB097	TTGGCGCGCCTGCTGTAGATATGTCTTGTGTGTAAGGG
JB099	AAAACTGCAGTTGTTTAAACAAGCTAAGCCCGGTC
	CGCCTTTTTTAGGCACGTGCCTTAATTAAGGGGCGCGCCAAT
	TCGAATTCCCGGGCCGGATCCCG
JB102	GGCCTAAAAAAGGATCGATACCGTCGACCTCGAG
JB103	GGGCTTAGCCGAGCGTCCCAAAACCTTCTC
JB153	TTGGCGCGCCATGGTGAGCAAGGGCGAGGA
JB155	CCTTAATTAACTACTTGTACAGCTCGTCCATGCC
JB156	TTGGCGCGCCATGTCTAAAGGTGAAGAATTATTCA
JB158	CCTTAATTAATTATTTGTACAATTCATCCATACCA
JB160	TTGGCGCGCCATGGTGAGCAAGCAGATCCTGAAG
JB161	CCTTAATTAATTACACCCACTCGTGCAGGC
JB162	ACATGTGCATGCACTGATCACGGGCAAAAGTGCGTATATAT
JB163	CAACCCAAGCTTTTAGTTTCGGGTTCCATTGTGG
JB164	GACACGCGTCGACCTGAGGTGTCTCACAAGTGC
JB165	GACACATGCATGCCCGGATCGAGGTGGGCGGG
JB166	GACACATGCATGCGAGCTCCCGGATCGAGGTGGGCGGG
JB167	CGCGGATCCCTGAGGTGTCTCACAAGTGCC
JB168	GACACGCGTCGACCCGGATCGAGGTGGGCGGG
JB169	CACACACGAGCTCCTAGTCTAGACTGAGGTGTCTCACAAGTGCC

Table 9.1 continued

JB170	CGCGGATCCCCGGATCGAGGTGGGCGGG
JB171	CACACACGAGCTCCTGAGGTGTCTCACAAGTGCC
JB172	CTAGTCTAGACCGGATCGAGGTGGGCGGG
JB173	CCGGAATTCCTGAGGTGTCTCACAAGTGCC
JB174	CACACACGAGCTCCTGAGGTGTCTCACAAGTGC
JB177	CTGAGGTGTCTCACAAGTGCCGTGCAGTCCCGCCCCCAC TTGCTTCTCTTTGTGTGTAGTGTACGTACATTATCGAG ACCGTTGTTCCCGCCCACCTCGATCCGG
JB178	CCGGATCGAGGTGGGCGGGAACAACGGTCTCGATAATG TACGTACACTACACACAAAGAGAAGCAAGTGGGGGCGG GACTGCACGGCACTTGT GAGACACCTCAG
JB249	CATATGCGGTGTGAAATACCGC
JB250	GGAATTCGGGTACCAACTGCAGTTCGAAACT GGCCGTCGTTTTACAAC
JB251	AAGCTTCAGGCGCGCCATCATGGTCATAGCTGTTTCCTG
JB252	CCCACATGTTCTTTCCTGCGTTATCC
JB253	GGAATTCAACTGCAGACTGGCCGTCGTTTTACAAC
JB254	ACATACATGCATGCGGTACCCGGGCAAAAGTGCGTATAT ATACAA
JB289	GGAATTCCATATGTGTAAAACGACGGCCAGTTT CGAAGGTACCAAGGAAGCATGCCTGCAG AAGCTTGGGAAGA
JB290	TCTTCCCAAGCTTCTGCAGGCATGCTTCCTTGGTA CCTTCGAAACTGGCCGTCGTTTTACACATATGGAATTC
JB311	TTGGCGCGCCATGACCATGATTACGGATTCACTG
JB312	ACTGTTGGCGCGCCACCATGATTACGGATTCACTGGC
JB313	CCTTAATTAATTATTTTTGACACCAGACCAACTGG
LQ9	CCAATTGGTTCGAAATGTCCCAACTTGCCAAATT
LQ10	CCAATTGGTTCGAAATTGCACCCCAGCCAGACCG
LQ12	CCAATTGGTTCGAATTGCGTTTCGCTCCCACAC
LQ13	CCAATTGGTTCGAATCTGTAAAAAGTCTCTACAAG
LQ14	CCAATTGGTTCGAATTGTGGTTGGGACTTTAG
LQ15	CCAATTGGTTCGAATTTCTTTGTCTGGCCATC
LQ16	CCAATTGGTTCGAACCCACACTTGCCGTTAAG
LQ17	TTAAAGCTTAGAGACCGGGTTGGCG
LQ18	TATAAGCTTGGCGCGCCTCATTTTGAATGATTCTTATAC
LQ19	CCAATTGGTTCGAACGGGCAAAAGTGCGTAT
LQ20	TTGGCGCGCCTGAAGCTTTTAGTTTCG
LQ25	CCAATTGGAAGCTTATGTCCCAACTTGCCAAATT
LQ26	CCAATTGGAAGCTTATTGCACCCCAGCCAGACCG
LQ28	CCAATTGGAAGCTTTTGCGTTTCGCTCCCACAC
LQ29	CCAATTGGAAGCTTTCTGTAAAAAGTCTCTACAAG

Table 9.1 continued

LQ30	CCAATTGGAAGCTTTTGTGGTTGGGACTTTAG
LQ31	CCAATTGGAAGCTTTTTCTTTGTCTGGCCATC
LQ32	CCAATTGGAAGCTTCCCACACTTGCCGTTAAG

Figure 9.1: Construction of plasmids used in Chapter 2

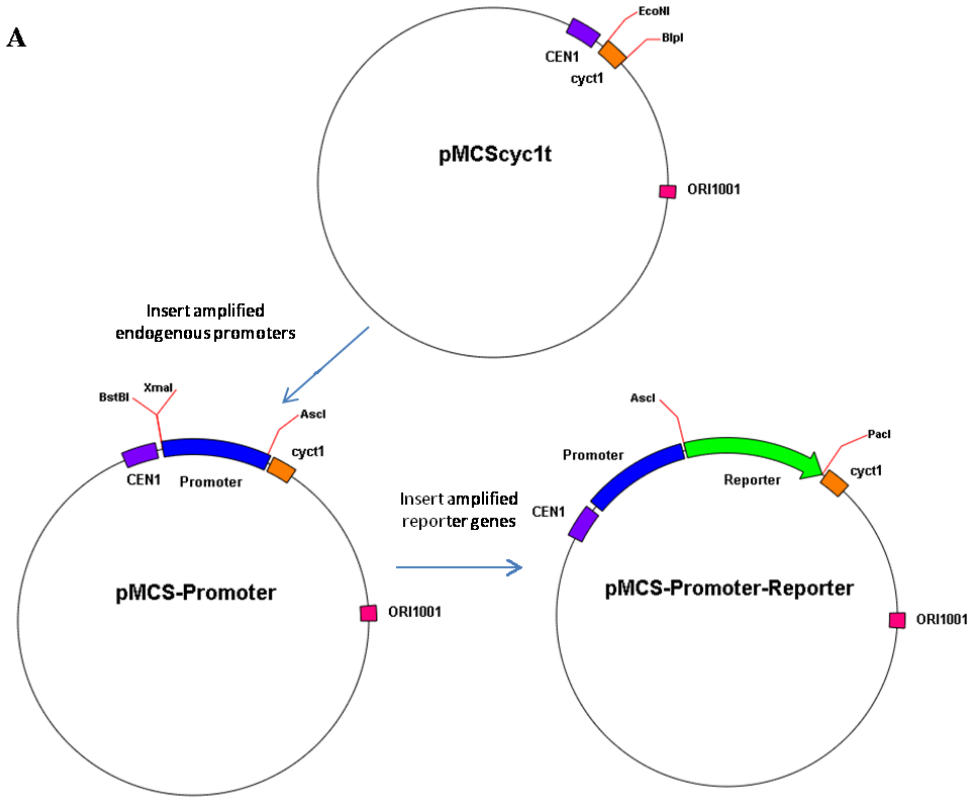
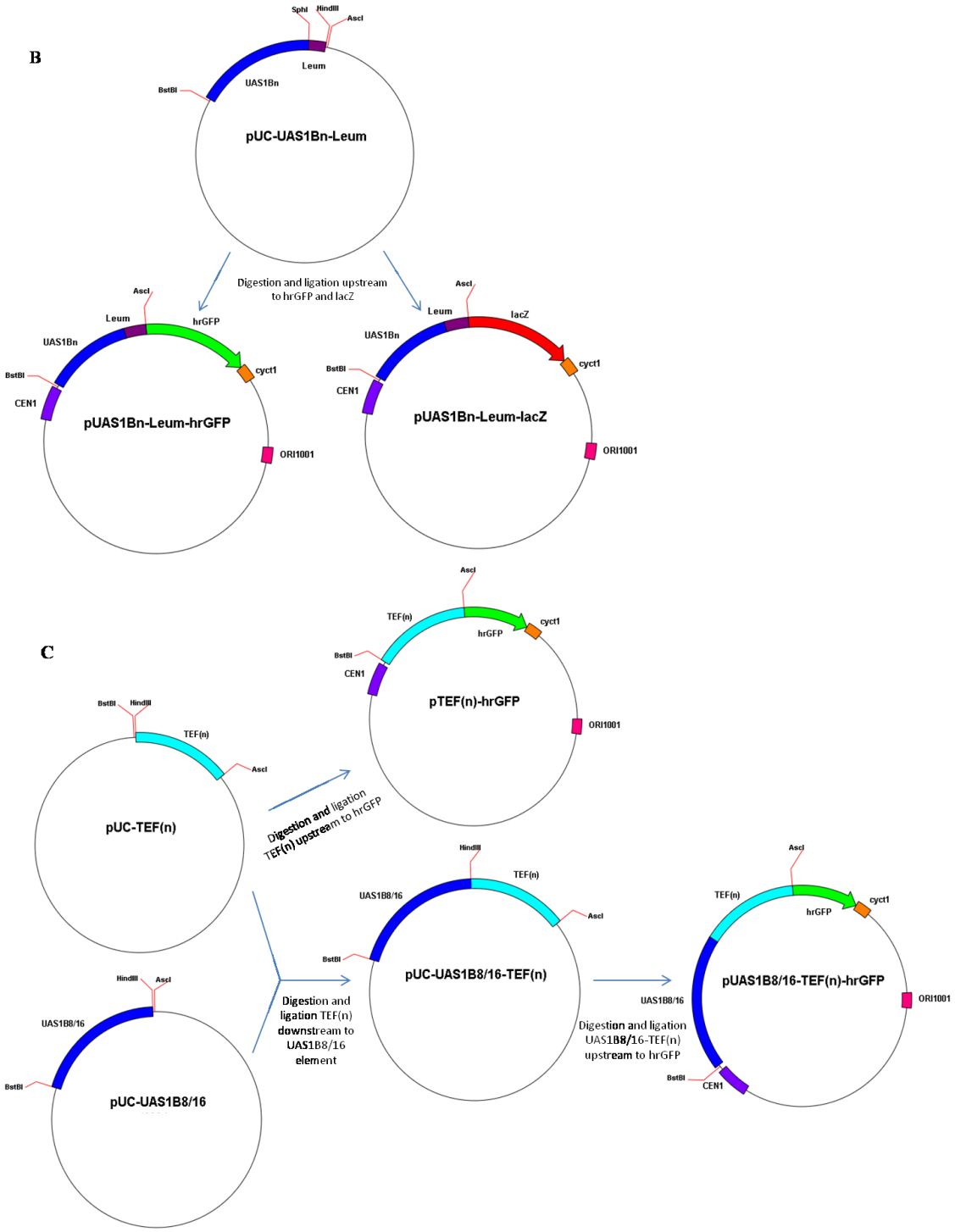


Figure 9.1 continued



(**A**) A simplified schematic picture is provided detailing the construction of endogenous promoter fluorescence cassettes (**B**) and the construction of UAS1B$_1$-Leum through UAS1B$_{32}$-Leum expression cassettes. (C) and the construction of TEF based promoters and expression cassettes.

9.2.3 Promoter characterization with flow cytometry

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological duplicate or triplicate) in YSC-LEU media for 48 hours at 30^{o}C with shaking, and then normalized to an OD_{600} of 0.03 in 2ml fresh YSC-LEU and incubated for another 48 hours at 30^{o}C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven. A time course of fluorescence values showed 48 hours to be an optimal incubation time for high expression levels from native and hybrid promoters (data not shown). To harvest, the cultures were spun down at 500g for five minutes, washed in cold water, and resuspended in 5ml ice cold water before testing with a FACSCalibur (BD Biosciences) using 488nm excitation; FL1 detector; and 10,000 cell count for hrGFP detection. Standard, optimized protocols were used for other reporter proteins tested in this study. The samples were kept on ice during the test and the data was analyzed using FlowJo software (Tree Star Inc., Ashland, OR) to compute mean fluorescence values. Day-to-day variability was mitigated by analyzing all comparable strains on the same day. An average fluorescence and standard deviation was calculated from the mean values of biological replicates.

9.2.4 Promoter characterization with β–galactosidase assay

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30^{o}C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven, and then normalized to an OD_{600} of 0.03 in 2ml fresh YSC-LEU and incubated for another 48

hours in the same conditions. The cultures were washed twice in 1ml Z buffer , resuspended in 1 ml of Z buffer, and their OD_{600} readings were recorded [106,107]. β-galactosidase assays were performed as described by Miller [106] using 10µl of chloroform-permeabilized cells, with a reaction length of 17 minutes.

9.2.5 Promoter characterization with qRT-PCR

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30°C with shaking, and normalized to an OD_{600} = 0.03 in 2ml fresh YSC-LEU media and incubated for another 48 hours at 30°C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven. The cells were pelleted and total RNA was extracted using the RiboPureTM-Yeast Kit (Ambion). 1000ng of RNA from each sample was used for a reverse transcription reaction with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). A 1.2µl sample from each reaction was used to set up a qPCR reaction (in triplicate) with FastStart SYBR Green Master (Roche) using primers

5'-TCAGCGACTTCTTCATCCAGAGCTTC-3'

and 5'-ACACGAACATCTCCTCGATCAGGTTG-3' as described in the manual with a non-template control. The reactions were run with Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems) using fast 96 well plates (Applied Biosystems). The data was analyzed with ABI 7900HTsequence detection systems (version 2.4; Applied Biosystems).

9.2.6 Plasmid stability test

Y. lipolytica PO1f strains containing plasmids pUAS1B$_{12}$-Leum-hrGFP or pUAS1B$_{16}$-Leum-hrGFP were grown for 48 hours from glycerol stock and thereafter subcultured in fresh YSC-LEU media at an OD_{600}=0.01 every 48 hours. After a total continuous culture time of 192 hours, corresponding to 36 cell doublings, yeast cells were miniprepped to extract the plasmid. Individual plasmids were isolated by transformation into *E. coli*, and sequencing and restriction enzymes digests of isolated plasmids were used to confirm the stability of the UAS1B$_{12}$-Leum and UAS1B$_{16}$-Leum promoters over this timeframe.

9.3 Materials and Methods for Chapter 3

9.3.1 Media

When testing on alternative carbon sources, glucose was replaced within the YSC media formulation by 20g/L sucrose (Acros Organics), 20g/L glycerol (Fisher Scientific), or 30g/L oleic acid (Sigma Aldrich)

9.3.2 Plasmid construction

Primer sequences can be found in **Table 9.2**. All *Y. lipolytica* plasmids were centromeric, replicative vectors derived from plasmid pSl16-Cen1-1(227) [98], which was modified to include a multi-cloning site, a hrGFP green fluorescent reporter gene (pIRES-hrGFP, Agilent), and a cyc1 terminator [210] to create plasmid pMCS-hrGFP. A

TEF promoter [45,203] was added to pMCS-hrGFP to form pMCS-TEF-hrGFP. Vectors pMCS-HrGFP, pMCS-TEF-HrGFP, pUC-UAS1B$_{16}$-TEF, pUC-UAS1B$_8$-TEF, pMCS-TEF(136)-hrGFP, pMCS-TEF-LacZ, and pUC-UAS1B$_2$-Leum have been described previously [2]. Unless stated otherwise, all PCRs utilized pMCS-TEF-hrGFP as template. All plasmids were sequenced confirmed.

9.3.2.1 Construction of UAS$_{TEF(n)}$-TEF and UAS$_{TEF(n)}$-Leum hrGFP expression cassettes (Figure 9.2a)

A UAS$_{TEF}$ fragment amplified by primers JB438/439 was SphI/HindIII digested and inserted in place of the UAS1B$_{16}$ fragment in pUC-UAS1B$_{16}$-TEF to form pUC-UAS$_{TEF(1)}$-TEF. A second UAS$_{TEF}$ fragment (primers JB440/437) was inserted into pUC-UAS$_{TEF(1)}$-TEF with BstBI/SphI to form pUC-UAS$_{TEF(2)}$-TEF. UAS$_{TEF(1)}$-TEF and UAS$_{TEF(2)}$-TEF promoters were gel extracted and inserted into pMCS-hrGFP with BstBI/AscI to create pMCS-UAS$_{TEF(1)}$-TEF-hrGFP and pMCS-UAS$_{TEF(2)}$-TEF-hrGFP, respectively. A final UAS$_{TEF}$ fragment (primers JB442/443) was inserted into pMCS-UAS$_{TEF(2)}$-TEF-hrGFP with XmaI/BstBI to form pMCS-UAS$_{TEF(3)}$-TEF-hrGFP, completing UAS$_{TEF(n)}$-TEF expression cassette construction.

