DISCLAIMER:

This document does not meet the
current format guidelines of
the Graduate School at
The University of Texas at Austin.

It has been published for
informational use only.

The Dissertation Committee for Heidi Rosemary Redden Certifies that this is the approved version of the following dissertation:

Developing synthetic, minimal promoters in *Saccharomyces cerevisiae*

Committee:

Karen Browning, Supervisor

Jeffrey Barrick

David Hoffman

Adrian Keatinge-Clay

Developing synthetic, minimal promoters in *Saccharomyces cerevisiae*

by

Heidi Rosemary Redden, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Dedication

To Lyle Lawrence Peña and future generations.

“The scientist is not a person who gives the right answers, he's one who asks the right questions.”

— Claude Lévi-Strauss

Acknowledgements

I would like to first acknowledge my committee members, Karen Browning, Jeffrey Barrick, David Hoffman, and Adrian Keatinge-Clay for taking the time to serve on my committee to provide me valuable feedback. I thank Hal Alper for his financial support from summer 2014 to fall 2016 made possible by a grant provided by the Air Force Office of Scientific Research. I also thank Alper for the use of his laboratory equipment, space and supplies for my experiments. Further funding was provided by a Dissertation Writing Fellowship for Spring 2016 and Parental Accommodation from CNS Parental Assistance funds for Fall 2016 and Spring 2017.

Next, I would like to thank all the members of the Alper lab, past and present for their help, and support over the years. I would like to first acknowledge the individuals who contributed directly to my work. John Blazeck and Leqian Liu allowed me to assist them with their *Yarrowia lipolytica* studies that I present in Chapter 2, and Matthew Deaner designed and carried out the experiments for the dCas9 promoter core amplifier presented in Chapter 4. While all of the Alper lab provided wonderful feedback and suggestions on my experiments, there are several members in particular I'd like to thank for their wisdom. Leqian Liu provided his knowledge in *E. coli* preparation, and John Leavitt gave me advice on undergraduate management. Furthermore, there are several members that made the monotony of science enjoyable. I would like to thank Nicholas Morse, Kelly Markham, and James Wagner for our distracting conversations about science over long hours of pipetting. Aside from the graduate students in our lab, I would

like to acknowledge several undergraduates. Rachel Woodard, Josh Damkroger, Rosemary Burke, and Monica Duque took the time out of their busy schedules to assist me. These individuals made the execution of large experiments possible. In particular, Rachel Woodard and Josh Damkroger spent hours under an obnoxious blue light picking hundreds of colonies for the experiments described in Chapter 5. Beyond members of the Alper lab, I would like to thank Richard Salinas for the many hours helping me sort all my promoter libraries, and for keeping the flow cytometers clean and working.

Finally, I would like acknowledge family and friends that one way or another made all this possible. I thank Dacia Leon for the many scientific arguments we have had over tacos, and Drew Tack for commiserating with me over failed experiments. Lastly, I thank my beautiful and lovely husband, Matt Peña for his encouragement and support. I would not be here if it were not for him.

Developing synthetic, minimal promoters in *Saccharomyces cerevisiae*

Heidi Rosemary Redden, PhD

The University of Texas at Austin, 2017

Supervisor: Karen Browning

Promoters enable synthetic biologists to manipulate protein expression at the DNA level. For this reason, promoters are essential for almost all applications aiming to engineer an organism. Unfortunately, promoters available for eukaryotic organisms are derived directly from the genome. Such promoters are large and result in substantial and therefore, difficult DNA insertions to express a heterologous multi-gene pathway. Furthermore, their high sequence homology provides the organism the opportunity to perform homologous recombination resulting in undesirable gene deletions. For these reasons, there is a critical demand for short promoters with low sequence homology to the organism's genome to continue synthetic biology advancements in eukaryotic hosts.

This work addresses the need for yeast promoters by engineering the promoter's two distinct DNA regions– the upstream activating sequence (UAS) and the downstream 3` area comprised of the promoter's core. The modularity of these regions is demonstrated in a non-conventional yeast, *Yarrowia lipolytica* by assembling multiple

native UAS in tandem with a core. In doing so, the strongest promoters ever reported for *Y. lipolytica* were created. Drawing from these lessons, the length of promoters in the popular host strain *Saccharomyces cerevisiae* was minimized. The core region is first addressed by establishing promoter libraries with minimized *de novo* cores. Synthetic cores are isolated from a short promoter library and are evaluated in six DNA contexts to establish nine minimal cores with modularity, robustness, and context independence. Second, the UAS region was minimized. To do so, a randomized region of DNA was hybridized upstream of a synthetic minimal core to construct 18 *de novo* libraries of promoters. From these libraries, 26 short constitutive and inducible UAS elements were isolated.

Collectively, this work highlights the utility of hybrid promoter engineering to increase the number of promoters available for host organisms, *Y. lipolytica* and *S. cerevisiae*. More importantly, it establishes a highly desirable set of 81 synthetic, minimal promoters of inducible and constitutive function that provides a 70-fold range of expression in *S. cerevisiae*. Furthermore, the workflow presented herein is generic enough for application in other eukaryotic host organisms to build their synthetic biology toolboxes.

Table of Contents

List of Abbreviations

UAS, upstream activating sequence

RNAP, RNA polymerase

ORF, open reading frame

TFBS, transcription factor binding site

TSS, transcription start site

PIC, preinitiation complex

TF, transcription factor

UTR, untranslated region

TBP, TATA-box binding protein

GTF, general transcription factors

SAGA, spt3-ada-gen5-acetyltransferase

FACS, fluorescence assisted cell sorting

sTF, synthetic transcription factor

IPTG, β-D-1-thiogalactopyranoside

Leum, leucine minimal

GBS, Gal4p binding site

sgRNA, single guide RNA

VPR, VP64-p65-rta

iCAPTR, isolation of conditionally activated parts for transcription regulation

List of Tables

List of Figures

Chapter 1: Introduction

For more than a millennium, *Saccharomyces cerevisiae* has been selectively cultivated and exploited by humankind for its robust growth, love of simple sugars, and production of valuable chemicals. Several favorable production traits make *S. cereivisiae* an attractive host. *S. cerevisiae*'s large cell size allows for easy recovery of desirable products. Compared to the bacterial host, *Escherichia coli,* yeast has a higher tolerance to by-products and acids, and lower temperature growth requirements[1]. Due to these traits, *S. cerevisiae* now serves as a key production platform for biofuels, nutraceuticals, industrial compounds, and therapeutic proteins[2,3]. To accomplish this, a broad and expanding array of genetic tools were developed to study and engineer yeast. Many of these tools are a direct result of the efforts put forth in synthetic biology.

The overarching goal of synthetic biology is to gain a better understanding of how to coordinate and regulate pathways and gene expression in organisms. In particular, the engineering of synthetic systems involves practical and rational design of disparate parts that all must work together to obtain a desired function. Biological hosts, such as *S. cerevisiae* and *E. coli* have distinct DNA-based tools (such as promoters, terminators, vector plasmids, genome integration sites) which can be utilized to achieve a desired cellular regulation. Each of these tools contributes to the overall synthetic biology toolbox of an organism.

The promoter, in particular is an indispensable synthetic biology part. A promoter serves to drive transcription of every gene in an organism. Unfortunately, the promoters

available for the eukaryotic host, *S. cerevisiae* are endogenous, lengthy, and dependent on native scaffolds. This is particularly problematic for eukaryotes because these hosts do not use polycistronic messages like their bacterial counterparts, and therefore, require a separate promoter for the expression of each gene. Compared to *E. coli* promoters of ~100 base pairs, the traditionally-used yeast promoter *GPD* (*TDH3*) is nearly seven times longer (**Figure 1.1**). Thus, to express a single gene of 1.5 Kb, an additional 1 Kb of regulatory DNA (between the promoter and terminator) is required increasing the DNA cargo load by over 60%. If an extensive, multi-gene heterologous pathway is required, this regulatory DNA could easily add up to tens of thousands of nucleotides, nearly ten times as much as needed in bacteria. Furthermore, numerous promoters would be required to express such a pathway. Given yeast's ability to perform homologous recombination, each of these promoters must be sequence unique. Therefore, to reduce the amount of regulatory DNA needed for expression and avoid homologous recombination, there is a critical need for diverse and minimal promoter sequences that do not resemble the host's genomic sequences.

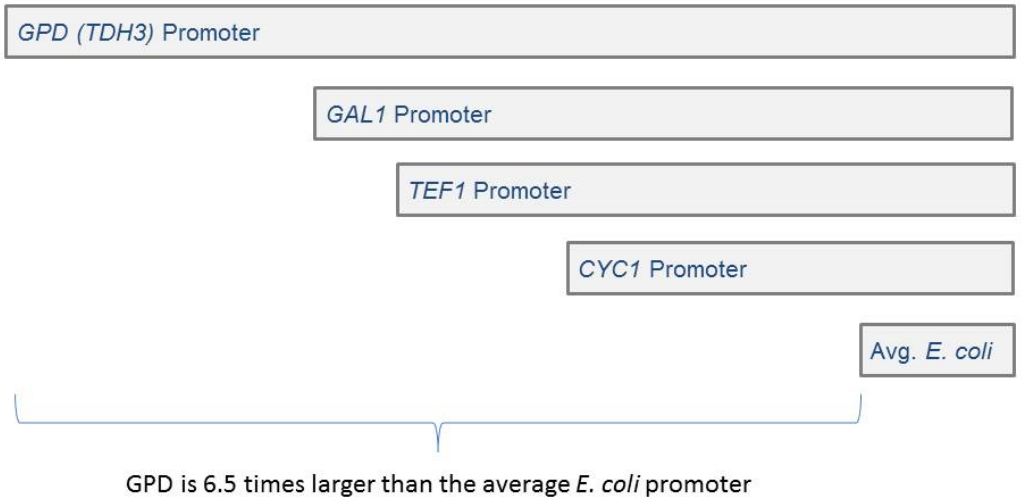


Figure 1.1: Length of commonly used promoters in metabolic engineering

Shown here to scale by DNA length, the commonly-used high strength *GPD* promoter is 6.5 times larger than the average *E. coli* promoter of 100 bp.

To date, this problem of bulky yeast promoters has been scarcely addressed in synthetic biology. Yeast promoter engineering has mainly focused on building promoters from large endogenous segments and native promoter scaffolds. Thus, the work in this dissertation seeks to resolve two main problems with current promoters available for *S. cerevisiae*: i) high sequence homology, and ii) overall large size. To do so, I exploit the current available biological insights on the native promoter's structure and mechanism. The next sections will discuss DNA architecture, critical mechanistic steps, and regulatory mechanisms of the eukaryotic promoter that must be considered in this work.

1.1 DNA ARCHITECTURE OF THE PROMOTER

Promoters are categorized by the type of RNA polymerase (RNAP) recruited to transcribe its nearby gene. There are three RNAPs: RNAPI, II and III. RNAPI and RNAPIII transcribe ribosome essential RNAs, rRNAs and tRNAs, while RNAPII transcribes mRNA. This work will focus only on RNAPII promoters as they are responsible for driving transcription of protein-coding genes. Promoters are identified as the region upstream of the first start codon of an open reading frame (ORF). While promoter elements deep within ORFs have been identified[4,5], these instances are extremely rare. Therefore, I will limit our discussion to promoters upstream of the first start codon. Architecturally, these promoters are comprised of an upstream activating sequence (UAS), and a promoter core (**Figure 1.2**)[6]. A core serves as a platform for RNAP binding, while an UAS provides additional stability and regulation to the RNAP. Distinct DNA sequences that characterize these regions in native promoters of *S. cerevisiae* will be discussed in detail in the following sections.

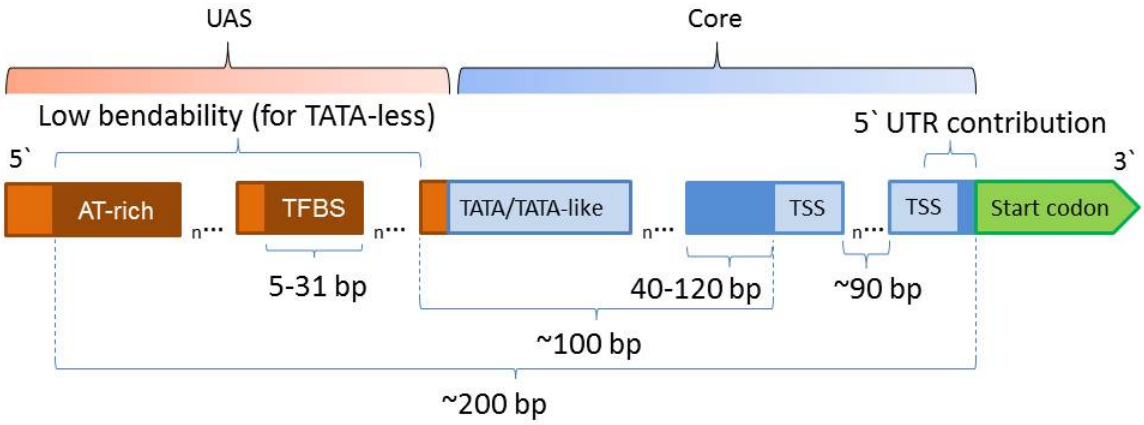


Figure 1.2: DNA architecture of a yeast genomic promoter

The promoter is comprised of two components, an UAS (orange) and a core (blue). The UAS may contain one or more short AT-rich tandem repeats, and one or more transcription factor binding site (TFBS). The core contains one or more TATA or TATA-like boxes and one or more transcription start site (TSS). The 3` end of the TSS contributes to the 5` UTR of the mRNA.

1.1.1 Necessary components of the promoter core

The core encompasses the 3` region of the promoter (**Figure 1.2**). The core performs two critical roles for transcription initiation: serves as the platform for the assembly of the pre-initiation complex (PIC), a complex of RNAPII and associating transcription factors (TF), and signals transcription to RNAPII. The core uses two essential DNA sites to carry out transcription initiation: a single AT-rich region (< 2% of native promoters contain two or more[7]) and at least one transcription start site (TSS) (**Figure 1.2**). The AT-rich region is traditionally referred to as the TATA-box and is located on the 5` end of the promoter core region. Although the consensus sequence is defined as TATAWAR[7] (where W represents A/T, and R represents G/C), about 80% of yeast promoters do not have this sequence[7]. In the literature, these promoters are termed

TATA-less, but more recently and appropriately, are referred to as TATA-like for harboring an AT-rich region with close resemblance (up to 2 mismatches) to the TATAWAWR sequence[8]. The presence of a bona fide TATA box in a promoter is noteworthy; genes driven by TATA promoters are distinct from those driven by TATA-like promoters in terms of transcriptional regulation, codon usage, network interaction, associated enzymatic pathways[9], and PIC assembly[7]. For example, a higher frequency of TATA boxes occur in promoters of genes with higher expression values[10] and under higher cellular regulation[7]. TATA-boxes are enriched in promoters that drive expression of genes related to heat stress, diauxic shift and sporulation[7,9]. Gene products from TATA promoters have a lower number of gene interactions, and their mRNAs are more efficiently translated due to a higher codon adaptation index[9]. Lastly, a bona fide TATA box confers directionality to a promoter. TATA-like promoters have a higher chance of driving expression in a bidirectional manner, and consequently, giving rise to non-coding RNAs[5]. Thus, the presence of a TATA box plays a substantial role for the promoter.

The other essential component of the core, the TSS, is located 40 to 120 bp downstream of the TATA or TATA-like box[11,12]. The TSS signals to the bound RNAPII to begin transcription (**Figure 1.2**). Unlike the TATA box, this sequence emerges more than once in a single yeast promoter[4]. The existence of multiple TSS in more than 99% of promoters has complicated efforts to identify any TSS consensus sequence[4,5,12]. Numerous sequences have been suggested including RRYRR, TCRA[13], YAWR[14] and most recently, $A(A_{rich})_5NYAWNN(A_{rich})_6$[12]. Currently, to accurately identify the site(s),

the 5` untranslated region (UTR) of the mRNA must be mapped experimentally[5]. Based on experimentally confirmed TSS, native promoters can be classified into two groups based on the usage of their TSS. One class utilizes a single dominant TSS, while the others' use is evenly distributed among the sites[4]. The use of multiple TSS allows the cell to control protein abundance at the transcript level; TSS' contribution to the 5` UTR affects mRNA translation efficiency and subsequently, protein abundance[15]. In fact, a study of 96 genes revealed that translation activity can be altered more than 100-fold depending on the 5` UTR isoform present in the mRNA[15]. Which TSS will be utilized during transcription is largely unknown, but a recent study suggests the distance between a TSS and the TATA box drives selection[16]. In this study, characterization of 7,536 TATA-containing synthetic core promoter variants revealed a usage bias for the TSS 88 to 66 bp downstream of the TATA box[16]. Thus, at least in this context, there appears to be an optimal arrangement for the core promoter's two essential elements, the TATA box and the TSS.

The core region provides an indispensable role for the promoter; a single contact point for PIC assembly, and transcription signals. However, by itself, the promoter core often drives weak expression. A much larger expanse of DNA found upstream of the core region provides additional regulation and strength. In the next section, I discuss the DNA features of this region in the genomic context. Although this region can also repress transcription, for sake of simplicity I will solely refer to it as the upstream *activating* sequence.

1.1.2 DNA features of Upstream Activating Sequences (UAS)

In the simplest sense, the UAS coordinates transcription of the downstream gene by influencing PIC assembly, and in doing so, affects promoter activity. In this section, the following DNA features found in native UAS regions are discussed: i) quantity and positioning of transcription factor binding sites (TFBS), ii) short tandem repeats, and iii) unique biophysical properties such as bendability and secondary structures (**Figure 1.2**).

Arguably, the most significant sequence element in an UAS region is the TFBS. TFBSs are short sequences varying in length from 5 to 31 nucleotides with a mean of 9.9[17] (**Figure 1.2**). TFBSs recruit TFs that help regulate PIC formation either through direct protein-protein interaction(s) or through histone modifications to alter nucleosome placement[18]. Most promoters contain one or two binding sites[19], with TATA promoters containing a lower number[18]. TFBS density peaks about ~100 bp upstream of the TSS[18] (**Figure 1.2**) with most restricted to the region 100 and 500 bp upstream of the start codon[20]. A recent study of various *S. cerevisiae* strains revealed that TFBSs found within 200 bp of the TSS were more evolutionary conserved than those at further distances[21]. This result taken together with the fact that very few TFBSs are found within 100 bp of the start codon[20] suggests a critical functional zone for TFs in the promoter. Further studies indicate that the placement and length of this zone is altered depending on the presence of a bona fide TATA box. Compared to TATA-like promoters, this TFBS-enriched zone is longer and found further upstream in TATA promoters[18]. A study suggests that the different arrangement of TFBS in TATA and TATA-like promoters is

due to how these promoter classes are regulated by nucleosomes[18]. I discuss this and other nucleosome regulation of promoters further in detail below.

TFBS are the not the only short regulatory sequences that appear in promoters. At least one AT rich tandem repeat of 2 to 6 nucleotides is present in 25% of all yeast promoters. Most of these repeats are located ~200 bp upstream of the start codon. They are distinct from known TFBS, but like TFBS, are inversely correlated with nucleosome density. Unlike TFBS, these repetitive sequences are difficult for the cell to replicate and consequently, encourage low fidelity. A recent study suggests this low fidelity may explain increased promoter evolution observed for promoters enriched with tandem repeats[22].

Aside from harboring short regulatory sequences discussed thus far, the UAS also contributes unique biophysical properties to the promoter such as bendability and secondary structure formation. Bendability is defined as the propensity of a triplet of base pairs to bend and is assessed by the increased ability of DNase I to cleave bendable DNA[23]. For promoters, this propensity to bend has been suggested to influence TF binding, nucleosome positions, and DNA looping. Regions of low bendability are unique to TATA-like promoters and are greatly enriched 100 to 200 bp upstream of the start codon, with peak rigidity at ~130 bp[24]. The increase in rigidity is thought to be necessary to enable the first step of PIC assembly, the binding of TATA-box binding protein (TBP) to the TATA-like sequence[24]. In addition, DNA in the promoter region is capable of forming secondary structures. Stabilized by Hoogsteen base pairing, four guanine tracts

effectively spaced can fold into a complex secondary structure called a G-quadruplex. These structures are highly enriched in genomic promoters and telomeres[25] and are well conserved across yeast species[26]. Specifically, G-quadruplexes are found within 500 bp of the start codon for non-essential genes[26]. The alternative DNA structure, Z-DNA, also seems to play a role in promoter function. Although mostly studied in human cell lines, Z-DNA has been demonstrated to influence promoter function in yeast[27]. Z-DNA can absorb the negative supercoils produced by nucleosome removal[27], DNA unwinding during transcription and chromatin remodeling[28]. The exact function and mechanism of these secondary structures in the promoter is still relatively unknown.

In summary, promoters contain unique DNA sequences that provide crucial contact points for binding factors to assemble. In doing so, they influence the formation of PIC, and subsequently, the transcription of the downstream gene. The next section discusses PIC assembly with a focus on the formation of the most critical step: the binding of the TBP to the core region.

1.2 First step in assembly of Preinitiation Complex (PIC) at the Promoter Core

RNAPII uses a set of general transcription factors (GTF) that function in promoter recognition, interaction with TF, DNA unwinding and TSS recognition[29-31]. These GTFs include TFIIA, TBP, TFIIB, TFIIF, TFIIE, and TFIIH. Before transcription occurs, RNAPII is positioned and stabilized onto the promoter DNA by GTFs to form PIC. This occurs in a stepwise fashion in which each GTF plays either an indispensable or

supplementary role. The first step in PIC assembly involves TBP recognition of the TATA-box (or TATA-like sequence), and is of particular importance as this step sets the stage for transcriptional strength[32,33]. For this reason, I will focus on this particular protein and the implications of binding to the promoter core.

TBP plays an indispensable, well-conserved and universal role in transcription initiation. In fact, TBP along with TFIIB can sufficiently transcribe certain genes[34]. The fact that archaebacteria also only require these two TFs[35,36] highlights TBP's longstanding evolutionary roots in transcription. Indeed, of all proteins involved in transcription initiation, TBP is the most conserved across a wide variety of species, showing a greater than 80% sequence identity[37]. Furthermore, TBP is required not just by RNAPII, but by all three eukaryotic RNA polymerases, making this protein a universal TF for all yeast promoters[38].

TBP's critical role is to bind to the TATA-box in the core promoter and induce a dramatic ~90 degree bend in the DNA. This bend allows the promoter DNA to eventually wrap around RNAPII for transcription initiation[39-41]. The affinity of TBP for the TATA box depends on the TATA box variant present and the site's flanking sequences[39]. Bona fide TATA boxes confer the highest affinity, while deviations from the TATAWAWR consensus sequence result in weaker binding. The strength of affinity is thought to determine which co-activator complex directs TBP to the TATA box, and consequently, the rate of transcription[33]. For TATA-containing promoters, the Spt-Ada-Gcn5-acetyltransferase (SAGA) complex directs TBP. However, TATA-like promoters,

TBP binds to the site as part of the TFIID complex[7,8]. These two co-activators are associated with specific gene expression profiles. Studies have revealed a relationship between SAGA regulated promoters and stress responsive genes[42]. In contrast, TFIID is involved in promoters that drive transcription of house-keeping genes[42]. Thus, the aforementioned significance of the TATA box or TATA like sequence in the promoter appears to function through the recruitment of the coactivator complex via the first step in transcription initiation, TBP binding.

In addition to their roles in TBP-TATA-box interactions, both TFIID and SAGA participate in chromatin remodeling[43]. In fact, SAGA has been suggested to influence nearly all RNAPII expressed genes through genome-wide histone modifications[44]. Indeed, nucleosome dynamics at the promoter level can have profound effects on the promoter's activity. In the next section, I discuss how nucleosomes serve to regulate gene expression by altering promoter strength.

1.3 Nucleosome Function in Promoter Regions

A nucleosome is a 147 bp tract of DNA wrapped around a histone octamer[45], and serves to compact DNA into chromatin. At any given time, nucleosomes sequester 80% of yeast's genome in their protein bundles[46]. Nucleosome positioning appears to be largely determined by DNA sequence[47]. In the yeast genome, most promoters[49], and to a lesser extent, terminators[48] are nucleosome-depleted. For promoters, this observed nucleosome-depletion pattern appears to be a fundamental feature of eukaryotic transcription[49], and evidence suggests nucleosome-depletion plays a role in promoter

strength. An analysis of nucleosome occupancy of genomic promoters showed an inverse correlation with promoter strength. In this analysis, 42% of nucleosome-depleted promoters regulated highly active genes[50]. Nucleosomes are thought to regulate transcription by physically blocking TF access to their binding sites thereby preventing activity at the promoter. Some studies have shown that TF access to binding sites can increase ~10 to 20-fold when the promoter is artificially depleted of nucleosomes[51,52]. A small number of TF, about 10 to 20, can actually compete with nucleosomes. In fact, certain TFBS, notably the Rap1p binding site, are more strongly correlated with nucleosome-depleted promoters[53]. Interestingly, these TFs have been observed to participate in protein-protein interactions with chromatin remodeling complexes, histones, and chromatin modification enzymes[54]. This suggests these TF are able to clear the promoter of nucleosomes by modifying their histone proteins.

Nucleosome positioning is not static, but rather highly dynamic. SAGA regulated promoters in particular tend to be occupied by nucleosomes at any given moment[55]. A genomic analysis of induced genes through the use of transcriptional activator Gcn4 revealed inconsistent phasing for SAGA-regulated promoters, which suggests these promoters are regulated by nucleosome repositioning[56]. Indeed, individual studies on stress-responsive promoters like *PHO5*[57,58] and *GAL1-10*[59] reveal that nucleosome positioning plays a critical role in activation of these promoters. In any case, studies have well established that nucleosomes are able to prevent transcription either transiently, as is the case with stress-induced genes, or more permanently, as that observed in chromosome

silencing, by sterically occluding access of binding factors to critical contact points in the DNA.

These studies and many others reveal the complicated nature of endogenous promoters. Tailored by evolution to function with a specific ORF, and in a particular genomic location, an endogenous promoter tends to be highly and unpredictably sensitive to DNA context. By utilizing the insights on promoter architecture and function gathered from molecular biology studies discussed above, numerous successful promoter engineering strategies have been developed to address the need for more orthologous expression.

1.4 PROMOTER ENGINEERING IN *SACCHAROMYCES CEREVISIAE*

Initially, promoters for synthetic biology purposes were isolated directly from the genome, mostly from the loci of the glycolytic pathway[60]. Decades later, these promoters are still used in many synthetic biology and metabolic engineering efforts[61-63]. Unfortunately, as discussed previously, endogenous promoters are inherently highly regulated by the cell. Thus, interest in the last decade has shifted to expand yeast's toolbox into the synthetic realm. In literature, the term "synthetic" is applied to a broad range of promoter sequences, but generally refers to promoters that are not directly isolated from the genome. This loose definition includes promoters highly homologous to the genome like those generated from PCR mutagenesis of native sequences. On the other extreme, this definition also includes promoters crafted completely *de novo*, which bear no resemblance to any sequences in the yeast genome. Thus, the degree of

artificiality of a proclaimed "synthetic" promoter can vary widely. In our discussion of synthetic promoters, I will refer to any promoter not amplified directly from the genome of any organism as "synthetic". With this definition, promoters have been produced through libraries, rationally designed based on rules established from studies on native promoters, and a combination of the two. These strategies have been successful in engineering constitutive promoters, those that continually drive expression, and inducible promoters, those that drive expression only under environments that deviate from established experimental yeast conditions. In this section I will review the application of these strategies to engineer both constitutive and inducible promoters in *S. cerevisiae*.

1.4.1 Promoter library construction

Promoter libraries are attractive for their ability to produce a wide diversity of variants which can be screened for desirable traits. Promoter libraries have been i) produced through mutagenesis techniques[63-67], ii) compiled from numerous synthesized sequences[16,68-70], and iii) generated by complete randomization of distinct promoter regions. In the application of mutagenesis techniques, constitutive endogenous promoter *TEF1* was mutagenized with error-prone PCR to produce a library of nearly 200 promoters in which 11 mutants were isolated with expression strengths ranging from 8% to 120% of the native[13,66]. However, mutagenesis generates sets of promoters with high sequence homology, which limits their use in eukaryotic organisms due to the ability of these hosts to perform homologous recombination. To circumvent high sequence homology, libraries can be constructed from a set of synthesized known sequences[16,68-70].

Most recent applications of this approach used microarrays to synthesize libraries of several thousand variants[16,69]. In these experiments, transcriptional effects of the quantity, quality and position of selected TFBS, nucleosome depleting sequences and core sequences were explored to elucidate general rules about promoter architecture[16,69]. However, producing libraries in this manner requires extensive design and data mining to create functional variants. With limited promoter architecture knowledge, completely randomizing certain promoter regions has proven succesful[71,72]. In one example, a 48 bp region of the *PFY1* promoter core was randomized to yield library members ranging from 91% to 11% of the native scaffold promoter[72]. This last approach can allow for generation of larger libraries and requires limited mechanistic understanding.

Any large promoter library regardless of the production method utilized requires selection. The most common selection involves the use of a fluorescent reporter and an *in vivo* sort by Fluorescence Assisted Cell Sorting (FACS)[69]. Unfortunately, library production and the subsequent selection process can add significant time and effort to a project that requires promoter(s) with specific traits. In certain scenarios, a rationally designed synthetic promoter is more appropriate.

1.4.2 Rational design of synthetic promoters

A rational design approach is often favored for the construction of inducible promoters due to the ability to manipulate the promoter's architecture, and avoid a selection process like that required for libraries. In rational designs, synthetic promoters

are stitched together using native and/or synthetic promoter elements such as TFBS[73-77], UAS[78-81], nucleosome disfavoring sequences [51,80,82,83], and promoter cores[76] (**Figure 1.3**).

