12-2017

Expanding the Genetic Toolbox to Improve Metabolic Engineering in the Industrial Oleaginous Yeast, *Yarrowia lipolytica*

Murtaza Shabbir Hussain
Clemson University

EXPANDING THE GENETIC TOOLBOX TO IMPROVE METABOLIC ENGINEERING IN THE INDUSTRIAL OLEAGINOUS YEAST, *YARROWIA LIPOLYTICA*

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Chemical & Biomolecular Engineering

---

by
Murtaza Shabbir Hussain
December 2017

Accepted by:
Mark A. Blenner, Committee Chair
Sapna Sarupria
Christopher Kitchens
Yi Zheng

ABSTRACT

The oleaginous yeast, *Yarrowia lipolytica*, is becoming a popular host for industrial biotechnology because of its ability to grow on non-conventional feedstocks and naturally accumulate significant amounts of lipids. With new genome editing technologies, engineering novel pathways to produce lipid-derived oleochemicals has become easier. The goal, however, is to expand the genetic toolbox to improve the efficiency of metabolic engineering such that production capacities could expand from proof-of-concept shake flasks to an industrial scale.

Building efficient metabolic circuits require controlling strength and timing of several enzymes in a metabolic pathway. One method to do this is through transcription – using suitable promoters to control the expression of genes that code for enzymes. Native promoters have limited application because of complex regulation and non-tunable expression. Engineering hybrid promoters alleviate these issues to obtain predictable and tunable gene expression. In *Y. lipolytica*, how to design these promoters is not fully understood, resulting in only a handful of engineered promoters to date.

In this work, we aim to develop tools for gene expression by investigating promoter architecture and designing tunable systems. In addition to Upstream Activating Sequences (UAS), tuning promoter strength can be achieved by varying sequence in the core promoter, TATA motif, and adjacent proximal sequences.

UASs can modulate transcription strength and inducibility, enabling controlled timing of expression. A promoter of the acyl-CoA oxidase 2 (POX2) from the β-oxidation pathway was truncated heuristically to identify oleic acid (OA) UAS

sequences. By fusing tandem repeats of the OA UAS elements, tunable yet inducible fatty acid hybrid promoters were engineered.

The current approaches to identify novel UAS elements in *Y. lipolytica* are laborious. Therefore, we investigated DNA accessibility through nucleosome positioning to determine if a relationship between POX2 UASs and DNA accessibility can be inferred. The goal is to eventually apply this approach develop newer hybrid promoters efficiently.

Finally, the hybrid fatty acid inducible promoter we developed was used to rationally engineering a *Y. lipolytica* strain capable of producing high amounts of free fatty acids. By localizing the fatty acyl / fatty aldehyde reductase in the peroxisome, we compartmentalized fatty alcohol production. This strategy led to upwards of 500 mg/L of fatty alcohols produced. It is a promising route to eventually make short to medium chain fatty alcohols in *Y. lipolytica* by utilizing the native β-oxidation machinery.

DEDICATION

I would like to dedicate this manuscript to my family. My father, Shabbir Hussain Hassanally, who made a lot sacrifices to make sure I received a good education. My mother, Romayn Sonia Klyn, for her love and attention to me even when I was several miles away. My older brother, Imran Shabbir Hussain, and younger sister, Mariyam Shabbir Hussain for being my best friends and for letting me be who I am and love me regardless. My older brother has always been there to remind me that I was smart and fought by my side when I felt personal pressure. My younger sister is an emotional rock to me. She has always looked up to me and I hope I have not disappointed her.

Most importantly, I want to dedicate this thesis and my Ph.D. to my dear aunt, Rosita Klyn, who passed away when I began graduate school at Clemson University. My whole life, she has been like another mother to me, cared for me when I was hurting, understanding my struggles when I felt I had to hide from the world and most importantly, helping me embrace who I am. To all our taxi and bus trips together, for believing in me and pushing me to pursue my dreams and for being so proud of me when I was accepted into graduate school. Losing her remains the hardest challenge. This thesis will be a reminder to never lose sight of the lessons she taught me and the legacy she left behind.

ACKNOWLEDGMENTS

First, I would like to sincerely acknowledge my Ph.D. advisor and mentor, Dr. Mark A. Blenner for his guidance, and facilitation of my research in his lab. His patience and dedication to my development as a researcher has enabled me to accomplish a significant body of novel work, contributing to the field of microbial genetic and metabolic engineering. Through trying times from my end, he continued to teach me important life lessons such as to heed criticism and be strong in the face of failure. His advice and patient guidance has made me the confident scientist I am today. In addition to research, he has always been there for professional development or life development. I am lucky to have had him as my mentor for five years.

For advising me whenever he had the opportunity and for providing very thoughtful guidance through most of my Ph.D. career, I thank Dr. Ian Wheeldon. His insight has been invaluable to my development. I will try to not overthink problems.

For showing me that sometimes success is an uphill battle and that if you persist long enough, it pays off, I thank Dr. Sapna Sarupria. Speaking to her ignites interesting life thoughts.

I thank my committee members, Dr. Chris Kitchens and Dr. Yi Zheng for taking the time to provide guidance on my research plan and advocating for my success.

Last but not least, my colleagues, Allison Yaguchi, Kyle Reed (former graduate student), Dr. Gabriel Rodriguez and Dr. Difeng Gao, Lauren Gambill, Spencer Smith, and Cory Schwartz who have helped me tremendously and whom I have had great research discussions and collaborations with, I thank you sincerely.

TABLE OF CONTENTS

Table of Contents (Continued) Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

Today, most industrial chemicals used to produce materials, plastics, surfactants, and solvents are derived from the non-renewable petroleum feedstock which leaves environmental problems and threats for human beings. As a result, the paradigm of research has shifted towards exploiting methods for production of sustainable and green products. Bio-based technology can be a good alternative to address this problem. While petrochemical based chemical production dominates much of the commodity chemical market, biochemical production has already shown promise for chemical production at commercial scale [1-3] requiring lower capital investments, therefore, providing a competitive edge. Furthermore, several commodity-scale chemicals have also been produced using biomass at costs lower than petrochemical processes [4, 5].

Industrial biotechnology is a rapidly expanding industry built upon a biological foundry to solve global challenges, offering new potential to meet demands for chemicals, fuel, and food with significantly reduced impact on the environment. The problems to be tackled can be broadly categorized in two sets, not mutually exclusive of one another. To minimize fossil-fuel based dependence for chemical production, a core challenge is to be able to tap into a biological, yet efficient means to utilize naturally abundant feedstocks such as plant biomass and sunlight as a fuel to drive the biochemical process. The second challenge is optimizing biological platforms for the production of selective chemicals in a cost-effective, economically feasible manner. To date, both concepts have been explored using plants, algae, and microbes.

1.1. Plants for Sustainable Chemical Production

Plants and algae are unique biological platforms because of their ability to harness energy from sunlight to drive photosynthesis making food in the form of carbohydrates to sustain physiological processes. Phototrophic systems, in theory, are a robust economically feasible process, however, there are limitations to harnessing solar energy. The theoretical efficiency is limited to the range of wavelengths applicable to photosynthesis and the quantum requirements of the photosynthetic process. The photosynthetically active radiation from solar energy is in the range of 400 and 700 nm, accounting for about 45% of the suns light energy. In combination with the quantum requirements for CO_2 fixation in photosynthesis, the theoretical maximum efficiency is around 11% of solar energy [6]. Practically, the magnitude of photosynthetic efficiency is further decreased due to reflection of the sun's wavelengths, respiration requirements for photosynthesis and the lack of optimal solar radiation. This drops efficiency to between 3% and 6% of the total solar energy that is harnessed. For crops, it should be noted that if only agriculturally relevant products such as seeds, fruits, and tubulars are considered rather than total biomass, this amounts to lower photosynthetic efficiencies.

In addition to growing crops for primary metabolites such as carbohydrates, fats, oils, and proteins, plants also naturally produce a variety of secondary metabolites which are industrially relevant and important for human health. Flavonoids, terpenoids, carotenoids, and phenolics are some of the compounds that encompass the broad range of chemicals that are produced in plants [7]. In recent decades, advances in genetic tools have improved the ability to engineer transgenic crops for not only maximizing yields of primary metabolites but increasing the landscape for production of valuable pharmaceuticals [8-10].

While promising, metabolic engineering of plants has its drawbacks. If sustainable chemical production to replace current methods is the goal, then exploiting plants as a production platform requires consideration of the long growth times and small production capacity of secondary metabolites. Chemical production can be maximized by growing more genetically engineered crops but the limitation is arable landmass creating competition with crops grown for food. From the approximately 2.3 billion acres of land in the United States, around 349 million acres are utilized for growing crops [11]. Increasing growth capacity of select crops for biochemical production would only increase competition for arable land against crops grown for food or be in direct competition with land mass used for urbanization.

Although plants may not be the best solution for biochemical production, there is a lot that can be learned about the biosynthetic pathways to produce a diversity of chemicals. These pathways can be applied to more feasible biological platforms such as microbes via heterologous expression of enzymes to engineer novel production pathways. Table 1.1 summarizes some of the more successful application of plant biosynthetic pathways in microbial systems for sustainable chemical production. *Escherichia coli* and *Saccharomyces cerevisiae* are the predominant conventional microbes for heterologous expression of plant pathways because of the wide array of genetic tools that facilitate engineering efforts. If transgenic crops were to become feasible for biochemical production in the future, there is still the backlash from the public perception about using genetically modified organisms (GMOs), particularly applying to transgenic crops that need to be addressed and resolved [12, 13].

1.2. Algae for Sustainable Chemical Production

Algae are phototrophic organisms requiring solar energy and CO_2, to grow and thrive in nutrient-depleted conditions. Unlike common crops such as wheat and barley, algae grow faster and occupy less space making it a better system for the development of sustainable, biorenewable production practices. The fact that some algae can gown in salt water conditions and tolerate wide pH conditions makes it more advantageous than agriculture practices that require large quantities of fresh water. Furthermore, algae can be grown more densely than plants reducing arable land use.

One of the major applications of algae in industrial biotechnology is biofuel production [14]. However, algae have also been explored extensively to produce high value-added chemicals. The world market of products from macroalgae has been estimated to be close to USD 6 billion per year while the retail price of products from microalgae was estimated at USD 5 to 6.5 billion [15, 16]. Hydrocolloids such as agar, alginate, and carrageenans are produced at industrial scale using algae. These chemicals are used as gelation and thickening agents in different food, pharmaceutical and biotechnological applications with a current estimated global value of over USD 1.1 billion [17]. Hydrocolloids are extracted from the cell wall of red seaweed algae at industrial scale since algae grow robustly and accumulate significant biomass.

Other valuable products that have been produced from algae at commercial scale include proteins, polyunsaturated fatty acids, the keto-carotenoid astaxanthin, and food dyes such as phycoerythrin and phycocyanin [18-21]. The ability to produce these chemicals economically is a result of their natural production capabilities and to a lesser effect, metabolic engineering efforts. The development of better genetic tools to facilitate efficient metabolic engineering could make algae a more promising host for sustainable chemical production. Currently, some of the tools available are transformation protocols, stable expression of transgenes [22], targeted microRNA mediated gene knockdown and silencing [23] and more recently, efficient CRISPR-Cas9 mediated genome engineering [24]. A majority of these genetic tools have been developed in the green algae, *Chlamydomonas reinhardtii*; however, they are being rapidly translated and developed for diatoms and other algal species that are of industrial importance.

One of the criticisms of engineering metabolic pathways in algae is the difficulty in scale-up that causes a loss in productivity. Although technically and economically more viable than crops, there is still a lot of work that needs to be done to make chemical production feasible and reduce the cost of downstream processes such as chemical extraction and separation. The future of microalgae scale-up requires optimization and design of advanced bioreactors and developing low-cost technologies for biomass harvesting, drying, and oil extraction. Furthermore, since algae show a strong dependence on a variety of environmental stress conditions, advancing the genetic toolbox to elicit more controlled expression under different stress signals is required to enhance metabolic engineering capabilities and make it a sustainable platform for chemical production. In the meantime, like plants, there are important biosynthetic pathways from algae that can be translated into microbes to produce a plethora of industrially relevant chemicals. Table 1.1 highlights some of these biosynthetic pathways engineered into microbes.

Table 1.1. Engineering biosynthetic pathways in conventional microbes for intricate biochemical synthesis using plant and algal enzymes.

Host	Enzymes for pathway biosynthesis	Product
E. coli	4-coumarate: CoA ligase: ***Lithospermum erythrorhizon* (gromwell)** stilbene synthase: ***Arachis hypogaea* (peanut)**	Stilbene synthesis [25]
E. coli	4-coumarate: CoA ligase: chalcone synthase: ***Glycyrrhiza echinata* (licorice)** chalcone isomerase: ***Pueraria lobata* (kudzu)** flavone synthase: ***Petroselinum crispum* (parsley)**	Flavone synthesis [25]
E.coli	Geranylgeranyl pyrophosphate synthase: Taxadiene synthase: taxadiene 5a-hydroxylase: ***Taxus brevifolia* (Pacific yew tree)**	Advancements in taxol synthesis [26]
E. coli	phytoene desaturase: ζ-carotene desaturase: ***Arabidopsis thaliana* (Thale cress)**	Pro-lycopene production [27]
S. cerevisiae	Δ4-desaturase: ***Euglena gracilis* (micro algae)** Δ5-desaturase: ***Phaeodactylum tricornutum* (diatom)** Δ5-elongase: ***Thalassiosira pseudonana (marine diatom)*** Δ6-elongase: ***Ostreococcus tauri (marine algae)***	Docosahexaenoic acid (22:ω3) production from C18:3 [28]
E. coli	norcoclaurine synthase: norcoclaurine 6-O-methyltransferase: coclaurine-N-methyltransferase: 3-hydroxy-Nmethylcoclaurine-4-O- methyltransferase: ***Coptis japonica (Japanese Goldthread)***	Reticuline production from dopamine
S. cerevisiae	Cytochrome P450 (CYP80G2): Corytuberine N-methyltransferase: ***Coptis japonica (Japanese Goldthread)***	benzylisoquinoline alkaloid biosynthesis from reticuline [29]

1.3. Exploiting Microbes for Biochemical Production

Microorganisms have been used in industrial biotechnology for decades. Bacteria, yeast, and fungi are predominant platforms for this purpose. A few of the earliest discoveries of microbial potential to making chemicals date back to the production of yogurt and cheese using probiotic microbe genus *Lactobacillus* [30], and production of the antibiotic, penicillin from the ascomycetes fungal genus, *Penicillium* [31]. Microbes have since been engineered to make numerous industrially relevant chemicals. In microbial metabolic engineering, *E. coli* and *S. cerevisiae* are used extensively as cellular factories to produce a diversity of commodity and specialty chemicals that can be produced in a renewable, eco-friendly manner. The potential to engineer metabolic pathways in these microbes is driven by the genetic toolbox that enables engineering, heterologous expression from well-characterized promoters with tunable strength and inducibility, and their natural propensity to grow and thrive [32-36].

1.3.1. Metabolic Engineering of Conventional Microbes

Advancements in metabolic engineering are driven by the ability to efficiently engineer microbes by developing biosynthetic pathways using nature's remarkable catalysts called enzymes. Having robust gene expression platforms and being able to fine-tune the expression of enzymes in a biosynthetic pathway is what contributes to improving the efficiency of metabolic pathways. While developing these novel pathways, a microbe's native regulatory mechanisms needs to be considered. The native regulatory mechanisms enable cells to maintain homeostasis that can occur on either a transcription, translation or metabolite level. While identification of these bottlenecks is crucial, having a library of genetic tools with various expression capabilities can alleviate issues at nodes to elegantly engineer foreign pathways with high efficiency.

As a result, *E. coli* and *S. cerevisiae* have both been established as safe, conventional microbes to produce industrially relevant chemicals. Butanol, for example, is a chemical feedstock that has gained much attention as a next-generation biofuel replacement to ethanol due to higher energy content and lower volatility. Previous attempts to produce improve butanol using *Clostridium* have been unsuccessful due to difficulty in genetically manipulating clostridial strains [37]. However, in recent years, it was demonstrated that 1-butanol can be produced in *E. coli* at titers as high as 30 g/L and 70% to 88% of its theoretical maximum via anaerobic fermentation [38]. Similar metabolic engineering feats have been accomplished with the production of 1,3-propanediol in *E. coli* attaining titers of 135 g/L and productivity of 3.5 g/L/h in a 10 L fed-batch reactor [1].

Saccharomyces cerevisiae has also been explored as a robust metabolic engineering platform to produce several chemicals relevant to industry [39, 40]. The development of an extensive genetic toolbox has enabled efficient production of natural compounds by transferring product-specific enzymes or entire metabolic pathways from other biological systems that either grow slow or are genetically intractable [41, 42]. Even though quantification metrics such as yields, titers and productivities may not always be up to par for industrial production, this microbe is best at demonstrating the ability to produce value-added specialty chemicals. Terpenoids, alkaloids, flavonoids and non-ribosomal peptides synthesis are some examples where biosynthetic pathways have been successfully constructed in *S. cerevisiae*.

Terpenoids are the largest class of naturally occurring molecules. The production of terpenoids has already been patented in the plant, *Arabidopsis thaliana* [43]. Although structurally diverse, terpenoids can be synthesized from two isoprene precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl diphosphate (DMAPP). The modification of IPP and its derivates to produce more complex terpentine molecules has been demonstrated in *S. cerevisiae*. One such molecule is artemisinic acid, a precursor to the antimalarial drug, artemisinin [44]. Polyketides are yet another molecule where structural complexity precludes chemical synthesis as an economically feasible route for large-scale production. *S. cerevisiae* has proven to be a promising host for pathway engineering complex polyketides because it already has already been leveraged to produce large amounts of fungal polyketides such as 6-methylsalicylic acid [45].

1.3.2 Genetic Regulation for Conventional Microbes

The development of genetic tools to fine-tune strength and timing of expression is what enables much of the metabolic engineering accomplishments in a microbe. The expression of an enzyme in a cell is determined by two biological processes, transcription and translation. Transcription is the genetic level control of expression that deals with controlling how much mRNA is produced from a gene. The primary step of transcription occurs at the promoter where specific proteins bind to upstream of the gene of interest to regulate the amount of the gene that is transcribed by RNA polymerase II. Therefore, at the base of enzyme expression is transcriptional regulation via promoters.

Promoter regulation is the most commonly studied approach to regulating gene expression. In both, *E. coli* and *S. cerevisiae*, a large library of promoters have been identified, studied and engineered for new properties [46-49]. Since eukaryotic gene regulation is the basis of this dissertation, a perspective on *S. cerevisiae* promoters and promoter engineering will be summarized. Native promoters are extensively well characterized in the microbe from constitutive to inducible [50]. Table 1.2 summarizes some of the commonly used native promoters in *S. cerevisiae*.

Table 1.2. Commonly used promoters used in metabolic engineering of *S. cerevisiae* [50-52]. ***** represents a subset of promoters from genes that are involved in global regulatory processes for cell survival. Other similar promoters can be found from genes involved in the expression of ribosomal proteins and chaperone proteins in *S. cerevisiae*.

Constitutive Native Promoters		
Promoter	Enzyme	Function
P_{PGK1}	3-phosphoglycerate kinase	1,3-bisphosphoglycerate to glycerate 3-phosphate
P_{TDH3}	Glyceraldehyde- 3-phosphate dehydrogenase (GAPDH)	Glyceraldehyde 3-phosphate to D-glycerate 1,3-bisphosphate
P_{TPI1}	Triosephosphate isomerase	Dihydroxyacetone phosphate to D-glyceraldehyde 3-phosphate
P_{ENO2}	Phospho-pyruvate hydratase	2-phosphoglycerate to phosphoenolpyruvate (PEP)
P_{ADH1}	Alcohol dehydrogenase	Alcohol to ketones
*P_{TEF1}	Translational elongation factor EF-1 alpha (TEF1-α)	Delivery of aminoacyl-tRNA to ribosome
*P_{TEF2}	Translational elongation factor EF-2 alpha (TEF2-α)	Delivery of aminoacyl-tRNA to ribosome
Inducible Native Promoters		
$P_{GAL1/GAL10}$	Galactokinase/UDP-glucose-4-epimerase	UDP Galactose to UDP-Glucose
P_{CUP1}	Copper thionein (Metallothionein)	Chelates copper at high concentration
P_{HXT7}	High-affinity hexose transporter	Active at low glucose concentration
P_{ADH2}	Alcohol dehydrogenase 2	Repressed in glucose
P_{PHO5}	Repressible acid phosphatase	Active under low inorganic phosphate

1.3.3. Promoter Engineering – A Page from the Conventional Yeast

Predictable expression from promoters is necessary for rational design and optimization of microbial cell factories; however, the use of native promoters in this context can be challenging because most promoters exhibit complex expression patterns. The complexity is a result of having spatially distributed regulatory motifs on a single promoter that recruit transcription factors to regulate transcription in response to changes to the environment (pH, temperature), nutrients (carbon, nitrogen) or cell physiology (early vs. late phase) etc. The *PIS1* gene in *S. cerevisiae* is an essential gene for *de novo* synthesis of the phospholipid, phosphatidylinositol and has been reported to have differential responses to fermentable versus non-fermentable carbon sources [53]. In addition, the *PIS1* promoter has regulatory regions for transcriptional factor *Sc*ROX1p, that represses promoters under hypoxic conditions [54]. Promoters form Sc*ACC1*, and genes involved in oxidative stress response have also shown similar complexities with multiple regulatory motifs on the promoter eliciting complex regulation to carbon conditions [55, 56].

The complexity of native regulatory systems, therefore, motivates the need to develop better promoters with predictable and defined expression patterns that can be then used to engineer more optimized metabolic processes. Engineering promoters in *S. cerevisiae* serve as the basis for understanding how yeast promoters function. Promoter engineering is an umbrella term used to describe many facets of developing new promoters.

One method is using error-prone PCR to create a library of random mutations on the native promoter and then screen for promoter function using a reporter gene (Figure 1.1). This has been demonstrated with P_{TEF1} (Table 1.2). Two hundred promoter mutants were screened that varied in expression strength from very weak to promoters two-fold stronger than the native promoter [57]. This approach to promoter engineering is efficient when high transformation efficiencies are attainable enabling a larger, more diverse library of mutants.

The other approach to promoter engineering in *S. cerevisiae* is based off rational design. By fusing modular elements, the core promoter and Upstream Activating Sequences (UASs), a library of "hybrid" promoters can be engineered (Figure 1.1). Variable levels of promoter inducibility and strength are achieved by changing the type and number of UAS sequences. In *S. cerevisiae* the authors showed that strongest constitutive native promoter, P_{TDH3} (Table 1.2), was made almost three-fold stronger by fusing disparate constitutive UAS elements from other promoters upstream of the native P_{TDH3} [58]. Galactose-inducible hybrid promoters spanning a fifty-fold dynamic range in galactose was demonstrated by fusing tandem *Sc*Gal4p binding upstream of a core promoter [58]. Finally, the catabolite repression of P_{GAL1} was alleviated by placing constitutive UAS sequences upstream of the inducible promoter [58].

Hybrid promoters are effective tools but the challenge lies in identifying UASs in native promoters. Determining UASs in yeast has traditionally relied on truncations of promoters fused to a reporter gene (Table 1.2). The loss of transcription activity is then correlated to a potential UAS element in the native promoter [59-61]. However, with today's high throughput technologies such as Chip-Seq and large transcription factor database for *S.* cerevisiae, the process of identifying transcription factor binding sites (TFBs) surpasses single promoter analysis to scanning all variations of the binding motifs in the genome [62].

One of the more interesting developments of *S. cerevisiae* hybrid promoter engineering is next-level regulation such as chimeric systems, fusing yeast UASs with bacterial operons to engineer dynamic regulation (Figure 1.1). Tet repression in a galactose-inducible promoter has been tested to investigate how number and positioning of tet repressor sites in the core promoter of a galactose-inducible promoter can affect gene expression [63]. Similarly, an *E. coli* FadR operator placed in the core promoter region of a yeast promoter fused to different UASs (inducible and constitutive) shows how dynamic promoter regulation can be used to only turn on metabolic pathways once there is sufficient buildup of the fatty acid precursor, relieving FadR repression to turn on the yeast promoter [64]. Studying these types of hybrid promoter systems in *S. cerevisiae* provides novel insight to building efficient metabolic circuits in eukaryotes and would better inform strategies for genetic tool development in newer, non-conventional microbes.

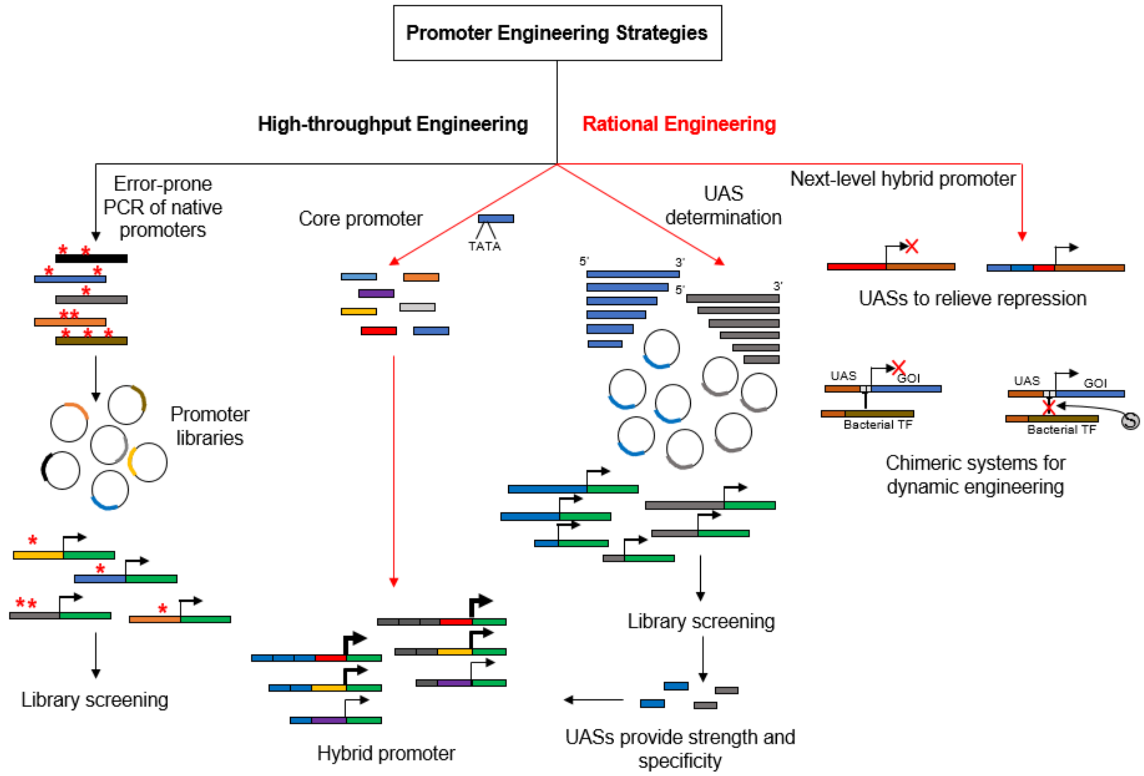


Figure 1.1. Schematic of three promoter engineering strategies used in *S. cerevisiae*. The heuristic approach is a more fundamental strategy for identifying promoter regions called Upstream Activating Sequences (UASs). Results from the first two methods can then be applied to experiments aimed at (1) determining specific transcription factors and binding motifs (2) developing tunable and dynamic hybrid promoters for metabolic engineering.

1.4. Non-Conventional Microbes in Biotechnology

In the past two decades, microbial engineering has mostly relied on conventional microbes for synthetic biology because of its genetic tractability and ease of genome editing. The development of bioinformatic tools during the same time enabled engineers to screen for suitable enzyme candidates to create pathways for target products. But, as biomolecular engineers, our goal is to develop a sustainable chemical engineering process within a biological context. In doing so, we are interested in two general parameters, successful transition of an engineered strain from the benchtop to a bioreactor and meeting key performance indicators, yield (mass of product / mass of substrate), titer (mass of product / reaction volume), and productivity (mass of product / reaction volume / time). Reaching such metrics for commercial feasibility can be challenging because synthetic circuits via heterologous pathways are not as reliable and robust as native pathways due to the microbe's innate regulation. This could lead to additional issues such as co-factor imbalances and allosteric inhibition that would need to be teased out using high throughput proteomics and metabolomics [65, 66].

Exploiting newer non-conventional microbes with capabilities to innately produce desired products has become increasingly popular in industrial biotechnology. This transition is greatly facilitated by tremendous advancements in synthetic biology over the past few years. Now, large-scale genome sequencing has become relatively inexpensive and coupled with the library of bioinformatic tools, developing genome-scale metabolic models for novel microbes has become more feasible and reliable. This has resulted in the exploration of non-model microbes to produce a wide variety of industrially relevant chemicals (Table 1.3). With little engineering efforts, sustainable production capacities can be achieved. Non-conventional microbes are also able to grow on less-traditional feedstocks to produce value-added products. This ability is very beneficial for not only developing economically feasible biochemical processes but also contributing to a cleaner environment by feeding on substrates found in waste streams. Amongst non-conventional yeasts, *Yarrowia lipolytica* leads the efforts to demonstrate the importance of exploring novel microbes for industrial biotechnology.

Table 1.3. Highlight of non-conventional microbes used for production of a wide range of industrially relevant products utilizing mainly non-traditional feedstocks

Microbe	Feed	Product	Titer (g/L)	Ref.
Aspergillus niger	Corn stover	Citric acid	100	[67]
Rhodosporidium toruloides	Sucrose	Fatty alcohols	8	[68]
Kluyverimyces marxianus	Wheat Straw	Ethanol	36	[69]
Trichosporon oleaginosus	Resorcinol	Lipids	1.64	[70]
Yarrowia lipolytica	Glycerol	Erythritol	220	[71]
Yarrowia lipolytica	Xylose	Citric acid	80	[72]
Yarrowia lipolytica	Glycerol	Citric acid	32	[73]
Xanthomonas campestris	Cassava starch	Xanthun gum	17	[74]

1.5. *Yarrowia lipolytica:* An Industrial Front Runner for Non-Conventional Microbes

The non-conventional oleaginous yeast, *Y. lipolytica*, has been used in industry for the past 60 years. Its broad applications include single cell protein production, citric acid production, and cell biomass as animal feed [75, 76]. Advances in synthetic biology and metabolic engineering have increased the overall utility of microorganisms by enabling custom built genetic engineering tools allowing for the manipulation of metabolism to produce valuable chemicals. The oleaginous property of *Y. lipolytica* allows this yeast to naturally accumulate lipids greater than 20% of dry cell weight [75]. This trait along with continued improvements in genetic engineering tools has led to increased interest in engineering this host to produce lipid-based products.

Over the years, significant effort has been made to understand the genetics of *Y. lipolytica* and to develop novel expression systems. Transformation protocols, basic expression cassettes, and gene deletion tools have been established for several years [77, 78]. More recently, episomal vectors, high expression synthetic promoters, and CRISPR-

Cas9 genome editing have been developed for use in *Y. lipolytica* [79-83]. This has resulted in accelerated metabolic engineering efforts in *Y. lipolytica*. In addition to high engineered lipid production capacities to reach titers close to 100 g/L [84], *Y. lipolytica* has been engineered to produce specific high-value products such as omega-3 fatty acids [85], dicarboxylic acids [86], polyhydroxyalkanoates (PHA) [87], itaconic acid [88], free fatty acids [89, 90], alkanes [90], esters [91], and alcohols [90, 92, 93].

Efforts have also been made to engineer *Y. lipolytica* for utilization of alternative sugars such as xylose, galactose, and starch [72, 94-100]. Alternative sugar substrates have become increasingly abundant due to advances in lignocellulose degradation [91]. The native xylose pathway of *Y. lipolytica* has been recently elucidated, which led to several studies achieving robust xylose utilization [72, 94, 95, 97].

1.5.1. Genetic Tool Development in Y. lipolytica

Efficient transformation protocols in tandem with access to a fully annotated and sequenced genomes of *Y. lipolytica* strains have greatly facilitated the development of genetic engineering tools over the past three decades [79, 101, 102]. These genetic engineering tools include but are not limited to the creation of hybrid, carbon responsive, and inducible promoter systems alongside quick and efficient genome editing techniques.

1.5.1.1. Transformation Methods for Y. lipolytica

Original transformation methods developed for *Y. lipolytica* utilized a PEG-protoplast transformation [103]. Advances to this protocol soon followed with a lithium acetate (LiAc) protocol adapted from *Saccharomyces cerevisiae* to obtain site-directed integrative transformation efficiencies of up to 1x10^4 transformants/ug of linearized DNA [104]. Modifications to the LiAc transformation and development of an electroporation protocol superseded the above technique, allowing for highly efficient replicative transformation [105]. To date, the LiAc protocol is the more commonly practiced method for transformation of plasmids and electroporation is more efficient for transforming linearized integrative vectors [77, 106]. Recently, improvements to the overall transformation efficiency of linearized integrative DNA has been accomplished by using a combinatorial approach of LiAc and electroporation, with yields reaching 2 x 10^4 transformants / μg of linearized DNA [107].

Although considerable work has been done to build transformation strategies in this oleaginous yeast, efficiencies remain about 2 orders of magnitude lower than the conventional yeast, *S. cerevisiae*. Despite this, it has been demonstrated that given the current efficiencies, *Y. lipolytica* could still be used as a suitable host for molecular evolution of proteins using both rational and directed evolution strategies [108, 109].

1.5.1.2. Native Promoters and Terminators for Regulating Gene Expression

Studies of metabolic pathways in *Y. lipolytica* have revealed several highly expressed native genes that are induced and/or repressed by different carbon sources and physiological conditions [110]. The best characterized native promoters are from the genes of Glycerol-3-phosphate dehydrogenase (*G3P*), isocitrate lyase (*ICL1*), 3-oxo-acyl-CoA thiolase (*POT1*), aceto-acetyl-CoA thiolase (*PAT1*), the acyl-CoA oxidases (*POX1*-

POX6), extracellular lipase 2 (*LIP2*), and alkaline extracellular protease (*XPR2*). A list of the native promoters commonly used and their responsiveness to different carbon conditions are summarized in Table 1.4.

In the scope of gene regulation, terminators are another important regulatory point since the sequences dictate completion of transcription and determine the half-life of synthesized mRNA. The terminator from the gene encoding cytochrome c oxidase from *S. cerevisiae* (ScCYC1t) is a commonly used terminator sequence in *Y. lipolytica* alongside native *Y. lipolytica* terminators [80-83, 95, 111, 112]. The use of short synthetic terminators to improve gene expression has also been explored in *Y. lipolytica* recently [112]. The synthetic terminator designs were first constructed in *S. cerevisiae* and translated into *Y. lipolytica* to attain a 2-fold improvement relative to the native TEF terminator of *Y. lipolytica.* Terminators and promoters can interact to form loops that regulate gene expression [113] however, this phenomenon has not been explored in *Y. lipolytica*.

Table 1.4. Commonly used native promoters for constitutive and native inducible expression in *Y. lipolytica* [110, 114-118]

Substrate(s)	Induced Promoter Systems	Repressed Promoter Systems
Glycerol	Glycerol-3-phosphate dehydrogenase (P_{G3P}) Translation Elongation Factor-1α (P_{TEF1})	Acyl-CoA-oxidase 2 (P_{POX2}) 3-oxo-acyl-CoA thiolase (P_{POT1})
Glucose	Glycerol-3-phosphate dehydrogenase (P_{G3P})	Acyl-CoA-oxidase 2 (P_{POX2}) 3-oxo-acyl-CoA thiolase (P_{POT1})
Oleic acid Ricinoleic acid methyl ester n-Decane	Acyl-CoA-oxidase 2 (P_{POX2}) Isocitrate lyase 1 (P_{ICL1}) 3-oxo-acyl-CoA thiolase (P_{POT1}) Translation Elongation Factor-1α (P_{TEF1})	Glycerol-3-phosphate dehydrogenase (P_{G3P})
Ethanol Acetone	Cytochrome P450 (P_{ICL1})	Acyl-CoA-oxidase 2 (P_{POX2}) Isocitrate lyase 1 (P_{ICL1}) Glycerol-3-phosphate dehydrogenase (P_{G3P})

1.5.1.3. Hybrid Promoters for Regulating Gene Expression

Native promoters in *Y. lipolytica* have been used to control metabolic pathways, however, the regulation associated with these promoters can be complex and exhibit unpredictable behavior. Furthermore, the need to construct both strong and varied promoter strengths for metabolic and pathway engineering is the desired outcome that cannot be met exclusively by native promoters. Over the past decade, there have been several lines of work aimed at developing and characterizing the next generation hybrid promoters that confer very high and tunable expression [119-121]; however, hybrid promoters that have programmed regulatory behavior have not been reported.

Hybrid promoters are created in *Y. lipolytica* by deconstructing the native promoters to identify upstream activating sequences (UAS) that confer transcriptional activation and creating repeats of these sequences in tandem to accomplish high levels of transcription. One commonly used UAS in *Y. lipolytica* is the UAS1B that was first isolated from the complexly regulated XPR2 promoter [122]. Functional dissection of the XPR2 promoter was used to identify the UAS1B sequence that was devoid of regulation by pH, nitrogen and peptone levels [123]. Placing 4 UAS1B elements tandemly in front of a minimal LEU2 core promoter containing a TATA box demonstrated the first efforts at creating a constitutive, synthetic hybrid hp4d promoter that conferred stronger transcriptional activation to the native P_{XPR2} [123, 124]. Since then, UAS1B elements have been used in tandem to create promoter libraries that exhibited more than a 400-fold increase in transcriptional levels relative to core promoters, effectively bypassing enhancer limitations associated with natural eukaryotic promoter systems [82]. We recently described that the UAS1B elements, while constitutive with respect to nitrogen, elicit carbon source-dependent regulation of expression with oleic acid being able to create the strongest transcriptional activation. Expression from the UAS1B hybrid promoters with glucose in the media is strong, albeit weaker than oleic acid. Using glycerol as a carbon source confers very weak transcriptional activation [80].

An interesting feature of the UAS1B hybrid promoters is that they are growth phase-dependent promoters. These promoters confer limited in gene expression during cell growth, or exponential phase, but increase significantly during late exponential / early stationary phase [80, 124]. This characteristic can be beneficial for heterologous protein production when segregating cell growth and protein expression. This segregation can contribute to cell productivity and alleviate toxic effects associated with heterologous protein expression. The UAS1B hybrid promoter would be less than optimal for metabolic engineering efforts to rewire pathways for substrate utilization, as the desired outcome, in this context, would be early phase enzyme expression while nitrogen is not depleted in the media. For this purpose, strong constitutive promoters from genes TEF1-α and RPS7 in *Y. lipolytica* is better suited [125].

Another widely used constitutive activator sequence used to create hybrid promoters is UAS(TEF), which was systematically dissected from truncations of the native TEF promoter [121]. These elements also demonstrated that tandem usage leads to a 4-fold increase in expression relative to the TEF (404) promoter. Other interesting features of the UAS (TEF) elements were the earlier growth phase transcriptional activation and a more consistent expression level independent of the carbon substrate (sucrose, glucose, glycerol and oleic acid) used compared to the UAS1B elements [121].

The discovery of UAS elements, which have become important modular tools for significantly improved gene expression in *Y. lipolytica*, have aided metabolic engineering and heterologous protein production efforts. However, there is little known to date about the regulatory sequences embedded within these sequences. A conceptual understanding of the enhancers and/or repressors within these elements could enable the development of smaller and more tightly regulatable/inducible hybrid promoters. The only example of a hybrid promoter designed from tandem enhancer sequences in *Y. lipolytica* is the use of Alkane Responsive Element 1 (ARE1) [126]. An n-decane inducible hybrid promoter could be designed by using tandem ARE1 sequences upstream of a minimal core promoter.

The work described this dissertation uses a heuristic approach to develop a fatty acid inducible promoter from fatty acid inducible native *YlPOX2* (YALI0F10857g) promoter. We revealed more than one UAS region in the *POX2* promoter and used tandem repeats of the different UAS_{POX} elements to create a tunable hybrid promoter devoid of carbon catabolite repression [116]. More recently, an erythritol/erythrulose hybrid promoter was developed from identified UAS elements in the erythrulose kinase (*YlEYK1*) gene (YALI0F1606g) of *Y. lipolytica* [127]. The approach to identifying UASs in the native promoter was different to what was shown with the POX2 promoter. Here, the nucleotide sequence of the *EYK1* promoter from different *Yarrowia* clades was aligned to discover two conserved motifs. Mutational studies of the conserved motifs revealed that these sites were responsive to erythritol and erythrulose [127]. Tunable expression was demonstrated using tandem repeats of the UAS_{EYK1}.

Kozak sequences present proximal to the ATG initiator codon could serve as another modular genetic component to control expression. These small sequences play a major role in the initiation of translation in eukaryotic systems [128, 129]. A commonly used Kozak sequence in *Y. lipolytica* that confers strong ribosome recognition affinity is

CCACC [130, 131]. Other modified Kozak sequences used for enhanced translational include AC(A/C)AAA [132, 133] and a CACA sequence [134]. In most examples of potential Kozak sequences used in *Y. lipolytica*, it is desired to have an A in the +3 position. A separate study showed that there was a strong bias towards having A/G at the -3 position and an A/C in the -2 position amongst 47 eukaryotic species [135]. However, more experimental data is required to determine whether this downstream consensus sequence improves the efficiency of translational initiation in *Y. lipolytica*.

1.5.1.4. Genome Editing Capacity for Yarrowia lipolytica

Double-stranded break (DSB) repair in yeast can occur via homologous recombination (HR), single strand annealing (SSA), and non-homologous end-joining (NHEJ) mechanisms such as micro-homology mediated end joining (MMEJ) and illegitimate recombination (IR) [136-138]. A comparative genetic analysis with DNA repair proteins in *S. cerevisiae* revealed that hemiascomycetous species such as *Y. lipolytica* would predominantly utilize the NHEJ pathway for DSB repair [139]. Previous studies aimed at using HR for genome editing and repair required up to 1 kb of homologous flanking fragments for site-directed gene insertion [140]. Long homologous flanking regions are required to yield ~ 50% for site-specific insertion frequency in *Y. lipolytica*. Otherwise, exogenous DNA would integrate randomly into the genome to repair DSB. This suggests that NHEJ is dominant over HR, although the two repair mechanisms are known to work independently in yeast [141].

To improve the frequency of HR in *Y. lipolytica*, the core component of the NHEJ pathway, the ku70 / ku80 heterodimer, was knocked out [142, 143]. In both papers cited above, the Δ*KU70* strain alone led to decreased transformation efficiencies but improved HR frequencies with 1 kb flanking homologies on both ends. Meanwhile, reducing the length of the flanking homology from 1000 bp to 50 bp did not have a dramatic effect HR frequencies, reducing it from 56% to 43%, respectively [142]. The use of short homology lengths (~50-40 bp) for homologous recombination has been demonstrated by using hydroxyurea to arrest and thereby enrich the cells in S-phase cells. This led to gene targeting frequencies between 4-9% in comparison to the untreated cells, where no targeted integration was observed [144].

Further improvements to HR efficiency for genome integration could be accomplished by using the CRISPR-Cas9 mediated system from *Streptococcus pyogenes* in *Y. lipolytica* [83]. Using CRISP-Cas9 directed HR, targeted gene integration occurred higher than 64% in the wildtype while in the Δ*KU70* strain, the frequency was 100%. Meanwhile, studies have shown low dependency of HR frequency to genomic loci [105, 143, 145].

The examples above describe improvements on strategies to perform scarless single copy integrations in *Y. lipolytica*. The use of an auxotrophic marker for selection of genome integration makes screening easy; however, the number of marker genes are limited in *Y. lipolytica*. In some instances, conserving the selectable marker is of interest for future applications and therefore one must rescue the marker post-integration. For this purpose, the Cre-Lox system has been explored in this yeast species [140]. The selectable marker in the disruption cassette is flanked by the LoxP/ LoxR sites which are 34 bp sequences containing 13 bp identical, inverted repeats separated by an 8 bp spacer [146]. Activation of the heterologous bacteriophage Cre-recombinase allows for the excision of the selectable marker after screening for site-directed integration, enabling for the marker to be used again. This editing mechanism, however, leaves a genomic scar but the CRISPR-Cas9 genome editing mechanism is a scar-free. An alternative means of marker recovery is to replace the selectable marker with an inactive gene by HR [85]. Using this method, single copy and multicopy integrations are possible. For multicopy integrations, the rDNA or zeta sites, which are repetitive DNA regions dispersed across the genome of *Y. lipolytica* can be targeted using flanking DNA regions homologous to these site to achieve as high as 30 copies of the gene per cell. [147, 148] Figure 1.2 summarizes the different modules of genetic engineering that have been investigated in *Y. lipolytica*.

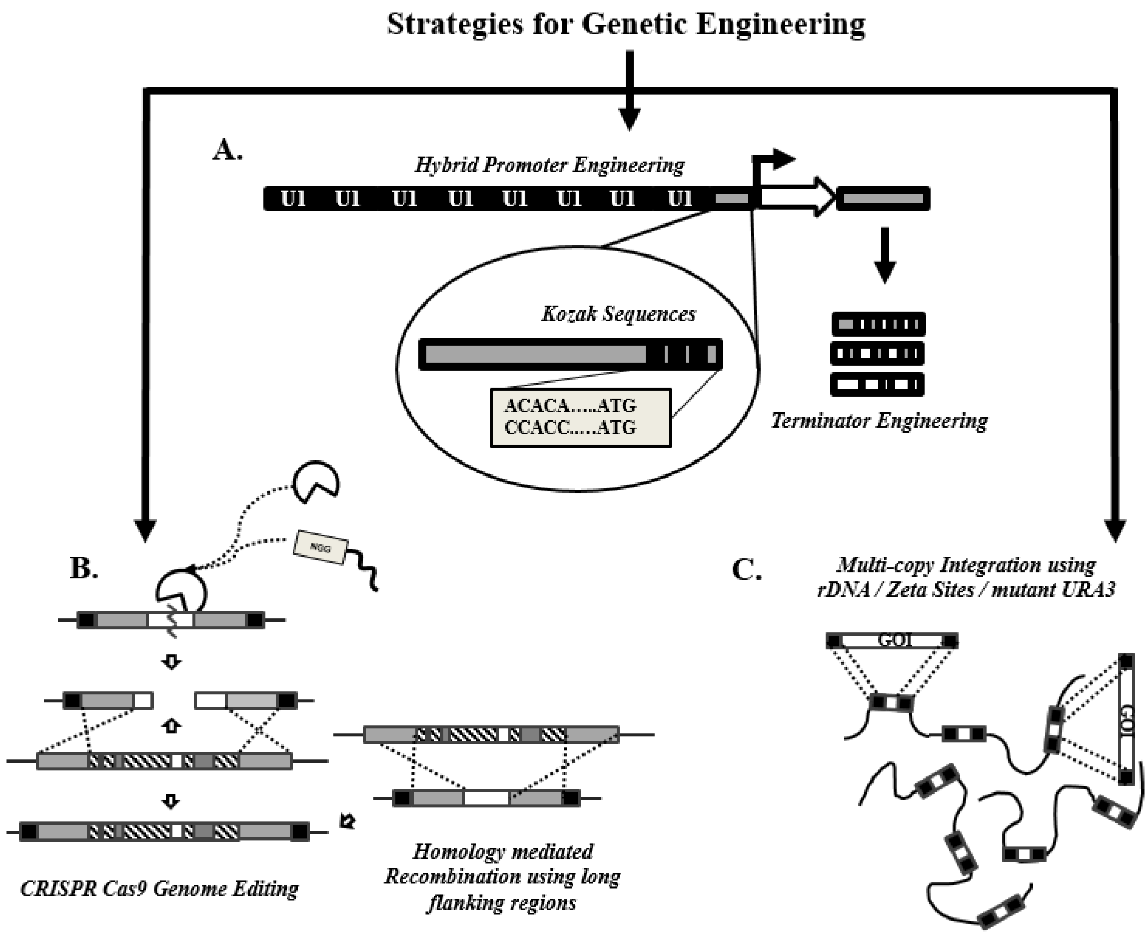


Figure 1.2. General modules applied towards the development of genetic tools and genome editing. A. At the promoter level, genetic tools can be engineered by constructing strong hybrid responsive or inducible systems. Kozak sequence modifications enable improved post-transcriptional expression. Furthermore, engineering synthetic promoters could help to improve mRNA stability and half-life, thereby improving expression levels. B. CRISPR Cas9 for knockouts and homologous recombination of DNA at high efficiencies. Homology mediated recombination requiring the use of large flanking DNA homologies are most efficient in Δ*ku70* strain. C. Multi-copy integration performed using zeta docking sites, rDNA sites or restoring URA3 function by multi-copy integration DNA fragments containing URA3 alleles with the gene of interest [118].

1.6. Metabolic Engineering in *Yarrowia lipolytica*

Lipid accumulation and degradation are two processes that are innately superior in *Y. lipolytica.* Further engineering of these two natural processes has been promising to produce biofuels and fatty acid derived bioproducts. This host has already been engineered to produce a number of key products via engineering its native oleaginous and lipolytic capabilities (Table 1.5) and therefore understanding the engineering scope of lipid accumulation and degradation is important to understanding how novel products can be produced using these pathways.

Table 1.5. Production of lipids and lipid-derived biochemicals in *Y. lipolytica*.

Products	Maximum Titer (g/L)	Reference
TAG Lipids	99 g/L	[84]
Free Fatty Acids (FFA)	10.4 g/L	[149]
Alcohols	2.15 g/L	[91]
Alkanes	23.3 mg/L	[91]
Fatty Acid Ethyl Esters (FAEE)	142.5 mg/L	[91]
Polyhydroxy Alkanoates (PHAs)	1.11 g/L	[150]
Itaconic Acid	4.6 g/L	[88]

1.6.1. Engineering Advanced Oleaginous Capabilities

Lipid accumulation can be induced by nitrogen limited conditions or by high carbon to nitrogen (C/N) ratios. It is proposed that nitrogen exhaustion leads to increased activity of AMP deaminase (*Yl*AMPD), which decreases the concentration of cytosolic AMP. The activity of AMP-dependent isocitrate dehydrogenase (IDH) is therefore inhibited, resulting in the accumulation of isocitrate. Citrate generated from accumulated isocitrate by aconitase then exits the mitochondria and is cleaved by cytosolic ATP-citrate lyase (*Yl*ACL1) for the generation of acetyl-CoA [151, 152]. From the metabolic overview in Figure 1.3., the first committed step of fatty acid synthesis is the carboxylation of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (*Yl*ACC1). NADPH generated by malic enzyme (ME) provides the reducing power for fatty acid synthesis. However, recent research has demonstrated that for *Y. lipolytica*, the pentose phosphate pathway is the major source for NADPH generation [153, 154]. When acetate is used as a substrate, NADPH for fatty acid synthesis is produced through gluconeogenesis and the oxidative pentose phosphate pathway [155]. A recent study showed no significant change in the *Yl*IDH expression during lipid accumulation. However, the gene encoding isocitrate lyase (ICL), which is involved in converting isocitrate to glyoxylate, was observed to be strongly up-regulated in *Y. lipolytica* [156].

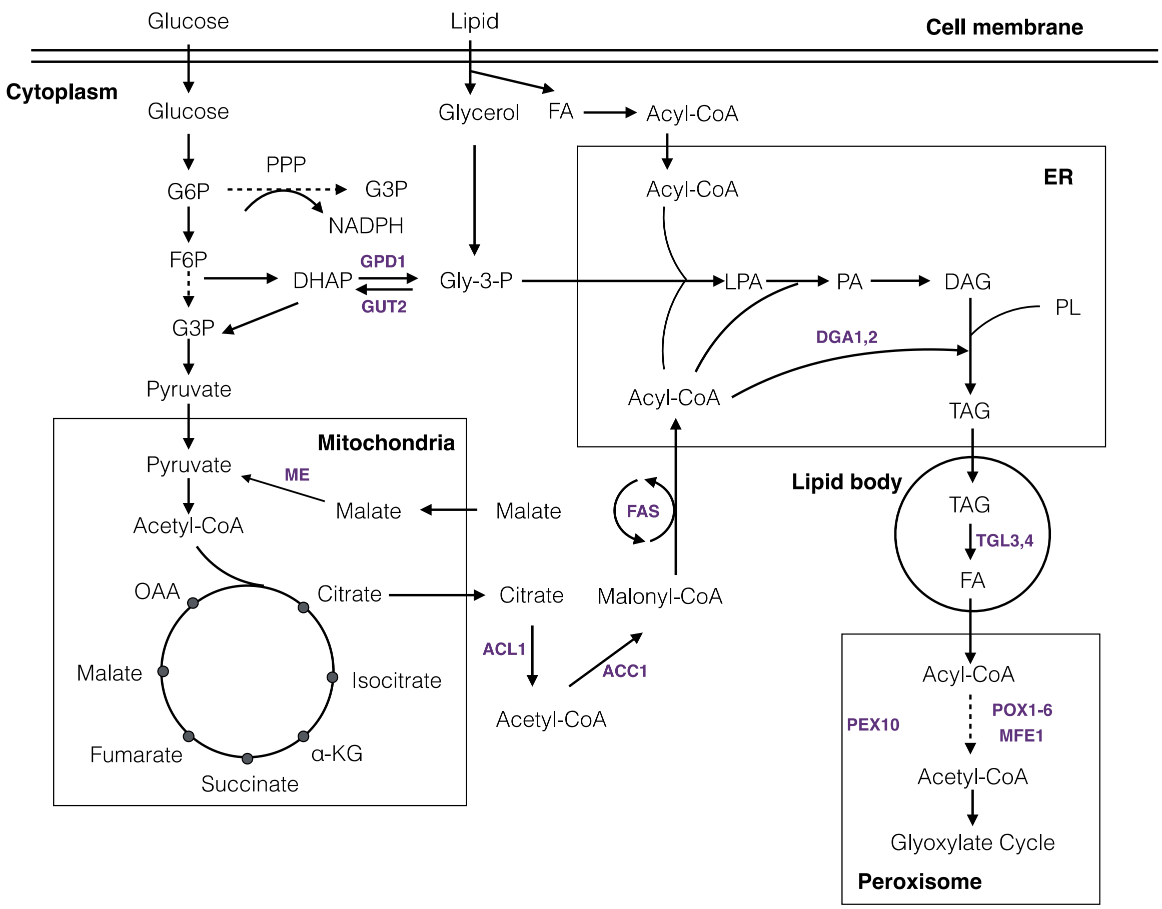


Figure 1.3. Lipid biosynthesis in *Y. lipolytica.* Triacylglyceride (TAG) biosynthesis in *Y. lipolytica*. GPD1: glycerol-3-phosphate dehydrogenase, GUT2: glycerol-3-phosphate dehydrogenase, ACL: ATP-citrate lyase, ACC1: acetyl-CoA carboxylase, ME: malic enzyme, DGA1&2: diacylglycerol acyltransferase, TGL3&4: triacylglyceride lyase, PEX10: peroxisomal biogenesis factor, POX1-6: Peroxisomal Acyl-CoA Oxidase, MFE: β-oxidation multi-function enzyme

TAG synthesis in *Y. lipolytica* involves three acyltransferases. The first step is the incorporation of fatty acyl-CoA into glycerol-3-phosphate (G3P) by glycerol-3-phosphate acyltransferase (*Yl*GPAT), forming lysophosphatidic acid (LPA). LPA and fatty acyl-CoA can be converted to phosphatidic acid (PA) by lysophosphatidic acid acyltransferase (*Yl*LPAT). The phosphate group is then removed from PA to form diacylglycerol (DAG) by phosphatidic acid phosphatase (*Yl*PAP). The last step of TAG synthesis is conducted by one of two types of diacylglycerol acyltransferases that incorporate an acyl group into a DAG. Diacylglycerol acyltransferase (*Yl*DGA1) transfers an acyl group from acyl-CoA to a DAG, while diacyltransferase (*Yl*PDAT) transfers an acyl group from a phospholipid to a DAG.