An EcoRI/SphI digested UAS$_{TEF}$ fragment amplified by primers JB436/437 was inserted in place of the UAS1B$_2$ fragment in plasmid pUC-UAS1B$_2$-Leum to create plasmid pUC-UAS$_{TEF(1)}$-Leum. A BstB1/EcoRI digested UAS$_{TEF}$ fragment (primers JB440/441) was inserted into pUC-UAS$_{TEF(1)}$-Leum to create pUC-UAS$_{TEF(2)}$-Leum. UAS$_{TEF(1)}$-Leum and UAS$_{TEF(2)}$-Leum promoters were inserted into pMCS-hrGFP with BstB1/AscI to create pMCS-UAS$_{TEF(1)}$-Leum-hrGFP and pMCS-UAS$_{TEF(2)}$-Leum-hrGFP,

respectively. A final UAS_{TEF} fragment (primers JB442/443) was inserted into pMCS-$UAS_{TEF(2)}$-Leum-hrGFP with XmaI/BstBI to form pMCS-$UAS_{TEF(3)}$-Leum-hrGFP, completing $UAS_{TEF(n)}$-Leum expression cassette construction.

9.3.2.2 Construction of the UAS_{TEF}(2)-$UAS1B_8$-TEF-hrGFP expression cassette (Figure 9.2b)

A BstBI/EcoRI digested UAS_{TEF} fragment amplified with primers JB440/441 was inserted 5' of the $UAS1B_8$ region of pUC-$UAS1B_8$-TEF to form pUC-UAS_{TEF}-$UAS1B_8$-TEF. Promoter UAS_{TEF}-$UAS1B_8$-TEF was extracted with BstBI/AscI and inserted into pMCS-hrGFP, and a final 5' UAS_{TEF} fragment was added with XmaI/BstBI to create pMCS-$UAS_{TEF(2)}$-$UAS1B_8$-TEF-hrGFP.

9.3.2.3 Dissection of the TEF upstream region

Twenty two overlapping fragments spanning the UAS_{TEF} region were inserted 5' of the TEF(136) minimal promoter in plasmid pMCS-TEF(136)-hrGFP with XmaI/BstBI digests to form plasmids pMCS-UAS_{TEF}#1-TEF(136)-hrGFP through pMCS-UAS_{TEF}#22-TEF(136)-hrGFP. Primer pairs JB442/443, JB442/508, JB442/509, JB442/510, and JB442/511 amplified fragments UAS_{TEF}#1through UAS_{TEF}#5. JB503/443, JB503/507, JB503/508, JB503/510, and JB503/511 amplified fragments UAS_{TEF}#6 through UAS_{TEF}#10. JB504/443, JB504/509, and JB504/511 amplified fragments UAS_{TEF}#11 through UAS_{TEF}#13. JB505/507, JB505/508, JB505/510, and JB505/511 amplified fragments UAS_{TEF}#14 through UAS_{TEF}#17. Finally, JB506/443,

JB506/507, JB506/508, JB506/510, and JB506/511 amplified fragments UAS_{TEF}#18 through UAS_{TEF}#22.

9.3.2.4 Construction of library of UAS_{TEF}#2$_{(n)}$-TEF(136) hrGFP expression cassettes (Figure 9.2c)

An initial UAS_{TEF}#2 region amplified by primers JB544/537 was inserted into plasmid pUC-UAS1B_8-TEF(136) in place of UAS1B_8 with EcoRI/HindIII to form plasmid pUC-UAS_{TEF}#2(1)-TEF(136). Second (primers JB546/539) and third UAS_{TEF}#2 (JB548/541) elements were inserted into pUC-UAS_{TEF}#2$_{(1)}$-TEF(136) with EcoRI/BamHI and then EcoRI/XbaI digests to create pUC-UAS_{TEF}#2$_{(2)}$-TEF(136) and then pUC-UAS_{TEF}#2$_{(3)}$-TEF(136), respectively. To enable plasmid construction, a multicloning site annealed together from JB630/631 was inserted into p416-MCS-yECitrine [3] to created plasmid pTMCS. A 737bp fragment containing three tandem UAS_{TEF}#2 elements in series amplified from pUC-UAS_{TEF}#2$_{(3)}$-TEF(136) plasmid DNA with primers JB620/621 was inserted twice sequentially into pTMCS, first using SalI-HF/HindIII then SphI-HF/EcoRI-HF to create plasmids pTMCS-UAS_{TEF}#2$_{(3)}$ and pTMCS-UAS_{TEF}#2$_{(6)}$. This same 737bp fragment, digested with SphI/EcoRI, was inserted into the three pUC-UAS_{TEF}#2$_{(1,\ 2,\ and\ 3)}$-TEF(136) plasmids to create plasmids pUC-UAS_{TEF}#2$_{(4,\ 5,\ and\ 6)}$-TEF(136). The six promoters - UAS_{TEF}#2$_{(1)}$-TEF(136) through UAS_{TEF}#2$_{(6)}$-TEF(136) - were BstBI/AscI extracted and inserted into pMCS-hrGFP to create pMCS-UAS_{TEF}#2$_{(1)}$-TEF(136)-hrGFP through pMCS-UAS_{TEF}#2$_{(6)}$-TEF(136)-hrGFP. Three or six UAS_{TEF}#2 tandem repeats were BstBI/XmaI extracted from pTMCS-UAS_{TEF}#2$_{(3)}$ or pTMCS-UAS_{TEF}#2$_{(6)}$ and inserted into pMCS-UAS_{TEF}#2$_{(6)}$-hrGFP to create pMCS-UAS_{TEF}#2$_{(9)}$-

TEF(136)-hrGFP and pMCS-UAS_{TEF}#$2_{(12)}$-TEF(136)-hrGFP, completing UAS_{TEF}#$2_{(n)}$-TEF(136)-hrGFP library construction.

9.3.2.5 Construction of the lacZ expression cassettes

The β-galactosidase gene encoded by *E. coli* lacZ [213] was gel extracted from pMCS-TEF-lacZ with AscI/PacI and inserted in place of hrGFP in certain pMCS- hrGFP-based plasmid series to generate lacZ expression cassettes.

9.3.2.6 Construction of mutant UAS_{TEF}#2 elements

Utilizing plasmid pUC-UAS_{TEF}#$2_{(1)}$-TEF(136) as DNA template for the Stratagene Quikchange mutagenesis kit, three putative transcription factor binding sites (TFBSs) were removed from UAS_{TEF}#2. Primers JB682/683 deleted a "tgtgt" motif to abrogate a putative NDT80p TFBS. Primers JB680/681 deleted a "ttaag" motif to remove a putative MCM1 TFBS, and primers JB684/685 deleted a "gccatc" motif to remove a GCRp putative TFBS. Sequential mutagenesis reactions created all combinations of NDT80p, MCM1p, and GCR1p TFBS deletion mutants in the pUC-UAS_{TEF}#$2_{(1)}$-TEF(136) background. These seven mutants were BstB1/AscI extracted and inserted into pMCS-hrGFP to complete mutant UAS_{TEF}#2 expression cassette construction.

Table 9.2: Primers used in Chapter 3

Primer Name	Primer Sequence (5'-3')
JB436	CGGAATTCAGAGACCGGGTTGGCGGC
JB437	GTCGACATGCATGCTCAGTAGTCTATTTTGCGTCCGGC
JB438	GTCGACATGCATGCAGAGACCGGGTTGGCGGC
JB439	TAGGTCCCAAGCTTTCAGTAGTCTATTTTGCGTCCGGC
JB440	CACTGGTTCGAAAGAGACCGGGTTGGCGGC
JB441	CGGAATTCTCAGTAGTCTATTTTGCGTCCGGC
JB442	TCCCAACCCGGGAGAGACCGGGTTGGCGGC
JB443	CACTGGTTCGAATCAGTAGTCTATTTTGCGTCCGGC
JB503	TCCCAACCCGGGATTTGTGTCCCAAAAAACAGCC
JB504	TCCCAACCCGGGCCCCAATTGCCCCAATTGA
JB505	TCCCAACCCGGGCCCAAATTGACCCAGTAGCG
JB506	TCCCAACCCGGGGGCCCAACCCCGGCGAG
JB507	CACTGGTTCGAATCTATTTTGCGTCCGGCATG
JB508	CACTGGTTCGAAGGTTACCCGGATGGCCAGA
JB509	CACTGGTTCGAAAAAGAAACTAGTACAAAGTCTGAAC
JB510	CACTGGTTCGAATGAACAAGCGTAGATTCCAGAC
JB511	CACTGGTTCGAAACTGCAGTACCCTACGCCC
JB537	AGTCAGAAGCTTGGTTACCCGGATGGCCAGA
JB539	CTGACGGATCCGGTTACCCGGATGGCCAGA
JB541	CTGACTCTAGAGGTTACCCGGATGGCCAGA
JB544	TAGTCGAATTCCTGACGGATCCAGAGACCGGGTTGGCGGC
JB546	TAGTCGAATTCCTGACTCTAGAAGAGACCGGGTTGGCGGC
JB548	TAGTCGAATTCCTGACGTCGACAGAGACCGGGTTGGCGGC
JB620	GACTGGTGCATGCCTGACGTCGACAGAGACC
JB621	GCGACTCGAATTCCCACAAAAGCTTGGTTACC
JB630	CGCGGTACCCGGGCATGCGCCAGGAATTCGTCGACC GCACAAGCTTCGAAGGCGCGCCAA
JB631	TTGGCGCGCCTTCGAAGCTTGTGCGGTCGACGAATTC CTGGCGCATGCCCGGGTACCGCG
JB680	CCCACACTTGCCGGGCGTAGGGTACT
JB681	AGTACCCTACGCCCGGCAAGTGTGGG
JB682	AGACCGGGTTGGCGGCGTATTCCCAAAAAAC
JB683	GTTTTTTGGGAATACGCCGCCAACCCGGTCT
JB684	GTACTAGTTTCTTTGTCTGCGGGTAACCAAGCTTTTGT
JB685	ACAAAAGCTTGGTTACCCGCAGACAAAGAAACTAGTAC

Figure 9.2: Construction of plasmids used in Chapter 3

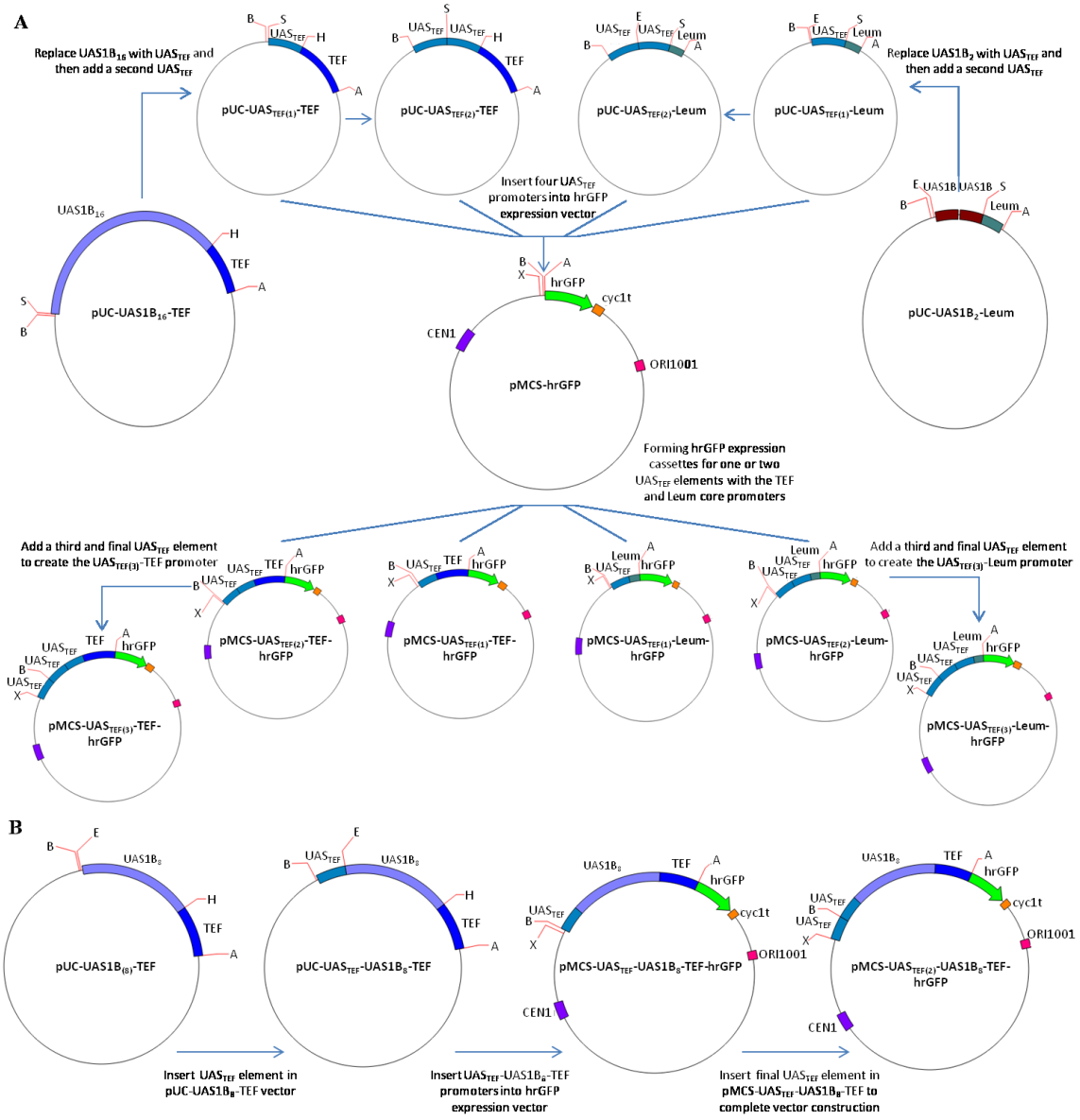
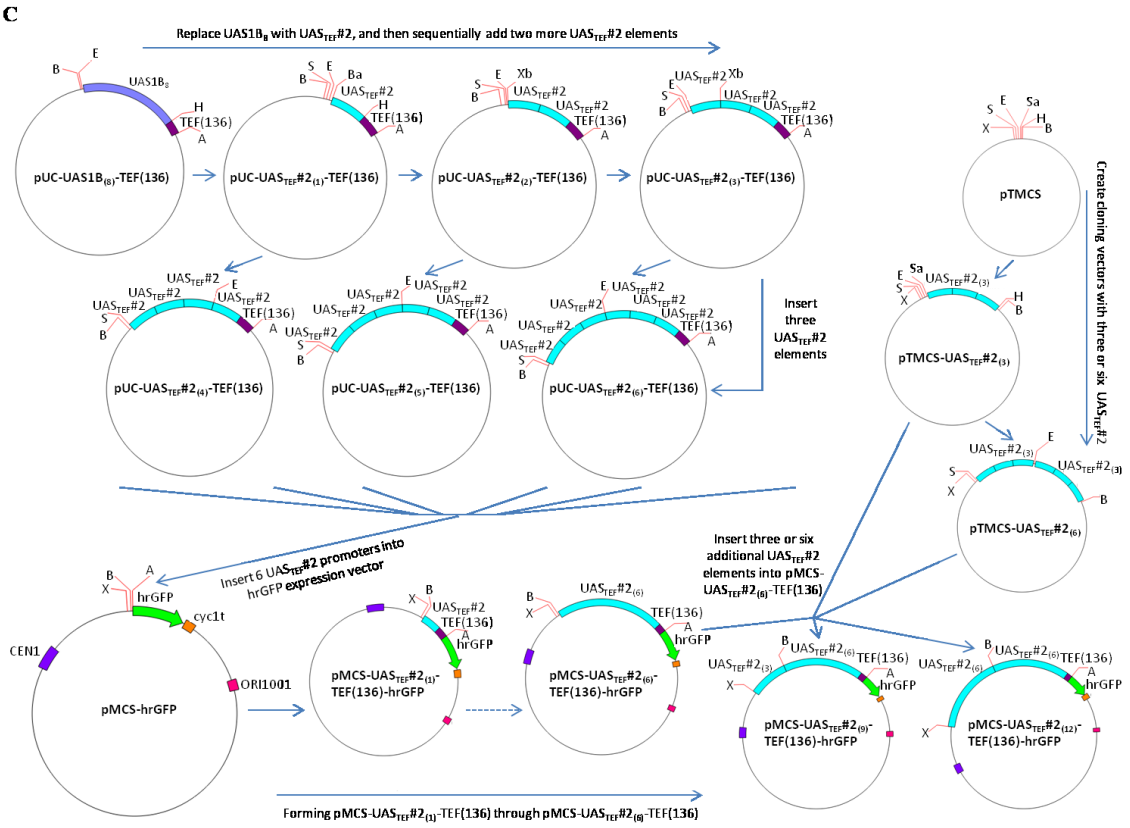


Figure 9.2 continued



(**A**) A schematic picture is provided detailing the construction of $UAS_{TEF(n)}$-TEF and $UAS_{TEF(n)}$-Leum promoter hrGFP fluorescence cassettes (**B**) and the construction of the $UAS_{TEF}(2)$-$UAS1B_8$-TEF-hrGFP expression cassette, (**C**) and the construction of a library of UAS_{TEF}#2 hrGFP expression cassettes. Restriction enzymes utilized are abbreviated as follows: A – AscI, Ba – BamHI, B – BstBI, E – EcoRI, H – HindIII, Sa – SalI, S – SphI, Xb – XbaI, X – XmaI

9.3.3 Promoter characterization by flow cytometry

The hrGFP green fluorescent protein has been validated as an ideal tool to assess promoter strength in *Y. lipolytica* at the single cell level [2], and thus was employed to assess relative promoter activity for the majority of results generated in this study. *Y. lipolytica* PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30°C with shaking, and then normalized to an OD_{600} of 0.01 in 2ml fresh YSC-LEU and incubated for another 48 hours (unless otherwise stated) at 30°C in a rotary drum (CT-7,

New Brunswick Scientific) at speed seven. A time course of fluorescence values showed 48 hours to be an optimal incubation time for high expression levels from native and hybrid promoters [2]. To harvest, the cultures were spun down at 4°C at 1000g for five minutes, washed, and resuspended in 1ml ice cold water before testing with a FACS Fortessa (BD Biosciences) using the GFP fluorochrome, a voltage of 319, and a 10,000 cell count for hrGFP detection. Samples were kept on ice during the test and the data was analyzed using FlowJo software (Tree Star Inc., Ashland, OR) to compute mean fluorescence values. Day-to-day variability was mitigated by analyzing all comparable strains on the same day. An average fluorescence and standard deviation were calculated from the mean values of biological replicates.