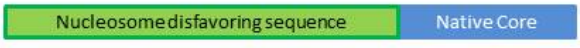
Promoter architecture	Reference
	78-81, this work
	84,85
	72,77,86-91
	68,69,71
	this work
	76
	74
	73
	70
	16,72, this work
	27,51,80,83

Figure 1.3: Promoter engineering strategies

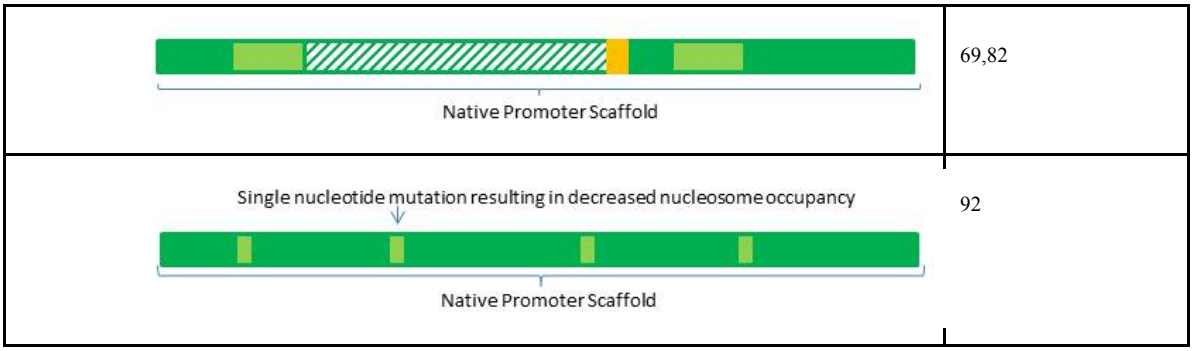


Figure 1.3: Promoter engineering strategies continued

A number of strategies have been used for promoter engineering. In these efforts, the promoter is often engineered as two distinct parts, the UAS region (orange) and the core region (blue). In some cases, an entire native promoter scaffold is utilized (green). TFBS (yellow) can be placed within the UAS region to activate the core, or in the core region to repress transcription. Nucleosome disfavoring sequences (light green) are used in the promoter region to increase transcript output. Regions of promoter constructs made with libraries are indicated by stripes. All promoters are unidirectional (towards the right in these depictions), except where indicated by two arrows.

TFBS represent the shortest promoter element, and are a popular choice as an activating sequence. One or more TFBS can be combined with promoter core(s) to craft full length functioning promoters. A TFBS may solicit the function of an endogenous TF[93], an imported TF[74,86], or a synthetic TF (sTF)[72,73,76,87,88,91] to impart an inducible or constitutive quality to the promoter. In most applications, TFBS are just paired with a single promoter core. However, recently two divergent core elements were used to create a bidirectional constitutive promoter[76]. In this case and many others, TFBS are positioned upstream of a promoter core to enhance transcription (**Figure 1.3**). However, TFBS can be placed within the core region to repress transcription[70,73,91] (**Figure 1.3**). For example, multiple core-positioned TFBS, activated by tetracycline-induced TFs were used to repress transcription of a synthetic inducible promoter[70]. In another effort, two sTF were imported to impart control on a synthetic promoter comprised of five upstream TFBS and one core-positioned TFBS. In doing so, an 8-fold induction under the presence of isopropyl β-D-1-thiogalactopyranoside (IPTG) and testosterone was achieved[73]. The fact that the position, not the identity of the TFBS in these examples dramatically affects the promoter activity highlights the importance of TFBS placement within the promoter. In most engineered promoters utilizing TFBS, the site(s) are blindly placed in close proximity to the core promoter. However, doing so risks steric occlusion of binding TF. To avoid this, native promoter scaffolds can be used as guide during construction. For example, a native pH inducible promoter was enhanced 5-fold by swapping out some TFBS with ones known to induce transcription in low pH

conditions[84] (**Figure 1.3**). While these successes and many others demonstrate the use of TFBS in promoter engineering, limitations do exist. The implementation of TFBS in synthetic promoters is limited by i) the availability of elucidated binding sites, ii) mechanistic understandings of the recruited TF and iii) unforeseen effects of neighboring sequences on necessary binding event(s).

Aside from TFBS, large expanses of native UAS regions have been used to activate promoter cores. The UAS and core regions can be teased from endogenous promoters and paired together to create hybrid promoters[78,79]. Like the use of multiple TFBS, multiple UAS regions can be place in tandem to increase transcriptional output of a core promoter[79,81] (**Figure 1.3**). In these synthetic promoters, endogenous TF are responsible for the strength. However, native UAS regions are not limited to endogenous TF activation. For example, *GAL* UAS regions were used to recruit imported estradiol-activated sTFs to an engineered synthetic promoter[78]. This engineered promoter allowed for improved zeanxanthin production in *S. cerevisiae* by 50-fold. Unfortunately, since native functional UAS regions can be hundreds of base pairs long, hybrid promoters constructed with them are significantly longer than other synthetic promoters. In engineering efforts, lengthy promoters are undesirable because they make expression cassette importations more difficult.

Nucleosome disfavoring sequences, on the other hand, are significantly shorter than UAS regions (just tens of basepairs), and easier to implement than TFBS. Nucleosome disfavoring sequences have been coupled with promoter cores[51,80,83], and

utilized in native promoter scaffolds to increase transcription output[69,82] (**Figure 1.3**). The most commonly used sequence is a homopolymeric tract of deoxyadenosine nucleotides, referred to as poly(dA:dT). These tracts are known to destabilize histone-DNA interactions necessary for nucleosomes formation[94,95]. Raveh-sadka *et al.* investigated the effects of poly(dA:dT) sequences on promoter function by altering their length, composition, and distance from several selected TFBS in a native promoter scaffold[82]. To do so, they designed and characterized 70 promoter variants. They concluded from these experiments that promoter strength can be tuned by manipulating the position of a 22 bp poly(dA:dT) tract upstream of a Gcn4p binding site. In this study, decreased distance between the two elements resulted in increased promoter strength[82]. These results were further corroborated in a similar, yet larger study of 777 designed promoter variants[69]. Aside from poly(dT:dA) tracts, there are other ways to manipulate nucleosome occupancy at the promoter to alter transcription strength. Our lab used nucleosome prediction software to reduce the predicted nucleosome occupancy score of the native constitutive promoters *TEF1* and *CYC1*. By introducing 5 to 71 mutations to *TEF1*, a 15-fold dynamic range in expression was achieved in a 15 member set of promoters[92]. This last example, in particular, demonstrates the utility of nucleosome depletion in tuning constitutive expression of promoters.

In these sections, I focused on DNA sequence manipulation to engineer promoters with desired functions. However, since recruited TFs are key proteins in promoter activity, simultaneously engineering both the promoter and the respective TF(s) is

common. Several examples of these TF-promoter engineered sets were briefly mentioned above. Nonetheless, since sTF can be a powerful, orthologous tool in manipulating gene expression at the transcription level, I will discuss common approaches to construct sTF and specifically, follow the incremental steps taken over two decades to create one of the most successful induction systems, the estradiol system.

1.5 TRANSCRIPTION FACTOR ENGINEERING IN *SACCHAROMYCES CEREVISIAE*

The most common strategy to engineer TFs is to fuse distinct protein domains derived from native TFs to create a TF with desired qualities. To do this, a DNA binding domain is utilized to provide binding properties to the sTF while a transcription activating or repressing domain is employed to supply transcription functionality. This modular construction allows for control over both the specificity and function of the TF. This design makes sTF a popular choice for constructing an orthologous induction system. A particularly successful system engineered with a sTF is the estradiol inducible system, and is the focus of this section.

Estradiol is an attractive inducing agent due to the limited number of unintended yeast gene activations. To construct an estradiol-activated TF, the human estrogen receptor domain is used as an activating domain. The activating domain is fused to a DNA binding domain to create a functioning TF, and to increase binding strength, an additional activation domain (VP16) is fused[96]. In one approach, the DNA binding domain of Gal4p is utilized[96]. In this application, a 200-fold activation is achieved via the native *GAL1* promoter in 8 to 11 hours after addition of estradiol. Over subsequent

years, the promoter used to drive expression of the sTF in this system was manipulated. A low strength promoter allowed for reduction of background expression levels or "leakiness"[97], and a *GAL1* promoter effected a feedback loop expression, where the resulting expressed Gal4p activated the *GAL1* promoter to increase expression of Gal4p[98]. However, since the promiscuous Gal4p DNA binding domain was being employed in these efforts, activation and repression of numerous untargeted genes persisted[99].

To address this, the Gal4p binding domain was swapped for a polydactyl zinc finger binding domain. Zinc finger DNA binding domains are the most commonly used DNA binding domain in sTF engineering efforts. These DNA binding domains are attractive due to their small size (about 30 amino acids), functional independence, and modular construction[100,101]. Each native zinc finger recognizes a 3-4 base pair unique DNA sequence and can be engineered to bind to sites other than their native recognition sequences[102-104]. When the zinc finger DNA binding domain is used to bind to a modified *GAL1* promoter, background expression of the estradiol induction system is significantly reduced[105]. However, the appeal of the Gal4p binding domain was strong activation. In this system, only a 50-fold induction is achieved despite reduced leakiness. Nonetheless, a 50-fold induction is still remarkable.

The system's major drawback is the price of the inducer. At ~$2000/g, estradiol represents one of the most expensive inducers, and effectively prices the induction system out of large scale synthetic biology applications[75]. Furthermore, the amount of

time and effort put forth to develop this system is rather notable as it is the product of two decades of improvements. The estradiol induction system not only represents the successes of TF engineering, but also the many hurdles that may be encountered when synthetically crafting a TF.

In conclusion, the advancements in promoter and TF engineering are relatively new. In fact, many of the synthetic promoters currently available were created during the work presented in this dissertation. This sharp rise in promoter engineering within in the last decade is the result of the increased demand for synthetic promoters of high quality. By moving away from native sequences, homologous recombination becomes less of a concern. However, none of the efforts explicitly address the length of the promoter, and very few create promoters *de novo*. In this regard, the need for additional short synthetic promoters for yeast remains.

1.6 SUMMARY

The work presented in this dissertation couples the current state of promoter engineering with our existing mechanistic and architectural understanding of the promoter. By doing so, I am able to address the need for novel promoters. My colleagues and I first highlight the utility of hybrid promoter engineering in the non-conventional yeast, *Yarrowia lipolytica*. Specifically, **Chapter 2** describes the use of multiple UAS in tandem to amplify and tune the transcriptional strength of endogenous promoters. Here, we create a set of *Y. lipolytica* promoters driving a continuum of expression and establish the strongest promoters ever reported in this organism. I draw

from lessons learned in this chapter to develop short, synthetic promoters in *S. cerevisiae*. In **Chapters 3 through 5** I describe individual minimization of the core and UAS regions of the promoter through the use of *de novo* libraries. Specifically, in **Chapter 3**, I establish synthetic, minimal cores by searching the sequence space, in terms of quality and length, between the TATA-box and TSS. To ensure robustness and modularity, I subject promoter cores to a series of tests. In **Chapter 4**, with the help of Matthew Deaner, I establish 20 succinct constitutive UAS elements through a four-fold approach involving the implementation of hybrid promoter libraries, UAS tandem assembly, unique DNA secondary structures and an easily programmable sTF. Finally, **Chapter 5** describes the development and application of a novel FACS-based isolation scheme for identification of inducible minimal hybrid promoters. In this final chapter, I present the utility of this workflow by successfully establishing seven synthetic maltose-inducible minimal promoters and one synthetic galactose-inducible minimal promoter. Thus, the work described herein contributes UAS and core parts to *S. cerevisiae* toolbox that can be modularly linked to craft a *de novo* set of promoters driving expression range of 70-fold. Moreover, this contributes a generic framework to synthetic biology that can be employed in other host organisms to expand their toolboxes.

Chapter 2: Tuning gene expression in *Y. lipolytica* by hybrid promoter approach

2.1 CHAPTER SUMMARY

We have successfully applied a hybrid promoter approach to produce high-expressing, tunable promoters in the nonconventional yeast, *Y. lipolytica*. We have overcome native expression limitations and provided a strategy for increasing native promoter capacity in a cellular system with ill-defined genetic tools. In doing so, this work creates the strongest promoters ever reported in *Y. lipolytica*, exhibiting eight-fold higher in protein levels compared with typically used endogenous promoters, and a range of more than 400 fold in mRNA levels. These results suggest native promoters in *Y. lipolytica* are limited and that this limitation can be partially or fully overcome through the addition of tandem copies of UAS[1].

2.2 INTRODUCTION

Developing and establishing a comprehensive suite of promoter elements in organisms with poorly defined genetic tools is essential for enabling metabolic and pathway engineering applications. The non-conventional yeast *Y. lipolytica* has received attention as a potential biofuels producing host. This yeast is a unique host for biochemical production and heterologous protein excretion due to its abilities to accumulate high levels of lipids[107-109], utilize hydrophobic and waste carbon sources[110-

---

[1] This work was published in "Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach" *Applied and Environmental Microbiology*. (2011) 7905-7914. It was a collaboration with first authors, John Blazeck and Leqian Liu. As second author, I assisted them in their work by constructing the expression cassettes for the promoter series, while they designed the cassettes, performed transcriptional analysis and ran all flow cytometry measurements, including conducting the plasmid stability test.

[112], and secrete native and heterologous proteins at high levels[113-115]. The availability of *Y. lipolytica*'s genome sequence[116,117] along with basic genetic tools such as transformation methods[118-120], gene knockouts [121], and both episomal[120,122-124] and integrative expression cassettes[125-127] enable metabolic engineering approaches. However, many of the methods in this organism rely on ill-defined genetic elements[111] especially in the area of promoters. One of the strongest promoters in *Y. lipolytica*, the *XPR2* promoter has complex requirements for induction that hinders its industrial applications [128]. Nevertheless, this promoter has been functionally analyzed to reveal a 105 base pair distal UAS fragment named UAS_{1B}[129,130]. Previously, between one to four tandem UAS_{1B} copies were fused to a *S. cerevisiae LEU2* core to create four increasingly strong hybrid promoters, named hp1d through hp4d[128]. As a result, the hp4d promoter has become a commonly used tool for heterologous protein expression in *Y. lipolytica*. A further re-analysis of these four promoters revealed a linear increase of promoter strength as a function of number of tandem UAS_{1B} elements. We sought to expand on this correlation to source promoters for *Y. lipolytica*.

In this chapter, we construct and characterize two sets of synthetic hybrid promoters: (i) an UAS_{1B}-*LEU* set in which between one and thirty two UAS_{1B} elements are tandem assembled to the minimal *LEU2* promoter region and (ii) an UAS_{1B}-*TEF* set in which eight or sixteen tandem UAS_{1B} elements are tandem assembled to varying sized *TEF* promoter regions. In doing so, this work creates the strongest characterized promoters in *Y. lipolytica* and the first ever reported capacity for tunable gene expression

in this system. Moreover, this work establishes hybrid promoter engineering as a generic approach for expanding the promoter toolbox in organisms with a limited tool set.

2.3 RESULTS

2.3.1 Characterization of endogenous promoters at the single-cell level

Prior studies of promoter strength in *Y. lipolytica* have relied on assaying whole cultures for protein expression level (using reporters such as β-galactosidase)[113,128]. It is commonly known that these methods can mask potential bimodal "on/off" distributions within the population. Thus, to avoid this complication, we sought to utilize a fluorescence-based assay using the *Y. lipolytica* plasmid pSl16-Cen1-1(227)[131]. All results generated in this study, except where indicated, employed derivatives of this replicative, ARS-CEN based plasmid. Since codon biases are known to limit translation in *Y. lipolytica*[132] and no fluorescent reporter protein has been previously used in *Y. lipolytica* to gauge promoter strength, we initially evaluated several available fluorescent reporter proteins. *TEF* promoter was used to drive expression of four different fluorescent proteins, yECitrine, EGFP, hrGFP and mStrawberry, and flow cytometry was performed to determine reporter functionality. Of these variants, only hrGFP imparted detectable fluorescence (**Figure 1a**). This gene, optimized for expression in mammalian cells, has the highest Codon Adaptation Index for *Y. lipolytica* of the four fluorescence genes, [133] indicating the closest compatibility with codon usage frequencies for this organism.

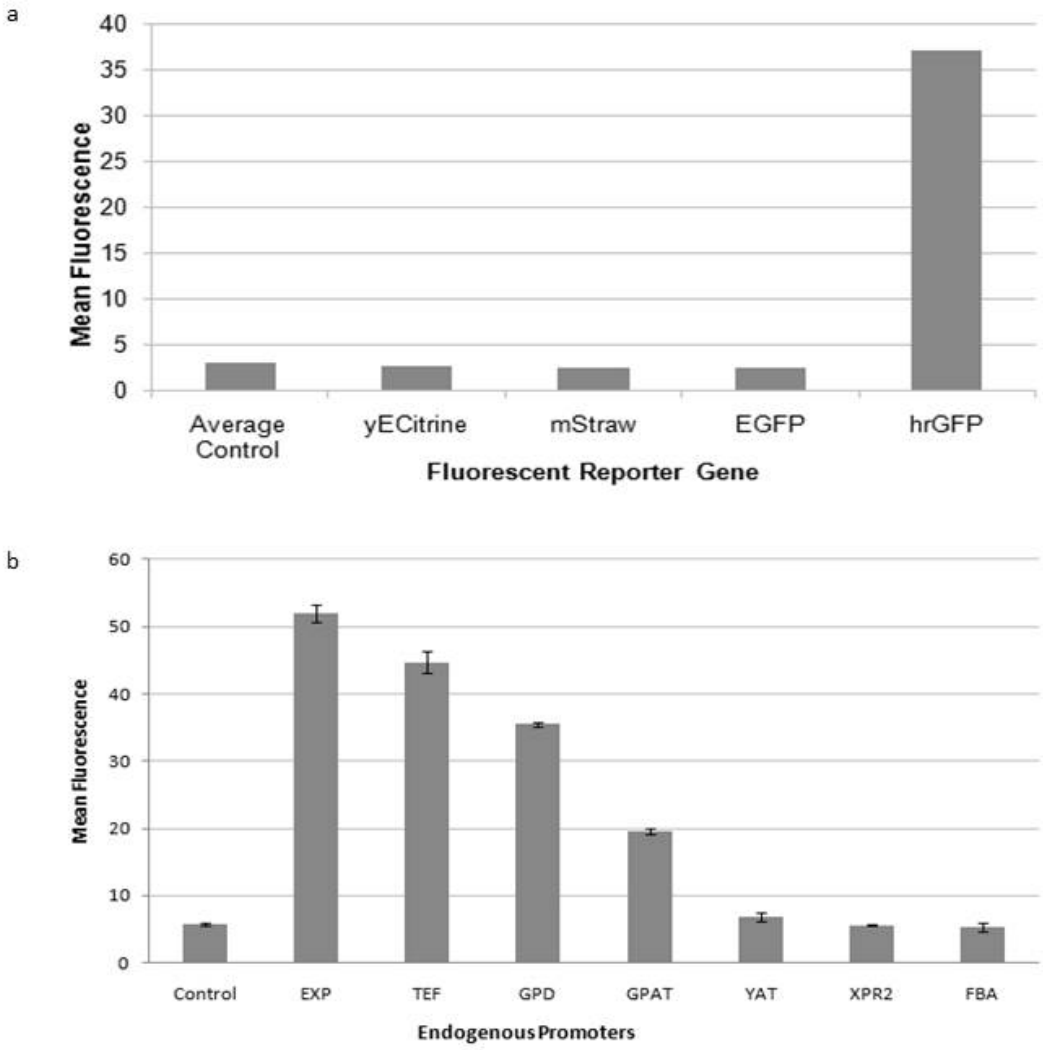


Figure 2.1: Evaluation of fluorescent proteins in Y. lipolytica

(a) The relative fluorescence levels of yE-Citrine, mStrawberry, EGFP, and hrGFP driven by the *TEF* promoter in *Y. lipolytica* were measured by flow cytometry. hrGFP was the only functional fluorescent protein for this system. (b) Endogenous promoters were used to drive expression of hrGFP (n=3). Error bars represent standard deviations from biological replicates. Control is *Y. lipolytica* strain PO1f transformed with pMCSCen1 with no fluorescent protein gene. *EXP* and *TEF* promoters were the strongest tested.

As a result, the hrGFP reporter gene was used to evaluate the promoter strength of seven previously identified endogenous *Y. lipolytica* promoters, *TEF*, *EXP*, *FBA*, *GPAT*, *GPD*, *YAT*, and *XPR2* (**Table 2.1**, **Figure 2.1b)**[113,134,135]. Based on this analysis, the relative ordering of promoters strengths is *EXP* > *TEF* > *GPD* > *GPAT* > *YAT* > *XPR2* > *FBA*. The low fluorescence values for even the strongest of these native promoters, *EXP* and *TEF*, highlight that even strong endogenous promoters in *Y. lipolytica* may be too low for metabolic engineering purposes. When each of these promoters was used in a plasmid-based construct, a bimodal fluorescence distribution was seen. The differential regulation patterns and small dynamic range of these endogenous promoters require a novel approach to enable metabolic engineering applications in this organism.

Promoter element name	Open reading frame regulated	YALI number	Basepair range	Reference number
EXP1	Export protein	YALI0C12034p	-999 to -1	[136]
GPAT	Glycerol-3-phosphate dehydrogenase	YALI0C00209p	-1130 to -1	[136]
GPD	Glyceraldehyde-3-phosphate dehydrogenase	YALI0C06369p	-931 to -1	[136]
TEF	Translation elongation factor EF-1α	YALI0C09141p	-406 to -1	[136]
YAT1	Ammonium transporter	YALI0E27203p	-775 to -1	[136]
FBA	Fructose-bisphosphate aldolase	YALI0E26004p	-830 to +171	[136]
XPR2	Alkaline extracellular protease	YALI0E26719p	-947 to -1	[128]
TEF(136)	Translation elongation factor EF-1α	YALI0C09141p	-136 to -1	This study
TEF(203)	Translation elongation factor EF-1α	YALI0C09141p	-203 to -1	This study
TEF(272)	Translation elongation factor EF-1α	YALI0C09141p	-272 to -1	This study
TEF(504)	Translation elongation factor EF-1α	YALI0C09141p	-504 to -1	This study
TEF(604)	Translation elongation factor EF-1α	YALI0C09141p	-604 to -1	This study
TEF(804)	Translation elongation factor EF-1α	YALI0C09141p	-804 to -1	This study

Table 2.1: List of promoter elements used in this study

TEF(1004)	Translation elongation factor EF-1α	YALI0C09141p	-1004 to -1	This study
Leum	β-isopropylmalate dehydrogenase	YALI0C00407p	-92 to +25	[128]
UAS1B	Alkaline extracellular protease	YALI0E26719p	-805 to -701	[128]

Table 2.1: List of promoter elements used in this study continued

The elements used in this study are listed with their names, open reading frames regulated, YALI accession numbers, and base pair ranges.

2.3.2 Creating and characterizing a hybrid promoter series using the UAS_{1B} element and minimal leucine core promoter

To bypass the expression limitations of endogenous promoters in *Y. lipolytica*, we evaluated the generalizable nature of hybrid promoters. In prior work, a nearly perfect positive linear correlation was previously detected between the number of tandem UAS_{1B} sequences and promoter outputs in the hp1d to hp4d promoter series[128]. Extrapolating this trend, we created a series of hybrid promoters by assembling between one and thirty-two tandem UAS_{1B} enhancer sequences to the *S. cerevisiae* leucine minimal promoter (Leum) to form promoters UAS_{1BX1}-Leum through UAS_{1BX32}-Leum (**Figure 3a**). This series was used to drive expression of hrGFP. Initially, an exponential increase in fluorescence was seen as the UAS_{1B} sequence count increased from one to eight. This trend became linear through nineteen tandem repeats (a total of 1995 bp of upstream activating sequences upstream of the core promoter). Finally, the output fluorescence seemed to be saturated through 32 tandem UAS_{1B} repeats (**Figure 2.2b**). This data strongly conformed to a Hill Cooperative Binding model (correlation coefficient of 0.95) and exhibited a high Hill Constant (3.889) which indicates a strong amount of binding cooperativity of the enhancer elements (**Figure 2.2b**). Specifically, this data was fit to the equation:

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

with the resulting coefficients of a= 0.794, Hill Coefficient= 3.889, and c= 10.146.

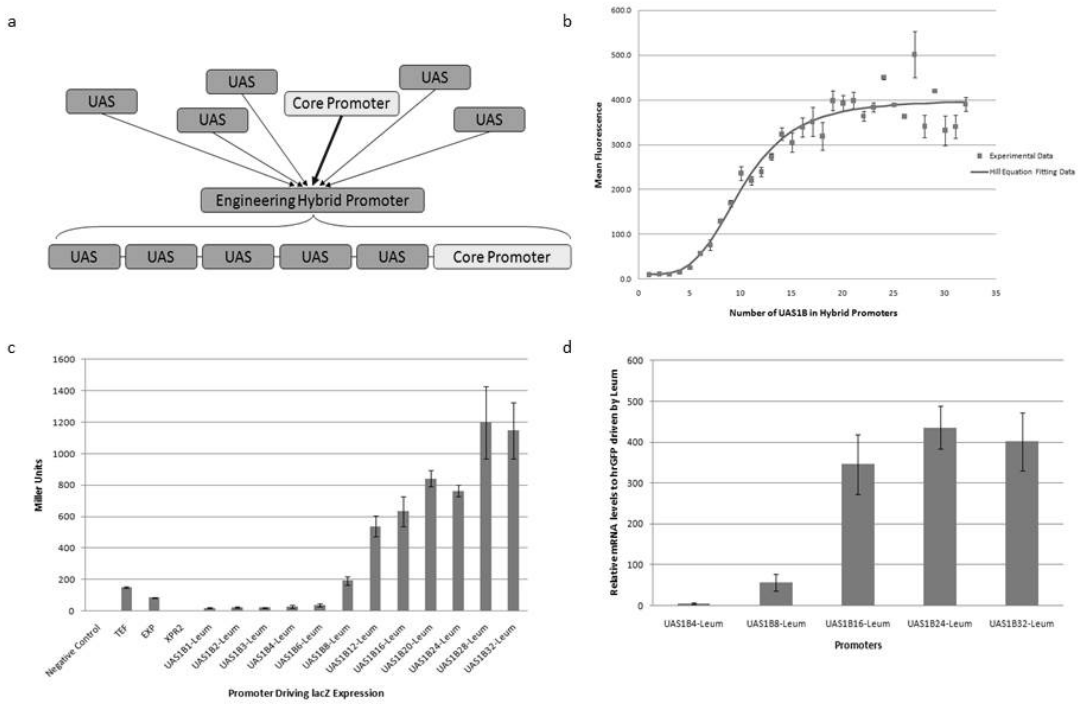


Figure 2.2: Development and characterization of a UAS_{1B}-Leum hybrid promoter set

(a) Hybrid promoter engineering involves the fusion of a core promoter with UAS elements. (b) Tandem assembly of 1 to 32 copies of UAS_{1B} to a minimal *LEU* promoter was used to drive expression of hrGFP, and resulting fluorescence was measured by flow cytometry. Error bars represent standard deviations from biological triplicates. The relative strengths of UAS_{1B}-Leum promoters were fit to a Hill equation (see equation 1), resulting in coefficients of a= 0.794, Hill Coefficient= 3.889, and c= 10.146 using Polymath (Willimantic, CT) software. (c) Endogenous and hybrid promoters were tested with a β-galactosidase reporter gene, yielding results similar to those of the hrGFP assay. Error bars represent standard deviations from biological triplicates. (d) qRT-PCR of select promoter constructs was used to calculate mRNA levels relative to those of the minimal Leum promoter. mRNA levels match fluorescent levels. Error bars represent standard deviations from biological triplicates.

2.3.3 Transcriptional analysis of the UAS$_{1B}$-leum hybrid promoter series

A transcriptional analysis was performed to confirm that the observed effect in fluorescence was indeed manifested at the transcriptional level. To do so, qRT-PCR analysis was employed using the hrGFP mRNA of select promoter constructs (Leum, UAS$_{1BX4}$-Leum, UAS$_{1BX8}$-Leum, UAS$_{1BX16}$-Leum, UAS$_{1BX24}$-Leum, and UAS$_{1BX32}$-Leum) (**Figure 2.2c**). Expression values were normalized to the mRNA level seen with only the minimal leucine promoter used to drive hrGFP. Indeed, the increase in mean fluorescence levels was strongly correlated with the increase in relative mRNA levels. The relative mRNA levels increased and likewise, plateaued for constructs with a high number of UAS$_{1B}$ repeats. Moreover, these results demonstrate an extraordinary 400-fold range of promoter strength in this series.

2.3.4 Utility and stability of the UAS$_{1B}$-leum hybrid promoter series

To ensure that the observed effect was independent of reporter gene, we sought to provide a further characterization of the UAS$_{1B}$-Leum promoter series with a separate reporter gene, the β-galactosidase gene encoded by *E. coli* lacZ. Native promoters *TEF*, *EXP*, and *XPR2* and hybrids UAS$_{1B}$-leum with 1, 2, 3, 4, 6, 8, 12, 16, 20, 24, 28, and 32 UAS$_{1B}$ copies were used to construct expression cassettes with lacZ in place of hrGFP. β-galactosidase assays were performed as described previously[137,138] with a maximum value of 1198 miller units generated by the UAS$_{1BX28}$-leum construct (**Figure 2.2d**). β-galactosidase assay results matched well with the hrGFP analysis and the data showed a strong positive statistical correlation (r^2=0.85). These results demonstrate that the

UAS1B-leum hybrid promoter series developed here is a generic tool for obtaining tuned gene expression in *Y. lipolytica*.

These hybrid promoters rely on a high number of tandem repeats, thus, genetic stability was evaluated. To accomplish this, promoters with 12 or 16 UAS_{1B} copies were tested on the basis of sequence fidelity after non-selective serial subculturing. These strains were subcultured for a total of 36 generations. After this process, cells were harvested and plasmids were isolated and sequenced to assess gene construct stability. In total, 20 separate plasmids with UAS_{1BX12}-Leum and 20 with UAS_{1BX16}-Leum promoters were evaluated. 17 out of 20 UAS_{1BX12}-Leum and 20 out of 20 UAS_{1BX16}-Leum were sequenced and restriction enzyme digest confirmed after 36 doublings. Only 3 out of 20 UAS_{1BX12}-Leum were truncated to UAS_{1BX3}-Leum, indicating instability. Thus, these promoters are suitably stable in *Y. lipolytica* for long-term expression and use. Collectively, this data suggests that the expression output from hybrid promoters can be altered by changing the number of UAS with a given core. Next, we sought to address the ability to alter the core promoter of this construct.

2.3.5 Generalizing the hybrid promoter approach by switching the core promoter

The data above suggests that tandem UAS elements may serve as modular expression amplifiers for a given promoter. Next, we sought to test the hypothesis that even native promoters in *Y. lipolytica* are enhancer limited and can be strengthened by adding additional UAS elements. To do so, we constructed new hybrid promoters containing either eight tandem UAS_{1B} sequences (UAS_{1BX8}) or sixteen tandem UAS_{1B}

sequences (UAS_{1BX16}) inserted 5' upstream of a series of different native *TEF*-based core promoters. Specifically, we amplified eight different regions of the *TEF* promoter spanning 136 bp and 1004 bp upstream of the ATG starting site from PO1f [128]genomic DNA (**Table 2.1**). Included in this set is the consensus 404 bp *TEF* promoter for *Y. lipolytica* as well as lengthened and truncated versions of this promoter. These eight core *TEF* promoters and their corresponding UAS_{1BX8} and UAS_{1BX16} hybrid promoters were tested and compared with the Leum, UAS_{1BX8}-Leum and UAS_{1BX16}-Leum constructs.