Early attempts to improve lipid accumulation focused on redirecting the carbon flux towards the glycerol pathway by deleting *GUT2*, preventing the reaction of glycerol-3-P to DHAP, and thereby generating more precursor glycerol-3-P for TAG synthesis. In this study, lipid degradation through β-oxidation was also hindered by deletion of *POX1-6* which encode six acyl-coenzyme A oxidases [157]. In *Y. lipolytica*, *DGA1* which encodes the DGA1 enzyme, and *DGA2,* encoding the DGAT2 enzyme are the only genes contributing to the acylation of DAG. The latter gene is suggested to be the major contributor to TAG synthesis. However, *DGA1* showed great potential in acyltransferase activity when expressed in the quadruple mutant strain under a strong constitutive promoter [158]. A push-and-pull strategy was developed by overexpression of both *ACC* and *DGA1* to enable high levels of lipid accumulation. Double expression of *ACC* and *DGA1* under the control of a strong TEF-intron promoter carries out the first and last step of TAG synthesis, providing an enhanced driving force to redirect the carbon flux toward lipid synthesis, resulting in an increased lipid content of 41.4% [159]. A further enhancement of lipid accumulation was achieved by simultaneous expression of *SCD* (delta-9 stearoyl-CoA desaturase) gene, *ACC1,* and *DGA1*. *SCD* was identified as rate limiting step and target for the metabolic engineering of lipid synthesis pathway by reverse engineering the mammalian cellular obese phenotypes. The high flux created by overexpression of *ACC* and *DGA1* is encouraged and sustained by preventing allosteric pathway inhibition. Overexpression of *SCD* enables the conversion of saturated to monounsaturated fatty acids, providing increased sequestration of the pathway products towards a lipid sink. Moreover, the engineered strain obtained other favorable phenotypes including fast growth, high sugar tolerance, and lipid productivity up to 22 g/l/d [160]. Several efforts have focused on redirecting carbon flux to fatty acid synthesis by modifying glucose repression regulators. Disruption of the *MIG1* gene, encoding a transcriptional regulator that binds to several glucose repression genes, enhanced lipogenesis through depression of several genes relevant to lipid synthesis including *GPD1*, *ICL*, *ME1* and *ACL1*, and through repression of β-oxidation genes including *MFE1* [161]. Another glucose repression regulator Snf1 from the Snf1/AMP-activated protein kinase (AMPK) pathway was identified as a lipid accumulation regulator. Deletion of *SNF1* led to the accumulation of lipid up to 2.6-fold higher than those of the wild-type [162]. Disruption of β-oxidation has been explored to prevent TAG degradation. Pex10p, encoded by the *PEX10* gene, is involved in peroxisome biogenesis. Deletion of *PEX10* in an eicosapentaenoic acid (EPA) producing strain, resulted in inactivation of β-oxidation and increased total lipid accumulation as well as EPA production [85].

The *MFE1* gene is another target for disruption of β-oxidation. Coupling deletion of *MFE1* and improvement of G3P synthesis increased both *de novo* and *ex novo* TAG synthesis [163]. Combinatorial multiplexing of several lipogenesis targets, including deletion of both *MFE1* and *PEX10* genes, overexpression of DGA1, and restoration of a complete leucine biosynthetic pathway, generated a significantly lipogenic strain with a lipid content of 74% [164]. This study also demonstrated that lipid accumulation could be uncoupled from nitrogen starvation and established links between leucine-mediated signaling and lipogenesis. In *Y. lipolytica*, the only source of cytosolic acetyl-CoA is from splitting citrate by ACL when the TCA cycle is repressed under nitrogen-limited conditions. Therefore, uncoupling lipid accumulation and nitrogen starvation can also be achieved by rewiring the acetyl-CoA pathway. Five alternative cytosolic acetyl-CoA pathways were engineered separately, including the pyruvate-acetate route, pyruvate-aldehyde route, pyruvate formate lyase, acetyl-CoA shuttling pathway, and nonoxidative pentose-phosphate pathway [91]. The engineered strains not only show improved lipid production but were also less sensitive to C/N ratio regulation. TGL3 and TGL4 are intracellular lipases responsible for the degradation of TAG in the lipid body. Deletion of the *TGL3* gene has a positive effect on preventing the degradation of TAGs in the later phases of lipid accumulation, and thus increased the overall lipid titer [165]. By combining TGL3 knockout with overexpression of a heterologous DGA1 (*R. toruloides*) and DGA2 (*Claviceps purpurea*), 77% lipid content and 0.21 g/g lipid yield were achieved in a batch fermentation.

Aside from rational metabolic engineering efforts, a rapid evolutionary metabolic engineering approach linked with a floating cell enrichment process was used to develop highly lipogenic strains. This screen led to a strain with a mutation of the succinate semialdehyde dehydrogenase, *UGA2*, achieving a high lipid content of 78% [166], and suggesting an important role of gamma-aminobutyric acid assimilation in lipogenesis. Another evolved strain had a mutant MGA2 protein, Mga2p, that served as a regulator of desaturase gene expression, and exhibited high lipid content with elevated unsaturated fatty acid levels. The mutant MGA2 regulator resulted in a drastically altered transcriptome, with glycolysis upregulated and the TCA cycle downregulated. This suggested that imbalance between glycolysis and the TCA cycle could serve as a driving force for lipogenesis [167].

1.6.2. Understanding FA Metabolism in Y. lipolytica

The process of fatty acid degradation for energy is *β*-oxidation and it primarily occurs within specialized organelles known as peroxisomes (Figure 1.4). The number, size, and content of peroxisomes vary with environmental and genetic stimuli. Some β-oxidation has also been reported to occur within the mitochondria [168]. The peroxisomal *β*-oxidation cycle consists of five major steps. First, the substrate of interest, often a fatty acid (FA), is transported into the peroxisome with the aid of Acyl-CoA Binding Proteins (ACBP) [168]. During transport, the FA is acetylated by the two peroxisomal acyl-CoA Synthases (PXA1/ PXA2) in an ATP-dependent reaction [169]. The newly acylated fatty acid is then desaturated by acyl-CoA oxidases (POX) at the vinyl position, consuming FAD^+ and producing H_2O_2 as a byproduct. *Yarrowia lipolytica* has six *POX* genes (*POX1-POX6*) which have been shown to different chain length and substrate specificities [170]. The newly formed desaturated FA-CoA ester is then hydrated across the double bond by Multi-Function Enzyme 2 (MFE2 – C domain) (encoded by the

MFE1 gene) such that the addition of a hydroxyl occurs at the *β*-carbon position, forming a 3-hydroxyacyl-CoA intermediate. From here, the MFE2 enzyme (A/B domains) acts again to oxidize the 3-hydroxy intermediate to 3-ketoacyl-CoA and forms NADH in the process. Finally, the 3-ketoacyl-CoA is cleaved at the alpha carbon by peroxisomal 3-oxyacyl-thiolase (POT1), releasing a molecule of acetyl-CoA and producing a fatty acyl-CoA which is two carbons shorter than the substrate that entered the cycle. From this point, the product that is now 2 carbons shorter, can loop back into the cycle beginning with the POX reaction.

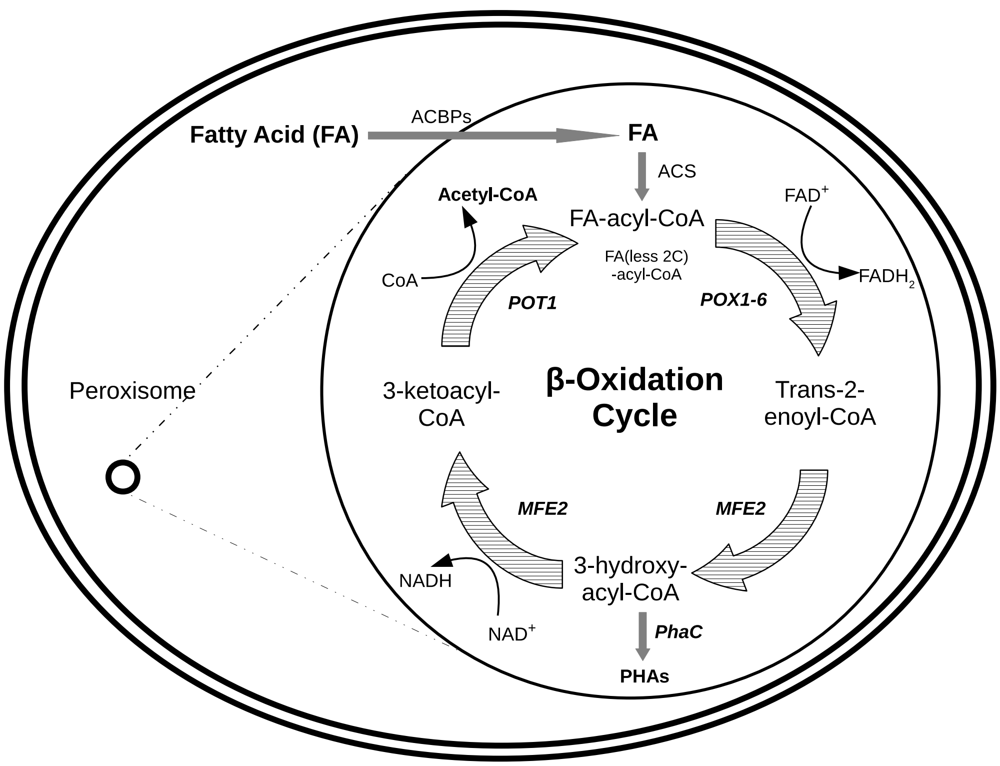


Figure 1.4. β-oxidation in *Y. lipolytica.* The enzymes involved in β-oxidation in *Y. lipolytica*. ACBP: Acyl-CoA Binding Proteins, ACS: acyl-CoA synthase, POX1-POX6: peroxisomal acyl-CoA oxidases, MFE2: multi-function enzyme 2, POT1: peroxisomal 3-oxoacyl-CoA thiolase, PhaC: polyhydroxyalkanoate synthase [118].

1.6.3. Advancing Metabolic Engineering of Alcohols and Other Oleochemicals

Recently several groups have begun producing other oleochemicals including alcohols, alkanes, and esters [91-93]. Medium and long-chain alcohols are used as moisturizers in cosmetics as well as lubricants and surfactants. Alcohol production in *Y. lipolytica* is typically achieved using a fatty acyl-CoA reductase and aldehyde reductase or carboxylic acid reductase (CAR) and aldehyde reductase. The bifunctional fatty acyl-CoA / aldehyde reductase from *Tyto alba* (TaFAR1) was used to enable the production of hexadecanol [92]. The deletion of the fatty alcohol oxidase (*FAO1*) gene from *Y. lipolytica* and increasing the copy number of the *Tafar1* gene lead to a ~5-fold increase in titers. Deletion of the *DGA1* gene responsible for TAG synthesis and the introduction of 5 copies of the Tafar1 gene led to a titer of ~690 mg/L from 160 g/L glucose after 6 days.

The production of 1-decanol was demonstrated using the FAR from *Arabidopsis thaliana* using a previously engineered *Y. lipolytica* for C8–C10 medium chain fatty acids [93]. The deletion of the *PEX10* gene while expressing FAR greatly increased 1-decanol titers by preventing peroxisome formation and thus alcohol degradation. A number of fatty acyl-ACP thioesterase (FAT) enzymes were also tested to release fatty acids from biosynthesis. The FAT enzyme from *Cuphea palustris* yielded the best decanol titers (550 mg/L). The majority of the decanol (~90%) was found to be secreted outside the cell and into the media.

More recently, a range of oleochemicals was produced by targeting various pathways to the different organelles involving fatty acid biosynthesis and degradation [91]. Fatty acid ethyl ester (FAEEs) production was achieved by expression of acetyltransferase, AtfA, from *Acinetobacter Baylyi*. When targeting this enzyme to the ER or Peroxisome, 136 mg/L and 111 mg/L of FAEE was produced, respectively, whereas only 7 mg/L was produced when targeted to the cytosol. Alkanes were produced using a similar organelle targeting approach by expressing the aldehyde-deformylating oxygenase (ADO) and CAR. Up to 23 mg/L of fatty alkanes were made by expressing CAR from *Mycobacterium marinum* and ADO from *Prochlorococcus marinus.* Alcohol production has also been demonstrated in *E. coli* were also produced by expression of *E. coli* fatty acyl-CoA synthetase, FadD, and *Marinobacter aquaeolei* FAR. In yeasts, however, the highest scale-up titers are reported in a 3 L bioreactor for *Y. lipolytica*, reaching 2.15 g/L.

[†] A part of this chapter is published in *AIMS Bioengineering* **2016**, *3(4), 493-514,* with co-authors Gabriel Rodriguez, Difeng Gao, Michael Spagnuolo, Lauren Gambill, and Mark Blenner

CHAPTER TWO

ENGINEERING PROMOTER ARCHITECTURE IN *YARROWIA LIPOLYTICA*

Abstract

Eukaryotic promoters have a complex architecture to control both the strength and timing of gene transcription spanning up to thousands of bases from the initiation site. This complexity makes rational fine-tuning of promoters in fungi difficult to predict; however, this very same complexity enables multiple possible strategies for engineering promoter strength. Here, we studied promoter architecture in the oleaginous yeast, *Yarrowia lipolytica*. While recent studies have focused on upstream activating sequences, we systematically examined various components common in fungal promoters. Here, we examine several promoter components including upstream activating sequences, proximal promoter sequences, core promoters, and the TATA box in autonomously replicating expression plasmids and integrated into the genome. Our findings show that promoter strength can be fine-tuned through the engineering of the TATA box sequence, core promoter, and upstream activating sequences. Additionally, we identified a previously unreported oleic acid-responsive transcription enhancement in the XPR2 upstream activating sequences, which illustrates the complexity of fungal promoters. The promoters engineered here provide new genetic tools for metabolic engineering in *Y. lipolytica* and provide promoter engineering strategies that may be useful in engineering other non-model fungal systems.

† A version of this chapter is published in *ACS Synthetic Biology* **2015**, *5*, *213-223*, with co-authors Lauren Gambill, Spencer Smith, and Mark Blenner.

Introduction

The complexity of eukaryotic promoter architecture is fundamental for the diverse pattern of gene expression that can be obtained from a relatively small number of transcription factors (TFs) [171-173]. Even in eukaryotes as exhaustively studied as *S. cerevisiae*, fine-tuned and predictive promoter design has been elusive.[58, 174, 175] As a result, metabolic engineering in eukaryotes has relied on a small number of well-defined endogenous promoters, such as the GAL1-10, TEF, and LEU2 promoters [176-178]. This problem is exacerbated by recent progress towards utilizing non-model yeasts from biochemical production [179-182]. In order to push titers, yields, and productivities to their limits, reaction fluxes need to be well-balanced, and even responsive to intermediate metabolite concentration [183-186]. Such advanced metabolic engineering strategies may be enabled if promoters were designed from the bottom up to have specific transcriptional activities. Central to the development of finely tuned promoters is a better understanding of how different promoters' elements influence promoter strength.

This work focuses on engineering promoters for the oleaginous and lipolytic hemiascomycetes yeast *Yarrowia lipolytica*. *Y. lipolytica* has long been studied as a model organism for dimorphism and as an alkane metabolizing yeast [187, 188]. It is also known to metabolize diverse substrates including fatty acids, triacylglycerides, glucose, and glycerol [189]. As an oleaginous yeast, it is able to accumulate greater than 20% of its mass as neutral lipids. Recent efforts to increase the lipid content of these cells grown on glucose have been successful, with resulting strains engineered to produce up to 90% w/w [160, 190]. As these lipids are useful precursors for biofuels, fatty acids, and fatty alcohols, *Y. lipolytica* has gained attention as a useful industrial microbe for the production of omega-3 oils, ricinoleic acid, and triacylglycerides [85, 191]. Recent efforts have greatly improved de novo production of TAGs for conversion to biodiesel [160, 190]. Other products produced include single cell protein, citric acid, lipase, lycopene, and α-ketoglutarate [192-196].

Overexpression of endogenous or heterologous enzymes to form new metabolic pathways requires functional promoters. Endogenous promoters are frequently used because promoter architecture in eukaryotes can be complex and transcriptional responses can be difficult to predict [197, 198]. Endogenous promoters used for overexpression include TEF1, FBA1, TDH1, GPM1, LEU2, POX2, XPR2. These promoters are typically over 1000 bp long and were identified by analysis of genomic and gene expression data. Unfortunately, this precludes tuning the level of gene expression since promoter strength is fixed by the endogenous promoter architecture. Without additional engineering, endogenous promoters cannot produce transcripts at levels higher than naturally occurring. Furthermore, the complex regulation of endogenous promoters is often ignored and can complicate metabolic engineering efforts.

The promoter strength is determined by several factors, including the TATA box, core promoter sequence, proximal promoter sequences, and enhancer regions in the upstream activating sequences (UAS) (Figure 2.1C). The most attention has been given to engineering hybrid promoters, built by combining repeats of UASs with downstream minimal promoters comprised of truncated promoters [81, 82, 199]. Madzak et al. [200] showed the promoter of extracellular protease (XPR2) can be described as two regions:

UAS1 and UAS2, with UAS2 being closer to the start codon and UAS1 being farther from the start codon. The endogenous XPR2 promoter is regulated by the media pH and nitrogen content. When only UAS2 was used to drive the expression of XPR2, the same pH and nitrogen regulation was observed. On the contrary, when only UAS1B, a 90 bp region of UAS1 was used to drive expression, transcription was independent of pH and nitrogen indicating regulatory features of the UAS were localized to UAS2, while UAS1 was a general amplifier of the downstream promoter. Blazeck et al.[82] made tandem repeats of UAS1B to drive expression of GFP from two minimal constitutive promoters, TEF and LEU. Increasing the number of repeats monotonically and cooperatively increased the transcription from the downstream core promoter. UAS repeats have been shown to be genetically stable in *Y. lipolytica* [82]. This modular architecture suggests that promoter strength and induction properties should be predictably engineerable using defined UASs, however, considerably less attention has been given to the TATA box, core promoter, and proximal promoter sequences.

In this study, we have taken a systematic look at the promoter architecture in order to engineer a new panel of hybrid promoters for metabolic engineering applications. We chose to study promoter structure in the context of autonomously replicating plasmids for two reasons. First, it allows us to study promoter architecture in the absence of epigenetic effects that commonly influence expression profiles in chromosomal DNA; and secondly, plasmids are useful vehicles for rapid testing of metabolic engineering strategies. We made several truncations to the acyl CoA oxidase (POX2) endogenous promoter and built hybrid promoters. We characterized a new substrate responsiveness from the XPR2 UAS1B sequences and showed that UAS1B enhancers are more induced by oleic acid compared to glycerol or glucose. We also tuned promoter strength through engineering of the TATA box and the proximal promoter regions. While TATA box engineering resulted in similar effects in different promoters, the proximal promoter sequences did not appear modular. In sum, these studies have helped elucidate the importance of each of the regions comprising the overall promoter architecture.

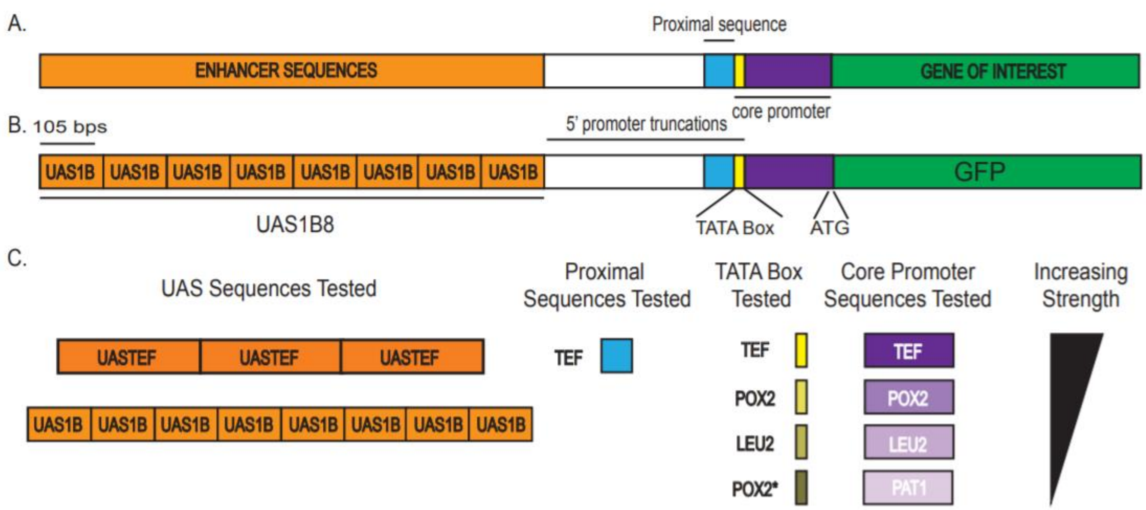


Figure 2.1. Promoter architecture. (A) Eukaryotic promoters contain a core promoter sequence that may have a TATA box, a proximal promoter sequence, and enhancer sequences located farther upstream. (B) Hybrid promoters were created by placing eight UAS1B sequences (UAS1B8) upstream of a promoter. The promoter consists of a core promoter and 5′ truncations of the native promoter to identify regulatory sequences upstream of the proximal and core promoter. Humanized Renilla reinformis GFP was used as a reporter to characterize promoter strength. (C) Different UAS, proximal, TATA box, and core promoter sequences tested throughout the course of this study. The promoter components are ordered from top to bottom in decreasing strength.

Experimental Procedures

Chemicals and Enzymes

All chemicals were purchased from Sigma unless otherwise stated. All restriction enzymes, DNA ligases, and DNA polymerases used for cloning and PCR were purchased from New England Biolabs (NEB) unless otherwise stated. Plasmid minipreps, PCR purifications, and gel extractions were done using the QIAprep spin miniprep kit and QIAquick PCR purification and gel extraction kit (Qiagen). Genomic DNA from *Y. lipolytica* was extracted using the E.Z.N.A. yeast DNA kit (Omega Biotek). All oligonucleotides and gBlocks were purchased from IDT.

Strains and Cultures

DH10β cells (NEB) were used for cloning and propagation of plasmids in Luria−Bertani (LB) media supplemented with 100 μg/mL ampicillin. *Y. lipolytica* strain PO1f (MATa leu2−270 ura3−302 xpr2−322 axp1) was used for GFP expression studies. Transformed was cultivated at 28 °C and 215 rpm in 20 mL volumes using 250 mL baffled flasks with YSC-LEU selective media consisting of 6.7 g/L YNB without amino acids (Difco), 0.69 g/L CSM-LEU (MP Biomedicals). Data in Figure 6 was collected in 2 mL volumes grown in 14 mL culture tubes under the same rotational speed and incubation temperature. The carbon source for YSC-LEU media contained either 2% (w/v) glucose (Sigma), 2% (v/v) glycerol (Fisher Scientific), or 2% (v/v) emulsified oleic acid (EMD Millipore) in 0.1% (v/v) Tween 80. Agar plates for post-transformation applications in *E. coli* and *Y. lipolytica* were prepared by adding 15 g/L agar to either LB or YSC-LEU media, respectively. *Y. lipolytica* transformations were done using the lithium acetate method as described previously [106].

General Plasmid Construction

Procedures for restriction enzyme digestions and PCR amplification were performed as recommended by supplier protocols. Ligations were incubated at room temperature for 20 min using T4 DNA ligase prior to transformation into DH10β *E. coli* using the heat shock method. Post-ligation transformants were grown overnight in LB media supplemented with 100 μg/mL ampicillin. All vectors used in this study were derived from a skeletal plasmid containing a centromeric site and autonomous replicative sequence (CEN/ ARS) pSL16-cen1-1(227) [173]. A gBlock containing a multiple cloning sites (MCS), hrGFP, and CYC1 terminator was synthesized (Supplementary Table 2.1) with flanking 5′ BamHI and 3′ HindIII restriction sites and ligated into the pSL16-cen1-1(227) skeletal vector to create pSL16-cen1-1(227)-MCS-hrGFP-CYC1t. Eight tandem repeats of UAS1B elements were PCR amplified from pUC-UAS1B8-TEF(136) [82] using primer pair F1/R1 (Supplementary Table 2.1). The F1 mutagenesis forward primer was used to introduce a BstBI restriction site and mutate the SphI site adjacent to 5′ of the UAS1B sequences, while the reverse primer retained the SphI site. The PCR amplified UAS1B8 fragment was ligated into the pSL16-cen1-1(227)-MCS-TEF(136)-hrGFP-CYC1t to create pSL16-UAS1B8-TEF(136)-hrGFP.

Construction of POX2 Promoter Truncation Plasmids

The native POX2 promoter, POX2 (2147 bp), and subsequent truncations, POX2 (1591 bp), POX2 (513 bp), POX2 (438 bp), POX2 (147 bp), and POX2 (100 bp), were PCR amplified from PO1f genomic DNA using primer pairs F2/R2, F3/R2, F4/R2, F5/R2, F6/R2, and F7/R2, respectively (Supplementary Table 2.2). The POX2 promoter truncations were ligated in place of the TEF(136) promoter using restriction sites SphI/AscI to make a series of POX2 hybrid promoters, pSL16-UAS1B8-POX2 (x bp)-hrGFP. Unless otherwise stated, all hybrid promoter constructs utilized 5′ SphI and 3′ AscI sites for ligation into the pSL16-UAS1B8- TEF(136)-hrGFP vector by replacing the TEF(136) minimal core promoter. To create a series of vectors containing truncations of the native POX2 promoter without the UAS1B elements, the series of POX2 promoter truncations were PCR amplified from PO1f genomic DNA using primer pairs F8/R2, F9/R2, F10/R2, F11/R2, F12/R2, and F13/R2 (Supplementary Table 2.2). The forward primers had a 5′ flanking XmaI site for ligation into the pSL16-cen1-1(227)-MCS-hrGFP-CYC1t vector to create the series of pSL16-POX2 (x bp)-hrGFP vectors.

Construction of Hybrid Core Promoter Plasmids

Four hybrid core promoter systems were constructed using core promoters of the TEF, LEU2, POX2, and PAT1 genes. Primer pairs F14/R3 and F15/R4 were used to amplify the TEF (111 bp) and LEU2 (78 bp) core TATA promoters, while synthesized oligonucleotide pairs F16/R5 and F17/R6 (Supplementary Table 2.2) were annealed to make the core TATA promoters, POX2 (48 bp) and PAT1 (61 bp), respectively. These promoters were digested and ligated into pSL16-UAS1B8-TEF-hrGFP in place of the TEF(136) promoter to construct the hybrid core promoters, pSL16-UAS1B8-POX2 (x bp)-hrGFP.

Construction of Hybrid TATA Box Promoter Plasmids

The TATA box in TEF core hybrid promoter was replaced with the POX2 and LEU2 TATA box motifs using mutagenesis forward primers F18 and F19 paired with R3. The template used for PCR was pSL16-UAS1B8-TEF (111 bp)-hrGFP. Mutations to the TATA box in the POX2 core promoter to substitute in the TEF and LEU2 TATA motifs were accomplished by purchasing synthesized oligonucleotides F20/R7 and F21/R8, respectively. pSL16-UAS1B8-POX2 (2147 bp)-hrGFP was digested to remove the POX2 (2147 bp) promoter, and the abovementioned annealed oligos were ligated in its place. Sequence and ligation-independent cloning (SLIC) [201] was used to replace the TATA box of the native POX2 promoter with the canonical TEF and the LEU2 TATA motifs. The vector template was pSL16-POX2 (2147 bp)-hrGFP. Primer pairs F22/R9 and F23/ R10 were used for the TEF TATA substitution, whereas F23/ R11 and F24/R10 were used to make LEU2 TATA motif substitutions (Supplementary Table 2.2). The base vector for mutagenesis and PCR amplification was pSL16-POX2 (2147 bp)-hrGFP.

Construction of Hybrid Proximal Promoter Plasmids

The effect of the proximal sequence in the TEF minimal core promoter was tested by placing proximal motif upstream of the POX2 and LEU2 core promoters, respectively.

Two separate gBlocks containing the respective core promoters and the 26 bp proximal sequence 5′ of the TEF core promoter, F25, and F36 were purchased. To test the effect of various TATA box associations with the TEF proximal sequence, gBlocks of the POX2 core promoters with modified TATA boxes containing the TEF proximal sequence (F32 and F33) were also purchased (Supplementary Table 2.2). SLIC was used to insert the oligonucleotides into a SphI/AscI double-digested pSL16-UAS1B8-POX2 (2147 bp)-hrGFP vector.

<u>Construction of UAS1B8 Hybrid Promoters for Genome Integration</u>

Hybrid promoter cassettes with TATA box modifications were made for integration into the leu2 locus of *Y. lipolytica* Δku70. The *Δku70* strain has previously been shown to improve the efficiency of homologous recombination events.51 pSL16-cen1-1(227) was digested with AatII to insert a 500 bp front-end homology to the leu2 locus and new restriction site, AvrII, in the vector. The 500 bp front-end homology was PCR amplified from the *Y. lipolytica* genome using SLIC primer pairs F26/R12 (Supplementary Table 2.2). The new vector containing the frontend homology was digested with HindIII-HF to insert a 500 bp back-end homology to the leu2 locus and a new restriction site, MfeI. The 500 bp back-end homology was PCR amplified from the *Y. lipolytica* genome using SLIC primer pairs F27/R13 (Supplementary Table 2.2). The new vector was termed pSL16-cen1-1(227) [LEU2 homologous ends]. The three hybrid TEF core promoter cassettes containing the TEF, LEU, and POX2 TATA boxes, respectively, were digested with BstBI and AatII, and an insert containing the LEU2 promoter, gene, and terminator was PCR amplified from the *Y. lipolytica* W29 genome using SLIC primer pairs F28/R14 and inserted into the double-digested vectors above (Supplementary Table 2.2). The purpose of this step was to remove the centromeric CEN1-1 DNA sequence that makes the plasmid replicative. These three vectors were then doubled-digested with AatII and HindIII to add in an insert containing leu2 homologous ends that were PCR amplified from pSL16-cen1-1(227) [LEU2 homologous ends] using primer pairs F27/R12 (Supplementary Table 2.2). Integration at the leu2 locus was verified by PCR and DNA sequencing.

<u>Construction of Vectors Containing Three Tandem UAS(TEF) Elements</u>

A pUC vector was used to first make three tandem UAS(TEF)#2 sequences33 prior to transferring these elements into a pSL16 hybrid vector. The UAS(TEF) from the first UAS(TEF) element was PCR amplified from the *Y. lipolytica* PO1f genome using SLIC primer pair F29/R15 (Supplementary Table 2.2) and inserted into pUC-UAS1B8(TEF136) (Addgene no. 44380) double-digested with SphI and BamHI to remove the 8UAS1B elements, creating pUC-UAS(TEF)#1. The second UAS(TEF) element was PCR amplified from the *Y. lipolytica* PO1f genome using SLIC primer pair F30/R16 (Supplementary Table 2.2) and inserted into the pUC-UAS(TEF)#1 vector double-digested with NdeI and BamHI to create pUC-UAS(TEF)#2. Finally, the third UAS- (TEF) element was PCR amplified from the *Y. lipolytica* PO1f genome using SLIC primer pair F30/R17 (Supplementary Table 2.2) and inserted into the pUC-UAS(TEF)#2 vector double-digested with NdeI and EcoRI to create pUC-UAS(TEF)#3. The three tandem UAS(TEF) elements were digested out of the pUC-UAS(TEF)#3 vector using SphI/BstBI and inserted into the double-digested pSL16-UAS1B8-TEF(TATA)-hrGFP, pSL16- UAS1B8-TEF(LEU2 TATA)-hrGFP, and pSL16-UAS1B8- TEF(POX2 TATA)-hrGFP core promoter hybrid cassettes to create counterpart vectors containing 3UAS(TEF) elements.

Flow Cytometry

Humanized *Renilla reinformis* GFP (hrGFP), codon optimized for *Y. lipolytica*, was used as the fluorescence reporter protein to measure promoter strength. Transformants from selective media plates were first propagated in 2 mL precultures for 48 h prior to inoculating 10 mL cultures at an OD600 of 0.3. The cultures were grown under constant agitation for 48 h, which was reported as the optimal incubation time for high expression levels.32 Prior to flow cytometry analysis, cultures were grown in glucose and glycerol and were spun at 12000xg for 2 min and resuspended in 0.1 M phosphate-buffered saline (PBS) at pH 7.4. Cultures grown in emulsified oleic acid were spun at 12 000gfor 2 min and suspended in YSC-LEU containing 5% (v/v) Tween 80. The resuspended pellet was agitated by vigorous shaking prior to centrifuging the sample at 12 000gfor 1 min and resuspending the pellet in 0.1 M PBS, pH 7.4. All samples were kept on ice during sample preparation and analysis. Flow cytometry analysis was performed using the BD Accuri C6 flow cytometer (BD Biosciences) with the standard detector, FL1 filter (533/30), used to capture fluorescence from the GFP fluorophore. The VIRTUALGAIN module in the BD Accuri C6 software was used to adjust peak position and account for normalizing gains across samples during analysis. Population gates were applied to account for the mean fluorescence from the GFP expressing population and negate autofluorescence.

Quantitative Reverse Transcription PCR (qRT-PCR)

Transformants grown under different carbon source conditions were subject to RNA extraction 48 h post-growth. The transformants across each of the selected cultures were normalized to an OD600 of 5 prior to RNA extraction procedures. The cells were pelleted, and total RNA was extracted using the E.Z.N.A yeast RNA kit (Omega Biotek). RNA extracts were placed in aliquots and stored at −80 °C until further use. For absolute RT-qPCR, a two-step protocol was employed. 500 ng of total RNA was used in cDNA synthesis that was performed using gene-specific priming with maxima reverse transcriptase (Thermo Scientific). 1.5 μL from the cDNA synthesis mix was subject to qPCR with the Maxima SYBR Green/Fluorescein qPCR master mix (Thermo Scientific). qPCR was performed in biological triplicates from the cDNA mix in a 96-well plate using a CFX Connect real-time (Bio-Rad). The primer pair, GFPF/ GFPR, used in RT-qPCR is listed in Supplementary Table 2.2. A standard curve was developed using a linearized vector containing the hrGFP gene to relate Cq values to copy number. This calibration curve was used to calculate mRNA copy numbers of qPCR analyzed samples.

Results

Oleic Acid Inducible Enhancers Are Upstream in the Native POX2 Promoter

While the whole POX2 promoter has proven to be useful for heterologous expression in *Y. lipolytica*, there is little known about the mechanism or localization of elements that confer oleic acid responsiveness in this relatively large promoter. To identify such oleic acid response elements (OREs) in the promoter, a series of 5′ deletions were made to the endogenous POX2 promoter, based on homology to *S. cerevisiae* OREs. These promoter truncations were placed upstream of a humanized *R. reinformis* GFP (hrGFP) reporter gene to quantify expression via cellular fluorescence (Figure 2.1A). The significant drop in fluorescence was observed from the POX2 (1591 bp) to the POX2 (513 bp) promoter. This indicates that most of the OREs reside in this 1 kb window, significantly upstream of the TATA box, where the preinitiation complex (PIC) is known to form (Figure 2.1A) in RNA polymerase II promoters.

UAS1B Sequences Act in a Distance-Dependent Manner

The UAS1B element in *Y. lipolytica* has previously been shown to be a constitutive transcriptional amplifier that is independent of the nitrogen content and pH of the media [199] and fusion of greater than four UAS1B elements in tandem could lead to a cooperative and significant amplification [199]. We combined eight tandem UAS1B (UAS1B8) repeats with truncations of the POX2 inducible promoter system and observed the strongest expression from the shortest POX2 truncations. Several short POX2 promoter truncations (100, 147, and 438 bp) resulted in minimal GFP fluorescence;

however, when they were combined with UAS1B8, these promoters were stronger than longer hybrid POX2 promoters. We observed that the UAS1B8 sequences conferred weaker transcriptional amplification as they were moved farther away from the gene (Figure 2.2B). A significant decrease in fluorescence is observed from the UAS1B8-POX2 (513 bp) to the UAS1B8-POX2 (1591 bp) hybrid promoter, which is an opposite effect to what was observed with the truncations of the native promoter.

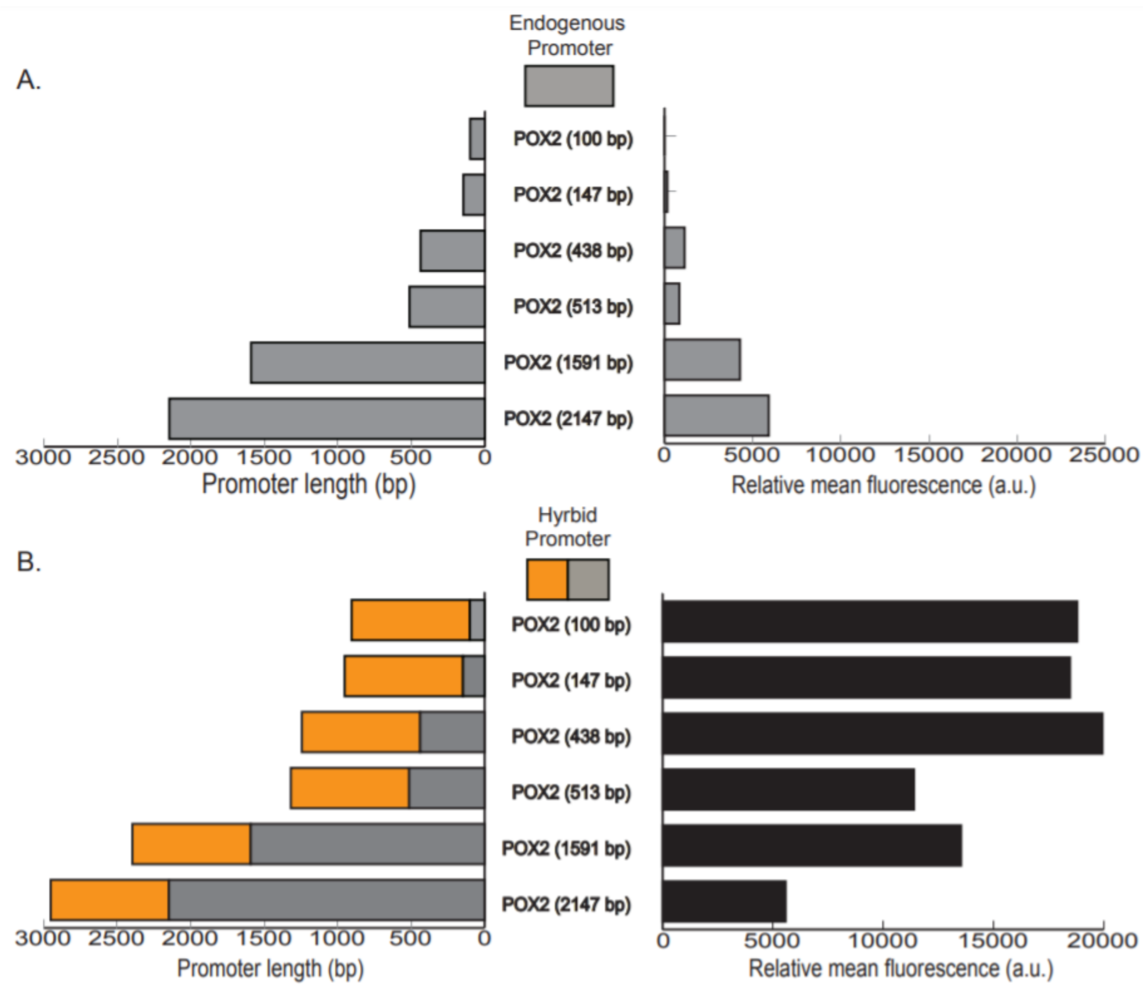


Figure 2.2. Truncated POX2 promoters and hybrid POX2 promoters. (A) Length of 5′ truncations of the native POX2 promoter (gray; left) and oleic acid induced fluorescence from the corresponding truncations of the endogenous POX2 promoter (right). (B) Eight UAS1B elements (orange) fused upstream of the 5′ POX2 promoter truncations (gray) to

create a series of POX2 hybrid promoter systems (left) and oleic acid induced fluorescence (black) from the corresponding hybrid POX2 promoter constructs (right).

Core Promoter Sequence Modulates Expression Level

A complete understanding of the mechanism for TFIID-dependent transcription and TSS localization in yeast promoters is still unclear; therefore, a range of TATA promoters containing different distributions of predicted initiator sequences were used in this study. Transcriptional effects from the types of initiator sequences (Supplementary Figure 2.1) and distances from TATA box still remain to be fully elucidated in yeast TATA-containing promoters [202]. Our hybrid POX2 promoter studies indicate that UAS1B8 sequences placed upstream of the core promoter truncated down to the TATA box still result in high levels of transcription and therefore we hypothesized that these sequences can similarly amplify other core promoters. The UAS1B8 elements were fused upstream of core promoters from the PAT1, POX2 inducible promoters and the TEF1-α, LEU2 constitutive promoters. The core promoters constructed in this chapter are designated as the truncated promoter sequence to a functional TATA box (Figure 2.3A), which acts as the binding site of the PIC in TATA-containing promoters [202]. The functional TATA boxes in these promoters were identified by scanning the promoter sequence between 40 and 100 bp from the start codon [203]. GFP reporter expression and quantitative PCR results indicate that, indeed, the UAS1B8 sequence enhances transcription from all core promoters tested and that transcript levels scale with core promoter strength, with TEF1 being the strongest and PAT1 and LEU2, the weakest. We also observed that the TEF1 core promoter is the longest functional core promoter (Figure 2.3A), whereas the shortest core promoter (POX2) had similarly high expression.

UAS1B from XPR2 Confers Substrate-Specific Responsiveness

The UAS1B sequence was originally described as lacking regulation by media conditions when compared to the full XPR2 promoter; however, an investigation of UAS1B hybrid promoter strength when using different carbon sources was not performed [199]. We compared the expression levels of GFP by flow cytometry and qPCR of cultures grown in either YSC media containing glucose, glycerol, or oleic acid. Oleic acid substrate resulted in the highest levels of promoter activation. UAS1B8 hybrid promoters were less activated in glucose and were minimally activated in glycerol. All four promoter systems exhibited a similar trend in substrate-specific responsiveness, largely following the pattern of core promoter strength (Figure 2.3B, C).

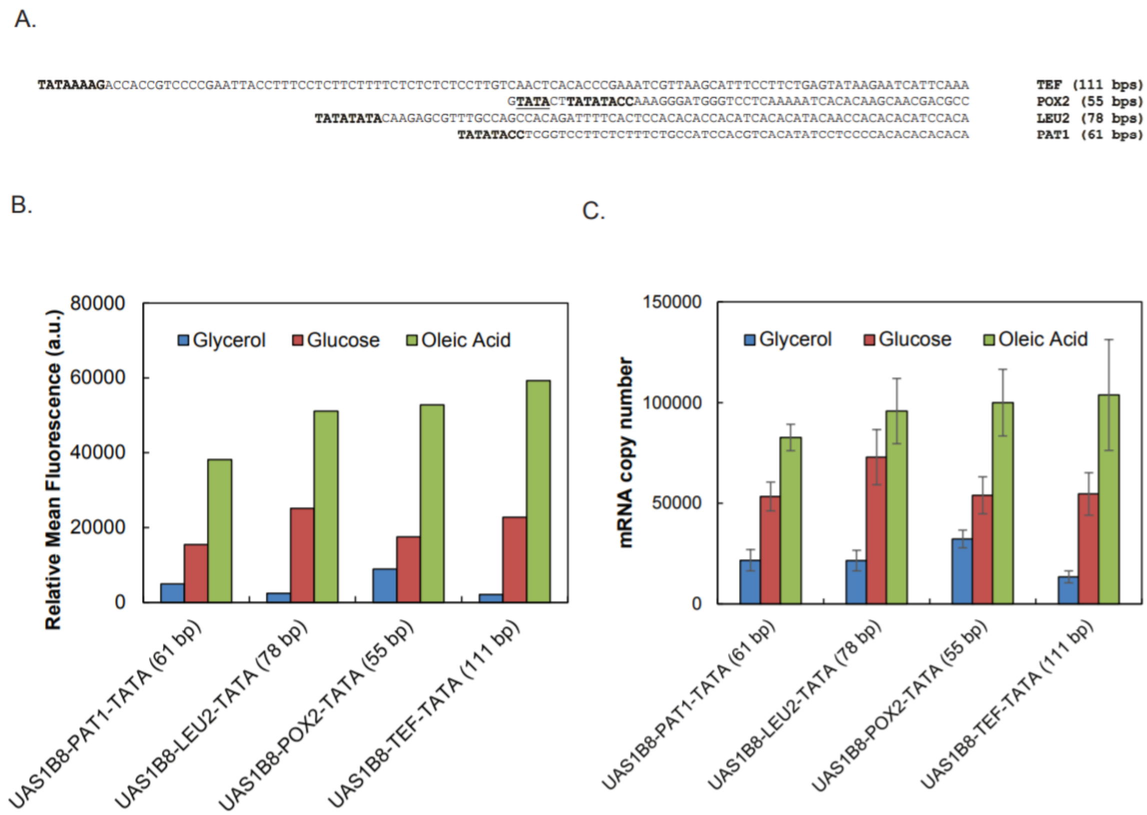
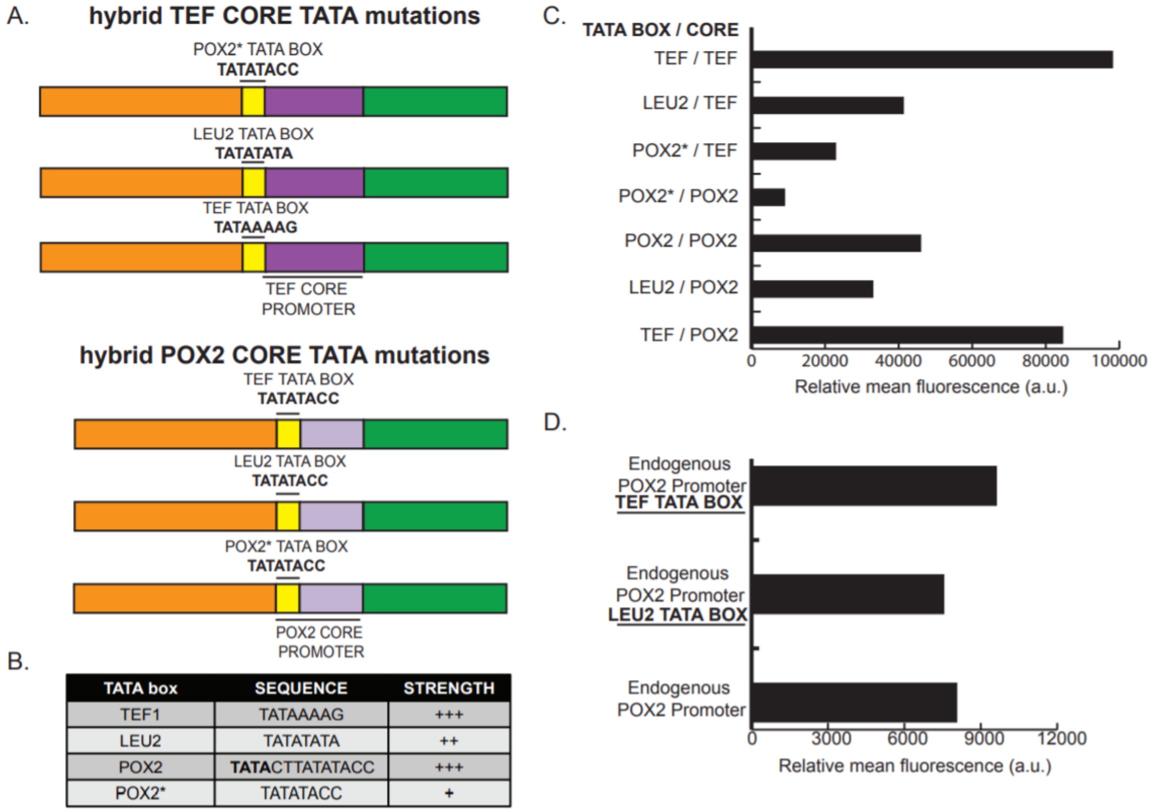


Figure 2.3. UAS1B8 enhancers are oleic acid responsive. (A) Core promoter sequences of TEF, POX2, LEU2, and PAT1 shows the variability of the TATA box sequence. (B, C) Substrate-specific transcriptional responsiveness (fluorescence, B; absolute qPCR, C) from the various hybrid core promoter constructs.

TATA Box Sequence Has a Large Influence on Expression Level

To study the effect of TATA elements on transcriptional regulation, we chose TATA boxes as 8 bp sequences, with 2 additional bases downstream of the 6 bp core TATA box from four sequentially different *Y. lipolytica* promoters. The TEF promoter contains a TATAAA[AG] sequence that is reported to be one of the strongest TATA box sequences in *S. cerevisiae*, while the LEU2 TATA box sequence, TATATA[TA], is also considered to be a strong sequence for TFIID binding, albeit to a weaker extent than the TEF TATA box [204]. Expression from these hybrid promoters was strong (Figure 2.4A, B, C). The POX2 TATA box sequence, TATACTTATATA, is not prevalent or highly uncommon in *S. cerevisiae* promoters; however, in *Y. lipolytica*, we observed strong expression from this hybrid promoter as well. When these TATA boxes were applied to the native POX2 promoter, we observed small changes in expression strength strongly suggesting that the strong UAS elements are predominantly important to confer strong expression (Figure 2.4D).



TATA box	SEQUENCE	STRENGTH
TEF1	TATAAAAG	+++
LEU2	TATATATA	++
POX2	**TATA**CTTATATACC	+++
POX2*	TATATACC	+

Figure 2.4. TATA box sequence impacts promoter strength. (A) Schematic of 8 bp TATA box mutations in the TEF (purple) and POX2 (lilac) hybrid core promoters. The consensus TEF TATA box, TATAAAAG, was substituted with lower affinity binding TATA boxes from the LEU2 and POX2 promoters. Furthermore, the POX2* TATA box was replaced with the LEU2 and the consensus TEF TATA box. (B) Qualitative representation of fluorescence strength from hybrid core promoters containing different core TATA boxes. (C) Fluorescence profiles for substitutions of the TATA boxes in the hybrid core promoters of POX2 and TEF. (D) Fluorescence profiles for when the TATA box from the endogenous POX2 (2147 bp) promoter was substituted with the TEF and LEU2 TATA boxes. * Indicates that a weaker, truncated version of the fully functional POX2 TATA box (TATACTTATATACC) was used in these experiments to create a series of hybrid promoter constructs with varying degrees of strength.

TEF Proximal Promoter Sequence Enhances Expression of Engineered Promoters

A comparison of the oleic acid-induced expression between the hybrid UAS1B8-TEF (136 bp) and UAS1B8-TEF (111 bp) transformants revealed that the UAS1B8-TEF (136 bp) showed a significant increase in fluorescence (Figure 2.5B). The difference between the two constructs is a 26 bp fragment adjacent to the 5′ end of the TATA box in the TEF core promoter. We wanted to understand if this proximal sequence could be used in a modular fashion for transcriptional enhancement in other TATA box containing promoters. We placed the TEF proximal sequence upstream of the hybrid POX2 core promoter containing a fully functional TATA box, UAS1B8-POX2 (TEF proximal) (Figure 2.5A). An insignificant increase in fluorescence was observed from this engineered promoter in comparison to the UAS1B8-POX2 (55 bp) construct without the TEF proximal sequence (Figure 2.5C), suggesting proximal sequences affect transcription in a nonmodular manner. The exact mechanism for how proximal sequences can enhance transcription is not well-understood and is likely context-dependent. To investigate this, we constructed UAS1B8 hybrid promoters containing the TEF proximal sequence upstream of TEF, LEU2, or POX2 core promoters. This data shows that the relative increase in transcriptional activity due to the TEF proximal sequence was insignificant for POX2 core promoters, higher in LEU2 core promoters, and highest in TEF core promoters (Figure 2.6A). Additionally, we showed that the TEF proximal sequence continues to exert no significant transcriptional enhancement when paired with the POX2 core promoter regardless of TATA sequence (Figure 2.6B), suggesting that proximal sequences interact with the core promoter sequence downstream of the TATA sequence. It is important to note that the effect of TATA sequence on GFP expression was preserved.