9.3.4 Promoter characterization with β–galactosidase assay

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30°C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven, and then normalized to an OD_{600} of 0.01 in 2ml fresh YSC-LEU and incubated for another 48 hours in the same conditions. The cultures were washed twice and resuspended in 1ml Z buffer, and their OD_{600} readings were recorded [106,107]. β-galactosidase assays were performed as described by Miller [106] using 25µl of chloroform-permeabilized cells, with a reaction time of 17 minutes.

9.3.5 Kinetic analysis of promoters

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30^{o}C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven, and then normalized to an OD_{600} of 0.01 in 2ml fresh YSC-LEU and incubated for either 24, 48, 72, or 96 hours in the same conditions. Cultures were inoculated such that all cultures (for the 24, 48, 72, and 96 hour time points) were harvested and tested via flow cytometry at the same time.

9.4 Materials and Methods for Chapter 4

9.4.1 Strains and media

The majority of yeast experiments were carried out in S. cerevisiae BY4741 (MATa; his3Δ1; leu2Δ0; met15Δ0; ura3Δ0) obtained from EUROSCARF, Frankfurt, Germany, excluding a single test assaying hybrid promoter strength in a *ura3*- mutant of Kyokai 7 Sake Yeast (NCYC479). Yeast strains were cultivated at 30^{o}C with constant agitation in either YSC-URA or YSC(gal)-URA media. YSC(gal)-URA utilized 20g/L galactose purchased from Fisher Scientific instead of glucose.

9.4.2 Plasmid construction

All plasmids containing expression cassettes were sequenced confirmed before transformation into *S. cerevisiae*. **Table 9.3** contains all primer sequences, and **Figure**

4.1 in the text of Chapter 4 illustrates the construction process for each promoter library, and a full description of plasmid construction follows. The CEN/ARS plasmid p416-TEF-yECitrine [30], containing the native P_{TEF} promoter from *S. cerevisiae*, the *CYC1* terminator, the *URA3* selectable marker, and the yellow fluorescent protein, yECitrine, was used as the basis for plasmid construction. Primers JB374 and JB375 were annealed together, digested with SacI-HF and XbaI, and inserted in place of the P_{TEF} promoter in a similarly digested p416-TEF-yECitrine plasmid to form plasmid p416-MCS-yECitrine, a plasmid containing the yECitrine reporter gene 3' of a multicloning site to enable plasmid construction.

9.4.2.1 Isolation of UAS elements and construction of plasmid expression cassettes containing core promoter regions

Four well-characterized *S. cerevisiae* endogenous promoters were isolated from plasmids described by Mumberg *et al.* [126,214], including promoters P_{GPD}, P_{TEF}, P_{CYC}, and P_{GAL}. These four promoters were amplified using primer pairs JB427/428, JB429/430, JB431/432, and JB512/513, respectively. A125 nucleotide minimal truncation of the LEU2 promoter, called P_{LEUM}, was amplified from plasmid p415-TEF using primers JB376/377 [38]. A 158 nucleotide truncation of the P_{CYC} promoter, called P_{CYC158}, was amplified using primers JB379/80. PCR amplicons of promoters P_{GPD}, P_{TEF}, P_{CYC}, P_{GAL}, P_{LEUM}, and P_{CYC158} were PmeI/XbaI digested and ligated into a similarly digested p416-MCS-yECitrine plasmid to create plasmids p416-MCS-Gpd, p416-MCS-Tef, p416-MCS-Cyc, p416-MCS-Gal, and p416-MCS-Leum, and p416-MCS- CYC158.

Three previously described upstream activating sequences and one novel UAS were utilized for expression enhancement in this study (Table 1 in the main text). The 240 basepair UAS_{CLB} [133], the 275 basepair UAS_{CIT} [134], and the 309 basepair UAS_{GAL} [29], have been described previously and were originally amplified from *S. cerevisiae* BY4741 genomic DNA using primer pairs JB417/418, JB395/396 and JB514/515, respectively. The novel UAS_{TEF}, containing 203 nucleotides between 198 and 401 basepairs upstream of the TEF1 start codon, was amplified from plasmid p416-TEF-yECitrine using primer pair JB520/521. PCR amplicons of UAS_{CLB}, UAS_{CIT}, UAS_{GAL}, and UAS_{TEF}, were HindIII/PmeI digested and inserted into a similarly digested plasmid p416-MCS-Leum to create plasmids p416-UAS_{CLB}-Leum, p416-UAS_{CIT}-Leum, p416-UAS_{GAL}-Leum, and p416-UAS_{TEF}-Leum. After sequence confirmation, these four plasmids were used for PCR template DNA for their corresponding UAS element.

9.4.2.2 Construction of LEUM-based and CYC158-based UAS-enhanced promoter libraries

A second UAS_{CIT} PCR fragment generated using primers JB415/416 was inserted into p416-UAS_{CIT}-Leum to create plasmid p416-$UAS_{CIT(2x)}$-Leum with a PacI/HindIII digest. A third UAS_{CIT} fragment generated using primers JB419/420 was inserted into p416-$UAS_{CIT(2x)}$-Leum with a PacI/AscI digest to create plasmid p416-$UAS_{CIT(3x)}$-Leum. An additional UAS_{CLB} fragment generated with primers JB425/426 was inserted into p416-$UAS_{CIT(3x)}$-Leum with a BamHI-HF/AscI digest to create plasmid p416-UAS_{CLB}-$UAS_{CIT(3x)}$-Leum. Following a similar construction process, two UAS_{CLB} fragments and one UAS_{CIT} fragment were amplified using primer pairs JB397/398, JB421/422, and

JB423/424, respectively, and then inserted one by one into plasmid p416-UAS_{CLB}-Leum to create plasmids p416-$UAS_{CLB(2x)}$-Leum, p416-$UAS_{CLB(3x)}$-Leum, and p416-UAS_{CIT}-$UAS_{CLB(3x)}$-Leum. Construction of the CYC158-based hybrid promoter library followed the construction process of the LEUM-based library exactly, and generated plasmids p416-UAS_{CIT}-CYC158, p416-$UAS_{CIT(2x)}$-CYC158, p416-$UAS_{CIT(3x)}$-CYC158, p416-UAS_{CLB}-$UAS_{CIT(3x)}$-CYC158, p416-UAS_{CLB}-CYC158, p416-$UAS_{CLB(2x)}$-CYC158, p416-$UAS_{CLB(3x)}$-CYC158, and p416-UAS_{CIT}-$UAS_{CLB(3x)}$-CYC158.

9.4.2.3 Construction of hybrid promoters with GPD1, TEF1, CYC1, or GAL1 core promoters

The following UAS constructs were gel extracted from p416-LEUM-based hybrid promoter plasmids using a BamHI-HF/PmeI Digest: UAS_{CLB}, $UAS_{CLB(2x)}$, $UAS_{CLB(3x)}$, $UAS_{CIT(3x)}$. These four enhancer constructs were inserted into BamHI-HF/PmeI digested p416-MCS-Gpd, p416-MCS-Tef, p416-MCS-Cyc, and p416-MCS-Gal vectors to form vectors p416-UAS_{CLB}-Gpd, p416-$UAS_{CLB(2x)}$-Gpd, p416-$UAS_{CLB(3x)}$-Gpd, p416-$UAS_{CIT(3x)}$-Gpd, p416-UAS_{CLB}-Tef, p416-$UAS_{CLB(2x)}$-Tef, p416-$UAS_{CLB(3x)}$-Tef, p416-$UAS_{CIT(3x)}$-Tef, p416-UAS_{CLB}-Cyc, p416-$UAS_{CLB(2x)}$-Cyc, p416-$UAS_{CLB(3x)}$-Cyc, p416-$UAS_{CIT(3x)}$-Cyc, p416-UAS_{CLB}-Gal, p416-$UAS_{CLB(2x)}$-Gal, p416-$UAS_{CLB(3x)}$-Gal, and p416-$UAS_{CIT(3x)}$-Gal.

The enhancer elements UAS_{GAL} and UAS_{TEF} were gel extracted from vectors p416-UAS_{GAL}-Leum and p416-UAS_{TEF}-Leum with a PacI/PmeI digest and inserted into similarly digested p416-MCS-Gpd, p416-MCS-Tef, p416-MCS-Cyc, and p416-MCS-Gal

vectors to create plasmids p416-UAS_{GAL}-Gpd, p416-UAS_{GAL}-Tef, p416-UAS_{GAL}-Cyc, and p416-UAS_{GAL}-Gal.

A HindIII/PmeI UAS_{GAL} digested fragment was gel extracted from vector p416-UAS_{GAL}-Gpd and inserted in place of the UAS_{CLB} fragment proximal to the LEUM core promoter in plasmid p416-$UAS_{CLB(3x)}$-Leum to form plasmid p416-$UAS_{CLB(2x)}$-UAS_{GAL}-Leum. A UAS_{GAL} element was PCR amplified using primers JB516 and JB517, digested with PacI/HindIII and inserted in place of the middle UAS_{CLB} element in plasmid p416-$UAS_{CLB(3x)}$-Leum to form plasmid p416-UAS_{CLB}-UAS_{GAL}-UAS_{CLB}-Leum. This same amplicon was inserted into a PacI/HindIII digest p416-MCS-Leum to form plasmid p416-$UAS_{GAL\text{-}A}$-Leum that had unused HindIII/Pme sites between the UAS_{GAL} sequence and the Leum promoter. A UAS_{GAL} element was PCR amplified using primers JB518 and JB519, digested with AscI/PacI and inserted in place of the UAS_{CLB} element furthest from the LEUM core promoter in plasmid p416-$UAS_{CLB(3x)}$-Leum to form plasmid p416-UAS_{GAL}-$UAS_{CLB(2x)}$-Leum.

A PacI/HindIII UAS_{CIT} digested fragment was gel extracted from vector p416-$UAS_{CIT(2x)}$-Leum and inserted into a similarly digested p416-UAS_{CLB}-Gpd vector to create plasmid p416-UAS_{CIT}-UAS_{CLB}-Gpd. A UAS_{TEF} fragment was PCR amplified using primers JB524 and JB525, digested with AscI/PacI, and inserted into a similarly digested p416-UAS_{CIT}-UAS_{CLB}-Gpd vector to create plasmid p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-Gpd. A UAS_{TEF}-UAS_{CIT}-UAS_{CLB} fragment was AscI/PmeI digested and gel extracted from plasmid p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-Gpd and inserted into an

AscI/PmeI digested p416-Leum plasmid to create plasmid p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-Leum.

Primers JB480C and JB481C were annealed together, BssHII digested, and inserted into a similarly digested p415TEF (insert mumberg reference) vector to form plasmid p415-MCS. Primers JB618 and JB619 were annealed, BamHI-HF/SacI-HF digested, and inserted into digested p415-MCS vector to form plasmid p415-6XMCS. AscI/PmeI digested fragments of UAS_{TEF}-UAS_{CIT}-UAS_{CLB} and $UAS_{CLB(3x)}$ were extracted from their respective p416-Leum based plasmids and inserted into digest p415-6XMCS to form plasmids p415-6XMCS-UAS_{TEF}-UAS_{CIT}-UAS_{CLB} and p415-6XMCS-$UAS_{CLB(3x)}$, respectively. BamHI-HF/SacI-HF fragments of UAS_{TEF}-UAS_{CIT}-UAS_{CLB} and $UAS_{CLB(3x)}$ were extracted from their respective p415-6XMCS based plasmids. The UAS_{TEF}-UAS_{CIT}-UAS_{CLB} fragment was inserted into plasmids p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-Gpd and p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-Leum to create plasmids p416-$(UAS_{TEF}$-UAS_{CIT}-$UAS_{CLB})_{2x}$-Gpd and p416-$(UAS_{TEF}$-UAS_{CIT}-$UAS_{CLB})_{2x}$-Leum, respectively, while the $UAS_{CLB(3x)}$ fragment was inserted into plasmids p416-$UAS_{CLB(3x)}$-Gpd to create plasmid p416-$UAS_{CLB(6x)}$-Gpd.

9.4.2.4 Construction of a library for tunable galactose-induced gene expression

A dissection of the key 54bp element within UAS_{GAL} was performed by inserting DNA products created using the following primer pairs behind the Leum promoter in plasmid p416-MCS-Leum with a HindIII/PmeI digest. G4BS1-4 was PCR amplified with primers JB585/586 from p416-MCS-GAL plasmid DNA and inserted as described to form plasmid p416-G4BS1-4-Leum. Primer pairs JB587/588, JB589/590, JB591/592,

JB593/594, JB595/596, JB597/598, JB599/600, and JB601/602 were annealed together and similarly inserted to form plasmids p416-G4BS13-Leum, p416-G4BS12-Leum, p416-G4BS34-Leum, p416-G4BS24-Leum, p416-G4BS1-Leum, p416-G4BS2-Leum, p416-G4BS3-Leum, and p416-G4BS4-Leum. Primer pair JB606/607 was annealed and inserted into plasmid p416-UAS_{GAL-A}-Leum with a HindIII/PmeI digest to form plasmid p416-UAS_{GAL}-A9-Leum. Fragments of the 5' region of P_{CYC1} generated with primer pairs JB603/604 and JB603/605 were inserted into the same backbone to create plasmids p416-UAS_{GAL}-CU1-Leum and p416-UAS_{GAL}-CU2-Leum, respectively.

Table 9.3: Primers used in Chapter 4

Primer	Sequence
JB374	gacacatgagctccgggatccttggcgcgccccttaattaacggaagcttcctgtttaaactagttctagactag
JB375	ctagtctagaactagtttaaacaggaagcttccgttaattaaggggcgcgccaaggatcccggagctcatgtgtc
JB376	agctttgtttaaaccaatattatttaaggacctattgttt
JB377	ctagtctagatagaatggtatatccttgaaatatatat
JB379	ctagtctagattagtgtgtgtatttgtgtttgcg
JB380	agctttgtttaaactgcgacgacacatgatcatat
JB395	gcagctcaagctttagagattactacatattccaacaag
JB396	agctttgtttaaacagggattgcgatcctcca
JB397	ccttaattaaagtggaattattagaatgaccactac
JB398	gcagctcaagcttggacaggcaccgaagttc
JB415	ccttaattaatagagattactacatattccaacaag
JB416	gcagctcaagcttagggattgcgatcctcca
JB417	gcagctcaagcttagtggaattattagaatgaccactac
JB418	agctttgtttaaacggacaggcaccgaagttc
JB419	ttggcgcgcctagagattactacatattccaacaag
JB420	ccttaattaaagggattgcgatcctcca
JB421	ttggcgcgccagtggaattattagaatgaccactac
JB422	ccttaattaaggacaggcaccgaagttc
JB423	cgggatcctagagattactacatattccaacaag
JB424	ttggcgcgccagggattgcgatcctcca
JB425	cgggatccagtggaattattagaatgaccactac
JB426	ttggcgcgccggacaggcaccgaagttc
JB427	agctttgtttaaacagtttatcattatcaatactcgccattt
JB428	ctagtctagaatccgtcgaaactaagtt