This new series of hybrid promoters was assayed via hrGFP fluorescence by flow cytometry. In the absence of UAS elements, it can be seen that the fluorescence value decreases for truncated promoters below the consensus *TEF* size as expected. Moreover, the full length and much larger *TEF* promoters are stronger than the minimal leucine promoter (**Figure 2.3a**). The UAS_{1BX8} and UAS_{1BX16} enhancer fragments in isolation do not confer any promoter activity (data not shown). When UAS are hybridized with *TEF* core promoters, a substantial increase in the net promoter strength was seen regardless of the *TEF* derivative utilized. The enhancement provided by UAS_{1BX8} was roughly half the value obtained by using UAS_{1BX16}. Moreover, these enhancements were seen for both more minimal and full length *TEF* promoter elements, even with the existence of naturally occurring UAS in the consensus and longer *TEF* promoters. Thus, this data suggests that even strong endogenous promoters like *TEF* are enhancer limited in *Y. lipolytica* and their expression capacity can be increased through additional UAS elements.

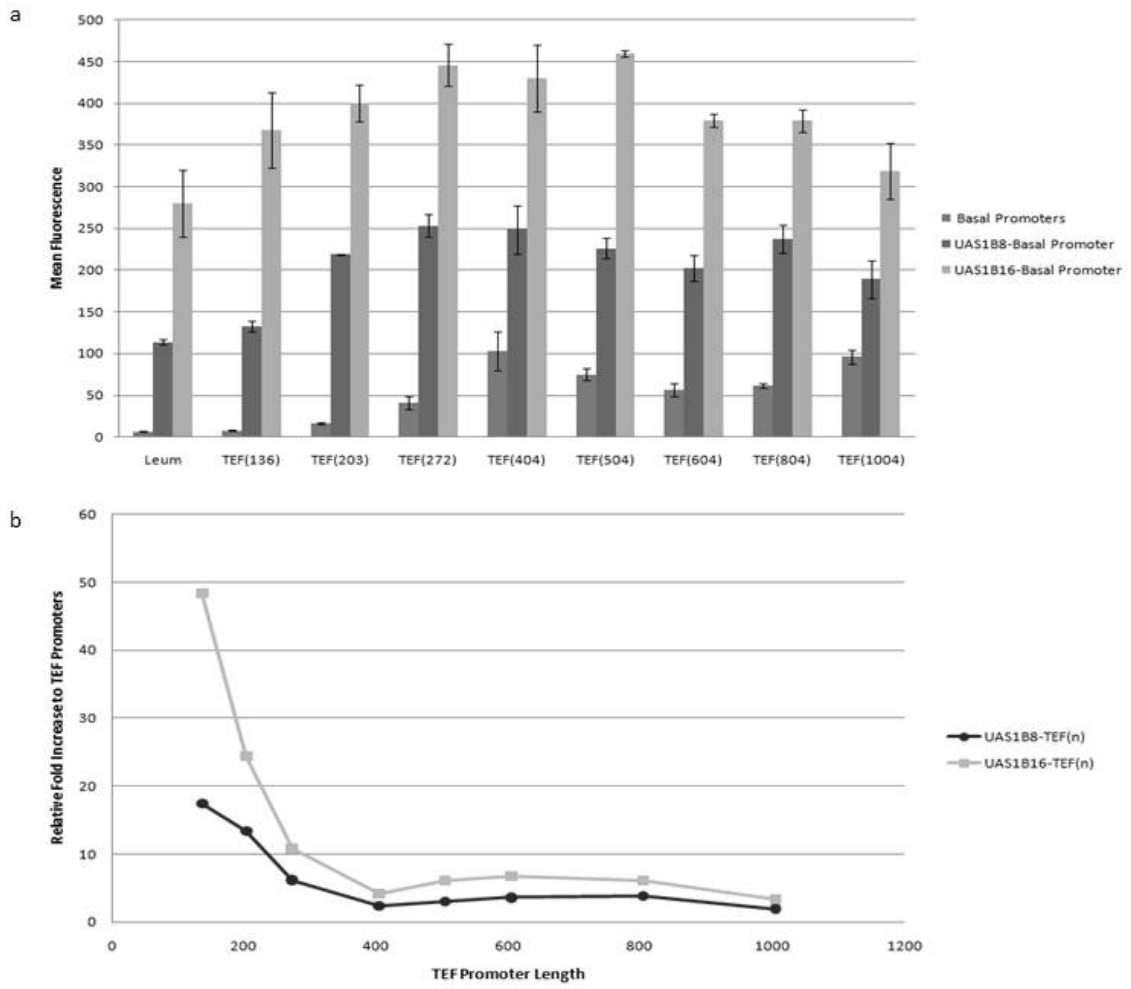


Figure 2.3: Expanding the hybrid promoter approach by altering the core promoter

(a) Characterization of relative promoter strengths for the core *TEF* promoters, UAS_{1BX8}-*TEF* promoters, and UAS_{1BX16}-*TEF* promoters using flow cytometry. These data were compared with data for the UAS_{1BX8}-Leum and UAS_{1BX16}-Leum promoters. (b) Tuning ability of UAS_{1BX8} and UAS_{1BX16} decreases as a function of core promoter length.

The amplification of expression imparted by the UAS_{1BX8} and UAS_{1BX16} was not identical for all promoters. The fold increase of the constructs relative to the UAS-free *TEF* core promoters is plotted in **Figure 2.3b**. The largest improvement is obtained when

the UAS elements were positioned closest to the core promoter. However, it is interesting to note that the entire promoter size (all UAS copies and core) reaches up to 3 kb for many of these; a region that is quite large for typical yeast constructs. In addition, many of these larger promoters were the best performing. Moreover, nearly all proved to be stronger than the corresponding UAS_{1BX8}-Leum and UAS_{1BX16}-Leum promoters, demonstrating the increased fitness of the *TEF*-based core promoter regions for strong hybrid promoter engineering. In this regard, this work demonstrates that tandem UAS elements serve to amplify the expression level imparted by the core promoter chosen.

2.4 Discussions and Conclusions

This study established a synthetic approach for tuning gene expression using a hybrid promoter approach. In doing so, this work created the strongest known promoters in the oleaginous yeast *Y. lipolytica* and allows for fine-tuned gene expression in this organism. The general strategy developed in this study utilizing tandem-assembled UAS could be applied to other organisms and further generalized by using other UAS. Finally, these results have the biological implication that the expression capacity for promoters (at least in *Y. lipolytica*) is enhancer limited. In this regard, this approach expands the quantity and quality of parts available for systems biology research[139,140].

The magnitude of these hybrid promoters can be seen by the relative mRNA range of more than 400-fold between the core promoter (Leum) and the maximum of UAS_{1BX24}-Leum. The strongest UAS_{1B}-Leum hybrid promoter exhibited a more than eight fold increase in promoter strength in terms of Miller units compared to the strong

endogenous promoters tested in this study. Although the expression driven by UAS_{1B4}-Leum in these studies was substantially lower than previously reported, this work still presents up to a fourfold increase in relative performance compared with the best reported endogenous promoters or previously constructed hybrid promoters[128,141]. This illustrates that multiple tandem repeats of the UAS_{1B} activate transcription to levels far stronger than those previously described and that this activation can occur at regions more than 2000 nucleotides upstream of the start codon (for $UAS_{1BX16\text{-}X32}$-Leum).

The UAS_{1BX8} and UAS_{1BX16} were shown to behave as synthetic amplifiers when tested with various *TEF* promoter derivatives. In this regard, we demonstrated that the ability of a UAS_{1B} to amplify expression is independent of the core promoter. However, the magnitude of amplification was dependent on the core promoter used. Thus, both the choice of UAS and core promoter contribute to the collective strength of the hybrid promoter. This observation raises the possibility of rationally designing hybrid promoters with specified expression strengths. The drastic increase in expression levels by both of the UAS_{1BX8} and UAS_{1BX16} across this series indicates that these genetic elements are portable, modular components that can generically alleviate native enhancer-limitation without disrupting endogenous regulation. The modularity of the UAS_{1B} insert and the strength of the UAS_{1B}-Leum and UAS_{1B}-*TEF* series advocate the use of hybrid promoter engineering as generic approach towards building stronger, fine-tuned promoters with interchangeable, modular components.

The hybrid promoter engineering described in this chapter is a valuable synthetic biology approach for the construction of high-level and fine-tuned promoters with interoperable promoter parts. By utilizing this approach, we expanded the metabolic engineering toolbox in *Y. lipolytica* and developed several novel promoter series - $UAS_{1BX1\text{-}X32}$-Leum, UAS_{1BX8}-*TEF*, and UAS_{1BX16}-*TEF* delivering a range of expression levels from low to the highest ever reported in this organism. In subsequent chapters, we use hybrid promoter engineering to expand the metabolic engineering toolbox in *S. cerevisiae*. Specifically, in **Chapter 4**, we draw from the tandem assemble approach described in this chapter to increase the transcriptional output of synthetic minimal cores.

Chapter 3: Synthetic minimization of promoter core elements in *S. cerevisiae*

3.1 CHAPTER SUMMARY

Synthetic promoters, especially minimally sized, are critical for advancing fungal synthetic biology. Fungal promoters often span hundreds of base pairs, nearly ten times the amount of bacterial counterparts. This size limits large-scale synthetic biology efforts in yeasts. To address this shortcoming, we establish a methodical workflow necessary to identify robust minimal core elements that can be linked with native UAS and TFBS to develop short, yet strong yeast promoters of a constitutive and inducible nature. To ensure robustness and independence, minimal core elements were subjected to a series of rigorous tests including i) transcriptional amplification by two native constitutive UAS, and a well-defined inducible TFBS, ii) expression of two ORFs, and iii) consistent transcriptional strength in two DNA contexts[2].

3.2 INTRODUCTION

The innate complexity of eukaryotic transcription makes organisms like *S. cerevisiae* quite distinct from their bacterial counterparts. This complexity is exemplified by considerably longer eukaryotic promoters compared to bacterial ones. *E. coli* promoters typically span well under 100 nucleotides[143] whereas native yeast promoters (especially those used in synthetic biology efforts) can stretch hundreds of base pairs[144,145]. Mechanistically, a longer stretch of DNA is needed to load and stabilize a

---

2 The work presented in this chapter was published in "The development and characterization of synthetic minimal yeast promoters" *Nature Communications* (2015) 7810. Sequences for synthetic core elements presented in this chapter are patented under the US patent #20160160299.

bulkier (~100 kDA larger) and more highly regulated eukaryotic RNAPII[146,147]. As such, two of the most commonly used yeast promoter elements, the *GAL1* inducible promoter and the strong *GPD* (*TDH3*) constitutive promoter span over 450 and 650 nucleotides respectively. Beyond the lack of sequence diversity, these fungal eukaryotic promoters make large-scale synthetic biology efforts cumbersome. Specifically, a single-gene circuit carrying a 1.5 Kb gene requires an additional 1 Kb of regulatory DNA (between the promoter and terminator) for appropriate expression, thus increasing the DNA cargo load by over 60%. If an extensive, multi-gene heterologous pathway is required, this regulatory DNA could easily add up to tens of thousands of nucleotides (nearly ten times as much as needed in bacteria). Thus, minimal sized fungal promoters are an essential and lacking tool in the field for fungal synthetic biology.

To date, this problem of bulky yeast promoters has been scarcely addressed in synthetic biology; yeast promoter engineering has mainly focused on building from larger, native scaffolds[63,66-69,71,82,92,148,149]. At the same time, synthetic promoter assembly via hybrid technologies[78,150,151] demonstrates promise in moving away from native scaffolds, but broadly still relies on endogenous parts. Thus, in this study, we address the lack of minimal yeast promoters by establishing minimal core elements that can be linked with native UAS to create fully functioning promoters. To develop these core elements, we isolated candidates from 15 million variants comprised only of two essential transcriptional features and subjugated them to a series of rigorous tests to establish nine robust, minimal core elements with truly modular and context independent functions

(**Figure 3.1**). These elements are highly unique both among each other and to any native genomic sequences in *S. cerevisiae.* We likewise demonstrate the ability to establish minimal, inducible promoters through this method with maximal expression levels similar to that of *GAL1*.

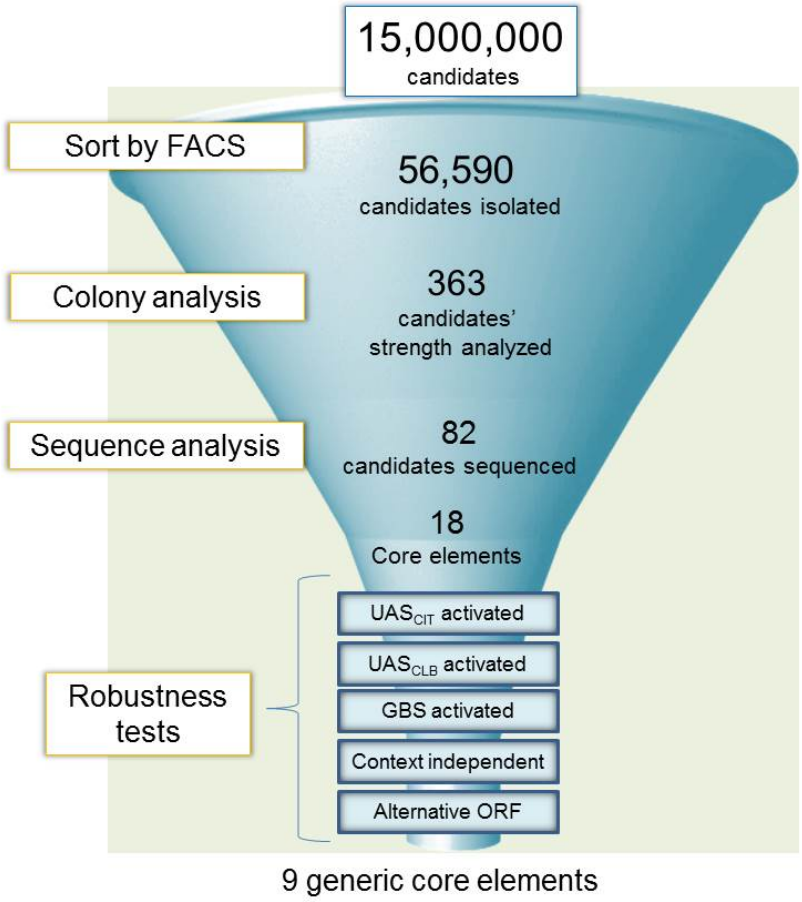


Figure 3.1: Isolation scheme of synthetic core elements

27 libraries of 15 million candidates were created. Libraries were sorted by FACS to pool the top 0.15% fluorescent cells. These sorted cells were subjected to colony analysis via flow cytometry. High strength candidates were sequenced. Due to the large number of multiple tandem insertions and duplicate candidates, only 18 core elements were present in the pool of 82 sequenced candidates. These 18 core elements were characterized under UAS_{CIT}, UAS_{CLB}, and Gal4p transcription factor binding site (GBS) activation. Lastly, context dependency of promoters created using synthetic cores was determined. All core elements were confirmed to function with an alternative ORF.

3.3 RESULTS

3.3.1 Creating a method for identifying minimal yeast promoters

To design minimal promoters, we established a plasmid-based core element scaffold (**Figure 3.2**) to determine the shortest length required for transcription and to serve as a platform for hybrid promoter technology. However, to establish minimal promoters, we created and evaluated ensembles of variably spaced TATA box-TSS core elements. Although the core elements contain all sequence components necessary for transcription initiation, very low expression is expected without UAS elements which provide the overall strength and regulation of a promoter[6,152]. Thus, we employed native UAS to provide strength to our core elements. Throughout this workflow, a series of stringency and robustness tests were invoked to establish core elements that are modular, context independent, robust, and interoperable (**Figure 3.1**).

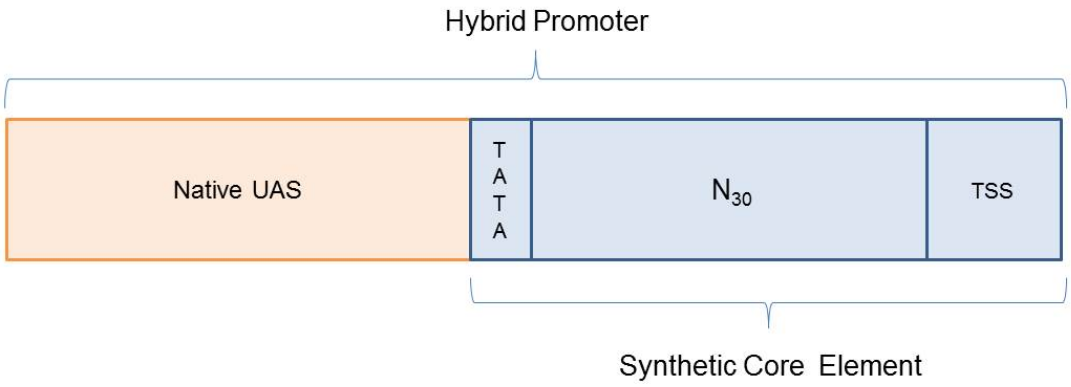


Figure 3.2: Synthetic core element scaffold

Minimal synthetic cores of 56 bp are paired with native UAS elements to isolate functioning core elements.

3.3.2 Core element designs

Our promoter minimization approach starts with an ensemble of core element scaffolds based on the necessary architectural elements of a promoter discussed earlier (**Figure 3.2**). We first sought to determine the minimum number of nucleotides required between the TATA box and the TSS to promote successful loading of the pre-initiation complex (PIC) and thus, transcription initiation by RNAP. As previously discussed, native spacing in *S. cerevisiae* is seen to span between 40 to 120 bp[11,12] . This lower limit is peculiar since the structure for yeast RNAP supports a minimal spacing of 30-31bp[153] matching the optimal spacing found in mammalian promoters[154]. Thus, multiple plasmid-based spacer libraries were synthesized containing 20 (N_{20}), 25 (N_{25}) and 30 (N_{30}) nucleotides between the TATA box and TSS using random oligonucleotides (**Figure 3.3a-b**). The impact of these spacers was assessed by linking these libraries to a fluorescent reporter protein (yECitrine), and assessing library function with flow cytometry. Interestingly, all libraries showed a lengthening in the histogram tails towards higher fluorescence when compared with a negative control (no fluorescent protein) (**Figure 3.4a-c**). However, only the N_{30} library exhibited a small population shift towards higher fluorescence indicative of a subpopulation of functional core elements (**Figure 3.3b)**.

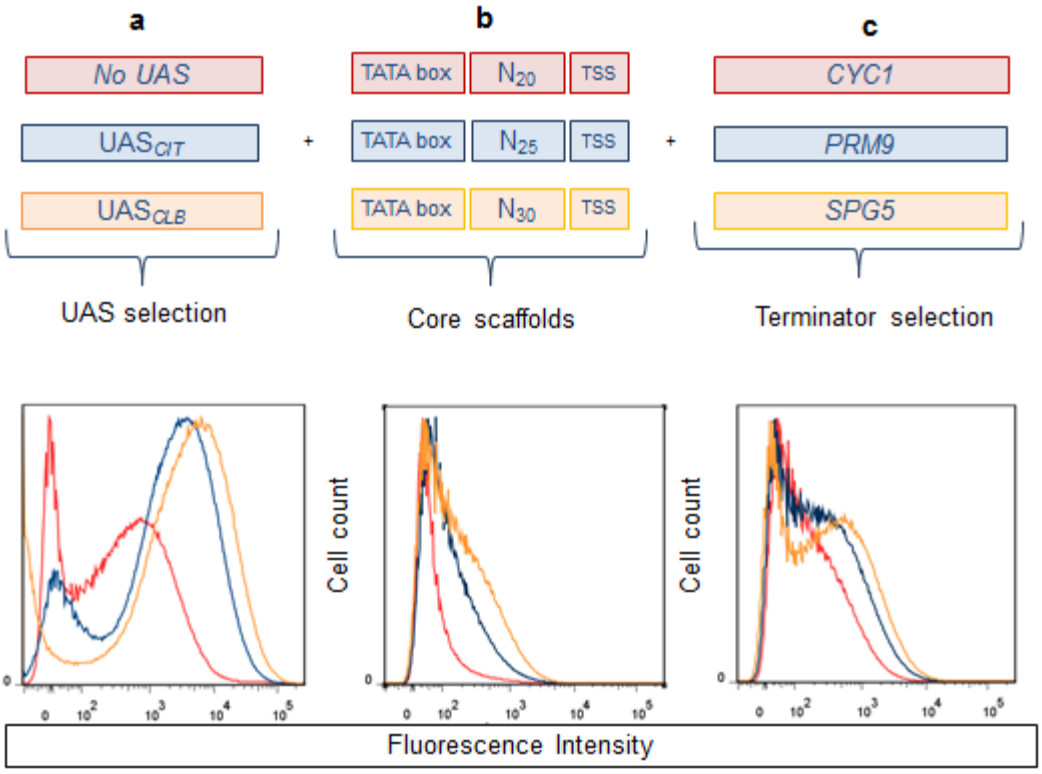


Figure 3.3: Promoter scaffold library assemblies

The color in the box outline in the top half of the figure corresponds to the color used in the histograms on bottom half of figure. (a) Two UAS (UAS$_{CLB}$ and UAS$_{CIT}$) were used to distinguish functional candidates from non-functioning candidates. All libraries showed a similar shift, with UAS$_{CLB}$ eliciting the strongest shift towards higher fluorescence. Libraries shown are N$_{30}$-*SPG5*, UAS$_{CIT}$-N$_{30}$-*SPG5*, and UAS$_{CLB}$ -N$_{30}$-*SPG5*. (b) Three lengths of cores were tested, where oligonucleotides of 20 bp (N$_{20}$), 25 bp (N$_{25}$), and 30 bp (N$_{30}$) were placed between TATA box and TSS. Libraries shown here are N$_{20}$-*CYC1*, N$_{25}$-*CYC1*, and N$_{30}$-*CYC1* without a UAS. An increase in the spacing results in a tail lengthening of the histograms. (c) Expression enhancing terminators (those thought to increase mRNA half-life) were used to distinguish functional candidates from non-functioning candidates. *SPG5* produced a select population shift greater than other terminators tested. Shown here are libraries N$_{30}$-*CYC1,* N$_{30}$-*PRM9*, and N$_{30}$-*SPG5*.

As the expected transcription function of core elements is slight, we sought to amplify the signal of functional minimal core elements. First, we established UAS-core element hybrid libraries in an effort to isolate core elements that can be substantially amplified with a UAS (**Figure 3.3a**). To do so, we employed a native UAS element previously demonstrated to be effective in yeast, a 275 bp sequence from the mitochondrial citrate synthase (*CIT1*) gene[150,155] referred to as UAS_{CIT}. Second, we sought to amplify the signal further by using an expression enhancing terminator, *SPG5*, shown to elevate mRNA concentration by increasing the transcript half-life[156] (**Figure 3.3c**). Both UAS_{CIT} and the *SPG5* terminator resulted in shifts in all libraries (**Figure 3.3a, 3.3c**), with the most dramatic, positive candidate shifts seen in the N_{30} library (**Figure 3.4a-c**). As a result, minimal core elements were identified from libraries comprised of UAS_{CIT} and *SPG5* terminator linked with the core element. Every library was sorted by FACS to pool the top ~0.15% expressing cells together, and subsequently only the N_{30} library resulted in robust promoters with a low frequency of multiple insertions (**Figure 3.4d-f**).

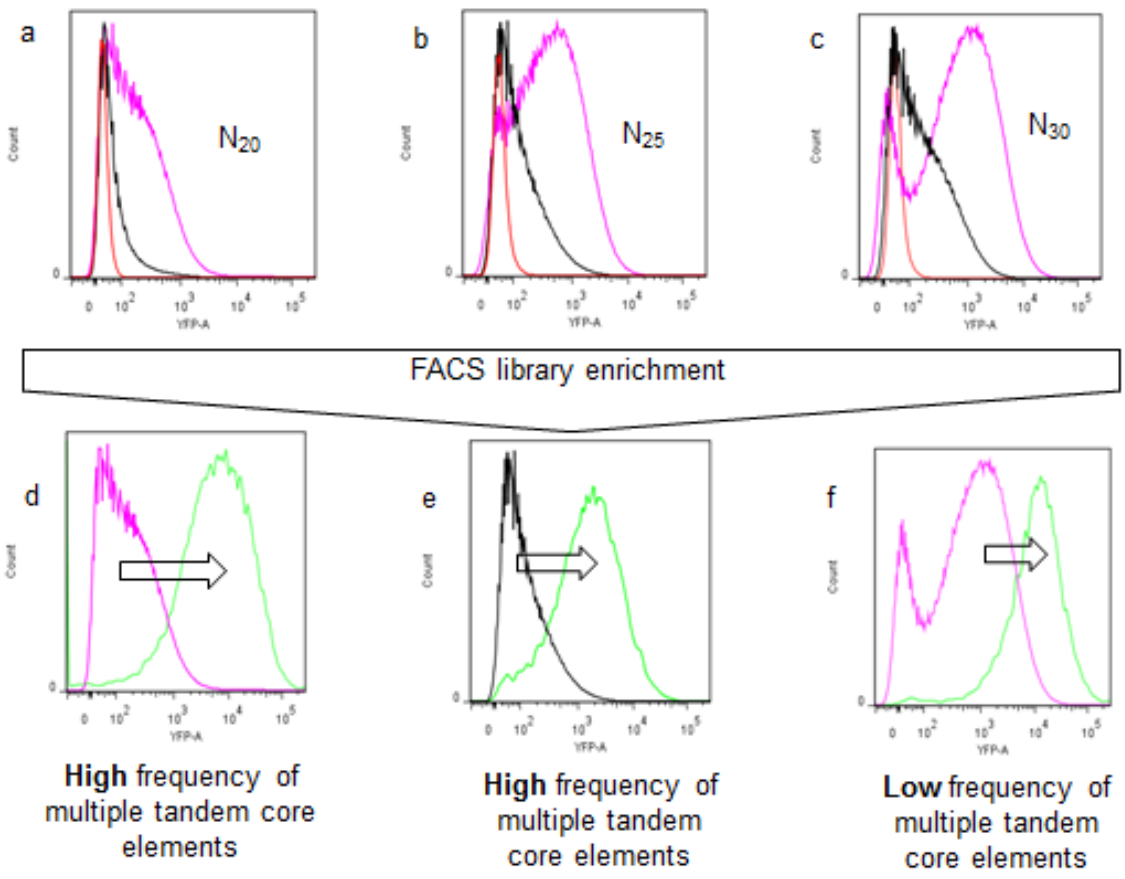


Figure 3.4: Histograms of select libraries before and after sorting by FACS

The top 0.15% fluorescent cells of each library was pooled together by fluorescence activated cell sorting (FACS). (a-c) Flow cytometry histograms of unsorted libraries were gathered and overlaid. Red histograms represent unsorted negative controls (no yECitrine) (a-c), (e). Black histograms represent unsorted libraries without a UAS; N_{20}, N_{25} and N_{30} for (a), (b, e) and (c) respectively. Purple histograms represent UAS_{CIT} coupled with N_{20}, N_{25} and N_{30} for (a, d), (b) and (c, f) respectively. Histograms of sorted libraries were gathered and overlaid with unsorted respective libraries (d-f). Arrow in (d-f) indicates histogram shift due to FACS with histogram to left that of unsorted population. Unsorted populations shown in panels (d-f) color correspond to unsorted populations in panels (a-c) respectively. Only a few select sorted libraries are shown. Although histograms shifts are present after all library sortings (d-f) (compared to unsorted libraries), only libraries with N_{30} spacing produced libraries with a low frequency of tandem insertions.

3.3.3 Isolation of putative cores elements

From the initial 15 million candidates synthesized as described above, we sought to isolate minimal core elements with desirable characteristics for synthetic biology applications. Specifically, these core elements should (i) be generically activated by any UAS or TFBS, (ii) function with alternative genes, and (iii) display little context dependence. Thus, a series of tests were used to narrow down the candidates to a set of nine robust generic core elements (**Figure 3.1**).

To begin, candidates were isolated from the enriched UAS_{CIT}-N_{30}-*SPG5* library obtained via FACS followed by isolated colony analysis, and sequencing. As noted earlier, this library had the lowest frequency of multiple insertions (**Figure 3.4f**) and thus, indicates that 30 bp may be the minimal spacing required between the TATA box and TSS for *S. cerevisiae*. Following isolation, characterization, quality control (reproducible upon retransformation and homogenous histograms), and sequencing, a total of 18 putative unique core elements were identified. As these elements were currently linked with a UAS_{CIT} upstream region, we removed this region and assessed the strength of the core element itself and found that indeed these core elements allowed for slight, but detectable, transcription (**Figure 3.5**).

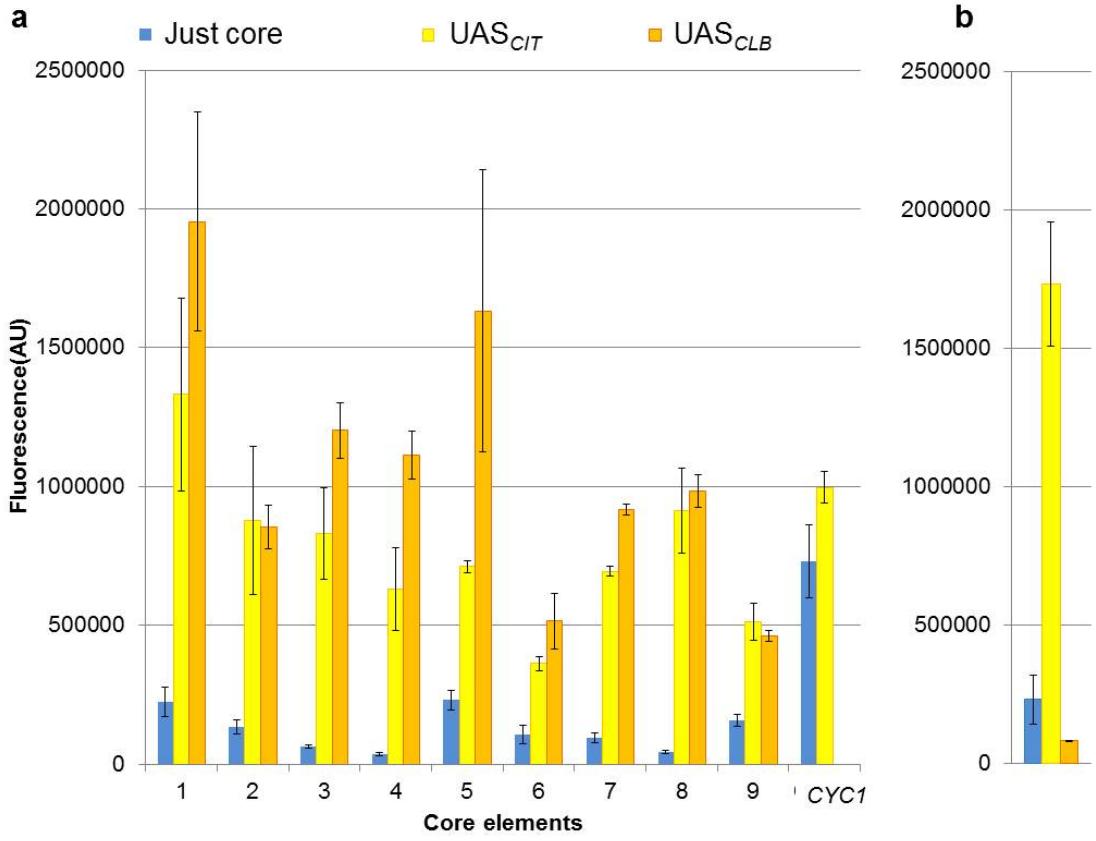


Figure 3.5: Core elements can be activated by UAS elements to create constitutive promoters

(a) UAS$_{CIT}$ and UAS$_{CLB}$ can activate all core elements. However, of the 18 core elements tested, three did not activate similarly with UAS$_{CIT}$ and UAS$_{CLB}$ (n=4). For comparative purposes, one core element rejected in this test is shown (n=3) (b). Error bars represent standard deviation among biological replicates.