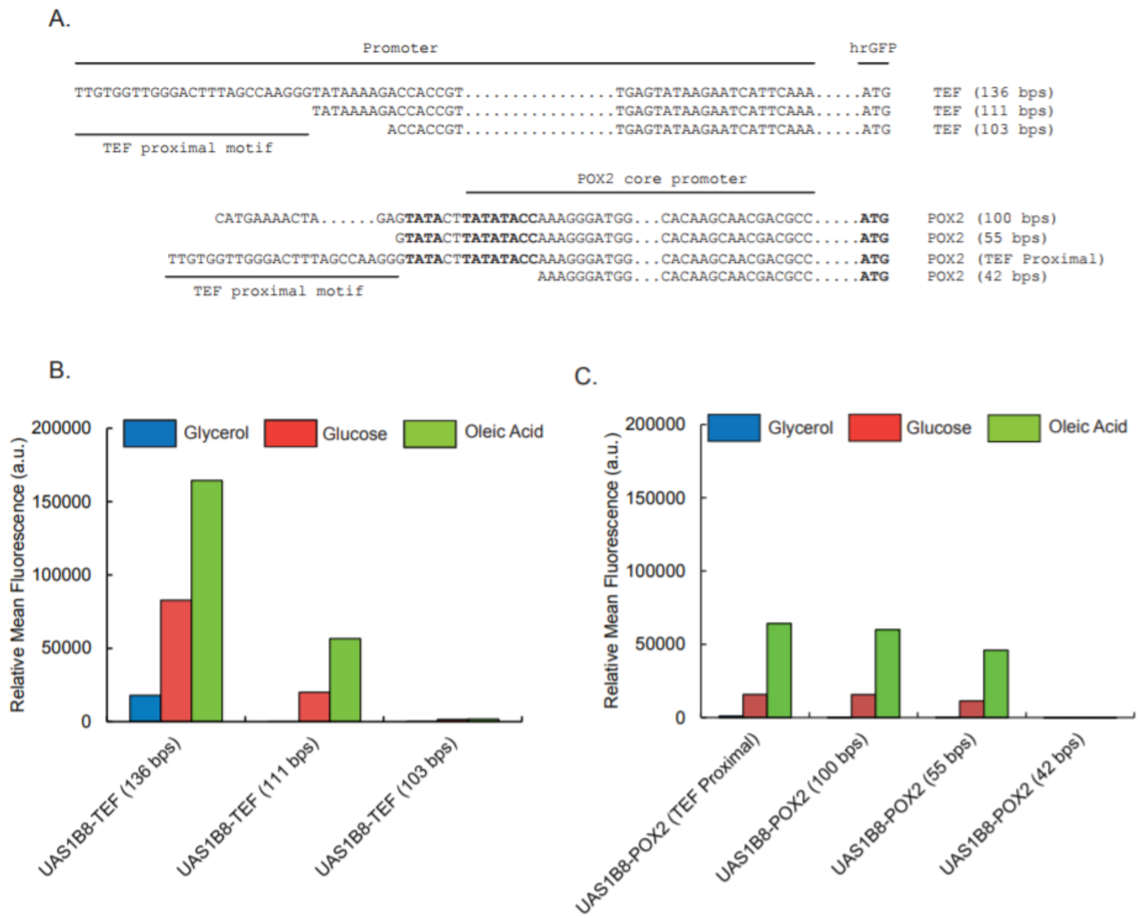


Figure 2.5. Proximal promoter sequences impact promoter strength but are not modular. (A) Schematic of TEF minimal core promoter with proximal promoter sequences adjacent to the core promoter. The schematic also shows insertion of the TEF proximal promoter motif directly upstream of the POX2 core promoter (UAS1B8-POX2 (TEF Proximal)). (B) Comparison of hrGFP fluorescence from TEF promoter including the proximal promoter sequence, including just the core promoter, and truncated past the TATA box. (C) Comparison of hrGFP fluorescence from POX2 core promoter TEF and native proximal sequence (TEF proximal and 100 bp) with core promoter (55 bp) and truncation (42 bp) that lacks the TATA box

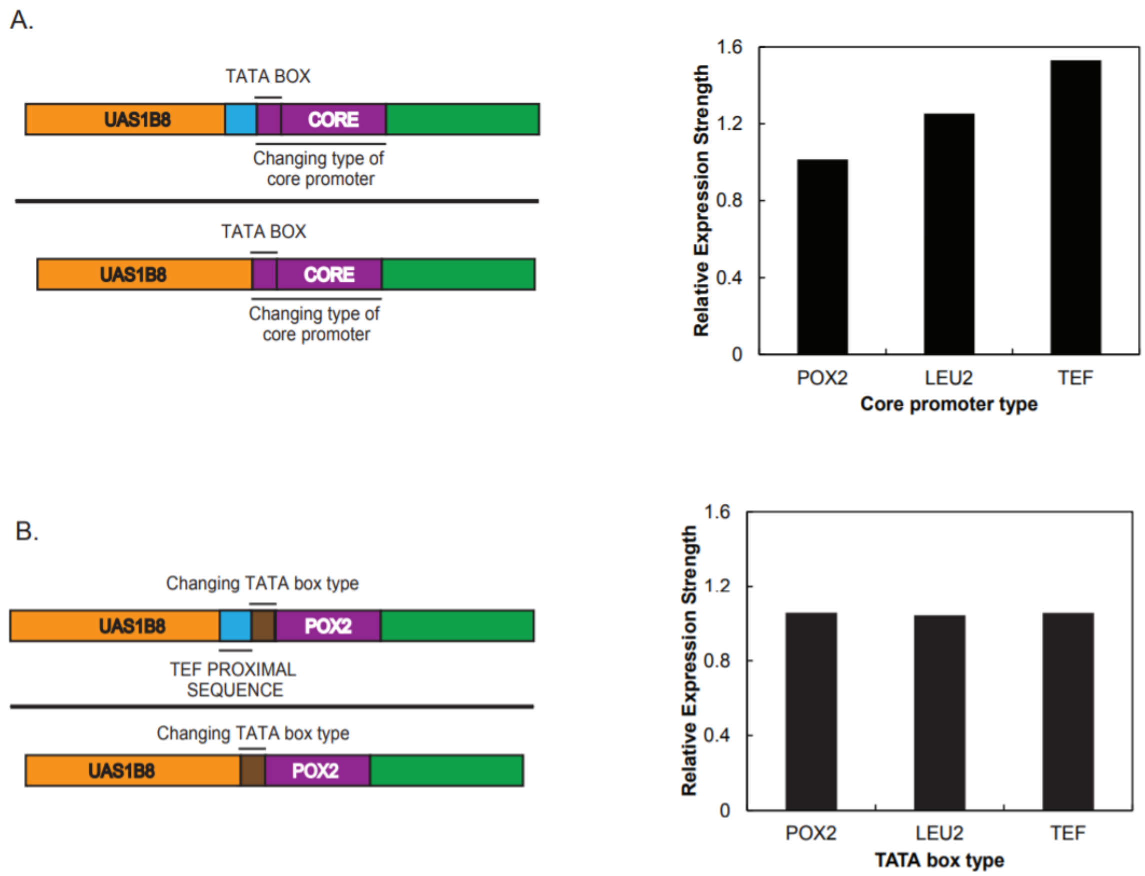


Figure 2.6. Modularity of the TEF proximal sequence in different promoter environments. (A) Comparison of hrGFP fluorescence from UAS1B8 hybrid promoters containing the TEF, LEU2, or POX2 core promoter with and without the TEF proximal sequence shows different levels of transcription enhancement based on core promoter sequence. (B) Comparison of hrGFP fluorescence from UAS1B8 hybrid promoters with POX2 core promoter sequence with and without the TEF proximal sequence shows that changing the TATA sequence does not alter the interaction between TEF proximal sequence and core promoter sequence.

Utility of the TATA Box Mutations in Different Episomal and Genomic Contexts

We investigated the utility of the TATA elements in transcriptional regulation of RNA polymerase II promoters. In hybrid promoters with three tandem repeats of UAS(TEF) and a TEF core promoter, the effect of TATA box sequences from TEF, LEU2, and POX2 followed the trend observed in hybrid promoters using UAS1B8 from

XPR2 (Figure 2.7A). Regardless of the type of UAS element used, the strongest expression levels were observed from the TEF TATA consensus box, followed by the LEU2 TATA box and the POX2 TATA box. The relative expression strengths among the TATA box modifications translate well across the UAS(TEF)3 and UAS1B8 elements, with the TEF hybrid core promoter containing the canonical TATAAAA box showing an approximately greater than 2-fold increase in comparison to the TEF hybrid core promoter with the LEU2 TATA motif. We also observed that the absolute promoter strength dropped approximately 2-fold when integrated into the genome, consistent with previous work showing lower expression of integrated promoters [205]. The relative differences in strength among the hybrid promoters with different TATA boxes follow the same trend in expression observed in UAS1B8 an UASTEF3 containing hybrid promoters. The TEF core promoter containing the TEF TATA box is still the strongest expressing system, followed by the LEU2 TATA box and the POX2 TATA box (Figure 2.7B). The decrease in promoter strength upon integration is consistent with additional repression of promoter strength due to genomic context.

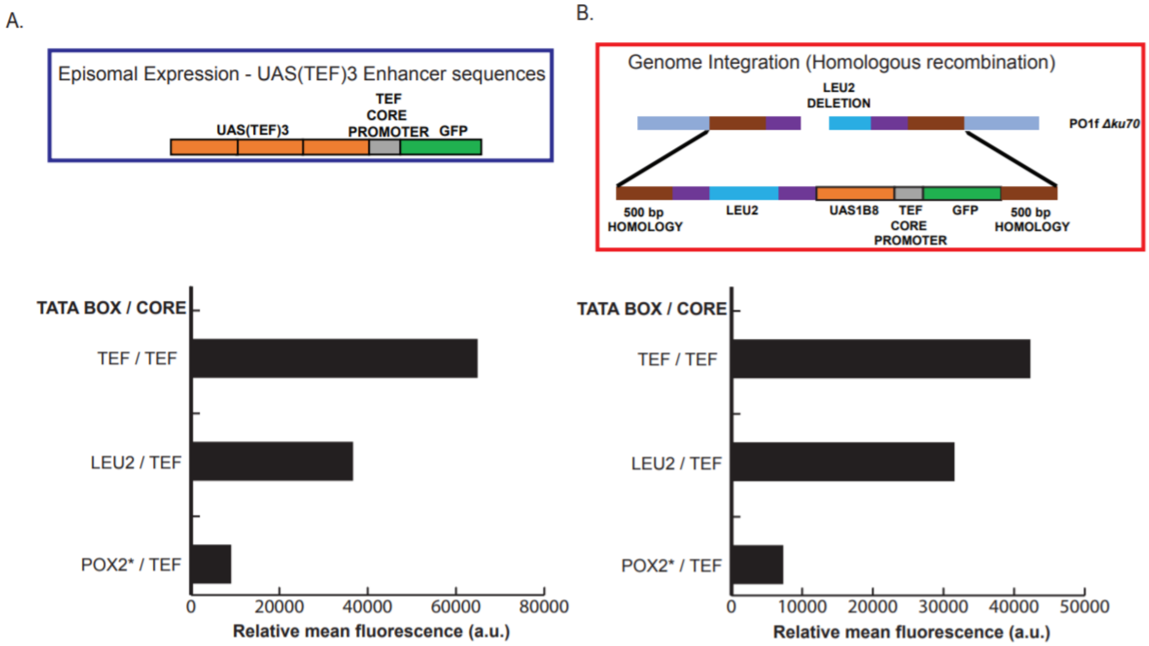


Figure 2.7. Modularity of the TATA box in different promoter environments. (A) hrGFP fluorescence from UAS(TEF)3 hybrid promoter containing TEF core promoter with either TEF, LEU2, or POX2* TATA box shows the same trend in promoter strength compared to identical episomal expression with UAS1B8 hybrid promoters. (B) hrGFP fluorescence from a UAS1B8 hybrid promoter containing TEF core promoter with either TEF, LEU2, or POX2* TATA box shows the same trend in promoter when integrated into the genome at the leu2 locus.

Table 2.1. Summary of promoters built from different component sequences

UAS TYPE	PROXIMAL SEQUENCE	TATA SEQUENCE	CORE PROMOTER	EXPRESSION STRENGTH
TATA modifications in TEF hybrid promoters (Episomal Expression)				
UAS1B8	NONE	TEF	TEF	++++
UAS1B8	NONE	LEU2	TEF	+++
UAS1B8	NONE	POX2*	TEF	++
UAS(TEF)3	NONE	TEF	TEF	++++
UAS(TEF)3	NONE	LEU2	TEF	+++
UAS(TEF)3	NONE	POX2*	TEF	++
TATA modifications in TEF hybrid promoters (Genomic Expression)				
UAS1B8	NONE	TEF	TEF	+++
UAS1B8	NONE	LEU2	TEF	++
UAS1B8	NONE	POX2*	TEF	+
TATA modifications in POX2 hybrid promoters (Episomal Expression)				
UAS1B8	NONE	POX2	POX2	+++
UAS1B8	NONE	TEF	POX2	++++
UAS1B8	NONE	LEU2	POX2	+++
UAS1B8	NONE	POX2*	POX2	+
TATA modifications in native promoters (Episomal Expression)				
POX2	POX2	POX2	POX2	+
POX2	POX2	TEF	POX2	+
POX2	POX2	LEU2	POX2	+
CORE PROMOTER modifications in hybrid promoters				
UAS1B8	NONE	TEF	TEF	++++
UAS1B8	NONE	LEU2	LEU2	+++
UAS1B8	NONE	POX2	POX2	+++
PROXIMAL SEQUENCE modifications in hybrid promoters				
UAS1B8	TEF	TEF	TEF	+++++
UAS1B8	TEF	LEU2	LEU2	++++
UAS1B8	TEF	POX2	POX2	+++
UAS1B8	TEF	LEU2	POX2	+++
UAS1B8	TEF	TEF	POX2	++++
Comparison of hybrid promoters				
UAS1B8	NONE	TEF	TEF	++++
UAS1B8	NONE	POX2	POX2	+++
UAS1B8	NONE	LEU2	LEU2	+++
UAS1B8	NONE	PAT1	PAT1	++

Discussion and Conclusion

Table 2.1 summarizes the library of promoters with varying degrees of expression that can be engineering by investigating promoter architecture. There are several components that can be used to fine-tune promoter strength and responsiveness. This chapter explores the contribution from each of these promoters to transcription.

The role of UASs sequences to promoter activation is studied in the context proximity to the core promoter and responsiveness to carbon. The native POX2 promoter is the most commonly described oleic acid inducible promoter in *Y. lipolytica* [110]. The inducibility and expression strength of the endogenous POX2 promoter have been well-characterized, and it has been used to drive inducible heterologous protein expression in *Y. lipolytica* [206, 207]. Although the promoter is weak, there is no understanding of where these UAS elements exist on the native promoter. From truncations, we observed that most of the ORE sites exist in a ~1000 bp region on the POX2 promoter. The lack of concordance between truncation data and predicted ORE sites is consistent with the observations of Poopanitpan *et al*. that showed predicted *S. cerevisiae* OREs in *Y. lipolytica* were nonfunctional [208]. Instead, the genetics of *Y. lipolytica* appear to be more closely related to filamentous fungi such as *Aspergillus nidulans* instead of *S. cerevisiae*.

In eukaryotes, enhancer sequences are located several hundred to thousands of kilobases upstream from a transcriptional start site (TSS). These sequences are hypothesized to recruit TFs to the TSS through a looping mechanism, suggesting that their position relative to the TSS is important for function. Tandem repeats of upstream activating sequences (UAS1B) from the XPR2 promoter have previously been used to create hybrid promoters (Figure 2.1B) that amplify downstream minimal promoter transcriptional activity [200]. Placing 8 tandem UAS1B sequences upstream of the POX2 truncation showed that as these sequences moved further away from the core promoter, the expression strength became weaker, elucidating that there is a distance dependence of UAS sequences for transcriptional activation. Interesting, we observed an increase in expression between the UAS1B8-POX2 (513 bp) and UAS1B8-POX2 (1591 bp) which correlated well with the increase in the expression between POX2 (513 bp) and POX2 (1591 bp) further supporting the potential for OREs in this ∼1 kb POX2 promoter fragment. These experiments suggest that transcriptional factors from enhancer regions of different promoter systems can work cooperatively to amplify transcription.

In addition to amplifying transcription, the UAS1B sequence from the XPR2 promoter contains elements that are highly responsive to regulation by oleic acid, in contrast to the notion that they are not regulated by media conditions. This observation underscores the need to be cautious about the possibility of unintended promoter regulation when engineering eukaryotic promoters.

Regulation at the core promoter level in eukaryotic systems can be complex as there are several elements within the core promoter that modulate promoter activity and strength. For example, in *S. cerevisiae* promoters, core promoters of both constitutive and regulated genes confer the highest activity when the sequence has a low G/C content, with T-rich motifs upstream of the TSS and A-rich motifs downstream of the TSS [209]. Such conditions are better suited for RNA polymerase II scanning downstream of the TATA box and selecting the most suitable TSS to initiate transcription. Our selection of TATA-containing core promoters in *Y. lipolytica* took into account sequences of varying degrees of G/C content, with the TEF1-α core promoter containing the lowest G/C content (~39%) and the PAT1 promoter having the highest percentage (~49%) (Figure 2.3A). Core promoters from POX2 and LEU2 both contain a G/C content that is approximately 43%. The TEF core promoter exhibits the strongest transcription while the PAT1 core promoter shows weakest expression but these promoters also contain distinctly different TATA elements that need to be considered. These results show that that promoter strength is a complex function of elements near core promoter sequence and distance of UAS sequences from the core promoter.

Distance from the TATA box and the core promoter strength can be used to fine-tune transcriptional levels, leading not only to multiple transcription levels but also to redundant transcriptional levels, as seen for UAS1B8-LEU2, UAS1B8-POX2, and UAS1BTEF (Figure 2.3B). However, to fully elucidate the role of initiator sequences in core promoters, we would need to test these sequences independently of the TATA boxes used as different TATA boxes can elicit different transcription strengths making it difficult to interpret the role of initiator sequences.

The TATA box is perhaps the most studied component of the eukaryotic promoter and is well-appreciated for its role determining the strength of eukaryotic core promoters, for example, in *S. cerevisiae* synthetic GAL4 enhanced promoter systems [210]. Two well-studied consensus core TATA boxes, TATAAA and TATATA, have previously been shown to have a high affinity to TFIID binding both in vitro and in vivo in *S. cerevisiae*, with the former being the strongest [204]. Furthermore, the nucleotide sequence that is immediately downstream TATA element has been shown to strongly affect transcription levels. Therefore, to test TATA elements, we considered 2 bp downstream of the 6 bp elements we considered TATA boxes based on comparison to *S. cerevisiae* sequences. The TATA motif from the TEF promoter showed strongest transcriptional strength while a mutation of one TATA box in the POX2 promoter that houses two TATA elements severely weakened transcription. This raises interesting questions about the most optimal TATA sequence for *Y. lipolytica*, which can be studied by increasing the sequence space from what we initially considered a TATA element in this study.

Modifications of the TATA box across hybrid promoters containing different UAS element types showed that expression from the different TATA boxes is independent of the UAS elements used to enhance expression. This suggests that promoter strength and therefore expression can be tuned using TATA box modifications independent of the type of enhancer used. To determine if the TATA box promoter tuning strategy would translate to engineered promoters integrated into the genome, hybrid UAS1B8-TEF core promoter cassettes containing the TATA box modifications were integrated into the genome at the leu2 locus and similar trends were observed, albeit, weaker than the hybrid promoter. This demonstrates the importance of strong UASs to amplify the variations of base transcription from TATA elements.

There was less success, however, in attaining predictable expression strength by translating proximal sequences from one core promoter to another. A potential reason for this lack of modularity is that the proximal sequence works cooperatively with the surrounding DNA sequences to provide a favorable nucleosome-depleted region near the TSS. In constitutive promoters, the DNA directly upstream of the TSS is typically nucleosome-depleted, whereas upstream of regulated promoters are more occupied [211]. A simple analysis of the POX2 and TEF core promoters clearly shows that, in addition to TATA box variability, there is a significant difference in the distribution and localization of the predicted yeast initiator sequences [212-214] and length of each core promoter (Supplementary Figure 2.1). We were unable to find any potential *S. cerevisiae* homologous transcriptional factors that could be associated with binding to the 26 bp proximal sequence.

This study has generated several new plasmids useful for episomal expression of genes in *Y. lipolytica*. We have identified a region of the native POX2 promoter that contains oleic acid responsive enhancers, we have shown that UAS1B8 repeats from XPR2 are sensitive to the carbon source, and we have examined the effect of the different core promoter and proximal promoter sequences on regulating transcript levels. We have also shown that by engineering different components of the promoter architecture a series of vectors with varying degrees of expression can be obtained (Table 2.1). Our study has shown that eukaryotic promoter engineering approaches can be focused at the enhancer region, at proximal sequences, the TATA box, and the core promoter, leading to a diverse and finely tunable level of gene expression. Future efforts to incorporate other responsive enhancers may lead to promoter engineering approaches that can finely tune both the transcription level as well as the environmental signal needed to activate transcription.

CHAPTER THREE

DEVELOPMENT OF A STRONG YET TUNABLE FATTY ACID INDUCIBLE PROMOTER IN *YARROWIA LIPOLYTICA*

Abstract

The engineering of *Yarrowia lipolytica* to accumulate lipids with high titers and productivities has been enabled with a handful of constitutive promoters for pathway engineering. However, the development of promoters that are both strong and lipid responsive could greatly benefit the bioproduction efficiency of lipid-derived oleochemicals in oleaginous yeast. In this study, we sought to engineer a fatty acid-regulated hybrid promoter for use in *Y. lipolytica*. We identified a 200 bp upstream regulatory sequence in the peroxisomal acyl CoA oxidase 2 (POX2) promoter. Further analysis of the promoter sequence revealed a regulatory sequence, that when used in tandem repeats, led to a 48-fold induction of gene expression relative to glucose and 4-fold higher than the native POX2 promoter. To date, this is the strongest inducible promoter reported in *Y. lipolytica*. Taken together, our results show that it is possible to engineer strong promoters that retain strong inducibility. These types of promoters will be useful in controlling metabolism and as fatty acid sensors.

† A version of this chapter is published in *Biotechnology Journal* **2017**, *12(10), 1-11,* with co-authors Ian Wheeldon, and Mark Blenner.

Introduction

Metabolic engineering of microorganisms for bioproduction greatly benefits from transcriptional control of native or heterologous genes [215, 216]. Precision control of gene expression enables rapid pathway optimization [217]. To that end, significant work has been put into developing libraries of promoters with predictable strength for a number of microorganisms [32, 215, 218]; however, in even well-characterized eukaryotes, such as *Saccharomyces cerevisiae*, the toolkit of promoters remains small compared to bacteria. Given the benefits of metabolic engineering in yeast [219, 220], additional focus on expanding the yeast promoter toolkit is warranted. Random mutagenesis of promoters has resulted in only modest improvements in promoter strength [215]. Similarly, it has been quite simple to identify loss-of-strength mutations in promoters [215, 218]. Rational approaches to increasing promoter strength have focused on hybrid promoters, where heuristically identified, well-defined DNA elements called upstream activating sequences (UASs) are placed in front of a core promoter sequence and can be used in a modular fashion to tune transcription strength [80-82, 199, 221]. While improving the strength of yeast promoters has been successful, the toolkit is currently lacking many options for inducible promoters.

Libraries of constitutive promoters with different strength can be used to tune gene expression; however, the expression level is statically set. Inducible promoters are useful genetic tools for metabolic engineering because, in addition to the benefits of being able to tune the transcriptional output, the expression level can be changed dynamically by metabolites or inducers. Inducible promoters can be especially important to improve the carbon flux efficiencies in a metabolic process by separating the growth phase from the production phase of a cell [222, 223]. Additionally, if the excess production of an enzyme or a product is toxic to the cell, it would be beneficial to have production switched off so cell growth would not be inhibited until the desired biomass is made [184, 186]. Inducible promoters used in metabolic engineering of yeast have largely been endogenous promoters responsive to small molecules such as copper, methionine, tryptophan, and phosphate [224-227]. Unfortunately, these promoters exhibit complex regulation patterns and are of modest strength at best. One notable exception is the Gal promoter from *S. cerevisiae*, which is both strongly repressed by glucose and strongly activated by galactose [224]. An upstream regulatory sequence (URS) from the GAL1-10 promoter has been used with a strong endogenous promoter, TDH, to confer galactose inducible control of the TDH promoter to create an inducible chimeric promoter [228]. Recently, a tryptophan inducible TDH promoter was engineered, using a mutant ARO80 transcription factor and tandem repeats of the URS_{ARO9}, resulting in promoters that are both strong and induced by tryptophan [229].

In recent years, there has been an increase in genetic tools for metabolic engineering in the oleaginous yeast, *Y. lipolytica*. The advancement of genetic tools [80, 82, 83, 230] has enabled metabolic engineering in this microorganism to produce a significant amount of FAs [89] and lipids [190, 231, 232] from different substrates [95, 233, 234]. Given the relative weakness of endogenous promoters, hybrid promoters have had great success in heterologous gene expression [82]. This success stems from the ability to tune expression strengths in hybrid systems. Whereas high expression is often desired [190], the option to tune transcriptional strength to lower levels can be critical to attaining optimal heterologous protein production [235]. An important set of hybrid promoters developed in *Y. lipolytica* contain the UAS1B element originally derived from the nitrogen and pH regulated XPR2 promoter [199, 200]. When used in tandem, the UAS1B elements provide enhancement in expression independent of nitrogen level and pH [81, 82, 199]. Another hybrid promoter used for metabolic engineering contains UAS_{TEF} from the constitutive promoter of the translation elongation factor 1-α (TEF1-α) gene [81]. Strong expression of genes is enabled by hybrid promoters; however, the current set of strong hybrid promoters is limited. Although tunable in expression strength, hybrid promoters sometimes elicit carbon source dependent regulation from the UAS, as was previously demonstrated using the POX2 UAS1B [80]. While this property might be used to create regulated promoters, the inducibility is neither rational nor tightly regulated. There has only been one report of a rationally designed inducible hybrid promoter in *Y. lipolytica*. This alkane inducible promoter was constructed from repeats of an alkane response element, but confers relatively weak expression when grown on alkanes [114, 115, 120, 236]. Therefore, development of hybrid promoters that are inducible to a biomolecule that is both readily metabolized in *Y. lipolytica* and acts as a precursor molecule for the synthesis of several other important biomolecules is desirable. Fatty acids are ideally suited for this task.

The native promoters from the peroxisomal acyl-CoA oxidase 2 (POX2), peroxisomal 3-ketoacyl-thiolase (POT1), and lipase 2 (LIP2) genes are commonly used as fatty acid inducible systems for metabolic engineering [117, 148, 206, 207, 237-239].

While the expression strength of the inducible promoter of POX2 has previously been characterized as strong [110, 148, 240], we show that the POX2 promoter is relatively weak compared to the beta-actin promoter and engineered hybrid promoters. In this work, we performed a series of truncations to identify fatty acid-responsive UASs and URSs that were combined to construct a library of hybrid promoters. A URS, called R1, was initially found to be critically important for fatty acid responsiveness and promoter strength. The amplifying effect of the R1 sequence was shown to be synergistic with a UAS, called A1. Tandem repeats of the R1 sequences from the POX2 promoter were used to create a strong fatty acid inducible system. This new hybrid promoter was approximately four-times the strength of the native POX2 promoter and smaller in size. This promoter had a 48-fold oleic acid induction of expression relative to glucose. The promoter was also induced by other fatty acids and lipids but remained strongly repressed in glycerol and glucose. Stronger inducible promoters were engineered by placing tandem R1 sequences upstream of either the TEF1-α promoter minimal promoter or the TEF1-α intron promoter, achieving expression levels greater than 10-fold higher than the native POX2 promoter. The hybrid promoter described here is the first engineered and strongest fatty acid inducible promoter system for *Y. lipolytica*.

Experimental Procedures

Strains & Culture Conditions

Plasmid propagation was performed using *Escherichia coli* DH10β competent cells (NEB). Transformations in *E. coli* were performed using standard methods [241].

strain PO1f (ATCC MYA-2613; *MATa leu2−270 ura3−302 xpr2−322 axp*) was purchased from ATCC. Transformations were done using the lithium acetate method as previously described with a minor modification for cell propagation after the transformation [80]. Briefly, following the heat shock step, the cells were mixed with 800 µL of 0.1 M lithium acetate, spun down at 6,000 x g for 2 minutes (4°C) and re-suspended in 100 µL of 0.1 M LiAc buffer prior to inoculating 2 mL cultures that were grown in 14 mL culture tubes. All cultures were grown at 215 rpm and 28 °C. Transformations and growth were performed in biological triplicates. was grown in Yeast Synthetic Complete media without leucine (YSC-LEU) comprised of 6.7 g/L yeast nitrogen base (YNB) without amino acids (Difco) and 0.69 g/L CSM-LEU (Sunrise Science Products). Carbon substrates used for characterization during the construction of the hybrid promoter were 2% (v/v) oleic acid (EMD Millipore) emulsified in 0.05% (v/v) Tween 80, and 2% (w/v) D-glucose. Other substrates that were used to characterize the substrate responsiveness of the hybrid promoter were all added at 30 mM, an equimolar concentration equivalent to 2% (v/v) oleic acid. These substrates include glycerol, linoleic acid (Alfa Aesar), triolein (Tokyo Chemical Industry), chicken fat (Animal Coproducts Research and Education Center, Clemson University) and n-decane. In cases where two carbon substrates were used, each was added at a final concentration of 15 mM. The hydrophobic substrates in the dual carbon experiments were emulsified with Tween 80 at a final concentration of 0.05% (v/v).

Promoter and Plasmid Design

All primers used to create the hybrid promoters designed in this study are presented in Supplementary Table 3.1. Unless otherwise stated, sequence and ligation independent cloning (SLIC) [201, 242] was used for cloning, and all PCRs were performed from genomic DNA. The base vector used for cloning was pSL16-CEN1-UAS1B8-POX2(100 bp)-hrGFP. [80] Our previous work identified a large region of the POX2 promoter that was important for POX2 promoter fatty acid responsiveness (-1590 bp to -513 bp) [80]. Within this region, we made periodic truncations using the primer pairs described in Section 1 of Table 3.1. The truncated POX2 promoter PCR fragments were inserted between the SphI/AscI restriction sites using SLIC. Hybrid promoter systems containing $UAS1_{POX2}$ and $UAS2_{POX2}$ (A1A2), $UAS2_{POX2}$ and $UAS3_{POX2}$ (A2A3) and $UAS1_{POX2}$ and $UAS3_{POX2}$ (A1A3) were constructed by performing a three-piece SLIC with each UAS and the base vector as described in Section 2 of Table 3.1. To construct A1A2A3, the A2 fragment was PCR amplified and inserted into the A1A3 vector at the MfeI site, as shown in Section 3 of Table 3.1. Inserts A1R1 and A1R1A2R2 were PCR amplified from genomic DNA and inserted in the base vector A1A3 digested with AvrII / SphI. Promoters containing $URS1_{POX2}$ (R1) and $URS2_{POX2}$ (R2) were constructed by PCR and cloned in between the SphI and AscI sites of the base vector (Section 4 of Table 3.1). The R1 sequence was inserted between into A2A3 and A1A3 respectively using the MfeI site (Sections 5 and 6 of Table 3.1). Additional R1 sequences were inserted to make A1(R1x2)A3 and A1(R1x3)A3 (Section 7 and 8 of Table 3.1). Construction of the R1A3 plasmid required the R1A3 insert be PCR amplified from the

A1R1A3 vector and inserted into the A1R1A3 base vector digested with AvrII / SphI. A1R1A1R1A3 was created by PCR amplifying the A1R1 sequence from gDNA and inserting it into the A1R1A3 vector digested with MfeI. To create the fatty acid inducible hybrid promoter containing the TEF(136) (Section 9 of Table 3.1), TEF(136)-hrGFP was PCRed from a previously constructed vector [243] and the A3 fragment was PCRed from the POX2 promoter. These fragments were inserted into a MfeI / AscI-digested hybrid fatty acid vector (Section 8 of Table 3.1). Construction of the fatty acid inducible promoter containing the TEF-intron utilized the same strategy instead, in this case, the TEF-intron-hrGFP vector had to first be constructed as described in Sections 10 and 11 of Table 3.1. The vector constructed in Section 11 of Table 3.1 was then used as the template to PCR the TEF-intron-hrGFP.

Table 3.1. Detailed list of vectors and primers used to construct hybrid promoters tested in this study.

	Starting Vector	Restriction Enzymes	Primer Pair(s)	Final Vector(s)
1	pSL16-UAS1B8-	SphI / AscI	F1 / R1	POX2(1590 bp)-hrGFP
	POX2(100 bp)-hrGFP		F2 / R1	POX2(1390 bp)-hrGFP
			F3 / R1	POX2(1190 bp)-hrGFP
			F4 / R1	POX2(990 bp)-hrGFP
			F5 / R1	POX2(790 bp)-hrGFP
			F6 / R1	POX2(540 bp)-hrGFP
			F7 / R1	POX2(513 bp)-hrGFP
			F8 / R1	POX2(430 bp)-hrGFP
			F9 / R1	POX2(100 bp)-hrGFP
2	pSL16-UAS1B8-	XmaI / SphI	F10 / R2	A1A2-POX2(100bp)-
	POX2(100 bp)-hrGFP		F11 / R3	hrGFP or
			F12 / R4	A1A2
			F13 / R5	A2A3-POX2(100bp)-
			F10 / R6	hrGFP or
			F11 / R5	A2A3
				A1A3-POX2(100bp)-
				hrGFP or
				A1A3
3	A1A3	MfeI	F11 / R4	A1A2A3-POX2(100bp)-
				hrGFP or A1A2A3
4	A1A3	AvrII / SphI	F10 / R11	A1R1-POX2(100bp)-
				hrGFP or
				A1R1
5	A1A3	AvrII / SphI	F10 / R12	A1R1A2R2-
				POX2(100bp)-hrGFP or
				A1R1A2R2
6	pSL16-UAS1B8-	SphI / AscI	F10 / R3	A1R1A2-POX2(100bp)-
				hrGFP or A1R1A2
	POX2(100bp)-hrGFP		F12 / R5	A2R2A3-POX2(100bp)-
				hrGFP or A2R2A3
7	A2A3	MfeI	F15/ R7	A2R1A3-POX2(100bp)-
				hrGFP or A2R1A3
8	A1A3	MfeI	F16 / R7	A1R1A3-POX2(100bp)-
				hrGFP or A1R1A3
9	A1A3	MfeI	F22 / R13	A1R1A1R1A3-Pox2(100
				bp)-hrGFP or
				A1R1A1R1A3
9	A1R1A3	MfeI	F17 / R7	A1(R1x2)A3-
				POX2(100bp)-hrGFP or

				A1(R1x2)A3
10	A1R1A3	AvrII / SphI	F21/ R5	R1A3-POX2(100bp)-hrGFP or R1A3
11	A1(R1x2)A3	MfeI	F17 / R7	A1(R1x3)A3-POX2(100bp)-hrGFP or A1(R1x3)A3
12	A1(R1x3)A3	MfeI / AscI	F18 / R8 F19 / R9	A1(R1x3)A3-TEF(136bp)-hrGFP
13	pSL16-UAS1B8-TEF(136)-hrGFP	PstI / NheI	gBlock®	pSL16-UAS1B8-TEF-intron
14	pSL16-UAS1B8-TEF-intron	BtgZI	F20 / R10	pSL16-UAS1B8-TEF-intron-hrGFP
15	A1(R1x3)A3-POX2(100bp)-hrGFP or A1(R1x3)A3	MfeI / NheI	F18 / R8 F19 / R9	A1(R1x3)A3-TEF-intron-hrGFP

RNA Extraction and Quantitative PCR

RNA extractions were performed on cell cultures that were grown until mid-exponential phase. Prior to RNA extraction, all cell cultures were normalized to an OD600 of 10 and 1 ml was used for the extraction using the Omega Biotek RNA Extraction Kit with the optional DNaseI digestion step. qPCR was done using the CFX Manager real-time machine from Bio-Rad. In accordance with MIQE guidelines, standard curves for quantification of hrGFP and beta-actin were created and the efficiency of each primer pair was calculated. Two housekeeping genes, beta-actin and TEF1-α were used as reference genes to initially validate analysis method. An equal mass of RNA from each of the samples was loaded. Protocols for qPCR conditions are described in protocols provided by the qPCR kit supplier. A relative quantification method was used to determine GFP expression. Standard curves were used to calculate copy numbers for the above-mentioned genes taking into account priming efficiency. The ratio of GFP mRNA copy number to beta-actin mRNA copy number was used to quantify changes in expression strength for the different POX2 promoter truncations. qPCR was used to measure GFP expression instead of flow cytometry because of its higher sensitivity.

GFP Fluorescence Analysis

During the development of the POX2 promoter, fluorescence spectroscopy with the Biotek Synergy MX fluorescence spectrophotometer was the method of choice for characterization of promoter strength. Cells grown in glucose were spun down at 6000 x g (4 °C) for 2 minutes and re-suspended in 0.1 M Phosphate Buffered Saline (PBS) (pH 7) while cell cultures from oleic acid were spun down, washed once with 0.1 M PBS containing 5% (v/v) Tween 80 and re-suspended in 0.1 M PBS. All cell cultures were grown for 36 hours, between mid and late exponential phase, where OD600 values across the samples were similar. The harvested cells were placed in 96 well plates and serial dilutions of the cells were performed to obtain an average fluorescence (Ex. 485 nm and Em. 510 nm). Serial dilutions were performed to obtain a fluorescence value in the linear range of detection without changing the gain for each experiment. The same gain was used for all measurements. The BD Accuri® C6 Flow Cytometer was used for promoter characterization with different carbon sources. In all flow cytometry measurements, 20,000 single cell events were counted and fluorescence was measured using the GFP channel. VirtualGain® was used to normalize the gain across all samples post-analysis. Fluorescent cell populations were gated and the same gate was used across all samples analyzed in each day. To obtain the specific mean fluorescence, the mean fluorescence of the non-fluorescent cells was subtracted from the mean fluorescence of the gated fluorescent cells.

Results

Identification of Fatty Acid Responsive Upstream Sequences in the POX2 Promoter

The most common approach for identifying UASs and URSs by measuring expression strength determined by a reporter gene from truncated promoters [81, 221, 244-246]. Such a description of the promoter architecture provides information about important DNA sequences in the promoter but leaves unanswered questions about the nature of these sequences and how they contribute to gene regulation patterns. Nevertheless, in systems such as *Y. lipolytica*, our lack of understanding of gene regulation prevents a more informed promoter engineering strategy *a priori*.

We previously made truncations to the POX2 promoter and identified a large region upstream (1590 bp) of the POX2 gene that is required for measurable transcription in glucose-free oleic acid media [80]. As a result, we chose to make truncations from the 5' end of the POX2(1590 bp) promoter at 200 bp intervals (Figure 3.2A). Our rationale for choosing these particular truncations was based on the identification of putative Por1p binding sites (Figure 3.1A). Por1p (YALI0D12628p) is a *Y. lipolytica* homolog of the fatty acid-responsive FarA transcriptional factor in *Aspergillus nidulans* [208, 247]. Homologs for *S. cerevisiae* fatty acid-responsive transcription factor Oaf1p do not exist in *Y. lipolytica* or other oleaginous yeast. Therefore, the well-studied *S. cerevisiae* regulatory system does not inform our work.

Quantitative PCR was used to detect changes in transcriptional profiles resulting from POX2 truncations (Figure 3.2B). In YSC-LEU media with oleic acid, there is a general decrease in expression strength with truncations moving towards to the core promoter. A four-fold decrease in mRNA copy number was observed between POX2 (1590 bp) and POX2 (1190 bp), suggesting the presence of an activating sequence we call $UAS1_{POX2}$ or A1. Another significant change in mRNA transcript was observed between POX2 (990 bp) and POX2 (540 bp); therefore, we call this region $UAS2_{POX2}$ or A2. A smaller drop in expression was observed between POX2 (438 bp) and the core promoter, POX2 (100 bp), defining a third activating sequence, $UAS3_{POX2}$ or A3. This truncation strategy also enabled us to identify regions in the native POX2 promoter where a single 200 bp truncation lead to a three-fold increase of transcriptional activity, as seen between the POX2 (1190 bp) and POX2 (990 bp) promoters. We call this upstream regulatory sequence $URS1_{POX2}$ or R1. A similar repressor sequence was observed for the truncation between POX2 (540 bp) and POX2 (438 bp), albeit to a weaker extent than R1, that we call $URS2_{POX2}$ or R2. Using this truncation strategy, we were able to map sequences in the native POX2 promoter that were potential fatty acid-responsive activating sequences and other sequences that appeared to behave as repressor sequences, although further investigation was required.

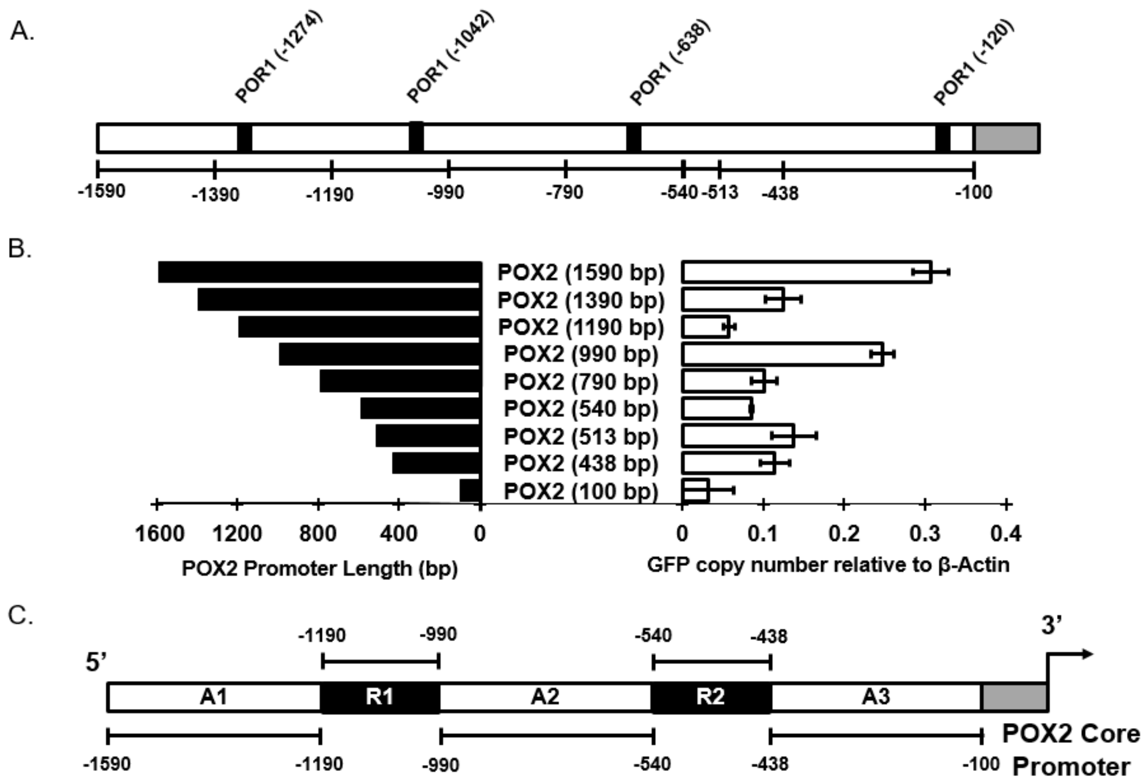


Figure 3.1. Identification of activating and regulatory sequences in the POX2 promoter. (A) Identification of hypothetical POR1p binding sites in the POX2 native promoter that were used to guide rational 5' truncations. (B) Schematic of systematic truncations were made to the native POX2 promoter to identify DNA fragments containing hypothetical fatty acid-responsive transcription factor binding sites. (C) Changes in truncated promoter strength were monitored with real-time PCR measurements of GFP mRNA relative to control beta-actin. Activating sequences (A1, A2, A3) were defined as those resulting in loss of transcriptional activity when truncated. Regulatory sequences (R1, R2) were defined as those resulting in either gain or constant transcriptional activity when truncated. The data are the average of mRNA copy number relative to beta-actin determined from biological triplicates. The error bars are the standard deviation of biological triplicates.

Discovery of a Fatty Acid Inducible Upstream Regulatory Sequence from POX2

By performing promoter truncations, 3 UASs and 2 URSs were identified in the native POX2 promoter. Our initial hypothesis was that by removing URSs, equivalent to combining the 3 UASs, would increase the fatty acid inducible expression. Furthermore, we reasoned the loss of UASs unimportant for fatty acid-regulated transcription would not greatly impact transcription. To test these hypotheses, we constructed promoters with various combinations of UAS in tandem (Figure 3.2A), expecting to obtain a promoter more strongly induced by oleic acid. When all three UASs were combined (A1A2A3), expression in YSC-LEU with oleic acid was diminished compared to the full native POX2 promoter (Figure 3.2B), suggesting the URSs might have a more complex role than could be predicted by the truncation experiments. This idea will be explored in section in the results following this section. When A3 was removed, resulting in promoter A1A2, transcription was further decreased, suggesting an important role for the A3 sequence. By restoring the A3 sequence and removing the A1 to create promoter A2A3, expression strength recovered to a value closer to that of the native POX2 promoter, further validating the importance of A3. When A2 was removed, resulting in promoter A1A3, gene expression was now comparable to the native POX2 promoter, suggesting that the A2 is dispensable and actually inhibitory in the context of these hybrid promoters. The A1A2 hybrid promoter that lacks the A3 sequences confers the weakest expression while A1A3, which is half the size of the native promoter, confers the strongest expression. These results suggest that the A1 and A3 sequence combine to provide an essential function for the A1A3 promoter. It should be noted that two tandem copies of the either A1 or A3 were tested and expression was significantly weaker than the native POX2 promoter (Supplementary Figure 3.1). Furthermore, these UAS sequences elicited a positional dependence as switching the order of A1A3 to A3A1 resulted in a significant drop in fluorescence (Supplementary Figure 3.1), suggesting again that substantial complexity exists in these systems. All promoters showed almost no expression when cells were cultured in YSC-LEU with glucose (Figure 3.2B).

From the initial 5’ truncation data, removal of either the R1 or R2 sequences conferred an increase in transcriptional activation suggesting their role as repressor sequences. However, because of the unexpected results from 5’ truncations, we created a series of 3’ truncations (Figure 3.2C). Promoter A1R1A2R2 has oleic acid-induced transcription similar to but lower than POX2 (Figure 3.2D), indicating the importance of A3. Further truncation to create A1R1A2 resulted in increased transcription, producing GFP similar to the POX2 promoter. This result is consistent with our previous findings that R2 is a regulatory sequence that acts as a repressor. The relative unimportance of the A2 sequence is further confirmed by oleic acid-induced expression from A1R1, which is similar to A1R1A2.

By comparing the results in Figure 2B and D, it became clear R1 also exhibited complex behaviors not predicted by the original truncation experiment. These data suggest that R1 can act as an activator when placed after the A1 sequence, contrary to the repression observed when the A1 sequence was removed from the native POX2 promoter during our truncation experiments (Figure 3.1). The same amplifying effect of the R1 sequence was not observed when the R2 sequence was placed in its natural position between the A2 and A3 sequences in promoter A2R2A3 (Supplementary Figure 3.1). The R1 sequence, when placed between the A2 and A3 sequences (A2R1A3) also did not confer strong activation of the promoter (Figure 3.2D), suggesting that there is a synergistic effect between the A1 and R1 sequence. The R1A3 promoter shows lower expression than A1R1, however, the difference is not statistically significant. The results here demonstrated the importance of the R1 sequence as an enhancer element when paired with the A1 and perhaps A3 sequence. This is also suggested by the A1R1 data in Figure 3.2D. To support this hypothesis, we created promoter A1R1A3, which combines the synergistic effects of A1R1 and the A3 sequence. This promoter was nearly three-fold stronger than the native POX2 promoter and serves as the foundation for building even stronger fatty acid inducible promoters.

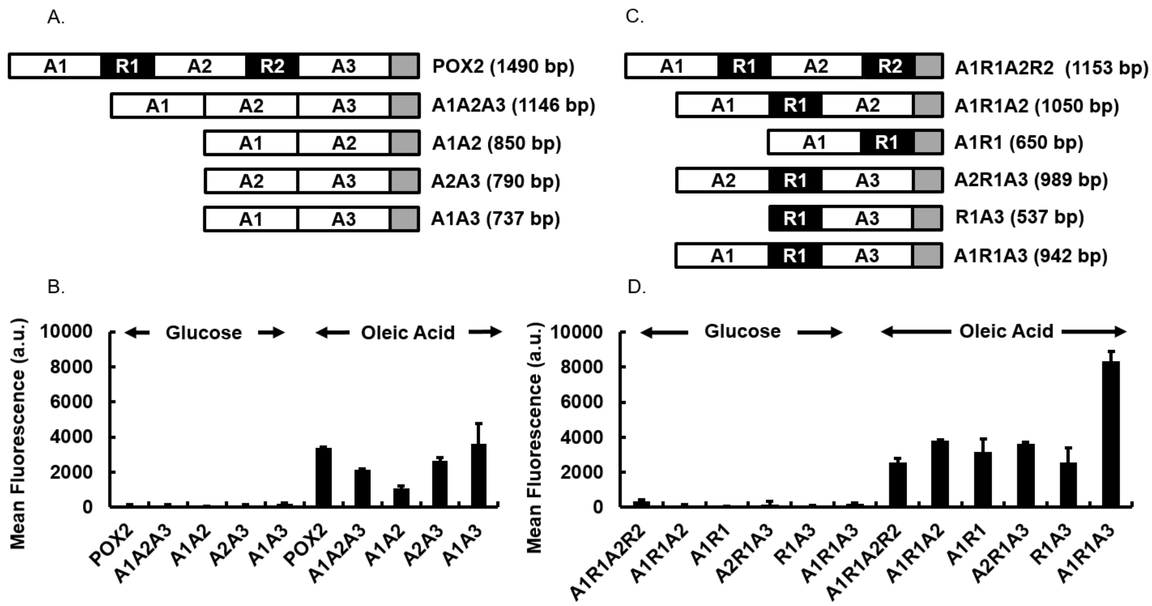


Figure 3.2. Discovery of the R1 UAS from the POX2 promoter. (A) New promoters were designed by combining parts of the POX2 promoter upstream of the POX2 core promoter. Activating (A1, A2, and A3) sequences were previously identified by 5' truncations. (B) Promoter strength is determined by expression of hrGFP and measured as mean fluorescence of an equal number of transformed cells grown in YSC-LEU with either glucose or oleic acid. Glucose samples are shown on the left and oleic acid on the right. A1 and A3 are important for oleic acid responsiveness, while A2 appears dispensable. (C) New promoters were designed by combining 5' and 3' truncations of the POX2 promoter upstream of the POX2 core promoter. Activating (A1, A2, and A3) sequences previously identified by 5' truncations. Regulatory (R1 and R2) sequences were previously defined by 5' truncations. (D) Promoter strength quantified by expression of hrGFP shows the R1 sequence results in strong expression in oleic acid media compared to glucose media. Glucose samples are shown on the left and oleic acid on the right. The R1 sequence appears to work synergistically with the A1 sequence. The data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviations of biological triplicates.

Engineering a Strong Fatty Acid Inducible Hybrid Promoter

Including the R1 in between the A1 and A3 sequences to form hybrid promoter A1R1A3 resulted in strong oleic acid activation compared to the native POX2 promoter and was a logical starting point to design a stronger fatty inducible promoter. Based on previous hybrid promoter work [81, 82, 248], we hypothesized that we could increase the oleic acid inducible transcriptional activation using tandem repeats of the R1 sequence. Therefore, hybrid promoters containing 0-3 copies of the R1 sequences were created (Figure 3.3A) and compared to the native POX2 promoter. The addition of each copy of the R1 sequence increased gene expression induced by oleic acid while the expression in glucose remained significantly and equally repressed, demonstrating the oleic acid inducible nature of the R1 sequence (Figure 3.3B). Addition of the first R1 sequence created a 2-fold increase in expression while subsequence additions of R1 sequences lead to about a 4-fold improvement in expression strength compared to the native POX2 promoter. Furthermore, we were able to improve the fold induction in oleic acid to 48-fold in the A1(R1x3)A3 hybrid promoter compared to the 19-fold induction in the native POX2 promoter (Figure 3.3C). Another promoter, (A1R1)x2A3 was created and had similar expression levels in oleic acid media as well as similar fold induction (Figure 3.3B, C). Given its larger size compared to A1(R1x3)A3, we chose to move forward with the smaller promoter. This new inducible hybrid promoter demonstrates the ability to engineer a tightly regulated oleic acid inducible switch and to tune the transcriptional output of the activated promoter.

The A1(R1x3)A3 promoter is already comparable in strength to strong hybrid promoters containing UAS1B8 and the POX2 core promoter (Supplementary Figure 3.2). We have previously shown that the strength of a hybrid promoter can be tuned by manipulating the modular elements of the promoter [80]. We demonstrate additional improvements to the A1(R1x3)A3 promoter by replacing the weaker POX2 core promoter with the stronger TEF(136) and TEF-intron core promoter (Supplementary Figure 3.2A), resulting in an additional two-fold and a three-fold increase in expression, respectively (Supplementary Figure 3.2B). Engineering the modular core promoter element allows us to tune the induction strength over a 10-fold range of expression; however, the increased expression in glucose led to a reduction in the fold induction (Supplementary Figure 3.2C). The core promoter is likely to exhibit some level of regulation mediated by regulatory TFs that bridge URS and the core promoter [249].