Table 9.3 continued

JB429	agctttgtttaaacatagcttcaaaatgtttctactcctt
JB430	ctagtctagaaaacttagattagattgctatgc
JB431	agctttgtttaaacatttggcgagcgttggttg
JB432	ctagtctagattagtgtgtgtatttgtgtttgc
JB480C	tgactgactgactgcgcgcgtcgacgtcagtcgagctcaatctagaggggatccttcccgggaacggccgagcgcgcagtgactgagttg
JB481C	caactcagtcactgcgcgctcggccgttcccgggaaggatcccctctagattgagctcgactgacgtcgacgcgcgcagtcagtcagtca
JB512	agctttgtttaaactagtacggattagaagccgc
JB513	ctctagtctagaatccggggtttttttct
JB514	gcagctcaagctttagtacggattagaagccgc
JB515	agctttgtttaaacctgttaatagatcaaaaatcatcgct
JB520	gcagctcaagcttatagcttcaaaatgtttctactcct
JB521	agctttgtttaaacgccttttttcgacgaagaaaaag
JB522	ccttaattaaatagcttcaaaatgtttctactcct
JB523	gcagctcaagcttgccttttttcgacgaagaaaaag
JB524	ttggcgcgccatagcttcaaaatgtttctactcct
JB525	ccttaattaagccttttttcgacgaagaaaaag
JB585	gctggtcaagcttcggattagaagccgccgag
JB586	agctttgtttaaaccggaggagagtcttccgtcg
JB587	ctgtcaagcttcggattagaagccgccgagcgggcgacagccctccggtttaaacaaagc
JB588	gctttgtttaaaccggagggctgtcgcccgctcggcggcttctaatccgaagcttgacag
JB589	gctggtcaagcttcggattagaagccgccgagggcgacagccgtttaaacacagtc
JB590	gactgtgtttaaacggctgtcgccctcggcggcttctaatccgaagcttgaccagc
JB591	gctgtcaagcttcgggcgacagccctccgacggaagactctcctccggtttaaacacctc
JB592	gaggtgtttaaaccggaggagagtcttccgtcggagggctgtcgcccgaagcttgacagc
JB593	gcagctgaagcttggcgacagccacggaagactctcctccggtttaaacacagtc
JB594	gactgtgtttaaaccggaggagagtcttccgtggctgtcgccaagcttcagctgc
JB595	gctggtcaagcttcggattagaagccgccggtttaaacacagtc
JB596	gactgtgtttaaaccggcggcttctaatccgaagcttgaccagc
JB597	gcagctgaagcttggcgacagccgtttaaacacagtc
JB598	gactgtgtttaaacggctgtcgccaagcttcagctgc
JB599	gctggtcaagcttcgggcgacagccctccggtttaaacacagtc
JB600	gactgtgtttaaaccggagggctgtcgcccgaagcttgaccagc
JB601	gctggtcaagcttcggaagactctcctccggtttaaacacagtc
JB602	gactgtgtttaaaccggaggagagtcttccgaagcttgaccagc
JB603	ccaaccaacgctcgccaaataagcttgaccagc
JB604	gactgtgtttaaactgcctgtatgtgtcagcactaaa
JB605	gactgtgtttaaaccatacagagcacatgcatgcca
JB606	gctggtcaagcttcatgtacttacccaattaggaaatttacatggtttaaacacagtc
JB607	gactgtgtttaaaccatgtaaatttcctaattgggtaagtacatgaagcttgaccagc
JB618	gtcagtgagagctcaggcgcgccgctagacatagtttaaacgggatcccatgcatg
JB619	catgcatgggatcccgtttaaactatgtctagcggcgcgcctgagctctcactgac

9.4.3 Promoter characterization with flow cytometry

S. cerevisiae strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-URA media for 48 hours at 30^{o}C with shaking, and then normalized to an OD_{600} of 0.01 in 2ml fresh YSC-URA or 2mls fresh YSC(gal)-URA. Strains reinoculated in YSC-URA were incubated for another 15 hours at 30^{o}C with shaking, while strains reinoculated in YSC(gal)-URA were given 24 hours before harvesting. To harvest, cultures were spun down at 4^{o}C at 1000g for five minutes, resuspended in 2ml ice cold water, and kept on ice until testing with a FACS Fortessa (BD Biosciences) using the YFP fluorochrome, a voltage of 308, and a 10,000 cell count for yECitrine detection. Data was analyzed using FlowJo software (Tree Star Inc., Ashland, OR) to compute mean fluorescence values. Day-to-day variability was mitigated by analyzing all comparable strains on the same day. An average fluorescence and standard deviation was calculated from the mean values of biological replicates.

9.4.4 Promoter characterization with β–galactosidase assay

S. cerevisiae BY4741 strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-URA media for 48 hours at 30^{o}C with shaking, and then normalized to an OD_{600} of 0.01 in 2ml fresh YSC-URA or 2mls fresh YSC(gal)-URA. Strains reinoculated in YSC-URA were incubated for another 15 hours at 30^{o}C with shaking, while strains reinoculated in YSC(gal)-URA were grown 24 hours before harvesting. Cultures were cooled in an ice bath for 30 minutes, washed twice in 1ml Z buffer, resuspended in 1ml Z buffer, and

measured for OD_{600} absorbance [106]. β-galactosidase assays were performed as described by Miller [106] using 50µl of chloroform-permeabilized cells, and a reaction length of 12.5 minutes. Average activity and standard deviation were calculated from biological replicates.

9.4.5 Promoter characterization with qRT-PCR

S. cerevisiae BY4741 strains, transformed with different plasmids, were inoculated directly from glycerol stock in YSC-URA media for 48 hours at 30°C with shaking, and then normalized to an OD_{600} of 0.01 in 2ml fresh YSC-URA or 2mls fresh YSC(gal)-URA. Strains reinoculated in YSC-URA were incubated for another 15 hours at 30°C with shaking, while strains reinoculated in YSC(gal)-URA were grown 24 hours before harvesting. The cells were pelleted and total RNA was extracted using the RiboPureTM-Yeast Kit (Ambion). 500ng of RNA from each sample was used as template for the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). A 3µl sample from each cDNA reaction was used to set up a qPCR reaction (in triplicate) with FastStart SYBR Green Master (Roche) using primers

5'- TTCTGTCTCCGGTGAAGGTGAA -3'

and 5'- TAAGGTTGGCCATGGAACTGGCAA-3' as described in the manual with a non-template control. The reactions were run with the ViiA™ 7 Real-Time PCR System using fast 96 well plates (Applied Biosystems). The data was analyzed with ViiA™ 7 Software (v. 1.0).

9.5 MATERIALS AND METHODS FOR CHAPTER 5

9.5.1 Cultivation and media

Y. lipolytica was cultivated at 30°C with constant agitation at 180rpm for all experiments. Fatty acid accumulation was promoted by cultivation in media with a high carbon to nitrogen ratio, called High C:N media, containing 160g/L glucose, CSM (-leucine, -uracil, or complete) supplement, 0.17g/L Yeast Nitrogen Base w/o amino acid and w/o $(NH_4)_2SO_4$ (Becton, Dickinson, and Company), and 0.2 g/L $(NH_4)_2SO_4$ (Fisher Scientific). Solid media for *E. coli* and *Yarrowia lipolytica* was prepared by adding 20g/L agar (Teknova) to liquid media formulations.

9.5.2 Plasmid construction

9.5.2.1 Construction of episomal Gmlox1 expression cassettes

A table of primer sequences can be found in **Table 9.4**. Plasmids pMCSCen1, pMCS-TEF-hrGFP, pMCS-EXP-hrGFP, pMCS-UAS1B$_{16}$-TEF-hrGFP, and pMCS-UAS1B$_{16}$-Leum-hrGFP have been described previously [2]. A codon optimized soybean lipoxygenase 1 gene [156], Gmlox1, was synthesized by Blue Heron and then amplified with primers JB408/JB409 and inserted into pMCSCen1, pMCS-TEF-hrGFP, and pMCS-EXP-hrGFP in place of hrGFP with an AscI/PacI digest to form pMCS-Gmlox1, pMCS-TEF-Gmlox1, and pMCS-EXP-Gmlox1, respectively. Gmlox1 was similarly inserted into pMCS-UAS1B$_{16}$-Leum-hrGFP after amplification by primers JB410/JB409 to create plasmid pMCS-UAS1B$_{16}$-Leum-Gmlox1.

9.5.2.2 Construction of integrative expression cassettes

Primers JB316/317 amplified a 149 basepair minimal terminator region from the XPR2 gene (xpr2tmin) from *Y. lipolytica* PO1f gDNA and flanked it with a 5' BamHI restriction enzyme site and 3' PmeI, RsrII, AvrII, SgfI, SalI sites. Xpr2tmin was digested with BamHI-HF/SalI-HF and inserted into pUC19 to form plasmid, pUC-xpr2tmin. Primers JB320/321 amplified a 578 basepair region of *Y. lipolytica* 26S rDNA [76] from the PO1f genome, flanked by 5' SacI and NotI restriction enzyme sites and 3' KpnI and XmaI sites[76]. This DNA fragment was inserted into pUC-xpr2tmin with a SacI-HF/XmaI digest to form plasmid pUCrDNA5'. Primers JB324/325 amplified a 822 basepair region of *Y. lipolytica* 26S rDNA [76] from the PO1f genome flanked by 5' SbfI and 3' NotI and HindIII sites. This fragment was inserted into pUCrDNA5' with a SbfI/HindIII digest to form plasmid pUCrDNA. Primers JB326/328 amplified the Ura3d1 allele [75] from *Y. lipolytica* strain IFO1659 [215] flanked by 5' KpnI and 3' BstBI, AscI, PacI and BamHI sites. The Ura3d1 fragment was inserted into pUCrDNA with a KpnI/BamHI-HF digest to form plasmid pUC-S1-nolox. To allow for marker retrieval, the Ura3d1 gene was replaced by an identical Ura3d1 gene flanked by lox sites using primers JB572/574 and a SacII/BstBI digest, forming plasmid pUC-S1. The strong UAS1B_{16}-Leum promoter [2] was then inserted into plasmid pUC-S1 with a BstBI/AscI digest to create plasmid pUC-S1-UAS1B_{16}-Leum. The UAS1B_{16}-TEF promoter [2] was similarly inserted into plasmid pUC-S1 to create plasmid pUC-S1-UAS1B_{16}-TEF. The Gmlox1 gene was gel extracted from plasmid pMCS-UAS1B_{16}-Leum-Gmlox1 and inserted into pUC-S1-UAS1B_{16}-Leum with AscI/PacI to form plasmid pUC-S1-UAS1B_{16}-Leum-Gmlox1. Primers JB608/609

amplified a Blue Heron codon optimized soybean hydroperoxide lyase gene [165-167], Gmhpl1, for insertion into pUC-S1-UAS1B$_{16}$-TEF with AscI/PacI to form plasmid pUC-S1-UAS1B$_{16}$-TEF-Gmhpl1.

9.5.2.3 Construction of knockout cassettes

Primers JB124/125 amplified a cyc1 terminator from plasmid p416-TEF-yECitrine [30,210] for insertion into plasmid pUC19 with XbaI/PstI to create plasmid pUCcyc1t. Primers JB126a/127 amplified a hygromycin resistance gene (Hph) from plasmid pAG34 [216] for insertion into plasmid pUCcyc1t to create plasmid pUCHphcyc1t. Primers JB138/146 amplified a EXP promoter regions from plasmid pMCS-EXP1-hrGFP [2] for insertion into plasmid pUCHphcyc1t to create plasmid pUCEXPHphcyc1t.

Primers JB199/200 were annealed together and inserted into pUC19 with an EcoRI-HF/SacI-HF digest to create plasmid pUC5'lox. Primers JB201a/202a were annealed together and inserted into pUC5'lox with a XbaI/HindIII digest to create plasmid pUC5'3'lox. Primers JB203a/204a amplified an EXP1-Hph-cyc1t hygromycin resistance cassette from plasmid pUCEXPHphcyc1t for insertion into plasmid pUC5'3'lox to create plasmid pKO. Primers JB347/348 amplified 3000 basepairs of genomic DNA 5' of the MFE1 gene (YALI0E15378g) using PO1f gDNA as template, while primers JB349/350 amplified genomic DNA 3' of the MFE1 gene. These fragments were inserted sequentially into plasmid pKO with an AscI/NotI-HF and a FseI/PmeI digest, respectively, to form plasmid pKOMFE1. Similarly, primers pairs JB564/565 and JB566/567 amplified 5' and 3' genomic DNA relative to the PEX10 gene (YALI0C01023g) to enable the construction of pKOPEX10.

9.5.2.4 Construction of cre-recombinase expressing plasmid

A cyc1 terminator regions was amplified from plasmid pUCcyc1t by primers JB151/152 for insertion into plasmid pMCSCen1 [2] to create plasmid pMCScyc1t. Then, the UAS1B$_{16}$-TEF promoter was gel extracted from plasmid pMCS-UAS1B$_{16}$-TEF-hrGFP [2] with a XmaI/AscI digested and inserted into plasmid pMCScyc1t to form plasmid pMCS-UAS1B$_{16}$-TEF. Primers JB149/150 amplified the Cre Recombinase gene from plasmid pSH47 [217] for insertion into pMCS-UAS1B$_{16}$-TEF with AscI/PacI to create plasmid pMCS-UAS1B$_{16}$-TEF-Cre.

Table 9.4: Primers used in Chapter 5:

Primer	Sequence (5' → 3')
JB124	gctctagaatcgataccgtcgacctcgag
JB125	aaaactgcagttgtttaaacaagctaagcccgagcgtcccaaaaccttctc
JB126a	cgggatccttggcgcgccccttaattaaatgggtaaaaagcctgaactcacc
JB127	gctctagattattcctttgccctcggacgag
JB138	ttggcgcgccggagtttggcgcccgttttt
JB146	ccttaattaatgctgtagatatgtcttgtgtgtaagg
JB149	ttggcgcgccatgtccaatttactgaccgtacac
JB150	ccttaattaactaatcgccatcttccagcag
JB151	ccttaattaatcatgtaattagttatgtcacgctt
JB152	ccggcgcgccgtttaaaccgagcgtcccaaaaccttctc
JB201a	tgctctagagtataacttcgtataatgtatgctatacgaagttattgggccggccagctttgtttaaacaagcttgggactag
JB202a	ctagtcccaagcttgtttaaacaaagctggccggcccaataacttcgtatagcatacattatacgaagttatactctagagca
JB199	cggaattcggcgcgccaagcggccgcaacctgcagggtataacttcgtataatgtatgctatacgaagttattggagctcgagtat
JB200	atactcgagctccaataacttcgtatagcatacattatacgaagttataccctgcaggttgcggccgcttggcgcgccgaattccg

Table 9.4 continued

JB203a	cggggtaccggagtttggcgcccgttttt
JB204a	cgcggatcccgagcgtcccaaaaccttctc
JB316	cgcggatccgcaattaacagatagtttgccggtg
JB317	accacgcgtcgacttgcgatcgcttcctaggttcggaccgtttgtttaaaccgaacaaagacgggattttgcc
JB320	actactacgagctcagcggccgcagatcttggtggtagtagcaaa
JB321	cccaacccgggaggtaccccgcgggtccggctgcca
JB324	gtcgctcgcctgcaggcagacactgcgtcgctccg
JB325	cccaagcttttgcggccgctgcttcggtatgataggaagagcc
JB326	ggggtaccctgcagactaaatttatttcagtctcc
JB328	cgcggatccttaattaattggttggcgcgccttggttttcgaactaacagttaatcttctggtaagcctc
JB347	ttggcgcgcctaagaattgccaaaacgtatctctt
JB348	aaggaaaaaagcggccgctgtgtgtgtgtgttgaataaataga
JB349	agtcagtcggccggccgctattatctgaccaagtgatacga
JB350	agctttgtttaaacgctgaggaggaaaaatccgta
JB408	ttggcgcgccatgttttccgcaggccacaa
JB409	ccttaattaattagatggagattgaattagggataccg
JB410	ttggcgcgccttttccgcaggccacaaaat
JB572	ccaccgcggataacttcgtataatgtatgctatacgaagttatctgcagactaaatttatttcagtctcc
JB574	cggttcgaaataacttcgtatagcatacattatacgaagttatctaacagttaatcttctggtaagcct
JB564	gataggcgcgccaggtcgtgggaaagacagatt
JB565	agcaaggttctgcggccgcgacgaagattccgaggggata
JB566	gagctagcaggccggccgatggaaggactagtcagcga
JB567	cagctcagtttaaacaagaagttctcacacagcagc
JB608	ttggcgcgccatgtctctccccccaccctc
JB609	ccttaattaattactttgttttctgaagagcagtg

9.5.3 Strain construction

Plasmids pUC-S1-UAS1B$_{16}$-Leum-Gmlox1 and pUC-S1-UAS1B$_{16}$-TEF-Gmhpl1 were digested with Not1-HF to create linear expression cassettes, devoid of *E. coli* DNA and surrounded by *Y. lipolytica* rDNA for targeted integrations. *Y. lipolytica* PO1f was transformed with these linearized expression cassettes individually and plated on YSC-URA dropout media to isolate strains PO1f-Gmlox1 and PO1f-Gmhpl1. The Ura3d1

marker was subsequently removed by transformation with plasmid pMCS-UAS1B$_{16}$-TEF-Cre followed by replica plating for selection of uracil auxotrophs, and then growth on permissive media to lose plasmid pMCS-UAS1B$_{16}$-TEF-Cre. Strain PO1f-Gmlox1 was further transformed with the S1-UAS1B$_{16}$-Leum-Gmlox1 cassette, followed by loss of the Ura3d1 marker as described above to yield strain PO1f-Gmlox1-Gmhpl1.

Knockout cassettes that targeted the MFE1 gene and PEX10 gene were amplified from plasmids pKOMFE1 and pKOPEX10 using primers JB347/350 and JB546/567, respectively. *Y. lipolytica* PO1f was transformed with these knockout cassettes, with selection on YPD-hygromycin plates, followed by loss of the hygromycin resistance cassette marker to yield strains PO1f ΔMFE1 and PO1f ΔPEX10. All strains were confirmed through gDNA extraction and five distinct PCR confirmations.