3.3.4 Tests to isolate robust, minimal core promoter elements

These 18 putative core elements were next assessed through a series of robustness tests (**Figure 3.1**). All compared fluorescence values for each test were gathered on the same day. To begin, we sought to evaluate the impact of an alternative, constitutive UAS. To do so, these core elements were linked with another previously used UAS[150],

the 240 bp sequence of the mitotic cyclin (*CLB2*) gene[157], termed the UAS_{CLB}. Many of the putative core elements were also able to be activated by UAS_{CLB} with a threshold of 2-fold increase in fluorescence (**Figure 3.5a**); however, some were not activated by this UAS and thus were removed from the candidate pool (**Figure 3.5b**). Of the 18 putative core elements, three were determined to not be functionally robust with respect to activation by an alternative, constitutive UAS element.

As a second robustness test for generalizability of these cores, we sought to demonstrate whether these core elements could be linked with a minimal galactose inducible UAS element to enable inducible promoter function in a minimal sequence space (**Figure 3.1**). To do so, these 18 candidate core elements were linked with 17 bp *GAL1*-derived Gal4p binding sites (GBS), UAS_{G4BS3} and UAS_{G4BS4} , as previously described[150]. As a means of linking these minimal core elements to a short UAS element, we evaluated variable spacing between the TATA box of the core element and the TFBS, and found that a neutral AT-rich 30 bp sequence (**Figure 3.6a, Figure 3.7a-b**) was required to avoid possible observed steric hindrances between the Gal4p transcription factor and TATA-box binding protein component of PIC (**Figure 3.6b-c**). In general, we found UAS_{G4BS4} induced expression under galactose better than UAS_{G4BS3} (**Figure 3.8**) with both UAS producing little to no effect under glucose (**Figure 3.9**). As with the UAS_{CLB}, most (**Figure 3.10a-b**), but not all (**Figure 3.10c**), putative core elements were able to be turned into inducible promoters with the addition of a UAS_{G4BS4}. Of the 18 putative core elements, two were determined to not be functionally robust with respect to

activation by both inducible UAS elements tested. More importantly, when UAS_{G4BS4} is combined with minimal core elements, the expression level in the induced mode for two of these promoters is comparable to that of the full native *GAL1* promoter, but in only 22% of the DNA sequence (**Figure 3.10a-b**). In fact, when a 54 bp cluster of sequence found in native *GAL1* containing three Gal4p binding sites is linked with all nine of the final core elements, most either reach *GAL1* strength or approach it (**Figure 3.11**).

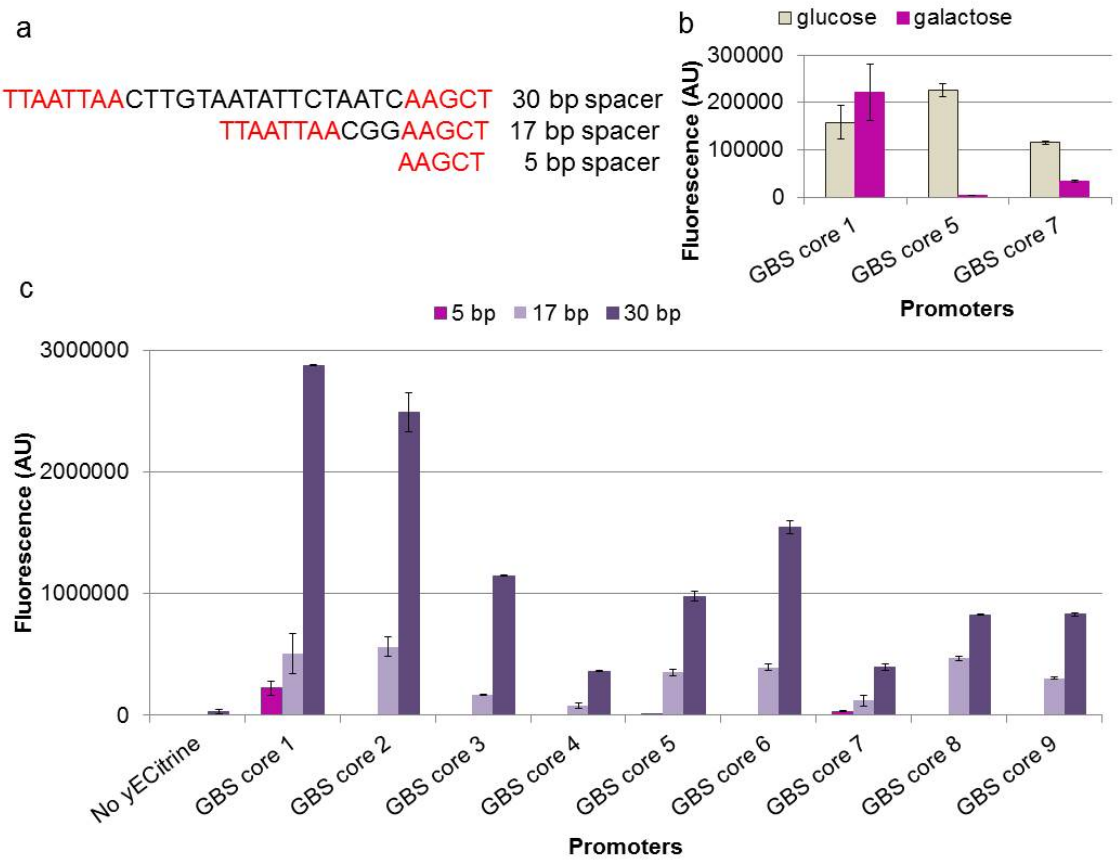


Figure 3.6: AT-rich spacer is required for galactose induction of core element by GBS

Experiments exploring spacing requirements indicates at least 30 bp spacing is required for Gal4p binding site (GBS) to perform. (a) Sequences of spacers used to distance GBS from core element. Restriction enzyme sites are indicated by red font. (b) No induction was observed when GBS was spaced just 5 bp upstream of the core element (n=3). In fact, two of the three promoters tested showed a reduction in expression under galactose induction. (c) Thus, GBS was positioned upstream of all core elements 17 bp and 30 bp, with the latter distance yielding the largest induction. Shown here is data for UAS_{G4BS4} coupled with core elements under galactose induction (n=3). All error bars represent standard deviation among biological triplicates.

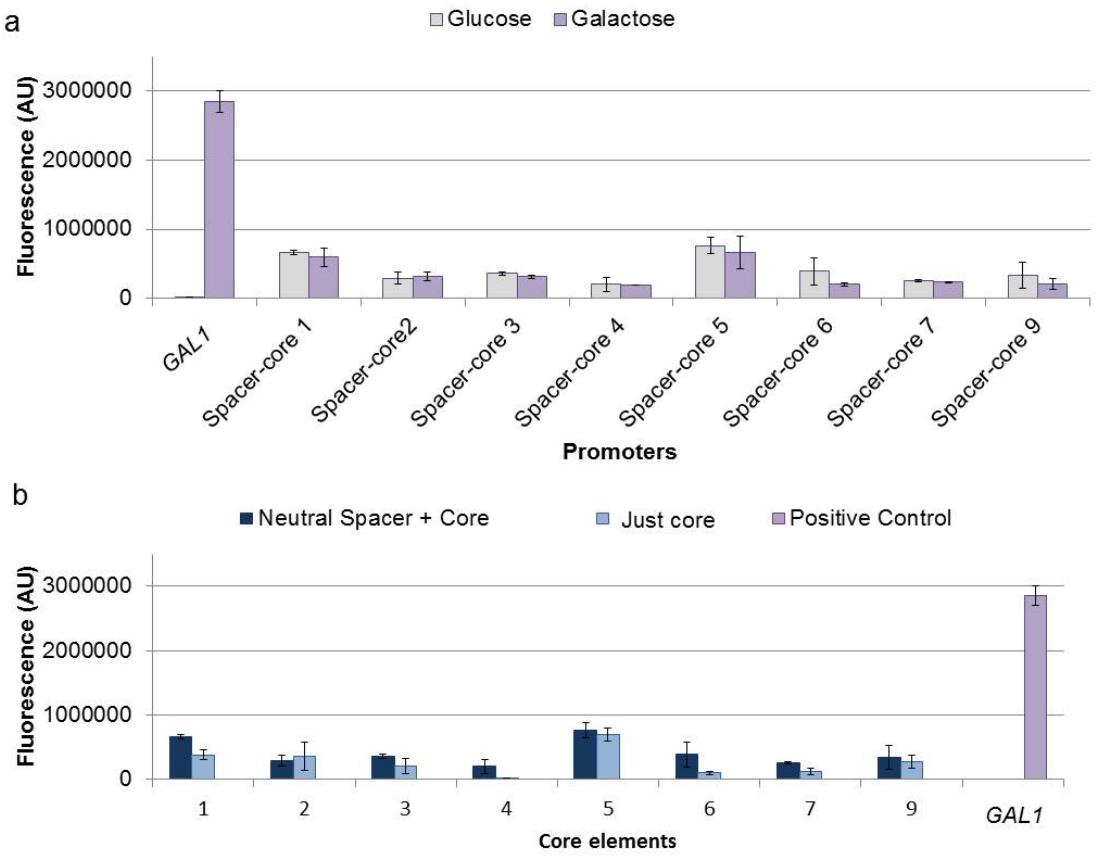


Figure 3.7: AT-rich spacer is neutral under glucose and galactose media for all core elements tested

AT-rich spacer was designed to be free of TFBS as determined by the TFBS database YEASTRACT[158] at the time of design, and free of TATA boxes and TATA-like boxes (up to 2 mismatches to TATAWAWR). (a) 30 bp neutral spacer does not respond to galactose induction (n=3). (b) 30 bp neutral spacer has little to no effect on the core element's expression under glucose (n=3). Error bars represent standard deviation among biological triplicates.

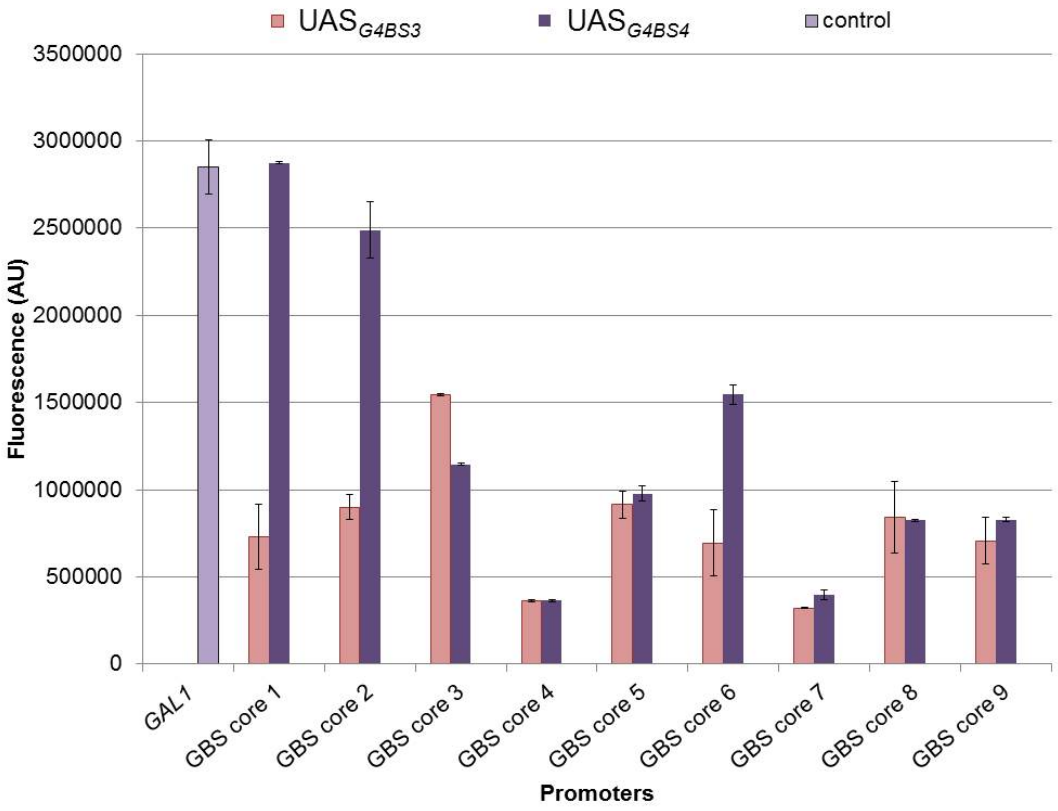


Figure 3.8: UAS$_{G4BS4}$ is a stronger amplifier than UAS$_{G4BS3}$ under 2% galactose

Induction by two Gal4p binding sites (GBS), Gal4BS3 (G4BS3) and Gal4BS4 (G4BS4) demonstrates both differential TFBS function as well as generic function of the core elements. *GAL1* promoter serves as the positive control. Error bars represent standard deviations among biological triplicates.

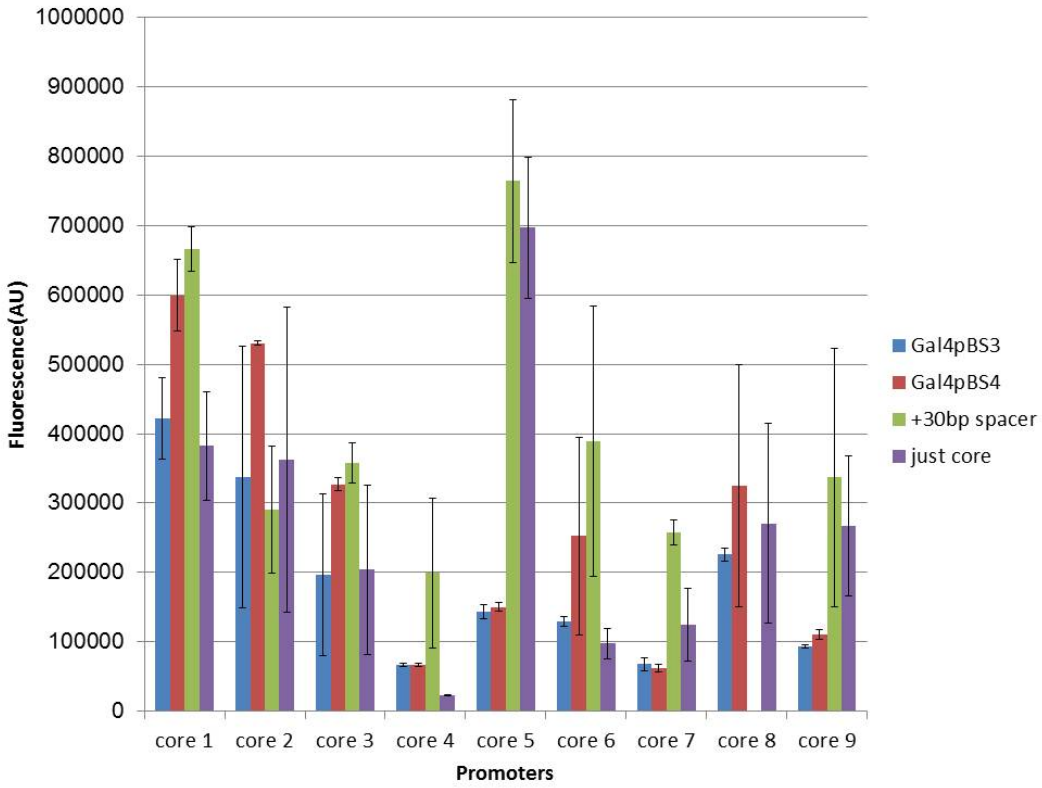


Figure 3.9: Effect of a Gal4 binding site on core elements under glucose

Both Gal4p binding sites, Gal4pBS3 and Gal4pBS4 have little to no effect on expression of core elements where glucose is sole carbon source. Error bars represent standard deviations among biological triplicates.

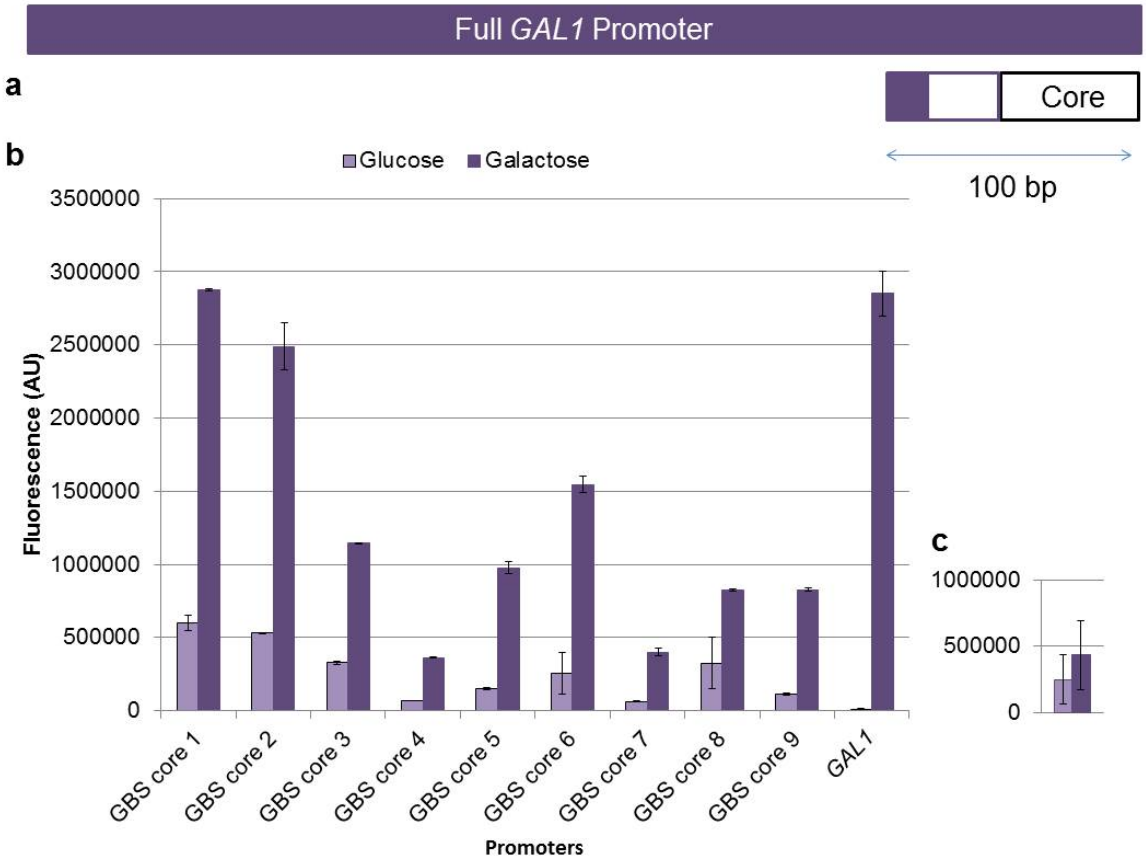


Figure 3.10: Core elements can be used to create strong inducible promoters

Core elements were paired with a Gal4p binding site (GBS). Specifically shown here is UAS_{G4BS4}. With the carbon source as galactose, promoters are induced. In some promoter pairings, promoter strength is that of full native *GAL1* promoter (b), but at a fraction of the length as shown in the scaled illustration (a). Of the 18 core elements tested in this construct, two did not activate with both UAS_{G4BS3} and UAS_{G4BS4} . For demonstrative purposes, one rejected core element is shown here (c). Error bars represent standard deviations among biological triplicates.

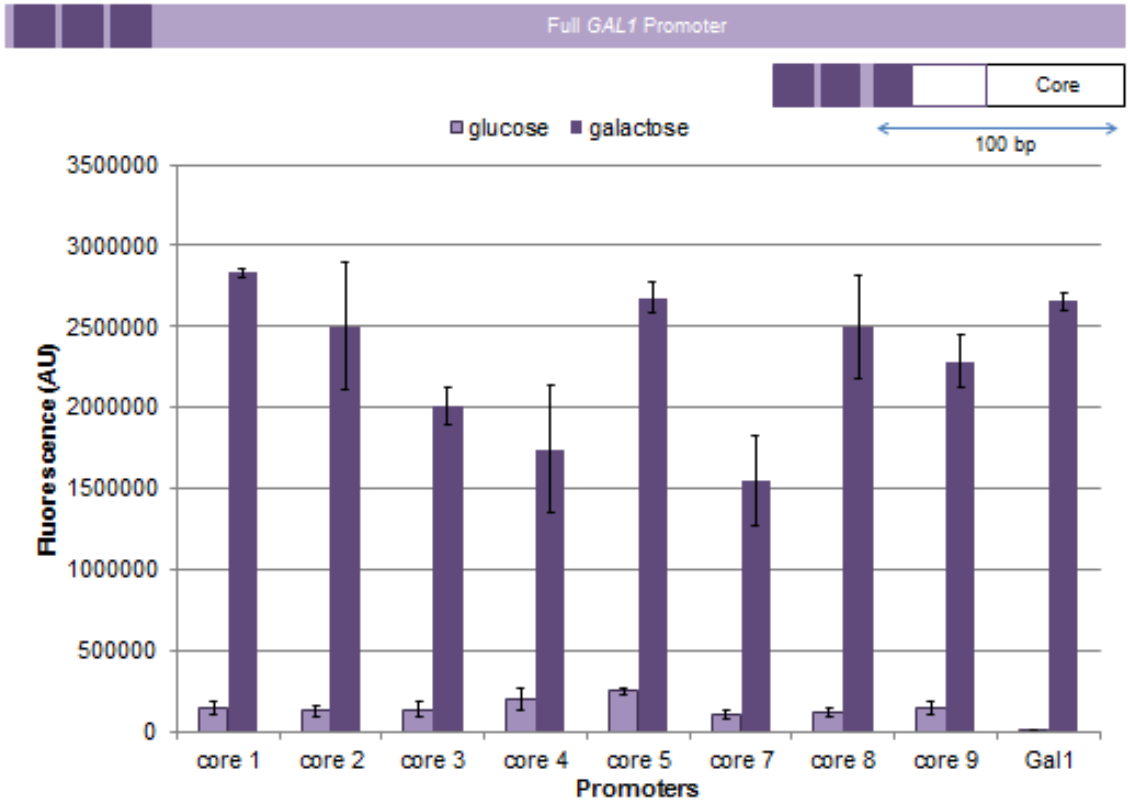


Figure 3.11: Multiple GBS provide strong induction

A 54 bp sequence stretch derived from the 5` end of the native *GAL1* promoter contains three Gal4p binding sites. These binding sites are represented by dark purple boxes. This 54 bp region was hybridized with core elements using an AT-rich spacer to distance the region from the core. Under galactose, promoters increase in strength approaching *GAL1*. Error bars represent standard deviation of biological triplicates. Promoters depicted as boxes are drawn to scale with respect to length. Error bars represent standard deviations among biological triplicates.

As a third robustness test, we sought to determine context dependency of core elements linked with UAS_{CIT} and UAS_{G4BS4}, as well as without any UAS (**Figure 3.1**). Promoters with limited context dependency provide a predictable and orthologous tool for simplified synthetic biology. Thus, to ensure our set of core elements is context

independent, we applied a simple test; we analyzed their expression strengths by flipping the entire expression cassette and re-introducing it back into the same plasmid location (**Figure 3.12**). Although this test does not confirm complete orthogonality, it certainly provides evidence for a more context independent promoter. Prior evidence within our lab supports the finding that native promoters (including *CYC1* shown in **Figure 3.12a**) can demonstrate drastically different transcriptional profiles in light of this test, thus highlighting the importance of developing context independent parts. Of the 18 putative core elements, five were determined to not function in a context independent manner making this test the most stringent one applied to the core elements (**Figure 3.12b**).

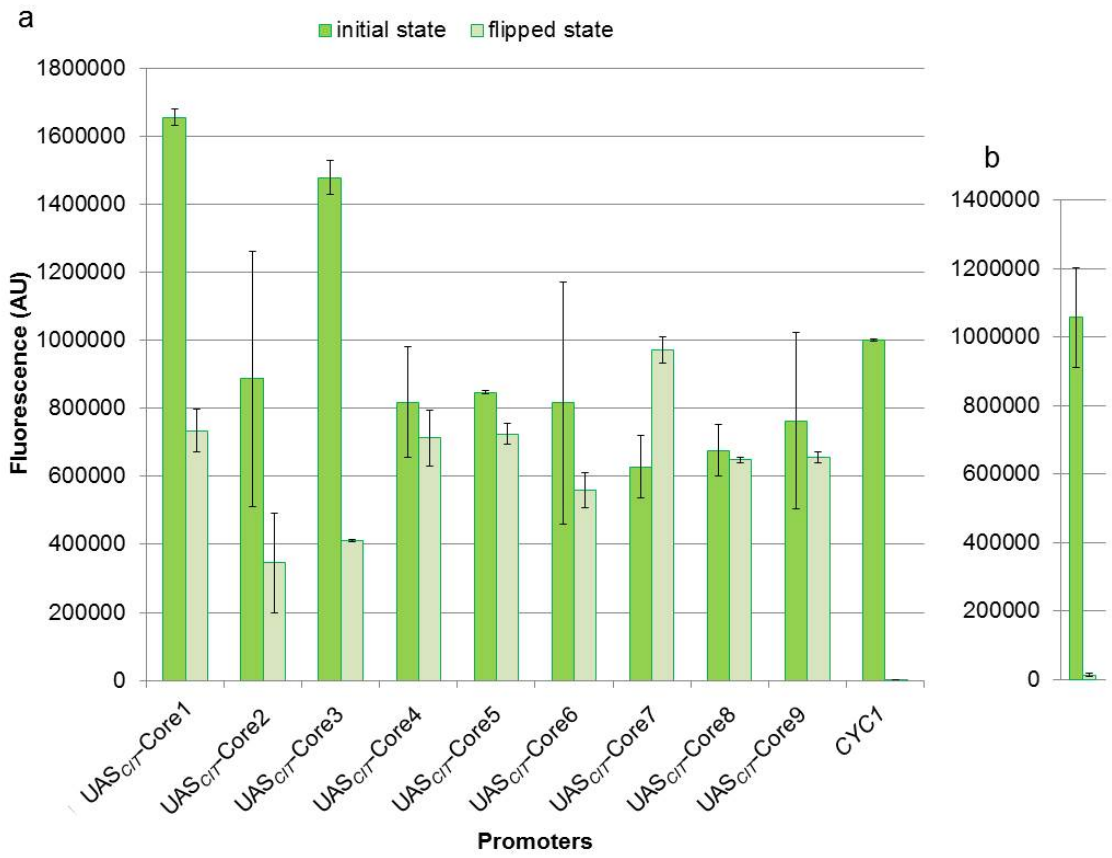


Figure 3.12: Promoters made using the core elements show less context specificity than commonly used CYC1 promoter

(a) *CYC1* promoter's function is completely abolished when positioned differently within plasmid, whereas most promoters showed little to no effect (n=3). Of the 18 core elements subjected to this test, four were determined to be context dependent and therefore, were rejected. For comparative purposes, one of these rejected core elements is shown (b) (n=3). Error bars represent standard deviation of biological replicates.

As a final robustness test, we sought to evaluate the impact of these core elements with different genes. Thus, an alternative reporter gene was used to ensure promoters could be successfully applied to other expression systems. Using *LacZ* as an alternative ORF, we demonstrate that all core elements with and without a UAS_{CIT} can induce

transcription of *LacZ* in a similar fashion as the fluorescent protein (**Figure 3.13**). Furthermore, we also show that fluorescence levels measured with flow cytometry correlate well with mRNA abundance levels indicating that the function of these elements is indeed at the transcription level (**Figure 3.14**). Thus, core elements isolated via this scheme were seen to function independently of the downstream gene.

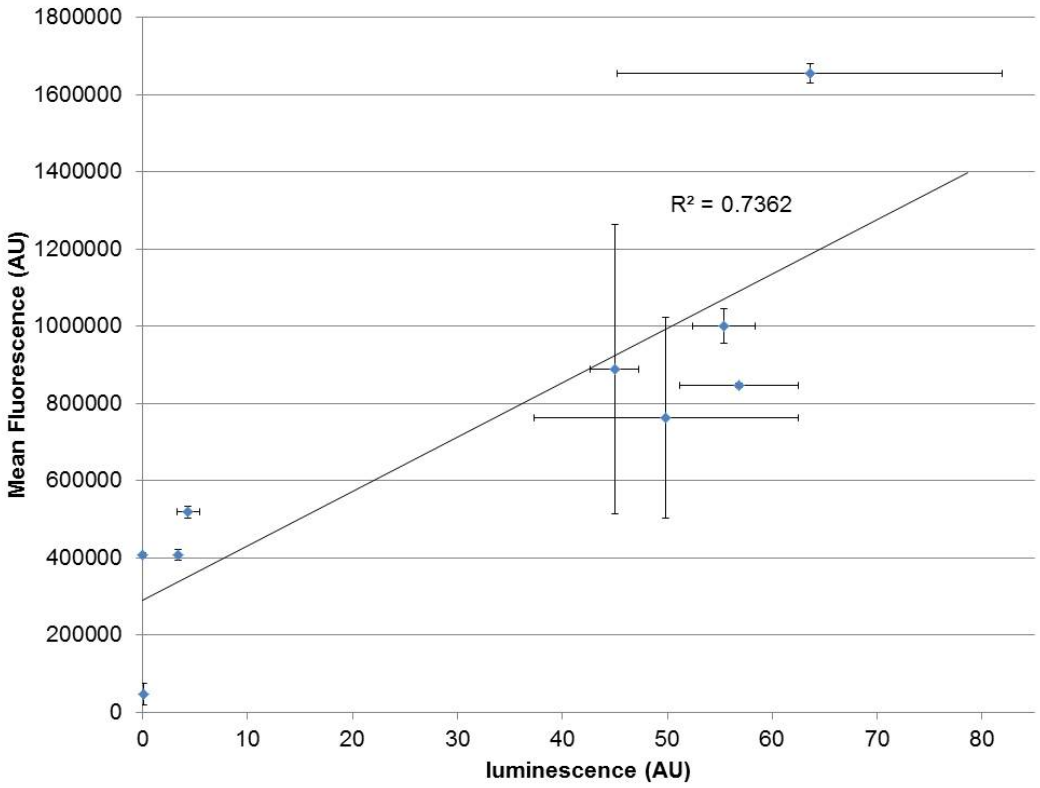


Figure 3.13: Mean fluorescence from flow cytometry parallels luminescence from LacZ assay under glucose growth

Select synthetic hybrid promoters were used to drive expression of β-galactosidase. Enzyme activity of β-galactosidase was quantified by measuring its product's luminescence. This data supports the premise that these promoter elements are independent of the gene being expressed (n=3). Error bars represent standard deviation of biological replicates.