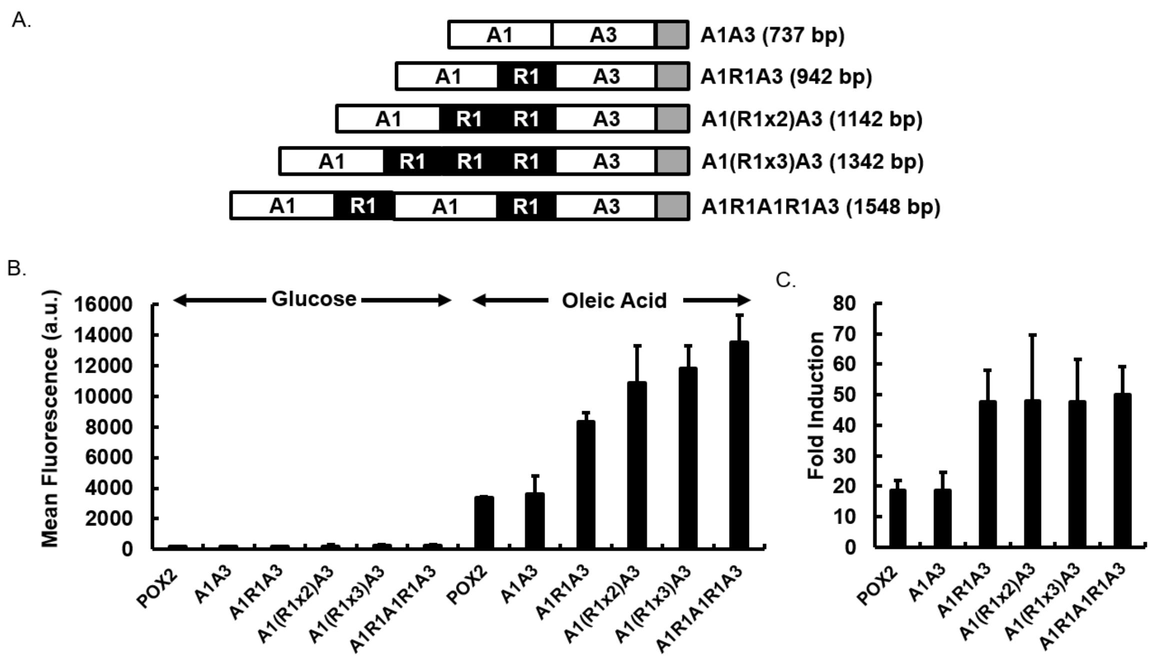


Figure 3.3. Engineering a strong fatty acid inducible hybrid promoter. (A) Schematic of hybrid promoters constructed with tandem repeats of the POX2 R1 sequence and A1R1 sequence. (B) Promoter strength is determined by expression of hrGFP and measured as mean fluorescence of an equal number of transformed cells. Glucose samples are shown on the left and oleic acid on the right. (C) Promoter induction using oleic acid as the carbon source relative to glucose as carbon source. In (B) and (C) the data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviation of biological triplicates.

<u>The A1(R1x3)A3 Hybrid Promoter is a Fatty Acid Sensor</u>

To better understand how different substrates affect transcription from the A1(R1x3)A3 hybrid promoter, we used flow cytometry to measure hrGFP expression controlled by A1(R1x3)A3 and compared it to the native POX2 promoter. Figure 4A shows that while both promoters were activated by various fatty acids (oleic and linoleic acid) and lipids (triolein and chicken fat), the inducible hybrid promoter A1(R1x3)A3 is consistently stronger than the native POX2 promoter. Linoleic acid elicited the highest transcriptional response for both promoters (Figure 3.4A). In all fatty acid and lipid substrates, the A1(R1x3)A3 promoter is two-to-four-fold stronger than the POX2 promoter (Figure 3.4B). Tween 80 was used as an emulsifier for fatty acid media and on its own does not elicit strong transcriptional activation.

When glucose, glycerol, or n-decane are used as the sole carbon source, the hybrid promoter remains strongly repressed, with only basal transcriptional activity similar to the native POX2 promoter. In media containing both glucose and oleic acid, the native POX2 promoter remained repressed, while the A1(R1x3)A3 hybrid promoter was strongly activated, suggesting the hybrid promoter is not catabolite repressed. Interestingly, in media containing both glycerol and oleic acid, both the native POX2 promoter and the A1(R1x3)A3 hybrid promoter were more strongly activated than oleic acid alone. The mechanism underlying this synergy remains unclear, however, a similar behavior was reported for the Lip2 promoter [117].

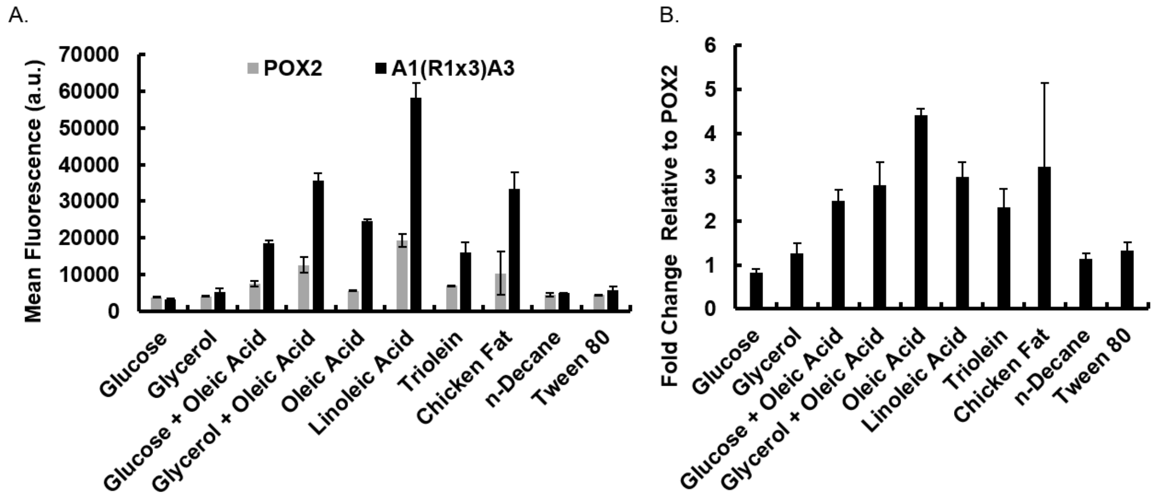


Figure 3.4. Substrate responsive induction of the A1(R1x3)A3 hybrid promoter. (A) Promoter strength was measured by hrGFP expression using flow cytometry. The A1(R1x3)A3 hybrid promoter is most strongly induced by linoleic acid and is strongly repressed by glucose and glycerol. Interestingly a combination of glycerol and oleic acid synergistically activated the hybrid promoter. (B) Fatty acids and mixtures of fatty acids with other carbon sources more strongly activate the hybrid promoter A1(R1x3)A3 compared to the native POX2 promoter. In (A) and (B), the data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviation of biological triplicates.

Fatty Acid Induction of the A1(R1x3)A3 Hybrid Promoter

In order to use the A1(R1x3)A3 promoter as an fatty acid inducible promoter, we grew cells to stationary phase in YSC-LEU glucose media and then induced the A1(R1x3)A3 promoter by titrating oleic acid into the media, at concentrations ranging from 0.25% (v/v) to 8%(v/v). The induction was measured using fluorescence spectroscopy. At all concentrations tested within this range, we observe nearly identical induction profiles of the hybrid promoter. This suggests the A1(R1x3)A3 promoter can be induced at oleic acid concentrations as low as 0.25% (v/v) during the stationary phase (Supplementary Figure 3.3).

Given the high sensitivity of the hybrid promoter, we determined if other fatty acids would similarly induce the hybrid promoter at this low concentration. We used 0.25% (v/v) of oleic (OA), linoleic (LA), arachidonic (ARA), and eicosapentaenoic (EPA) acids in YSC-LEU to induce the hybrid promoter in the stationary phase. The hybrid promoter is strongly and similarly induced by the different fatty acids (Figure 3.5A-D) while glucose did not induce GFP expression (Figure 3.5E). The fatty acid induction profiles for the A1(R1x3)A3 promoter, when induced at stationary phase with EPA, appears stronger than other fatty acids, however, the error associated with these stationary phase measurements makes the difference of low statistical confidence.

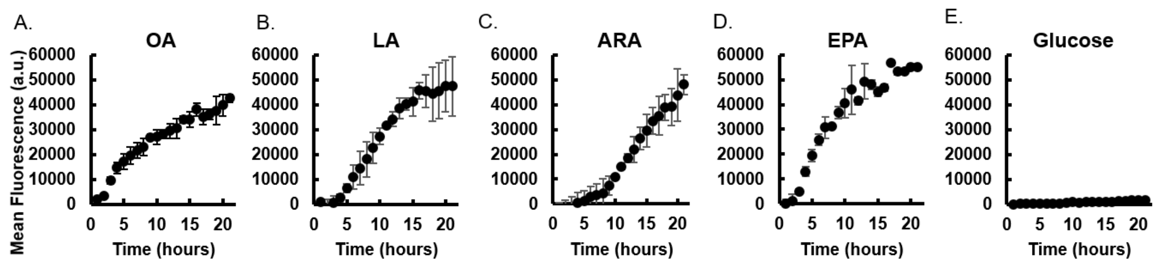


Figure 3.5. Inducibility of the A1(R1x3)A3 hybrid promoter by different fatty acids. Cells were grown to stationary phase with glucose and induced by the addition of 0.25% (v/v) of fatty acid. Induction was monitored over 20 hours using a fluorescence plate reader. Fatty acids used in this experiment include: (A) oleic acid (OA, 18:1), (B) linoleic acid (LA, 18:2), (C) arachidonic acid (ARA, 20:4 ω-6), (D) eicosapentaenoic acid (EPA, 20:5 ω-3) and (E) Glucose as a control shows no induction. The data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviation of biological triplicates.

Discussion and Conclusion

Our work has resulted in the development of a fatty acid inducible hybrid promoter for *Y. lipolytica*. UAS and URS sequences were initially identified using a rational truncation strategy; however, our subsequent experiments resulted in different conclusions about the roles of the A1 and R1 sequences when tested in isolation of the native POX2 promoter. Therefore, the truncation approach is useful in identifying functionally important sequences of promoters, but in this case, fails to correctly identify how these sequences will work in different contexts. This analysis showed that the R1 sequence contains a fatty acid response element and that it acts synergistically with the A1 sequence. Hybrid promoters with tandem repeats of the R1 sequence lead to increased transcriptional strength in oleic acid media (four-fold stronger than native POX2) while maintaining tight repression in glucose media (48-fold induced by oleic acid). These hybrid promoters are strongly activated by a variety of long chain fatty acid and lipids. Interestingly, the engineered hybrid promoter is not catabolite repressed in contrast to the native POX2 promoter and carbon metabolism promoters [227, 250-252].

The only comparable work in *Y. lipolytica* focuses on alkane responsive elements (AREs). Prior studies have identified a URS in the ALK1 gene (responsible for alkane oxidation in alkane metabolism) that contains an ARE that binds the Yas1p/Yas2p TFs [120]. A hybrid promoter containing three copies of the ARE1 sequence were placed upstream of the LEU2 core promoter, resulting in 6-fold activation on n-decane compared to glucose. By comparison, the A1(R1x3)A3 promoter is 48-fold activated by fatty acids compared to glucose. While differences in assay methods prevent a direct comparison of the strength of these two promoters, we have shown alkane promoters are significantly weaker than the beta-oxidation promoters (Supplementary Figure 3.4). This finding opens opportunities to engineer additional responsive hybrid promoter tools for *Y. lipolytica.*

Enabled by a deeper understanding of its genetics, there has been more work engineering strong inducible promoters in *S. cerevisiae*. For example, the most well studied inducible yeast promoter is the Gal1-Gal10 system. It exhibits remarkably low basal transcription in glucose-containing media and activated up to four orders of magnitude by galactose. These levels of regulation are determined by the combination of six repressing operator sites that overlap four Gal4p binding sites [253]. Analogously, further improvements to our fatty acid-responsive hybrid promoter may be possible once the TFs that bind to the R1 sequence of POX2 are identified.

We demonstrated that combining the hybrid promoter containing R1 repeats with other core promoters, including the TEF(136) and TEF-intron core promoters increased both the basal level expression and the induced expression. This result was expected in light of work combining different *S. cerevisiae* UAS sequences. For example, combining a constitutive UAS from CYC1 with the Gal1 or Gal10 promoter elements resulted in galactose regulated expression [228]. Stronger hybrid promoters were engineered by placing the UAS_{gal} upstream of weak core promoters (pLEU and pCYC) leading to glucose repression and galactose activation; however, when UAS_{gal} was placed upstream of strong core promoters (pTEF and pTDH3), a higher level of basal expression in glucose was observed [58]. This study also directly used individual Gal4p binding sites to further tune and enhance promoter regulation. More recently, ultra-strong and tryptophan regulated promoters were created by placing 5 UAS_{aro} sequences upstream of the TDH core promoter [229]. These promoters were 1.7-fold stronger than the TDH promoter and had 14-fold induction by tryptophan. Future efforts to increase promoter strength and maintain strong inducibility would benefit from an additional focus on upstream sequences, transcription factor binding sites, and less emphasis on strong core promoters.

UASs and URSs from native promoters have traditionally been identified through truncation studies [81, 200, 244]. While this method proves to be a solid foundation for identifying the parts needed for hybrid promoters, it fails to always capture the complexity of eukaryotic transcription regulation, which is controlled by the association of multiple transcriptional factors to their cognate binding sites [254, 255]. Because of the heuristic way in which UASs and URSs are identified, they are inherently subject to context-dependent behaviors. Our work demonstrates the difficulty encountered as a result of this disconnect. The UAS/URS sequences contain multiple and often overlapping transcription factor binding sites enabling higher strength transcription or regulation of transcription [228]. Unlike *S. cerevisiae*, there are few studies on *Y. lipolytica* TFs or transcription factor binding site motifs, and at least with respect to fatty acid metabolism, TFs in *S. cerevisiae* do not always have homologs in *Y. lipolytica*. To date, there has been only one transcription factor associated with fatty acid regulation. Deletion of the POR1 gene causes some growth defect on oleic acid and a reduction in POX2 mRNA expression [208]. As POR1p is a homolog of FarA from *A. nidulans*, POR1p may bind a similar DNA sequence. Putative POR1p binding sites were found in the R1 region, suggesting a significant role of POR1 in regulating the A1(R1x3)A3 promoter. A better understanding of TFs and their binding sites may lead to a more direct identification UAS/URS sequences and more rapid design of regulated hybrid promoters.

The A1(R1x3)A3 promoter is induced by a several different long chain fatty acids, including OA, LA, ARA, and EPA. Interestingly, stationary phase induction (Figure 3.5) of this promoter was nearly identical for OA, LA, and ARA, whereas when cells were grown on the fatty acids (Figure 3.4), LA more strongly induced the promoter. EPA caused the greatest induction of the hybrid promoter, which was unexpected since EPA is not synthesized by *Y. lipolytica*. When measuring induction during growth, the hybrid promoter was not repressed by glucose or glycerol when co-fed oleic acid, consistent with recent observations for the native Lip2 promoter; however, the PO1f strain used in our study does not co-utilize glucose and oleic, so the results shown by Sassi et al. [117] showing Lip2 promoter induction was strongest with a 40/60 mixture of glucose and oleic acid (w/w), are not likely to work in all strains. These works do suggest a potential strategy for fatty acid inducible gene expression when using glucose or glycerol as a substrate.

The fatty acid-regulated hybrid promoter created in this study represents a significant advance in the toolkit for engineering. We demonstrated the identification of a URS, R1, and its construction into a fatty acid-regulated promoter significantly stronger than the native POX2 promoter. We also showed that the hybrid promoter design can lead to the tuning of both the transcriptional output as well as the inducibility of the promoter. This promoter system is one of the strongest identified and is the strongest inducible promoter for *Y. lipolytica*. This regulated promoter has great promise for use as a sensor for strain engineering applications, for dynamic regulation of heterologous gene expression, or as an inducible promoter for toxic genes. We anticipate further development of regulated and strong promoters to expand the genetic engineering tools available for *Y. lipolytica*.

CHAPTER FOUR

USING DNA ACCESSIBILITY AS A ROBUST APPROACH TO ENGINEER NOVEL HYBRID INDUCIBLE PROMOTERS

Abstract

Inducible promoters are powerful genetic tools in metabolic engineering as they allow for an added layer of control of metabolic processes. In the yeast, *Yarrowia lipolytica*, these types of promoter systems are poorly characterized. Here, we investigated a novel mechanism to engineer hybrid inducible systems by measuring changes in DNA accessibility that can result from the interaction of regulatory transcription factors with DNA and chromatin structure. The Acyl CoA oxidase 2 promoter (POX2) in *Y. lipolytica* is one of the best understood inducible systems, activated by fatty acids and repressed by either glucose or glycerol. The DNA binding sites of the POR1p transcription factor were mapped to the native POX2 promoter. These binding sites were found at the edges of regions where changes in DNA accessibility was visible under induced and non-induced conditions. The functionally important binding sites on the POX2 promoter were elucidated via site-directed mutagenesis studies. The role of POR1p and CFU1p, both implicated in the regulation of beta-oxidation genes, were knocked out. The findings presented here provide new insight into the development of hybrid, tunable inducible promoters in a more robust and efficient manner.

[¥] This work will be included in a future publication with co-authors Scott Anglin, Sara Edgecomb, and Mark Blenner.

Introduction

Advancements in synthetic biology have revolutionized the field of microbial cell factories for sustainable chemical production. The ability to efficiently screen, sequence and annotate genomes of new microbes has enabled the shift from traditional hosts such as *Escherichia coli* and *Saccharomyces cerevisiae* to non-conventional microbes bearing unique native traits. The growth of *Trichosporon oleaginosus* on aromatics to accumulate lipids [70] or the ability of *Y. lipolytica* to tolerate many environmental factors, above all pH [256] or grow on hydrocarbons such as alkanes to produce lipids [257] are a few examples of superior features of non-conventional yeasts.

With recent success in developing genome editing tools, [83, 140, 230], engineering metabolic pathways and developing stable strains in *Y. lipolytica* has become easier. Making these pathways efficient requires tunable expression of enzymes to guide flux of metabolites. To date, only a handful of hybrid promoters exist to enable this process [80-82, 116, 127, 217]. The most commonly used hybrid promoters are constitutive providing one dimension of control that is expression strength. If we are to improve pathway efficiencies, more means to control enzyme expression by selectively inducing enzymes at different times during growth is important. This can be accomplished with inducible promoters.

Inducible promoter systems have gained interest as a “next-level” genetic tool that can be used to modulate expression strength and control the timing of expression. By doing so, these systems can significantly improve the efficiency of microbial engineering processes [258, 259]. In *Y. lipolytica*, tunable and inducible promoters are poorly understood because there is insufficient knowledge on promoters that demonstrate inducibility. Furthermore, transcription factors and mechanisms for regulation of such promoters are not known. Understanding how inducible promoters work could motivate rational engineering of novel metabolite and non-metabolite hybrid inducible promoter systems.

To date, strategies to engineer hybrid promoters have relied on heuristic methods such as truncating the promoter from either the 5’ or 3’ end and coupling it to a reporter gene coding for beta-galactosidase or green fluorescent protein to quantitatively measure the expression strength [81, 116, 122, 221]. Any loss or gain in reporter activity from truncations was used to identify the promoter regions as either an enhancer or repressor sequence. Heuristic approaches can fail for two reasons. First, there is the likelihood of truncating within the regulatory region of the promoter that can affect the affinity of a transcription factor to the binding domain. This approach also does not account for complex regulatory mechanisms typical of eukaryotic promoters resulting from the interaction of two or more transcription factor binding sites [260, 261]. Second, heuristic approaches do not account for DNA accessibility that results from dynamic changes in histone modifications to regulate gene expression, which is covered below.

In this chapter, we describe a new method for the development of next-generation inducible, yet tunable promoter systems by investigating changes in DNA accessibility of the promoter on the genome under conditions that induce or repress gene expression. Nucleosomes are a basic unit of DNA packaging in eukaryotes, made up of 147 bp of DNA wrapped by an octamer of core histones, H2A-H2B dimers flanking two core H3-

H4 histone tetramers [262, 263]. The final arrangement of eukaryotic DNA is a supercoiled structure called chromatin that is comprised of multiple nucleosomes separated by short linker fragments [264]. The role of nucleosomes in the genome has been well studied over the years. Structurally, nucleosomes bring about the first level of genomic compaction, which is 180 bps and then facilitates self-assembly into higher-order compaction of DNA inside the nucleosome [265, 266]. This structural arrangement protects DNA from damage and the positioning of nucleosomes plays a crucial role in transcriptional regulation and DNA replication [267-269].

DNA and histone methylation is known to influence nucleosome position and structure, thereby regulating transcription. CpG DNA methylation represses promoters by either recruiting transcriptional repressors to bind to the methyl CpG moiety [270] or inhibiting binding of transcriptional activators from binding to the promoter region [271]. Histone methylation of arginine (R) and lysine (K) residues play a key role in transcriptional regulation by altering chromatin organization. The methylated amino acids K4, K9, and R19 on histones H3 or H4 can either cause transcriptional activation of repression at the promoter of genes [272, 273]. In *S. cerevisiae*, H3K4, H3K36, and H3K79 methylation correlates with transcriptional activation [274]. Histone acetylation is also associated with transcriptional regulation. Remodeling of the PHO5 and PHO8 promoters in *S. cerevisiae* were initiated by acetylation of the H2B histone, which recruits ATP-dependent remodelers to evict the nucleosomes creating a nucleosome-free region (NFR) for transcriptional activation by PHO4 [275, 276]. Similarly, other histone modifications can cause nucleosome "sliding" or "loosening" to enable transcriptional activation (Figure 4.1) [277].

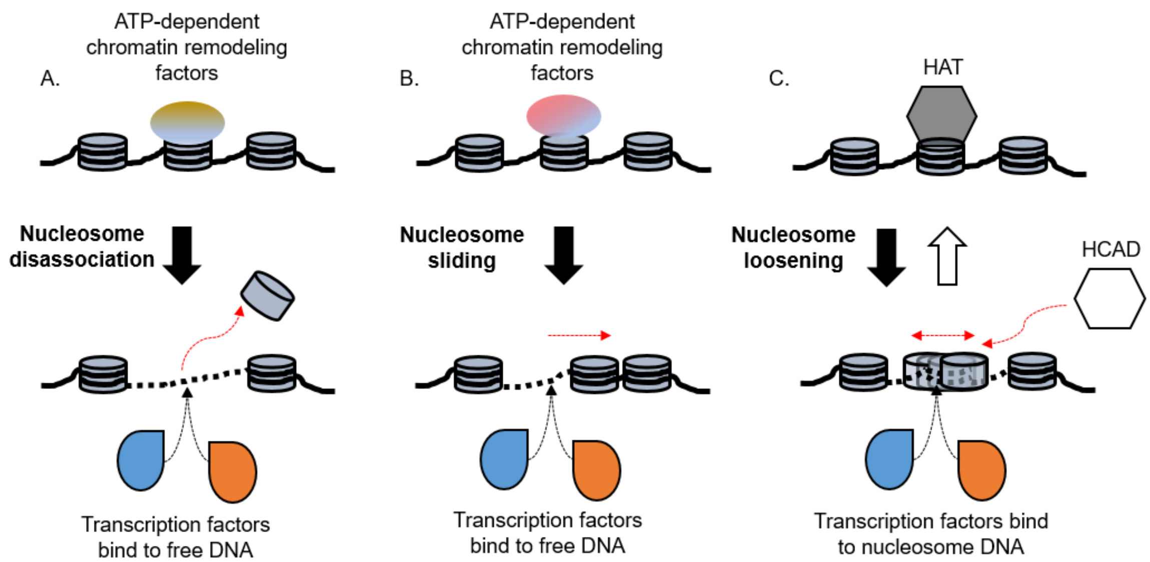


Figure 4.1. Three mechanisms by which nucleosome remodeling can occur to accommodate transcription factor binding. Mechanisms A and B rely on *trans* factors (TF) like ATP-consuming chromatin remodelers that cause either (A) complete nucleosome disassociation or (B) sliding events. In mechanism C, nucleosome structure is altered via histone-modifying enzymes. Acetylation of amino acid residues via histone acetyltransferases (HATs) increases the electronegative charge of the histone residue causing repulsion between DNA and histones, leading to a "loosening" effect of the nucleosome thereby making it more accessible to transcription factors. The process is reversible by histone deacetylases (HDACs) [277].

Nucleosome sliding is a well documented mechanism in *S. cerevisiae* [278-280]. There are essentially two models that have been proposed to explain how this occurs, the twist diffusion and the loop/ bulge propagation model [281]. In brief, the twist diffusion model explains a stepwise movement of a single base pair from the linker that shifts into the nucleosomal DNA space causing the nucleosomes to twist or untwist to accommodate the change [282]. Meanwhile, with the loop/ bulge propagation model, larger segments from the linker are transferred to the nucleosome DNA space creating a loop that causes the histone core to quickly diffuse to the opposite end to remain wrapped [283]. The detached nucleosome DNA can now be exposed to transcriptional regulators. This mechanism relies on the intrinsic dynamic nature of nucleosomes. Once the proteins bind to the DNA, it can control the level of nucleosome occupancy at their sides [284].

The affinity of transcription factors to DNA binding sites is another influence on transcription regulation. The PHO84 promoter of *S. cerevisiae* contains five copies of the PHO4p binding site dispersed across its promoter and not all motifs bear similar affinities. Mutational analysis of the 6 bp binding motifs revealed that only three motifs are important for transcriptional activation, but only two are necessary for full regulation of the *PHO84* gene [285]. The regions flanking PHO4p binding motifs affected its binding affinity. Therefore, even though some binding sites may be positioned in NFRs, if these sites have a low binding affinity, no transcriptional activation will be observed until the high-affinity sites are relieved from nucleosome compaction under phosphate starvation conditions. Figure 4.2 depicts a model for how chromatin may influence gene expression by differentially regulating the accessibility of the PHO4 binding sites on the promoter. Similar remodeling mechanisms have been shown in other *S. cerevisiae* promoters such as GAL1, SUC2, and CUP1 [286, 287].

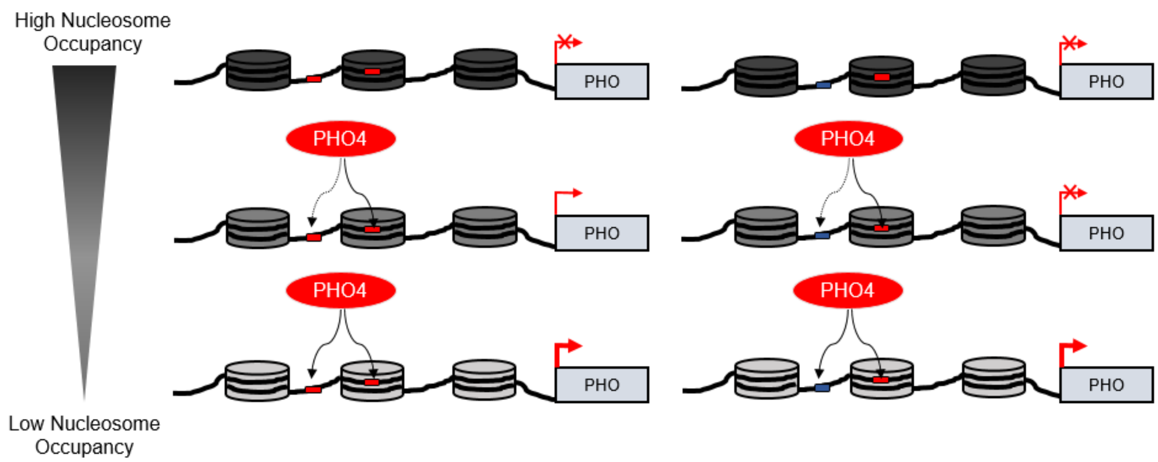


Figure 4.2. Model depicting transcription factor affinity and nucleosome positioning on differentially regulating PHO binding motifs [277]. Red arrows represent transcription and thickness of arrows represent the strength of transcription. **X** represents no transcription. As nucleosome occupancy reduces in and around transcription factor binding sites, high-affinity binding sites (red) become more accessible to PHO4 binding resulting in some transcription activation. The same is not true for low-affinity sites (blue) that require high levels of nucleosome depression to have a cooperative effect with high-affinity binding sites and elicit strong transcriptional activation.

As described above, each transcription factor can have an inherent affinity to the DNA and yet its binding probability be affected by competition with nucleosomes. Therefore, to understand how nucleosomes are affected by transcription factors, it would be ideal to study a system where all transcription factors regulating a promoter are known. Understanding the effect of specific transcription factors on nucleosome positioning can be leveraged to building hybrid promoters regulated via transcription factor-mediated nucleosome remodeling. This is one interesting application for next-level hybrid promoters to modulate gene expression, however in *Y. lipolytica*, this is not possible because there is no well-established promoter with known regulatory elements or an understanding of the mechanisms that dictate nucleosome positioning.

Another application for studying nucleosome positioning is to identify Upstream Activating Sequences (UASs) rapidly, promoting the development of hybrid promoters with tunable strength and inducibility. We hypothesize that UASs can be identified by observing nucleosome repositioning during transition from repression to induction. The *POX2* promoter in *Y. lipolytica* is a good platform to establish design principles since our previous work have already elucidated critical activator regions in the native promoter [116] that can be mapped to the nucleosome profile. The mapping suggested that the putative transcription factor binding sites from POR1p [208] align to regions adjacent to significant changes in nucleosome occupancy observed between oleic acid and glucose samples. This suggests some remodeling may be occurring in POX2 regulation. Of the four 6 bp binding sites in the POX2 promoter, three mapped near highly nucleosome occupied regions while the fourth site did not. Mutation to the putative POR1p binding motifs revealed that two of the three sites that mapped to occupied regions were crucial for full transcriptional activation. Knockout of POR1p and CFU1p, two transcriptional factors implicated in the regulation of beta-oxidation caused a significant loss in GFP expression from the native POX2 promoter. The POX2 promoter in the knockout strain, PO1f*Δpor1Δcfu1*, generally had higher DNA occupancy throughout the promoter. Although the profiles of DNA accessibility between the wild-type and knockout strain had similarity trends, the knockout strain had widened nucleosome coverage. The broadening of nucleosome coverage in oleic acid was more evident, particularly at the upstream-most site. This suggests that different mechanisms of regulation may govern transcription factor-promoter interactions. A comparison of the profiles between the wild-type and knockout strain in different carbon conditions revealed new regulatory regions that can be used to re-engineer the POX2 promoter, showing promising application as a strategy to develop other regulatable promoters in *Y. lipolytica*.

Experimental Methods

Chemicals and Reagents

All restriction enzymes used in cloning were purchased from New England Biolabs (NEB, Ispwich, MA) unless otherwise stated. Oligos designed for all experiments were synthesized by Eurofins Genomics or IDTDNA. Plasmid minipreps were performed using the Zyppy™ Plasmid miniprep kit (Zymo Research, Irvine, CA). PCR purifications and restriction digest purifications were performed using the DNA Clean and Concentrator kit (Zymo Research). Micrococcal nuclease solution (Thermo Scientific, Waltham, MA) was used to prepare mononucleosomal DNA from gDNA. Digestion of all proteins from cell extract was performed using Proteinase K (Thermo Scientific). A bicinchoninic assay (BCA) (Thermo Scientific) was used to quantify protein in the cell lysate. RNA digestions were performed using RNase A (Thermo Scientific). gDNA extraction was performed using the E. Z. N. A. Yeast DNA Kit (Omega Biotek, Norcross, GA).

Plasmids and DNA Cloning

Construction of the base plasmid consisting of the POX2 native promoter (1540 bp) fused to a GFP reporter gene that has been described elsewhere [80]. Mutations were made to the putative POR1p binding sites identified on native promoter as shown in

Figure 4.3. There are four putative binding sites labeled 1, 2, 3, and 4 that were found in the A1, R1, A2, and A3 regions identified in Chapter 3, respectively [116]. All cloning was performed using sequence and ligation independent cloning (SLIC) method unless otherwise stated. The base vector used was pSL16-POX2(1590bp)-hrGFP, harboring the native POX2 promoter, in this study referred to as pSL16-POX2-hrGFP. Table 4.1 summarizes the cloning strategy used to make all vectors containing individual and combinatorial mutations to the POR1p binding sites. Briefly, the base vector and all preceding constructs containing the desired mutations were first digested with AatII to linearize DNA prior to PCR. The primers used for amplification of the vector can be found in Supplementary Table 4.1. First, four individual mutations were made to the POR1p binding sites in the A1, R1, A2 and A3 sequences of the native promoter. Combinatorial mutations were then created using these vectors as the base vectors until all four binding sites were mutated in the native POX2 promoter.

Table 4.1: Cloning strategy to create vectors containing mutations to the respective POR1p binding sites in the A1, R1, A2 and A3 sequence.

Base Vector	***Mutation***	***Primers***	***Final Vector***
pSL16-POX2-hrGFP	A1	F1/R1 F2/R2	pSL16-POX2 [A1 mutation]-hrGFP
	R1	F1/R1 F3/R3	pSL16-POX2 [R1 mutation]-hrGFP
	A2	F1/R1 F4/R4	pSL16-POX2 [A2 mutation]-hrGFP
	A3	F1/R1 F5/R5	pSL16-POX2 [A3 mutation]-hrGFP
pSL16-POX2 [A1 mutation]-hrGFP	R1	F1/R1 F3/R3	pSL16-POX2 [A1R1 mutation]-hrGFP
	A2	F1/R1 F4/R4	pSL16-POX2 [A1A2 mutation]-hrGFP
	A3	F1/R1 F5/R5	pSL16-POX2 [A1A3 mutation]-hrGFP
pSL16-POX2 [A1R1 mutation]-hrGFP	A2	F1/R1 F4/R4	pSL16-POX2 [A1R1A2 mutation]-hrGFP
pSL16-POX2 [A1A2 mutation]-hrGFP	A3	F1/R1 F5/R5	pSL16-POX2 [A1A2A3 mutation]-hrGFP
pSL16-POX2 [A1R1A2 mutation]-hrGFP			pSL16-POX2 [A1R1A2A3 mutation]-hrGFP
pSL16-POX2 [R1 mutation]-hrGFP			pSL16-POX2 [R1A3 mutation]-hrGFP

Transformation, Cell Culture, and Media Formulations

Plasmid propagation and cloning were performed using *E. coli* DH10β competent cells (NEB). The heat shock transformation method was used. All transformations of *Y. lipolytica* (MATa leu2−270 ura3−302 xpr2−322 axp1) were performed using the lithium acetate method as previously described [106] with a minor modification to the final step. In brief, the cells were heat-shocked for 10 minutes at 39ºC and then mixed with 1.2 mL 0.1M LiAc (pH 6.0) solution. The mixture was centrifuged at *3,000xg* for 2 mins and the pellet was re-suspended in 0.1 mL of 0.1 M LiAc (pH 6.0) and 0.03 mL of the mixture was transferred into the desired auxotrophic media to propagate cell growth. The auxotrophic media prepared was Yeast Complete Synthetic (YSC) media – LEU using 6.7 g/L Yeast Nitrogen Base (YNB) w/o amino acids (BD Diagnostic, Hunt Valley, MD) and 0.69 g/L Complete Supplement Mixture deficient in leucine (CMS-LEU) (MP Biomedicals, Solon, OH). Yeast Peptone Dextrose (YPD) or Oleic Acid (YPO) media was prepared using 5 g/L Yeast Extract (BD Diagnostics), Peptone (BD Diagnostics) and either 2% (w/v) Glucose (Sigma Aldrich, St. Louis, MO) or 2% (v/v) Oleic Acid (EMD Millipore, Burlington, MA) emulsified with 5% Tween80 (Sigma Aldrich), respectively. To test the effects of mutations and transcriptional factor knockouts, all cultures were performed in 2 mL cultures in 14 mL culture tubes. For mononucleosome extraction, 100 mL cultures were grown in a 250 mL baffled flask.

CRISPR-Cas9 to Create Knock-out Strains

The PO1f strain was used for all experiments and further engineering. CRISPR-Cas9 genome editing [83] was used to create all knockout strains used in this study. The CRISPR vector was modified to contain a NsiI restriction site used to insert the gRNAs of interest. gRNA oligos containing SLIC overhangs (Supplementary Table 4.2) were annealed together prior to cloning into the NsiI CRISPR vector that was digested with NsiI-HF. The strains PO1f*Δpor1*, PO1f*Δcfu1* and PO1f *Δpor1Δcfu1* contained frameshift mutations (Supplementary Figure 4.2). The sites for gRNA targeting sequences were designed using Benchling software (http://www.benchling.com) to insert the frameshift in 5' region the coding sequence of the respective proteins. Verification primers shown in

Supplementary Table 4.2 were used to perform colony PCR and screen colonies containing the desired mutation and Sanger sequencing was used to verify the mutations.

Mononucleosome DNA Preparation

Adaptations were made to a mononucleosome DNA preparation protocol previously described for *S. cerevisiae* resulting in >90% mononucleosome DNA [288]. In brief, the respective strains were grown in either YPD or YPO until early exponential phase (OD600 of ~20-25) and measured using the cuvette function on the NanoDrop (Thermo Scientific). Cells were chilled on ice prior to centrifugation at *3,000xg* and washed twice with 1x Phosphate Buffered Saline (PBS). The cells were then resuspended in nuclease digestion buffer (10 mM Tris-HCl (pH 8) and 1 mM $CaCl_2$). For cell lysis with the bead beater (BioSpec Products, Bartlesville, OK), our modification was seven rounds of beating, each cycle for a minute with a 2-minute pause in between at 4ºC. Protein quantification was performed using a Bradford assay. It is essential to run an optimization step by varying MNase concentration to obtain a final mononucleosome DNA product (>90%) which can be detected and analyzed as ~146 bp on a 2% DNA gel run at ~80V (constant voltage, 1xTAE buffer). It should be noted that a histone immunoprecipitation step was not performed therefore we cannot definitively call this a nucleosome preparation method, rather technique that provides a DNA accessibility profile.

PCR-tiling and PCR to determine C_T of each Amplification

PCR tiling primers for the POX2 promoter were designed using the PCT Tiler v1.42 online tool [289]. The minimum and maximum temperature for primer annealing were set at 60ºC and 63ºC, respectively. The specifications for minimum and maximum amplicon length were 100 and 130 bps. A key parameter was designing overlapping primer pairs where each consecutive primer starts 30 base pairs after the beginning of the previous primer to obtain a good overlap between each of the primers tiling the promoter. This resulted in 44 primers pairs constructed for the tiling experiment (Supplementary Table 4.3). PerfeCTa SYBR Green FastMix (Quantabio) was used for qPCR reactions in the CFX Connect real-time (Bio-Rad). The reaction volume for each set-up was 20 μL and total primer concentration in each well was 0.4 μM. The efficiency of each primer was calculated using a four-point standard curve with gDNA as the template strand. The efficiency curves were then used to convert C_T of each reaction to copy number. For each tiling experiment, the copy numbers were normalized to a range between 0 and 1 by dividing the copy number by the maximum copy number within the experimental set to provide a nucleosome occupancy profile between 0 and 1.

Fluorescence Spectroscopy

To test the effect of mutations on the POX2 promoter, each of the vectors harboring mutations as described in Table 4.1 were transformed into the PO1f wildtype strain and grown in YSC-LEU media containing 2% (w/v) glucose for 48 hours until cells reached stationary phase. Cells from a glucose culture were spun down and resuspended in YSC-LEU media with 2% (v/v) oleic acid at a starting OD_{600} of 0.2 and allowed to grow for around 36 hours determined to be the optimal time for POX2 expression (Supplementary Figure 4.1). Cells from the oleic acid culture were spun down and washed once in 1xPBS containing 5% (v/v) Tween 80 to remove residual oleic acid and finally resuspended in 1x PBS solution (pH 7.0). After OD_{600} normalization across all cultures, 0.2 mL of the cultures were loaded into 96 well black plates and serial dilutions of this culture were made to calculate the fluorescence as described in the methods section elsewhere [116]. The excitation/emission wavelengths used were 485 / 510 nm with a bandwidth of 9 in the Biotek Synergy MX fluorescence spectrophotometer. Since the strength of the native POX2 promoter is inherently weak, the gain was set to 157 to obtain reliable measurements of fluorescence. To determine the effects of the transcription factor knockouts, the respective strains were transformed with the A1R1x3A3-GFP vector and grown in minimal media containing glucose until early stationary phase was reached. Cells were then spun down and resuspended in minimal media containing 2% (v/v) oleic acid and transferred into to a 48-well plate. Fluorescence from the GFP reporter was measured each hour with continuous shaking for 18 hours while the cells remained at stationary phase. For these experiments, the gain was set to 140 since the hybrid promoter is significantly stronger than the native POX2 promoter.

Screening for Growth in Oleic Acid

PO1f wild-type and strains, PO1f*Δpor1* and PO1f*Δcfu1*, in biological replicates, were grown in minimal synthetic media supplemented with 100 mg/L L-leucine (BD

Diagnostics) (YSC) and 2% (w/v) glucose until stationary phase. Four quadrants were drawn on YSC plates containing 2% (v/v) emulsified oleic acid and a second set of plates with 2% (w/v) glucose. Cells from stationary phase cultures were spread on each of the plates using an inoculating loop and grown at 28ºC for 96 hours (4 days) to observe growth.

Results

Nucleosome Profiling of POX2 Promoter

Activation of PHO promoters occur due to extensive remodeling and loss of nucleosome at the upstream activation sites harboring transcription factor binding sites causing activation of the promoters under cellular phosphate starvation. We wanted to investigate whether a similar remodeling process can be observed in promoters of *Y. lipolytica*. The POX2 promoter in *Y. lipolytica* is a fatty acid inducible promoter that is activated by oleic acid and repressed in either glucose or glycerol [110, 116]. The 6 bp putative binding site for POR1p (CCTCGG) were mapped to the POX2 promoter (Figure 4.3). These motifs, labeled as 1, 2, 3, and 4, are located in the A1, R1, A2, and A3 UAS regions previously identified via a heuristic approach described in Chapter 3 [116].

All four binding motifs are in regions of the promoter where there is low nucleosome density. Binding site 1 is near a region that transitions to a higher nucleosome occupancy in oleic acid. Meanwhile, in oleic acid, the nucleosome profile near binding site 2 shifts to the right, away from the binding motif. Binding site 3 is located in a ~90 bp nucleosome-free trough. To the far right of the binding site, however, there is a significant increase and shift in nucleosome position in oleic acid. The nucleosome profile near motif 4 has a higher density of nucleosomes in oleic acid compared to glucose. At this point, it is difficult to conclude mechanisms that cause these transitions but note that binding sites 1, 2, and 4 are nearest to the edge of transitions in nucleosome occupancy between glucose and oleic acid. From PHO5 regulation, UASs do not have to be wrapped within nucleosomes to be inaccessible but sitting near regions of high nucleosome density can also prevent accessibility of transcriptional factors [226]. Nucleosome remodeling around the region relieves compaction and makes the binding site more accessible. We could hypothesize a similar mechanism to explain the shifts in nucleosome profiles from glucose to oleic acid.

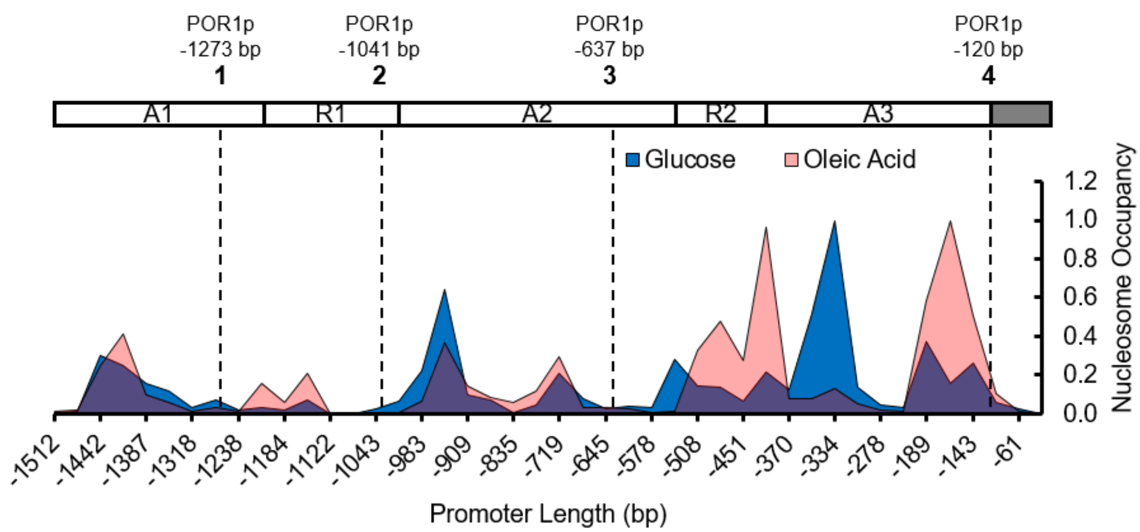


Figure 4.3. Nucleosome occupancy for native POX2 promoter in the PO1f WT strain grown in YP Glucose (blue) and YP Oleic acid (pink) mapped to POR1p binding sites on the native POX2 promoter. Putative binding sites for POR1p, 1, 2, 3, and 4 are found in A1, R1, A2, and A3 regions of the POX2 promoter previously mapped out [116]. Purple regions indicate overlap of nucleosome profile for both culture conditions.

Determination of High-Affinity Binding Motifs on the POX2 Promoter

Another important factor to consider alongside nucleosome positioning is binding site affinity. In the *S. cerevisiae* PHO5 promoter, all upstream binding sites for transcription factor promoter activation are not bound by nucleosomes, nor do all binding sites share the same affinity for transcription factor binding [285, 290]. Low and high-affinity sites binding sites are located in the nucleosome-free linker regions and set the threshold for promoter activation. This provides an initial binding site for activation during induction although the threshold is higher for low-affinity binding sites [291].

Having mapped the putative POR1p motifs to DNA accessibility profiles in glucose and oleic acid, we wanted to determine the high and low-affinity sites in the POX2 promoter. The palindromic CCTCGG POR1p recognition sequences were mutated from GC rich to AT-rich (TAAATA) (Figure 4.4A). Vectors containing a GFP reporter gene driven by a POX2 promoter harboring individual and combinatorial mutations were episomally expressed in *Y. lipolytic* and fluorescence was measured after 36 hours of growth in oleic acid. From individual mutations, sites 2 and 4 were most critical for expression from the POX2 promoter. Mutating site 1 or 3 had minimal effect on GFP expression (Figure 4.4B). Interestingly, mutating both the 1 and 2 motifs had a stronger effect on expression strength compared to mutation of 2 alone suggesting there is a dependence of binding site 1 on 2. The same effect was not observed when mutations to site 1 and 3 were made, resulting in approximately the same strength as the promoter harboring a mutation at site 1. The most striking observation, however, was the near complete loss in fluorescence with a mutation to sites 2 and 4, demonstrating the importance of these two putative binding motifs (Figure 4.4B). Both 2 and 4 sit in regions of high DNA accessibility in either glucose or oleic acid, therefore we can hypothesize that these motifs may have high affinity to POR1p binding.

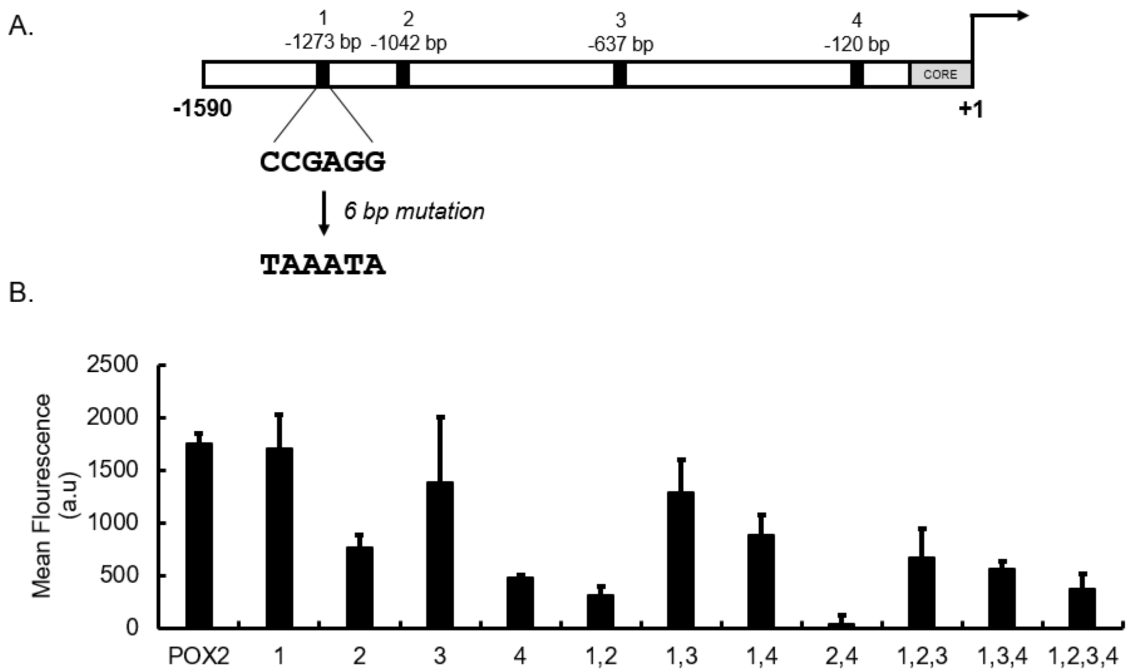


Figure 4.4. (A) Native POX2 promoter with mapped 1, 2, 3, and 4 binding motifs. Mutation to the POR1p consensus binding motif was changed from CG rich to AT-rich. (B) GFP fluorescence of POX2 promoters with individual and combinatorial mutation of the POR1p binding sites. The error is represented as a standard deviation of replicate size n=3.

Screening for Transcription Factors

Nucleosome positioning on promoters to regulate transcription is influenced by competition with transcription factors [292]. Therefore studying how nucleosome form with and without and transcription factors can elucidate mechanisms for nucleosome positioning to activate transcription in a UAS region of a promoter. To first study this mechanism requires us to know the transcription factors and binding motifs on a UAS element. We use the POX2 promoter that is the most characterized promoter system in *Y. lipolytica* where it is understood that POR1p is involved in transcriptional activatation and putative binding sites are known. The goal is to eventually translate our knowledge of nucleosome positioning in UAS regions of the POX2 promoter to identify newer UAS elements in other promoter systems that would enable of hybrid promoter engineering efforts.

Our previous results suggest that putative binding POR1p motifs, 1, 2, and 4 are important for transcriptional activation. Theses motifs are found on UAS elements, A1, R1, and A3 used to construct a fatty acid inducible hybrid promoter four times stronger than the native POX2 promoter [116]. Therefore, we use the strong hybrid promoter to determine the significance of POR1p and identify other transcription factors necessary in association with POR1p to activate transcription in oleic acid.

Our first strategy was to knockout POR1p in *Y. lipolytica* and investigate its effect on growth and transcriptional activation from the hyrid promoter harboring important UAS elements. The strain harboring a non-functional POR1p, PO1f*Δpor1*, led to partial growth in oleic acid (Figure 4.5A). This result complements previous work showing similar effects of ΔPOR1p on beta-oxidation [208]. Knocking out POR1p resulted in only partial loss in expression from the hybrid promoter compared to the PO1f wild-type (Figure 4.5C). A similar effect has been observed with the native POX2 promoter [208]. This suggests that POR1p may not be the only transcription factor activating transcription from UASs in the POX2 promoter.

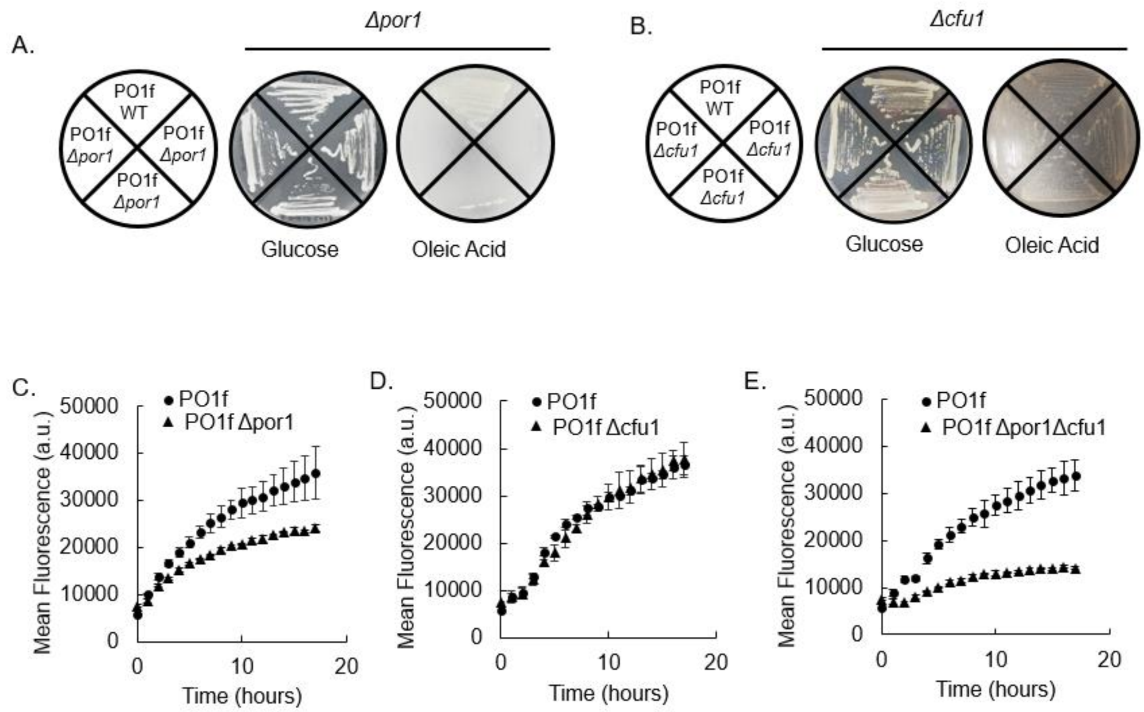


Figure 4.5. Identification of critical transcription factors. (A) PO1f*Δpor1* steaked out on synthetic media plates containing glucose and oleic acid shows severe growth defects in oleic acid. (B) No growth defects are observed in either carbon sources for PO1f*Δcfu1*. (C, D, E) GFP fluorescence experiments for strains PO1f*Δpor1*, PO1f*Δcfu1* and PO1f*Δcfu1Δpor1* grown in glucose until stationary phase and then induced with 2% (v/v) oleic acid to measure A1R1x3A3-GFP promoter strength over time. The error is represented as a standard deviation with replicate size n=3.

Therefore, we investigated other potential coregulators of UAS activity in oleic acid. The transcription factor, ADR1p, is known to be a positive regulator for transcription of genes encoding peroxisomal proteins [293]. The homolog of ADR1p was identified as CFU1p (Control of Fatty Acid Utilization) in *Y. lipolytica* (unpublished report). Our BLAST analysis showed no sequence homology of *Yl*CFU1p to *Sc*ADR1p (Supplementary Figure 4.3A). The closest homology was to transcription factor, TDA9p, from the pentose utilizing yeast, *Sugiyamaella lignohabitans*, a relative to *Y. lipolytica*

[294]. However, a BLAST of *Yl*CFU1p against the *S. cerevisiae* database revealed that the *Sc*ADR1p, chain A, containing the GATA zinc-finger binding domain shared 65% sequence homology to a similar domain in *Yl*CFU1p (Supplementary Figure 4.3B). To date, the effect of *Yl*CFU1p on either beta-oxidation or POX2 regulation is unknown. We were interested to determine whether CFU1p had any effect on the UAS elements from the POX2 promoter. Knocking out CFU1p did not result in any growth defect on either media (Figure 4.5B) and no effect on the hybrid promoter (Figure 4.5D). This suggests that CFU1p alone does not have an direct effect on beta-oxidation or promoter regulation.