9.5.4 Cultivation conditions and quantification of pentane yield

Y. lipolytica strains were inoculated directly from glycerol stock in biological triplicate in appropriate media (YSC-LEU for strains containing plasmid, and YSC for strains not containing plasmid) for 48 hours at 30^{o}C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven. These strains were reinoculated at an OD$_{600}$ of 0.01 in 100mL of fresh media (composition varies and is specified in text) and incubated for up to 192 hours in 1000mL Kimax* GL 45 glass media/storage bottles (Fisher Scientific), sealed with open top GL-45 Caps with 34mm apertures (Chemglass Life Sciences) fitted with PTFE seals (Chemglass Life Sciences). The cultures were grown inside 1L glass bottles sealed with PTFE septa to prevent loss of pentane product and to allow for headspace sampling of the product.

Pentane production was quantified by injecting 100μL headspace from the 100mL cultures into a GC-FID (Agilent Technologies 6890 Network GC System) equipped with an Agilent HP-5 column (5% phenyl-95% methylsiloxane - product number 19091J-413) and measuring the peak area at 3.34 minutes. Briefly, the following settings were used: Detector Temp = 250^{o}C, He Flow = 45.0 mL/min, Oven Initial Temp = 50^{o}C, Oven Final Temp = 140^{o}C, Ramp = 20^{o}C/min. Standards were used to detect and quantify *n*-pentane (Fisher scientific) and hexanal (Acros Organics) production.

9.5.5 Lipid quantification and fatty acid profile analysis

Y. lipolytica strains were cultivated as described above for eight days in 1L glass bottles. Lipids from ~20-30 OD equivalents were extracted following the procedure described by [218] and modified for yeast [219]. Dried lipids were transesterified with N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (Sigma-Aldrich) following the procedure of [220], and 2 μL samples were injected into a GC-FID (Agilent Technologies 6890 Network GC System) equipped with an Agilent HP-5 column (5% phenyl-95% methylsiloxane - product number 19091J-413) to analyze fatty acid fractions. Briefly, the following settings were used: Detector Temp = 300^{o}C, He Flow = 1.0 mL/min, Oven Temp = 80°C for 2 min, increased at 30°C/min to 200 °C, increased at 2°C/min to 229 °C, increased at 1°C/min to 232°C, increased at 50°C/min to 325°C. Fatty acid standards for C16:0 palmitic acid, C16:1(n-7) palmitoleic acid, C18:0 stearic acid, C18:1 (n-9) oleic acid, and C18:2 (n-6) linoleic acid were purchased from Sigma-Aldrich, transesterified, and analyzed by GC to identify fatty acid peaks.

9.6 MATERIALS AND METHODS FOR CHAPTER 6

9.6.1 Strains and media

Y*arrowia lipolytica* strain PO1f (ATCC # MYA-2613), a leucine and uracil auxotroph devoid of any secreted protease activity (Madzak et al., 2000), was used as the base strain for all studies. **Table 6.2** contains a complete list of PO1f derivatives produced and used in Chapter 6. *Y. lipolytica* was cultivated at 30^{o}C with constant agitation. 2mL cultures of *Y. lipolytica* used in large-scale screens were grown in a rotary drum (CT-7, New Brunswick Scientific) at speed seven, and larger culture volumes were shaken in flasks at 225rpm or fermented in a bioreactor.

Lipid accumulation response towards media formulation was investigated by cultivation in varying concentrations of glucose and nitrogen. These media formulations contained .79g/L CSM, 1.7g/L Yeast Nitrogen Base w/o amino acid and w/o $(NH_4)_2SO_4$ (Becton, Dickinson, and Company), between 20g/L and 160g/L glucose, and between 0.2g/L and 5g/L ammonium sulfate - $(NH_4)_2SO_4$ (Fisher Scientific) – which corresponds to between 0.055g/L and 1.365g/L ammonium. Minimal media formulations utilized for bioreactor fermentations typically contained 80g/L glucose and 6.7g/L Yeast Nitrogen Base w/o amino acids (1.7g/L YNB and 5g/L $(NH_4)_2SO_4$). When utilizing alternative carbon sources, glucose was replaced by 80g/L arabinose (Fisher Scientific), 80g/L fructose (Alfa Aesar), 80g/L galactose (Fisher Scientific), 80g/L glycerol (Fisher Scientific), 80g/L mannose (Alfa Aesar), 80g/L maltose (Acros Organics), 80g/L ribose (MP Biomedicals), 80g/L sucrose (Acros Organics), or 80g/L Xylose (Acros Organics). Solid media for *E. coli* and *Yarrowia lipolytica* was prepared by adding 20g/L agar

(Teknova) to liquid media formulations. Leucine (MP Biomedicals) and isoleucine (Sigma Aldrich) supplementation was used to analyze the effect of leucine biosynthetic capacity. Leucine was added at a concentration of 0.8g/L or 1.6g/L, while isoleucine was added at concentration of 1.6g/L. Inhibition of the TOR protein was caused by supplementation with 200ng/mL rapamycin (LC Laboratories).

9.6.2 Plasmid construction

Primer sequences can be found in the **Table 9.5**. All *Y. lipolytica* episomal plasmids were centromeric, replicative vectors derived from plasmid pSl16-Cen1-1(227) [98] after it had been modified to include a multi-cloning site, a hrGFP green fluorescent reporter gene (pIRES-hrGFP, Agilent) driven by the strong UAS1B$_{16}$-TEF promoter [2], and a cyc1 terminator [210] to create plasmid pMCS-UAS1B$_{16}$-TEF-hrGFP. Integrative plasmids were derived from plasmids pUC-S1-UAS1B$_{16}$-Leum or pUC-S1-UAS1B$_{16}$-TEF [4] that contained 5' and 3' rDNA integrative sequences surrounding the following elements - (from 5' to 3') a uracil section marker surrounded by LoxP sites for marker retrieval, the strong UAS1B$_{16}$-Leum or UAS1B$_{16}$-TEF promoter, AscI and PacI restriction enzyme sites for gene insertion, and a XPR2 minimal terminator. These integrative plasmids were also designed to contain two identical NotI restriction enzyme sites directly outside of the rDNA regions so that plasmid linearization would simultaneously remove *E. coli* pUC19-based DNA. All plasmids containing expression cassettes were sequenced confirmed before transformation into *Y. lipolytica*. All auxotrophic or antibiotic selection markers used for chromosomal integrations were

flanked with LoxP sites to allow for retrieval of integrated markers with the pMCS-UAS1B$_{16}$-TEF-Cre replicative vector [4].

9.6.2.1 Construction of episomal expression cassettes

The following genes were PCR amplified from *Y. lipolytica* PO1f gDNA and inserted into vector pMCS-UAS1B$_{16}$-TEF-hrGFP in place of hrGFP with an AscI/PacI digest: AMPD, ACL subunit 1 (ACL1), ACL subunit 2 (ACL2), MEA1, DGA1, and DGA2 with primers JB387/388, JB402/404, JB405/407, AH020/021, JB911/912, and JB913/914, respectively. This formed plasmids pMCS-UAS1B$_{16}$-TEF-AMPD, pMCS-UAS1B$_{16}$-TEF-ACL1, pMCS-UAS1B$_{16}$-TEF-ACL2, pMCS-UAS1B$_{16}$-TEF-MEA, pMCS-UAS1B$_{16}$-TEF-DGA1, and pMCS-UAS1B$_{16}$-TEF-DGA2.

A leucine marker containing plasmid expressing the Cre-Recombinase gene, pMCS-UAS1B$_{16}$-TEF-Cre, has been described previously [4].

9.6.2.2 Construction of integrative expression cassettes

The following genes were gel extracted from the previously constructed episomal expression vectors and inserted into vector pUC-S1-UAS1B$_{16}$-TEF with an AscI/PacI digest: AMPD, ACL subunit 1 (ACL1), ACL subunit 2 (ACL2), MEA1, DGA1, and DGA2. This formed plasmids pUC-S1-UAS1B$_{16}$-TEF-AMPD, pUC-S1-UAS1B$_{16}$-TEF-ACL1, pUC-S1-UAS1B$_{16}$-TEF-ACL2, pUC-S1-UAS1B$_{16}$-TEF-MEA1, and pUC-S1-UAS1B$_{16}$-TEF-DGA1, and pUC-S1-UAS1B$_{16}$-TEF-DGA2. The loxP-surrounded uracil marker of these integrative plasmids was replaced with a loxP-surrounded leucine marker created by amplification of pMCSCen1 template with primers JB862/863 followed by

insertion using a BstBI/SacII digest. These plasmids enabled integrative selection with leucine auxotrophy and co-expression of two enzymes without marker retrieval. These leucine marker integrative plasmids were dubbed plasmids pUC-S2-UAS1B$_{16}$-TEF-AMPD, pUC-S2-UAS1B$_{16}$-TEF-ACL1, pUC-S2-UAS1B$_{16}$-TEF-ACL2, pUC-S2-UAS1B$_{16}$-TEF-MEA1, and pUC-S2-UAS1B$_{16}$-TEF-DGA1, and pUC-S2-UAS1B$_{16}$-TEF-DGA2.

ACL1 and ACL2 were similarly inserted into pUC-S1-UAS1B$_{16}$-Leum with primers JB403/404 and JB406/407, respectively, to form plasmids pUC-S1-UAS1B$_{16}$-Leum-ACL1 and pUC-S1-UAS1B$_{16}$-Leum-ACL2.

9.6.3 Strain construction

All strains were confirmed through gDNA extraction and PCR confirmation and are listed in **Table 6.2**. We previously constructed two markerless single-gene deletion strains in the *Y. lipolytica* PO1f background, PO1f-Δmfe1 and PO1f-Δpex10, deficient in their β-oxidation and peroxisomal biogenesis capacity, respectively [4]. Following our previous protocol, the PEX10 gene was deleted from strain PO1f-Δmfe1 to form the markerless double mutant PO1f-Δmfe1-Δpex10. These four strains, referred to as PO1f, *pex10*, *mfe1*, and *pex10 mfe1* were utilized as backgrounds for single and double overexpression of the AMPD, ACL1, ACL2, MEA, DGA1, and DGA2 proteins, including variation in selective marker utilized, i.e., leucine (chromosomal or episomal expression cassette) vs. uracil (chromosomal expression cassette). Integrative cassettes were linearized, transformed into the four background strains, and selected for on appropriate dropout plates. Integrative vectors without open reading frames to express,

pUC-S1-UAS1B$_{16}$-TEF and pUC-S2-UAS1B$_{16}$-TEF, were utilized to create strains with leucine, uracil, or both leucine and uracil prototrophies, but without enzyme overexpression **(Table 6.2)**.

Table 9.5: Primers used in Chapter 6

JB387	TTGGCGCGCCatgccgcagcaagcaatgg
JB388	CCTTAATTAAttaaccatgcagccgctcaaac
JB402	TTGGCGCGCCatgtctgccaacgagaacat
JB403	TTGGCGCGCCtctgccaacgagaacatctc
JB404	CCTTAATTAActatgatcgagtcttggccttg
JB405	TTGGCGCGCCATGTCAGCGAAATCCATTCACG
JB406	TTGGCGCGCCTCAGCGAAATCCATTCACGAG
JB407	CCTTAATTAATTAAACTCCGAGAGGAGTGGAA
JB862	CCAccgcggataacttcgtataatgtatgctatacgaagttatgagtctttattggtgatgggaaga
JB863	CGGTTCGAAataacttcgtatagcatacattatacgaagttatcagtcgccagcttaaagatatcta
JB911	cattcaaaGGCGCGCCatgactatcgactcacaatactaca
JB912	gcGGATCCTTAATTAAttactcaatcattcggaactctgg
JB913	cattcaaaGGCGCGCCATGGAAGTCCGACGACGAAA
JB914	gcGGATCCTTAATTAACTACTGGTTCTGCTTGTAGTTGT
AH020	GACTGGCGCGCCATGTTACGACTACGAACCATGC
AH021	GTCCTTAATTAACTAGTCGTAATCCCGCACATG

9.6.4 Fatty acid characterization by nile red staining coupled with flow cytometry or fluorescence microscopy:

Nile Red (MP Biomedicals) is commonly utilized to stain oleaginous cellular material and can be coupled with fluorescence flow cytometry to gauge relative lipid content [182]. *Y. lipolytica* strains were routinely inoculated from glycerol stock in biological triplicate in appropriate media for 72 hours at 30^{o}C with shaking. Cell concentrations were normalized to a specific OD_{600} for reinoculation in fresh media and further incubation. For assays in which the effect of media formulation was not being

investigated, this media contained .79g/L CSM, 1.7g/L Yeast Nitrogen Base w/o amino acid and w/o $(NH_4)_2SO_4$, 80g/L carbon source, and 5g/L ammonium sulfate, as this formulation was shown to strongly induce lipid accumulation in the highest lipid producing strains. For large experiments, 2mL cultures were utilized to test large number of cultures and were inoculated to an OD_{600} =2.5, and larger volume cultures were inoculated to an OD_{600} = 0.1. Cultures were incubated for two to eight days at 30^{o}C with constant agitation. 2mL cultures were incubated in a rotary drum (CT-7, New Brunswick Scientific) at speed seven. Flasks were shaken at 225rpm in a standing incubator, and bioreactors were agitated by rotor at no less that 225rpm and no more than 800rpm. To harvest, one OD_{600} unit of culture was spun down at 1000g for three minutes and resuspended in 500 µL Phosphate Buffered Saline solution (PBS) (Sigma Aldrich). 6µL of 1mM Nile Red (dissolved in DMSO) was added, and then cells were incubated in the dark at room temperature for 15 minutes. Cells were spun down at 1000g for three minutes, resuspended in 800µL ice cold water, spun down again, and resuspended again in 800µL ice cold water. 300µL of stained cells was added to 1ml ice cold water and tested with a FACS Fortessa (BD Biosciences), a voltage of 350, a 10,000 cell count, a forward scatter of 125, a side scatter of 125, and the 535LP and 585/42BP filters for fluorescence detection using the GFP fluorochrome. Samples were kept on ice and in the dark during the test and fluorescence data was analyzed using FlowJo software (Tree Star Inc., Ashland, OR) to compute mean fluorescence values. Day-to-day variability was mitigated by analyzing all comparable strains on the same day. An average fluorescence and standard deviation were calculated from the mean values of biological replicates.

Stained cells were routinely examined with fluorescence microscopy under a 100X oil immersion objective using the FITC channel on an Axiovert 200M microscope (Zeiss).

9.6.5 Lipid quantification and fatty acid profile analysis

Lipids from 500μL culture were extracted following the procedure described by [218] and modified for yeast [219]. Dried lipids were transesterified with N-tert-Butyldimethylsilyl-N-methyltrifluoroacetamide (Sigma-Aldrich) following the procedure of (Paik et al., 2009), and 2 μL samples were injected into a GC-FID (Agilent Technologies 6890 Network GC System) equipped with an Agilent HP-5 column (5% phenyl-95% methylsiloxane - product number 19091J-413) to analyze fatty acid fractions. Briefly, the following settings were used: Detector Temp = 300^{o}C, He Flow = 1.0 mL/min, Oven Temp = 80°C for 2 min, increased at 30°C/min to 200 °C, increased at 2°C/min to 229 °C, increased at 1°C/min to 232°C, increased at 50°C/min to 325°C. Fatty acid standards for C16:0 palmitic acid, C16:1(n-7) palmitoleic acid, C17:0 heptadecanoic acid, C18:0 stearic acid, C18:1 (n-9) oleic acid, and C18:2 (n-6) linoleic acid were purchased from Sigma-Aldrich, transesterified, and analyzed by GC to identify fatty acid peaks. Similarly, C16:1(n-9) palmitoleic acid was purchased from Cayman Chemical and used a standard. GC-MS confirmation of fatty acid type was performed with a Thermo Scientific TSQ Quantum Triple Quad using chemical ionization and a 1.5mL/minute flow rate.

9.6.6 Ammonium quantification

1 mL of culture was heated to 80^{o}C for 15 minutes, and then centrifuged at 17,900 x g for 3 minutes. Supernatant was stored at 4^{o}C for less than 1 week, and ammonium concentration was determined using the R-Biopharma Ammonium Assay kit following the manufacturer's instructions. Ammonium Assay kit accuracy was assessed by measuring ammonium concentration in solutions with varying concentrations of Yeast Nitrogen Base w/o amino acids. Minor necessary adjustments were made using the resulting standard curve.

9.6.7 Bioreactor fermentations

Typically, bioreactor fermentations were run in minimal media (described above) as batch processes. However, one fermentation included a spike of an additional 80g/L glucose at the 72 hour timepoint, and another had a doubled media formulation that contained 160g/L glucose and 13.4g/L YNB w/o amino acids. All fermentations were inoculated to an initial OD_{600} = 0.1 in 1.5L of media. Dissolved oxygen was maintained at 50% of maximum by varying rotor speed between 250rpm and 800rpm with a constant air input flow rate of 2.5v/v/m (3.75v/m). PH was maintained at 3.5 or above with 1M NaOH, and temperature was maintained at 28^{o}C. 10-15 mL samples were taken every twelve hours, and fermentations lasted 6-7 days. We ran several fermentations with suboptimal conditions before settling on the above parameters.