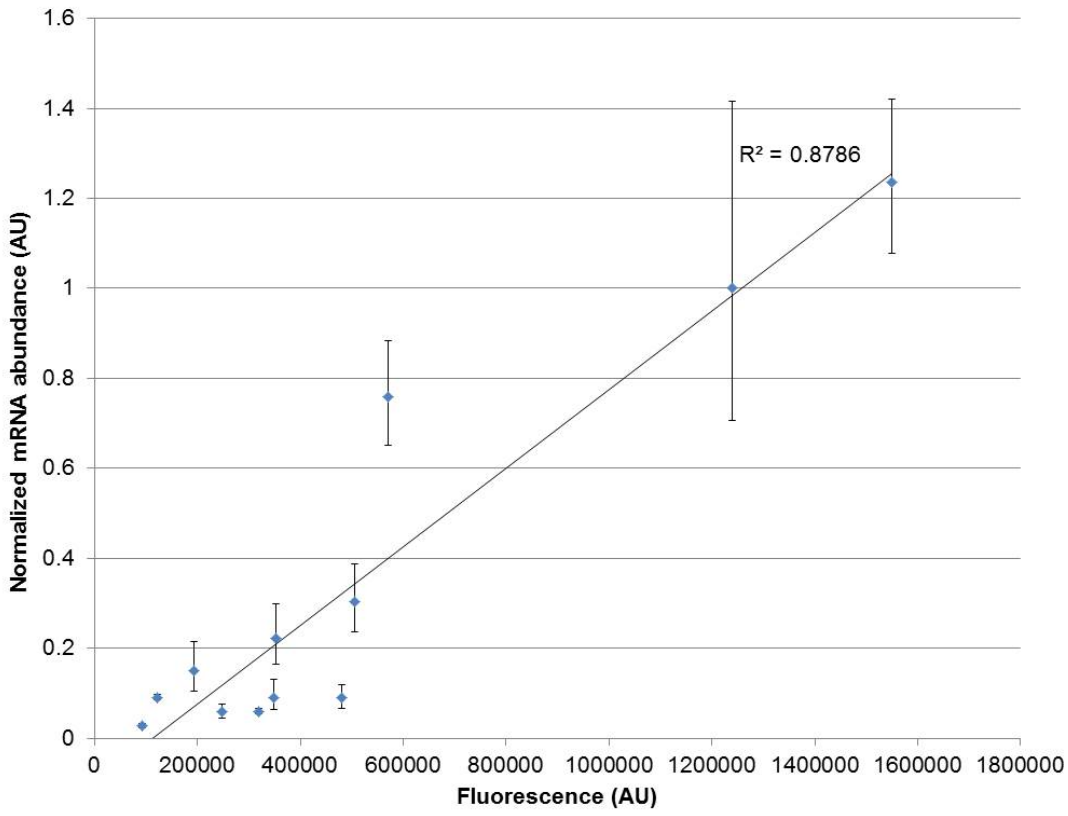


Figure 3.14: Fluorescence driven by synthetic promoters correlate with mRNA abundance

mRNA levels of fluorescent proteins driven by selected synthetic hybrid promoters were measured by qRT-PCR. mRNA levels correlate to fluorescence levels. This correlation supports the premise that these elements are functioning at the transcriptional level (n=3). Error bars represent standard deviation of technical triplicates.

3.3.5 Final set of 9 generic, minimal core elements

In the workflow described above, we sorted through 15 million elements to isolate 18 putative core elements. Through a series of robustness tests to isolate core elements with desirable characteristics which could be used in synthetic biology efforts, we narrowed our usable set to just nine core elements (**Figure 3.1**). This indicates that not

only do most random sequences between the TATA box and TSS fail to function properly, but that many core elements are unable to function in a generic manner (especially with respect to genetic context), highlighting the exceptionality of the core elements selected in this research. Although the core elements share common desirable characteristics at the phenotypic level, their sequences are quite dissimilar spanning a wide range of GC content from 47-70% (**Figure 3.15**), and possessing a diversity of TFBS, both in quantity, quality and directionality as assessed by the transcription TFBS database YEASTRACT[158] (**Figure 3.15**). Finally, the sequence homology to the genome is low among the set as none of them match any sequences found in the genome of *S. cerevisiae* (**Figure 3.15**). Thus, this set of 9 minimal core promoters are functionally similar and robust, yet sequence diverse.

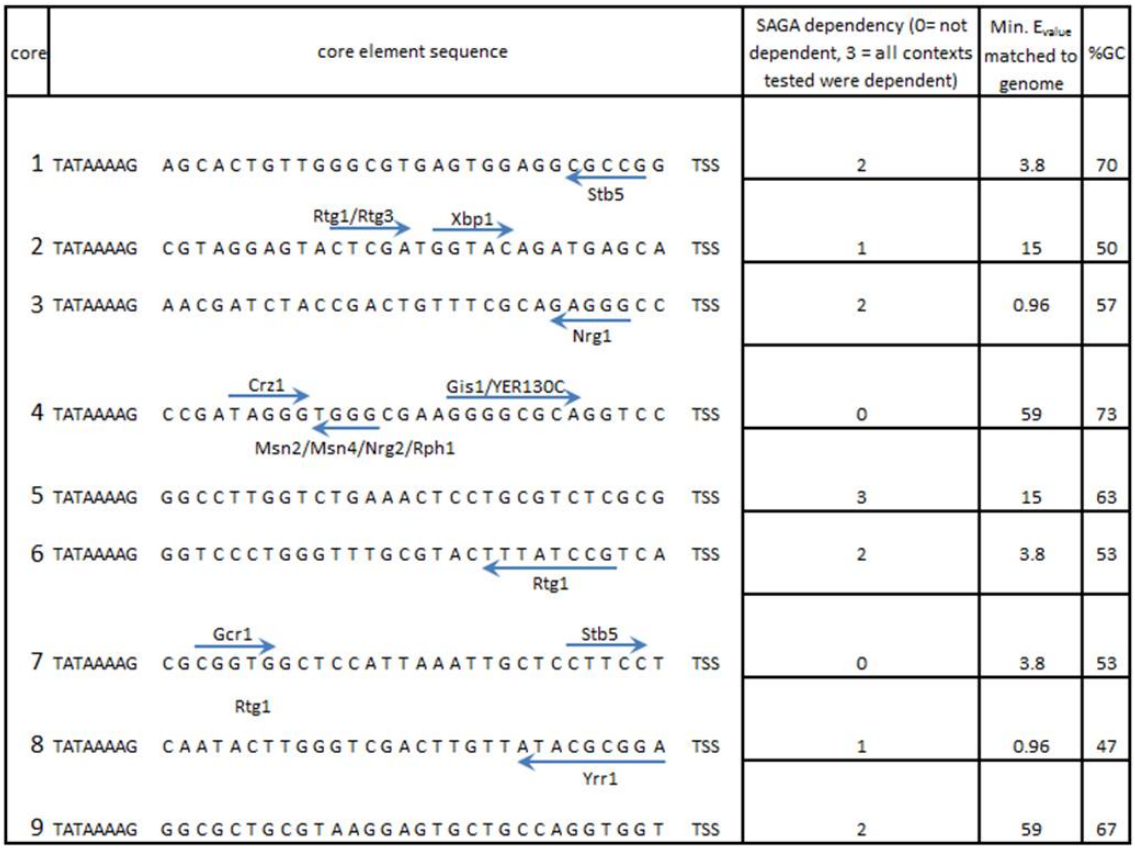


Figure 3.15: Minimal core elements and UAS sequences

Predicted TFBS are indicated by arrows where direction of arrow designates orientation of site. Forward TFBS above reference sequence and reverse TFBS below reference sequence. TF that bind to these sequences are labeled as such. (a) Core elements are distinct from one another spanning a %GC content of 47 to 73. The quantity, quality and directionality of TFBS as determined by the TFBS database YEASTRACT[158] vary greatly.

3.4 Discussions and Conclusions

Prior to this study, yeast promoters required hundreds of base pairs to achieve high transcriptional capacity. Through the methodical workflow developed here, it was possible to identify robust, minimal core elements that can be linked with native UAS and TFBS. By using just a single native TFBS, we were able to establish galactose

inducible promoters at lengths comparable to bacterial promoters. In doing so, we successful reduced the galactose inducible promoter by 75%. Moreover, the framework developed here can allow for continued advancements in minimization of orthologous synthetic promoters in fungal hosts.

The 9 minimal core elements described here were identified from a pool of 15 million candidates (**Figure 3.1**). On a sequence basis, each of these minimal core elements is distinct from both each other and the native *S. cerevisiae* sequence. To further elucidate potential distinct mechanisms of these elements, it was hypothesized that many of these synthetic promoters created with these core elements would use the SAGA complex. A critical component of the SAGA complex is the Spt3p subunit that is essential for transcriptional activation[159,160]. Testing the SAGA complex interaction of core elements hybridized with a UAS (UAS_{CIT}, UAS_{G4BS4}) and without a UAS was done in a *Δspt3* strain. Whereas the *Δspt3* mutation removed functionality of the *GAL1* native promoter as expected[161], the impact on each of these core elements by themselves, and assembled into assorted promoters varied (**Figure 3.15**). While this experiment does not conclusively prove linkage to Spt3p dependent transcription, the marked difference for each promoter suggest that different transcription initiation machinery is utilized across this set of core elements. This characteristic makes this set an excellent tool for generating diverse promoter engineering variants.

Finally, this workflow highlights the importance of robustness tests for promoters. Specifically, we find that generic and interoperable promoters require both robustness to

alternative UAS elements and especially genomic context. Through this work here, half of the putative core elements failed these robustness tests and were thus eliminated from the final collection. The most stringent test, context flipping, demonstrates the importance of genomic context in performance (and often, mischaracterization) of synthetic parts[162]. Such robustness analysis is important for future synthetic part evaluation. Moreover, the largest culling seen in the putative pool came at the colony analysis level. Specifically, constructs simply identified through FACS did not always function in a robust manner with respect to growth phase, homogenous expression, and re-transformation.

In this work, we simplified the core element and created a set of truly modular, context independent and completely synthetic core elements that will lay the framework for additional promoter engineering discussed in **Chapters 4 and 5**. Moreover, the generic workflow presented in this chapter can be followed in alternative fungal hosts to further expand their synthetic biology toolboxes.

Chapter 4: Construction and minimization of the UAS

4.1 Chapter Summary

With success at identifying minimal core elements, we sought to further minimize fungal promoters by shortening the UAS region of the promoter. A four-faceted approach was taken to isolate succinct and modular synthetic UAS that can activate the synthetic minimal core elements isolated and characterized in **Chapter 3**. First, distinct 10 bp UAS elements from a single library are isolated and tandem assembled to create a strong UAS. Second, synthetic minimal UAS elements in 10 bp increments are built using a library-based approach to obtain a variety of UAS elements that activate a synthetic core to a continuum of strengths. Third, unique G-rich sequences are isolated from a well-defined library, and fourth, a rationally designed dCAS9 promoter amplifier is applied to synthetic core elements. With these approaches, promoters as short as those annotated in *E. coli* are established, making these promoters the shortest constitutive synthetic promoters described to date in yeast.[3]

4.2 Introduction

In **Chapter 3**, the promoter core was minimized to 56 bp by decreasing the spacing between the TATA box and TSS. However, for native promoters, the UAS region is considerably longer than the promoter core. For commonly used constitutive promoters *GPD*, *CYC1* and *TEF1,* this region contributes approximately 87%, 79% and

---

[3] Part of the work presented in this chapter was published in "The development and characterization of synthetic minimal yeast promoters" *Nature Communications* (2015) 7810. Sequences of a select few of minimal UAS elements are patented under US patent #20160160299.

77% of the promoter DNA capacity respectively (assuming the 5` end of the core begins at the TATA-box or TATA-like sequence closest to the start codon). In terms of absolute DNA length, the UAS for *GPD*, *CYC1* and *TEF1* is 570, 228, and 308 bp respectively. Given that a single 17 bp TFBS can be used to amplify the synthetic core elements (**Figure 3.10**), I hypothesized that this region of hundreds of base pairs could be narrowed to just a few tens.

While the core contains the necessary sequences for PIC assembly, the UAS region is responsible for recruiting additional factors which influence the stability of the complex. In doing so, the UAS contributes constitutive and inducible qualities to the promoter. Native UAS' size is due to a wide range of TFBS distribution and other regulatory elements ranging from 100 to 500 bp upstream of the start codon[20,22]. At one extreme, our experiments in **Chapter 2** indicate the UAS can function at distances approaching 2,000 bp upstream of the start codon, while at the other extreme, as demonstrated in **Chapter 3**, a single TFBS for Gal4p spaced just 95 bp upstream of the start codon can amplify core elements. Interestingly, for all native promoters, a peak TFBS density has been observed at ~100 bp upstream of the TSS[18]. This suggests that Gal4pBS is not unique in the ability to function in close proximity. Furthermore, since most native promoters contain only one or two TFBS[19], and the average length for a TFBS[17] is ~10 nucleotides, the span of a UAS could theoretically be just tens of base pairs. Thus, it seems reasonable that a UAS can be significantly smaller than the hundreds of base pairs found in commonly used promoters.

Here, eight promoter libraries, both well-defined and highly-variable, and a rationally designed promoter are used to engineer a set of constitutive yeast promoters (**Figure 4.1a-e**). With these approaches, 72 completely synthetic minimal promoters driving a 70-fold expression range from strengths lower than the weak promoter *CYC1* to ones approaching that of the strong promoter *GPD* (*TDH3*) are assembled. Furthermore, a set of 23 *de novo* minimal UAS elements are established. By pairing these UAS elements with the synthetic short promoter cores described in **Chapter 3**, the shortest promoters described to date in a yeast system are assembled.

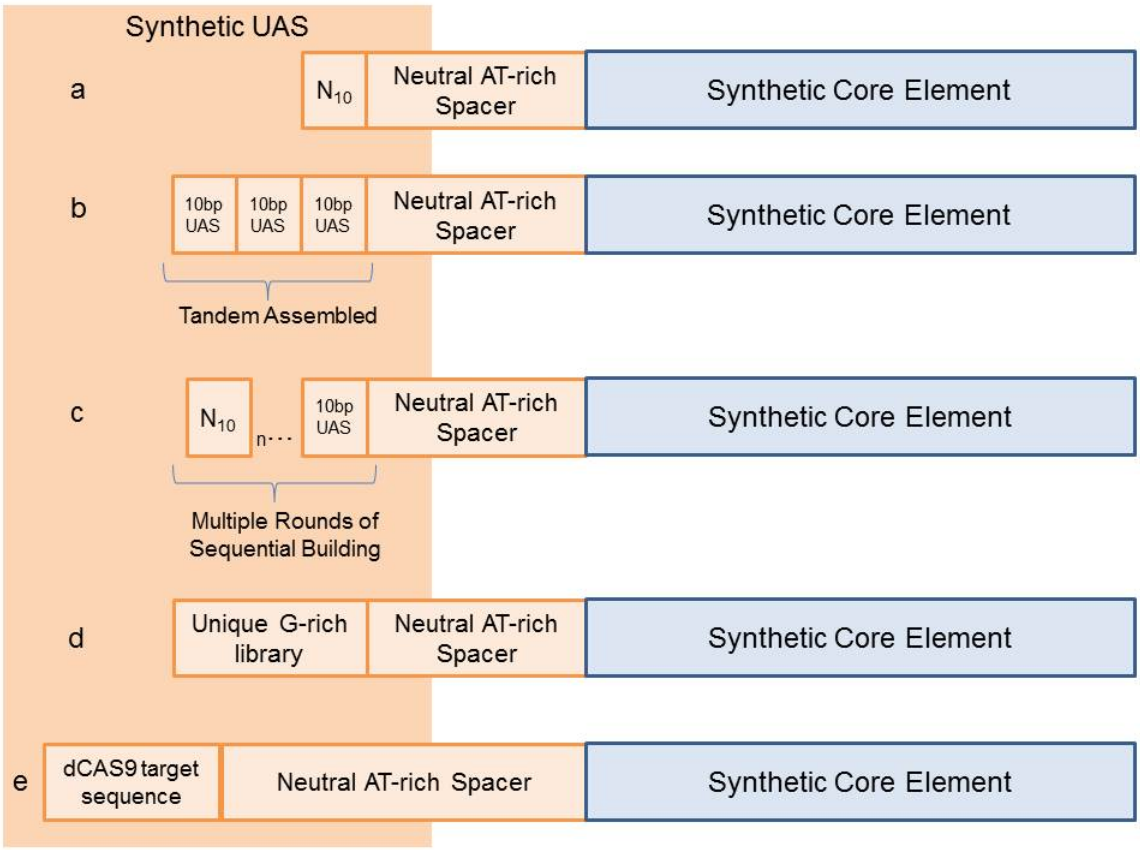


Figure 4.1: Four fold approach to synthetic minimal UAS engineering

Four approaches are used to engineer minimal UAS elements. (a) Functional 10 bp UAS elements are isolated from a N_{10}-core 1 library and then, (b) tandem assembled to create stronger promoters. (c) Functional and longer UAS elements are isolated from N_{10}-UAS-core 1 libraries using previously established synthetic minimal UAS elements. (d) A well-defined unique G-rich library is used as source of UAS elements. (e) A modified dCAS9-based sTF is harnessed to amplify synthetic cores via dCAS9 target sequence. In all promoters, an AT-rich neutral spacer is utilized to distance UAS elements from the core region. This is done to avoid steric occlusion of potential activating factors. All boxes are to scale where the synthetic core element represents 56 bp.

4.3 RESULTS

4.3.1 Construction of 10 bp synthetic minimal UAS element library

I hypothesized that the minimal length for an UAS element is the span of a single TFBS. Since most TFBS are within 10 bp[17], I reasoned a random N_{10} library hybridized with a promoter core could serve as a source of minimal UAS elements (**Figure 4.1a**). In this library construction, synthetic core 1 was chosen to serve as the promoter core for the core's robust activation as demonstrated in **Chapter 3**. To enable sufficient binding of a potential TF to the UAS library, the same AT-rich neutral spacer used to distance the TFBS for Gal4p from the core (**Figure 3.6**) as described in **Chapter 3** was employed. As a result, the promoter library scheme follows the general architecture of native yeast promoters; the potential TFBS is positioned ~100 bp from the TSS[18,163].

4.3.2 Isolation and characterization of 10 bp synthetic minimal UAS elements

Using yECitrine as a fluorescent reporter, N_{10}-core 1 library of 1.3 million candidates was sorted by FACS (**Figure 4.2**). The top 0.15% of fluorescent cells from one million measured fluorescent events were collected. This enriched pool is estimated to be composed of 140 UAS candidates. The strength of these candidates was analyzed. To do so, the FACS enriched pool of cells was plated and 119 colonies were picked and grown into liquid minimal media. Strength of the UAS candidates in these liquid cultures was determined by flow cytometry. Plasmids from 28 highly fluorescent cultures were isolated and sequenced. All 28 plasmids were retransformed into fresh BY4741 background, and fluorescence of biological triplicates in liquid culture was measured by

flow cytometry to confirm strength. Of the 28 candidates, only six were found to activate core 1. In all, from a pool of 1.3 million candidates, six minimal UAS elements were identified. These six UAS are referred to as UAS_A, UAS_B, UAS_C, UAS_D, UAS_E, and UAS_F. When hybridized with core 1, rank in activating strength of minimal UAS elements is as follows: $UAS_C > UAS_B > UAS_E > UAS_A \approx UAS_D \approx UAS_F$ (**Figure 4.3**). In fact, minimal UAS_C activates core 1 better than native UAS_{CIT}. Thus, this strategy was proven to be effective in isolating minimal UAS elements that activate a synthetic core to high promoter outputs.

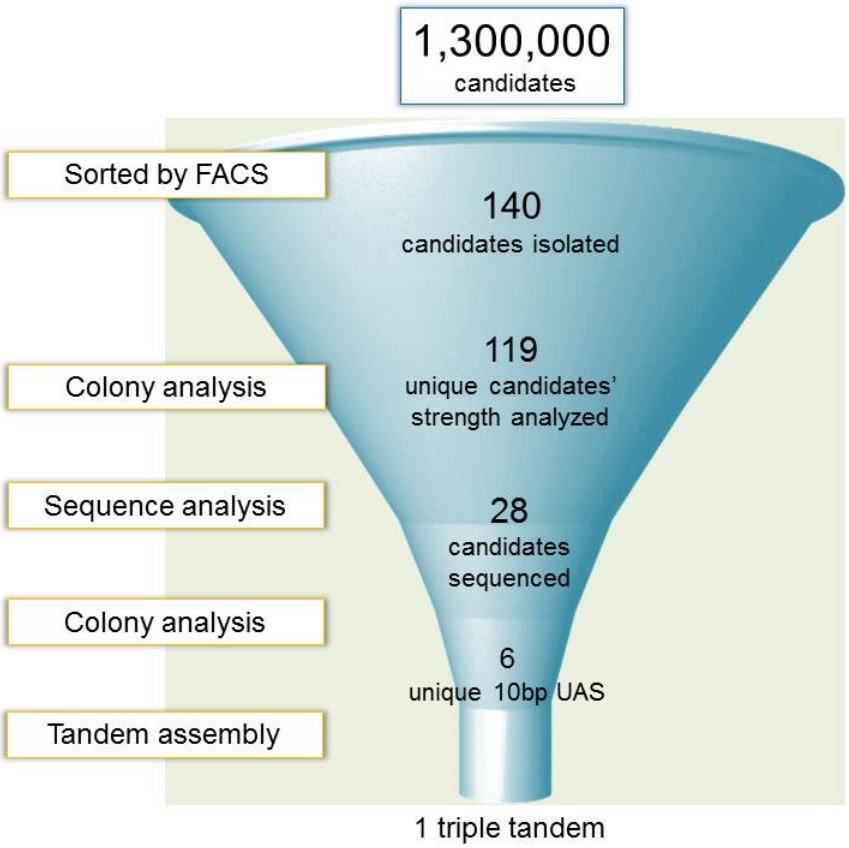


Figure 4.2: Isolation and tandem assembly of 10 bp UAS

One library of 1.3 million UAS candidates was sorted by FACS. Colony analysis was performed on 119 candidates resulting in 28 apparent high functioning candidates. Following sequence analysis and additional characterizations, candidates were narrowed to just a pool of six UAS. Of these six, three were assembled in tandem to yield a high strength UAS element.

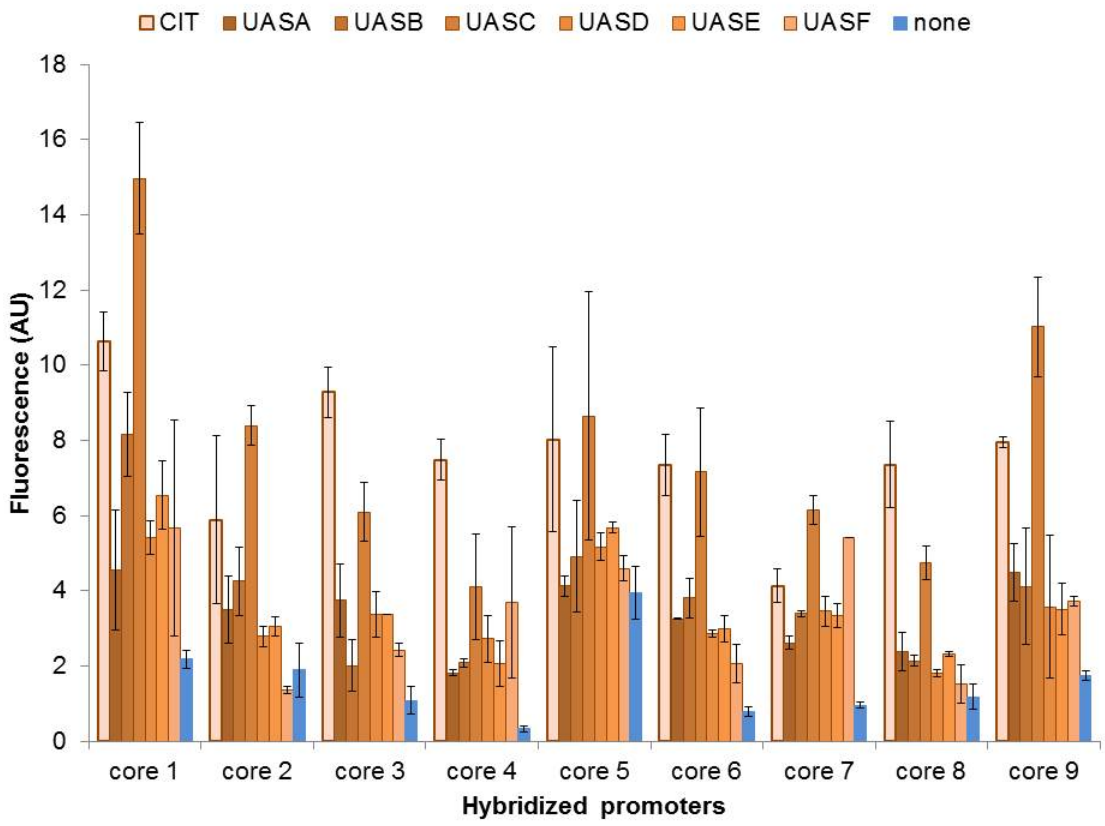


Figure 4.3: 10 bp UAS activate other synthetic core elements

10 bp UAS elements derived from a N_{20} core 1 hybrid promoter library function with other synthetic core elements. Activation by UAS elements is not consistent across core elements indicating activation bias of the UAS elements for promoter cores. Error bars represent standard deviation among biological triplicates.

To investigate whether the identified minimal UAS would activate other cores, all six UAS elements were hybridized with all nine synthetic core elements to create 36 promoters. Promoter strength was assessed with flow cytometry. With this experiment, I demonstrated that these minimal UAS can independently activate many of our minimized core elements (**Figure 4.3**). However, the above mentioned ranking of UAS elements is not consistent across core elements, indicating synergy between certain pairs of UAS and

cores. Core 8 was the least functional with this set of UAS elements, while core 1 was the most. Core 1 being the most functional is not surprising; core 1 served as the promoter core in the UAS library scaffold. Thus, there appears to be a bias of the UAS elements for the specific core element utilized in the library. Nonetheless, UAS_C was the strongest in every promoter context regardless of core element present, suggesting it is possible to have an UAS element that is completely modular with respect to the nine synthetic core elements.

4.3.3 Tandem assembly of 10 bp UAS elements to establish stronger UAS

Next, we sought to demonstrate how these functional minimal UAS elements can be linked in tandem and then hybridized with core elements to establish high strength, short promoters in yeast (**Figure 4.1b**). To do so, the six UAS elements were tandem assembled to construct 14 unique, longer UAS. These sequences were comprised of two, three or four 10 bp minimal UAS elements. Specifically, we investigated the functionality of tandem assembled UAS with one another, and multiples of the same UAS. We also investigated the arrangement of UAS elements in a tandem assembly. To evaluate their core activating capabilities, all of the tandem-assembled UAS were hybridized with synthetic core 1 to create full length promoters. To confirm the results were not specific to core 1, a select few of tandem-assembled UAS were also hybridized with synthetic core 2. Resulting promoters were used to drive expression of yECitrine, and promoter strength was assessed with flow cytometry.

None of the multiples of the same UAS tested resulted in stronger promoters (**Figure 4.4a**), and the arrangement of UAS severely affected the promoter output (**Figure 4.4b**). In all, of the 14 UAS constructs tested, four resulted in stronger promoters with each additional 10 bp UAS element (**Figure 4.4-4.5**). Ten showed no increase in activation with additional UAS (**Figure 4.4-4.5**), while three actually resulted in a weaker promoter (**Figure 4.4b**). For the select few tandem-assembled UAS tested with another core element, we demonstrate that the relative strengths of tandem-assembled UAS elements are the same (**Figure 4.5a-b**). This suggests the observed failure of additional UAS elements to increase transcriptional output is not due to the specific core element utilized in the promoter construct. Of the four UAS that elicited a stronger promoter, one was exceptionally strong. A triple tandem-assembled UAS activated core element 1 to 70% of the strength of the strong constitutive yeast promoter, *GPD* (*TDH3*) (**Figure 4.6a-b)**. Comprised of just three 10 bp minimal UAS, this high strength tandem assembly is just 18% of the DNA space of *GPD* (*THD3*) (**Figure 4.6a**). This experiment highlights the utility of UAS tandem-assembly even in the construction of hybrid promoters comprised of entirely synthetic parts.