Eukaryotic regulation is complex and usually involves the association of more than one transcription factor. Therefore, we tested the effect on expression of the hybrid fatty acid inducible promoter in the PO1f*Δpor1Δcfu1* strain. A noticeably weaker fluorescence signal compared PO1f*Δpor1* was observed in oleic acid (Figure 4.5C). This may suggest that that CFU1p is a coregulator requiring POR1p as the main transcriptional activator in the UAS elements. Since we were unable to get complete loss of fluorescence from the hybrid promoter suggests that may be other transcription factors involved in oleic acid activation from the UAS elements that still needs to be identified. Chapter 6 details experiments to deterimine these transcription factors.

DNA Accessibility in Strains Devoid of Transcription Factors

Although, we were not able to identify all transcription factors involved in oleic acid activation, we identified two important transcription factors, POR1p and CFU1p, in combination seem to impact activity of the UASs from the POX2 promoter in oleic acid.

Therefore, we wanted to investigate if deleting these transcription factors would relieve the competition with nucleosomes and result in more identical DNA accessibility profiles in glucose and oleic acid. We anticipate the difference in DNA accessibility between the wild-type and the PO1f*Δpor1Δcfu1* strain in oleic acid would enable us to identify UAS regions in the POX2 promoter. Using a PCR tiling array, DNA accessibility of the native POX2 promoter was mapped in PO1f*Δpor1Δcfu1* strain (Figure 4.6).

With the exception of nucleosome density, as represented by the area under the graph, and the region between -578 bp to -278 bp, there is better overall convergence of the nucleosome profiles in the between the two carbon conditions in the knockout strain (Figure 4.6) compared to the profiles in the wildtype strain (Figure 4.3). The DNA accessibility near binding motif 1 is similar between oleic acid and glucose. Meanwhile, the nucleosome profile shift noted in the wild-type strain next to binding site 2 is less prominent due to better convergence of the oleic acid and glucose nucleosome data. At motif 4, nucleosome spread is greater and density is higher under both conditions relative to the wild-type. Given that motif 4 is close to the core promoter, it would be interesting to investigate how the increase in nucleosome density near the core promoter would affect transcriptional regulation since the core promoter contains the crucial TATA docking site for basal transcription machinery.

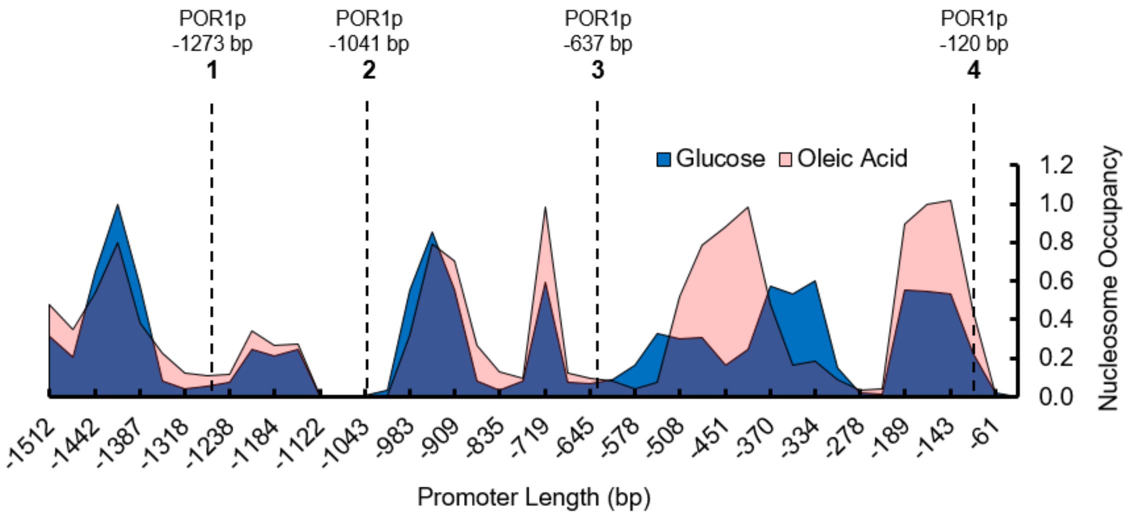


Figure 4.6. Nucleosome occupancy for native POX2 promoter in the PO1f*Δpor1Δcfu1* strain grown in YP Glucose (blue) and YP Oleic acid (pink) mapped to POR1p binding sites on the native POX2 promoter. Purple regions indicate overlap of nucleosome profile for both culture conditions.

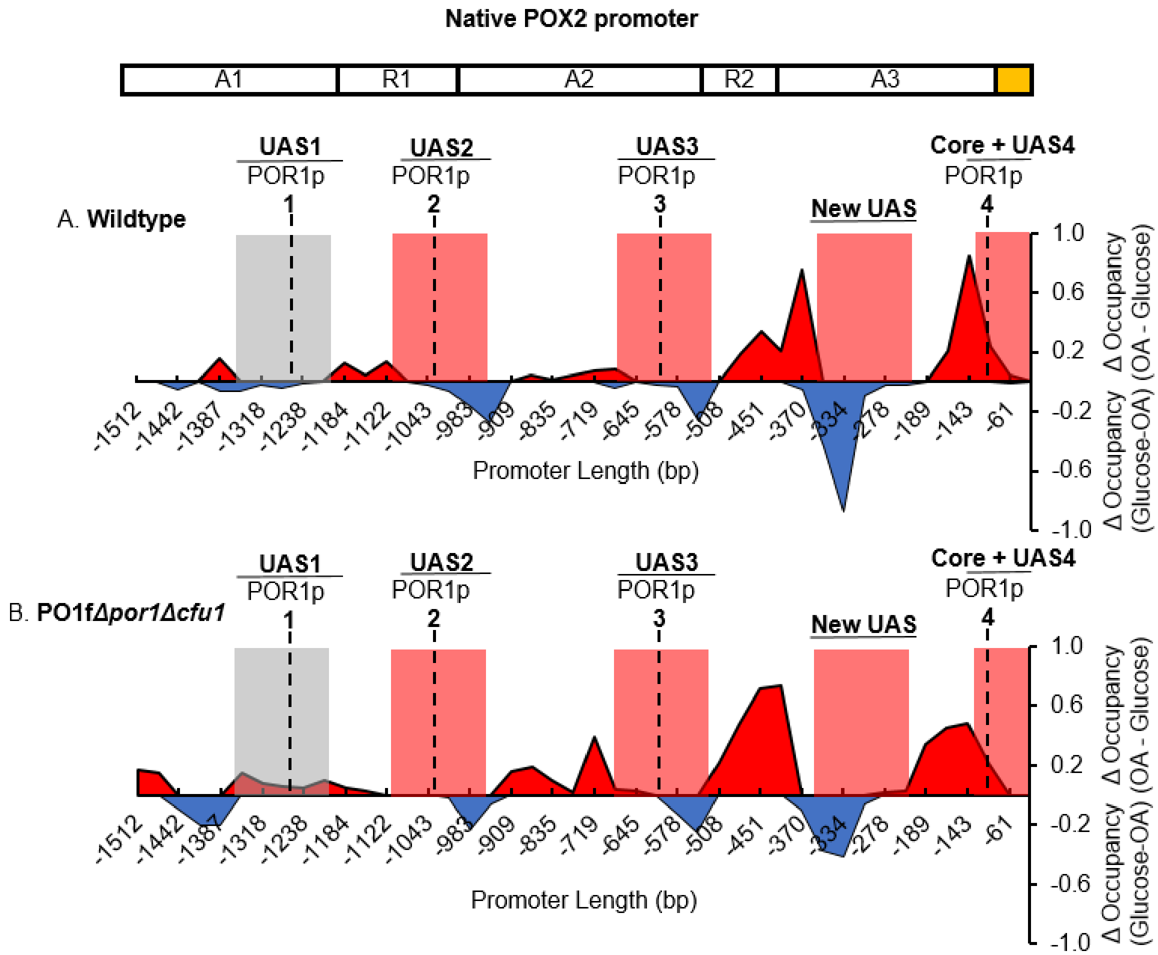


Figure 4.7. Re-engineering POX2 promoter from differences in DNA accessibility in glucose and oleic acid. POX2 promoter from (A) wildtype strain and (B) PO1f*Δcfu1Δpor1* strain showing higher nucleosome occupancy in oleic acid (OA) on top axis and higher nucleosome occupancy in glucose on lower axis. Putative POR1p binding sites are mapped out to DNA accessibility profile located in previously identified UASs, A1, R1, A2, and A3 elements on the POX2 promoter. Grey and red regions represent UAS elements selected from differences in DNA accessibility between both carbon sources. Yellow represents POX2 core promoter.

Discussion and Conclusion

The goal of this work is to determine if it is possible to predict locations of UAS elements in native inducible promoters by investigating changes in DNA accessibility on promoters under different conditions. If feasible, this method will enable quicker development of hybrid promoters with tunable and predictable expression. A strategy like this is especially beneficial for non-conventional microbes such as *Y. lipolytica* where there is a need to develop more genetic tools in spite of a limited understanding of transcriptional regulation of native promoters. RNA seq can be used to determine inducible and constitutive native promoters but does not inform us how promoters are regulated. Identifying short segments of UASs in long eukaryotic promoters aids in the engineering of better expression platforms.

Developing tunable promoters using DNA accessibility is a novel approach that has not been attempted to date. Constructing promoters in this context requires a mechanistic understanding of how nucleosomes re-position or displace to accommodate transcription factors under inducible conditions. Nucleosome positioning on promoters is largely influenced by DNA diversity that can favor or disfavor nucleosome formation [295, 296]. For example, in core promoter regions next to the TATA box, the AT-rich nucleotide region promotes weak base-pair interaction and facilitates DNA unwinding for basal transcriptional activation. The AT-rich abundance leads to a low propensity for nucleosome formation leading to NFRs in core promoters [297]. This is also observed in the tiling experiments for the native POX2 promoter where the core promoter region (-61 to +1), being ~58% AT-rich, has a sharp drop in nucleosome occupancy (Figures 4.3 and

4.6). However, upstream of the promoter where transcription activators or repressors bind to amplify or repress genes, the understanding of molecular interplay between histones, nucleosome remodeling complexes, specific and non-specific TFs is less clear.

The POX2 promoter is one the more well-studied native fatty acid inducible promoters in *Y. lipolytica* [110, 116, 208] and is a good candidate promoter to test the hypothesis. POR1p has previously been shown to be important for POX2 expression in oleic acid is a close homolog of the fatty acid regulator FarA in *Aspergillus nidulans* [208]. Putative binding sites of the *Yl*POR1p were mapped on the native POX2 promoter in the wild-type strained tiled to determine nucleosome occupancy in glucose and oleic acid. The four binding sites, 1,2, 3, and 4 were in low nucleosome occupancy regions for both glucose and oleic acid but 1, 2, and 4 were at the edge of where transitions of nucleosome occupancy in oleic acid were observed, either an increase or a shift in profile.

While it has been established that not all transcription factor binding sites bear similar importance to transcriptional regulation in the PHO promoters of *S. cerevisiae*, we wanted to determine which of the putative motifs contributed most to transcriptional activation of the POX2 promoter. Binding site 1 was interesting because mutating the site did not contribute to transcriptional loss although in tandem with a mutation binding site 2 or 4 resulted in greater than a four-fold drop in strength. Mutation to 2 and 4 caused almost near loss of expression. These results inform us that these sites are important but the reasons are less clear. It would be important to determine if the mutation removes the POR1p binding site on the DNA or whether there are there are changes in DNA accessibility that makes the binding sites within the region less accessible. Nucleosome profiling of the POX2 promoter harboring these mutations can answer this question. The region near or at binding site 1 has demonstrated cooperativity to binding motifs 2 and 4. This result aligns well with our previous finding that demonstrated the importance of UASs A1, R1 and A3, harboring sites 1, 2, and 4, respectively, to confer strong transcriptional activation [116]. Binding site 4, ~60 bp away core promoter, also had a deleterious impact on POX2 expression. Combined with the mutation to binding site 2, caused near loss of expression. In the future, it would be important to investigate how surrounding sequences could impact the affinity of the transcription factor to the binding site in a nucleosome independent context [285].

Knockout studies of the POR1p and CFU1p transcription factors had different effects on growth and expression of the hybrid promoter. PO1f*Δpor1* caused severe growth defects on oleic and lowered transcription output from the hybrid promoter while PO1f*Δcfu1* had no effect. The double mutant, PO1f*Δpor1Δcfu1* showed more significant down-regulation of the hybrid promoter but did not eliminate expression completely. More transcription factors would need to be screened to determine the mechanism for regulation.

We observed a better convergence of the nucleosome profiles of the PO1f*Δpor1Δcfu1* strains in glucose and oleic acid except for the 300 bp region between -578 and -378 bps. For future work, nucleosome profiling in the PO1f*Δpor1* strain in glucose and oleic acid is required to determine the effects ΔCFU1p has on the profile. At this point, the role of *Yl*CFU1p is less understood and although, it contains a conserved DNA binding domain (Supplementary Figure 4.3B), whether it directly binds to the

POX2 promoter is unknown. Preliminary results presented here show that the CFU1p alone may not have a direct association with POX2 expression or growth (Figure 4.5 B, D) however, further work is required to elucidate its role. Similarly, *Yl*POR1p has not been shown to directly bind to the POX2 promoter in oleic acid. This is a key experiment needed to validate both transcription factors as nuclear receptors for oleic acid induction.

Figure 4.7A and B presents a new analysis of DNA accessibility in the wildtype and PO1f*Δpor1Δcfu1* strains, respectively. The calculated difference in occupancy between both substrates enables better visualization of which regions are more occupied in glucose versus oleic acid and vice versa. The objective of this analysis is to select regions in the POX2 promoter that could be potential UAS elements. In the grey UAS1 region containing putative POR1p binding site 1, there is generally a lower nucleosome occupancy in both two carbon conditions. However, once the POR1p and CFU1p are knocked out in the mutant strain, PO1f∆*por1*∆*cfu1*, we observed a bleed over of nucleosomes into the region. This selection of this region as a potential UAS is made on the hypothesis that knocking out competition from these oleic acid inducible transcription factors would now favor nucleosome formation in oleic acid. The transitions seen here may hint at one mechanism for how enhancer sites are regulated.

UAS2 and UAS3 elements shaded in red contains putative POR1p binding sites 2 and 3. The selection of these regions is based on the hypothesis that oleic acid-induced transcription would result in lower nucleosome occupancy in oleic acid than in glucose. Applying a similar principle helps us identify another potential UAS region (New UAS) that could be tested. Furthermore, we selected UAS4 based on Figure 4.4B data that shows it is the most critical binding motif for transcriptional regulation and it sits in proximity to the core promoter. These UASs can be used to engineer a new fatty acid hybrid promoter that would be different to the heuristic approach applied in Chapter 3. Ultimately, we would like to apply this approach to dissecting new UASs from other promoters to engineer more hybrid promoters in *Y. lipolytica*.

Here, we make predictions for new UASs based off DNA accessibility profiles that account for nucleosomes and transcription factor binding to the DNA. From these DNA accessibility profiles, it would not be possible to definitively conclude changes as a result of nucleosome eviction or sliding as transcription factors are also crosslinked to DNA. This can prevent MNase cleaving DNA at the transcription factor bound sites, thereby we may see more nucleosome protected regions. However, if transcription factor binds to DNA wrapped in a nucleosome, which has been reported as another mechanism, then predictions made from this dataset would be valid. At this moment, we do not know which mechanisms are more predominant, therefore, the work presented here needs to be complemented with profiling of pure mononucleosome DNA that can be achieved with chromatin immunoprecipitation using histone-specific antibodies [298]. This could also provide insight into histone modifications that contribute to transcriptional regulation.

The work highlighted in this chapter is aimed towards developing methodologies towards designing new regulatable promoters using predictions from DNA accessibility. The nucleosome mapping approach is applied on the POX2 promoter that has previously been studied to identify regulatory regions (A1, R1, A2, A3) and transcription factors (POR1p, CFU1p) associated with FA-induced expression. The nucleosome profile of the wild-type strain under the different conditions enables mapping of key binding sites with respect to changes in nucleosome profile. Most of the binding sites are located near or within regions of high nucleosome occupancy in oleic acid growth conditions. The importance of these binding motifs is tested using via mutational studies. Next, transcription factors, *Yl*POR1p and *Yl*CFU1p are knocked out, leading to repressed expression from the POX2 promoter. The POX2 promoter was tiled in the knockout strain which showed a significant broadening of nucleosome profiles in oleic acid relative to the wild-type strain while maintaining a similar profile in glucose. The difference in nucleosome between oleic acid and glucose between the wild-type and knockout strain aids new strategies to re-engineer a hybrid promoter from POX2 elements.

CHAPTER FIVE

A FATTY ACID RESPONSIVE PROMOTER USED TO GUIDE ENGINEERING OF FATTY ALCOHOL PRODUCTION IN THE OLEAGINOUS YEAST, *YARROWIA LIPOLYTICA*

Abstract

Fatty alcohols are an important class of oleochemicals with a wide range of industrial applications from biofuels to surfactants and detergents. A microbial platform capable of producing biorenewable fatty alcohols may be competitive with current production capacities. From an engineering perspective, *Yarrowia lipolytica* is a promising microbe for fatty alcohol production because of its natural ability to synthesize and metabolize lipids that are precursors to fatty alcohols. The strategy described here utilizes a fatty acid-responsive promoter developed in Chapter 3 to detect intracellular pools of free fatty acids. A strain capable of producing upwards of 1 g/L intracellular free fatty acids was first engineered. By expressing a heterologous fatty acyl-CoA/aldehyde reductase from *Marinobacter aquaeolei*, MAACR, localized to the peroxisome, we produced over 530 mg/L fatty alcohols in the engineered strain, a yield close to 7.5 mg/g glucose. While there is more work that needs to be done to make production more efficient, this is first time peroxisomal targeted fatty alcohol production has been demonstrated in *Y. lipolytica*, a promising approach to produce short to medium chain fatty alcohols in the future.

[¥] This work will be included in a future publication with co-authors Michael Spagnuolo, Matthew Brabender, Cory Schwartz, Ian Wheeldon, and Mark Blenner.

Introduction

Fatty alcohols are an important class of biological molecules that can range from as short as 4-6 carbons to as many as 22-26 carbons. Depending on the chain length, fatty alcohols can have a wide range of applications in industry ranging from personal care & cosmetics, soaps & detergents, textiles, oil, and gas. The global fatty alcohols market is estimated at USD 4.7 billion in 2017 and projected to reach USD 6 billion by 2022 [299].

Traditional means to produce these aliphatic compounds via catalysis route can be costly and require energy-intensive reactions. Furthermore, synthesis of unique metal catalysts is required to produce fatty alcohols with different chain length specificities [300, 301]. In comparison, biological means of fatty alcohol production bypass these issues because in nature, several naturally occurring biocatalysts have chain length specificity but also enable the production of these biomolecules in a more energy friendly sustainable manner by utilizing cheap, cost-effective feedstock [65, 302].

In recent years, microbes have been explored as an alternative route for fatty alcohol production [303-305]. *S. cerevisiae* and *E. coli* are target hosts because of the wide array of genetic tools, ease of genome editing, and an extensive body of literature that already exists well establishing the metabolic pathways. Yeast systems are more industrially relevant because of the ability to grow them to high density at large-scale fermentation, resistance to phage infections, and higher tolerance to toxic inhibitors and products [306].

There are essentially two main pathways to produce fatty alcohols in yeast (Figure 5.1). One strategy is the conversion of fatty acyl-CoA to fatty alcohol via a single four-electron reduction step using a bifunctional acyl-CoA / aldehyde reductase. The commonly used enzymes are *Ta*FAR1 *(Tyo alba)*, *Mm*FAR1 *(Mus musculus)* and MAACR *(Marinobacter aquaeoli)* [65, 91, 92, 307]. The other strategy is to convert fatty acids to aldehydes via a carboxylic acid reductase (CAR) and then convert the aldehydes to fatty alcohols via an alcohol dehydrogenase or aldehyde reductase.

The earliest work in *S. cerevisiae* relied on pushing flux to fatty acyl-CoA from acetyl-CoA and using a fatty acyl-CoA reductase, *Mm*FAR1, to convert fatty acyl-CoA to fatty alcohol [90]. This led to almost 100 mg/L fatty alcohols and the production yield was 5 mg/g (glucose and galactose). Another strategy was to block triacylglycerol synthesis by deleting *DGA1*, accumulating fatty acyl-CoAs in the cytosol [308]. The fatty acyl-CoA was converted to fatty alcohols using a fatty acyl-CoA reductase from *Ta*FAR1 [308]. The authors used the word “yield” to describe titers, reporting 100 mg/L fatty alcohols produced in the engineered strain. If total substrate consumption is considered, the yield would be 1.6 g/g galactose.

It was not until recently that d’Espaux *et. al.* showed significant improvements in fatty alcohol production. First, the push to fatty acyl-CoA from acetyl-CoA was engineered by overexpressing an acetyl CoA carboxylase (*Sc*ACC1), the first committed step in fatty acid synthesis. Proteomics revealed that this enzyme was naturally present in low abundance. Many competing pathways for the fatty acyl-CoA were blocked, the more significant knockout being *Sc*DGA1. A strong pull on fatty acyl-CoA pools for fatty alcohol production was accomplished by overexpressing *Mm*FAR1 [65]. They determined that *Mm*FAR1 was a more efficient four-electron reducing reductase than

TaFAR1. Titers of 1.2 g/L and yield of 70 mg/g glucose were reported (Table 5.1). This is the best fatty alcohol yield reported to date in *S. cerevisiae* using glucose as substrate. This experiment, however, was performed in rich YPD media, which is not suitable for industrial production because of its high costs, complex make up, and variable composition.

The other strategy in *S. cerevisiae* was the conversion of free fatty acids to fatty alcohols. A pull towards free fatty acid production was engineering by over-expressing a heterologous ATP-dependent citrate lyase, *Mm*ACL1 *(Mus musculus)*, to convert citrate to acetyl CoA, and then overexpressing the endogenous *Sc*ACC1p. Intracellular fatty acid pools were increased by blocking fatty acid activation and degradation [309]. Conversion of fatty acids to fatty alcohols was accomplished by overexpressing a heterologous *Mycobacterium marinum* CAR, *Mm*CAR, an endogenous aldehyde dehydrogenase, *Sc*ADH5. A build-up of C18 fatty aldehydes suggested that overexpression of *Sc*ADH5 was not sufficient, therefore, MAACR was overexpressed to contribute to more efficient fatty aldehyde reduction. This strategy produced only 120 mg/L fatty alcohols in shake flasks and a low yield of 6.3 mg/g glucose [309].

More recently, inducible promoters have been demonstrated to fine-tune metabolic pathways to dynamically control fatty alcohol production. A glucose-repressible promoter from the *HXT1* gene was used to control the expression of *FAA1*, encoding for an enzyme that converts fatty acids to fatty acyl CoA. Fatty alcohol accumulation improved by ~41% while FFAs decreased by ~63% relative to the control strain [310]. In a separate study, a bifunctional reductase from *Arabidopsis* *thaliana*, *At*FAR, and upstream enzymes were placed under the control of a galactose-inducible promoter. This enabled separation of growth from production and prevented the build-up of toxic compounds from the mevalonate pathway. The engineered strain produced close to 85 mg/L of C22 alcohols, about four times higher than fatty alcohol titers in the strain containing constitutive expression promoters [311].

In addition to producing fatty alcohols in high enough titers, it is important to design a microbial production platform that is efficient and provides the means to manipulate chain length specificity. The global demand for short to medium chain fatty alcohols (C6-C14) accounts for more than half of the fatty alcohol market volume and is projected to increase in the next few years [312]. This is due to the broad range of industrial applications for medium-chain fatty alcohols while short-chain fatty alcohols are specialty chemicals because they are not abundant in nature. As a result, these fatty alcohols have a larger global demand resulting in higher market costs. Our analysis shows that given the current market value for fatty alcohols and glucose, lower yields of short and medium chain fatty alcohols would need to be produced to "break even" on the cost of glucose (Table 5.1). This does not account for production and separation costs suggesting that we would have to produce higher than the break-even yield (Table 5.1) to eventually make the process economically feasible.

Table 5.1. Average market value of fatty alcohol based on chain length and yield required to meet the cost of glucose [313]. Shaded in grey is the market price for un-refined glucose [314].

Fatty Alcohol Range	Fatty Alcohol Price (cents / g alcohol)	Break-Even Yields (mg alcohol / g glucose)
Short (C6-C8)	32	180
Medium (C8-C14)	25	250
Long (>C14)	16	400

One method to create short-chain fatty alcohols is compartmentalizing the production pathway to the peroxisome. This strategy can be used to produce shorter chain fatty acids that can then be reduced to alcohols [65, 315]. In *S. cerevisiae* short to medium chain fatty alcohols (C6-C12) have been produced in the peroxisome with a high degree of efficiency (43 mg/g glucose) compared to other studies [315]. The other strategy to create shorter chain fatty alcohols is premature termination of fatty acid elongation by using chain length specific acyl-thioesterases that convert fatty acyl-ACP (Acyl Carrier Protein) to fatty acids, which can then be converted to fatty alcohols via a reductase. The latter strategy has been explored in *Y. lipolytica* to produce C10 fatty alcohols [93]. *Y. lipolytica* is a promising host for production of fatty alcohols because of its ability to produce lipids [157, 162, 257, 316, 317] and fatty acids [149], which can serve as precursors for fatty alcohols.

By expressing bacterial and plant acyl-ACP thioesterase, fatty acid elongation was terminated to produce a large abundance of C10 fatty acids [318]. Overexpression of native fatty acyl CoA synthase, *Yl*FAA1, converted C10 fatty acids to fatty acyl-CoAs followed by *At*FAR overexpression to produce upwards of 500 mg/L of C10 fatty alcohols in the cytosol with a yield of ~10 mg/g glucose [93]. The strain lacked peroxisomes which resulted in four times weaker growth than the wildtype, W29 strain. We have observed that knocking out the peroxisomes in the PO1f auxotrophic strain causes weaker growth and slower growth rates, which for industrial purposes makes the strain more difficult to use. Table 5.2 summarizes other accomplishments to date of fatty alcohol production in *Y. lipolytica* and a comparison the best engineering feats in *S. cerevisiae*. The theoretical yield is calculated based on the total glucose supplied to the media while yield coefficient is calculated based on how much glucose was consumed.

To date, the highest fatty alcohol production achieved from *Y. lipolytica* via scale up into a 3-L bioreactor is 2.5 g/L [91] while in 40 mL shake flask fermentations, the titer was 205 mg/L [91]. Various pathways were accessed for fatty alcohol production but the most promising strategy was pushing flux from fatty acids to fatty acyl-CoA using a heterologous fatty acyl-CoA synthetase from *E. coli*, *Ec*FadD, and pulling on fatty acyl-CoA pools to fatty alcohols using MAACR reductase [91].

In this work, we compartmentalized fatty alcohols production inside the peroxisome that houses all enzyme necessary for beta-oxidation (Figure 5.2). This strategy, which has not been reported in *Y. lipolytica* before provides access to the machinery to tune fatty acid chain length by manipulating the *POX1-6* genes. The native fatty acid synthesis pathway creates predominantly C16 and C18 fatty acids while our goal is to produce short and medium chain fatty alcohols (C6-C14). In addition to achieving chain length specification, localizing the pathway into the peroxisome allows spatial compartmentalization of the enzymes that facilitate the transfer of metabolites from one enzyme to the next. This generates a more efficient assembly line for the process and has been shown in eukaryotes to increase overall metabolic output [319] and prevent undesired side reactions that hinder production efficiency.

We first localized MAACR into the peroxisome by fusing it to a peroxisomal targeting sequence (PTS). The first strategy relied on increasing the cytosolic fatty acyl-CoA pool by knocking out expression of *Yl*DGA1. We hypothesized that this would create a cellular increase fatty acyl-CoA concentrations that would also translate into the peroxisome. The second strategy that showed more success, was increasing the intracellular free fatty acid pools. We accomplished this by blocking fatty acid activation (Δ*Yl*FAA1) and beta-oxidation inside the peroxisome (Δ*Yl*MFE1). The peroxisomal fatty acyl-CoA synthetase, *Yl*AAL1, was kept in-tact. The fatty acid hybrid promoter reported the strongest GFP expression from this strain and thin layer chromatography (TLC) showed that >50% of the intracellular lipid content were free fatty acids. This strategy resulted in titers of 530 mg/L although the yield was a little over 7 mg/g glucose suggesting that more work needs to be done to improve the efficiency of the pathway.

Table 5.2. Comparison of fatty alcohol production in *S. cerevisiae* and *Y. lipolytica*. Highlighted in blue are peroxisomal targeted strategies for *S. cerevisiae* and *Y. lipolytica*. Yield coefficient is yield based on glucose consumed while yield was calculated with how much glucose was initially added to the media. N.R. means not reported data.

Host	Titer Shake Flask (mg/L)	Titer Bioreactor (mg/L)	Yield (mg/g glucose)	Yield Coefficient (mg/g glucose)	Ref.
S. cerevisiae	1200	6000	58	70	[65]
S. cerevisiae	330	1100	16.5	30	[320]
S. cerevisiae	837	1300	40	43	[315]
Y. lipolytica	690	N.R.	4.3	18	[92]
Y. lipolytica	550	N.R.	11	N.R.	[93]
Y. lipolytica	167	N.R.	5.5	N.R.	[321]
Y. lipolytica	205	2500	3.4	N.R.	[91]
Y. lipolytica	530	N.R.	7	12	This study

Experimental Methods

Chemicals and Reagents

All restriction enzymes used in cloning were purchased from New England Biolabs (NEB, Ipswich, MA) unless otherwise stated. Oligos designed for all experiments were synthesized by Eurofins Genomics. Plasmid minipreps were performed using the Zyppy™ Plasmid miniprep kit (Zymo Research, Irvine, CA). PCR purifications and restriction digest purifications were performed using the DNA Clean and Concentrator kit (Zymo Research). Fatty alcohol and fatty acid standards were purchased from Nu-Check Prep. All chemicals were purchased from Sigma Aldrich unless otherwise stated.

Transformation, Cell Culture and Media Formulations

Plasmid propagation and cloning were performed using *E. coli* DH10β competent cells (NEB). The heat shock transformation method was used. All *Y. lipolytica* transformations were performed using the lithium acetate method as previously described [106] with a minor modification to the final step. In brief, post heat-shock for 10 minutes at 39ºC, transformed cells were mixed with 1.2 mL of 0.1M LiAc (pH 6.0) solution prior to spinning down at *3,000xg* for 2 mins and re-suspending the pellet in 0.1 mL of 0.1 M LiAc (pH 6.0). 0.03 mL of the mixture was transferred into 2 mL of the desired synthetic media containing 6.7 g/L yeast nitrogen base (YNB) without amino acids (BD Diagnostics, Hunt Valley, MD) and complete synthetic media either without leucine (0.69 g/L CSM-LEU) or without leucine and uracil (0.67 g/L CSM-LEU-URA) (Sunrise Science Products, San Diego, CA). The cells were precultured in media containing 20 g/L glucose and grown in culture tubes to propagate cell growth for 40-48 hours. The cells were then transferred into their respective auxotrophic cultures at an initial OD_{600} of 0.2.

For fatty alcohol production, experiments were performed in synthetic culture media containing 1.7 g/L YNB without amino acids and ammonium sulfate (Difco) and CSM-LEU or CSM-LEU-URA. Final ammonium sulfate concentrations in the culture were 2.5 g/L and 5 g/L to accomplish C:N molar ratios 60:1 and 30:1, respectively. Each of the cultures contained 80 g/L glucose. Total culture volume was 20 mL grown in 50 mL baffled flasks. The cultures were overlaid with a 10% (v/v) dodecane layer to capture fatty alcohol excreted and prevent volatilization [322]. Cells were grown at 215 rpm for 5 days prior to harvesting cells for fatty alcohol quantification.

Plasmids and DNA Cloning

The TEF(404)-Intron and AAL1 were cloned into pSL16-cen1-1(227) ARS/CEN with a URA3 selection marker using primer pairs F1/R1 and F2/R2. To create the MAACR-PTS vector, the plasmid pSL16-UAS1B8-GFP-ScCYC1t referenced in Chapter 2 was digested with BssHI/NheI to clone a PCR amplified MAACR-PTS using SLIC primers F3/R3. Primers can be found in Supplementary Table 5.1. For intracellular fatty acid detection, a plasmid containing A1R1x3A3-GFP (created in our lab) harboring a leucine auxotrophic was transformed in each of the strains.

CRISPR-Cas9 to Create Knockout Strains

The base strain used for all subsequent knockouts was PO1f (MATa leu2−270 ura3−302 xpr2−322 axp1). CRISPR-Cas9 [83] was used to create the knockouts strains described in this chapter. The gRNA sequences used to create each knockout can be found in Supplementary Table 5.1. CRISPR-Cas9 / gRNA cassettes were made using a base vector that was redesigned to contain a NsiI cut sight where all subsequent gRNA were cloned by using SLIC. The design of the modified CRISPR vector containing the NsiI cloning site was described in Chapter 4. Table 5.3 highlights the strains engineered for fatty alcohol production and intracellular fatty acid reporter experiments. The POf1*Δfao1* strain was created as the base strain for all fatty alcohol experiments. gRNA oligos were first annealed together prior to using SLIC to clone the gRNA sequence into CRISPR vector digested with NsiI. The same strategy was used to create all strains except for the PO1f*Δpex10* and PO1f*Δmfe1Δfaa1Δpex10* where two gRNA sequences were used to remove the *PEX10* gene instead creating a frameshift mutation. First, two separate plasmids were cloned containing gRNA sequences that cut upstream and downstream of *PEX10*. The annealed gRNA oligonucleotides were F8/R8 and F9/R9, respectively. Next, the CRISPR plasmid harboring the gRNA sequence that cuts downstream of *PEX10* was PCR amplified using primer pairs F10/R10 to obtain the SCR1-tRNA-gRNAPEX10down-tracrRNA. This insert was cloned into the CRISPR vector containing the gRNA that cuts upstream of *PEX10* that was digested with a single enzyme XmaI, placing the second gRNA sequence in the same expression cassette harboring the first gRNA sequence. Sanger sequencing was used to verify screened colonies post-transformation. A greater than 50% cut efficiency of the CRISPR system was achieved using this strategy.

Table 5.3. List of engineered strains used for fatty alcohol production and tested for intracellular fatty acid production. Not in study – reported in Supplementary Figure 5.1.

Fatty alcohol production strains	
PO1f (Wild Type)	Not tested
PO1f*Δfao1* + *MAACR* (leucine)	Not tested
PO1f*Δfao1* + *MAACR* (leucine) + *AAL1* (uracil)	In study
PO1f*Δdga1Δfao1* + *MAACR* (leucine)	In study
PO1f*Δdga1Δfao1* + *MAACR* (leucine) + *FAA1* (uracil)	In study
PO1f*Δdga1Δfao1* + *MAACR* (leucine) + *TGL4* (uracil)	In study
PO1f*ΔmfeΔfaa1* + *MAACR* (leucine)	In study
PO1f*ΔmfeΔfaa1* + *MAACR* (leucine) + *AAL1* (uracil)	In study
Intracellular fatty acid sensor strains	
PO1f (Wild Type)	Not tested
PO1f + *hrGFP FA sensor* (leucine)	In study
PO1f*Δdga1* + *hrGFP FA sensor* (leucine)	In study
PO1f*Δmfe1* + *hrGFP FA sensor* (leucine)	In study
PO1f*Δmfe1* + *hrGFP FA sensor* (leucine) + *FAA1* (uracil)	Not in study
PO1f*Δpex10* + *hrGFP FA sensor* (leucine)	In study
PO1f*Δfaa1* + *hrGFP FA sensor* (leucine)	In study
PO1f*ΔmfeΔfaa1* + *hrGFP FA sensor* (leucine)	In study
PO1f*ΔmfeΔfaa1Δpex10* + *hrGFP FA sensor* (leucine)	In study
PO1f*ΔmfeΔfaa1* + *hrGFP FA sensor* (leucine) + *AAL1* (uracil)	In study
PO1f*ΔmfeΔfaa1* + *hrGFP FA sensor* (leucine) + *DGA1* (uracil)	Not in study
PO1f*ΔmfeΔfaa1* + *hrGFP FA sensor* (leucine) + *TGL4* (uracil)	Not in study

Fluorescence Spectroscopy for Intracellular Fatty Acid Detection

The Biotek Synergy MX fluorescence spectrophotometer was used for all fluorescent studies. Cells were inoculated into 48-well plates containing 0.250 mL of synthetic media prepared as described above. The spectrophotometer was set to gain 140 with fast orbital shaking. Cells were grown in plates and for ~60 hours past stationary phase with cell growth and fluorescence measured every hour using an excitation/emission wavelength 485/510 nm and a bandwidth of 9.

Total Lipid and Free Fatty Acid Extraction

To identify and quantify lipids in cell biomass, extracted cellular lipids were transesterified to FAMEs as described previously with minor modifications [95]. Briefly, 1 mL cell culture was harvested and spun down at 13,000 rpm for 3 minutes at 25 °C. 100 μL glyceryl triheptadecanoate at a concentration of 2 mg/mL methanol was added to the cell pellet as an internal standard. Lipids were transesterified to FAMEs with 500 μL of 0.5 N sodium methoxide followed by 30 min of vortexing at 2,000 rpm. The solution was neutralized with 40 μL sulfuric acid. FAMEs were extracted by adding 850 μL hexane followed by 20 min of vortexing at 2,000 rpm. The mixture was centrifuged for 1 min at 8,000 rpm, and 800 μL of the organic layer was collected for GC-FID analysis. For free fatty analysis, the blot on the TLC plate (Millipore, Burlington, MA) is scraped off and run subject to the same fatty acid methylation and quantification process as described in this section.

Thin Layer Chromatography for Determination of Percentage Lipid Classes

Cell cultures, normalized to OD_{600} of 17, were collected and centrifuged. The same extraction protocol was employed for cellular lipid and fatty alcohol extraction as described in the fatty alcohol methods section below with a minor modification. At the final step, instead of re-suspending the sample in ethyl acetate, the dried lipid extract was re-suspended in 0.1 mL hexane to be run on TLC. A protocol for silica plate-based thin layer chromatography has been described before [323, 324]. TLC plates were activated by heating the plate immediately before use for 10 min at 105 °C to remove the water.

The solvent system was prepared by thoroughly mixing hexane/diethyl ether/acetic acid (70/30/1, v/v) and poured into the chamber to a level up to approximately 1 cm from the bottom. The chamber was closed to enable solvent saturation. A pencil line was drawn on the plate approximately 2 cm from the bottom. The lipid sample was then applied to the plate using either a microsyringe or a sample applicator device incubated at room temperature for 1-2 min to allow for the hexane to evaporate. The plate was quickly placed inside the saturated chamber, standing vertically and submerged in ~ 1cm of solvent. At the end of separation (once the solvent has migrated to the top of the plate), the plate was dried by passing gently passing nitrogen gas over the plate. The plate was then placed in a chamber containing crystal of iodine (non-destructive) for several minutes until yellow or brown spots appear. A pencil was used to demarcate lipid regions of interest for further processing. An image of the chromatogram was analyzed using the Image J software [325] to determine proportions of the lipid fractions on the plate.

Fatty Alcohol Extraction

Ten mL of culture from each flask were spun down in centrifuge tubes at 3000xg for 10 minutes. The dodecane overlay (top layer) was extracted and directly used in GC-FID analysis to determine fatty alcohols in the dodecane layer. To extract fatty alcohols from the supernatant, 2 mL of ethyl acetate was added to 10 mL of the supernatant in a closed glass vial and allowed to shake for 2 hours. The mixture was then incubated at 4ºC overnight to facilitate separation of both layers. One mL of ethyl acetate was extracted for GC-FID analysis. To extract fatty alcohols from the cell pellet, 10 mL of cell culture was centrifuged and washed once with 1xPBS (Sigma) to prevent any carry-over from the supernatant. Fatty alcohol, including lipids, were extracted as described before [159, 326]. Briefly, cells were resuspended and in 1 mL of 2:1 (v/v) Chloroform: Methanol mixture and vortexed at 5000xg for 1 hour. The mixture was spun down and 0.9 mL of the liquid was extracted. 0.225 mL of 0.85% (w/v) NaCl (Saline solution) was added and vortexed for 5 minutes. The mixture was then centrifuged and chilled on ice for 5 minutes to improve the separation of the organic and aqueous phase. The aqueous layer (top phase) was removed and the organic layer was evaporated under vacuum and room temperature overnight. Ethyl acetate (0.9 mL) was added to each of the dried tubes to resuspend the fatty alcohols for GC-FID analysis. Each of the samples was spiked with 0.1 mL of 0.5 mg/mL C17:0 alcohol, heptadecanol, prior to any of the above-mentioned extraction processes.

GC-FID for Fatty Acid and Fatty Alcohol Analysis

Both methylated FFAs and fatty alcohols were quantified using GC-FID (Agilent 7890B). For fatty acid analysis, FAME species were separated on an Agilent J&W DB-23 capillary column (30 m × 0.25 mm ×.0.15 μm), with helium carrier gas at a flow rate of 1 mL/min. The temperature of the oven started at 175°C and was ramped with a gradient of 5°C/minute until 200°C. The FID was operated at a temperature of 280 °C with a helium makeup gas flow of 25 mL/min, hydrogen flow of 30 mL/min, and airflow of 300 mL/min. Fatty alcohols were separated on an Agilent DB-Waxter capillary column (30 m × 0.25 mm × 0.15 μm) using helium carrier gas at a flow rate of 1 mL/min.

The initial temperature of the oven was 100 °C and was ramped to 220°C at a rate of 20°C per minute and held for 2 minutes. The temperature was then ramped to 300°C at a rate of 20°C per minute and held for 5 mins. The temperature of the inlet was maintained at 250°C and the injection volume was 1uL with a split ratio of 1:10.

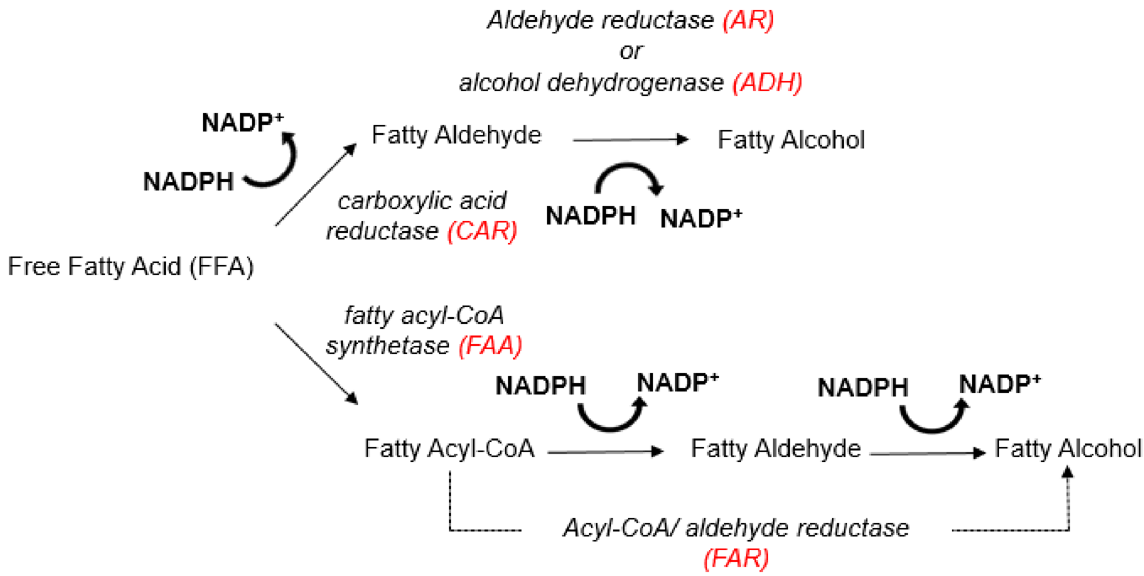


Figure 5.1. Two commonly used pathways in yeast to produce fatty alcohols. The first pathway (above) utilizes two two-electron reducing steps to convert fatty acids to fatty alcohols. The second pathway (below) utilizes a single four-electron reducing enzyme to convert fatty acyl-CoA to fatty alcohol.

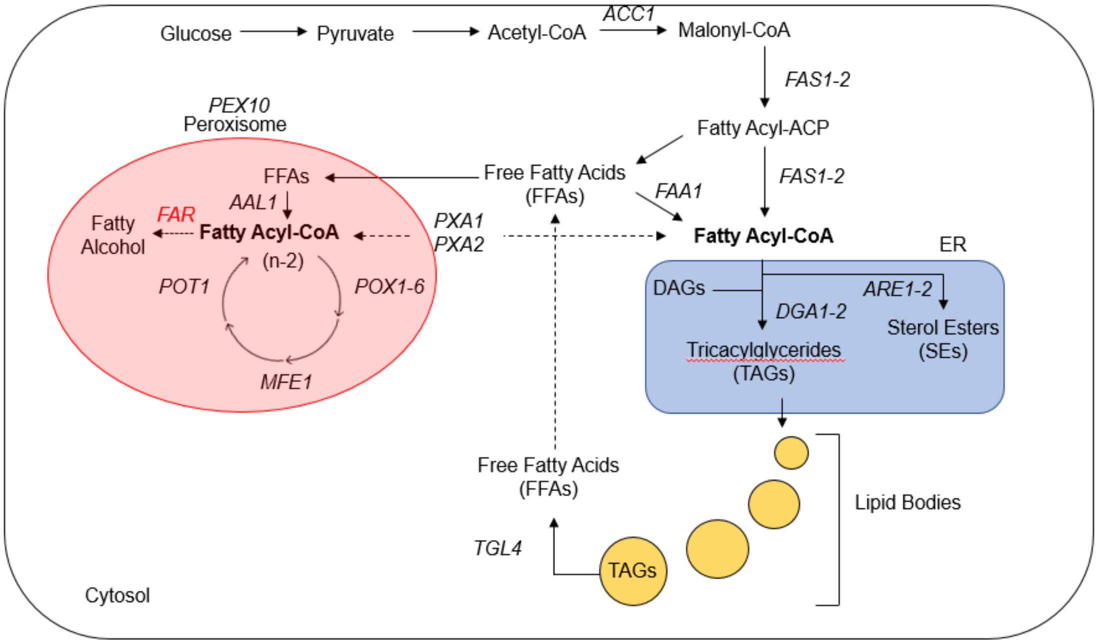


Figure 5.2. Fatty alcohol production pathway starting with glucose as the primary carbon source. Cytosolic fatty acyl-CoA can either be transported and produced in the peroxisome or shuttled into the ER for neutral lipid synthesis (TAGs, SEs).

Results

Increasing Cytosolic Fatty Acyl-CoA pools

The first attempt at engineering a strain to increase peroxisomal fatty acyl-CoA pools was to knockout *DGA1* (Figure 5.2) thereby reducing the loss of cytosolic acyl-CoAs to TAG synthesis. This engineered strain, PO1f*Δdga1Δfao1*, has been previously used for fatty alcohol production resulting in greater than 600 mg/L produced with a yield of 18 mg/g glucose [92]. However, their strategy used a different FAR that showed strong specificity towards C16 fatty alcohol production. To reach the reported titers, five copies of the FAR had to be expressed in the cytosol using media containing very little nitrogen (0.273 g/L) and very high glucose (160 g/L) amounting to a C:N ratio of

~1000:1. Severely retarded growth was reported that they attributed the cause being fatty alcohol production inhibiting growth.

Our engineered strain employed an episomally expressed a single FAR from *Marinobacter aquaeolei VT8*, MAACR, that included a C terminus PTS that should target it to the peroxisome. MAACR is known to have a broader range of activity towards fatty acyl-CoA substrate [327] with high activity toward short-chain acyl-CoAs (C8-C12), which is of special interest to our future goals. Furthermore, lower glucose concentrations (80 g/L) and ~10 times higher nitrogen conditions than Chen. *et. al.* [92] were used (C:N of 60:1) that resulted in better overall cell growth.

The initial hypothesis was that increasing the cytosolic fatty acyl-CoA pools should increase overall cellular fatty acyl-CoAs including that in the peroxisome. Cytosolic acyl-CoAs exists at the core of the metabolic pipeline and can be directed to various pathways such as beta-oxidation in the peroxisome, desaturation, elongation in the endoplasmic reticulum (ER), production of storage molecules such as TAGs in the ER then form into lipid bodies or conversion to FFAs that are excreted from the cell (Figure 5.2). The base engineered strain, PO1f*Δdga1Δfao1* with peroxisomal targeted fatty alcohol biocatalysis produced ~150 mg/L total fatty alcohols (Figure 5.3 A). To further increase the cytosolic acyl-CoA concentrations, a gene encoding for fatty acyl-CoA synthesis from free fatty acids, *FAA1*, was over-expressed. Surprisingly, this strain showed less fatty alcohol production dropping the total titers by ~ 33% compared to the original engineered strain.

High C:N ratios have been shown as a mechanism to upregulate pathways for lipid accumulation in *Y. lipolytica* [328, 329]. Although the DGA1 knockout should dramatically drop total neutral lipid accumulation, the high C:N ratio should still upregulate pathways favoring TAG accumulation. To release TAGs, the native TGL4 lipase was overexpressed. After 5 days of growth in synthetic media, over 200 mg/L fatty alcohols were produced, a ~25% increase compared to the base engineered strain (Figure 5.3A). The results indicate that increasing cytosolic fatty acid pools rather than fatty acyl-CoAs were more beneficial to peroxisomal fatty alcohol production. In either of the three engineered strains, the fatty alcohol distribution did not change (Figure 5.3B), showing higher percentages of C18 alcohols produced. This result is unique because C18 fatty acyl-CoA specificity has never been reported with MAACR before.

We were also able to detect C12 and to a lesser extent C10 fatty alcohols in the dodecane layer (Supplementary Figure 5.2 B) amounting to a total yield of greater than 100 mg/L for each strain (Supplementary Figure 5.2A). Since no C14 alcohols were detected, we hypothesize the production of the C12 and C10 medium chain fatty alcohols was a result of the active alkane metabolism in *Y. lipolytica*, oxidizing C12 alkanes to fatty acids that are then converted to fatty alcohols. C10 fatty alcohols are a result of a single cycle of beta-oxidation to produce C10 fatty acyl-CoAs that are converted to fatty alcohols. Although not initially expected, this result suggests that peroxisomal localized MAACR is a promising strategy to capitalize on the native beta-oxidation pathway to produce fatty alcohols. This result needs to be validated with dodecane fed as the sole carbon source to determine whether C10 and C12 alcohol production were a result of alkane oxidation and not beta-oxidation of fatty acids synthesized from glucose.

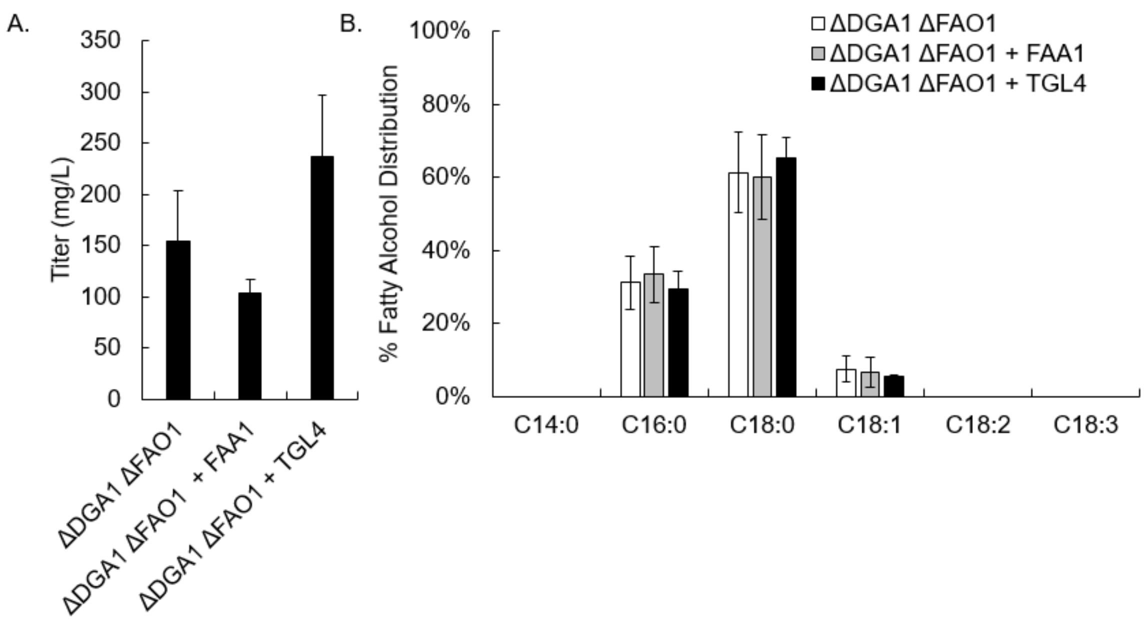


Figure 5.3: (A) Peroxisomal fatty alcohol production from increasing cytosolic acyl-CoA pools by knocking out DGA1p, a critical enzyme in TAG synthesis, or over-expressing the primary fatty acyl-CoA synthetase enzyme, FAA1p. TGL4 lipase over-expression, an enzyme that metabolizes TAGs to fatty acids in the cytosol, shows best titers. (B) Distribution of fatty alcohols produced in the three engineered strains. Samples were run in technical replicates (n=3) and represented error is a standard error.

Utilizing a Fatty Acid Responsive Promoter to Detect Intracellular Fatty Acid Production

In the second strategy, we wanted to build intracellular free fatty acid pools instead of fatty acyl-CoA to determine if this would improve fatty alcohol production. Therefore, our first goal was to engineer a strain capable of producing high amounts of free fatty acids. In Chapter 3, a fatty acid GFP based sensor was engineered to show high sensitivity and induction when extracellular fatty acids were used in the media. We were interested in testing whether the hybrid promoter can also be used to detect real-time changes in fatty acid pools produced intracellularly. If successful, this could serve as a powerful tool for engineering strains for high fatty acid production.