9.6.8 Transesterification

Y. lipolytica lipid reserves were transesterified using acid-promoted direct methanolysis of cellular biomass [221]. 1L of *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1p fermented in a bioreactor for seven days as described above was washed twice in 400mL water. Cells were dried on a hotplate at 140^{o}C for 3 hours. The dried cell mass was transesterified with 2% w/v H_2SO_4 in 200mL methanol at a fast boil with reflux and constant agitation for 72 hours. The reaction mixture was centrifuged to remove cellular debris. FAMEs (fatty acid methyl esters) were extracted from the supernatant by adding 0.2 volumes water, mixing, centrifuging, and removing the polar phase. Additional FAMEs were extracted in the polar phase with a second extraction using 0.4 volume of water. FAMEs were washed in 1 volume of water and analyzed with TLC and GC.

9.6.9 Thin layer chromatography

A thin layer chromatography of transesterified lipid product was run on a silica gel 60 TLC plate with hexane (Fischer Scientific) - diethyl ether (Sigma Aldrich) - acetic acid (Mallinckrodt Baker, Inc) at 40:10:1 using 6µl of 1.25 µg/µl standard mixture and 6µl of 1 µg/µl sample and then stained with vanillin (Sigma Aldrich).

9.6.10 Protein extraction

Protein content from 0.5 to 1.0 mL of culture was extracted using the Pierce BCA Protein Assay Kit following the manufacturer's instructions. Protein concentration (mg/mL) was normalized per mL of culture and per culture OD_{600} to normalize to the

individual cellular level. PO1f leucine$^+$ uracil$^+$ and *pex10 mfe1* leucine$^+$ uracil$^+$ DGA1p strains were analyzed in this manner after fermentation in a bioreactor.

9.6.11 Glucose quantification

Supernatant was diluted 1:10 and glucose concentration was quanitified using a YSI Life Sciences Bioanalyzer 7100MBS.

9.7 MATERIALS AND METHODS FOR CHAPTER 7

9.7.1 Media

Cultivation for itaconic acid production entailed the following: Y*arrowia lipolytica* strains were cultivated for two days at 30^{o}C with constant agitation in 2mL cultures and then reinoculated to an OD600 = 0.005 in 15mL media in flasks and shaken for three days at 30^{o}C at 225rpm.

Itaconic acid accumulation response towards media formulation was investigated by cultivation in varying concentrations of glucose and nitrogen. These media formulations contained .79g/L CSM, 1.7g/L Yeast Nitrogen Base w/o amino acid and w/o $(NH_4)_2SO_4$ (Becton, Dickinson, and Company), and the following concentrations of glucose and ammonium sulfate - 20g/L and 5g/L ammonium sulfate, 20g/L glucose and 0.2g/L ammonium sulfate, 80g/L and 5g/L ammonium sulfate. Minimal media formulations utilized 20g/L glucose, 6.7g/L Yeast Nitrogen Base w/o amino acids (1.7g/L YNB and 5g/L $(NH_4)_2SO_4$), and uracil supplementation at .02g/L if necessary.

9.7.2 Plasmid construction

Primer sequences can be found in **Table 9.6**. Four gBlocks gene fragments (Integrated DNA Technologies) were designed to encompass the intronless CAD1 gene sequence from *Aspergillus terreus* with at least 50 nucleotide overlap between each gBlock and with the p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-P_{GPD} vector backbone [3]. Primers JB931/932, JB933/934, JB935/936, and JB937/938 were used to PCR amplify the four gBlocks. Amplified gBlock DNA fragments and linearized p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-P_{GPD} vector backbone were transformed into *S. cerevisiae* BY4741 following Hegemann's yeast transformation protocol [217] to enable homologous recombination mediated gene assembly [222]. Plasmid p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-P_{GPD}–AtCAD1 was isolated from transformed BY4741 with a yeast miniprep, transformed into *E. coli*, miniprepped, and sequence confirmed.

Primers JB1050 JB1051 were used to amplify the CAD1 gene from plasmid p416-UAS_{TEF}-UAS_{CIT}-UAS_{CLB}-P_{GPD}–AtCAD1 and insert it into the pUC-S2-$UAS1B_{16}$-TEF and pMCS-$UAS1B_{16}$-TEF chromosomal and episomal expression vectors (respectively) with an AscI/PacI digest to form plasmids pUC-S2-$UAS1B_{16}$-TEF-CAD1 and pMCS-$UAS1B_{16}$-TEF-CAD1.

Table 9.6: Primers used in Chapter 7

Primers	Sequence (5'--> 3')
JB931	gtattgattgtaattctgtaaatctatttc
JB932	cttgctgcaaagaccgcaggaaggacaatgcttgcagagtgtaggggggcttcgctgtgg
JB933	tttcatacaggctacggagcttgacgactaccacagcgaagccccctacactctgcaag
JB934	gaggctctctgccgttgccc
JB935	ttcttgggggactgttggcc
JB936	agatgaagtaaccttcctggccagatc
JB937	ccgtccagctggtcgaccag
JB938	ctccttcctttttcggttagagcggatgtgggggggagggcgtgaatgtaa
JB1050	gagtggcgcgccatgaccaaacaatctgcgg
JB1051	gcacttaattaattataccagtggcgatttca

9.7.3 Strain construction

All strains were confirmed through gDNA extraction and PCR confirmation. We previously constructed and AMPD chromosomal expression strain utilizing the uracil auxotrophic marker (**Table 9.5**), referred to in Chapter 7 as PO1f AMPD. Episomal (pMCS-UAS1B$_{16}$-TEF-CAD1) and chromosomal (NotI-HF linearized pUC-S2-UAS1B$_{16}$-TEF-CAD1) expression cassettes were transformed in *Y. lipolytica* PO1f and PO1f AMPD to form four strains: PO1f CAD (episomal and chromosomal) and PO1f AMPD CAD (episomal and chromosomal).

References

1 Blazeck, J. *et al.* Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **97**, 3037-3052, doi: DOI: 10.1007/s00253-012-4421-5 (2013).

2 Blazeck, J., Liu, L., Redden, H. & Alper, H. Tuning Gene Expression in *Yarrowia lipolytica* by a Hybrid Promoter Approach. *Applied and Environmental Microbiology* **77**, 7905-7914 (2011).

3 Blazeck, J., Garg, R., Reed, B. & Alper, H. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering* **109**, 2884-2995, doi:doi: 10.1002/bit.24552 (2012).

4 Blazeck, J., Liu, L., Knight, R. & Alper, H. Heterologous production of pentane in the oleaginous yeast *Yarrowia lipolytica*. *Journal of Biotechnology* **165**, 184-194, doi:10.1016/j.jbiotec.2013.04.003 (2013).

5 Tevz, G., Bencina, M. & Legisa, M. Enhancing itaconic acid production by *Aspergillus terreus*. *Applied Microbiology and Biotechnology* **87**, 1657-1664, doi:10.1007/s00253-010-2642-z (2010).

6 Becker, J., Zelder, O., Hafner, S., Schroder, H. & Wittmann, C. From zero to hero-Design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. *Metab. Eng.* **13**, 159-168, doi:10.1016/j.ymben.2011.01.003 (2012).

7 Lutke-Eversloh, T. & Bahl, H. Metabolic engineering of *Clostridium acetobutylicum*: recent advances to improve butanol production. *Current Opinion in Biotechnology* **22**, 634-647, doi:10.1016/j.copbio.2011.01.011 (2011).

8 Barth, G. & Gaillardin, C. in *Nonconventional Yeasts in Biotechnology: A Handbook* Vol. 1 (ed Klaus Wolf) Ch. 10, 313-388 (Springer, 1996).

9 Papanikolaou, S. *et al.* Biosynthesis of lipids and organic acids by *Yarrowia lipolytica* strains cultivated on glucose. *European Journal of Lipid Science and Technology* **111**, 1221-1232, doi:10.1002/ejlt.200900055 (2009).

10 Alper, H. & Stephanopoulos, G. Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat. Rev. Microbiol.* **7**, 715-723, doi:10.1038/nrmicro2186 (2009).

11 Blazeck, J. & Alper, H. Systems metabolic engineering: Genome-scale models and beyond. *Biotechnology Journal* **5**, 647-659, doi:10.1002/biot.200900247 (2010).

12 Hahn, S. & Young, E. T. Transcriptional Regulation in *Saccharomyces cerevisiae*: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. *Genetics* **189**, 705-736, doi:10.1534/genetics.111.127019 (2011).

13 Liu, X., Bushnell, D. A., Wang, D., Calero, G. & Kornberg, R. D. Structure of an RNA Polymerase II-TFIIB Complex and the Transcription Initiation Mechanism. *Science* **327**, 206-209, doi:10.1126/science.1182015 (2010).

14 Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annu. Rev. Biochem.* **72**, 449-479, doi:10.1146/annurev.biochem.72.121801.161520 (2003).
15 Juven-Gershon, T., Hsu, J. Y. & Kadonaga, J. T. Perspectives on the RNA polymerase II core promoter. *Biochemical Society Transactions* **34**, 1047-1050 (2006).
16 Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**, 699-709, doi:10.1016/s0092-8674(04)00205-3 (2004).
17 Monteiro, P. T. *et al.* YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **36**, D132-D136, doi:10.1093/nar/gkm976 (2008).
18 Badis, G. *et al.* A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. *Molecular Cell* **32**, 878-887, doi:10.1016/j.molcel.2008.11.020 (2008).
19 Ferretti, V. *et al.* PReMod: a database of genome-wide mammalian cis-regulatory module predictions. *Nucleic Acids Research* **35**, D122-D126, doi:10.1093/nar/gkl879 (2007).
20 Zhu, C. *et al.* High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Research* **19**, 556-566, doi:10.1101/gr.090233.108 (2009).
21 Zhu, J. & Zhang, M. Q. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* **15**, 607-611, doi:10.1093/bioinformatics/15.7.607 (1999).
22 Keseler, I. M. *et al.* EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Research* **33**, D334-D337, doi:10.1093/nar/gki108 (2005).
23 Cheng, J. X., Floer, M., Ononaji, P., Bryant, G. & Ptashne, M. Responses of four yeast genes to changes in the transcriptional machinery are determined by their promoters. *Current Biology* **12**, 1828-1832, doi:10.1016/s0960-9822(02)01257-5 (2002).
24 Michel, D. How transcription factors can adjust the gene expression floodgates. *Prog. Biophys. Mol. Biol.* **102**, 16-37, doi:10.1016/j.pbiomolbio.2009.12.007 (2009).
25 Michel, D. Fine tuning gene expression through short DNA-protein binding cycles. *Biochimie* **91**, 933-941, doi:10.1016/j.biochi.2009.03.022 (2009).
26 Bassel, J. & Mortimer, R. Genetic order of galactose structural genes in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **108**, 179-183 (1971).
27 Murphy, K. F., Balazsi, G. & Collins, J. J. Combinatorial promoter design for engineering noisy gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12726-12731, doi:10.1073/pnas.0608451104 (2007).
28 Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Molecular Systems Biology* **3**, 11 (2007).

29 West, R. W., Yocum, R. R. & Ptashne, M. *Saccharomyces cerevisiae* Gal1-Gal10 Divergent Promoter Region: Location and Function of the Upstream Activating Sequence UAS_G. *Molecular and Cellular Biology* **4**, 2467-2478 (1984).
30 Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *P Natl Acad Sci USA* **102**, 12678-12683, doi:DOI 10.1073/pnas.0504604102 (2005).
31 Nevoigt, E. *et al.* Engineering of promoter replacement cassettes for fine-tuning of gene expression in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology* **72**, 5266-5273, doi:10.1128/aem.00530-06 (2006).
32 Nevoigt, E. *et al.* Engineering promoter regulation. *Biotechnology and Bioengineering* **96**, 550-558, doi:10.1002/bit.21129 (2007).
33 Deboer, H. A., Comstock, L. J. & Vasser, M. The Tac Promoter - a Functional Hybrid Derived from the Trp and Lac Promoters. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **80**, 21-25 (1983).
34 Haldimann, A., Daniels, L. L. & Wanner, B. L. Use of new methods for construction of tightly regulated arabinose and rhamnose promoter fusions in studies of the *Escherichia coli* phosphate regulon. *Journal of Bacteriology* **180**, 1277-1286 (1998).
35 Sengstag, C. & Hinnen, A. A 28-bp sement of the *Saccharomyces cerevisiae* PHO5 upstream activator sequence confers phosphate control to the Cyc1-LacZ gene fusion. *Gene* **67**, 223-228, doi:10.1016/0378-1119(88)90399-x (1988).
36 Rosenberg, S. & Tekamp-Olson, P. Enhanced yeast transcription employing hybrid GAPDH promoter region constructs. United States patent 5089398 (1992).
37 Guarente, L. & Hoar, E. Upstream activation sites of the Cyc1 gene of *Sacchomyces cerevisiae* are active when inverted but not when placed downstream of the TATA box. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81**, 7860-7864 (1984).
38 Guarente, L., Lalonde, B., Gifford, P. & Alani, E. Distinctly Regulated Tandem Upstream Activation Site Mediate Catabolite Repression of the CYC1 Gene of *Saccharomyces cerevisiae*. *Cell* **36**, 503-511 (1984).
39 Guarente, L., Yocum, R. R. & Gifford, P. A Gal10-Cyc1 hybrid yeast promoter identifies the Gal4 regulatory region as an upstream site. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **79**, 7410-7414 (1982).
40 Johnston, M., Flick, J. S. & Pexton, T. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **14**, 3834-3841 (1994).
41 Szabo, R. Dimorphism in *Yarrowia lipolytica*: Filament formation is suppressed by nitrogen starvation and inhibition of respiration. *Folia Microbiologica* **44**, 19-24 (1999).
42 Cervantes-Chavez, J. A., Kronberg, F., Passeron, S. & Ruiz-Herrera, J. Regulatory role of the PKA pathway in dimorphism and mating in *Yarrowia*

lipolytica. *Fungal Genetics and Biology* **46**, 390-399, doi:10.1016/j.fgb.2009.02.005 (2009).
43 Guevaraolvera, L., Calvomendez, C. & Ruizherrera, J. The role of polyamine metabolism in dimorphism in *Yarrowia lipolytica*. *Journal of General Microbiology* **139**, 485-493 (1993).
44 Zinjarde, S. S., Kale, B. V., Vishwasrao, P. V. & Kumar, A. R. Morphogenetic behavior of tropical marine yeast *Yarrowia lipolytica* in response to hydrophobic substrates. *Journal of Microbiology and Biotechnology* **18**, 1522-1528 (2008).
45 Muller, S., Sandal, T., Kamp-Hansen, P. & Dalboge, H. Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. *Yeast* **14**, 1267-1283 (1998).
46 Ángel Domínguez, E. F., Manuel Sánchez, Francisco J. González, Flor M Pérez-Campo, Susana García, Ana B. Herrero, Avelino San Vicente, Juan Cabello, Marciano Prado, Francisco J. Iglesias, Altino Choupina, Francisco J. Burguillo, Luis Fernández-Lago, M. Carmen López. Non-conventional yeasts as hosts for heterologous protein production. *International Microbiology* **1**, 131-142 (1998).
47 Madzak, C., Gaillardin, C. & Beckerich, J. M. Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *Journal of Biotechnology* **109**, 63-81, doi:10.1016/j.jbiotec.2003.10.027 (2004).
48 Swennen, D. *et al.* Secretion of active anti-Ras single-chain Fv antibody by the yeasts *Yarrowia lipolytica* and *Kluyveromyces lactis*. *Microbiology-Sgm* **148**, 41-50 (2002).
49 Nicaud, J. M. *et al.* Protein expression and secretion in the yeast Yarrowia lipolytica. *Fems Yeast Research* **2**, 371-379 (2002).
50 Santiago-Gomez, M. P. *et al.* Characterization of purified green bell pepper hydroperoxide lyase expressed by *Yarrowia lipolytica*: Radicals detection during catalysis. *Enzyme and Microbial Technology* **41**, 13-18, doi:10.1016/j.enzmictec.2006.11.017 (2007).
51 Park, C. S., Chang, C. C., Kim, J. Y., Ogrydziak, D. M. & Ryu, D. D. Y. Expression, secretion, and processing of rice alpha-amylase in the yeast *Yarrowia lipolytica*. *Journal of Biological Chemistry* **272**, 6876-6881 (1997).
52 Ryu, H. M., Kang, W. K., Kang, H. A. & Kim, J. Y. Secretion of active urokinase-type plasminogen activator from the yeast *Yarrowia lipolytica*. *Biotechnology and Bioprocess Engineering* **8**, 162-165, doi:10.1007/bf02940274 (2003).
53 Madzak, C., Treton, B. & Blanchin-Roland, S. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *Journal of Molecular Microbiology and Biotechnology* **2**, 207-216 (2000).
54 Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35-44, doi:10.1038/nature02579 (2004).