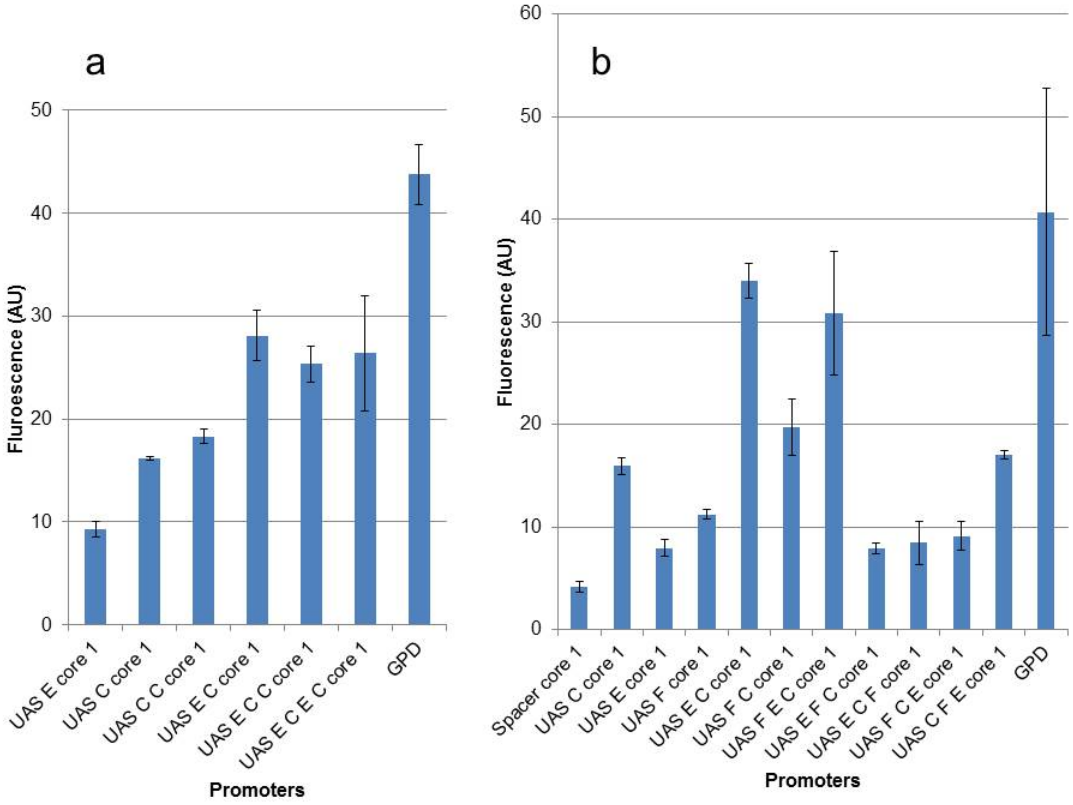


Figure 4.4: Tandem assembly of minimal UAS

Minimal UAS elements are tandem assembled and hybridized with core 1 to yield full length promoters. Flow cytometry data for each pane was gathered in the same experiment. Error bars represent standard deviation among biological triplicates.

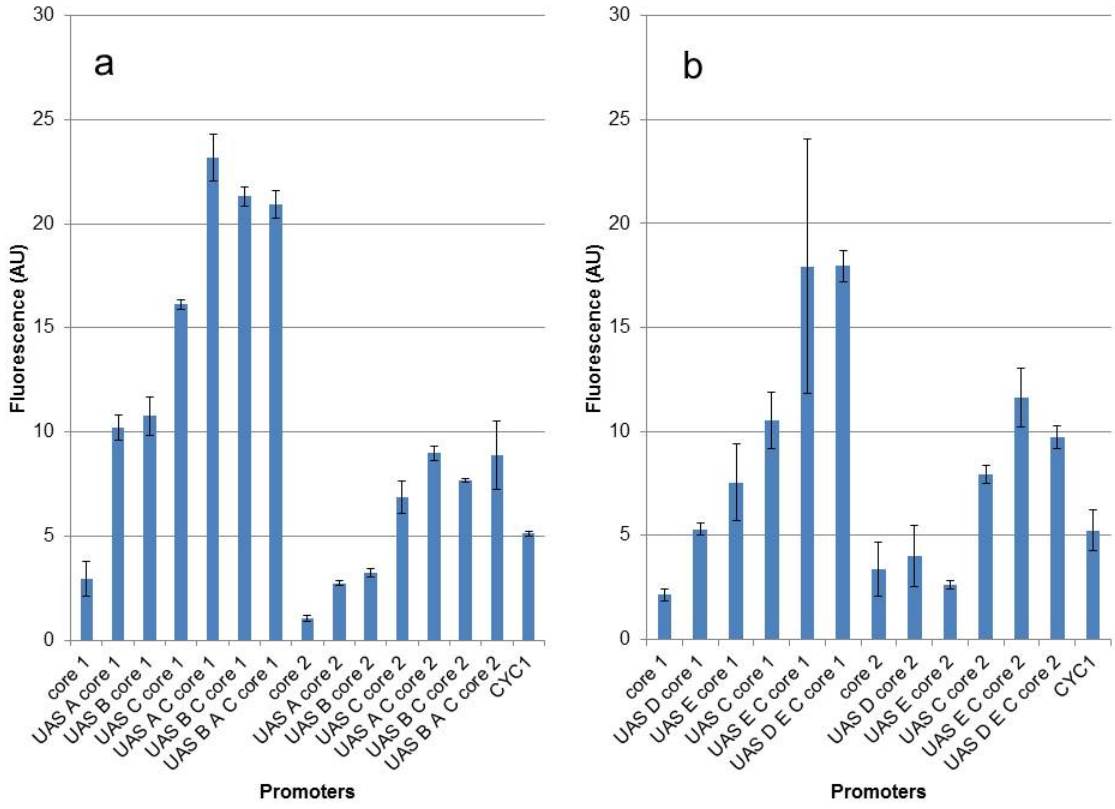


Figure 4.5: Tandem assembled UAS hybridized with two different cores

Minimal UAS elements are tandem assembled and hybridized with core 1 and core 2 to yield full length promoters. Flow cytometry data for each pane was gathered in the same experiment. Error bars represent standard deviation among biological triplicates.

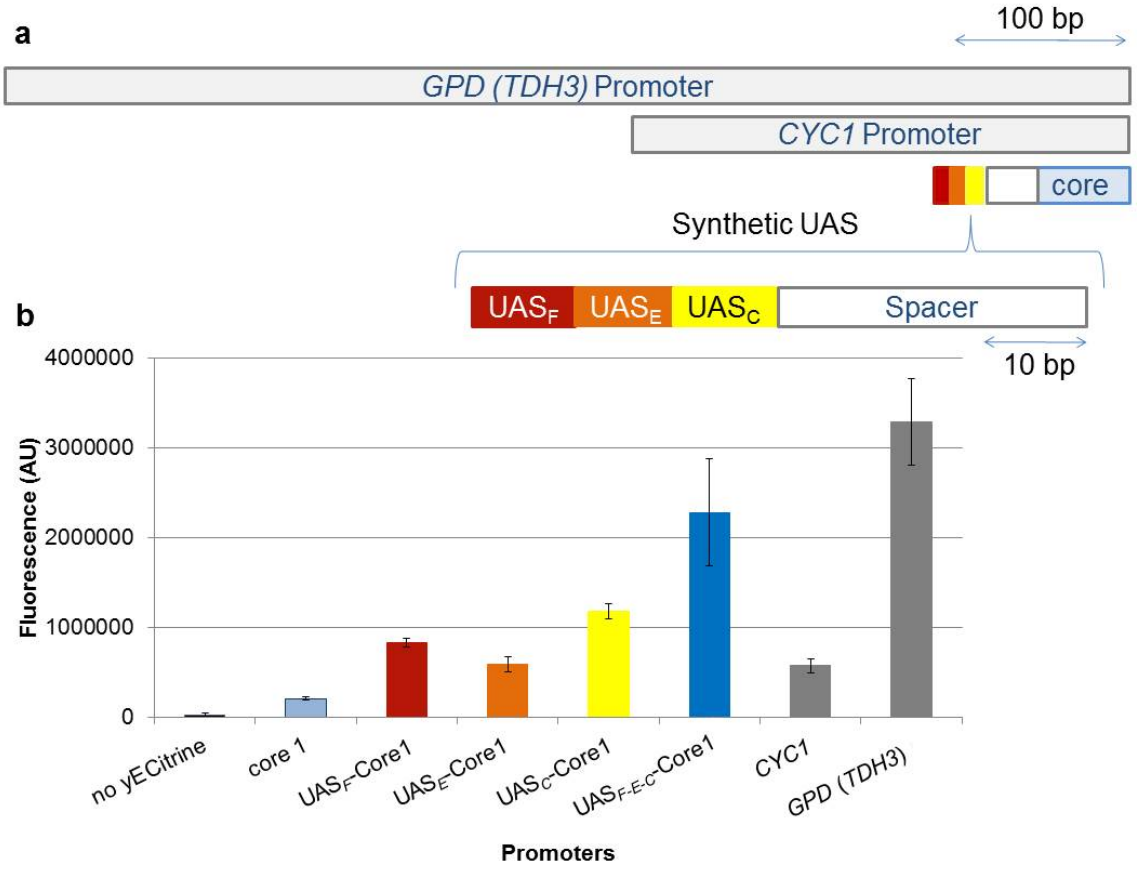


Figure 4.6: Synthetic tandem assembled UAS can activate core elements to yield high strength constitutive promoters

(a) Length of promoters is illustrated to scale. All synthetic UAS (UAS$_F$, UAS$_E$ and UAS$_C$) are positioned upstream of core element using an AT-rich neutral 30 bp spacer. (b) Synthetic UAS elements, UAS$_F$, UAS$_E$ and UAS$_C$ can activate synthetic core 1 to the strength of promoter *CYC1*. When UAS elements are tandem-assembled and then hybridized to core 1, strengths approaching *GPD* (*TDH3*) can be obtained. This data was published in *Nature Communications* in 2015[142]. Error bars represent standard deviation among biological triplicates.

4.3.4 Employment of 10 bp UAS elements in promoter library scaffolds

Although we were able to tandem assemble 10 bp UAS elements to yield a single strong UAS (**Figure 4.6b**), most combinations of tandem assemblies did not result in higher activation of core 1 and core 2 (**Figure 4.4-4.5**). We reasoned that individually

building on the six unique 10 bp UAS elements with tailored upstream sequences may be a more successful approach to construct stronger promoters (**Figure 4.1c**). To custom build on the 10 bp UAS elements, a N_{10} library was tandem assembled with each minimal UAS element. These linked UAS were then hybridized with core 1 to construct six hybrid promoter libraries of N_{10}-UAS-core 1. By utilizing libraries to build on the minimal UAS elements, cooperative sequences are allowed to be isolated. These sequences may reinstate a cryptic TFBS perturbed during hybridization, strengthen binding of the recruited TF to the minimal UAS element, and/or install a spacing sequence in the event another TFBS is present in the library sequence space.

To avoid disrupting putative TFBS identified by YEASTRACT database [158], N_{10} was hybridized upstream of these identified sites (**Figure 4.7a**). Specifically, nucleotides predicted to participate in a TFBS that overlapped with the 5` region upstream of the minimal UAS were included the N_{10}-UAS-core library. To keep UAS size to a minimum, all but one of the nucleotides of the TFBS were included in the library. Thus, if the TFBS was indeed being utilized it would still be present in 25% of the resulting library. For UAS without predicted overlapping TFBS (UAS_E and UAS_C), N_{10} was hybridized immediately upstream. Constructed UAS libraries were hybridized with core 1 and resulting promoter libraries were used to drive expression of yECitrine.

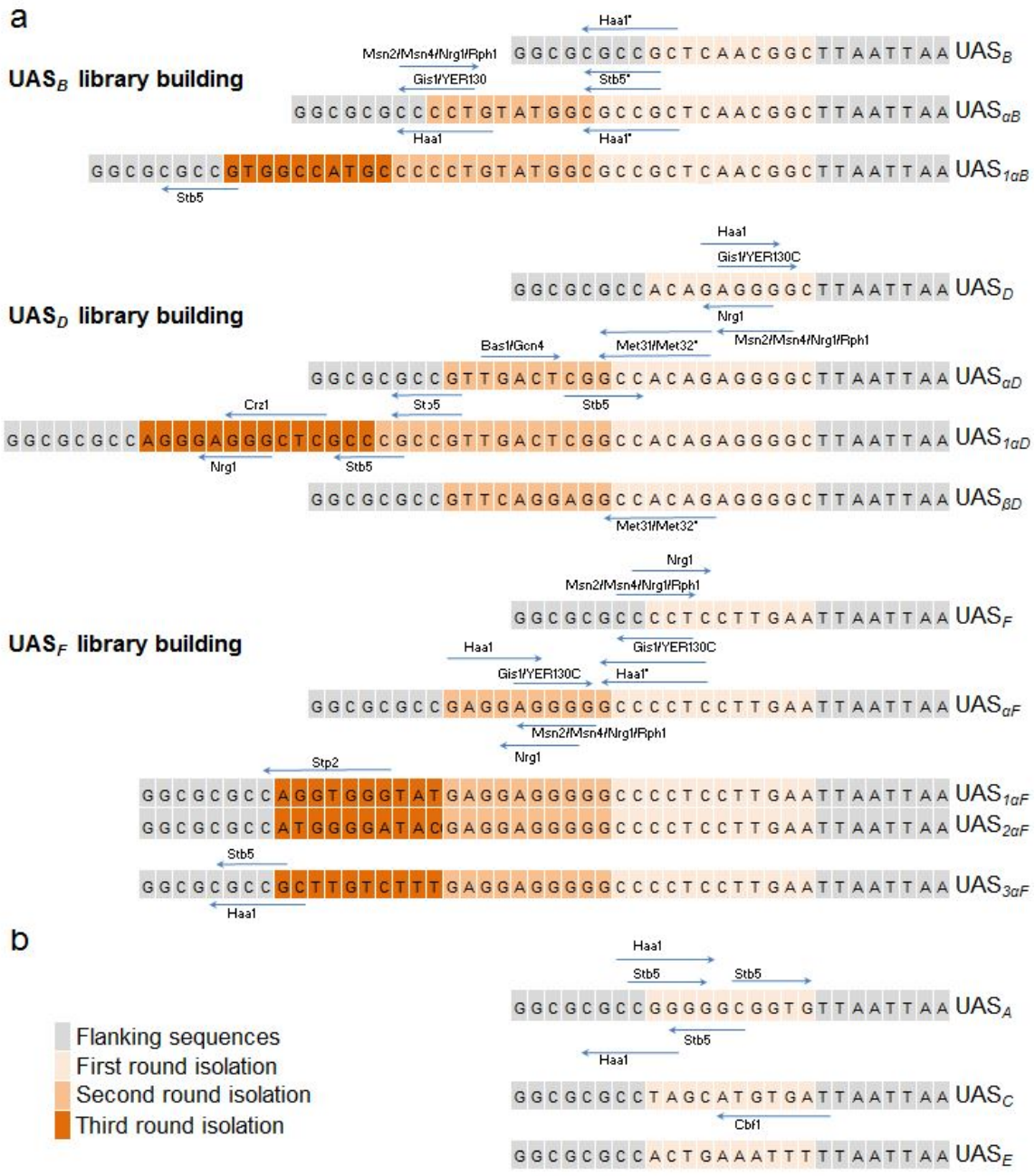


Figure 4.7: Predicted TFBS of synthetic minimal UAS elements

Figure 4.7: Predicted TFBS of synthetic minimal UAS elements continued

Numerous TFBS were predicted using YEASTRACT database[158]. Arrows indicate length, placement and direction of TFBS within the UAS. TF predicted to bind to these sequences are as labeled. (a) Of the six 10 bp minimal UAS, only three (UAS_B, UAS_D, and UAS_F) could be used to construct stronger minimal promoters. Color of sequence indicates origin of sequence. Sites identified in first round of isolation are still predicted to be present in second round of isolation despite lack of annotation. Some TFBS were predicted to overlap into flanking regions (upstream and downstream sequences). For some of these sites, they were disrupted with N_{10} library additions, but were regained during the next round of enrichment. Such sites are indicated with *. For example, predicted Met31p/Met32p binding site in UAS_D sequence appeared again in both $UAS_{\alpha D}$ and $UAS_{\beta D}$. (b) UAS that did not function in a N_{10}-UAS-core scaffold.

4.3.5 Isolation of stronger UAS from tandem-assembled libraries

Using FACS, the top 0.15% of fluorescent cells from 1 million fluorescent events was pooled from each N_{10}-UAS-core hybrid promoter library. The enriched pool was plated, 96 colonies were picked and flow cytometry was performed on liquid cultures to identify high strength individuals. High strength candidates were retransformed and promoter strength was confirmed in biological triplicate by flow cytometry.

Four stronger promoters were successfully isolated from hybrid promoter libraries constructed with UAS_B, UAS_D and UAS_F (**Figure 4.7a**). However, stronger promoters could not be isolated from hybrid promoter libraries comprised of UAS_A, UAS_C and UAS_E (**Figure 4.7b**). Newly identified promoters are now comprised of a UAS region of roughly 20 bp and synthetic core 1. To identify even stronger promoters, UAS regions ($UAS_{\alpha B}$, $UAS_{\alpha D}$, $UAS_{\alpha F}$, and $UAS_{\beta D}$) from these four promoters were employed in additional libraries of N_{10}-UAS-core. As with the first construction, putative 5` overlapping TFBS in the UAS region were preserved when hybridizing N_{10} upstream of the UAS element. Stronger promoters were successfully isolated from hybrid promoter libraries constructed with $UAS_{\alpha B}$, $UAS_{\alpha D}$, and $UAS_{\alpha F}$ (**Figure 4.7a**). However, stronger promoters could not be established from the hybrid promoter library N_{10}-$UAS_{\beta D}$-Core1 (**Figure 4.7a**). In all, nine UAS elements capable of activating synthetic core 1 were isolated and characterized from six hybrid promoter libraries of N_{10}-UAS-core. Five of the nine UAS activated core 1 significantly better than native UAS_{CLB} (**Figure 4.8-4.9**).

One UAS was particularly strong; $UAS_{1\alpha D}$ activated core 1 to 68% of *GPD* strength (**Figure 4.8-4.9**).

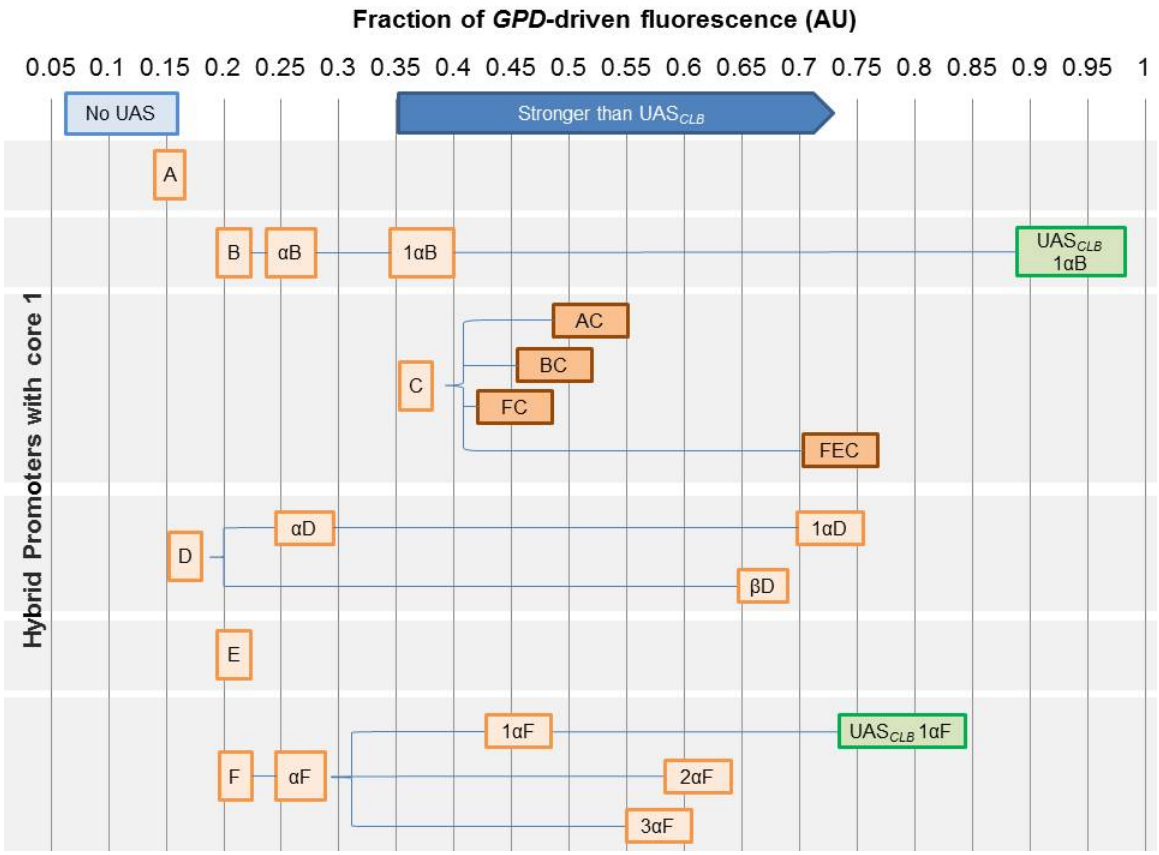


Figure 4.8: Construction of stronger UAS

10 bp UAS elements derived from a N_{10} synthetic core hybrid promoter library were used to build longer and stronger UAS. 10 bp UAS were employed in N_{10}-UAS-core library scaffolds (light orange boxes). UAS_B, UAS_D, and UAS_F were successfully used in two rounds of N_{10}-UAS-core library scaffolds. UAS_A, UAS_C, and UAS_E did not produce functional promoters in N_{10}-UAS-core library scaffolds. UAS_C could be tandem assembled with other 10 bp UAS (dark orange boxes). Promoter strength is indicated by the placement of the left hand side of the box (ie. UAS_C is 0.35 of *GPD* strength). Nomenclature refers to rounds applied (ie. 1αD and 1αF are not identical in sequence, but rather have undergone two rounds in a library scaffold. Name assignment (α, β, 1, 2) is arbitrary.) Two UAS, $UAS_{1\alpha B}$ and $UAS_{1\alpha F}$ could be further amplified with UAS_{CLB} (green boxes). No UAS is defined as spacer-core 1.

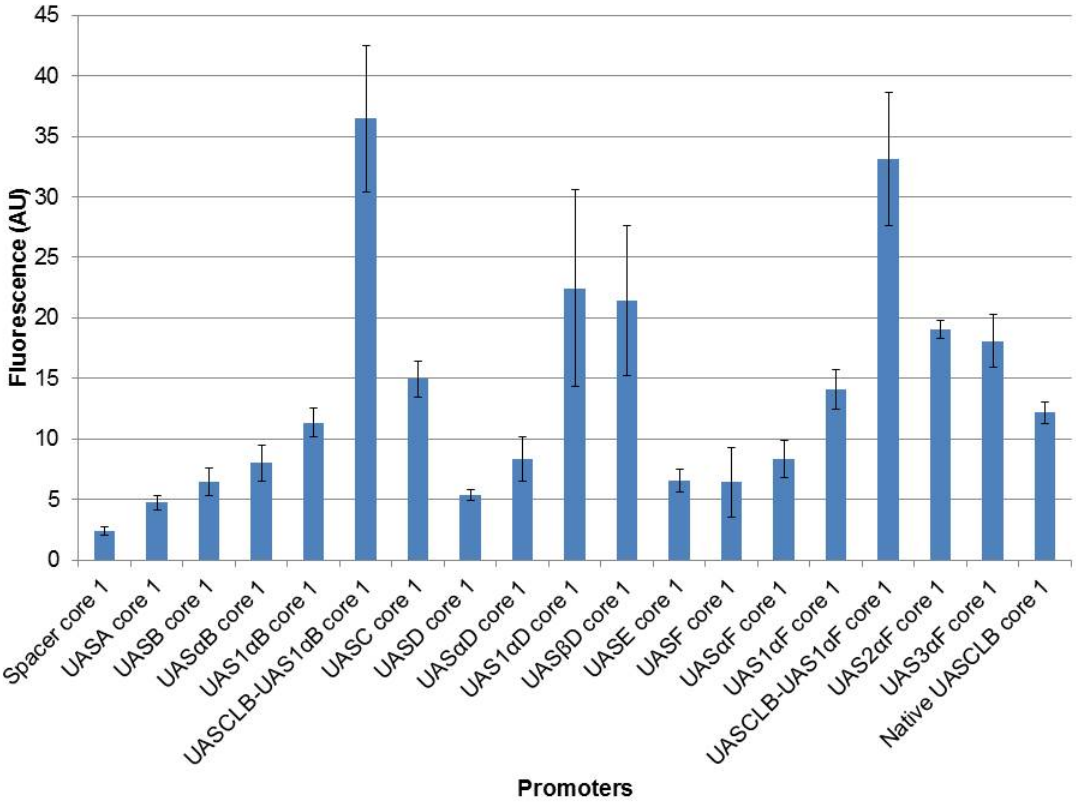


Figure 4.9: Strength of UAS elements

Strength of hybrid promoters is assessed by fluorescence. Error bars represent standard deviation among biological triplicates. Nomenclature refers to rounds applied (ie. 1αD and 1αF are not identical in sequence, but rather have undergone two rounds in a library scaffold. Name assignment (α, β, 1, 2) is arbitrary.)

While this approach was able to isolate strong UAS elements that function with a short minimal core, the resulting set has high sequence homology. For example, the sequences of $UAS_{1\alpha F,}$ $UAS_{2\alpha F,}$ and $UAS_{3\alpha F}$ are ~70% identical since all of these elements contain $UAS_{\alpha F}$ sequence. Furthermore, the success rate of this approach to establish novel UAS elements was rather low. First, only half of the six 10 bp UAS elements could be used in a N_{10}-UAS-core hybrid promoter library. Second, even after exhaustive

library screening (sorting through 1 million measured fluorescent events and characterization of 96 selected colonies from each enriched pool), only one to three promoters could be established from each library. Thus, we sought to use another approach to increase the promoter toolbox with UAS elements.

4.3.6 Construction of G-rich library

Libraries discussed thus far have been rather simple. A region upstream of the core is completely randomized. In addition to keeping the size to a minimum, these randomized regions were also kept short to minimize the size of the library, thereby enabling a more complete screening of all possible sequences. For example, a randomized region of 10 base pairs yields roughly one million possible sequences. However, I reasoned that the number of these sequences that could activate our synthetic core may be too small to isolate with our current isolation scheme. Thus, I sought to use a more defined library sequence containing a higher percentage of sequences that could potentially activate our synthetic core. I wanted a library that was still relatively short in length (<30 bp) with a limited number of possible sequences (<10 million). To do so, I turned to sequences with a high potential to form G-quadruplexes (GQ) *in vivo*. Studies show G-rich sequences that have the potential to form complex secondary structures called GQ are highly enriched in genomic promoters[25]. Since they are well conserved across yeast species[26], I hypothesized they have the potential for cross-species functionality. Thus, G-rich sequences poised to be an intriguing source of UAS elements.

While their mechanistic role in promoter regions is largely unknown, I hypothesized their short sequence could be exploited as a minimal UAS element.

There exists great flexibility in the kinds of sequences that can form G-quadraplexes. In general, potential GQ-forming sequences require at least four guanine tracts separated by linkers[164]. Tracts are defined as three or more guanine residues in succession with longer tracts being more stable[164]. For our library, tracts of three guanines were chosen to keep overall length to a minimum. Linkers range from one to seven nucleotides with shorter conferring increased stability[164,165], and greater variation in topology[166]. Linkers were randomized and a consistent linker length of 3 bp was chosen to construct a library of appropriate size. A randomized region shorter than 3 bp would yield a small number of possible sequences. To further aid in flexibility, flanking regions of the G-rich sequence were also randomized. Thus, the UAS library had the following sequence: $NG_3N_3G_3N_3G_3N_3G_3N$. This library was hybridized with synthetic core 1 (**Figure 4.1d**). To ensure any potential secondary structures did not interfere with RNAPII binding at the TATA box, a spacer was used to distance the library from the core. Specifically, the same AT-rich neutral spacer used to distance the TFBS for Gal4p from the core (**Figure 3.6**) as described in **Chapter 3** was used in the library construction.

4.3.7 Isolation and characterization of G-rich UAS elements

The final G-rich library was comprised of up to $2.9X10^6$ unique candidates (as determined by the number of *E. coli* transformants), and was used to drive the expression

of yECitrine fluorescent reporter. Top 0.1% of two million fluorescent events from the library was sorted by FACS under exponential and stationary growth phases. Of 96 colonies analyzed, three G-rich UAS elements derived from the stationary growth phase sort were confirmed to bestronger than the weak promoter *CYC1*. Further characterization reveals that under exponential phase, these promoters are weak, similar to that of the weak constitutive promoter *CYC1*. However, under stationary growth phase, the measured fluorescence levels dramatically increase. $UAS_{G\text{-}richB}$- core1 promoter in particular results in a 227% increase in fluorescence from exponential growth to stationary growth, approaching 61% the fluorescence driven by the strong constitutive promoter, *GPD* (*TDH3*) (**Figure 4.10a**). Furthermore, it does so at a fraction of the length (**Figure 4.1b**).

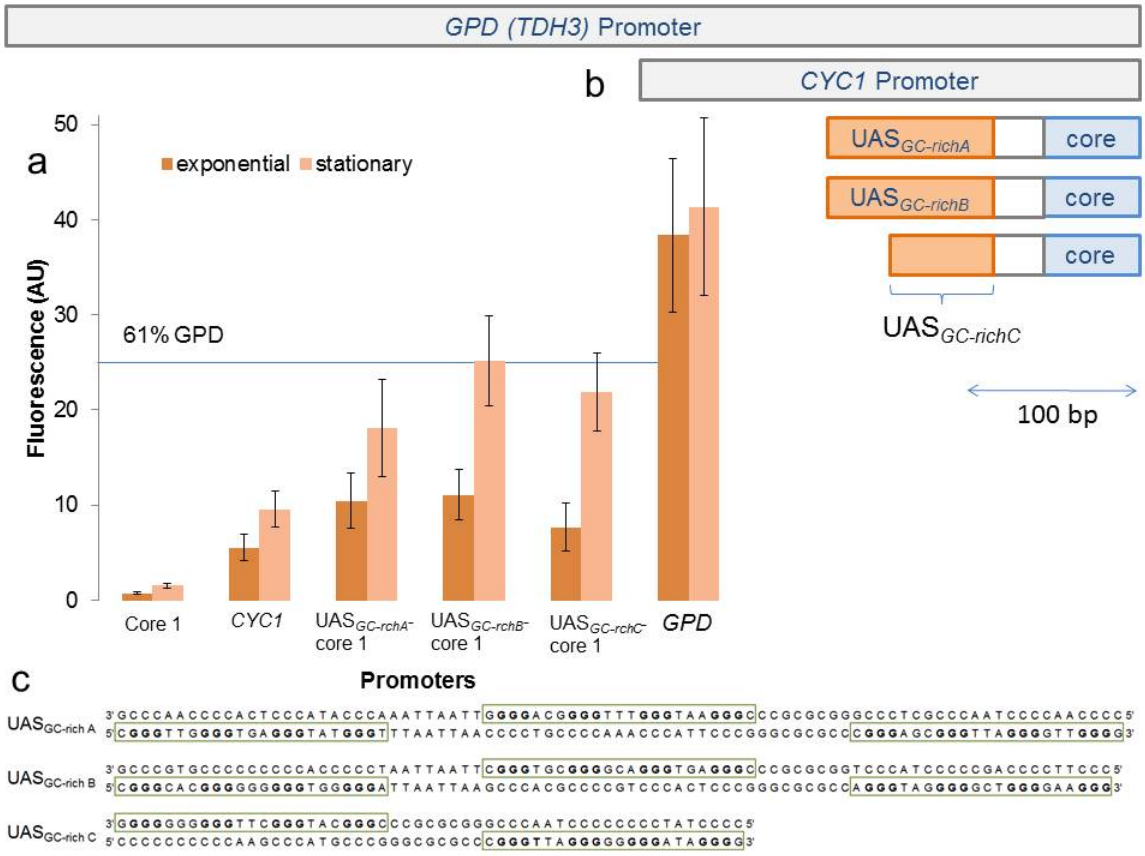


Figure 4.10: Unique G-rich UAS elements display ramp-up characteristics

(a) G-rich UAS elements hybridized to synthetic core 1 are smaller than commonly used *GPD* (*TDH3*) and *CYC1* promoters. Promoters shown here are to scale with 100 bp as indicated. (b) G-rich UAS elements hybridized to core 1 result in higher fluorescence levels in later growth stages. Error bars represent standard deviation of biological triplicates. *GPD* (*TDH3*)-driven fluorescence remains constant in contrast. In exponentially-growing cells, G-rich UAS elements drive low expression, similar to that of *CYC1*. With increased growth, a G-rich UAS element activates core 1 to reach 61% of the fluorescence attained with promoter *GPD* (*TDH3*). (c) Sequences of G-rich UAS elements are diverse. GQ-potential forming sequences are boxed. Multiple GQ-forming potential sequences are present although libraries were not intentionally made so. All contain at least one GQ-potential forming sequence on their non-template strand (3`→5`). Error bars represent standard deviation among biological triplicates.