The episomally transformed hybrid fatty acid promoter fused to a GFP reporter was used to detect intracellular fatty acids in various engineered strains with knockouts and TLC was used to determine relative percentages of the two more important lipid classes, FFAs and TAGs (Figure 5.4A). Normalized fluorescence is a measure of cell fluorescence per OD_{600}, an indication of GFP accumulation in each cell over time. We wanted to determine if this signal would correspond to fatty acid accumulation over time. The PO1f wildtype strain shows no fluorescence and can be correlated to very little fatty acids as observed on TLC plates. The knockout strains, PO1f*Δdga1* and PO1f*Δpex10* both shows a slight change in fluorescence profile relative to the WT and also produced slightly higher free fatty acids. The two highest fatty acid producing strains were the PO1f*Δfaa1* and PO1f*Δmfe1Δfaa1* which resulted in intracellular fatty acid percentages of 30% and 54%, respectively. The intracellular GFP fluorescence profile in the PO1f*Δfaa1* first drops slightly and then increases suggesting an accumulation of fatty acids over time. The highest and most stable GFP expression profile was observed from the PO1f*Δmfe1Δfaa1* which correlated well with the high proportions of free fatty acids on TLC plates.

Interestingly, the PO1f*Δmfe1* strain also showed a dramatic increase in fluorescence followed by a sharp drop, although fatty acid abundance from TLC does not explain this phenomenon if we hypothesize that intracellular fatty acids can be correlated to *in vivo* GFP expression from the fatty acid inducible promoter. This led us to believe that there is complexity associated with detection that needs further understanding.

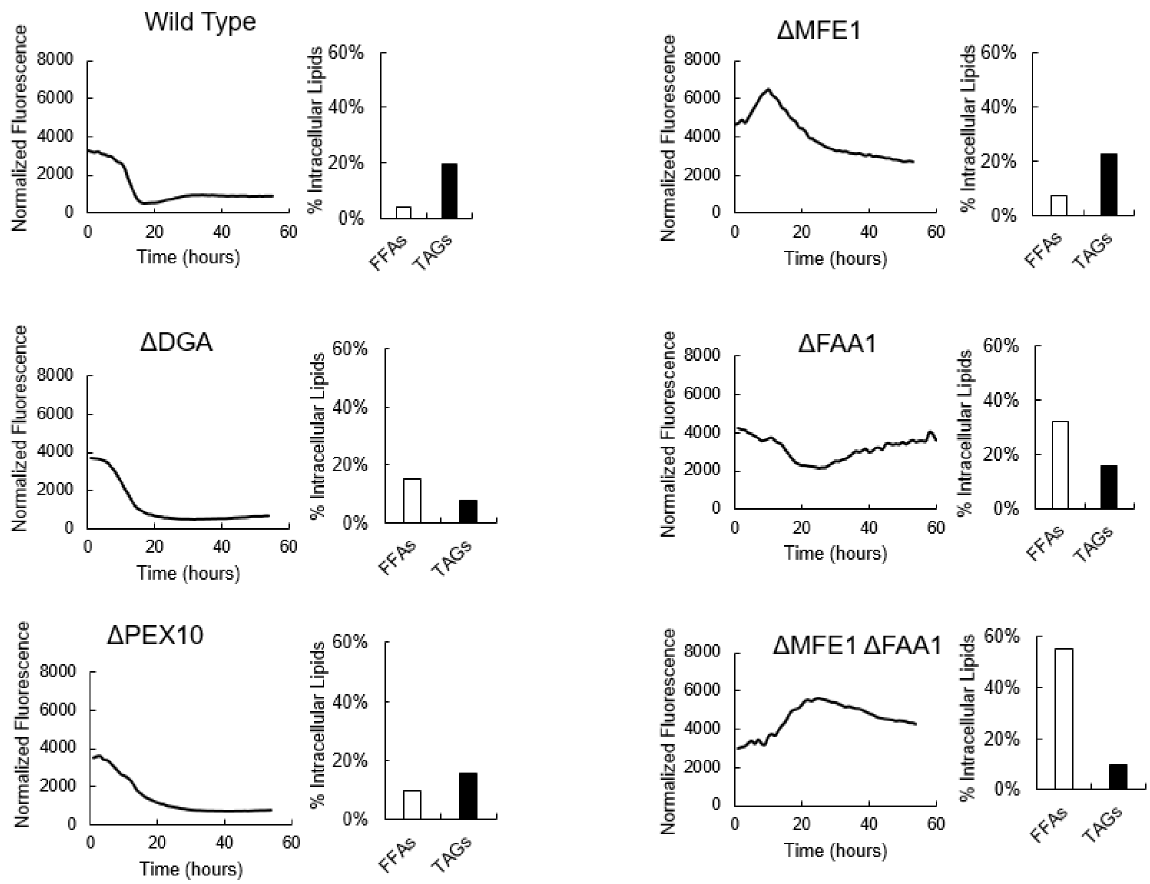


Figure 5.4. Fatty acid sensor (A1R1x3A3-GFP) used to detect intracellular fatty acid pools in *Y. lipolytica* engineered strains. The experiment was performed in a 48-well plate with cell growth and fluorescence recorded over time. Normalized fluorescence represents the ration of cell fluorescence to cell density to provide fluorescence quantification per OD_{600}. Densiometric analysis on TLC plates is performed to quantify percentages of fatty acids and TAGs in each of the engineered strains.

Characterization of Fatty Acid Hybrid Reporter

In Figure 5.4, the fluorescence profile PO1f*Δmfe1* shows a transient increase in GFP signal prior to tailing off rapidly. TLC profiles, however, show minimum proportions of fatty acids but the highest accumulation of TAGs from all engineered strains. From Figure 5.2, knocking out *MFE1*, a gene encoding the second step of beta-oxidation inside the peroxisome, should prevent fatty acyl-CoAs from shunting through the beta-oxidation pathway. The notable spike in expression from PO1f*Δmfe1* does bring into question whether this behavior could be related to a transient flux of fatty acyl-CoAs or free fatty acids into the peroxisome that cannot be degraded and is therefore transported back into the cytosol to be activated to TAGs. This raises two questions. Is the transient spike in GFP a result of transient fatty acid or fatty acyl-CoA pools inside the peroxisome? Is the hybrid promoter reporting fatty acids, fatty acyl-CoAs or both?

First, to test the impact of peroxisomes on GFP expression, cellular peroxisome formation was knocked out by deleting the peroxisome biogenesis factor, *PEX10* gene in the engineered strain that showed highest free fatty acid accumulation, PO1f*Δmfe1Δfaa1*. As a result, we observed a dramatic drop in fluorescence from the PO1f*Δmfe1Δfaa1Δpex10* (Figure 5.5A). TLC analysis showed a similar quantity of intracellular free fatty acids although the PO1f*Δmfe1Δfaa1Δpex10* had ~7-fold higher TAG accumulation. In the absence of peroxisomes, fatty acids were directed into storage as TAGs. This result suggests that intact peroxisomes are required for activation of the hybrid promoter. The regulatory process behind this association needs to be investigated further.

Next, we wanted to determine if the peroxisomal dependent intracellular hybrid promoter was detecting fatty acids or was it also exhibiting responsiveness to the activated form, fatty acyl-CoA. To test this hypothesis, the native peroxisomal fatty acid synthase gene, *YlAAL1*, was over-expressed to activate free fatty acids into fatty acyl-CoAs inside peroxisome (Figure 5.2). *Yl*AAL1 was reported to be a peroxisomal fatty acyl-CoA synthetase [330]. This experiment was also performed in the high fatty acid producing strain. AAL1 overexpression caused a drop in intracellular fluorescence over time (Figure 5.5B). From TLC, total free fatty acid pools remained the same between the two strains but ~5 times more intracellular TAGs were produced in the engineered strain overexpressing AAL1 (Figure 5.5B). This suggests that there is increased activation of fatty acyl-CoA activation resulting in the assimilation into TAGs. The experiment, however, falls short in answering if the hybrid promoter responds to fatty acids and/or fatty acyl-CoAs in the peroxisome. Metabolite analysis of fatty acids and fatty acyl-CoAs inside the peroxisome is required to know which exists in more abundance. The experiments required to elucidate the process will be discussed further in Chapter 6.

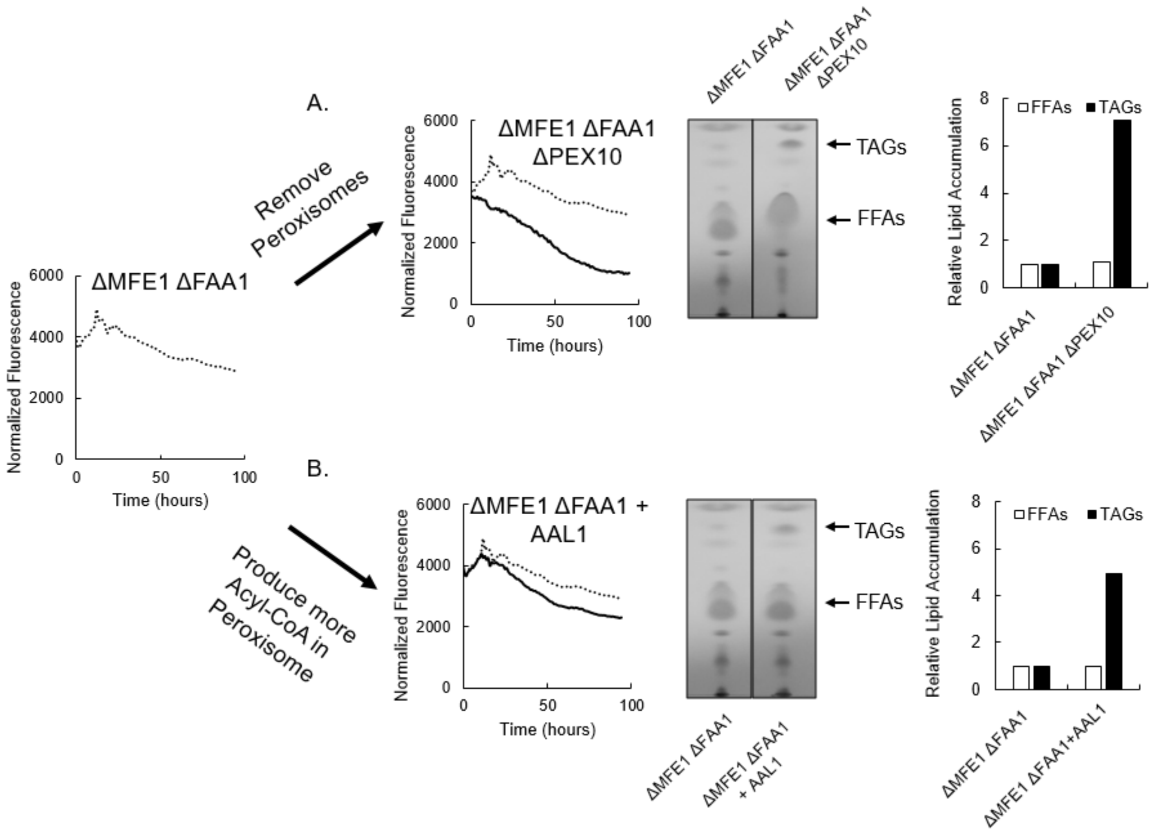


Figure 5.5. Characterization of intracellular sensor A1R1x3A3-GFP. (A) Removal of peroxisome via ΔPEX10p shows a significant drop in fluorescence. From the densitometric analysis of TLC plates, the fatty acid percentage remains the same between both strains while TAG accumulation increases 7-fold in PO1f*Δmfe1Δfaa1Δpex10* strain. (B) Over-expression of AAL1 in high fatty acid producing strain, PO1f*Δmfe1Δfaa1*, causes the fluorescence signal to drop over time. Densiometric analysis on TLC plates shows similar percentages of FFAs while TAG accumulation increases 5-fold. Fluorescence experiments were performed in a 48-well plate with cell growth and fluorescence measured over time. Normalized fluorescence represents the ration of cell fluorescence to cell density to provide fluorescence quantification per OD_{600}. Cells from growth experiments were harvested for TLC experiment.

Fatty Alcohol Production from Engineered Fatty Acid Producing Strain

PO1f*Δmfe1Δfaa1* has previously been shown to produce large amounts of intracellular and extracellular fatty acids surpassing 2 g/L combined [89] making it a good starting point for a fatty alcohol production platform. Analysis of the fatty acid distribution of the intracellular pool revealed that there is a predominant percentage of C14 and C16 fatty acids produced which is different to the fatty acid profiles of total lipids, containing predominantly unsaturated fatty acids, C16:1 and C18:1 (Supplementary Figure 5.3 A, B). Therefore, we anticipated C14 alcohol production. However, to our surprise, no C14 fatty alcohols were produced.

The media used had a lower C:N ratio of 30:1 to facilitate peroxisome formation. High nitrogen concentration switches the cells global regulatory circuit from lipid accumulation to fatty acid degradation which requires peroxisomes [331]. In the wildtype overexpressing *AAL1*, very little fatty alcohols were produced, predominantly being C16:0 and C18:0 (Figure 5.6 A, B). In PO1f*Δmfe1Δfaa1*, we observed a 10-fold improvement in fatty alcohol production reaching 530 mg/L and 14% of dry cell weight (Supplementary Figure 5.5). In addition to a predominant abundance of C16:0 and C18:0, small proportions of monounsaturated alcohol peaks C16:1, and C18:1 were detected.

In attempts to further increase the fatty acyl-CoA pools inside the peroxisome, *AAL1* was over-expressed. Surprisingly, this perturbation to the pathway resulted in a 35% drop in fatty alcohol production. The hybrid promoter was previously used to test GFP expression in PO1f *Δmfe1 Δfaa1* over-expressing *AAL1* (Figure 5.5B). We observed fluorescence drop over time relative to PO1f*Δmfe1Δfaa1*. Whether this effect can be attributed to an increase in peroxisomal fatty acyl-CoA pools needs to be determined. However, higher TAG accumulation was observed in the *AAL1* over-expressing strain suggesting higher levels of fatty acid activation to fatty acyl-CoA.

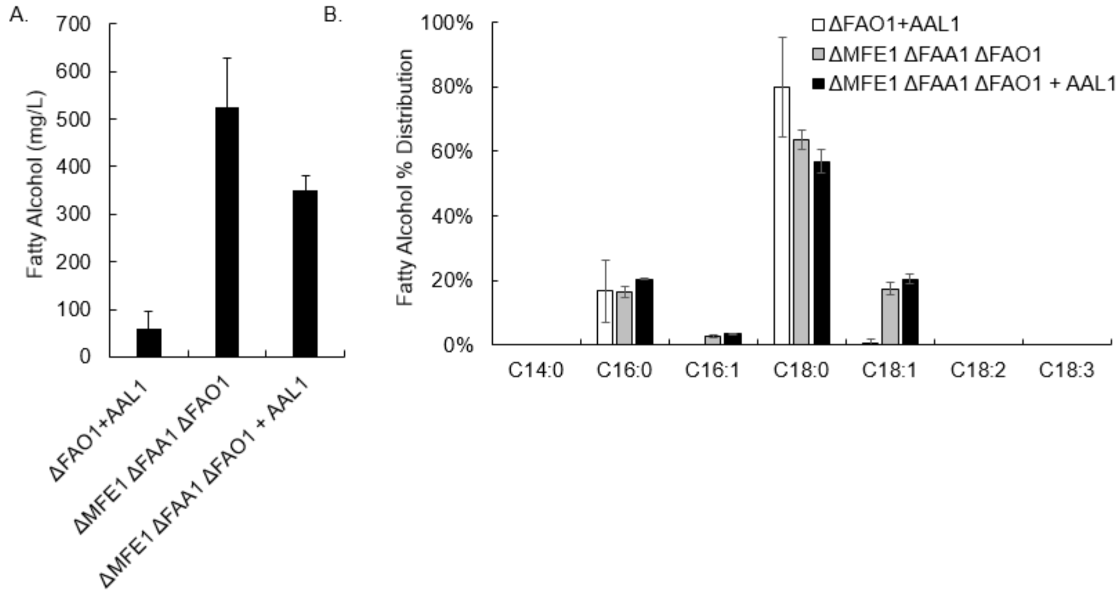


Figure 5.6. (A) Peroxisomal fatty alcohol production from increasing cytosolic fatty acid pools and beta-oxidation by knocking out fatty acid synthase, FAA1 and multifunctional enzyme, MFE1 produced highest titers exceeding 500 mg/L fatty alcohols. Over-expression of AAL1 in engineered strains leads to a ~30% drop in titers. (B) Percentage fatty alcohol distribution of all three strain show similar profile except the engineered strains, PO1f*Δmfe1Δfaa1Δfao1* and PO1f*Δmfe1Δfaa1Δfao1* with AAL1 over-expression shows production of unsaturated fatty alcohols, C16:1 and C18:1. Samples were run in technical replicates (n=3) and represented error is a standard error.

Discussion and Conclusion

Developing microbial platforms to produce fatty alcohols has garnered a lot of interest in recent years with the exploitation of several microbes ranging from bacterial systems such as *E. coli*, cyanobacteria and *Marinobacter aquaeoli* [322, 332-334] to yeasts, predominantly *S. cerevisiae* [308, 310, 315, 335]. One of the benefits of exploring yeast systems for this function is that metabolic processes are naturally separated into specialized yet distinct subcellular compartments called organelles. For example, in yeasts, beta-oxidation of fatty acids is localized into the peroxisome or transported into the ER to be stored in the form of TAGs. Furthermore, fatty acids produced in the cytosol via the fatty acid biosynthesis are bound to acyl-CoA binding proteins and can be transported either into the peroxisome or the ER. In comparison to conventional engineering of pathways in the cytosol, compartmentalization provides additional advantages such as enabling faster reaction rates due the ability to concentrate metabolites and enzymes. One of the more predominant advantages of pathway localization is the capacity to segregate the biochemical process of interest from competing pathways and mitigating regulatory responses at the protein level [336].

The metabolic strategy we propose is a novel method to explore fatty alcohol production in *Y. lipolytica* inside the peroxisome. This provides the ability to highjack the hydrocarbon chain processing capabilities via beta-oxidation to eventually produce shorter chain fatty alcohols. By increasing the cytosolic fatty acid pools instead of the cytosolic fatty acyl-CoA pools, we were able to produce more than 2-fold higher fatty alcohol titers in the peroxisome. These results suggest that fatty acids may enter the peroxisome more readily than fatty acyl-CoA. The observation can be validated by quantifying fatty acid and fatty-acyl CoA pools in the peroxisome.

PO1f*Δmfe1Δfaa1* strain produced the most amounts of fatty acids as evidenced by the hybrid promoter signal and TLC analysis. We believe this strategy of over-producing fatty acids enables better diffusion of fatty acids to enter the peroxisome for fatty alcohol conversion.

In attempts to improve the efficiency of fatty alcohol production inside the peroxisome, *AAL1* gene was over-expressed to pull fatty acids towards fatty acyl-CoA production, which are direct precursors to MAACR. This strategy, however, led to a ~30% drop in fatty alcohol titers and increase in TAGs as evidenced by TLC. Given that AAL1 expression is peroxisomal, we could at this point only hypothesize that higher production of fatty acyl-CoA is imbalanced by a less active MAACR leading to activation of fatty acyl-CoA export from the peroxisome to the cytosol to be available for TAG synthesis. This bottleneck can be overcome by improving the activity of FARs, regulating peroxisomal fatty acyl CoA pools, and blocking TAG synthesis.

To improve the fatty alcohol reductase activity, a codon optimized version of the *MAACR* gene will be expressed. We anticipate better translation efficiency of the gene, thereby increasing enzyme concentration and overall kinetics of the reaction. Codon-optimized versions of other FARs, particularly, TaFAR1 and MmFAR1 should be tested. TaFAR1 has already been shown to be functional in *Y. lipolytica* [92]. Meanwhile, MmFAR1, from the mouse, *Mus musculus*, is yet another fatty alcohol that has proven to show high activity when expressed in *S. cerevisiae*, greater than TaFAR1 [65]. There is no published literature to date on its activity in *Y. lipolytica*. The activity and fatty alcohol profiles generated from these FARs in our engineered strain may provide new alcohol distributions and improve the overall efficiency of the final production step, thereby improving yields.

The PO1f*Δmfe1Δfaa1* strain produced significant C14 fatty acids (Supplementary Figure 5.3) although, with MAACR, no C14 fatty alcohols were produced. We hypothesize MAACR has a lower affinity to C14 when there is a high abundance of C16 fatty acyl-CoA. Furthermore, lower specificity of MAACR to C14 aldehydes has been demonstrated before [332]. The inability to convert abundant C14 fatty acids to fatty alcohols affects the yields and titers of the process and therefore motivates our work to investigate additional FARs.

We can attempt to engineer higher accumulation of fatty acyl-CoA inside the peroxisome but this is challenging because there is little known about the mechanism of fatty acyl-CoA transport and regulation to and from the peroxisome [169]. The dimeric ATP- dependent transporter, YlPxa1p/ Pxa2p, is hypothesized to transport fatty acyl-CoA into the peroxisome, whether this transport is reversible is unknown. This transporter has been knocked out previously to increase cytosolic acyl-CoA pools, however, no change to fatty alcohol production was observed [92]. This result suggests that fatty acyl-CoA transport to the peroxisome may be more regulated than we currently understand. Another approach would be to place the expression of *YlAAL1* under the expression of the fatty acid hybrid promoter. Instead of attempting to create stagnant fatty acyl-CoA pools, dynamically regulating the expression of the protein by placing the gene under a fatty acid hybrid promoter could improve process efficiency.

Since we are observing more TAG accumulation due to peroxisomal *AAL1* over-expression, it would also be beneficial to block TAG formation via Δ*DGA1* in the engineered PO1fΔ*mfe1*Δ*faa1*Δ*fao1*. This strategy could push fatty acyl-CoA flux back into the peroxisome to improve production efficiency inside the peroxisome.

While glucose may be a convenient substrate, it may not ideal from a standpoint of making fatty alcohol production economically viable (Table 5.2). The advantage to using *Y. lipolytica* as the model yeast to engineer fatty alcohol production as opposed to *S. cerevisiae* provides several benefits, notably, its ability to uptake and metabolize hydrocarbons efficiently. This opens opportunities to utilize cheaper and more abundant alternatives as feed such as crude glycerol from the biodiesel waste stream or rendered animal fats. Reducing feed cost can bias economic profitability of the process.

Two alternate methods have been explored for localized production of fatty alcohols inside the peroxisome of *Y. lipolytica*. In the first pathway, we attempted to increase cytosolic acyl-CoA pools by knocking out the lipid storage capabilities. Higher fatty acyl-CoA pools were expected inside the peroxisome, however, this strategy proved less efficient with the production of slightly over 200 mg/L of fatty alcohols. In the next strategy, that proved more promising, higher fatty acid production was engineered in *Y. lipolytica* by knocking out the cytosolic acyl-CoA activation step and preventing fatty acid degradation via beta-oxidation. Using a fatty acid inducible promoter fused to a GFP reporter gene, real-time production of intracellular fatty acids was monitored. Using this strategy, close to 600 mg/L of fatty alcohols were produced. Attempts to further increase production were tested by overexpressing the fatty acyl-CoA production pathway inside the peroxisome, however, this led to a decrease in fatty alcohol production with increased TAG accumulation. The work presented here highlights the need to explore mechanisms to improve the efficiency of fatty acyl-CoA conversion to fatty alcohol, better characterization of the hybrid promoter so it could be utilized more efficiently, and engineering fatty acyl-CoA pools inside the peroxisome. This work is the first of its kind to engineer localized fatty alcohol biosynthesis in *Y. lipolytica*.

CHAPTER SIX

Conclusions

The work in this dissertation establishes tools to improve engineering efforts in the industrial yeast, *Yarrowia lipolytica*. Metabolic pathways in microbes were constructed with enzymes and controlling the time and strength of expression can improve the efficiency of biochemical processes. A large part of this work focused on understanding and developing better gene expression systems in *Y. lipolytica*. Hybrid promoters with tunable and predictable strengths can be engineered by combining different elements that constitute a promoter, namely, UASs, proximal sequences, TATA element and core promoters. In particular, UAS elements in hybrid promoters were used to control strength and timing of expression. Identifying UASs, however, is a limitation in non-conventional yeasts such as *Y. lipolytica* where there is little known about regulatory elements such as transcription factors or its binding motifs. Therefore, we explored a novel method to screen for new UASs from differences in DNA accessibility profiles in native promoters under different conditions. We explored this concept with the POX2 promoter and anticipate using it to determine novel UAS elements in other native promoters. Finally, a fatty acid hybrid promoter detailed in this dissertation was used to guide the engineering of a fatty acid producing strain of *Y. lipolytica*. By localizing the final step of fatty alcohol biosynthesis to the peroxisome, the engineered strain shows promise for future metabolic engineering geared to short and medium chain fatty alcohol production. There are still a number of interesting opportunities for further investigation of hybrid promoters and application of these promoters to improve fatty alcohol production efficiency.

Future Work

In Chapter 1, libraries of hybrid promoters were built by investigating the architecture of native promoters in *Y. lipolytica.* In addition to Upstream Activating Sequences (UASs), we studied the role of sequences near the transcriptional start site, namely the proximal, TATA box and core promoters from four promoter systems upstream of the genes, *YlPOX2*, *YlPAT1*, *YlLEU2*, and *YlTEF1-α*. The goal of this work was to elucidate how each these elements contributed to promoter activity, which of the elements had modular properties and most importantly, which elements contributed the most to promoter activity.

The TATA box is a highly conserved element across all eukaryotic species and is the recruitment site for the pre-initiation complex (PIC) machinery that initiates transcription [337]. Although the TATA element itself is conserved, there is a lot of sequence variability surrounding this element that can affect transcription [338]. The TATA element from the *TEF1-α* promoter elicited the strongest transcriptional response of the four promoters tested. Another interesting result was that the POX2 promoter contained two TATA boxes. Removing one TATA element severely weakened transcription. Our experiments were limited to four promoters with an 8 bp window around the TATA sequence.

I recommend increasing this sequence space to test more base pair diversity around the TATA box. It would be beneficial to determine if there is a maximum threshold for transcriptional activation using TATA elements by varying sequence diversity around the conserved TATA box. This information can be leveraged to increase the strength of inherently weak native promoters or applied to the development of hybrid promoters. The dual TATA box activity of the POX2 promoter strongly suggests that adjacent sequence can play a pivotal role for PIC affinity to the TATA element.

This experiment can be performed on one core promoter such as the POX2 core promoter than can be used as the control of the experiment. Oligo mixes of ~60 bp can be synthesized with 6-8 random base pairs (N) before and after a conserved TATA box. Each end in the mix of forward oligos should have the same 20 bp sequence while the ends of reverse oligo mix should be complementary to the ends of the forward oligos. This would allow the oligo mix to be PCR amplified. The PCR amplified product can then be cloned downstream of a UAS sequence and upstream of a GFP reporter in the pSL16 shuttle vector. By transforming these vectors into *E. coli* DH10β cells, a library of vectors with variations to the sequence before and after the conserved TATA element can be created and sequence verified via Sanger sequencing. As a first round, around twenty different sequence variations can be transformed into *Y. lipolytica* and fluorescence can be measured in 48-well plates to determine which sequence provides the strongest expression relative to POX2 core promoter. Sequences of the optimal expressing core promoters from *Y. lipolytica* can be screened using colony PCR and Sanger sequencing

The initiator sequences in the core promoter are another under-studied promoter element that was not covered in Chapter 2. In Supplementary Figure 2.1, we identified potential initiator regions in the four promoters based on homology to metazoan sequences. In *S. cerevisiae*, RNA polymerase II performs a downstream scan in search of transcriptional start sites (TSSs) [339] that results in transcriptional initiating occurring 40-120 bps downstream of the TATA box [197]. Given that core transcriptional machinery is strongly conserved in eukaryotes [340], it is likely that the identified initiator regions in *Y. lipolytica* have some impact the transcription process. Therefore, the significance of initiator sequences for transcriptional regulation needs to be further examined.

The experimental set up to test for the effect of initiator sequences is similar to the process explained above for TATA elements. In this instance, we have mapped out putative initiator sequences in four core promoters, therefore, we have an idea of the sequence length that needs to be randomized using oligos.

In Chapter 3, we set out to engineer the first hybrid fatty acid inducible promoter by identifying a key promoter element for transcription, enhancer binding regions, found within UASs. The focus, however, was to search for UAS elements that elicited inducible activation towards fatty acids. The *POX2* promoter of *Y. lipolytica* was used because it was induced by fatty acids and repressed in glucose and oleic acid [110]; however, its inducibility was weak, which makes its application in metabolic engineering limited. A heuristic approach was used to dissect the native promoters and fused to a GFP fluorescence reporter to quantify expression strength. We identified three oleic acid-inducible UASs, A1, R1, and A3 via truncations of the promoter. Tandem repeats of R1 enabled strong activation when oleic acid was used as an extracellular inducer. Substituting in stronger core promoters led to higher activation of the hybrid promoter. The strongest hybrid promoter had an ~10-fold increased activation in oleic acid compared to the native POX2 promoter.

The promoter was truncated such that we conserved the putative *Yl*POR1p binding sites within the truncations. Furthermore, in Chapter 4, we discuss DNA accessibility in relation to where the POR1p motifs are located on the promoter. However, to date, there has been no direct line of evidence suggesting that POR1p physically binds the POX2 promoter. BLAST analysis of the POR1p reveals a conserved DNA binding and activator domain belonging to the zinc finger transcription factor family.

We should first demonstrate that POR1p binds to the POX2 promoter, and to determine DNA binding specificities to leverage this information towards building a library of hybrid promoters using smaller, well defined yet optimal UASs. As discussed in this dissertation, a commonly used approach to building stronger hybrid promoters relies on using tandem repeats of UASs. If the binding affinity of a transcription factor is weak in a UAS, then the tandem repeats of a UAS improves transcription strength by increasing number of transcription factors, thereby increasing the strength of transcription. The more efficient method would be to use few UASs containing motifs with higher binding affinity sites. The question remains as to how we search for these motifs.

To answer this question, I recommend an *in vivo* approach to determining POR1p DNA binding motifs by not only looking at the POX2 promoter but the whole genome. Chromatin Immunoprecipitation (ChiP) combined with either PCR tiling arrays or high throughput sequencing (ChiP-seq) can be used to locate POR1p binding sites in the POX2 promoter or the entire genome. POR1p attached to an epitope tag (Myc, V5 or HA) can be immunoprecipitated using antibodies after crosslinking the cells in formaldehyde (Figure 6.2). Proteinase digest of the pull-down should expose DNA protected by POR1p.

In addition to elucidating the role of POR1p, we are also interested in identifying other co-regulators that associate with POR1p to mediate transcription. For this, I recommend ChiP combined to tandem mass spectrometry (ChiP-MS). The crosslinked and immunoprecipitated protein sample is subjected to enzymatic or chemical degradation to produce peptides that can then be sequenced using MS. The peptide sequences can be blasted against the *Y. lipolytica* protein database to identify the transcription factors.

Usually, ChiP this may yield a larger than desired library of proteins due to non-specific binding of the antibody or proteins. Therefore, there may be a need to narrow down the pool of proteins that are potential co-regulators. For this purpose, I recommend a yeast 2-hybrid promoter using POR1p as the bait and other identified proteins (preys) fused to a well-characterized activation domain. Using this system, we can elucidate the mechanism of oleic transcription activation and apply this system to engineer new types of hybrid promoters in *Y. lipolytica*.

In Chapter 4, we wanted to investigate whether UASs in native promoters can be identified by scanning DNA accessibility on the promoter. To test this, we used the fatty acid inducible POX2 promoter that was characterized in Chapter 3. The conclusion from this chapter was that we could begin to investigate new UASs based on differences in DNA accessibility between conditions that induce or repress transcription. Choosing UASs using this method reduces the number of combinations of UASs to be tested compared to using a truncation method in Chapter 3 that can generate many truncated sequences.

We would like to test this in other native promoters where there is no understanding of transcription factors or putative binding sites. We hope to determine whether differences in DNA accessibility across the promoter can guide the selection of UASs. We are currently in the process of testing this on another promoter that is induced by glycerol [224] from the gene, glycerol kinase (*YlGUT1*) that converts glycerol to glycerol-3-phosphate (G3P) [341]. Prior to publishing work in Chapter 4, there are other aspects of the project I would recommend testing, detailed below.

Recently, it was reported that there is an inherent bias in mononucleosome DNA preparation using micrococcal nuclease (MNase) [342]. Higher MNase activity by using higher concentrations of MNase releases mononucleosome DNA from regions of low DNA accessibility while low MNase activity has been used to map positions of “fragile” nucleosomes, which are nucleosomes harboring post-translational modification that make its interaction with DNA weak [343]. Crosslinking the histones to DNA can mitigate this issue. The approach that we used to “standardize” occupancy maps was to test more than one MNase concentration and choosing a sample that showed a >90% mononucleosome DNA fraction but this still does not remove the inherent bias that MNase cleavage can have on reliably determining nucleosome probability.

The purpose of nucleosome profiling is to quantify the probability of a nucleosome forming in a region of the promoter in a DNA. When two nucleosomes are close to one another, and if the probability of this event happening is high, our data can be skewed by the MNase digestion affinity to separate the two nucleosomes because they are in proximity to one another. This leads to different populations of MNase-resistant and MNase-sensitive nucleosome that biases the DNA accessibility profiles (Figure 6.2).

My recommendation to troubleshooting this issue is (1) test nucleosome profiles for more than one MNase digestion condition and see if the DNA accessibility profiles are reproducible (2) attempt another method such as sonication to shear DNA into mononucleosome fragments after crosslinking, and (3) using ChiP to separate association of transcription factors and nucleosomes in DNA accessibility profiles.

When MNase digestions are performed, more than one digestion condition can provide greater than >90% mononucleosome DNA fractions. At this point, it is left up to the discretion of the experimenter to select a sample to move onto PCR tiling for DNA accessibility profiles. My recommendation here is that if we have more than one digestion condition that yielded >90% mononucleosome DNA fraction, then we should do DNA accessibility on all MNase digested samples with very high mononucleosome DNA and investigate the reproducibility of the profiles. We need to access the level of variability and determine whether this could significantly affect our experimental goals, i.e. to efficiently identify UAS elements on the promoter.

My second recommendation is to attempt another method to retrieve mononucleosome DNA. If the chemical means such as MNase activity causes technical bias, then a physical process of preparing mononucleosome DNA such as sonication should be used to investigate if it is a better solution.

My final recommendations for Chapter 4 are to separate the difference between transcription factor binding to the DNA and nucleosomes occupancy. To obtain DNA accessibility profiles in Chapter 4, we used formaldehyde crosslinking to covalently bind the nucleosomes to the DNA, however, this also covalently binds transcription factors to the DNA (Figure 6.1B). How transcription factors bind to DNA in the presence of nucleosomes can vary depending on the type of mechanism. In *S. cerevisiae* alone, nucleosome eviction [277] and sliding [280] have been reported. More recently, nucleosome “loosening” due to post-translational modifications enable transcription factors to access binding motifs wrapped in nucleosome [342]. This mechanism does not change nucleosome profiles. Therefore, looking at the DNA accessibility data in the presence and absence of transcription factors can provide some mechanistic insight into transcription factors association with nucleosomes to regulate gene expression.

One method I propose is to investigate DNA accessibility without crosslinking with formaldehyde (Figure 6.1C). Transcription factors are known to have weaker DNA interactions that nucleosome. Without crosslinking, transcription factors would disassociate from the DNA easily. This experiment will be difficult to interpret because crosslinking fixes nucleosomes to DNA that are otherwise dynamic. Without crosslinking, nucleosome can move during MNase treatment but the extent of this effect is not known in *S. cerevisiae* or *Y. lipolytica*. Without crosslinking, we would also not be protecting fragile nucleosomes on the promoter. The results from this experiment compared to the crosslinking data would be beneficial to predict promoter regions where nucleosomes have undergone post-translational modifications or regions protected by transcription factor complexes. Without performing the experiment, it is difficult to determine how important this data would be to better elucidate transcriptional regulation via nucleosomes and transcription factors.

ChiP for histones is a better-suited method prevent the interference of histones being dynamic and yet provide us with nucleosome occupancy data in the absence of transcription factor-DNA interactions. Cross-linking is possible in this step and immunoprecipitation is done using histone antibodies or beads targeted at histone modifications [344]. The difference in nucleosomes profiles between of histone modifications between repressed and induced conditions can provide insight into potential UASs.

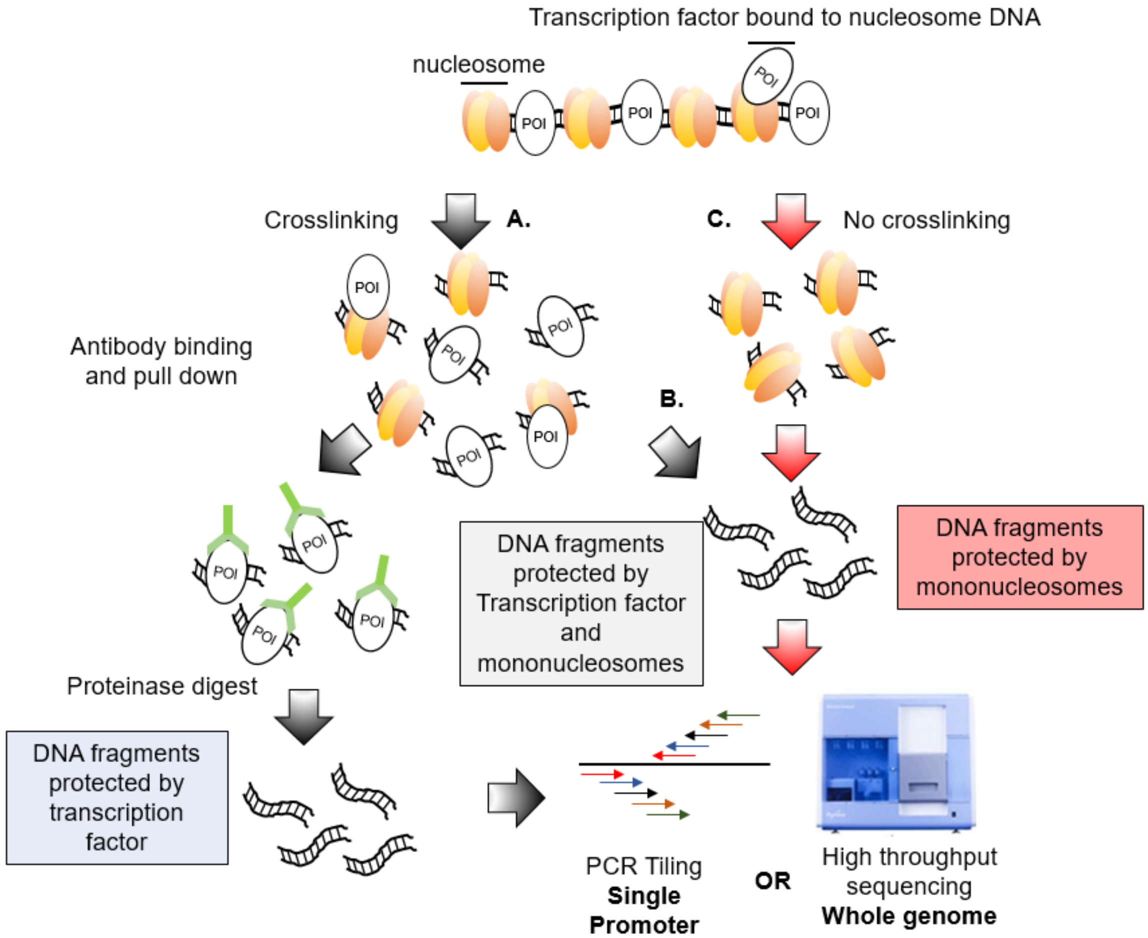


Figure 6.1. ChiP can be used for the detection of protected DNA regions either protected by nucleosomes and transcription factors via crosslinking or nucleosome protected DNA by not crosslinking linking the reaction. Protected DNA on a single locus of the genome can be analyzed using PCR Tiling or protected DNA across the entire genome can be sequencing using a high throughput sequencing technique.

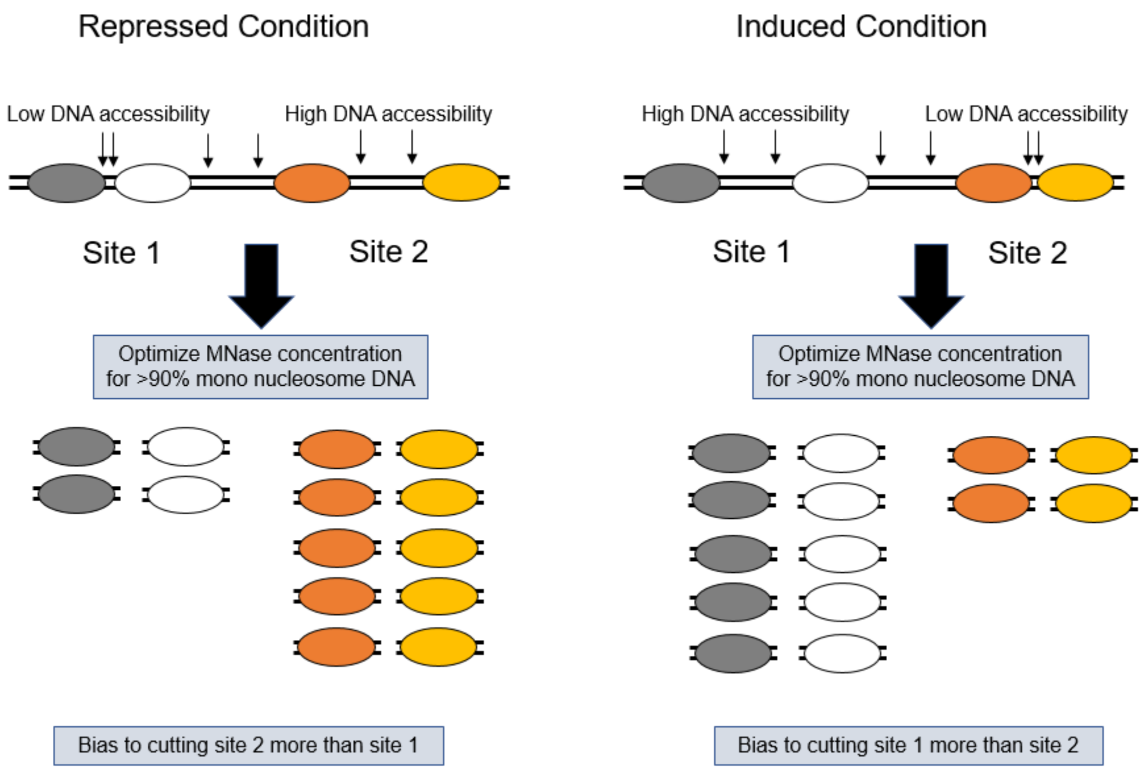


Figure 6.2. MNase digestion can bias DNA accessibility profiles based on cutting efficiency of the MNase. This biases nucleosome occupancy data that should capture probability of finding a nucleosome in a given region of the promoter. ↓ shows MNase cleavage sites

In Chapter 5, a novel pathway for fatty alcohol production is examined in *Y. lipolytica* by docking the final step of the enzymatic reaction, conversion of fatty acyl-CoA to fatty alcohol, inside the peroxisome using a peroxisome targeting signal (PTS). The enzyme used to convert fatty acyl-CoA to fatty alcohols is a fatty acyl-CoA / aldehyde reductase (FAR) from *Marinobacter aquaeoli*, MAACR. We demonstrate that by engineering the strain to produce intracellular fatty acids, upwards of 500 mg/L fatty alcohols can be produced, although additional work needs to be done to improve the efficiency of the process. The current process yields 7 mg fatty alcohols / g glucose. The future work for Chapter 5 will address two key issue. (1) Improving the efficiency of fatty alcohol biosynthesis in *Y. lipolytica*, and (2) characterization of the fatty acid hybrid promoter for future strain engineering applications.

The work demonstrated in Chapter 5 is a promising route for fatty alcohol production, however, there are different aspects of the metabolic pathway that need to be addressed to improve its efficiency. These are separated into modules as depicted in Figure 6.3.

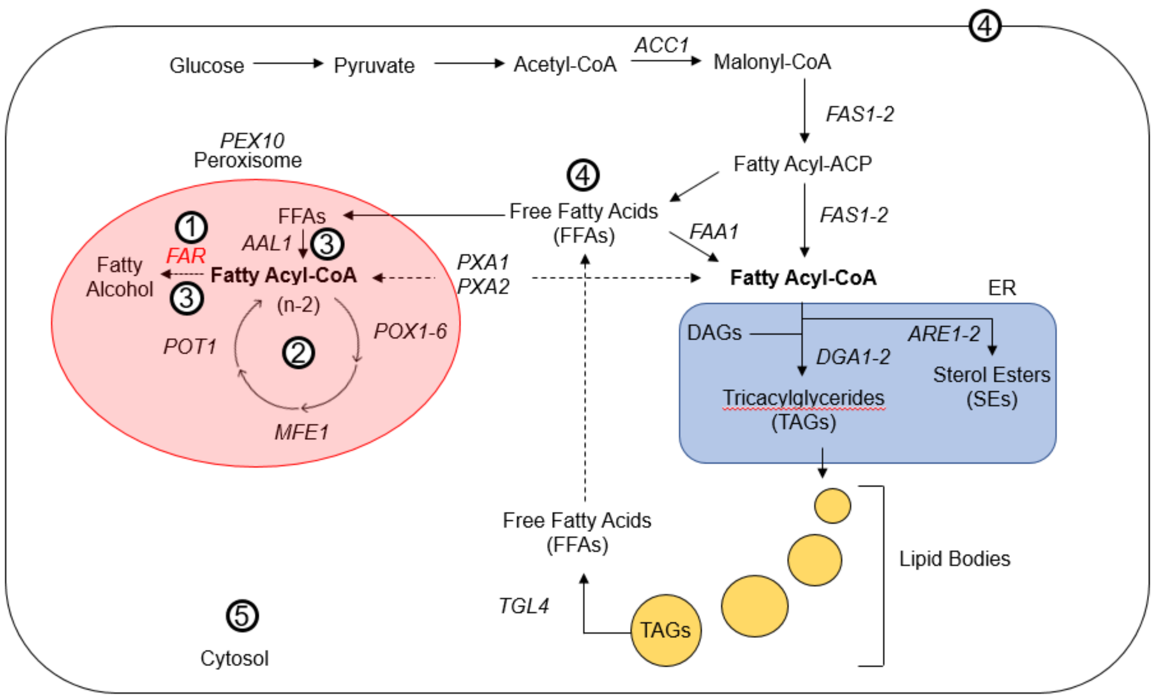


Figure 6.3. Five modules identified to improve the efficiency of engineering fatty alcohol production in *Y. lipolytica*.

Module 1:

Improving final reaction step. Figure 5.6A showed that attempting to increase fatty acyl-CoA inside the peroxisome, by overexpressing *Yl*AAL1, led to lower fatty alcohol production. Fatty acyl-CoA are direct precursors for MAACR therefore, we anticipated that this strategy would yield better fatty alcohol titers. One hypothesis to explain this result is that the expression of MAACR is not optimal because the MAACR is not codon optimized for *Y. lipolytica*. The MAACR is being expressed using the UAS1B8-TEF hybrid promoter, shown to be strong and constitutively expressed [82]. However, codon biases in microbes can have a large impact on the efficiency of gene expression [345]. We believe that codon optimization of MAACR for *Y. lipolytica* should relieve this bottleneck.

Additionally, two more FARs, MmFAR1 (mouse) and TaFAR1 (barn owl) will be tested in *Y. lipolytica* to observe if better fatty alcohol production and different chain length specificities can be attained. We are unable to produce C14 alcohols although the engineered strain, PO1fΔ*mfe1*Δ*faa*1Δ*fao*1, produces a large abundance of C14 fatty acids (Supplementary Figure 5.3B). One of the goals is to investigate whether changing the FAR could improve selectivity for C14 fatty alcohol production.

Module 2: *Chain shortening using native β-oxidation*.

One of our primary objectives for compartmentalizing fatty alcohol synthesis to the peroxisome was to use the β-oxidation pathway to produce short and medium chain fatty alcohols. The work described in Chapter 5 does not test chain shortening since the *MFE1* knockout in our engineered strain impairs beta-oxidation. To test fatty alcohol distribution via natural beta-oxidation, MFE1 function will be restored in the engineered strain to create PO1fΔ*faa1*Δ*fao1*. This strain capable of beta-oxidation and yet accumulates fatty acids. Biasing the chain length specificity to short and medium-chain fatty alcohols can be accomplished by knocking out the *POX3* gene. *POX3* has been shown to be a short chain fatty acyl-CoA oxidase [170].

Module 3: *Dynamic regulation of fatty alcohol production.*

Our current engineering strategy utilizes two constitutively hybrid promoters to express *AAL1* and *MAACR* genes encoding enzymes for fatty acyl-CoA and fatty alcohol production, respectively. To increase peroxisomal fatty acyl-CoA, *AAL1* was placed under the expression of an early phase, strong constitutive promoter, $P_{TEF\text{-}intron}$ from the *YlTEF* gene of *Y. lipolytica*. The expression of *MAACR* was driven by UAS1B8-TEF, a strong late phase hybrid promoter. While *AAL1* over-expression should have created more fatty acyl-CoA precursor for MAACR, a negative effect on fatty alcohol production was observed in PO1fΔ*faa1*Δ*mfe1*Δ*fao1 + AAL1* relative to the engineered strain without *AAL1* over-expression. An inadequate pull on fatty acyl-CoA pools as described in module 1 coupled to up-regulation of pathways to transport fatty acyl-CoA outside the peroxisome to maintain homeostasis could be a reason for the result we observe. I propose using a fatty acid inducible promoter to regulate *AAL1* expression thereby using intracellular fatty acids to dynamically regulate fatty acyl-CoA production inside the peroxisome (Figure 6.4). In addition, it would be beneficial to test if improvement to fatty alcohol production is attainable via combinations of constitutive and hybrid promoters to drive expression of the *AAL1* and *FAR* genes.

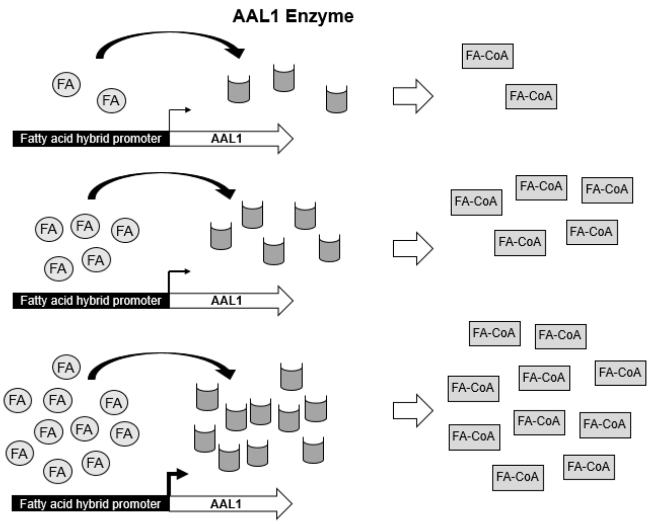


Figure 6.3. Using fatty acids to dynamically regulate fatty acyl-CoA pools inside the peroxisome should prevent fatty acyl-CoA build-up. Coupled with constitutive expression of a codon-optimized FAR, the process should improve fatty alcohol production efficiency. FA = Fatty Acids; FA-CoA = Fatty acyl-CoA; AAL1 = fatty acyl CoA synthetase

Module 4: *Maintaining high intracellular fatty acid pools.*

While PO1fΔ*mfe1*Δ*faa*1 and PO1fΔ*mfe1*Δ*faa*1+*AAL1* creates large amounts of intracellular fatty acids (>1 g/L), we also observed high levels of extracellular fatty acids secreted to the media (Figure 6.4)

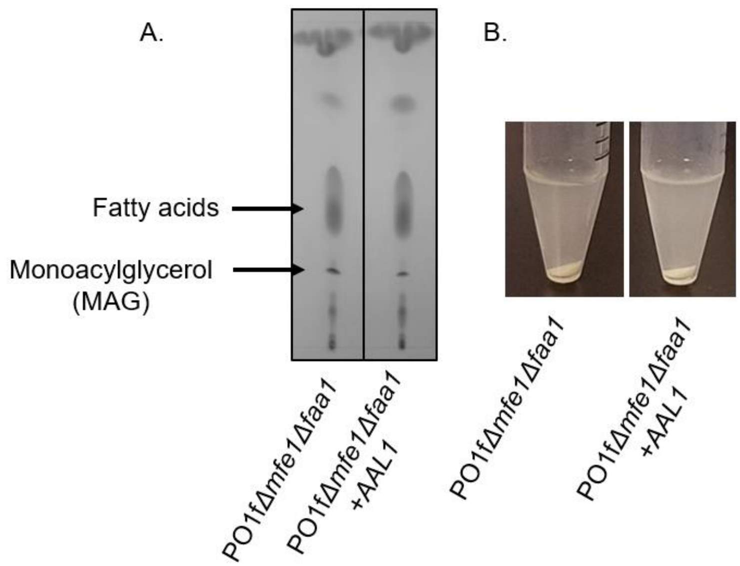


Figure 6.4. (A) Thin Layer Chromatography (TLC) for analysis of lipids in secreted to the media in strains PO1fΔ*mfe1*Δ*faa*1 and PO1fΔ*mfe1*Δ*faa*1+*AAL1*. (B) Visible turbidity in media due to fatty acids for both strains.

The same analysis has not been performed on the engineered strain overexpressing MAACR, therefore, validating similar extracellular fatty acid accumulation is necessary prior to engineering the strain to accumulate intracellular fatty acids.

To date, there has been no published record for fatty acid transporters in *Y. lipolytica*, therefore, how fatty acids are transported in and out of peroxisomes or cells are unknown. Our results suggest that there may be no current limitation to fatty acid transport into the peroxisome to be activated for fatty alcohol production. However, we would need to first identify transporters responsible for fatty acid transport in and out of the cell to engineer a strain capable of higher intracellular fatty acid accumulation. Another strategy would be to bind the internal free fatty acids to fatty acid binding proteins. Fatty acid binding proteins have been characterized in *Y. lipolytica* [346] and binding free fatty acids to these proteins are required for activation of fatty acids to fatty acyl-CoA [169]. Currently, there are no data to ascertain whether binding free fatty acids to fatty acid binding proteins would prevent fatty acids from being secreted to the extracellular matrix. Therefore, this strategy needs to be experimentally tested.