55 Andre, A. *et al.* Biotechnological conversions of bio-diesel-derived crude glycerol by *Yarrowia lipolytica* strains. *Engineering in Life Sciences* **9**, 468-478, doi:10.1002/elsc.200900063 (2009).
56 Barth, G. & Gaillardin, C. Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *Fems Microbiology Reviews* **19**, 219-237 (1997).
57 Fickers, P. *et al.* Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *Fems Yeast Research* **5**, 527-543, doi:10.1016/j.femsyr.2004.09.004 (2005).
58 Rywinska, A., Musial, I., Rymowicz, W., Zarowska, B. & Boruczkowski, T. Effect of agitation and aeration on the citric acid production by *Yarrowia lipolytica* grown on glycerol. *Preparative Biochemistry & Biotechnology* **42**, 279-291, doi:10.1080/10826068.2012.656868 (2012).
59 Beopoulos, A. *et al. Yarrowia lipolytica* as a model for bio-oil production. *Prog. Lipid Res.* **48**, 375-387, doi:10.1016/j.plipres.2009.08.005 (2009).
60 Papanikolaou, S. & Aggelis, G. Modeling lipid accumulation and degradation in *Yarrowia lipolytica* cultivated on industrial fats. *Curr Microbiol* **46**, 398-402 (2003).
61 Dulermo, T. & Nicaud, J. M. Involvement of the G3P shuttle and beta-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab. Eng.* **13**, 482-491, doi:10.1016/j.ymben.2011.05.002 (2011).
62 Beopoulos, A., Chardot, T. & Nicaud, J. M. *Yarrowia lipolytica*: A model and a tool to understand the mechanisms implicated in lipid accumulation. *Biochimie* **91**, 692-696, doi:10.1016/j.biochi.2009.02.004 (2009).
63 Beopoulos, A. *et al.* Control of Lipid Accumulation in the Yeast *Yarrowia lipolytica*. *Applied and Environmental Microbiology* **74**, 7779-7789, doi:10.1128/aem.01412-08 (2008).
64 Papanikolaou, S. & Aggelis, G. Lipid production by *Yarrowia lipolytica* growing on industrial glycerol in a single-stage continuous culture. *Bioresour Technol* **82**, 43-49 (2002).
65 Athenstaedt, K. *et al.* Lipid particle composition of the yeast *Yarrowia lipolytica* depends on the carbon source. *Proteomics* **6**, 1450-1459 (2006).
66 Sherman, D. *et al.* Génolevures complete genomes provide data and tools for comparative genomics of hemiascomycetous yeasts. *Nucleic Acids Research* **34**, D432-D435, doi:10.1093/nar/gkj160 (2006).
67 Davidow, L. S. *et al.* Integrative Transformation of the Yeast *Yarrowia lipolytica*. *Current Genetics* **10**, 39-48 (1985).
68 Chen, D. C., Yang, B. C. & Kuo, T. T. One Step Transformation of Yeast in Stationary Phase. *Current Genetics* **21**, 83-84 (1992).
69 Fickers, P., Le Dall, M. T., Gaillardin, C., Thonart, P. & Nicaud, J. M. New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *Journal of Microbiological Methods* **55**, 727-737, doi:10.1016/j.mimet.2003.07.003 (2003).

70 Vernis, L. *et al.* An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology* **17**, 1995-2004 (1997).
71 Vernis, L. *et al.* Only centromeres can supply the partition system required for ARS function in the yeast *Yarrowia lipolytica*. *Journal of Molecular Biology* **305**, 203-217, doi:10.1006/jmbi.2000.4300 (2001).
72 Fournier, P. *et al.* Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4912-4916 (1993).
73 Matsuoka, M. *et al.* Analysis of regions essential for the function of chromosomal replicator sequences from *Yarrowia lipolytica*. *Molecular & General Genetics* **237**, 327-333 (1993).
74 Juretzek, T. *et al.* Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast* **18**, 97-113 (2001).
75 Ledall, M. T., Nicaud, J. M. & Gaillardin, C. Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Current Genetics* **26**, 38-44 (1994).
76 Vanheerikhuizen, H. *et al.* Heterogeneity in the ribosomal RNA genes of the yeast *Yarrowia lipolytica* - Cloning and analysis of two size classes of repeats. *Gene* **39**, 213-222 (1985).
77 Schirmer, A., Rude, M. A., Li, X. Z., Popova, E. & del Cardayre, S. B. Microbial Biosynthesis of Alkanes. *Science* **329**, 559-562, doi:10.1126/science.1187936 (2010).
78 Shi, S. B., Valle-Rodriguez, J. O., Khoomrung, S., Siewers, V. & Nielsen, J. Functional expression and characterization of five wax ester synthases in *Saccharomyces cerevisiae* and their utility for biodiesel production. *Biotechnology for Biofuels* **5**, 1-10.
79 Sanders, T. H., Pattee, H. E. & Singleton, J. A. Aerobic Pentane Production by Soybean Lipoxygenase Isozymes. *Lipids* **10**, 3 (1975).
80 Cheesbrough, T. M. & Kolattukudy, P. E. Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from *Pisum sativum*. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6613-6617 (1984).
81 Chuang, L. T. *et al.* Co-expression of heterologous desaturase genes in Yarrowia lipolytica. *New Biotech.* **27**, 277-282, doi:10.1016/j.nbt.2010.02.006 (2010).
82 Pattee, H. E., Singleto.Ja & Johns, E. B. Pentane production by peanut lipoxygenase. *Lipids* **9**, 302-306 (1974).
83 Bonnarme, P. *et al.* Itaconate Biosynthesis in *Aspergillus terreus*. *Journal of Bacteriology* **177**, 3573-3578 (1995).
84 Botham, P. A. & Ratledge, C. Biochemical explanation for lipid accumulation in Candida-107 and other oleaginous microorganisms. *Journal of General Microbiology* **114**, 361-375 (1979).
85 Ratledge, C. Regulation of lipid accumulation in oleaginous micro-organisms. *Biochemical Society Transactions* **30**, 1047-1050 (2002).

86 Alfeel, W., Chirala, S. S. & Wakil, S. J. Cloning of the yeast FAS3 gene and primary structure of yeast acetyl-CoA carboxylase. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4534-4538 (1992).
87 Kennedy, E. P. Biosynthesis of Complex Lipids. *Federation Proceedings* **20**, 934-940 (1961).
88 Beopoulos, A. *et al.* Identification and characterization of DGA2, an acyltransferase of the DGAT1 acyl-CoA:diacylglycerol acyltransferase family in the oleaginous yeast *Yarrowia lipolytica*. New insights into the storage lipid metabolism of oleaginous yeasts. *Applied Microbiology and Biotechnology* **93**, 1523-1537, doi:10.1007/s00253-011-3506-x (2012).
89 Beopoulos, A., Nicaud, J. M. & Gaillardin, C. An overview of lipid metabolism in yeasts and its impact on biotechnological processes. *Applied Microbiology and Biotechnology* **90**, 1193-1206, doi:10.1007/s00253-011-3212-8 (2011).
90 Hong, S., Sharpe, P., Xue, Z., Yadav, N. & Zhu, Q. Peroxisome biogenesis factor protein (pex) disruptions for altering the content of polyunsaturated fatty acids and the total lipid content in oleaginous eukaryotic organisms. USA patent (2012).
91 Zhu, Q. *et al.* in *SIMB Annual Meeting & Exhbition* (Washington, D.C., 2012).
92 Tai, M. & Stephanopoulos, G. Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. *Metab. Eng.* **15**, 1-9, doi:10.1016/j.ymben.2012.08.007 (2013).
93 Pfleger, B. F., Pitera, D. J., D Smolke, C. & Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology* **24**, 1027-1032, doi:10.1038/nbt1226 (2006).
94 Rosenberg, S. & Tekamp-olson, P. Enhanced yeast transcription employing hybrid GAPDH promoter region constructs. United States patent (1992).
95 Mukai, H. *et al.* Yeast promoter and process for preparing heterologous protein. Europe patent (1992).
96 Madzak, C., Blanchin-Roland, S., Otero, R. R. C. & Gaillardin, C. Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. *Microbiology-Uk* **145**, 75-87 (1999).
97 Blanchinroland, S., Otero, R. R. C. & Gaillardin, C. 2 Upstream Activation Sequences Control the Expression of the Xpr2 Gene in the Yeast *Yarrowia-Lipolytica*. *Molecular and Cellular Biology* **14**, 327-338 (1994).
98 Yamane, T., Sakai, H., Nagahama, K., Ogawa, T. & Matsuoka, M. Dissection of centromeric DNA from yeast *Yarrowia lipolytica* and identification of protein-binding site required for plasmid transmission. *J Biosci Bioeng* **105**, 571-578, doi:S1389-1723(08)70113-0.
99 Gasmi, N., Fudalej, F., Kallel, H. & Nicaud, J. M. A molecular approach to optimize hIFN alpha 2b expression and secretion in *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology* **89**, 109-119, doi:10.1007/s00253-010-2803-0.
100 Haas, J., Park, E. C. & Seed, B. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Current Biology* **6**, 315-324 (1996).

101 Nicaud, J. M., Fabre, E., Becherich, J. M., Fournier, P. & Gaillardin, C. Cloning, sequencing and amplification of the alkaline extracellular protease (XPR2) gene of the yeast *Yarrowia lipolytica*. *Journal of Biotechnology* **12**, 285-297 (1989).
102 Damude, H. G. H., DE, US), Gillies, Peter John (Landenberg, PA, US), Macool, Daniel Joseph (Philadelphia, PA, US), Picataggio, Stephen K. (Landenberg, PA, US), Pollak, Dana Walters M. (Media, PA, US), Ragghianti, James John (Bear, DE, US), Xue, Zhixiong (Chadds Ford, PA, US), Yadav, Narendra S. (Chadds Ford, PA, US), Zhang, Hongxiang (Chadds Ford, PA, US), Zhu, Quinn Qun (West Chester, PA, US). High eicosapentaenoic acid producing strains of *Yarrowia lipolytica*. United States patent (2006).
103 Sumita, T. *et al.* YlALK1 encoding the cytochrome P450ALK1 in *Yarrowia lipolytica* is transcriptionally induced by n-alkane through two distinct cis-elements on its promoter. *Biochem Biophys Res Commun* **294**, 1071-1078.
104 Endoh-Yamagami, S., Hirakawa, K., Morioka, D., Fukuda, R. & Ohta, A. Basic helix-loop-helix transcription factor heterocomplex of Yas1p and Yas2p regulates cytochrome P450 expression in response to alkanes in the yeast *Yarrowia lipolytica*. *Eukaryotic Cell* **6**, 734-743, doi:10.1128/ec.00412-06 (2007).
105 Yamagami, S., Morioka, D., Fukuda, R. & Ohta, A. A basic helix-loop-helix transcription factor essential for cytochrome P450 induction in response to alkanes in yeast Yarrowia lipolytica. *Journal of Biological Chemistry* **279**, 22183-22189, doi:10.1074/jbc.M313313200 (2004).
106 Miller, J. H. *Experiments in molecular genetics*. (Cold Spring Harbor Laboratory, 1972).
107 Gaillardin, C. & Ribet, A. M. LEU2 directed expression of beta-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica*. *Curr Genet* **11**, 369-375 (1987).
108 Juretzek, T., Wang, H., Nicaud, J. M., Mauersberger, S. & Barth, G. Comparison of Promoters Suitable for Regulated Overexpression of β-Galactosidase in the Alkane-Utilizing Yeast *Yarroiwa lipolytica*. *Biotechnol. Bioprocess Eng.* **5**, 320-326 (2000).
109 Blazeck, J. & Alper, H. Promoter Engineering: Recent Advances in Controlling Transcription at the Most Fundamental Level. *Biotechnology Journal* **8**, 46-58, doi:10.1002/biot.201200120 (2013).
110 Blanchin-roland, S., Otero, R. R. C. & Gaillardin, C. Two Upstream Activation Sequences Control the Expression of the XPR2 Gene in the Yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology* **14**, 327-338 (1994).
111 Drazinic, C. M., Smerage, J. B., Lopez, M. C. & Baker, H. V. Activation mechanism of the multifunctional transcription factor repressor-activator protein 1 (Rap1p). *Molecular and Cellular Biology* **16**, 3187-3196 (1996).
112 Sun, J. *et al.* Cloning and characterization of a panel of constitutive promoters for applications in pathway engineering in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*, doi:10.1002/bit.2448 (2012).
113 Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943, doi:10.1038/nature04640 (2006).

114 Lu, C. F. & Jeffries, T. Shuffling of promoters for multiple genes to optimize xylose fermentation in an engineered *Saccharomyces cerevisiae* strain. *Applied and Environmental Microbiology* **73**, 6072-6077, doi:10.1128/aem.00955-07 (2007).
115 Holland, M. J. & Holland, J. P. Isolation and identification of yeast messenger ribonucleic acids coding for enolase, glyceraldehyde-3-phosphate dehydrogenase, and phosphoglycerate kinase. *Biochemistry* **17**, 4900-4907, doi:10.1021/bi00616a007 (1978).
116 Holland, J. P. & Holland, M. J. Structural comparison of two non-tandemly repreated yeast glyceraldehyde-3-phosphate dehydrogenase genes. *Journal of Biological Chemistry* **255**, 2596-2605 (1980).
117 Gatignol, A., Dassain, M. & Tiraby, G. Cloning of *Saccharomyces cerevisiae* promoters using a probe vector based on phleomycin resistance. *Gene* **91**, 35-41, doi:10.1016/0378-1119(90)90159-o (1990).
118 Reifenberger, E., Freidel, K. & Ciriacy, M. Identification of novel HXT genes in *Saccharomyces cerevisiae* reveals the impact of individual hexose transporters on glycolytics flux. *Molecular Microbiology* **16**, 157-167, doi:10.1111/j.1365-2958.1995.tb02400.x (1995).
119 Guarente, L. Yeast promoters and LacZ fusing designed to study expression of cloned genes in yeast. *Methods in Enzymology* **101**, 181-191 (1983).
120 Walfridsson, M., Hallborn, J., Penttila, M., Keranen, S. & Hahnhagerdal, B. Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose-phosphate pathway enzymes transketolase and transaldolase. *Applied and Environmental Microbiology* **61**, 4184-4190 (1995).
121 Wisselink, H. W. *et al.* Engineering of *Saccharomyces cerevisiae* for efficient anaerobic alcoholic fermentation of L-arabinose. *Applied and Environmental Microbiology* **73**, 4881-4891, doi:10.1128/aem.00177-07 (2007).
122 Jeppsson, M., Johansson, B., Jensen, P. R., Hahn-Hagerdal, B. & Gorwa-Grauslund, M. F. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant *Saccharomyces cerevisiae* strains. *Yeast* **20**, 1263-1272, doi:10.1002/yea.1043 (2003).
123 Li, A. M. *et al.* Construction and characterization of bidirectional expression vectors in *Saccharomyces cerevisiae*. *Fems Yeast Research* **8**, 6-9, doi:10.1111/j.1567-1364.2007.00335.x (2008).
124 Gari, E., Piedrafita, L., Aldea, M. & Herrero, E. A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*. *Yeast* **13**, 837-848, doi:10.1002/(sici)1097-0061(199707)13:9<837::aid-yea145>3.0.co;2-t (1997).
125 Da Silva, N. A. & Srikrishnan, S. Introduction and Expression of Genes for Metabolic Engineering Applications in *Saccharomyces cerevisiae*. *FEMS Yeast Research* **12**, 197-214, doi:doi: 10.1111/j.1567-1364.2011.00769.x (2012).

126 Mumberg, D., Muller, R. & Funk, M. Yeast Vectors for the Controlled Expression of Heterolgous Proteins in Different Genetic Backgrounds. *Gene* **156**, 119-122, doi:10.1016/0378-1119(95)00037-7 (1995).
127 Guarente, L. & Ptashne, M. Fusion of *Escherichia coli* LacZ to the cytochrome-C gene of *Sacchomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **78**, 2199-2203 (1981).
128 Steen, E. J. *et al.* Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microbial Cell Factories* **7**, 1-8, doi:10.1186/1475-2859-7-36 (2008).
129 Shiba, Y., Paradise, E. M., Kirby, J., Ro, D. K. & Keasing, J. D. Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of isoprenoids. *Metab. Eng.* **9**, 160-168, doi:10.1016/j.ymben.2006.10.005 (2007).
130 Maury, J. *et al.* Reconstruction of a bacterial isoprenoid biosynthetic pathway in *Saccharomyces cerevisiae*. *Febs Letters* **582**, 4032-4038, doi:10.1016/j.febslet.2008.10.045 (2008).
131 Adams, B. G. Induction of glactokinase in *Saccharomyces cerevisiae* - kinetics of induction and glucose effects. *Journal of Bacteriology* **111**, 308-315 (1972).
132 Hawkins, K. M. & Smolke, C. D. The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **281**, 13485-13492, doi:10.1074/jbc.M512317200 (2006).
133 Van Slyke, C. & Grayhack, E. J. The essential transcription factor Reb1p interacts with the CLB2 UAS outside of the G(2)/M control region. *Nucleic Acids Research* **31**, 4597-4607, doi:10.1093/nar/gkg638 (2003).
134 Rosenkrantz, M., Kell, C. S., Pennell, E. A., Webster, M. & Devenish, L. J. Distinct upstream activation regions for glucose-repressed and derepressed expression of the yeast citrate synthase gene CIT1. *Current Genetics* **25**, 185-195 (1994).
135 Harper, C. & Liccione, J. (ed Agency for Toxic Substances and Disease Registry - U.S. Department of Health and Human Services) 1-224 Centers for Disease Control and Prevention (1995).
136 Faroon, O., Mandell, D. & Navarro, H. (ed U.S. Department of Health and Human Services - Agency for Toxic Substances and Disease Registry) 1-150 (Centers for Disease Control and Prevention, 1995).
137 Risher, J., Bittner, P. & Rhodes, S. (ed U.S. Department of Health and Human Services - Agency for Toxic Substances and Disease Registry) 1-200 Centers for Disease Control and Prevention (1998).
138 IARC. Vol. 45 159-201 World Health Organization, International Agency for Research on Cancer, Lyon, France (1989).
139 Lu, P.-Y. Vol. 4 Wright-Patterson Air Force Base, OH (1989).