4.3.8 dCAS9 amplification of core elements

For additional activating sequences, we turned to a rational design approach. The experiments described in this section were performed by Matthew Deaner. In our rational design, we used a TFBS to amplify synthetic cores. We chose a binding site that recruits an easily programmable and orthologous sTF, a dCAS9-based TF. dCas9 is an RNA-guided endonuclease whose endonuclease activity has been removed through mutations[167]. dCAS9, like CAS9 can be targeted to a specific sequence through tight binding of the associated single guide RNA (sgRNA) to a complementary sequence[168]. The ability to target binding with an RNA sequence made this protein an attractive TF source. Engineers were able to successfully engineer a dCAS9-based sTF by fusing dCAS9 with a tripartite transcription activator VP64-p65-rta (VPR)[169]. This sTF can be easily targeted by either altering (i) the sgRNA sequence expressed to complement a desired DNA sequence, or (ii) the targeted DNA binding sequence to complement the expressed sgRNA sequence. In our promoter construction, we sought to use the sgRNA site as a minimal UAS due to the site's short size. A high performing sgRNA site can be as small as 20 bp due to the high binding affinity of the dCAS9-associated sgRNA for the site[170]. The high affinity makes the system highly orthologous, and thus an excellent synthetic biology tool to exploit for promoter engineering efforts.

In our rational design, we positioned a sgRNA binding site 30 bp and 50 bp upstream of minimal core elements 1, 3, 5, 8, and 9. We showed that the 50 bp arrangement resulted in the greatest amplification (**Figure 4.11a**). Positioning the

sgRNA site just an additional 20 bp upstream, from 30 bp to 50 bp doubled the strength of all promoters constructed. This dramatic difference in expression indicates a high sensitivity to location with respect to a promoter core, and may be due to steric hindrances of the dCAS9 TF and RNAPII and its associating TFs. By distancing sgRNA site 50 bp upstream of synthetic core 5, we crafted a promoter reaching 60% of the strength of *GPD* (*TDH3*) (**Figure 4.11a**). Moreover, we demonstrated that this short sgRNA site along with expression of dCAS9 TF can amplify transcriptional capacity of cores 670 to 1258% (**Figure 4.11b**). This is the first time a dCAS9 artificial TF was employed in a hybrid promoter manner to amplify the transcriptional output of a synthetic promoter core.

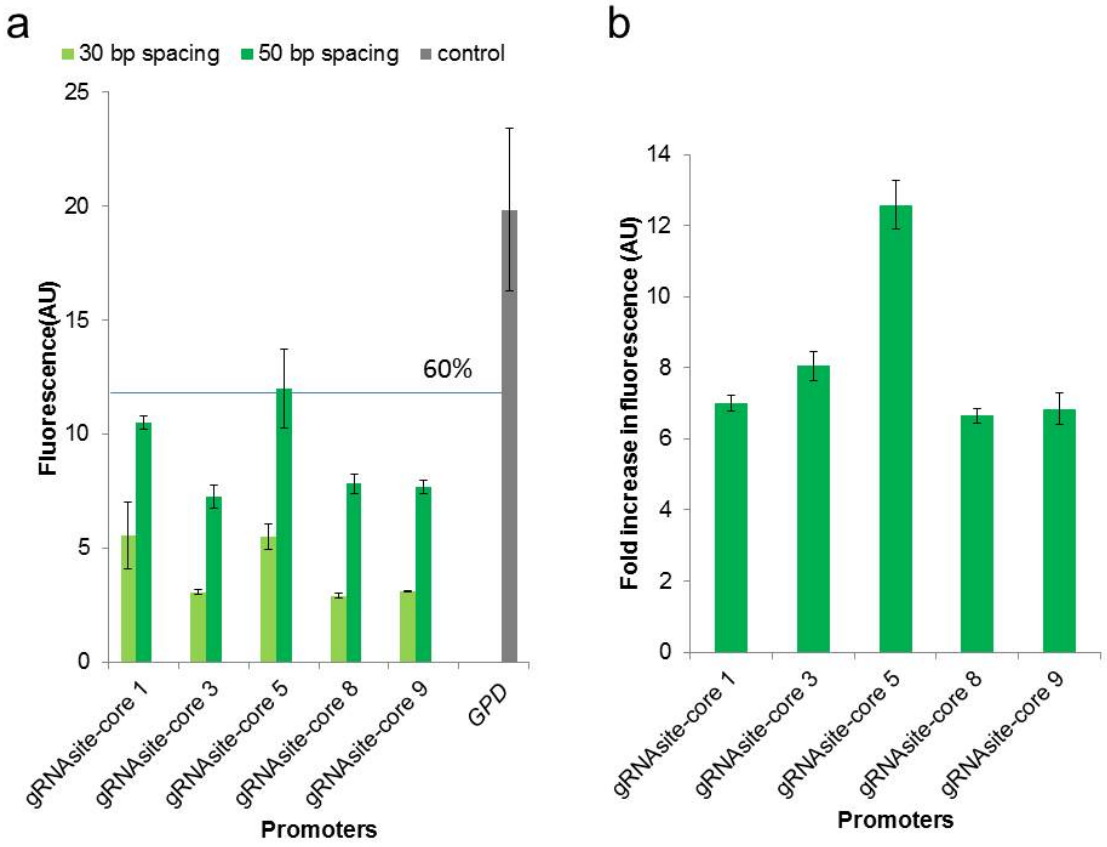


Figure 4.11: dCAS9 amplification of core elements

(a) 20 bp sgRNA site was positioned 30 bp and 50 bp upstream of selected synthetic core elements 1, 3, 5, 8, and 9. 50 bp spacing of sgRNA site allows for higher amplification of core elements. (b) When sgRNA is expressed, hybrid promoters with sgRNA site positioned 50 bp upstream of synthetic core elements increase expression of yECitrine. Error bars are standard deviations of biological triplicates.

4.4 Discussions and Conclusions

By drawing from four sources of synthetic minimal UAS elements, a set of interoperable and modular UAS elements was established (**Figure 4.1a-e**). By hybridizing this unique set with synthetic minimal core elements, the shortest constitutive

promoters were constructed in *S. cerevisiae*, greatly expanding the promoter toolbox for this host organism.

The first source of synthetic minimal UAS elements was a pool of 1.3 million 10 bp UAS candidates (**Figure 4.2**). From this pool, a set of six constitutive unique 10 bp UAS elements was isolated. Modularity of the elements as distinct activations parts was demonstrated. First, they were hybridized with all nine synthetic minimal core elements to synthesize full length functional promoters (**Figure 4.3**). Second, they were tandem-assembled with both synthetic and native UAS elements to create even stronger activating sequences (**Figure 4.8**). Hybridized with core 1, these full length promoters approach strengths of *GPD* (*TDH3*).

Next set of UAS candidates were sourced from rounds of N_{10}-UAS-core hybrid promoter library scaffolds. Six 10 bp modular UAS elements described above were employed in two rounds of N_{10}-UAS-core hybrid promoter library scaffolds totaling 7.4 million candidates. In doing so, nine additional constitutive minimal UAS elements were created, with a wide range of sequence diversity and predicted TFBS (**Figure 4.8a-b**). Interestingly, sequence analysis revealed a positional bias of predicted TFBS towards the 5` end of the UAS. This bias may indicate a TFBS generally requires more than a 30 bp spacing from the core for optimum performance. Indeed, for native TATA promoters, TFBS are generally found an additional ~25 bp upstream than those in TATA-like promoters[18]. Overall, this method of UAS element construction yielded a promoter 68% of the strength of *GPD* (*TDH3*) in less than 20% of the sequence space.

To further add to the synthetic minimal UAS element tool box, G-rich library comprised of four guanine tracts with high potential to form a well-conserved promoter GQ was employed. G-rich UAS elements derived from this library were eventually sequenced albeit with much difficulty (**Figure 4.10c**). Indeed, the sequence complications encountered were indicative of strong secondary structure formation. Consistent with formation of strong secondary structures, poor DNA yields of the double stranded G-rich oligonucleotide were observed. The addition of a GC stabilizer such as DMSO during double stranding did not remedy the problem, but actually, decreased DNA yields. This decreased double stranding efficiency is consistent with the G-rich oligonucleotide forming GQs *in vitro*, as GQs are reportedly stabilized by DMSO[171]. Although the library was synthesized with guanine tracts on the template strand, due to the occurrence of multiple insertions, all three G-rich UAS have at least one set of guanine tracts on the non-template strand (**Figure 4.10c**). This lies in contrast to native promoters where GQ tend to be on the template strand[25]. Given the synthetic nature of our promoter scaffold, it is possible that the GQ is only functional on the non-template strand.

In collaboration with Matthew Deaner, a sgRNA site was tapped as a source of minimal UAS to add to our synthetic promoter toolbox. dCAS9-based sTF was enlisted to bind to a 20 bp sgRNA site and cause amplification of selected core elements. In all promoters constructed, dCAS9-based sTF functioned to amplify the core elements. In fact, we were able to assemble short minimal promoters at 60% of the strength of *GPD* in

less than 20% of the sequence space. Interestingly, there was a dramatic increase in amplification when the sgRNA site was positioned from 30 bp to 50 bp upstream of the core elements. Thus, there is potential room for improvement by investigating the optimal spacing of the bound dCAS9 to RNAPII and its associating TFs within the promoter. Nonetheless, at 50 bp spacing of the sgRNA site, there was a 670 to 1258% amplification of core elements (**Figure 4.11b**). The fact this dCAS9 promoter amplifier functioned in all promoters assembled highlights its modularity and its potential to amplify other cores.

In all, 18 UAS elements were derived from eight libraries (**Table 4.1a-e**). 2 additional were rationally designed; 1 tandem assembled triple UAS, and 1 dCAS amplifier. When hybridized with minimal core elements, 73 fully synthetic promoters were assembled that drive a 70-fold expression range (from core 4 to $UAS_{F\text{-}E\text{-}C}$-core 1) reaching levels as high as 70% of *GPD* (*TDH3*) in less than 20% of the sequence space (**Figure 4.8**). When hybridized with endogenous UAS_{CLB}, 90% of the strength of *GPD* (*TDH3*) was obtained (**Figure 4.8**). Lastly, all of the strategies utilized are applicable to other organisms to build their toolboxes.

Chapter 5: Establishment of an isolation scheme for the identification of inducible promoters in *Saccharomyces cerevisiae*.

5.1 CHAPTER SUMMARY

In *S. cerevisiae*, molecular biology investigations and synthetic biology efforts are dampened by the lack of minimally-sized inducible promoters. To address this, we sought to establish an isolation scheme for the identification of minimal promoters that can be induced with inexpensive small molecules present in the media. Such promoters can allow engineers and scientists alike the flexibility to turn on transcription as desired while keeping costs and DNA importations to a minimum. In this chapter, I design a high-throughput FACS-based isolation scheme, iCAPTR (Isolation of Conditionally Activated Parts for Transcription Regulation) and in a pilot test of iCAPTR, isolate a completely synthetic galactose-inducible UAS that can activate a synthetic core to maximum levels as high as the native Gal4pBS. With further improvements to iCAPTR, I demonstrate the potential of iCAPTR through the establishment of seven maltose-inducible minimal promoters that provide induction at a continuum of strengths.

5.2 INTRODUCTION

Manipulation of gene expression is useful for a variety of scientific fields. For metabolic engineers, inducing expression of a burdensome gene in a robust growth phase can greatly decrease and even completely eliminate toxicity of a protein and/or its products. For molecular biologists, gene induction provides a way to mechanistically understand cell biology and physiology by observing the effects the abundance of a protein has on a biological system. Although numerous specialty systems are available to

allow for this kind of cellular control at the DNA level[172], the majority require modification of the gene of interest, either by (i) adding RNA-based switches for translational control[173-177], (ii) fusing additional protein domains to direct cellular localization or allow for active site manipulation, or (iii) splitting it for expression of two inactive protein fragments that regain their activity upon light or small molecule induced association[178]. These sophisticated systems require significant engineering of the protein of interest, and potentially, the background strain. I argue that the oldest and simplest, yet most effective method is to employ inducible promoters.

A list of inducible promoters is available for use in yeast[179], with native promoters, *GAL1*, *CUP1* and *ADH2* being the most commonly used[16] (**Table 5.1**). *GAL1* remains the most popular for two qualities. First, *GAL1* can deliver tight-regulation of expression. In the presence of glucose, the *GAL1* promoter is completely repressed, and in the presence of galactose, a 1000-fold increase in expression has been observed[180]. Second, *GAL1* can also provide rapid induction of the gene of interest. The rapid increase in expression can be achieved in just four hours[180]. However, carbon-sourced promoters like *GAL1* are not desirable for several reasons. First, since these promoters can be highly repressed in the presence of glucose, as is the case with both *GAL1* and *ADH2* (**Figure 5.1**), they require a laborious washout of glucose to obtain maximum induction. Secondly, this necessary carbon swap can greatly alter the transcriptome and metabolome[181,182], and as a consequence, complicate metabolic engineering efforts and mechanistic understandings of cellular biology. Lastly, since the inducer can be

metabolized, the strength of the promoter can wane unless the inducer is consistently supplied[181]. A popular alternative to carbon-sourced promoters is the native Cu^{2+}-inducible *CUP1* promoter, which can be induced 20-fold (**Table 5.1**)[183]. However, in addition to being leaky (driving expression without inducer present), *CUP1*'s inducer can perturb enzyme activities[184], potentially interfering with heterologous enzymes. Regardless of their induction mechanism, all native inducible promoters, like their native constitutive cousins are rather long. For the most commonly-used ones, they are over 450 bp long (**Table 5.1**).

Promoter	Commonly Used Inducer*	Max. fold Induction**	Length (bp)***	Pros	Cons
GAL1	Galactose	1000[180]	453[180]	Tightly controlled, rapid induction[185]	Requires low glucose, weakens as galactose is consumed[185]
ADH2	<1%glucose	200 [186]	573[186]	Highest max transcript output[185]	Requires low glucose[185]
CUP1	Cu^{2+}	50[187]	450[188]	Not carbon-sourced, highest after diauxic shift[189]	Leaky[185], inducer perturbs enzyme activities[184]

Table 5.1: Commonly used native inducible promoters

*Other inducers are possible. For example, *CUP1* can also be induced with low pH[190], and DNA damage/oxidative stress[191]. Commonly used inducer is typically the easiest to apply and often stimulates the strongest induction. **Maximum fold induction ever reported in literature. Value may be lower when applied. For example, in our studies *GAL1* only induced 400-fold (**Figure 5.3**). ***Length used in literature to date Actual functional length may be shorter.

Another way to control expression at the DNA level is to recruit an imported TF to a targeted promoter for transcription manipulation. One of the most effective exogenous TF/promoter systems is the estradiol-activated sTF and its targeted promoter. Two general versions of the estradiol-activated sTF exist. One utilizes the Gal4p DNA binding domain[96], while the other relies on the more orthologous and programmable zinc finger DNA binding domain[105]. Both target the *GAL1* promoter, either the native

sequence or a modified version of the promoter. These sTF are able to induce transcription 200-fold and 50-fold, respectively, with the latter delivering tight-regulation. Although the estradiol-activated system offers an orthogonal way to manipulate protein expression at the DNA level, the system has drawbacks. Aside from the high cost of the small molecule inducer, estradiol ($2,000/g [75]), this system requires a significant amount of DNA importation for application. In addition to the lengthy target promoter DNA of over 450 bp, an entire expression cassette for the sTF is necessitated.

Thus, for similar reasons to shorten the constitutive promoter, I sought to minimize the inducible promoter. To do so, I propose to isolate minimal UAS elements that when hybridized with a minimal synthetic core activate transcription in the presence of inexpensive small molecules (galactose, caffeine, itaconic acid, aspirin, eugenol, vanillin, calcium, copper and maltose). In this chapter, I establish a high-throughput FACS-based isolation scheme, iCAPTR (Isolation of Conditionally Activated Parts for Transcription Regulation). During the pilot test of iCAPTR, I isolate a completely synthetic, minimal, galactose-inducible UAS element that activates synthetic core 1 to the same maximum expression as a native Gal4pBS. To further demonstrate the utility of iCAPTR, I isolate a set of synthetic, minimal, maltose-inducible promoters with the strongest driving expression as high as the strong constitutive *TEF1* promoter.

5.3 RESULTS

5.3.1 Library design for iCAPTR

To isolate inducible promoters, I used a library-based approach similar to that applied in **Chapter 4**. Specifically, I hybridized a randomized region to a synthetic core element to create a hybrid promoter library. By sorting this library under a desired inducible condition for strong promoters, I hypothesized that I could obtain a sequence that is activated under that specific condition (**Figure 5.1**). For the "inducer condition", I sought to use small molecules present in the liquid culture media. In **Chapter 4**, I demonstrated that a N_{10} region in a hybrid promoter library scaffold provides sufficient sequence space to obtain synthetic constitutive UAS elements. However, for the inducible minimal promoter library, the randomized UAS region was lengthened to 20 bp. Given that native promoters with dynamic expression profiles tend to be TATA-promoters[10], and that TATA-promoters tend to have longer TFBS sites[18], I hypothesized that any potential activator would require more than 10 bp to bind. Thus, we established a N_{20}-core hybrid promoter library. The same AT-rich spacer used in the constitutive hybrid promoter library described in **Chapter 4** was used to distance the N_{20} upstream of the core, and synthetic core 1 was chosen for its robust activation as demonstrated in **Chapter 3**.

5.3.2 Pilot test of iCAPTR: Isolation of galactose inducible UAS

To test iCAPTR, I first attempted to isolate a synthetic, minimal galactose-inducible promoter. Gal4p is the TF responsible for the dramatic induction of native

GAL promoters. Work presented in **Chapter 3** demonstrated that Gal4p can activate our synthetic core elements via a Gal4pBS under galactose conditions. With a consensus binding site of $CGGN_{11}CCG$[192], I hypothesized a Gal4pBS could be isolated from a N_{20} library under galactose growth conditions. To test the hypothesis, N_{20}-core 1 hybrid promoter library containing up to $2X10^6$ unique candidates was used. This library was used to drive the expression of the fluorescent reporter yECitrine. In the first step of iCAPTR, the highest fluorescent cells from N_{20}-core 1 hybrid promoter were sorted with FACS under the inducer (**Figure 5.1**). Specifically, the top 0.1% of two million measured fluorescent events was pooled to isolate the highest functioning candidates (**Figure 5.1**). Sorting was performed with 2% galactose as the sole sugar source (**Figure 5.2a**). In step 2 of iCAPTR, the sorted cells were recovered in 2% glucose minimal media liquid culture, plated on inducer (2% galactose) and 90 strongly fluorescent colonies were picked qualitatively (**Figure 5.1**). For step 3, fluorescence of the 90 cultures was measured under inducing (2% galactose) and non-inducing (2% glucose) conditions by flow cytometry (**Figure 5.1**). For the control, core 1 was linked with the AT-rich spacer used to distance the N_{20} library from core in the scaffold. The control was subjected to the same sorting conditions to compensate for any selection bias that is not a result of the library sequence upstream of the core in the plasmid.

Under galactose, 70 of the 90 cultures characterized in step 3 of iCAPTR were higher in expression than our control (spacer-core 1) (**Figure 5.2a**). Of those 70, 17 induced more than 10-fold (**Figure 5.2a**). These were analyzed in biological triplicate in

step 4 of iCAPTR after retransformation into fresh BY4741 to gain a more accurate assessment of expression strength (**Figure 5.1**). In step 4, only 2 of the 17 were confirmed to be inducible under galactose (**Figure 5.2a**). In fact, a 20 bp UAS element was obtained that responded to galactose just as well as the native Gal4pBS present in *GAL1* promoter (**Figure 5.2b**). Sequence BLAST of this UAS element reveals a completely synthetic sequence (BLAST E_{value} of 6.1). Although the UAS does not match any yeast genomic sequences, the sequence does follow the Gal4pBS consensus sequence (**Figure 5.2c**).

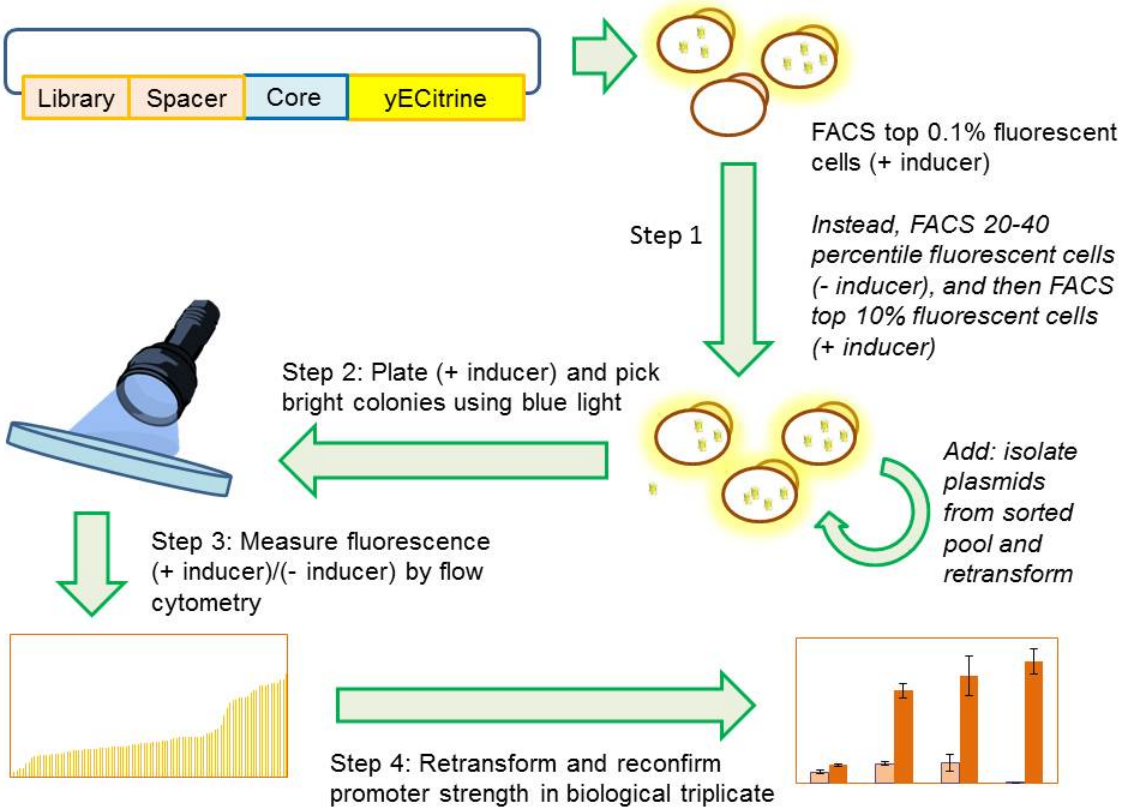


Figure 5.1: Outline of iCAPTR

After a pilot test of iCAPTR with galactose as inducer, improvements to step 1 (shown in italics) were performed. In step 1, a N_{20}-core 1 hybrid library on plasmid was transformed into *S. cerevisae* and sorted by FACS. In the pilot test, top 0.1% of fluorescent cells were sorted by FACS with inducer present. After pilot, this step was altered to two tandem sorts: 20-40 percentile sort without inducer, and then, top 10% with inducer. This step was further improved with a plasmid isolation and yeast retransformation step. In step 2, the sorted cells were recovered overnight, plated with inducer, and strongly fluorescent colonies were picked qualitatively with a blue light. Picked colonies were grown in liquid media, and underwent one -80° C freeze/thaw cycle before having fluorescence measured by flow cytometry with and without inducer (Step 3). In step 4, plasmids were extracted from cultures with highest ratio of fluorescence with inducer and fluorescence without inducer, and retransformed into fresh BY4741 background to have induction strength confirmed in biological triplicate.

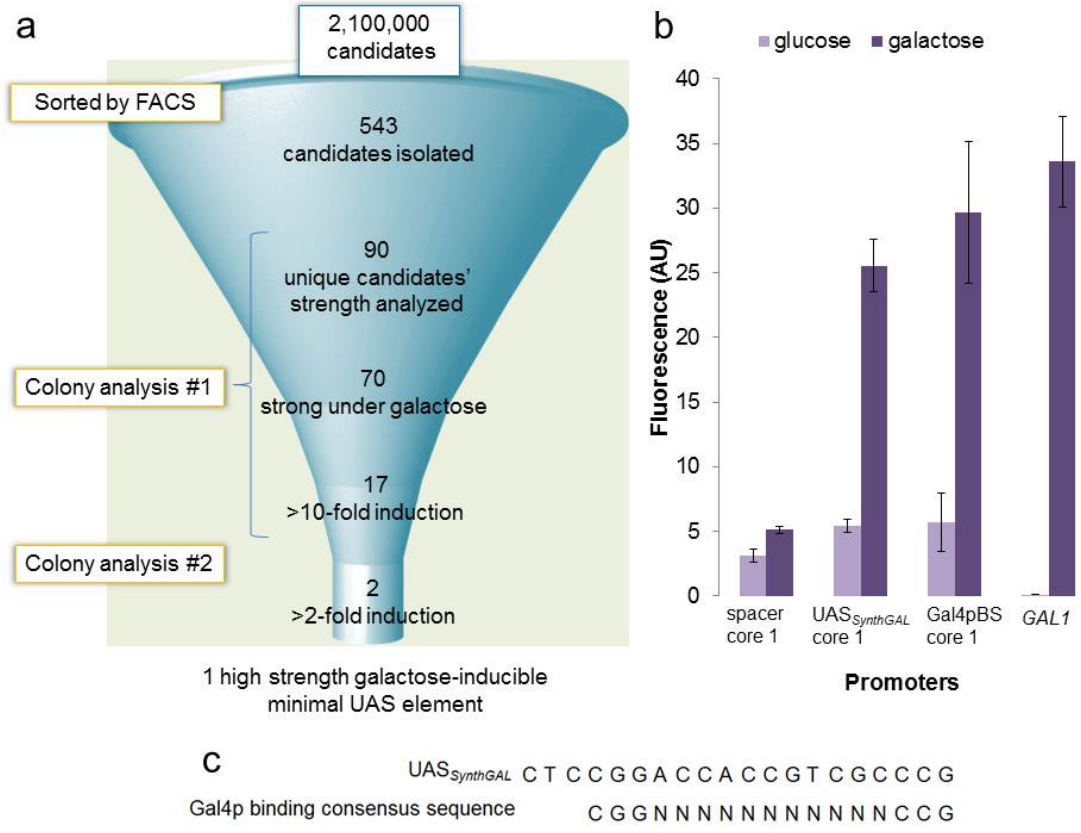


Figure 5.2: One synthetic high strength galactose-inducible UAS element was derived from pilot test of iCAPTR

iCAPTR (**Figure 5.1**) was successful in isolating not just a galactose-inducible UAS element, but a completely synthetic UAS element. (a) Top 0.1% of 2 million cells from a library pool of up to 2.1 million unique candidates were sorted by FACS with 2% galactose as the sole sugar source. Colony analysis was performed on 90 candidates. Under galactose, 70 of the 90 were stronger than spacer-core 1 control under 2% galactose. Of those 70, 17 induced more than 10-fold during the initial screen. After retransformation, 2 induced more than 2-fold with one candidate driving high maximum expression under galactose. (b) One galactose-inducible UAS element, $UAS_{SynthGAL}$ was shown to be strongly induced under galactose conditions, functioning similar to the strongest native Gal4pBS. Error bars represent standard deviations among biological triplicates. (c) Sequence of $UAS_{SynthGAL}$ follows consensus sequence for Gal4pBS. However, actual sequence is completely synthetic with an E_{value} BLAST score of just 6.1.

5.3.2 Selection of inducer set

With the establishment of an isolation scheme, a set of eight inducers was selected. I selected compounds that as a set affect a spectrum of cellular pathways, are readily available and have low cost (**Table 5.2**). Additionally, the compounds needed to be soluble in aqueous solution at the effective concentration. This ensured the inducers could be easily added to liquid media for induction. I included small molecules, vanillin, calcium, copper, and maltose known to induce genomic promoters, *ADH6*/*ADH7*, *FKS2*, *CUP1*, and *MAL* complex locus respectively.