Module 5: *Investigating fatty alcohol production in the cytosol.*

Our immediate goal is to produce fatty alcohols inside the peroxisome. We accomplished this knocking out fatty acid activation to fatty acyl-CoA in the cytosol (Δ*Yl*FAA1p) while maintaining this function in the peroxisome (*Yl*AAL1p). However, we have yet not tested cytosolic fatty alcohol production capabilities of the engineered strain by expressing MAACR in the cytosol of *Y. lipolytica*. This would establish if fatty alcohol production predominantly occurs in the peroxisome for the current engineered strain. From an efficiency standpoint, this is important because we want to leverage peroxisomes for chain shortening and therefore, if significant fatty alcohols are produced in the cytosol, we would need to revisit our strategy and re-engineer the strain. Furthermore, this experiment would inform us whether there is a need to delete fatty alcohol oxidase deletion, Δ*YlFAO1*, which is cytosolic. FAO1p is responsible for the degradation of fatty alcohols.

Final Recommendation: *Further characterization of the fatty acid hybrid promoter*

In Chapter 3, we demonstrate the development of a hybrid promoter that is inducible by extracellular fatty acids while in Chapter 5, we utilize this hybrid promoter to profile real-time changes in intracellular fatty acid. What has not been established to date is whether the response from the hybrid promoter is directly related to fatty acids or products that are a result of fatty acid being processed. Activation of fatty acids to fatty acyl-CoA is the first step of β-oxidation. In chapter 5, we show that by overexpressing the AAL1 enzyme that catalyzes this reaction in the peroxisome, we can observe a reduction in fluorescence from the hybrid promoter (Figure 5.5B). This experiment still does not tease out if the hybrid promoter is responding to concentrations of fatty acids, fatty acyl-CoA or both.

Therefore, I recommend two experiments to characterize the responsiveness of the hybrid promoter. First, we need to knock out the fatty acid activation step inside the peroxisome (Δ*AAL1*) in the fatty acid producing strain, PO1fΔ*mfe1*Δ*faa*1. AAL1 is the predominant fatty acyl-CoA synthetase, which should severely limit fatty acyl-CoA production. As a result, the fluorescence signal from the hybrid promoter should increase if the promoter is responding to fatty acids. Liquid chromatography coupled to mass spectrometry (LC-MS) can then be used to determine fatty acid and fatty acyl-CoA inside the peroxisome and total cellular pools. The same analysis should be used on the current engineered strain, PO1fΔ*mfe1*Δ*faa*1 to determine how these metabolite pools change in comparison to PO1fΔ*mfe1*Δ*faa*1Δ*aal*1.

APPENDICES

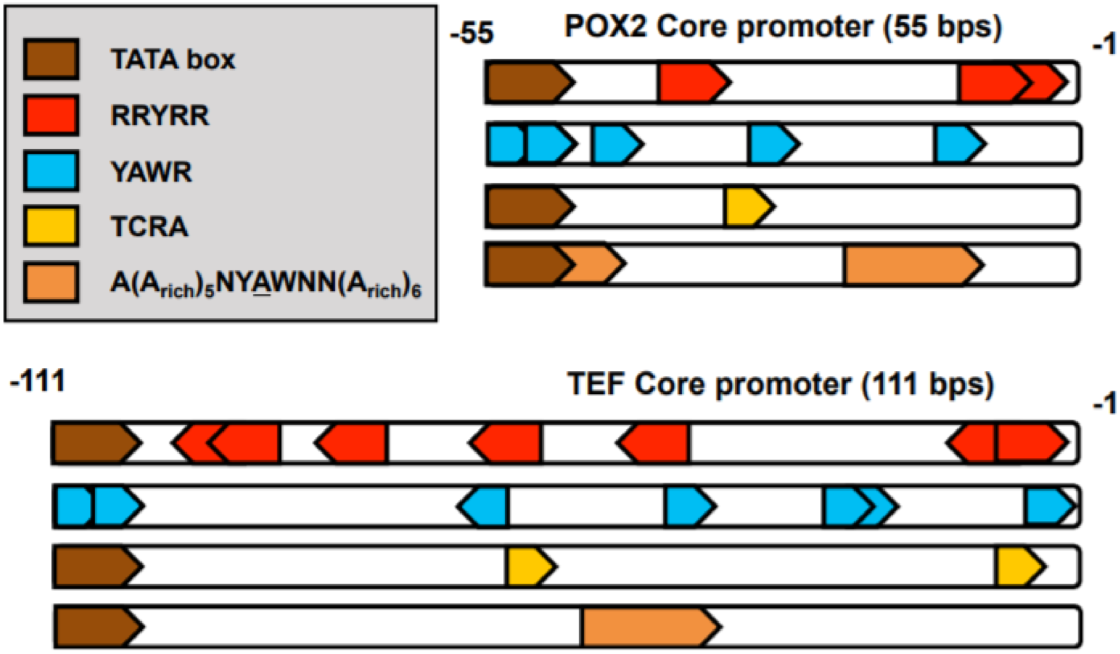
Appendix A

Supplementary Tables for Chapter 2

Supplementary Table 2.1 gBlock® and primer pairs used to construct base hybrid promoter containing restriction sites for cloning promoters.

Name	Sequence
gBlock®	GATCCCCCGGGTTCGAAGCTAGCCCTAGGGGCGCGCCATGGTG AGCAAGCAGATCCTGAAGAACACCGGCCTGCAGGAGATCATGA GCTTCAAGGTGAACCTGGAGGGCGTGGTGAACAACCACGTGTT CACCATGGAGGGCTGCGGCAAGGGCAACATCCTGTTCGGCAAC CAGCTGGTGCAGATCCGCGTGACCAAGGGCGCCCCCCTGCCCT TCGCCTTCGACATCCTGAGCCCCGCCTTCCAGTACGGCAACCGC ACCTTCACCAAGTACCCCGAGGACATCAGCGACTTCTTCATCCA GAGCTTCCCCGCCGGCTTCGTGTACGAGCGCACCCTGCGCTACG AGGACGGCGGCCTGGTGGAGATCCGCAGCGACATCAACCTGAT CGAGGAGATGTTCGTGTACCGCGTGGAGTACAAGGGCCGCAAC TTCCCCAACGACGGCCCCGTGATGAAGAAGACCATCACCGGCC TGCAGCCCAGCTTCGAGGTGGTGTACATGAACGACGGCGTGCT GGTGGGCCAGGTGATCCTGGTGTACCGCCTGAACAGCGGCAAG TTCTACAGCTGCCACATGCGCACCCTGATGAAGAGCAAGGGCG TGGTGAAGGACTTCCCCGAGTACCACTTCATCCAGCACCGCCTG GAGAAGACCTACGTGGAGGACGGCGGCTTCGTGGAGCAGCAC GAGACCGCCATCGCCCAGCTGACCAGCCTGGGCAAGCCCCTGG GCAGCCTGCACGAGTGGGTGTAAGCTAGCCTCATGTAATTAGT TATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAA CCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTAT TTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTATATT TCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAA CATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGA AGGCTTTAATTTGCA
F1	ACAATTTCGAATGCGGTACCCGAATTCCT
R1	AAGCTTCTGCAGGCATGC

Supplementary Figures for Chapter 2



Supplementary Figure 2.1. Identification of predicted transcription initiation sites in the TEF and POX2 core promoter region.

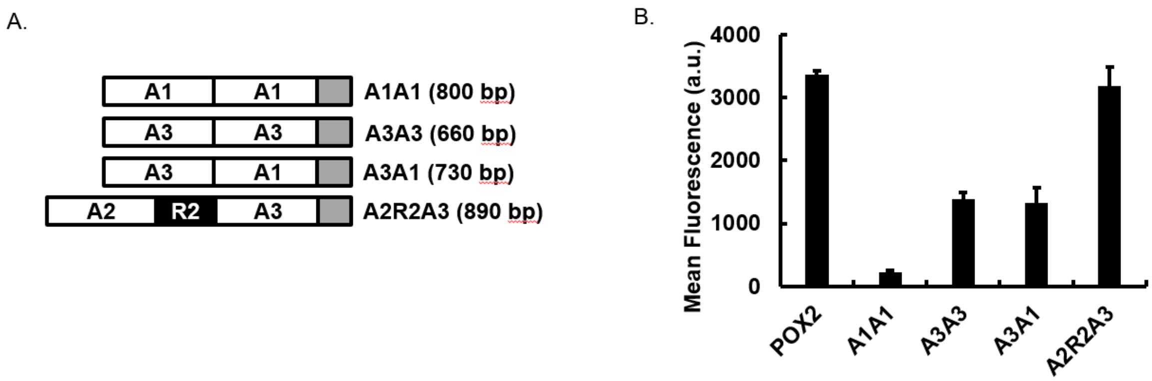
Appendix B

Supplementary Tables for Chapter 3

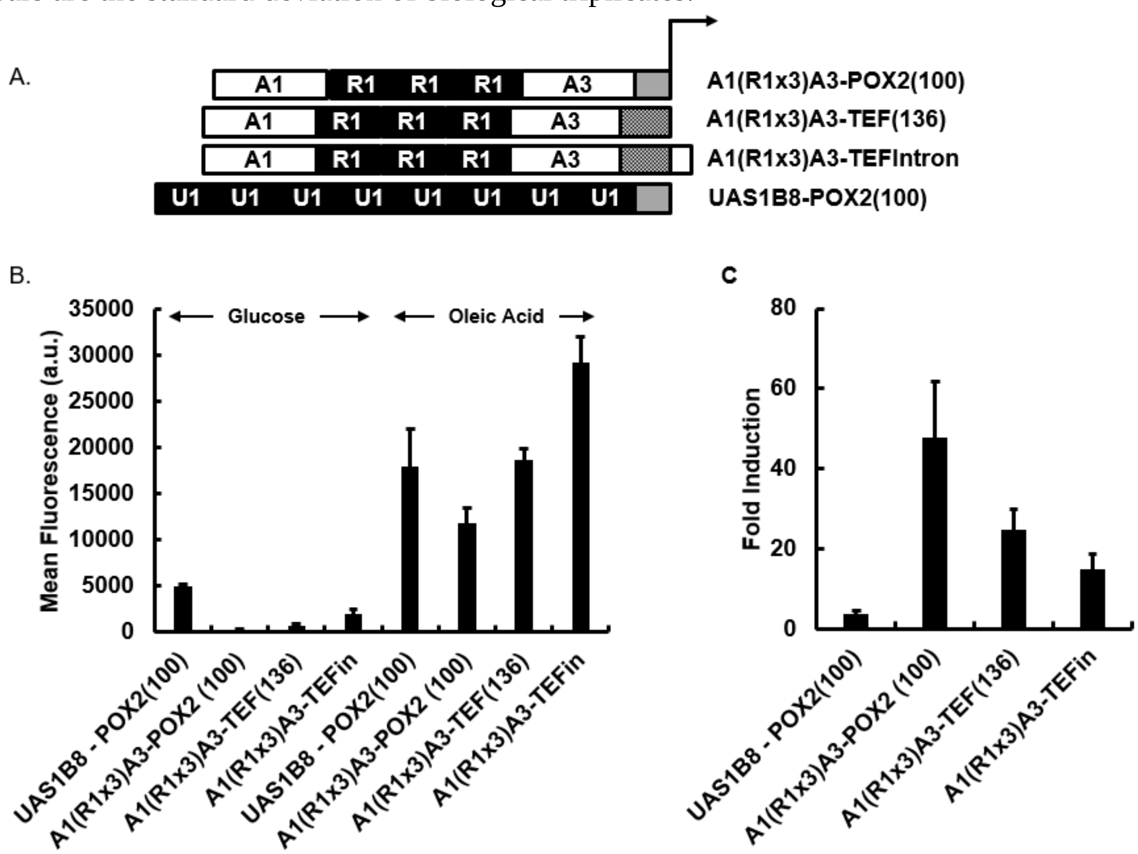
Supplementary Table 3.1. Detailed list of vectors and primers used to construct hybrid promoters tested in this study.

F1	CCGGGTTCGAAGGTACCAAGGAAGCATGCGATATTCCGGTCCCG AAACCCGAT
F2	CCGGGTTCGAAGGTACCAAGGAAGCATGCTTCTCCCCCCTTTCA CACTCTG
F3	CCGGGTTCGAAGGTACCAAGGAAGCATGCCCGTCTCCTCTATAT GTGTATCCG
F4	CCGGGTTCGAAGGTACCAAGGAAGCATGCAAGTGAGACTGGCG ATCGG
F5	CCGGGTTCGAAGGTACCAAGGAAGCATGCGAGAAGCGATCGCC CGTC
F6	CCGGGTTCGAAGGTACCAAGGAAGCATGCGGTACCAGCGGGAG GTTAC
F7	CCGGGTTCGAAGGTACCAAGGAAGCATGCGGGATACCGGAATA ACCCTGGCT
F8	CCGGGTTCGAAGGTACCAAGGAAGCATGCATGTTTGTTTTTCCG ATCTTTCGG
F9	CCGGGTTCGAAGGTACCAAGGAAGCATGCCATGAAAACTATAA CCTAGACTACACG
F10	CTAAATTTGATGAAAGGGGGATCCCCCGGGTTCCTAGGGATATT CCGGTCCCGAAACCC
F11	TTGACGTGGTGAATGTCGCCCGTTCTCACGTGACAAGTGAGACT GGCGATC
F12	CCCTAAATTTGATGAAAGGGGGATCCCCCGGGTTCCTAGGCAAG TGAGACTGGCGATC
F13	AAGTATATTGAATGTGAACGTGTACAATATCACAATTGGACATG TTTGTTTTTCCGA
F14	GAATGTCGCCCGTTCTCACGTGAGCATGCAATTGGACATGTTTG TTTTTCCGATCTTT
F15	AAAGTATATTGAATGTGAACGTGTACAATATCACCCCGTCTCCT CTATATGTGTATCCG
F16	GTGGTGAATGTCGCCCGTTCTCACGTGAGCATGCCCCGTCTCCT CTATATGTGTATCCG
F17	GCACAAGGGGTAGGCGAATGGTACGATTCCGCCCCGTCTCCTCT ATATGTGTATCC
F18	AAGGGGTAGGCGAATGGTACGATTCCGCAATTGGACATGTTTGT TTTTCCGA
F19	TTCTCCCTCGGCTCTCGGTATTTCAGCATGCTTGTGGTTGGGACT TTAGCC

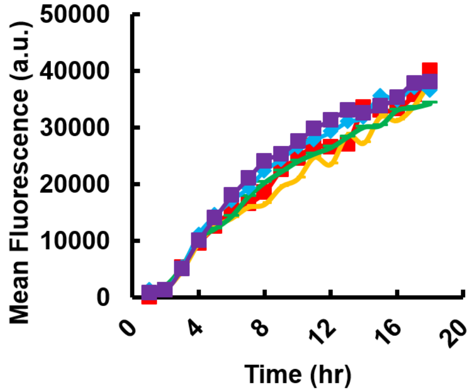
F20	GACCAGCACTTTTTGCAGTACTAACCGCAGGTGAGCAAGCAGAT CCTGAAGAACACC
F21	GATGAAAGGGGGATCCCCCGGGTTCCTAGGCCCGTCTCCTCTAT ATGTGTATCC
F22	TTGATGAAAGGGGGATCCCCCGGGTTCCTAGGGATATTCCGGTC CCGAAACCCCAGAGT
R1	CTTCAGGATCTGCTTGCTCACCATGGCGCGCCGGCGTCGTTGCTT GTGTGAT
R2	CACCAACCCTTCTCCCGATCGCCAGTCTCACTTGTCACGTGAGA ACGGGCGAC
R3	CGTGTAGTCTAGGTTATAGTTTTCATGGCATGCTGATATTGTACA CGTTCACATTCAAT
R4	CGAAAGATCGGAAAAACAAACATGTCCAATTGTGATATTGTAC ACGTTCACATTCA
R5	GTGTAGTCTAGGTTATAGTTTTCATGGCATGCTGAAATACCGAG AGCCGAGG
R6	AAGATCGGAAAAACAAACATGTCCAATTGCATGCTCACGTGAG AACGGGCGAC
R7	GCCCGAAAGATCGGAAAAACAAACATGTCCAATTGCGGAATCG TACCATTCGC
R8	CCCTTGGCTAAAGTCCCAACCACAAGCATGCTGAAATACCGAGA GCCGAGG
R9	GCGTGACATAACTAATTACATGAGGCTAGCTTACACCCACTCGT GCAGG
R10	ATGTAAGCGTGACATAACTAATTACATGAGTTACACCCACTCGT GCAGGCTGCC
R11	GTAGTCTAGGTTATAGTTTTCATGGCATGCGCGGAATCGTACCA TTCGC
R12	GTAGTCTAGGTTATAGTTTTCATGGCATGCGGGCAGTGACGGAA ACGACA
R13	GGTTTCGGGACCGGAATATCCCTAGGCGGAATCGTACCATTCGC CTACCC
gBlock®	CCCGCCCACCTCGATCCGGGCATGCCTGCAGAAGCTTTTGTGGT TGGGACTTTAGCCAAGGGTATAAAAGACCACCGTCCCCGAATTA CCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACCCGAAA TCGTTAAGCATTTCCTTCTGAGTATAAGAATCATTCAAAATGGT GAGTTTCAGAGGCAGCAGCAATTGCCACGGGCTTTGAGCACAC GGCCGGGTGTGGTCCCATTCCCATCGACACAAGACGCCACGTCA TCCGACCAGCACTTTTTGCAGTACTAACCGCAGGGCGCGCCTGC ATCGCATTGGATAGCCATTCTCCGAGTGTTTTAGCGTTAATTAA AACCACAGAGCATAAAGAGAACCTCTAGCTGGCGATGCTTTGCT AGCCTCATGTAATTAGTTATGTCACGCT



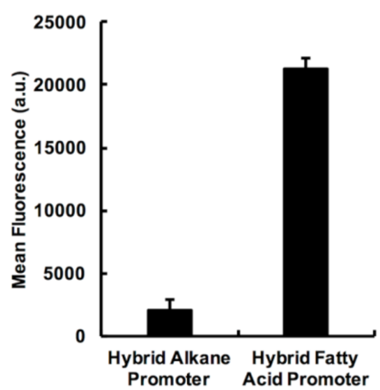
Supplementary Figure 3.1. Additional hybrid promoters tested. A. New promoters were designed by combining parts of the POX2 promoter upstream of the POX2 core promoter. Activating (A1, A2, A3) sequences and regulatory (R2) sequences were defined by 5’ truncations. B. Promoter strength is determined by expression of hrGFP and measured as mean fluorescence of an equal number of transformed cells. The data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviation of biological triplicates.



Supplementary Figure 3.2. Engineering stronger fatty acid-responsive promoters based on the A1(R1x3)A3 hybrid promoter. A. Schematic of different promoters based on the A1(R1x3)A3 promoter. B. Promoter strength is determined by expression of hrGFP and measured as mean fluorescence of an equal number of transformed cells**.** Fluorescence from hybrid promoters containing TEF core and intron sequence improves oleic acid expression relative to the POX2 native promoter by 5 fold and 10 fold, respectively. Glucose samples are shown on the left and oleic acid on the right. C. Promoter induction using oleic acid as the carbon source relative to glucose as carbon source. In (B) and (C) the data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviation of biological triplicates.



Supplementary Figure 3.3. Sensitivity and inducibility of the A1(R1x3)A3 hybrid promoter. Titration of the A1(R1x3)A3 hybrid promoter with 0.25% - 8.0% oleic acid resulted in identical activation. The data are the average of mean fluorescence measurements from biological triplicates. The error bars are the standard deviation of biological triplicates. Red is 0.25%, Orange is 0.5%, Green is 1.0%, Cyan is 4.0%, Purple is 8.0%.



Supplementary Figure 3.4. Promoter strength is determined by expression of hrGFP and measured as mean fluorescence of an equal number of transformed cells grown in 2% n-decane or 2% oleic acid.

Appendix C

Supplementary Tables for Chapter 4

Supplementary Table 4.1: Primer sequences used to create the vectors containing mutations to the POR1p binding sites.

Oligo	Sequence	Description
F1	ATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAA CCATTATTATCATGACATT	Vector Primer Forward
F2	GATGGAGAGCGCCAGACGAGCAGAATAAATAGAC AGCGGATCGGGGGAGGGCTGT	A1 Mutation Forward
F3	AGACGAACAAGTGATAGGCCGAGAGTAAATAACGA GGTGGAGTGCACAAGGGGTAG	R1 Mutation Forward
F4	GTTAAGCTTGTAGCGAATTTCGCTAAATAACATCACC CCATACGACGGACACA	A2 Mutation Forward
F5	GTCCAAATACCCCCGTTTATTCTCTAAATACTCTCGG TATTTCACATGAAAACTATA	A3 Mutation Forward
R1	TAATGTCATGATAATAATGGTTTCTTAGACGTCAGGT GGCACTTTTCGGGGAAAT	Vector Primer Reverse
R2	ACAGCCCTCCCCCGATCCGCTGTCTATTTATTCTGCTC GTCTGGCGCTCTCCATC	A1 Mutation Reverse
R3	CTACCCCTTGTGCACTCCACCTCGTTATTTACTCTCG GCCTATCACTTGTTCGTCT	R1 Mutation Reverse
R4	GTCCGTCGTATGGGGTGATGTTATTTAGCGAAATTC GCTACAAGCTT	A2 Mutation Reverse
R5	TTTCATGTGAAATACCGAGAGTATTTAGAGAATAA ACGGGGGTATTTG	A3 Mutation Reverse

Supplementary Table 4.2. gRNA oligos used to create CRISPR based vectors and verification primers used in colony PCR.

gRNA	Sequence	Purpose
gRNA_POR1_F	AATTCCGGGTCGGCGCAGGTTGACGTACT CACAAGCCCTGGAAGCTCGGTTTTAGAGC TAGAAATAGCAAGTTA	Forward Oligo
gRNA_POR1_R	TAACTTGCTATTTCTAGCTCTAAAACCGAG CTTCCAGGGCTTGTGAGTACGTCAACCTGC GCCGACCCGGAAT	Reverse Oligo
gRNA_CFU1_F	AATTCCGGGTCGGCGCAGGTTGACGTGAG GTCACAGAGACCCCCGAGGTTTTAGAGCT AGAAATAGCAAGTTAA	Forward Oligo
gRNA_CFU1_R	TTAACTTGCTATTTCTAGCTCTAAAACCTC GGGGGTCTCTGTGACCTCACGTCAACCTG CGCCGACCCCGGAAT	Reverse Oligo

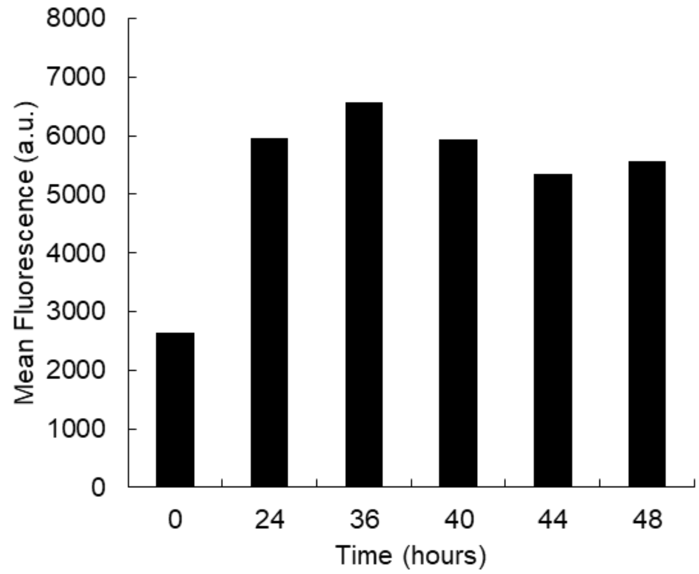
POR1_Ver_F	ATGTCTTCCAAGGTCAAAGAGGAG	Colony PCR
POR1_Ver_R	CCTGCTGAAAGTGCATGAGC	
CFU1_Ver_F	ATGCTTCCCGAGCTGGAAA	
CFU1_Ver_R	GCAAATGGGGCACTGGAATG	

Supplementary Table 4.3. Primer pairs used to tile the POX2 native promoter.

	Forward Primer	Reverse Primer
1	CACGGTGGGACGTGTCTG	CACGTGGCCCAAAAGCTC
2	CACGGTGGGACGTGTCTG	GTGTGAAGCCGGGAGGTC
3	GACCTCCCGGCTTCACAC	AAATGTGGGGGCAGATTCA
4	GTGACCTCCCGGCTTCAC	TGTGGGGGCAGATTCAGA
5	CGGCTTCACACGTGGTTG	GAGATAAAATGTGGGGGCAGA
6	TCTGAATCTGCCCCCACA	CGGTGCGGGGTTTATGTA
7	CCCCCTTTCACACTCTGCT	ACACCTAACGGCGGCTTC
8	CCGCACCGTTTGGAACTC	GGCGCTCTCCATCTGACA
9	CCGCCGTTAGGTGTGTCA	CGTCAACAGTGCCCTTCG
10	GCAGAACCGAGGGACAGC	TGAGAACGGGCGACATTC
11	CGAAGGGCACTGTTGACG	CCAAACAAAGAGGCGGATACA
12	GAATGTCGCCCGTTCTCA	CCAAACAAAGAGGCGGATACA
13	TCCGCCTCTTTGTTTGGTT	CAGGACGATGCAGATGTCTACTTT
14	TCCGCCTCTTTGTTTGGTT	TGGAATGCAGGACGATGC
15	CCACCCCAATCACATGCT	CCTCGGCTCTCGGCCTAT
16	CCTGCATTCCATCCCACA	CCATTCGCCTACCCCTTG
17	AGAGCCGAGGACGAGGTG	CCCGATCGCCAGTCTCAC
18	CAAGGGGTAGGCGAATGG	ATCCCCCATGACCACCAA
19	CGATCGGGAGAAGGGTTG	TCCGCTACTCGTAGTGGTTTTT
20	TGGTCATGGGGGATAGAATTT	CTTGCACTCCCACCATTGC
21	ACCACTACGAGTAGCGGATTTG	CTTGCACTCCCACCATTGC
22	GCAATGGTGGGAGTGCAA	CCGTATTGCCCCGTTTCT
23	AGAAACGGGGCAATACGG	TCGGACTTGTGGCGATTG
24	CGTCTGTTCAATCGCCACA	TGGCGCTTGTCCAGTATGA
25	CCTGTCAATCATGGCACCAC	CTCTGGCGCTTGTCCAGT
26	ACTGGACAAGCGCCAGAG	CCGTCGTATGGGGTGATG
27	GAATTTCGCCCTCGGACA	CAGCTACAATAAGAGAGGCTGTTTG
28	CACCCCATACGACGGACA	CCCGCTGGTACCTGATATTG
29	CAGGTACCAGCGGGAGGT	CCATCTCCAAGCCAGGGTTA
30	CAGGTACCAGCGGGAGGT	TGGACCGACCATCTCCAA
31	CCCTGGCTTGGAGATGGT	TCCAATTGGGGCAGTGAC
32	CCCTGGCTTGGAGATGGT	GCCCGAAAGATCGGAAAA

33	TCCGTGTCGTTTCCGTCA	AGCAACAGTCCAGGAGACAGA
34	GCGCCCTCTCCTTGTCTC	GAACCAATGGAGGCCAAAG
35	GCGCCCTCTCCTTGTCTC	GACGGGGAGGAACCAATG
36	TGGACTGTTGCTACCCCATT	AGGGAAACCATGCAACCAT
37	CATTGGTTCCTCCCCGTCT	AACATGTGACTGTGGGGAAAA
38	CCCGTCTTTCACGTCGTC	CCCCTTGCACGTCAAAATTA
39	GGGGTCTAGATGGAGGCCTAA	GCCTTGCACCATGTCCAC
40	ATTGGGGCGAGAAACACG	CGGAAAAGCGTCGAATCA
41	CGTGGACATGGTGCAAGG	CGAGAGCCGAGGGAGAAT
42	GGTTGATTCGACGCTTTTCC	CGAGAGCCGAGGGAGAATA
43	TATTCTCCCTCGGCTCTCG	GGTTGCCCGTGTAGTCTAGGT
44	CCCTCGGCTCTCGGTATT	GCGTCGTTGCTTGTGTGA

Supplementary Figures for Chapter 4



Supplementary Figure 4.1. Flow Cytometry data for expression of native POX promoter fused to GFP.

POR1 gRNA Target CGAGCTTCCAGGGCTTGTGAG
POf1Δ*por1* **CGA-CTTCCAGGGCTTGTGAG**

CFU1 gRNA Target CTCGGGGGTCTCTGTGACCTC
POf1Δ*cfu1* **CTCGGG--TCTCTGTGACCTC**

CFU1 gRNA Target CTCGGGGGTCTCTGTGACCTC
POf1Δ*por1*Δ*cfu1* **CTC--------TCTGTGACCTC**
***Mutation made to *CFU1* in PO1fΔ*por1* strain**

Supplementary Figure 4.2. Frameshift mutations to create POf1*Δpor1*, POf1*Δcfu1*, and POf1*Δpor1Δcfu1* strains.

A.

Description	Max score	Total score	Query cover	E value
YALI0D18678p [Yarrowia lipolytica CLIB122]	2388	2388	100%	0.0
YALIA101S06e08922g1_1 [Yarrowia lipolytica]	2351	2351	100%	0.0
hypothetical protein YALI1_D23479g [Yarrowia lipolytica]	2346	2346	100%	0.0
Tda9p [Sugiyamaella lignohabitans]	556	751	61%	5e-176
putative dna binding regulatory protein [Diplodia seriata]	186	274	51%	1e-45
hypothetical protein NADFUDRAFT_84492 [Nadsonia fulvescens var. elongata DSM 6958]	556	756	80%	4e-173
CYFA0S02e10374g1_1 [Cyberlindnera fabianii]	186	321	54%	3e-45
putative transcription factor TDA9 [Cyberlindnera fabianii]	188	323	56%	5e-46
hypothetical protein AJ80_02223 [Polytolypa hystricis UAMH7299]	188	318	40%	5e-46
related to DNA-binding protein [Phialocephala subalpina]	189	316	39%	3e-46
hypothetical protein LY89DRAFT_392739 [Phialocephala scopiformis]	194	322	53%	7e-48
DNA binding regulatory protein AmdX [Histoplasma capsulatum G186AR]	192	320	39%	3e-47
zinc finger protein [Paracoccidioides brasiliensis Pb18]	197	324	39%	9e-49

B.

S. cerevisiae ADR1 DNA binding domain	VCLTCTRSFARLEHLKRHERSHTKERPFQCPICERCFARRDLLLRHRQKLHA
Y. lipolytica CFU1p	VCEVCTRAFARQEHLKRHYRSHTNEKPYPCGLCNRAFTRRDLLIRHAQKIHS

Supplementary Figure 4.3. (A) BLAST of *Yl*CFU1p (YALI0D18678) against fungal database did not reveal *Sc*ADR1p as strong homolog. (B) 65% sequence similarity between *Sc*ADR1p and *Yl*CFU1p DNA zinc finger binding domain.

Appendix D

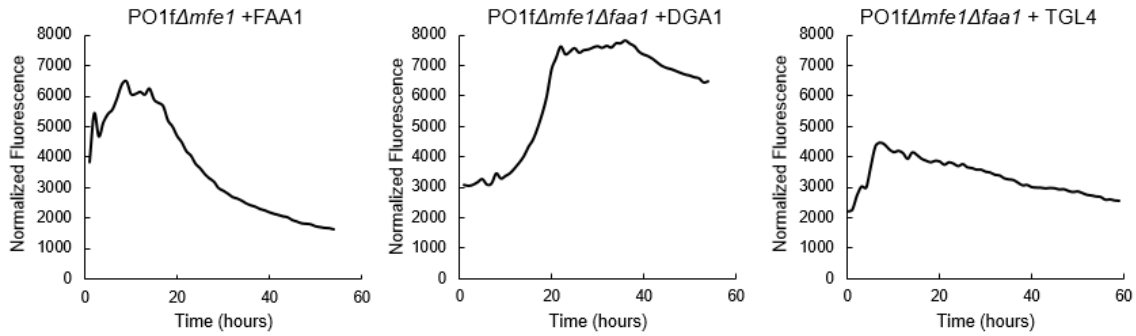
Supplementary Tables for Chapter 5

Supplementary Table 5.1: gRNA oligos for CRISPR vectors and primers for PCR amplification of genes tested in the study.

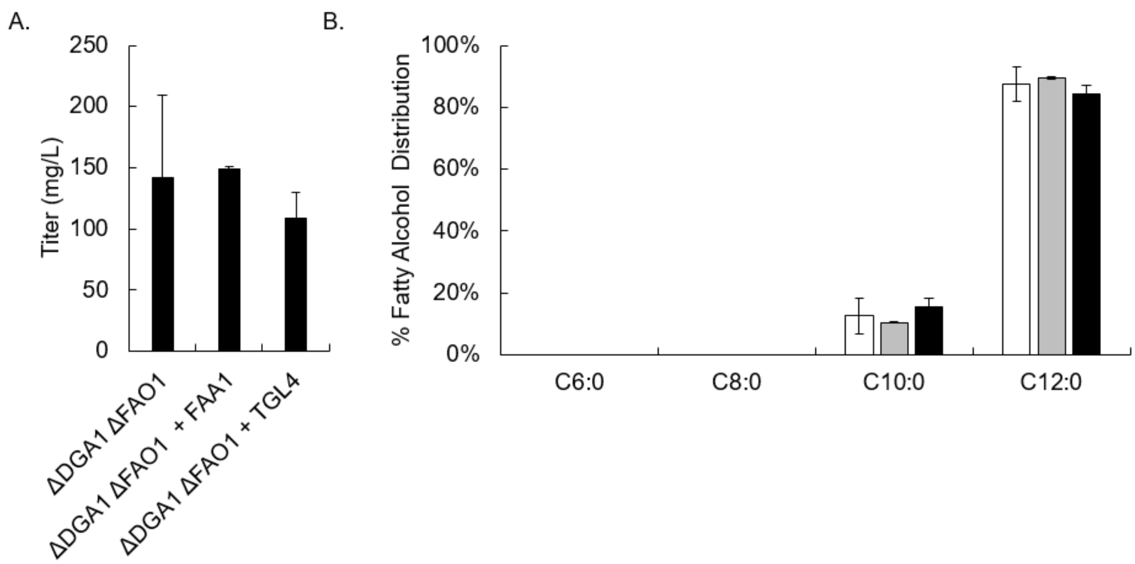
Oligo	Sequence	Description
F1	GAGGCCCAGATCCTCTAGAGTCGAAG CGGCCGCAGACCGGGTTGGCGGCGTAT	TEF(404)-Intron
F2	GCACTTTTTGCAGTACTAACCGCAGCC CCAAATCATTCACAAATCTGC	AAL-PTS
F3	TATAAGAATCATTCAAAGGCGCGCAT ATGGCCATTCAGCAGGTCCATC	MAACR-PTS
F4	GGTCGGCGCAGGTTGACGTAGTTGTTC GTTCCACCTCCAAGTTTTAGAGCTAGA AATAGC	Forward Oligo
F5	GGTCGGCGCAGGTTGACGTATACTACC CTCTGGACGTCCGTTTTAGAGCTAGAA ATAGCA	Forward Oligo
F6		Forward Oligo
F7	GGTCGGCGCAGGTTGACGTAAGATATG AAGATCTACACCAGTTTTAGAGCTAGA AATAGC	Forward Oligo
F8	TTCCGGGTCGGCGCAGGTTGACGTAGATA GGATATCTGCAACCCG	Forward Oligo
F9	TTCCGGGTCGGCGCAGGTTGACGTATGCT GCATAGCAGTGCACAG	Forward Oligo
F10	ATTTGATGAAAGGGGGATCCCCCGG CCCAGTTGCAAAAGTTGACA	Dual gRNA vector
R1	CTGCGGTTAGTACTGCAAAAAGTGC TGGTCG	TEF(404)-Intron
R2	GACATAACTAATTACATGAGGCTAG CTTACAACTTACTCACATCAATGCC	AAL-PTS
R3	ACATAACTAATTACATGAGGCTAGTTAATT AATTAGAGCTTAGCGGCAGCCTTTTTTC	MAACR-PTS
R4	TGCTATTTCTAGCTCTAAAACTTGGAGGTG GAACGAACAACTACGTCAACCTGCGCCGAC	Reverse Oligo
R5	CCTCGGGCTCCGAACTTAACGTTTTAGAGC TAGAAATAGCAAGTTA	Reverse Oligo
R6	AACTTGCTATTTCTAGCTCTAAAACCGGG TTGCAGATATCCTATC	Reverse Oligo
R7	TGCTATTTCTAGCTCTAAAACTGGTGTAGA TCTTCATATCTTACGTCAACCTGCGCCGAC	Reverse Oligo

R8	AACTTGCTATTTCTAGCTCTAAAACCGGG TTGCAGATATCCTATC	Reverse Oligo
R9	AACTTGCTATTTCTAGCTCTAAAACCTG TGCACTGCTATGCAGCA	Reverse Oligo
R10	GCTTCCTTGGTACCTTCGAACCCGG AAAAGCACCGACTCGG	Dual gRNA vector

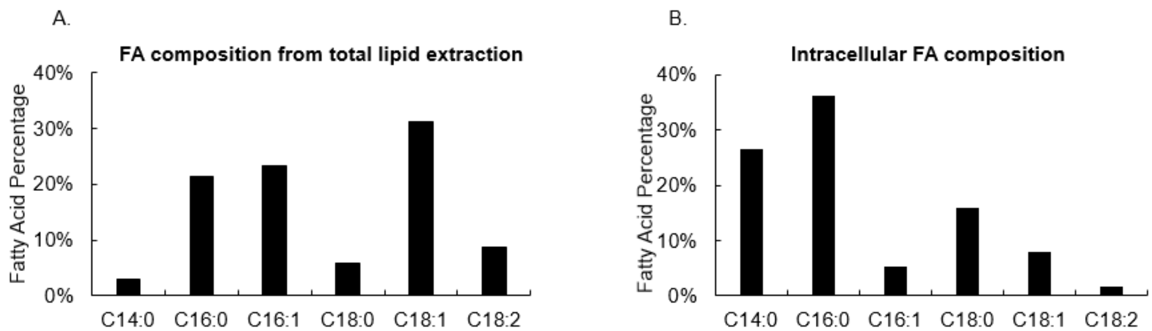
Supplementary Figures for Chapter 5



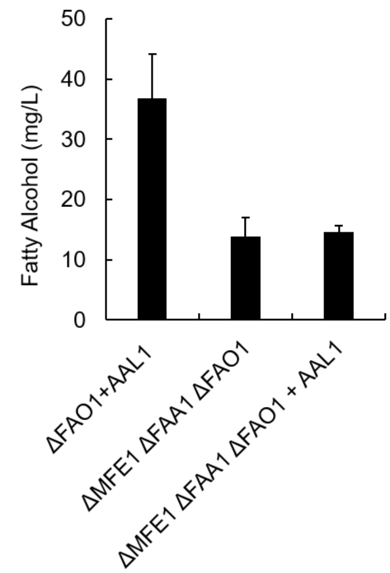
Supplementary Figure 5.1. GFP expressing strains from Table 5.3.



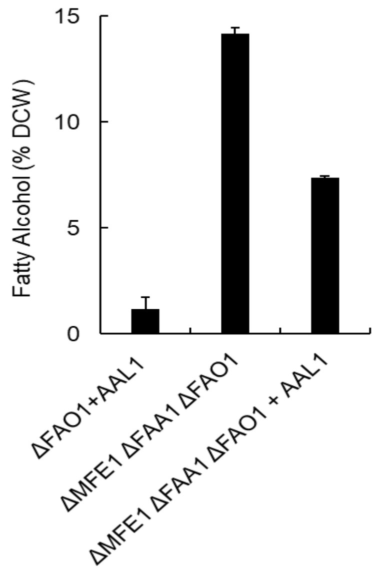
Supplementary Figure 5.2. Fatty alcohol production in the PO1f *ΔDGA1ΔFAO1 ΔDGA1ΔFAO1 + FAA1* and *ΔDGA1ΔFAO1 + TGL4* strains due to oxidation of dodecane resulting in dodecanol (C12:0). Some decanol (C10:0) is observed due to intact β-oxidation pathway.



Supplementary Figure 5.3. (A) Fatty acid distribution from whole cell lipid analysis of PO1f*Δfaa1Δmfe1* strain. (B) Fatty acid distribution from TLC extracted FFA component of PO1f*Δfaa1Δmfe1* strain showing higher proportions of C14:0 FFAs.



Supplementary Figure 5.4. Fatty alcohol production in the PO1f*Δfao1*, PO1f*Δfaa1Δmfe1 Δfao1* and PO1f*Δfaa1Δmfe1Δfao1 + AAL1* due to oxidation of dodecane (C12:0) to produce dodecanol.



Supplementary Figure 5.5. Fatty alcohol production as a percentage of dry cell weight (DCW) in the PO1f*Δfao1*, PO1f*Δfaa1Δmfe1 Δfao1* and PO1f*Δfaa1Δmfe1Δfao1 + AAL1*.

REFERENCES

[1] Nakamura, C. E., Whited, G. M., Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* 2003, *14*, 454-459.
[2] Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., *et al.*, Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol* 2011, *7*, 445-452.
[3] Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., *et al.*, High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature* 2013, *496*, 528-532.
[4] Hermann, B. G., Blok, K., Patel, M. K., Producing bio-based bulk chemicals using industrial biotechnology saves energy and combats climate change. *Environ Sci Technol* 2007, *41*, 7915-7921.
[5] Hackett, M., *Safe and Sustainable Chemicals Report*, IHS Chemical, Menlo Park, CA 2011.
[6] *Biological Energy Production*, Agriculture and Consumer Protection, FAO Agricultural Services Bulletins 1997.
[7] Freiesleben, S. A. a. J., A. K., Correlation between Plant Secondary Metabolites and Their Antifungal Mechanisms–A Review. *Medicinal & Aromatic Plants* 2014, *3*, 154.
[8] Roberts, S. C., Production and engineering of terpenoids in plant cell culture. *Nat Chem Biol* 2007, *3*, 387-395.
[9] Su, P., Cheng, Q., Wang, X., Cheng, X., *et al.*, Characterization of eight terpenoids from tissue cultures of the Chinese herbal plant, Tripterygium wilfordii, by high-performance liquid chromatography coupled with electrospray ionization tandem mass spectrometry. *Biomed Chromatogr* 2014, *28*, 1183-1192.
[10] Jaakola, L., Hohtola, A., Effect of latitude on flavonoid biosynthesis in plants. *Plant Cell Environ* 2010, *33*, 1239-1247.
[11] Carter, J., *Western Watersheds Project*, Watersheds Messenger 2002.
[12] Frewer, L. J. S., R, Ethical concerns and risk perceptions associated with different applications of genetic engineering: Interrelationships with the perceived need for regulation of the technology. *Agriculture and Human Values* 1995, *12*, 48-57.
[13] Robinson, J., Ethics and transgenic crops: a review. *Electronic Journal of Biotechnology* 1999, *2*.
[14] Menetrez, M. Y., An overview of algae biofuel production and potential environmental impact. *Environ Sci Technol* 2012, *46*, 7073-7085.
[15] Pulz, O., Gross, W., Valuable products from biotechnology of microalgae. *Appl Microbiol Biotechnol* 2004, *65*, 635-648.
[16] McHugh, D. J., School of Chemistry, University College; University of New South Wales; Australian Defence Force Academy FAO Fisheries Technical Paper 2003.
[17] Bixler, H. J. P., H., A Decade of Change in the Seaweed Hydrocolloids Industry. *Journal of Applied Phycology* 2010, *2017*.
[18] Spolaore, P., Joannis-Cassan, C., Duran, E., Isambert, A., Commercial applications of microalgae. *J Biosci Bioeng* 2006, *101*, 87-96.
[19] Brennan, L., Owende, P., Biofuels from Microalgae - A Review of Technologies for Production, Processing, and Extraction of Biofuels and Co-Products. *renewable and sustainable energy review* 2010, *14*, 557.

[20] Ma, R. Y. N., Chen, F., Enhanced production of free trans-astaxanthin by oxidative stress in the cultures of the green microalga Chlorococcum sp. *Process Biochem* 2001, *36*, 1175-1179.
[21] Roman, R. B., Alvarez-Pez, J. M., Fernandez, F. G. A., Grima, E. M., Recovery of pure B-phycoerythrin from the microalga Porphyridium cruentum. *J Biotechnol* 2002, *93*, 73-85.
[22] Rasala, B. A., Chao, S. S., Pier, M., Barrera, D. J., Mayfield, S. P., Enhanced genetic tools for engineering multigene traits into green algae. *PLoS One* 2014, *9*, e94028.
[23] Zhao, T., Wang, W., Bai, X., Qi, Y., Gene silencing by artificial microRNAs in Chlamydomonas. *Plant J* 2009, *58*, 157-164.
[24] Shin, S. E., Lim, J. M., Koh, H. G., Kim, E. K., *et al.*, CRISPR/Cas9-induced knockout and knock-in mutations in Chlamydomonas reinhardtii. *Sci Rep* 2016, *6*, 27810.
[25] Katsuyama, Y., Funa, N., Miyahisa, I., Horinouchi, S., Synthesis of unnatural flavonoids and stilbenes by exploiting the plant biosynthetic pathway in Escherichia coli. *Chem Biol* 2007, *14*, 613-621.
[26] Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., *et al.*, Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* 2010, *330*, 70-74.
[27] Bartley, G. E., Scolnik, P. A., Beyer, P., Two Arabidopsis thaliana carotene desaturases, phytoene desaturase and zeta-carotene desaturase, expressed in Escherichia coli, catalyze a poly-cis pathway to yield pro-lycopene. *Eur J Biochem* 1999, *259*, 396-403.
[28] Meyer, A., Kirsch, H., Domergue, F., Abbadi, A., *et al.*, Novel fatty acid elongases and their use for the reconstitution of docosahexaenoic acid biosynthesis. *J Lipid Res* 2004, *45*, 1899-1909.
[29] Minami, H., Kim, J. S., Ikezawa, N., Takemura, T., *et al.*, Microbial production of plant benzylisoquinoline alkaloids. *Proc Natl Acad Sci U S A* 2008, *105*, 7393-7398.
[30] Ehsandoost, E., Gholami, S. P., Nazemi, M., Effect of denak (Oliveria decumbens Vent) on growth and survival of Lactobacillus acidophilus and Bifidobacterium bifidum for production of probiotic herbal milk and yoghurt. *Pak J Biol Sci* 2013, *16*, 2009-2014.
[31] Moyer, A. J., Coghill, R. D., Penicillin; the laboratory scale production of penicillin in submerged cultures by Penicillium notatum Westling (NRRL 832). *J Bacteriol* 1946, *51*, 79-93.
[32] De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K., Vandamme, E. J., Construction and model-based analysis of a promoter library for E-coli: an indispensable tool for metabolic engineering. *Bmc Biotechnol* 2007, *7*.
[33] Jensen, P. R., Hammer, K., Artificial promoters for metabolic optimization. *Biotechnology and Bioengineering* 1998, *58*, 191-195.
[34] Jeschek, M., Gerngross, D., Panke, S., Combinatorial pathway optimization for streamlined metabolic engineering. *Curr Opin Biotech* 2017, *47*, 142-151.
[35] Tan, S. Z., Prather, K. L., Dynamic pathway regulation: recent advances and methods of construction. *Curr Opin Chem Biol* 2017, *41*, 28-35.

[36] Vickers, C. E., Williams, T. C., Peng, B., Cherry, J., Recent advances in synthetic biology for engineering isoprenoid production in yeast. *Curr Opin Chem Biol* 2017, *40*, 47-56.
[37] Rao, A., Sathiavelu, A., Mythili, S., Genetic Engineering In BioButanol Production And Tolerance. *Braz. Arch. Biol. Technol.* 2016, *59*.
[38] Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., *et al.*, Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Appl Environ Microbiol* 2011, *77*, 2905-2915.
[39] Jullesson, D., David, F., Pfleger, B., Nielsen, J., Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnol Adv* 2015, *33*, 1395-1402.
[40] Nielsen, J., Keasling, J. D., Engineering Cellular Metabolism. *Cell* 2016, *164*, 1185-1197.
[41] Hawkins, K. M., Smolke, C. D., Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. *Nat Chem Biol* 2008, *4*, 564-573.
[42] Noda, S., Kondo, A., Recent Advances in Microbial Production of Aromatic Chemicals and Derivatives. *Trends Biotechnol* 2017, *35*, 785-796.
[43] Du, J., Shao, Z., Zhao, H., Engineering microbial factories for synthesis of value-added products. *J Ind Microbiol Biotechnol* 2011, *38*, 873-890.
[44] Chang, M. C., Keasling, J. D., Production of isoprenoid pharmaceuticals by engineered microbes. *Nat Chem Biol* 2006, *2*, 674-681.
[45] Wattanachaisaereekul, S., Lantz, A. E., Nielsen, M. L., Nielsen, J., Production of the polyketide 6-MSA in yeast engineered for increased malonyl-CoA supply. *Metab Eng* 2008, *10*, 246-254.
[46] Yuan, T., Guo, Y. K., Dong, J. K., Li, T. Y., *et al.*, Construction, characterization and application of a genome-wide promoter library in Saccharomyces cerevisiae. *Front Chem Sci Eng* 2017, *11*, 107-116.
[47] Meurer, M., Chevyreva, V., Cerulus, B., Knop, M., The regulatable MAL32 promoter in Saccharomyces cerevisiae: characteristics and tools to facilitate its use. *Yeast* 2017, *34*, 39-49.
[48] Zhou, S. H., Du, G. C., Kang, Z., Li, J. H., *et al.*, The application of powerful promoters to enhance gene expression in industrial microorganisms. *World J Microb Biot* 2017, *33*.
[49] Shimada, T., Tanaka, K., Ishihama, A., The whole set of the constitutive promoters recognized by four minor sigma subunits of Escherichia coli RNA polymerase. *Plos One* 2017, *12*.
[50] Peng, B. Y., Williams, T. C., Henry, M., Nielsen, L. K., Vickers, C. E., Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb Cell Fact* 2015, *14*.
[51] Weinhandl, K., Winkler, M., Glieder, A., Camattari, A., Carbon source dependent promoters in yeasts. *Microb Cell Fact* 2014, *13*, 5.
[52] Ozcan, S., Johnston, M., Function and regulation of yeast hexose transporters. *Microbiol Mol Biol Rev* 1999, *63*, 554-569.

[53] Anderson, M. S., Lopes, J. M., Carbon source regulation of PIS1 gene expression in Saccharomyces cerevisiae involves the MCM1 gene and the two-component regulatory gene, SLN1. *J Biol Chem* 1996, *271*, 26596-26601.
[54] Gardocki, M. E., Lopes, J. M., Expression of the yeast PIS1 gene requires multiple regulatory elements including a Rox1p binding site. *J Biol Chem* 2003, *278*, 38646-38652.
[55] Traven, A., Jelicic, B., Sopta, M., Yeast Gal4: a transcriptional paradigm revisited. *EMBO Rep* 2006, *7*, 496-499.
[56] He, X. J., Fassler, J. S., Identification of novel Yap1p and Skn7p binding sites involved in the oxidative stress response of Saccharomyces cerevisiae. *Mol Microbiol* 2005, *58*, 1454-1467.
[57] Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., Tuning genetic control through promoter engineering. *P Natl Acad Sci USA* 2005, *102*, 12678-12683.
[58] Blazeck, J., Garg, R., Reed, B., Alper, H. S., Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnol Bioeng* 2012, *109*, 2884-2895.
[59] Loy, C. J., Lydall, D., Surana, U., NDD1, a high-dosage suppressor of cdc28-1N, is essential for expression of a subset of late-S-phase-specific genes in Saccharomyces cerevisiae. *Molecular and Cellular Biology* 1999, *19*, 3312-3327.
[60] 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* 1994, *25*, 185-195.
[61] West, R. W., Yocum, R. R., Ptashne, M., Saccharomyces-Cerevisiae Gal1-Gal10 Divergent Promoter Region - Location and Function of the Upstream Activating Sequence Uasg. *Molecular and Cellular Biology* 1984, *4*, 2467-2478.
[62] Lefrancois, P., Gallagher, J. E. G., Snyder, M., Global Analysis of Transcription Factor-Binding Sites in Yeast Using ChIP-Seq. *Yeast Genetics: Methods and Protocols* 2014, *1205*, 231-255.
[63] Murphy, K. F., Balazsi, G., Collins, J. J., Combinatorial promoter design for engineering noisy gene expression. *Proc Natl Acad Sci U S A* 2007, *104*, 12726-12731.
[64] Teo, W. S., Chang, M. W., Development and characterization of AND-gate dynamic controllers with a modular synthetic GAL1 core promoter in Saccharomyces cerevisiae. *Biotechnol Bioeng* 2014, *111*, 144-151.
[65] d'Espaux, L., Ghosh, A., Runguphan, W., Wehrs, M., *et al.*, Engineering high-level production of fatty alcohols by Saccharomyces cerevisiae from lignocellulosic feedstocks. *Metab Eng* 2017, *42*, 115-125.
[66] Wasylenko, T. M., Stephanopoulos, G., Metabolomic and (13)C-metabolic flux analysis of a xylose-consuming Saccharomyces cerevisiae strain expressing xylose isomerase. *Biotechnol Bioeng* 2015, *112*, 470-483.
[67] Zhou, P. P., Meng, J., Bao, J., Fermentative production of high titer citric acid from corn stover feedstock after dry dilute acid pretreatment and biodetoxification. *Bioresour Technol* 2017, *224*, 563-572.
[68] Fillet, S., Gibert, J., Suarez, B., Lara, A., *et al.*, Fatty alcohols production by oleaginous yeast. *J Ind Microbiol Biot* 2015, *42*, 1463-1472.