140 Watkins, J. & Krukonis, V. Aero Propulsion and Power Directorate, Wright Research and Development Center, Air Force Systems Command, Wright-Patterson Air Force Base, OH. (1991).
141 Kizer, L., Pitera, D. J., Pfleger, B. F. & Keasling, J. D. Application of functional genomics to pathway optimization for increased isoprenoid production. *Applied and Environmental Microbiology* **74**, 3229-3241, doi:10.1128/aem.02750-07 (2008).
142 Pitera, D. J., Paddon, C. J., Newman, J. D. & Keasling, J. D. Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.* **9**, 193-207, doi:10.1016/j.ymben.2006.11.002 (2007).
143 Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* **21**, 796-802, doi:10.1038/nbt833 (2003).
144 Paradise, E. M., Kirby, J., Chan, R. & Keasling, J. D. Redirection of flux through the FPP branch-point in Saccharomyces cerevisiae by down-regulating squalene synthase. *Biotechnology and Bioengineering* **100**, 371-378, doi:10.1002/bit.21766 (2008).
145 Savage, T. J., Hristova, M. K. & Croteau, R. Evidence for an elongation/reduction/C1-elimination pathway in the biosynthesis of n-heptane in xylem of Jeffrey pine. *Plant Physiology* **111**, 1263-1269 (1996).
146 Dennis, M. & Kolattukudy, P. E. A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5306-5310 (1992).
147 Brash, A. R. Lipoxygenases: Occurrence, functions, catalysis, and acquisition of substrate. *Journal of Biological Chemistry* **274**, 23679-23682, doi:10.1074/jbc.274.34.23679 (1999).
148 Porta, H. & Rocha-Sosa, M. Plant lipoxygenases. Physiological and molecular features. *Plant Physiology* **130**, 15-21, doi:10.1104/pp.010787 (2002).
149 Bate, N. J. *et al.* Molecular characterization of an *Arabidopsis* gene encoding hydroperoxide lyase, a cytochrome P-450 that is wound inducible. *Plant Physiology* **117**, 1393-1400, doi:10.1104/pp.117.4.1393 (1998).
150 Garssen, G. J., Vliegenthart, J. F. & Boldingh, J. An anaerobic reaction between lipoxygenase, linoleic acid and its hydroperoxides. *Biochem J* **122**, 327-332 (1971).
151 Sanders, T. H., Pattee, H. E. & Singleton, J. A. Lipoxygenase isoenzymes of peanut. *Lipids* **10**, 681-685 (1975).
152 Andrianarison, R. H., Beneytout, J. L. & Tixier, M. An Enzymatic Conversion of Lipoxygenase Products by a Hydroperoxide Lyase in Blue-Green Algae (Oscillatoria sp.). *Plant Physiol* **91**, 1280-1287 (1989).
153 Vick, B. A. & Zimmerman, D. C. Metabolism of Fatty Acid Hydroperoxides by *Chlorella pyrenoidosa*. *Plant Physiol* **90**, 125-132 (1989).

154 Bourel, G. *et al.* Fatty acid hydroperoxide lyase of green bell pepper: cloning in *Yarrowia lipolytica* and biogenesis of volatile aldehydes. *Enzyme and Microbial Technology* **35**, 293-299, doi:10.1016/j.enzmictec.2003.12.014 (2004).
155 Minor, W. *et al.* Crystal structure of soybean lipoxygenase L-l at 1.4 angstrom resolution. *Biochemistry* **35**, 10687-10701, doi:10.1021/bi960576u (1996).
156 Sanders, T. H., Pattee, H. E. & Singleton, J. A. Aerobic pentane production by soybean lipoxygenase isozymes. *Lipids* **10**, 568-570 (1975).
157 Mlickova, K. *et al.* Lipid accumulation, lipid body formation, and acyl coenzyme A oxidases of the yeast *Yarrowia lipolytica*. *Appl Environ Microbiol* **70**, 3918-3924.
158 Thierry, D. & Nicaud, J. M. Involvement of the G3P shuttle and beta-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab. Eng.* **13**, 482-491, doi:10.1016/j.ymben.2011.05.002 (2011).
159 Wang, H. J. J. *et al.* Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the n-alkane-assimilating yeast *Yarrowia lipolytica*. *Journal of Bacteriology* **181**, 5140-5148 (1999).
160 Blazeck, J. & Alper, H. Promoter Engineering: Recent Advances in Controlling Transcription at the Most Fundamental Level. *Biotechnology Journal*, doi:10.1002/biot.201200120 (2012).
161 Aggelis, G. & Komaitis, M. Enhancement of single cell oil production by *Yarrowia lipolytica* growing in the presence of *Teucrium polium* L. aqueous extract. *Biotechnology Letters* **21**, 747-749, doi:10.1023/a:1005591127592 (1999).
162 Papanikolaou, S. *et al.* Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by *Yarrowia lipolytica*. *Current Microbiology* **52**, 134-142, doi:10.1007/s00284-005-0223-7 (2006).
163 Makri, A., Fakas, S. & Aggelis, G. Metabolic activities of biotechnological interest in *Yarrowia lipolytica* grown on glycerol in repeated batch cultures. *Bioresource Technology* **101**, 2351-2358, doi:10.1016/j.biortech.2009.11.024 (2010).
164 Gabaldon, T. Peroxisome diversity and evolution. *Philosophical Transactions of the Royal Society B-Biological Sciences* **365**, 765-773, doi:10.1098/rstb.2009.0240 (2010).
165 Guttikonda, S. K. *et al.* Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. *BMC Plant Biol.* **10**.
166 Matoba, T., Hidaka, H., Kitamura, K., Kaizuma, N. & Kito, M. Contribution of hydroperoxide lyase activity to *n*-hexanal formation in soybean. *Journal of Agricultural and Food Chemistry* **33**, 856-858, doi:10.1021/jf00065a022 (1985).
167 Gardner, H. W., Weisleder, D. & Plattner, R. D. Hydroperoxide lyase and other hydroperoxide-metabolizing activity in tissues of soybean, *Glycine max*. *Plant Physiology* **97**, 1059-1072, doi:10.1104/pp.97.3.1059 (1991).

168 Matsui, K., Shibutani, M., Hase, T. & Kajiwara, T. Bell pepper fruit fatty acid hydroperoxide lyase is a cytochrome P450 (CYP74B). *Febs Letters* **394**, 21-24, doi:10.1016/0014-5793(96)00924-6 (1996).
169 Pitera, D. J., Newman, J. D. & Keasling, J. D. Optimizing a heterologous mevalonate pathway for the production of terpenoids in *Escherichia coli*. *Abstracts of the General Meeting of the American Society for Microbiology* **104**, 471 (2004).
170 Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K. & Uemura, H. DGA1(diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the snf2 disruptant of *Saccharomyces cerevisiae*. *Biochemical Journal* **408**, 61-68, doi:10.1042/bj20070449 (2007).
171 Curran, K. & Alper, H. Expanding the Chemical Palate of Cells by Combining Systems Biology and Metabolic Engineering. *Metab. Eng.*, doi:10.1016/j.ymben.2012.04.006 (2012).
172 Curran, K. A., Leavitt, J., Karim, A. & Alper, H. S. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. *Metab. Eng.* **15**, 55-66, doi:10.1016/j.ymben.2012.10.003 (2013).
173 Lee, J. W. *et al.* Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chemical Biology* **8**, 536-546, doi:10.1038/nchembio.970 (2012).
174 Zhang, F. Z., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* **30**, 354-360, doi:10.1038/nbt.2149 (2012).
175 Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86-89 (2008).
176 Li, Q., Du, W. & Liu, D. H. Perspectives of microbial oils for biodiesel production. *Applied Microbiology and Biotechnology* **80**, 749-756, doi:10.1007/s00253-008-1625-9 (2008).
177 Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-U182, doi:10.1038/nature08721 (2010).
178 Lu, X. F., Vora, H. & Khosla, C. Overproduction of free fatty acids in *E. coli*: Implications for biodiesel production. *Metab. Eng.* **10**, 333-339, doi:10.1016/j.ymben.2008.08.006 (2008).
179 Rodolfi, L. *et al.* Microalgae for Oil: Strain Selection, Induction of Lipid Synthesis and Outdoor Mass Cultivation in a Low-Cost Photobioreactor. *Biotechnology and Bioengineering* **102**, 100-112, doi:10.1002/bit.22033 (2009).
180 Alvarez, H. M., Mayer, F., Fabritius, D. & Steinbuchel, A. Formation of intracytoplasmic lipid inclusions by *Rhodococcus opacus* strain PD630. *Archives of Microbiology* **165**, 377-386, doi:10.1007/s002030050341 (1996).
181 Blazeck, J. *et al.* Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **97**, 3037-3052, doi:10.1007/s00253-012-4421-5 (2013).

182 Greenspan, P., Mayer, E. P. & Fowler, S. D. Nile red: a selective fluroscent stain for intracelluluar lipid droplets. *Journal of Cell Biology* **100**, 965-973, doi:10.1083/jcb.100.3.965 (1985).
183 Hammond, E. G., Johnson, L. A., Su, C., Wang, T. & White, P. J. in *Bailey's Industrial Oil and Fat Products* (ed F. Shahidi) Ch. 13, 577-653 (John Wiley & Sons, Inc., 2005).
184 Werpy, T. & Petersen, G. Top Value Added Chemicals from Biomass: Volume I—Results of Screening for Potential Candidates from Sugars and Synthesis Gas. (U.S Department of Energy, 2004).
185 Tsai, Y. C., Huang, M. C., Lin, S. F. & Su, Y. C. Method for the production of itaconic acid using *aspergillus terreus* solid state fermentation. United States patent (2000).
186 Tate, B. E. in *Kirk-Othmer Encyclopedia of Chemical Technology* (eds M. Grayson & E. Eckroth) 865–873 (1981).
187 Okabe, M., Lies, D., Kanamasa, S. & Park, E. Y. Biotechnological production of itaconic acid and its biosynthesis in *Aspergillus terreus*. *Applied Microbiology and Biotechnology* **84**, 597-606, doi:10.1007/s00253-009-2132-3 (2009).
188 Nuss, P. & Gardner, K. H. Attributional life cycle assessment (ALCA) of polyitaconic acid production from northeast US softwood biomass. *International Journal of Life Cycle Assessment* **18**, 603-612, doi:10.1007/s11367-012-0511-y (2013).
189 Itaconix, LLC. NIFA, New Hampshire (2009).
190 Kinoshita, K. Über die Production von Itaconsäure und Mannit durch einem neuen Schimmelpilz *Aspergillus itaconicus*. *Acta Phytochim* **5**, 271-287 (1932).
191 Li, A., Pfelzer, N., Zuijderwijk, R. & Punt, P. Enhanced itaconic acid production in *Aspergillus niger* using genetic modification and medium optimization. *Bmc Biotechnology* **12**, (2012).
192 Kautola, H., Rymowicz, W., Linko, Y. Y. & Linko, P. Itaconic acid production by immobilized *aspergillus terreus* with varied metal additions. *Applied Microbiology and Biotechnology* **35**, 154-158 (1991).
193 Yahiro, K., Takahama, T., Park, Y. S. & Okabe, M. Breeding of *Aspergillus terreus* Mutant TN-484 for Ataconic Acid Production with High Yield. *Journal of Fermentation and Bioengineering* **79**, 506-508, doi:10.1016/0922-338x(95)91272-7 (1995).
194 Park, Y. S., Itida, M., Ohta, N. & Okabe, M. Itaconic acid production using an air-lift bioreactor in repeated batch culture of *Aspergillus terreus*. *Journal of Fermentation and Bioengineering* **77**, 329-331, doi:10.1016/0922-338x(94)90245-3 (1994).
195 Kanamasa, S., Dwiarti, L., Okabe, M. & Park, E. Y. Cloning and functional characterization of the cis-aconitic acid decarboxylase (CAD) gene from *Aspergillus terreus*. *Applied Microbiology and Biotechnology* **80**, 223-229, doi:10.1007/s00253-008-1523-1 (2008).

196 Gruer, M. J., Artymiuk, P. J. & Guest, J. R. The aconitase family: Three structural variations on a common theme. *Trends Biochem.Sci.* **22**, 3-6, doi:10.1016/s0968-0004(96)10069-4 (1997).
197 Radakovits, R., Jinkerson, R. E., Darzins, A. & Posewitz, M. C. Genetic Engineering of Algae for Enhanced Biofuel Production. *Eukaryotic Cell* **9**, 486-501, doi:10.1128/ec.00364-09 (2010).
198 Tornoe, J., Kusk, P., Johansen, T. E. & Jensen, P. R. Generation of a synthetic mammalian promoter library by modification of sequences spacing transcription factor binding sites. *Gene* **297**, 21-32, doi:10.1016/s0378-1119(02)00878-8 (2002).
199 Juven-Gershon, T., Cheng, S. & Kadonaga, J. T. Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917-922, doi:10.1038/nmeth937 (2006).
200 Lanza, A. M., Dyess, T. J. & Alper, H. S. Using the Cre/lox system for targeted integration into the human genome: loxFAS-loxP pairing and delayed introduction of Cre DNA improve gene swapping efficiency. *Biotechnology Journal* **7**, 898-908, doi:10.1002/biot.201200034 (2012).
201 Courchesne, N. M. D., Parisien, A., Wang, B. & Lan, C. Q. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *Journal of Biotechnology* **141**, 31-41, doi:10.1016/j.jbiotec.2009.02.018 (2009).
202 Reed, B., Blazeck, J. & Alper, H. Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *Journal of Biotechnology* **158**, 75-79, doi:10.1016/j.jbiotec.2012.01.028 (2012).
203 Damude, H. G. H., Gillies, Peter John, Macool, Daniel Joseph, Picataggio, Stephen K., Pollak, Dana Walters M., Ragghianti, James John, Xue, Zhixiong, Yadav, Narendra S. , Zhang, Hongxiang, Zhu, Quinn Qun. High eicosapentaenoic acid producing strains of *Yarrowia lipolytica*. United States patent 20060115881 (2006).
204 Damude, H. *et al.* Docosahexaenoic acid producing strains of *Yarrowia lipolytica*. USA patent (2006).
205 Brat, D., Weber, C., Lorenzen, W., Bode, H. B. & Boles, E. Cytosolic re-localization and optimization of valine synthesis and catabolism enables increased isobutanol production with the yeast *Saccharomyces cerevisiae*. *Biotechnology for Biofuels* **5**, (2012).
206 Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nature Biotechnology* **31**, 335-+, doi:10.1038/nbt.2509 (2013).
207 Pitula, J. S. *et al.* Selective inhibition of the citrate-to-isocitrate reaction of cytosolic aconitase by phosphomimetic mutation of serine-711. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10907-10912, doi:10.1073/pnas.0404308101 (2004).
208 Puigbo, P., Bravo, I. G. & Garcia-Vallve, S. CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct* **3**, (2008).

209 Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* **28**, 292-292 (2000).
210 Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119-122, doi:0378111995000377 (1995).
211 Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast* **21**, 661-670, doi:Doi 10.1002/Yea.1130 (2004).
212 Kalnins, A., Otto, K., Ruther, U. & Muller-Hill, B. Sequence of the lacZ gene of Escherichia coli. *EMBO J* **2**, 593-597 (1983).
213 Kalnins, A., Otto, K., Ruther, U. & Mullerhill, B. Sequence of the lacZ gene of *Escherichia coli*. *Embo Journal* **2**, 593-597 (1983).
214 Mumberg, D., Muller, R. & Funk, M. Regulatable Promoters of *Saccharomyces cerevisiae* - Comparision of Transcriptional Activity and Their Use for Heterologous Expression. *Nucleic Acids Research* **22**, 5767-5768, doi:10.1093/nar/22.25.5767 (1994).
215 Tabuchi, T., Tahara, Y., Tanaka, M. & Yanagiuti, S. . Preliminary experiments on the mechanism of citrate fermentation in yeasts. *Nippon Nogeikagaku Kaishi* **47**, 617-622 (1973).
216 Goldstein, A. L. & McCusker, J. H. Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast* **15**, 1541-1553, doi:10.1002/(sici)1097-0061(199910)15:14<1541::aid-yea476>3.0.co;2-k (1999).
217 Guldener, U., Heck, S., Fiedler, T., Beinhauer, J. & Hegemann, J. H. A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Research* **24**, 2519-2524, doi:10.1093/nar/24.13.2519 (1996).
218 Folch, J., Lees, M. & Stanley, G. H. S. A simple method for the isolation and purification of total lipids from animal tissues. *Journal of Biological Chemistry* **226**, 497-509 (1957).
219 Schneiter, R. & Daum, G. Extraction of yeast lipids. *Methods in Molecular Biology* **313**, 41-45 (2006).
220 Paik, M. J., Kim, H., Lee, J., Brand, J. & Kim, Y. R. Separation of triacylglycerols and free fatty acids in microalgal lipids by solid-phase extraction for separate fatty acid profiling analysis by gas chromatography. *Journal of Chromatography A* **1216**, 5917-5923, doi:10.1016/j.chroma.2009.06.051 (2009).
221 Liu, B. & Zhao, Z. Biodiesel production by direct methanolysis of oleaginous microbial biomass. *Journal of Chemical Technology and Biotechnology* **82**, 775-780, doi:10.1002/jctb.1744 (2007).
222 Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Research* **37**, Article No.: e16, doi::10.1093/nar/gkn991 (2009).