[69] Zafar, S., Owais, M., Ethanol production from crude whey by Kluyveromyces marxianus. *Biochem Eng J* 2006, *27*, 295-298.
[70] Yaguchi, A., Robinson, A., Mihealsick, E., Blenner, M., Metabolism of aromatics by Trichosporon oleaginosus while remaining oleaginous. *Microb Cell Fact* 2017, *16*, 206.
[71] Mironczuk, A. M., Furgala, J., Rakicka, M., Rymowicz, W., Enhanced production of erythritol by Yarrowia lipolytica on glycerol in repeated batch cultures. *J Ind Microbiol Biotechnol* 2014, *41*, 57-64.
[72] Ledesma-Amaro, R., Lazar, Z., Rakicka, M., Guo, Z., *et al.*, Metabolic engineering of Yarrowia lipolytica to produce chemicals and fuels from xylose. *Metab Eng* 2016, *38*, 115-124.
[73] Abghari, A. S., S., Engineering Yarrowia lipolytica for Enhanced Production of Lipid and Citric Acid. *Fermentation* 2017, *3*, 34-56.
[74] Silva, G. S., Assis, D. J., Druzian, J. I., Oliveira, M., *et al.*, Impact of Preservation Conditions on Fatty Acids, Xanthan Gum Production and Other Characteristics of Xanthomonas campestris pv. mangiferaeindicae IBSBF 2103. *Indian J Microbiol* 2017, *57*, 351-358.
[75] Goncalves, F. A., Colen, G., Takahashi, J. A., Yarrowia lipolytica and its multiple applications in the biotechnological industry. *ScientificWorldJournal* 2014, *2014*, 476207.
[76] Zhu, Q., Jackson, E. N., Metabolic engineering of Yarrowia lipolytica for industrial applications. *Curr Opin Biotechnol* 2015, *36*, 65-72.
[77] Cordero Otero, R., Gaillardin, C., Efficient selection of hygromycin-B-resistant Yarrowia lipolytica transformants. *Appl Microbiol Biotechnol* 1996, *46*, 143-148.
[78] Chen, D. C., Beckerich, J. M., Gaillardin, C., One-step transformation of the dimorphic yeast Yarrowia lipolytica. *Appl Microbiol Biotechnol* 1997, *48*, 232-235.
[79] Liu, L. Q., Otoupal, P., Pan, A., Alper, H. S., Increasing expression level and copy number of a Yarrowia lipolytica plasmid through regulated centromere function. *Fems Yeast Res* 2014, *14*, 1124-1127.
[80] Shabbir Hussain, M., Gambill, L., Smith, S., Blenner, M. A., Engineering Promoter Architecture in Oleaginous Yeast Yarrowia lipolytica. *Acs Synth Biol* 2016, *5*, 213-223.
[81] Blazeck, J., Reed, B., Garg, R., Gerstner, R., *et al.*, Generalizing a hybrid synthetic promoter approach in Yarrowia lipolytica. *Appl Microbiol Biot* 2013, *97*, 3037-3052.
[82] Blazeck, J., Liu, L., Redden, H., Alper, H., Tuning gene expression in Yarrowia lipolytica by a hybrid promoter approach. *Appl Environ Microbiol* 2011, *77*, 7905-7914.
[83] Schwartz, C. M., Hussain, M. S., Blenner, M., Wheeldon, I., Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. *Acs Synth Biol* 2016, *5*, 356-359.
[84] Qiao, K., Wasylenko, T. M., Zhou, K., Xu, P., Stephanopoulos, G., Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. *Nat Biotechnol* 2017, *35*, 173-177.
[85] Xue, Z., Sharpe, P. L., Hong, S. P., Yadav, N. S., *et al.*, Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. *Nature biotechnology* 2013, *31*, 734-740.

[86] Smit, M. S., Mokgoro, M. M., Setati, E., Nicaud, J. M., alpha,omega-Dicarboxylic acid accumulation by acyl-CoA oxidase deficient mutants of Yarrowia lipolytica. *Biotechnol Lett* 2005, *27*, 859-864.
[87] Haddouche, R., Poirier, Y., Delessert, S., Sabirova, J., *et al.*, Engineering polyhydroxyalkanoate content and monomer composition in the oleaginous yeast Yarrowia lipolytica by modifying the ss-oxidation multifunctional protein. *Appl Microbiol Biotechnol* 2011, *91*, 1327-1340.
[88] Blazeck, J., Hill, A., Jamoussi, M., Pan, A., *et al.*, Metabolic engineering of Yarrowia lipolytica for itaconic acid production. *Metab Eng* 2015, *32*, 66-73.
[89] Ledesma-Amaro, R., Dulermo, R., Niehus, X., Nicaud, J. M., Combining metabolic engineering and process optimization to improve production and secretion of fatty acids. *Metab Eng* 2016, *38*, 38-46.
[90] Runguphan, W., Keasling, J. D., Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid-derived biofuels and chemicals. *Metab Eng* 2014, *21*, 103-113.
[91] Xu, P., Qiao, K., Ahn, W. S., Stephanopoulos, G., Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc Natl Acad Sci U S A* 2016, *113*, 10848-10853.
[92] Wang, G., Xiong, X., Ghogare, R., Wang, P., *et al.*, Exploring fatty alcohol-producing capability of Yarrowia lipolytica. *Biotechnol Biofuels* 2016, *9*, 107.
[93] Rutter, C. D., Rao, C. V., Production of 1-decanol by metabolically engineered Yarrowia lipolytica. *Metab Eng* 2016, *38*, 139-147.
[94] Ryu, S., Hipp, J., Trinh, C. T., Activating and Elucidating Metabolism of Complex Sugars in Yarrowia lipolytica. *Appl Environ Microbiol* 2015, *82*, 1334-1345.
[95] Rodriguez, G. M., Hussain, M. S., Gambill, L., Gao, D., *et al.*, Engineering xylose utilization in Yarrowia lipolytica by understanding its cryptic xylose pathway. *Biotechnol Biofuels* 2016, *9*, 149.
[96] Ledesma-Amaro, R., Nicaud, J. M., Metabolic Engineering for Expanding the Substrate Range of Yarrowia lipolytica. *Trends Biotechnol* 2016, *34*, 798-809.
[97] Li, H., Alper, H. S., Enabling xylose utilization in Yarrowia lipolytica for lipid production. *Biotechnol J* 2016, *11*, 1230-1240.
[98] Ledesma-Amaro, R., Dulermo, T., Nicaud, J. M., Engineering *Yarrowia lipolytica* to produce biodiesel from raw starch. *Biotechnol Biofuels* 2015, *8*, 148.
[99] Celinska, E., Bialas, W., Borkowska, M., Grajek, W., Cloning, expression, and purification of insect (*Sitophilus oryzae*) alpha-amylase, able to digest granular starch, in *Yarrowia lipolytica* host. *Appl Microbiol Biotechnol* 2015, *99*, 2727-2739.
[100] Yang, C. H., Huang, Y. C., Chen, C. Y., Wen, C. Y., Heterologous expression of Thermobifida fusca thermostable alpha-amylase in *Yarrowia lipolytica* and its application in boiling stable resistant sago starch preparation. *J Ind Microbiol Biotechnol* 2010, *37*, 953-960.
[101] Dujon, B., Sherman, D., Fischer, G., Durrens, P., *et al.*, Genome evolution in yeasts. *Nature* 2004, *430*, 35-44.

[102] Magnan, C., Yu, J., Chang, I., Jahn, E., *et al.*, Sequence Assembly of *Yarrowia lipolytica* Strain W29/CLIB89 Shows Transposable Element Diversity. *PLoS One* 2016, *11*, e0162363.
[103] van Heerikhuizen, H., Ykema, A., Klootwijk, J., Gaillardin, C., *et al.*, Heterogeneity in the ribosomal RNA genes of the yeast Yarrowia lipolytica; cloning and analysis of two size classes of repeats. *Gene* 1985, *39*, 213-222.
[104] Davidow, L. S., Apostolakos, D., O'Donnell, M. M., Proctor, A. R., *et al.*, Integrative transformation of the yeast *Yarrowia lipolytica*. *Current Genetics* 1985, *10*, 39-48.
[105] Barth, G., Gaillardin, C., *Yarrowia lipolytica In: Nonconventional Yeasts in Biotechnology (Wolf, K., Ed.), A Handbook*, Springer, Berlin, Heidelberg, NewYork 1996.
[106] Barth, G., and Gaillardin, C., *Yarrowia lipolytica*, *Nonconventional Yeasts in Biotechnology*, Springer, Berlin, Heidelberg 1996, pp. 313-388.
[107] Wang, J. H., Hung, W., Tsai, S. H., High efficiency transformation by electroporation of *Yarrowia lipolytica*. *J Microbiol* 2011, *49*, 469-472.
[108] Duquesne, S., Bordes, F., Fudalej, F., Nicaud, J. M., Marty, A., The yeast *Yarrowia lipolytica* as a generic tool for molecular evolution of enzymes. *Methods in molecular biology* 2012, *861*, 301-312.
[109] Theerachat, M., Emond, S., Cambon, E., Bordes, F., *et al.*, Engineering and production of laccase from Trametes versicolor in the yeast Yarrowia lipolytica. *Bioresour Technol* 2012, *125*, 267-274.
[110] Juretzek, T., Wang, H.J., Nicaud, J.M., Mauersberger, S., Barth, G., Comparison of Promoters Suitable for Regulated Overexpression of Beta-Galactosidase in the Alkane-Utilizing Yeast Yarrowia lipolytica. *Biotechnology & Bioprocess Engineering* 2000, *5*, 320-326.
[111] Gao, S., Han, L., Zhu, L., Ge, M., *et al.*, One-step integration of multiple genes into the oleaginous yeast *Yarrowia lipolytica*. *Biotechnology letters* 2014, *36*, 2523-2528.
[112] Curran, K. A., Morse, N. J., Markham, K. A., Wagman, A. M., *et al.*, Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS synthetic biology* 2015, *4*, 824-832.
[113] Laine, J. P., Singh, B. N., Krishnamurthy, S., Hampsey, M., A physiological role for gene loops in yeast. *Genes Dev* 2009, *23*, 2604-2609.
[114] Mori, K., Iwama, R., Kobayashi, S., Horiuchi, H., *et al.*, Transcriptional repression by glycerol of genes involved in the assimilation of n-alkanes and fatty acids in yeast Yarrowia lipolytica. *Fems Yeast Res* 2013, *13*, 233-240.
[115] Sumita, T., Iida, T., Yamagami, S., Horiuchi, H., *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 Bioph Res Co* 2002, *294*, 1071-1078.
[116] Shabbir Hussain, M., Wheeldon, I., Blenner, M. A., A Strong Hybrid Fatty Acid Inducible Transcriptional Sensor Built From Yarrowia lipolytica Upstream Activating and Regulatory Sequences. *Biotechnol J* 2017.

[117] Sassi, H., Delvigne, F., Kar, T., Nicaud, J. M., *et al.*, Deciphering how LIP2 and POX2 promoters can optimally regulate recombinant protein production in the yeast Yarrowia lipolytica. *Microb Cell Fact* 2016, *15*.
[118] Hussain, M. S., Rodriguez, G. M., Gao, D. F., Spagnuolo, M., *et al.*, Recent advances in bioengineering of the oleaginous yeast Yarrowia lipolytica. *Aims Bioengineering* 2016, *3*, 493-514.
[119] Wang, H. J., Le Dall, M. T., Wach, Y., Laroche, C., *et al.*, Evaluation of acyl coenzyme A oxidase (Aox) isozyme function in the n-alkane-assimilating yeast Yarrowia lipolytica. *J Bacteriol* 1999, *181*, 5140-5148.
[120] 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. *J Biol Chem* 2004, *279*, 22183-22189.
[121] Blazeck, J., Reed, B., Garg, R., Gerstner, R., *et al.*, Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*. *Applied microbiology and biotechnology* 2013, *97*, 3037-3052.
[122] Blanchin-Roland, S., Cordero Otero, R. R., Gaillardin, C., Two upstream activation sequences control the expression of the XPR2 gene in the yeast Yarrowia lipolytica. *Mol Cell Biol* 1994, *14*, 327-338.
[123] Madzak, C., Blanchin-Roland, S., Cordero Otero, R. R., Gaillardin, C., Functional analysis of upstream regulating regions from the Yarrowia lipolytica XPR2 promoter. *Microbiology* 1999, *145 (Pt 1)*, 75-87.
[124] 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* 2000, *2*, 207-216.
[125] 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* 1998, *14*, 1267-1283.
[126] 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*. *The Journal of biological chemistry* 2004, *279*, 22183-22189.
[127] Trassaert, M., Vandermies, M., Carly, F., Denies, O., *et al.*, New inducible promoter for gene expression and synthetic biology in Yarrowia lipolytica. *Microb Cell Fact* 2017, *16*, 141.
[128] Kozak, M., Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic acids research* 1984, *12*, 857-872.
[129] Kozak, M., Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 1986, *44*, 283-292.
[130] Merkulov, S., van Assema, F., Springer, J., Fernandez Del Carmen, A., Mooibroek, H., Cloning and characterization of the *Yarrowia lipolytica* squalene synthase (SQS1) gene and functional complementation of the Saccharomyces cerevisiae erg9 mutation. *Yeast* 2000, *16*, 197-206.

[131] Zhang, B., Rong, C., Chen, H., Song, Y., *et al.*, De novo synthesis of trans-10, cis-12 conjugated linoleic acid in oleaginous yeast Yarrowia lipolytica. *Microb Cell Fact* 2012, *11*, 51.
[132] Kramara, J., Willcox, S., Gunisova, S., Kinsky, S., *et al.*, Tay1 protein, a novel telomere binding factor from *Yarrowia lipolytica*. *The Journal of biological chemistry* 2010, *285*, 38078-38092.
[133] Farrel, C., Mayorga, M., Chevreux, B., in: Ltd, D. N. P. (Ed.) 2013.
[134] Gasmi, N., Fudalej, F., Kallel, H., Nicaud, J. M., A molecular approach to optimize hIFN alpha2b expression and secretion in *Yarrowia lipolytica*. *Applied microbiology and biotechnology* 2011, *89*, 109-119.
[135] Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H., Miura, K., Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes. *Nucleic acids research* 2008, *36*, 861-871.
[136] Kanaar, R., Hoeijmakers, J. H., van Gent, D. C., Molecular mechanisms of DNA double strand break repair. *Trends in cell biology* 1998, *8*, 483-489.
[137] Pastink, A., Lohman, P. H., Repair and consequences of double-strand breaks in DNA. *Mutation research* 1999, *428*, 141-156.
[138] Pastink, A., Eeken, J. C., Lohman, P. H., Genomic integrity and the repair of double-strand DNA breaks. *Mutation research* 2001, *480-481*, 37-50.
[139] Loeillet, S., Palancade, B., Cartron, M., Thierry, A., *et al.*, Genetic network interactions among replication, repair and nuclear pore deficiencies in yeast. *DNA repair* 2005, *4*, 459-468.
[140] 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. *J Microbiol Methods* 2003, *55*, 727-737.
[141] Van Dyck, E., Stasiak, A. Z., Stasiak, A., West, S. C., Binding of double-strand breaks in DNA by human Rad52 protein. *Nature* 1999, *398*, 728-731.
[142] Verbeke, J., Beopoulos, A., Nicaud, J. M., Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains. *Biotechnology letters* 2013, *35*, 571-576.
[143] Kretzschmar, A., Otto, C., Holz, M., Werner, S., *et al.*, Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. *Curr Genet* 2013, *59*, 63-72.
[144] Vasiliki, T., Elena, B., Gregory, S., Joe, S. A., Improved Gene Targeting through Cell Cycle Synchronization. *PloS one* 2015.
[145] Isaac, R. S., Jiang, F., Doudna, J. A., Lim, W. A., *et al.*, Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* 2016, *5*.
[146] Hamilton, D. L., Abremski, K., Site-specific recombination by the bacteriophage P1 lox-Cre system. Cre-mediated synapsis of two lox sites. *Journal of molecular biology* 1984, *178*, 481-486.
[147] Mauersberger, S., Wang, H. J., Gaillardin, C., Barth, G., Nicaud, J. M., Insertional mutagenesis in the n-alkane-assimilating yeast Yarrowia lipolytica: generation of tagged mutations in genes involved in hydrophobic substrate utilization. *J Bacteriol* 2001, *183*, 5102-5109.

[148] Pignede, G., Wang, H. J., Fudalej, F., Seman, M., *et al.*, Autocloning and amplification of LIP2 in Yarrowia lipolytica. *Appl Environ Microbiol* 2000, *66*, 3283-3289.
[149] Ledesma-Amaro, R., Dulermo, R., Niehus, X., Nicaud, J. M., Combining metabolic engineering and process optimization to improve production and secretion of fatty acids. *Metabolic engineering* 2016, *38*, 38-46.
[150] Gao, C., Qi, Q., Madzak, C., Lin, C. S., Exploring medium-chain-length polyhydroxyalkanoates production in the engineered yeast *Yarrowia lipolytica*. *Journal of industrial microbiology & biotechnology* 2015, *42*, 1255-1262.
[151] Ratledge, C., Regulation of lipid accumulation in oleaginous micro-organisms. *Biochemical Society transactions* 2002, *30*, 1047-1050.
[152] Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J. L., *et al.*, Yarrowia lipolytica as a model for bio-oil production. *Prog Lipid Res* 2009, *48*, 375-387.
[153] Zhang, H., Zhang, L., Chen, H., Chen, Y. Q., *et al.*, Regulatory properties of malic enzyme in the oleaginous yeast, *Yarrowia lipolytica,* and its non-involvement in lipid accumulation. *Biotechnology letters* 2013, *35*, 2091-2098.
[154] Wasylenko, T. M., Ahn, W. S., Stephanopoulos, G., The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metab Eng* 2015, *30*, 27-39.
[155] Liu, N., Qiao, K., Stephanopoulos, G., 13C Metabolic Flux Analysis of acetate conversion to lipids by Yarrowia lipolytica. *Metab Eng* 2016, *38*, 86-97.
[156] Morin, N., Cescut, J., Beopoulos, A., Lelandais, G., *et al.*, Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *Yarrowia lipolytica*. *PloS one* 2011, *6*, e27966.
[157] Beopoulos, A., Mrozova, Z., Thevenieau, F., Le Dall, M. T., *et al.*, Control of lipid accumulation in the yeast Yarrowia lipolytica. *Appl Environ Microbiol* 2008, *74*, 7779-7789.
[158] Beopoulos, A., Haddouche, R., Kabran, P., Dulermo, T., *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* 2012, *93*, 1523-1537.
[159] Tai, M., Stephanopoulos, G., Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. *Metab Eng* 2013, *15*, 1-9.
[160] Qiao, K., Imam Abidi, S. H., Liu, H., Zhang, H., *et al.*, Engineering lipid overproduction in the oleaginous yeast *Yarrowia lipolytica*. *Metabolic engineering* 2015, *29*, 56-65.
[161] Wang, Z. P., Xu, H. M., Wang, G. Y., Chi, Z., Chi, Z. M., Disruption of the MIG1 gene enhances lipid biosynthesis in the oleaginous yeast *Yarrowia lipolytica* ACA-DC 50109. *Biochimica et biophysica acta* 2013, *1831*, 675-682.
[162] Seip, J., Jackson, R., He, H., Zhu, Q., Hong, S. P., Snf1 is a regulator of lipid accumulation in Yarrowia lipolytica. *Appl Environ Microbiol* 2013, *79*, 7360-7370.

[163] 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*. *Metabolic engineering* 2011, *13*, 482-491.
[164] Blazeck, J., Hill, A., Liu, L., Knight, R., *et al.*, Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 2014, *5*, 3131.
[165] Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E. H., *et al.*, Engineering of a high lipid producing *Yarrowia lipolytica* strain. *Biotechnology for biofuels* 2016, *9*, 77.
[166] Liu, L., Pan, A., Spofford, C., Zhou, N., Alper, H. S., An evolutionary metabolic engineering approach for enhancing lipogenesis in *Yarrowia lipolytica*. *Metabolic engineering* 2015, *29*, 36-45.
[167] Liu, L., Markham, K., Blazeck, J., Zhou, N., *et al.*, Surveying the lipogenesis landscape in *Yarrowia lipolytica* through understanding the function of a Mga2p regulatory protein mutant. *Metabolic engineering* 2015, *31*, 102-111.
[168] Poirier, Y., Antonenkov, V. D., Glumoff, T., Hiltunen, J. K., Peroxisomal beta-oxidation--a metabolic pathway with multiple functions. *Biochimica et biophysica acta* 2006, *1763*, 1413-1426.
[169] Dulermo, R., Gamboa-Melendez, H., Ledesma-Amaro, R., Thevenieau, F., Nicaud, J. M., Unraveling fatty acid transport and activation mechanisms in *Yarrowia lipolytica*. *Biochimica et biophysica acta* 2015, *1851*, 1202-1217.
[170] Haddouche, R., Delessert, S., Sabirova, J., Neuveglise, C., *et al.*, Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards beta-oxidation and polyhydroxyalkanoate biosynthesis in Yarrowia lipolytica. *Fems Yeast Res* 2010, *10*, 917-927.
[171] Maston, G. A., Evans, S. K., Green, M. R., Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006, *7*, 29-59.
[172] Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., *et al.*, The sequence of the human genome. *Science* 2001, *291*, 1304-1351.
[173] 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* 2008, *105*, 571-578.
[174] Curran, K. A., Crook, N. C., Karim, A. S., Gupta, A., *et al.*, Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nat Commun* 2014, *5*.
[175] Nevoigt, E., Kohnke, J., Fischer, C. R., Alper, H., *et al.*, Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol* 2006, *72*, 5266-5273.
[176] Gaillardin, C., Ribet, A. M., LEU2 directed expression of beta-galactosidase activity and phleomycin resistance in Yarrowia lipolytica. *Curr Genet* 1987, *11*, 369-375.
[177] Mumberg, D., Muller, R., Funk, M., Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* 1995, *156*, 119-122.
[178] Schneider, J. C., Guarente, L., Vectors for expression of cloned genes in yeast: regulation, overproduction, and underproduction. *Methods Enzymol* 1991, *194*, 373-388.

[179] Caspeta, L., Nielsen, J., Toward systems metabolic engineering of Aspergillus and Pichia species for the production of chemicals and biofuels. *Biotechnol J* 2013, *8*, 534-544.
[180] Goshima, T., Negi, K., Tsuji, M., Inoue, H., *et al.*, Ethanol fermentation from xylose by metabolically engineered strains of Kluyveromyces marxianus. *J Biosci Bioeng* 2013, *116*, 551-554.
[181] Harner, N. K., Wen, X., Bajwa, P. K., Austin, G. D., *et al.*, Genetic improvement of native xylose-fermenting yeasts for ethanol production. *J Ind Microbiol Biotechnol* 2015, *42*, 1-20.
[182] Liu, L., Redden, H., Alper, H. S., Frontiers of yeast metabolic engineering: diversifying beyond ethanol and Saccharomyces. *Curr Opin Biotechnol* 2013, *24*, 1023-1030.
[183] Wu, S. G., He, L., Wang, Q., Tang, Y. J., An ancient Chinese wisdom for metabolic engineering: Yin-Yang. *Microb Cell Fact* 2015, *14*, 39.
[184] Xu, P., Li, L., Zhang, F., Stephanopoulos, G., Koffas, M., Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc Natl Acad Sci U S A* 2014, *111*, 11299-11304.
[185] Farmer, W. R., Liao, J. C., Improving lycopene production in Escherichia coli by engineering metabolic control. *Nature Biotechnology* 2000, *18*, 533-537.
[186] 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* 2012, *30*, 354-U166.
[187] Roy, P. K., Singh, H. D., Bhagat, S. D., Baruah, J. N., Characterization of Hydrocarbon Emulsification and Solubilization Occurring during the Growth of Endomycopsis-Lipolytica on Hydrocarbons. *Biotechnology and Bioengineering* 1979, *21*, 955-974.
[188] Ruiz-Herrera, J., Sentandreu, R., Different effectors of dimorphism in Yarrowia lipolytica. *Arch Microbiol* 2002, *178*, 477-483.
[189] Fickers, P., Benetti, P. H., Wache, Y., Marty, A., *et al.*, Hydrophobic substrate utilisation by the yeast Yarrowia lipolytica, and its potential applications. *Fems Yeast Res* 2005, *5*, 527-543.
[190] Blazeck, J., Hill, A., Liu, L. Q., Knight, R., *et al.*, Harnessing Yarrowia lipolytica lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 2014, *5*.
[191] Beopoulos, A., Verbeke, J., Bordes, F., Guicherd, M., *et al.*, Metabolic engineering for ricinoleic acid production in the oleaginous yeast Yarrowia lipolytica. *Appl Microbiol Biot* 2014, *98*, 251-262.
[192] Anastassiadis, S., Aivasidis, A., Wandrey, C., Citric acid production by Candida strains under intracellular nitrogen limitation. *Appl Microbiol Biot* 2002, *60*, 81-87.
[193] Fickers, P., Nicaud, J. M., Gaillardin, C., Destain, J., Thonart, P., Carbon and nitrogen sources modulate lipase production in the yeast Yarrowia lipolytica. *J Appl Microbiol* 2004, *96*, 742-749.
[194] Zhou, J. W., Yin, X. X., Madzak, C., Du, G. C., Chen, J., Enhanced alpha-ketoglutarate production in Yarrowia lipolytica WSH-Z06 by alteration of the acetyl-CoA metabolism. *J Biotechnol* 2012, *161*, 257-264.

[195] Ryu, S., Labbe, N., Trinh, C. T., Simultaneous saccharification and fermentation of cellulose in ionic liquid for efficient production of alpha-ketoglutaric acid by Yarrowia lipolytica. *Appl Microbiol Biot* 2015, *99*, 4237-4244.
[196] Matthaus, F., Ketelhot, M., Gatter, M., Barth, G., Production of Lycopene in the Non-Carotenoid-Producing Yeast Yarrowia lipolytica. *Appl Environ Microb* 2014, *80*, 1660-1669.
[197] Lubliner, S., Keren, L., Segal, E., Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Res* 2013, *41*, 5569-5581.
[198] Weingarten-Gabbay, S., Segal, E., The grammar of transcriptional regulation. *Hum Genet* 2014, *133*, 701-711.
[199] 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. *J Mol Microb Biotech* 2000, *2*, 207-216.
[200] Madzak, C., Blanchin-Roland, S., Otero, R. R. C., Gaillardin, C., Functional analysis of upstream regulating regions from the Yarrowia lipolytica XPR2 promoter. *Microbiol-Uk* 1999, *145*, 75-87.
[201] Li, M. Z., Elledge, S. J., Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* 2007, *4*, 251-256.
[202] Erb, I., van Nimwegen, E., Transcription factor binding site positioning in yeast: proximal promoter motifs characterize TATA-less promoters. *PLoS One* 2011, *6*, e24279.
[203] Singer, V. L., Wobbe, C. R., Struhl, K., A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes Dev* 1990, *4*, 636-645.
[204] Wobbe, C. R., Struhl, K., Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. *Mol Cell Biol* 1990, *10*, 3859-3867.
[205] Flagfeldt, D. B., Siewers, V., Huang, L., Nielsen, J., Characterization of chromosomal integration sites for heterologous gene expression in Saccharomyces cerevisiae. *Yeast* 2009, *26*, 545-551.
[206] Nicaud, J. M., Madzak, C., van den Broek, P., Gysler, C., *et al.*, Protein expression and secretion in the yeast Yarrowia lipolytica. *Fems Yeast Res* 2002, *2*, 371-379.
[207] Nthangeni, M. B., Urban, P., Pompon, D., Smit, M. S., Nicaud, J. M., The use of Yarrowia lipolytica for the expression of human cytochrome P450CYP1A1. *Yeast* 2004, *21*, 583-592.
[208] Poopanitpan, N., Kobayashi, S., Fukuda, R., Honuchi, H., Ohta, A., An ortholog of farA of Aspergillus nidulans is implicated in the transcriptional activation of genes involved in fatty acid utilization in the yeast Yarrowia lipolytica. *Biochem Bioph Res Co* 2010, *402*, 731-735.
[209] Lubliner, S., Keren, L., Segal, E., Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Research* 2013, *41*, 5569-5581.

[210] Mishra, A. K., Vanathi, P., Bhargava, P., The transcriptional activator GAL4-VP16 regulates the intra-molecular interactions of the TATA-binding protein. *J Biosci* 2003, *28*, 423-436.
[211] Jiang, C., Pugh, B. F., Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 2009, *10*, 161-172.
[212] Furter-Graves, E. M., Hall, B. D., DNA sequence elements required for transcription initiation of the Schizosaccharomyces pombe ADH gene in Saccharomyces cerevisiae. *Mol Gen Genet* 1990, *223*, 407-416.
[213] Hahn, S., Hoar, E. T., Guarente, L., Each of three "TATA elements" specifies a subset of the transcription initiation sites at the CYC-1 promoter of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* 1985, *82*, 8562-8566.
[214] Zhang, Z., Dietrich, F. S., Mapping of transcription start sites in Saccharomyces cerevisiae using 5' SAGE. *Nucleic Acids Res* 2005, *33*, 2838-2851.
[215] Alper, H., Fischer, C., Nevoigt, E., Stephanopoulos, G., Tuning genetic control through promoter engineering (vol 102, pg 12678, 2005). *P Natl Acad Sci USA* 2006, *103*, 3006-3006.
[216] Shabbir Hussain, M., Rodriguez, G., D., G., M., S., *et al.*, Recent Advances in Metabolic Engineering of Yarrowia lipolytica. *AIMS Bioengineering* 2016, *3*, 493-514.
[217] Blazeck, J., Alper, H. S., Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnol J* 2013, *8*, 46-+.
[218] Qin, X. L., Qian, J. C., Yao, G. F., Zhuang, Y. P., *et al.*, GAP Promoter Library for Fine-Tuning of Gene Expression in Pichia pastoris. *Appl Environ Microb* 2011, *77*, 3600-3608.
[219] Nielsen, J., Larsson, C., van Maris, A., Pronk, J., Metabolic engineering of yeast for production of fuels and chemicals. *Curr Opin Biotech* 2013, *24*, 398-404.
[220] Da Silva, N. A., Srikrishnan, S., Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae. *Fems Yeast Res* 2012, *12*, 197-214.
[221] Guarente, L., Yocum, R. R., Gifford, P., A Gal10-Cyc1 Hybrid Yeast Promoter Identifies the Gal4 Regulatory Region as an Upstream Site. *P Natl Acad Sci-Biol* 1982, *79*, 7410-7414.
[222] Zaslaver, A., Mayo, A. E., Rosenberg, R., Bashkin, P., *et al.*, Just-in-time transcription program in metabolic pathways. *Nat Genet* 2004, *36*, 486-491.
[223] Holtz, W. J., Keasling, J. D., Engineering Static and Dynamic Control of Synthetic Pathways. *Cell* 2010, *140*, 19-23.
[224] Weinhandl, K., Winkler, M., Glieder, A., Camattari, A., Carbon source dependent promoters in yeasts. *Microb Cell Fact* 2014, *13*.
[225] Wimalarathna, R. N., Pan, P. Y., Shen, C. H., Co-dependent recruitment of Ino80p and Snf2p is required for yeast CUP1 activation. *Biochem Cell Biol* 2014, *92*, 69-75.
[226] Korber, P., Barbaric, S., The yeast PHO5 promoter: from single locus to systems biology of a paradigm for gene regulation through chromatin. *Nucleic Acids Research* 2014, *42*, 10888-10902.

[227] Solow, S. P., Sengbusch, J., Laird, M. W., Heterologous protein production from the inducible MET25 promoter in Saccharomyces cerevisiae. *Biotechnol Prog* 2005, *21*, 617-620.
[228] West, R. W., Chen, S. M., Putz, H., Butler, G., Banerjee, M., Gal1-Gal10 Divergent Promoter Region of Saccharomyces-Cerevisiae Contains Negative Control Elements in Addition to Functionally Separate and Possibly Overlapping Upstream Activating Sequences. *Gene Dev* 1987, *1*, 1118-1131.
[229] Leavitt, J. M., Tong, A., Tong, J., Pattie, J., Alper, H. S., Coordinated transcription factor and promoter engineering to establish strong expression elements in Saccharomyces cerevisiae. *Biotechnol J* 2016, *11*, 866-876.
[230] Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M., Wheeldon, I., Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. *Acs Synth Biol* 2017, *6*, 402-409.
[231] Qiao, K., Abidi, S. H. I., Liu, H. J., Zhang, H. R., *et al.*, Engineering lipid overproduction in the oleaginous yeast Yarrowia lipolyticale. *Metab Eng* 2015, *29*, 56-65.
[232] Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E. H., *et al.*, Engineering of a high lipid producing Yarrowia lipolytica strain. *Biotechnol Biofuels* 2016, *9*.
[233] Hu, P., Chakraborty, S., Kumar, A., Woolston, B., *et al.*, Integrated bioprocess for conversion of gaseous substrates to liquids. *P Natl Acad Sci USA* 2016, *113*, 3773-3778.
[234] Shaw, A. J., Lam, F. H., Hamilton, M., Consiglio, A., *et al.*, Metabolic engineering of microbial competitive advantage for industrial fermentation processes. *Science* 2016, *353*, 583-586.
[235] Dulermo, R., Brunel, F., Dulermo, T., Ledesma-Amaro, R., *et al.*, Using a vector pool containing variable-strength promoters to optimize protein production in Yarrowia lipolytica. *Microb Cell Fact* 2017, *16*, 31.
[236] Thevenieau, F., Le Dall, M. T., Nthangeni, B., Mauersberger, S., *et al.*, Characterization of Yarrowia lipolytica mutants affected in hydrophobic substrate utilization. *Fungal Genetics and Biology* 2007, *44*, 531-542.
[237] Garofano, A., Eschemann, A., Brandt, U., Kerscher, S., Substrate-inducible versions of internal alternative NADH: ubiquinone oxidoreductase from Yarrowia lipolytica. *Yeast* 2006, *23*, 1129-1136.
[238] Gasmi, N., Ayed, A., Ammar, B. B. H., Zrigui, R., *et al.*, Development of a cultivation process for the enhancement of human interferon alpha 2b production in the oleaginous yeast, Yarrowia lipolytica. *Microb Cell Fact* 2011, *10*.
[239] Hong, S. P., Seip, J., Walters-Pollak, D., Rupert, R., *et al.*, Engineering Yarrowia lipolytica to express secretory invertase with strong FBA1IN promoter. *Yeast* 2012, *29*, 59-72.
[240] De Pourcq, K., Vervecken, W., Dewerte, I., Valevska, A., *et al.*, Engineering the yeast Yarrowia lipolytica for the production of therapeutic proteins homogeneously glycosylated with Man(8)GlcNAc(2) and Man(5)GlcNAc(2). *Microb Cell Fact* 2012, *11*, 53.

[241] Green, M. R., Sambrook, J., Sambrook, J., *Molecular cloning : a laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2012.
[242] Jeong, J. Y., Yim, H. S., Ryu, J. Y., Lee, H. S., *et al.*, One-step sequence- and ligation-independent cloning as a rapid and versatile cloning method for functional genomics studies. *Appl Environ Microbiol* 2012, *78*, 5440-5443.
[243] Blazeck, J., Liu, L. Q., Redden, H., Alper, H., Tuning Gene Expression in Yarrowia lipolytica by a Hybrid Promoter Approach. *Appl Environ Microb* 2011, *77*, 7905-7914.
[244] Ding, W. T., Zhang, G. C., Liu, J. J., 3 ' Truncation of the GPD1 Promoter in Saccharomyces cerevisiae for Improved Ethanol Yield and Productivity. *Appl Environ Microb* 2013, *79*, 3273-3281.
[245] Eckert, S. E., Muhlschlegel, F. A., Promoter regulation in Candida albicans and related species. *Fems Yeast Res* 2009, *9*, 2-15.
[246] Lopez-Estrano, C., Gopalakrishnan, A. M., Semblat, J. P., Fergus, M. R., *et al.*, An enhancer-like region regulates hrp3 promoter stage-specific gene expression in the human malaria parasite Plasmodium falciparum. *Bba-Gene Struct Expr* 2007, *1769*, 506-513.
[247] Hynes, M. J., Murray, S. L., Duncan, A., Khew, G. S., Davis, M. A., Regulatory genes controlling fatty acid catabolism and peroxisomal functions in the filamentous fungus Aspergillus nidulans. *Eukaryot Cell* 2006, *5*, 794-805.
[248] Li, R., Stimulation of DNA replication in Saccharomyces cerevisiae by a glutamine- and proline-rich transcriptional activation domain. *J Biol Chem* 1999, *274*, 30310-30314.
[249] Butler, J. E., Kadonaga, J. T., The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* 2002, *16*, 2583-2592.
[250] Koller, A., Valesco, J., Subramani, S., The CUP1 promoter of Saccharomyces cerevisiae is inducible by copper in Pichia pastoris. *Yeast* 2000, *16*, 651-656.
[251] de Winde, J. H., Grivell, L. A., Global regulation of mitochondrial biogenesis in Saccharomyces cerevisiae. *Prog Nucleic Acid Res Mol Biol* 1993, *46*, 51-91.
[252] Pinkham, J. L., Guarente, L., Cloning and molecular analysis of the HAP2 locus: a global regulator of respiratory genes in Saccharomyces cerevisiae. *Mol Cell Biol* 1985, *5*, 3410-3416.
[253] Finley, R. L., Jr., Chen, S., Ma, J., Byrne, P., West, R. W., Jr., Opposing regulatory functions of positive and negative elements in UASG control transcription of the yeast GAL genes. *Mol Cell Biol* 1990, *10*, 5663-5670.
[254] Barrett, L. W., Fletcher, S., Wilton, S. D., Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell Mol Life Sci* 2012, *69*, 3613-3634.
[255] Struhl, K., Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 1999, *98*, 1-4.
[256] Egermeier, M., Russmayer, H., Sauer, M., Marx, H., Metabolic Flexibility of Yarrowia lipolytica Growing on Glycerol. *Front Microbiol* 2017, *8*, 49.
[257] Thevenieau, F., Beopoulos, A., Desfougeres, T., Sabirova, J., *et al.*, Uptake and Assimilation of Hydrophobic Substrates by the Oleaginous Yeast Yarrowia lipolytica, in:

Timmis, K. N. (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology*, Springer Berlin Heidelberg, Berlin, Heidelberg 2010, pp. 1513-1527.
[258] Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., *et al.*, Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 2006, *440*, 940-943.
[259] Maury, J., Asadollahi, M. A., Moller, K., Schalk, M., *et al.*, Reconstruction of a bacterial isoprenoid biosynthetic pathway in Saccharomyces cerevisiae. *FEBS Lett* 2008, *582*, 4032-4038.
[260] Rohde, J. R., Trinh, J., Sadowski, I., Multiple signals regulate GAL transcription in yeast. *Mol Cell Biol* 2000, *20*, 3880-3886.
[261] Holstege, F. C., Jennings, E. G., Wyrick, J. J., Lee, T. I., *et al.*, Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 1998, *95*, 717-728.
[262] Arents, G., Burlingame, R. W., Wang, B. C., Love, W. E., Moudrianakis, E. N., The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. *Proc Natl Acad Sci U S A* 1991, *88*, 10148-10152.
[263] Finch, J. T., Lutter, L. C., Rhodes, D., Brown, R. S., *et al.*, Structure of nucleosome core particles of chromatin. *Nature* 1977, *269*, 29-36.
[264] Kornberg, R. D., Structure of chromatin. *Annu Rev Biochem* 1977, *46*, 931-954.
[265] Felle, M., Hoffmeister, H., Rothammer, J., Fuchs, A., *et al.*, Nucleosomes protect DNA from DNA methylation in vivo and in vitro. *Nucleic Acids Res* 2011, *39*, 6956-6969.
[266] McGinty, R. K., Tan, S., Nucleosome structure and function. *Chem Rev* 2015, *115*, 2255-2273.
[267] Workman, J. L., Kingston, R. E., Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Annu Rev Biochem* 1998, *67*, 545-579.
[268] Belotserkovskaya, R., Oh, S., Bondarenko, V. A., Orphanides, G., *et al.*, FACT facilitates transcription-dependent nucleosome alteration. *Science* 2003, *301*, 1090-1093.
[269] Berger, S. L., Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 2002, *12*, 142-148.
[270] Takai, D., Jones, P. A., Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* 2002, *99*, 3740-3745.
[271] Tate, P. H., Bird, A. P., Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* 1993, *3*, 226-231.
[272] Kouzarides, T., Histone methylation in transcriptional control. *Curr Opin Genet Dev* 2002, *12*, 198-209.
[273] Ingvarsdottir, K., Edwards, C., Lee, M. G., Lee, J. S., *et al.*, Histone H3 K4 demethylation during activation and attenuation of GAL1 transcription in Saccharomyces cerevisiae. *Mol Cell Biol* 2007, *27*, 7856-7864.
[274] Martin, C., Zhang, Y., The diverse functions of histone lysine methylation. *Nat Rev Mol Cell Biol* 2005, *6*, 838-849.
[275] Reinke, H., Gregory, P. D., Horz, W., A transient histone hyperacetylation signal marks nucleosomes for remodeling at the PHO8 promoter in vivo. *Mol Cell* 2001, *7*, 529-538.

[276] Gossett, A. J., Lieb, J. D., In vivo effects of histone H3 depletion on nucleosome occupancy and position in Saccharomyces cerevisiae. *PLoS Genet* 2012, *8*, e1002771.
[277] Jansen, A., Verstrepen, K. J., Nucleosome positioning in Saccharomyces cerevisiae. *Microbiol Mol Biol Rev* 2011, *75*, 301-320.
[278] Suganuma, T., Gutierrez, J. L., Li, B., Florens, L., *et al.*, ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. *Nat Struct Mol Biol* 2008, *15*, 364-372.
[279] Lomvardas, S., Thanos, D., Nucleosome sliding via TBP DNA binding in vivo. *Cell* 2001, *106*, 685-696.
[280] Fazzio, T. G., Tsukiyama, T., Chromatin remodeling in vivo: evidence for a nucleosome sliding mechanism. *Mol Cell* 2003, *12*, 1333-1340.
[281] Bowman, G. D., Mechanisms of ATP-dependent nucleosome sliding. *Curr Opin Struct Biol* 2010, *20*, 73-81.
[282] van Holde, K., Yager, T., Models for chromatin remodeling: a critical comparison. *Biochem Cell Biol* 2003, *81*, 169-172.
[283] Langst, G., Becker, P. B., Nucleosome remodeling: one mechanism, many phenomena? *Biochim Biophys Acta* 2004, *1677*, 58-63.
[284] Polach, K. J., Widom, J., A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J Mol Biol* 1996, *258*, 800-812.
[285] Ogawa, N., Saitoh, H., Miura, K., Magbanua, J. P., *et al.*, Structure and distribution of specific cis-elements for transcriptional regulation of PHO84 in Saccharomyces cerevisiae. *Mol Gen Genet* 1995, *249*, 406-416.
[286] Shen, C. H., Leblanc, B. P., Alfieri, J. A., Clark, D. J., Remodeling of yeast CUP1 chromatin involves activator-dependent repositioning of nucleosomes over the entire gene and flanking sequences. *Mol Cell Biol* 2001, *21*, 534-547.
[287] Fedor, M. J., Kornberg, R. D., Upstream activation sequence-dependent alteration of chromatin structure and transcription activation of the yeast GAL1-GAL10 genes. *Mol Cell Biol* 1989, *9*, 1721-1732.
[288] Rizzo, J. M., Bard, J. E., Buck, M. J., Standardized collection of MNase-seq experiments enables unbiased dataset comparisons. *BMC Mol Biol* 2012, *13*, 15.
[289] Gervais, A. L., Marques, M., Gaudreau, L., PCRTiler: automated design of tiled and specific PCR primer pairs. *Nucleic Acids Res* 2010, *38*, W308-312.
[290] Venter, U., Svaren, J., Schmitz, J., Schmid, A., Horz, W., A nucleosome precludes binding of the transcription factor Pho4 in vivo to a critical target site in the PHO5 promoter. *EMBO J* 1994, *13*, 4848-4855.
[291] Lam, F. H., Steger, D. J., O'Shea, E. K., Chromatin decouples promoter threshold from dynamic range. *Nature* 2008, *453*, 246-250.
[292] Segal, E., Widom, J., From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat Rev Genet* 2009, *10*, 443-456.
[293] Simon, M., Adam, G., Rapatz, W., Spevak, W., Ruis, H., The Saccharomyces-Cerevisiae Adr1 Gene Is a Positive Regulator of Transcription of Genes Encoding Peroxisomal Proteins. *Molecular and Cellular Biology* 1991, *11*, 699-704.

[294] Bellasio, M., Peymann, A., Steiger, M. G., Valli, M., *et al.*, Complete genome sequence and transcriptome regulation of the pentose utilizing yeast Sugiyamaella lignohabitans. *Fems Yeast Res* 2016, *16*.
[295] Drew, H. R., Travers, A. A., DNA bending and its relation to nucleosome positioning. *J Mol Biol* 1985, *186*, 773-790.
[296] Lowary, P. T., Widom, J., New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 1998, *276*, 19-42.
[297] Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I. K., *et al.*, Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* 2008, *4*, e1000216.
[298] Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., *et al.*, Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* 2010, *143*, 470-484.
[299] Report Linker 2017, p. 142.
[300] Monick, J. A., Fatty alcohols. *Journal of the American Oil Chemists' Society* 1979, *56*, 853A-860A.
[301] Deshpande, V. M., Ramnarayan, K., Narasimhan, C. S., Studies on ruthenium-tin boride catalysts II. Hydrogenation of fatty acid esters to fatty alcohols. *Journal of Catalysis* 1990, *121*, 174-182.
[302] Aro, E.-M., From first generation biofuels to advanced solar biofuels. *Ambio* 2016, *45*, 24-31.
[303] Akhtar, M. K., Dandapani, H., Thiel, K., Jones, P. R., Microbial production of 1-octanol: A naturally excreted biofuel with diesel-like properties. *Metab Eng Commun* 2015, *2*, 1-5.
[304] Liu, Y., Chen, S., Chen, J., Zhou, J., *et al.*, High production of fatty alcohols in Escherichia coli with fatty acid starvation. *Microb Cell Fact* 2016, *15*, 129.
[305] Fillet, S., Gibert, J., Suarez, B., Lara, A., *et al.*, Fatty alcohols production by oleaginous yeast. *J Ind Microbiol Biotechnol* 2015, *42*, 1463-1472.
[306] Hong, K. K., Nielsen, J., Metabolic engineering of Saccharomyces cerevisiae: a key cell factory platform for future biorefineries. *Cell Mol Life Sci* 2012, *69*, 2671-2690.
[307] Wang, W., Wei, H., Knoshaug, E., Van Wychen, S., *et al.*, Fatty alcohol production in Lipomyces starkeyi and Yarrowia lipolytica. *Biotechnol Biofuels* 2016, *9*, 227.
[308] Tang, X., Chen, W. N., Enhanced production of fatty alcohols by engineering the TAGs synthesis pathway in Saccharomyces cerevisiae. *Biotechnol Bioeng* 2015, *112*, 386-392.
[309] Zhou, Y. J., Buijs, N. A., Zhu, Z., Qin, J., *et al.*, Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat Commun* 2016, *7*, 11709.
[310] Teixeira, P. G., Ferreira, R., Zhou, Y. J., Siewers, V., Nielsen, J., Dynamic regulation of fatty acid pools for improved production of fatty alcohols in Saccharomyces cerevisiae. *Microb Cell Fact* 2017, *16*, 45.
[311] Yu, T., Zhou, Y. J., Wenning, L., Liu, Q., *et al.*, Metabolic engineering of Saccharomyces cerevisiae for production of very long chain fatty acid-derived chemicals. *Nat Commun* 2017, *8*, 15587.

[312] *Fatty Alcohols Market Analysis By Product, Application To 2022*, Grand View Research, http://www.grandviewresearch.com 2016.
[313] Taylor, J., Popat, N., Lin, Y. P., ICIS Pricing 2014.
[314] American Sugar Alliance, https://sugaralliance.org 2017.
[315] Sheng, J., Stevens, J., Feng, X., Pathway Compartmentalization in Peroxisome of Saccharomyces cerevisiae to Produce Versatile Medium Chain Fatty Alcohols. *Sci Rep* 2016, *6*, 26884.
[316] Mlickova, K., Roux, E., Athenstaedt, K., d'Andrea, S., *et al.*, Lipid accumulation, lipid body formation, and acyl coenzyme A oxidases of the yeast Yarrowia lipolytica. *Appl Environ Microbiol* 2004, *70*, 3918-3924.
[317] Tenagy, Park, J. S., Iwama, R., Kobayashi, S., *et al.*, Involvement of acyl-CoA synthetase genes in n-alkane assimilation and fatty acid utilization in yeast Yarrowia lipolytica. *Fems Yeast Res* 2015, *15*, fov031.
[318] Rutter, C. D., Zhang, S., Rao, C. V., Engineering Yarrowia lipolytica for production of medium-chain fatty acids. *Appl Microbiol Biotechnol* 2015, *99*, 7359-7368.
[319] al-Habori, M., Microcompartmentation, metabolic channelling and carbohydrate metabolism. *Int J Biochem Cell Biol* 1995, *27*, 123-132.
[320] Feng, X., Lian, J., Zhao, H., Metabolic engineering of Saccharomyces cerevisiae to improve 1-hexadecanol production. *Metab Eng* 2015, *27*, 10-19.
[321] Wang, W., Wei, H., Knoshaug, E., Van Wychen, S., *et al.*, Fatty alcohol production in Lipomyces starkeyi and Yarrowia lipolytica. *Biotechnol Biofuels* 2016, *9*.
[322] Youngquist, J. T., Schumacher, M. H., Rose, J. P., Raines, T. C., *et al.*, Production of medium chain length fatty alcohols from glucose in Escherichia coli. *Metab Eng* 2013, *20*, 177-186.
[323] Petrie, J. R., Shrestha, P., Mansour, M. P., Nichols, P. D., *et al.*, Metabolic engineering of omega-3 long-chain polyunsaturated fatty acids in plants using an acyl-CoA Delta6-desaturase with omega3-preference from the marine microalga Micromonas pusilla. *Metab Eng* 2010, *12*, 233-240.
[324] Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., *et al.*, Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc Natl Acad Sci U S A* 2000, *97*, 6487-6492.
[325] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012, *9*, 676-682.
[326] Folch, J., Lees, M., Sloane Stanley, G. H., A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem* 1957, *226*, 497-509.
[327] Willis, R. M., Wahlen, B. D., Seefeldt, L. C., Barney, B. M., Characterization of a fatty acyl-CoA reductase from Marinobacter aquaeolei VT8: a bacterial enzyme catalyzing the reduction of fatty acyl-CoA to fatty alcohol. *Biochemistry* 2011, *50*, 10550-10558.
[328] Rakicka, M., Lazar, Z., Dulermo, T., Fickers, P., Nicaud, J. M., Lipid production by the oleaginous yeast Yarrowia lipolytica using industrial by-products under different culture conditions. *Biotechnol Biofuels* 2015, *8*, 104.

[329] Lazar, Z., Dulermo, T., Neuveglise, C., Crutz-Le Coq, A. M., Nicaud, J. M., Hexokinase--A limiting factor in lipid production from fructose in Yarrowia lipolytica. *Metab Eng* 2014, *26*, 89-99.
[330] Dulermo, R., Gamboa-Melendez, H., Ledesma-Amaro, R., Thevenieau, F., Nicaud, J. M., Yarrowia lipolytica AAL genes are involved in peroxisomal fatty acid activation. *Biochim Biophys Acta* 2016, *1861*, 555-565.
[331] Ageitos, J. M., Vallejo, J. A., Veiga-Crespo, P., Villa, T. G., Oily yeasts as oleaginous cell factories. *Appl Microbiol Biot* 2011, *90*, 1219-1227.
[332] Liu, A., Tan, X., Yao, L., Lu, X., Fatty alcohol production in engineered E. coli expressing Marinobacter fatty acyl-CoA reductases. *Appl Microbiol Biotechnol* 2013, *97*, 7061-7071.
[333] Yao, L., Qi, F., Tan, X., Lu, X., Improved production of fatty alcohols in cyanobacteria by metabolic engineering. *Biotechnol Biofuels* 2014, *7*, 94.
[334] Soltani, M., Metzger, P., Largeau, C., Effects of hydrocarbon structure on fatty acid, fatty alcohol, and beta-hydroxy acid composition in the hydrocarbon-degrading bacterium Marinobacter hydrocarbonoclasticus. *Lipids* 2004, *39*, 491-505.
[335] Wenning, L., Yu, T., David, F., Nielsen, J., Siewers, V., Establishing very long-chain fatty alcohol and wax ester biosynthesis in Saccharomyces cerevisiae. *Biotechnol Bioeng* 2017, *114*, 1025-1035.
[336] Avalos, J. L., Fink, G. R., Stephanopoulos, G., Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat Biotechnol* 2013, *31*, 335-341.
[337] Smale, S. T., Kadonaga, J. T., The RNA polymerase II core promoter. *Annu Rev Biochem* 2003, *72*, 449-479.
[338] Rhee, H. S., Pugh, B. F., Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 2012, *483*, 295-301.
[339] Kuehner, J. N., Brow, D. A., Quantitative analysis of in vivo initiator selection by yeast RNA polymerase II supports a scanning model. *J Biol Chem* 2006, *281*, 14119-14128.
[340] Kornberg, R. D., The eukaryotic gene transcription machinery. *Biol Chem* 2001, *382*, 1103-1107.
[341] Silverman, A. M., Qiao, K., Xu, P., Stephanopoulos, G., Functional overexpression and characterization of lipogenesis-related genes in the oleaginous yeast Yarrowia lipolytica. *Appl Microbiol Biotechnol* 2016, *100*, 3781-3798.
[342] Mieczkowski, J., Cook, A., Bowman, S. K., Mueller, B., *et al.*, MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun* 2016, *7*, 11485.
[343] Xi, Y., Yao, J., Chen, R., Li, W., He, X., Nucleosome fragility reveals novel functional states of chromatin and poises genes for activation. *Genome Res* 2011, *21*, 718-724.
[344] Wei, G., Hu, G., Cui, K., Zhao, K., Genome-wide mapping of nucleosome occupancy, histone modifications, and gene expression using next-generation sequencing technology. *Methods Enzymol* 2012, *513*, 297-313.

[345] Newman, Z. R., Young, J. M., Ingolia, N. T., Barton, G. M., Differences in codon bias and GC content contribute to the balanced expression of TLR7 and TLR9. *Proc Natl Acad Sci U S A* 2016, *113*, E1362-1371.
[346] Dell'Angelica, E. C., Ermacora, M. R., Santome, J. A., Purification and partial characterization of a fatty acid-binding protein from the yeast, Yarrowia lipolytica. *Biochem Mol Biol Int* 1996, *39*, 439-445.