Prospective inducer	Cellular effects	Concentration used	
		Exponential sort (C_{5-10})	Stationary sort (C_{40-50})
Caffeine	Alters cell wall architecture via Pkc1p-mediated cell integrity pathway. Also targets Tor1 kinase[193].	500 µM	6 mM
Itaconic acid	Petrochemical precursor	1.0 g/L	4.5 g/L
Aspirin	Increases mitochondrial and cytosolic superoxides, increases oxidation of mitochondrial NAD(P)H[194].	7 µM	500 µM
Eugenol	Increases cytosolic Ca^{2+}[195], interferes with permeases that transport aromatic and branched-chain amino acids through cytoplasmic membrane[196].	25 µM	100 µM
Vanillin*[197]	Induces formation of cytoplasmic messenger ribonucleoproteins granules[198], activates oxidative stress-responsive Yap1p[199] and general stress responsive Msn2p[200]	2 mM	N/A
Calcium*[201]	Calcineurin-dependent pathways[202]	10 mM	200 mM
Copper*[188]	Oxidative stress, reduced Cu-transporter expression[203]	100 µM	500 µM
Maltose*[204]	Activates maltose catabolism genes in *MAL* loci[205]	2 % w/w**	N/A

Table 5.2: List of prospective inducers

Table 5.2: List of prospective inducers continued

A set of prospective inducers was chosen based on their high availability, low cost, and wide spectrum of cellular effects. *Known endogenous promoters are induced by these small molecules. Concentrations used based on survival assays performed. Inducer concentrations used for libraries sorted at exponential growth phase are those that cause 5-10% decrease in maximum cell culture density (C_{5-10}). Inducer concentrations used for libraries sorted at stationary growth phase are those that cause a 40-50% decrease in maximum cell culture density (C_{40-50}). Vanillin inducer was used as a case study to improve isolation scheme. Stationary phase not sorted for this inducer. **No survival assay performed for maltose inducer. Concentration used is well established amount for induction of *MAL* genes[204]. Stationary phase not sorted for maltose.

5.3.3 Determination of exposed inducer concentrations through survival assays

Once a set of inducers was chosen, exposure concentrations were established (**Table 5.2**). Survival assays were performed in minimal media to determine the growth defects incurred by each inducer at varying concentrations. Growth defects indicate the inducer is perturbing cellular functions, a prerequisite for an inducer to function. Two concentrations of inducer were determined with these assays: i) that which reduced maximum cell density by 5-10% ($C_{5\text{-}10}$), and ii) that which reduced maximum cell density by 40-50% ($C_{40\text{-}50}$). I hypothesized that the lowest concentration of inducer should be used to minimize cellular effects and costs. Higher, more cellular disruptive concentrations were used when no inducible promoters were attained with lower concentrations. Since maltose is a carbon source, a survival assay was not performed with this compound. 2% w/w maltose was used based on the traditional concentration used to induce *MAL* promoters[204].

5.3.4 Length and time of inducer exposure to promoter library

I next chose the length of inducer exposure before sorting and the growth phase at which to sort the cells. Cells were sorted under the inducer at two different growth phases depending on the concentration of inducer used (**Table 5.2**). Under inducer concentrations of $C_{5\text{-}10}$, cells were sorted under exponential growth phase. I hypothesized that cells would be most sensitive to the effects of selected inducer during rapid growth and thus, increasing the chances that the inducer signal is encoded by the cells internally. As mentioned previously, if no inducible promoter was isolated using the lower, less

toxic concentration, the higher concentration ($C_{40\text{-}50}$) was applied. Yeast cells exposed to inducer concentrations of $C_{40\text{-}50}$ were done so at stationary phase due to the toxicity of the inducer. I found that cells exposed to inducer concentrations of $C_{40\text{-}50}$ at low cell densities incurred extreme growth defects. Thus, it was pertinent that the cells reach a robust growth phase before exposure. Furthermore, the length of exposure was adjusted depending on the inducer concentration. Cells exposed to inducers at $C_{40\text{-}50}$ were done so at stationary growth phase for 4 hours. In contrast, cells exposed to the lower, less toxic concentrations of C_5-10 were done so for 14 hours.

5.3.5 Introduction of FACS under non-inducing conditions removes constitutive candidates from library

Before attempting to derive inducible promoters that respond to the selected set of inducers, the isolation scheme needed to be improved. I acknowledge two exceptional qualities of a Gal4p-activated UAS element. First, Gal4p is able to activate promoter cores to high maximum transcription outputs as demonstrated in **Chapter 3**. This unique quality means galactose-inducible promoters could easily be present in the top 0.1% fluorescent cells of the hybrid promoter library under galactose conditions. However, this may not be the case for promoters induced by the presence of the selected inducer compounds. Therefore, to ensure potential inducible candidates are not left out of the FACS-enriched pool, FACS was relaxed from the top 0.1% to the top 10%. However, doing so increases the enriched pool size by 100-fold. A pool that large means only a small fraction could be assessed in the initial screen and increases the odds of missing an inducible candidate.

This brings us to the second exceptional quality of Gal4p: high DNA binding promiscuity. To successfully derive an inducible promoter with iCAPTR, the chances of an inducer-activated TF binding to the hybrid-promoter library needs to be 1 in 2 million, if not considerably higher. Assuming Gal4p can activate a synthetic core via binding to any sequence within $CGGN_{11}CCG$, there are theoretically 30 possible binding sites in 2 million sequences of a randomized 20 bp region. However, this high frequency may not be the case with the set of selected inducers, and that potential binding sites for inducer-activated TFs in the library would be largely outnumbered by sites recruiting constitutively expressed TF. Thus, I reasoned that a selection under non-inducing conditions would need to be included in iCAPTR. This selection would not only reduce the size of the enriched pool further, but would also remove unwanted constitutive promoters from the FACS-enriched pool.

Thus, N_{20}-core 1 hybrid promoter libraries were subjected two sorts: i) isolation of highly fluorescent cells under inducing conditions and ii) isolation of lowly fluorescent cells under non-inducing conditions. To isolate 'off' promoters under non-inducing conditions, the bottom 30% fluorescent cells from two million recorded fluorescent events without inducer present were sorted. I investigated whether to follow a low/(-) inducer sort with a high/(+) inducer sort (L/H), or a high/(+) inducer sort with a low/(-) inducer sort (H/L). During this investigation, vanillin was used as an inducer at $C_{5\text{-}10}$ (2 mM). Between sortings, cells of the enriched library were allowed to recover in liquid media, and diluted to low cell density to sort again at exponential growth phase. Cells

were exposed to inducer for 14 hours as they grew up to exponential phase. Both sorts were done with cells in exponential growth phase.

H/L sorting resulted in acquisition of constitutive promoters (**Figure 5.3a-b**). 21 bright colonies were picked qualitatively. Of the 21 analyzed by flow cytometry, the expression strength of six with the highest induction was retested (**Figure 5.3a**). Plasmids were isolated, retransformed and induction strength was measured again with flow cytometry. Little to no induction was observed for all six (**Figure 5.3b**). Since H/L sorting resulted in constitutive promoters, I decided this was not the appropriate sorting method to use.

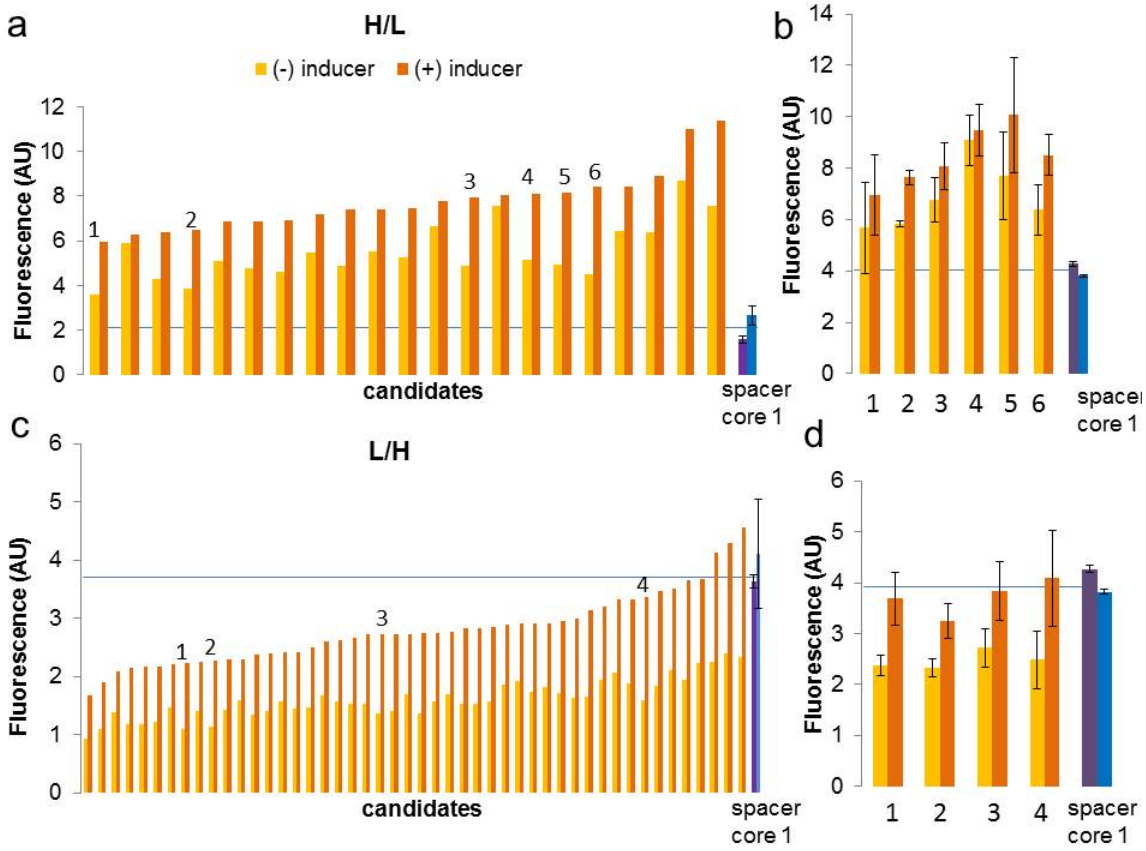


Figure 5.3: Removal of constitutive promoters from enriched pool

An additional selection was included in the isolation scheme to remove constitutive promoters during FACS enrichment. (a) Sorting highly fluorescent cells (top 10%) with inducer followed by a sort of lowly fluorescently cells (bottom 30%) without inducer (H/L) resulted in UAS candidates with constitutive function (n=1) as confirmed after retransformation (n=3) (b). (c) Sorting lowly fluorescent cells (bottom 30%) without inducer followed by a sort of highly fluorescent cells (top 10%) with inducer (L/H) resulted in UAS candidates repressing expression without inducer present (n=1) as confirmed after retransformation (n=3) (d). Inducer used here is vanillin (2 mM). *Indicates which candidates in (a) and (c) were retransformed into fresh BY4741 and strength confirmed in (b) and (d) respectively. Spacer core 1 control was performed in triplicate for (a) and (c). Controls were subjected to same sorting conditions. All error bars represent standard deviations among biological replicates.

L/H sorting, on the other hand yielded weak repressors (**Figure 5.3c-d**). 50 bright colonies were picked qualitatively. Of the 50 analyzed, the induction strength of a select

four was tested as described above (**Figure 5.3c**). All four promoters analyzed showed an increase in strength under inducer (**Figure 5.3d**). However, maximum expression was the same as the spacer-core 1 control, and in absence of inducer, fluorescence expression was weakened 28 to 40% compared to spacer-core 1 (**Figure 5.3d**). Thus, this indicates that the promoter is possibly recruiting a repressor under non-inducing conditions which dissociates under inducing conditions. In fact, repressing activity was observed for all candidates during the first colony analysis. All 50 candidates selected exhibited strengths lower than spacer-core 1 when inducer was not present (**Figure 5.4a**). While there is still an observed induction, given the low strength of the core already, these promoters used as is may not be useful for synthetic biology efforts. Nonetheless, L/H sorting did technically yield inducible promoters. For this reason, a double sorting method was introduced to step 1 of the isolation scheme as indicated by italicized text in **Figure 5.1**.

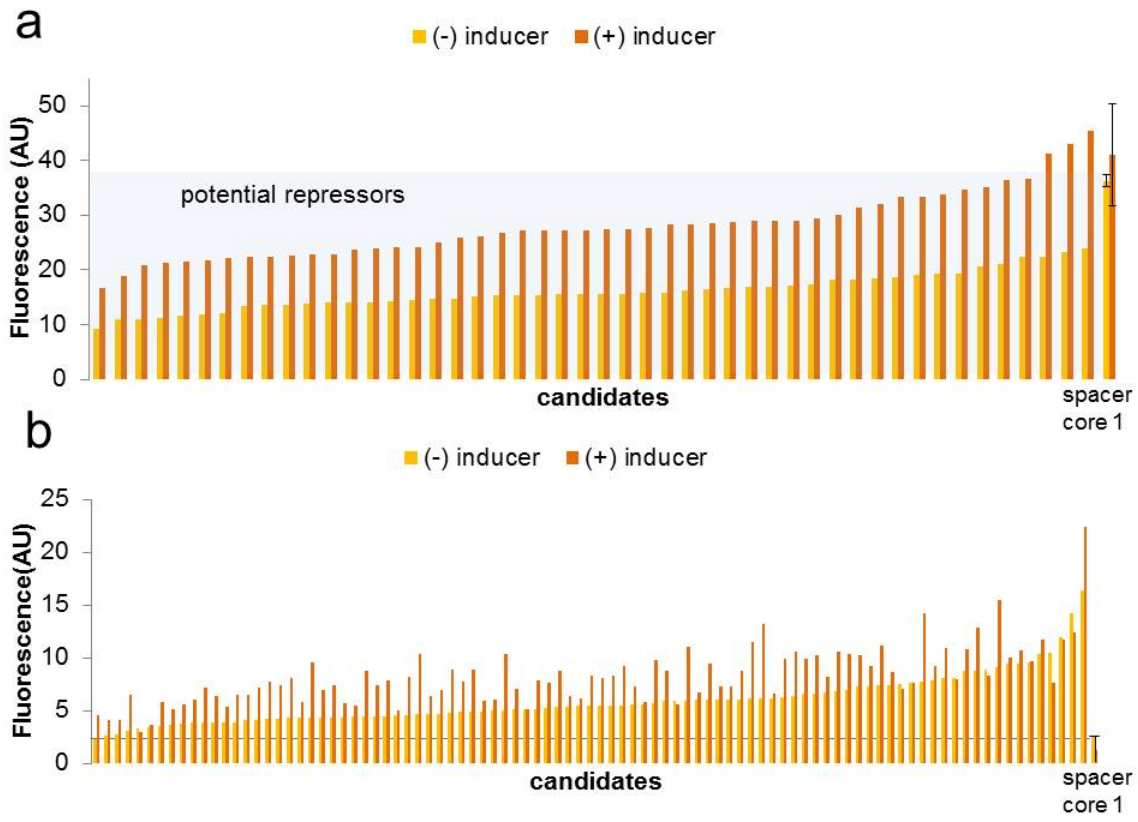


Figure 5.4: Repressors avoided by excluding bottom 20% fluorescent cells in low sort in iCAPTR

To avoid repressors, double sorting in step 1 of iCAPTR was modified. (a) When sorting the bottom 30% fluorescent cells without inducer followed by a sort of highly fluorescent cells, promoters weaker than control (spacer core 1) under non-inducing conditions were obtained in first initial strength assessment (step 3). Under inducing conditions, these candidates induced to the strength of the control or less. (b) When the 20 to 40 percentile of fluorescent cells are sorted without inducer, thereby avoiding the bottom 20% of fluorescent cells, no candidates lower than control (spacer core 1) are present under non-inducing conditions in the first initial strength assessment (step 3). Vanillin (2 mM) is inducer used. Controls were performed in triplicate and error bars represent standard deviation among these biological triplicates. Controls were subjected to same sorting practice as respective candidates. Fluorescent values for candidates represent one fluorescent measurement.

5.3.6 Exclusion of extremely lowly fluorescent cells from enriched library under non-inducing conditions avoids repressed promoters

In order to further improve the isolation scheme, I sought to remove repressed promoters from the FACS-enriched pool of lowly fluorescent cells under non-inducing conditions in step 1 (**Figure 5.1**). To remove them, I hypothesized that the lowest fluorescent cells need to be excluded from the enriched pool. To test the hypothesis, I used vanillin at $C_{5\text{-}10}$ (2 mM) as an inducer. Specifically, I sorted out the 20 to 40 percentile fluorescent cells from N_{20}-core hybrid promoter library without inducer present. Cells were recovered overnight in liquid minimal media without inducer. Cells were then diluted to low cellular concentrations (0.015 OD_{600}), and exposed to inducer until culture reached exponential phase (14 hours). 10% of the highest fluorescent cells from one million recorded events were isolated. An enriched pool of cells was recovered overnight in liquid minimal media without inducer, and plated onto agar with inducer and highest fluorescent colonies were selected qualitatively (step 2) (**Figure 5.1**). Select colonies were grown in liquid culture and fluorescence was measured by flow cytometry at exponential phase (step 3).

With this experiment, excluding the bottom 20 percent fluorescent cells from the enriched pool during non-inducing conditions did avoid repressed candidates (**Figure 5.4a-b**). All selected candidates from a low sort where the bottom 20% was included were found to be lower in strength than spacer-core 1 control in absence of inducer (**Figure 5.4a**). When this bottom 20% of fluorescent cells is excluded, no candidates are found to be lower in strength than control (**Figure 5.4b**). These results show that by

excluding the extremely lowly fluorescent cells from the FACS-enriched pool, repressed candidates under non-inducing conditions are avoided. For this reason, FACS enrichment during non-inducing conditions was altered to 20 to 40 percentile in iCAPTR (**Figure 5.1**). This modification to the double sort in step 1 of iCAPTR is indicated in italicized text of **Figure 5.1**.

5.3.7 Introduction of retransformation step of enriched library yields accurate candidate strength assessment in initial screen

Next, I sought to address the inconsistency of promoter strengths observed during the first assessment. Specifically, I frequently observed induction traits of candidates during the first initial screen of step 3 of iCAPTR only to fail to confirm this strength after retransformation into a fresh BY4741 background. As an example, candidates UAS_{vanA}-core 1 and UAS_{vanB}-core 1 were initially shown to be inducible 2-fold by vanillin (**Figure 5.5**). However, when plasmids are isolated from these candidates and retransformed into a fresh BY474 background, induction is no longer observed (**Figure 5.5**). I hypothesize the source of this transient promoter activity observed in the first initiation assessment is largely the result of FACS. It is well known that numerous interrelated factors are associated with a cell's heterologous protein abundance including the initial amount inherited from the mother[206], cell size[207], growth rate[207], age[208], endogenous protein levels[207], epigenetic modifications[209], and plasmid copy number [210]. In the experiments, I observed that a specific phenotype was favored during FACS. Large, granular cells (as indicated by the forward and side light scatter data gathered during flow cytometry) were overrepresented in the enriched pool of highly fluorescent

cells. Furthermore, cells exposed and sorted under vanillin, in particular had a more robust growth in the presence of the inducer, even after many generations. In this instance, the background strain has possibly undergone long-term transcriptional alterations to deal with the stress caused by the inducer. Thus, the resulting pool of yeast cells after FACS and exposure to the inducer, even after many generations appears to poorly reflect the expression profile of promoters in newly transformed cells.

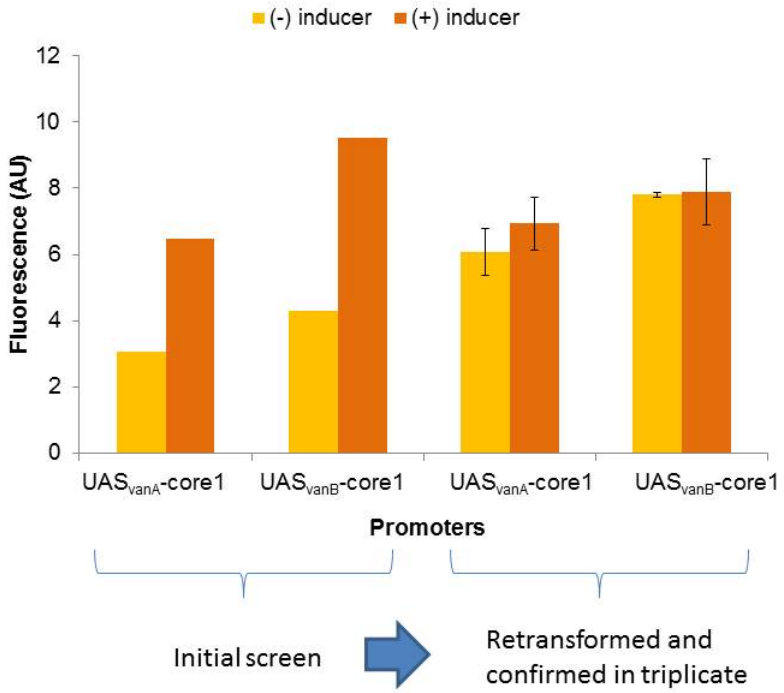


Figure 5.5: Initial screen reflects transient promoter strength

Promoter candidates UAS_{vanA}-core 1 and UAS_{vanB}-core 1 were identified as inducible promoters during the initial screen (n=1). However, after retransformation of plasmid into a fresh BY4741 background, the promoter no longer appears to induce (n=3). Error bars represent standard deviations among biological replicates. Vanillin (2 mM) is used as inducer.

Thus, to gain a more accurate assessment of promoter strength during the initial screen, plasmids from FACS-enriched cells were isolated and transformed into a fresh BY4741 background during step 1 of iCAPTR. To test the effects of this retransformation step, I used vanillin at $C_{5\text{-}10}$ (2 mM) as an inducer. Indeed, results show

that the introduction of this step to iCAPTR does allow the first colony analysis to reflect more consistent expression strength of the promoter (**Figure 5.6a-b)**. Without retransformation, most candidates selected during the first colony analysis exhibited a 1.5-fold increase in presence of inducer (**Figure 5.6a**). With retransformation, only 3% showed above 1.5-fold increase (**Figure 5.6b**). Further colony analysis showed that none of these are actually inducible. Thus, a retransformations step was adopted in step 1 of iCAPTR and is indicated by italicized text in **Figure 5.1**.

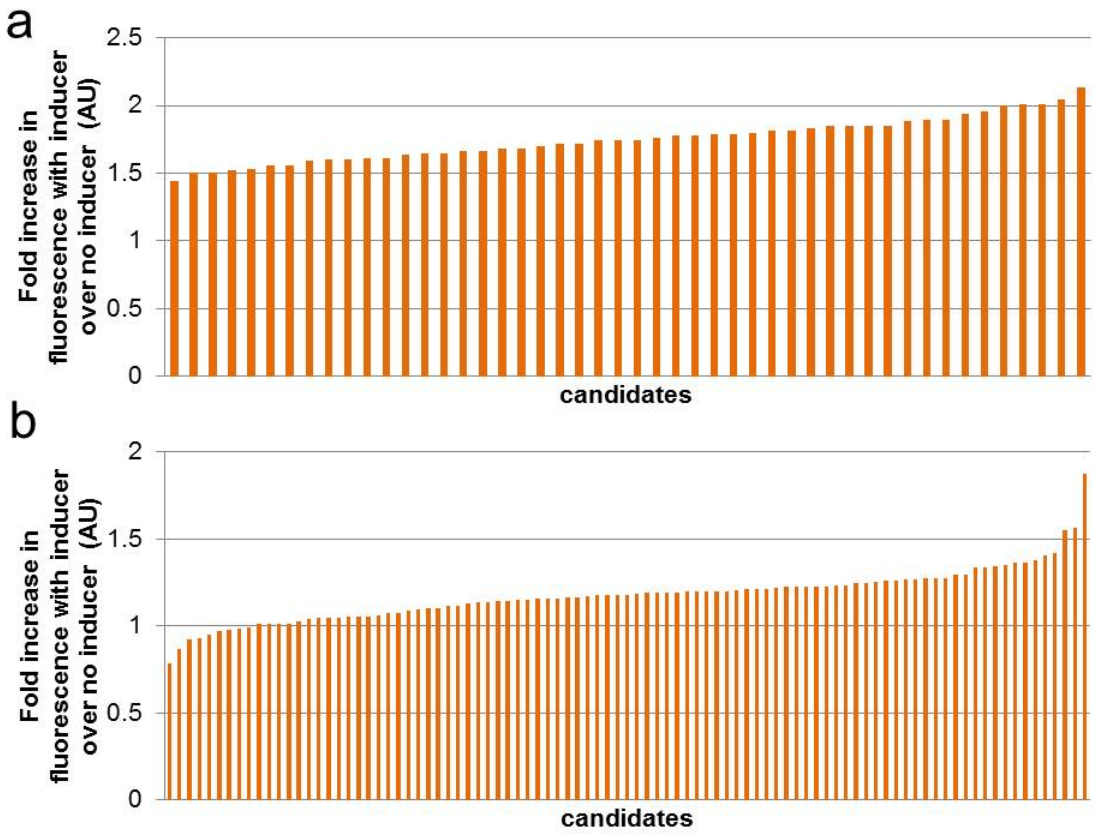


Figure 5.6: Introduction of retransformation step to iCAPTR provides a more accurate assessment of candidates' strengths

To gain a more accurate assessment of candidate expression strength in the first colony analysis, enriched libraries were retransformed into BY4741. (a) Without retransformation, most candidates (n=1) selected displayed a 1.5-fold increase in presence of inducer. (b) With introduction of retransformation step, only 3% of candidates showed above 1.5-fold increase (b). With further colony analysis, none of these candidates are actually inducible. Inducer used here is vanillin (2 mM).

5.3.8 Final implementation of iCAPTR

Improved iCAPTR was applied to inducers listed in **Table 5.2**. Of the 8 inducers tested, only two had candidates with more than a 2-fold induction observed in the first colony analysis. Copper had one but was unable to be confirmed when analyzed in

biological triplicate in step 4. 29 candidates with a 2-fold induction or higher in the presence of 2% maltose were obtained during step 3 (**Figure 5.7a**). 8 of the 29 were selected for strength confirmation in biological triplicate due to either on a high maximum expression or high fold induction. All 8 tested proved to induce in step 4 of iCAPTR. Two were able to drive high maximum expression (**Figure 5.7b**).

Unexpectedly, the AT-rich neutral spacer employed to distance the UAS from the core drives transcription under maltose. Core 1 without spacer does not increase in strength under maltose (**Figure 5.8**). Due to this cryptic maltose-inducibility of the library scaffold, one candidate mimicking the induction profile of spacer-core1 was discarded. The final set of seven UAS elements exhibits a range of maximum expression from as high as *TEF1* to as low as core 1 by itself (**Figure 5.8**). In particular, two are of relatively high strength. They contribute to the already maltose-inducibility of the spacer. In fact, $UAS_{SynthMALA}$ and $UAS_{SynthMALB}$ increased the strength 2.0 and 2.3 times that of spacer-core 1 under 2% maltose approaching strengths of *TEF1* (**Figure 5.7b**). To identify putative TFBS, sequences were analyzed using the YEASTRACT database[158]. In this analysis, Adr1pBS was identified (**Figure 5.7c**). Interestingly, Adr1p plays a key role in the function of *ADH2* promoter[211]. Incidentally, this promoter represents one of the commonly-used inducible promoters (**Table 5.1**).

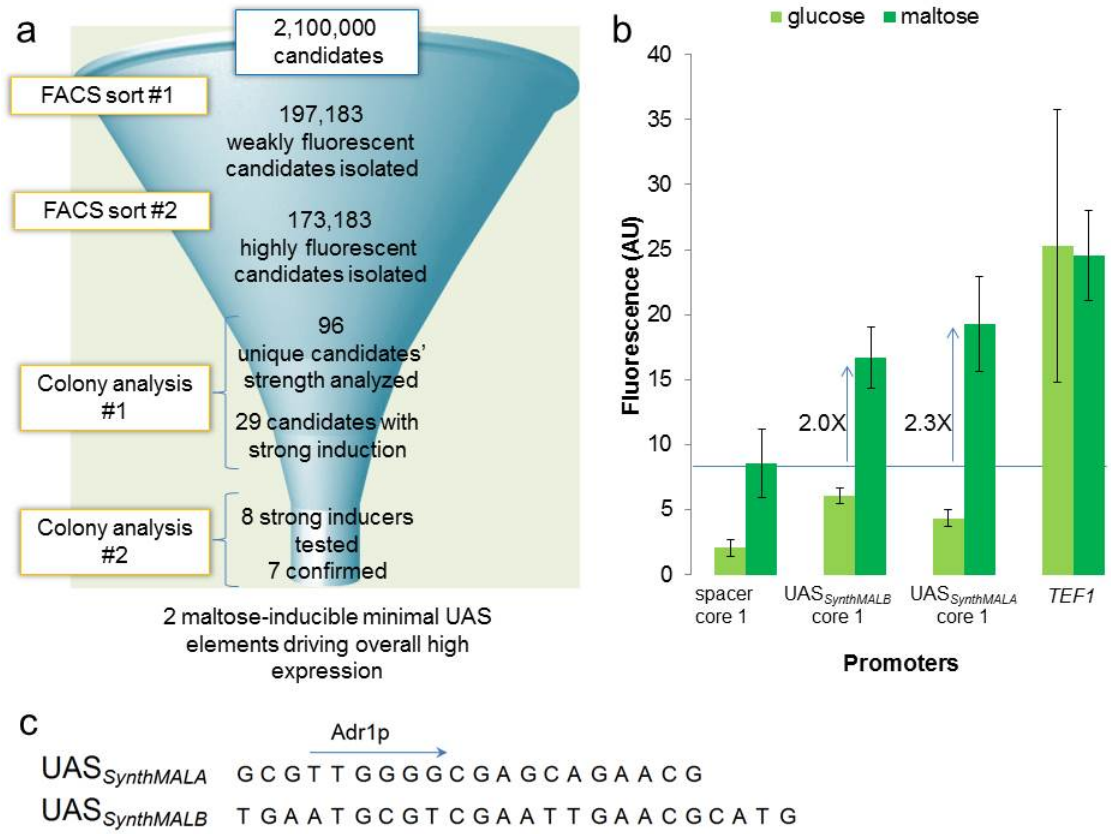


Figure 5.7: Two synthetic maltose-inducible promoters were derived using iCAPTR

$UAS_{SynthMALA}$ and $UAS_{SynthMALB}$ were derived from a library with iCAPTR using maltose as an inducer. (a) A library containing up to $2.1X10^6$ unique sequences was sorted by FACS twice; 20-40% percentile of fluorescent events followed by top 10% percent of fluorescent cells. An initial colony analysis was performed on 96 candidates under inducing (2% maltose) and non-inducing (2% glucose) conditions. 29 of these candidates showed >2 fold induction. A second colony analysis was performed on 8 of the 29. 7 of the 8 were confirmed to be inducible UAS elements. 2 of the 7 drove strong expression under inducing conditions. (b) Under 2% maltose as inducer, synthetic minimal promoters drive expression as strong as *TEF1* (n=3). Synthetic promoters were 2.0X and 2.3X stronger under maltose than spacer core 1. Error bars represent standard deviation among biological triplicates. (c) Sequences of maltose-inducible minimal UAS are completely synthetic (E_{value} = 0.39). Adr1pBS is predicted by YEASTRACT[158] to be present in $UAS_{SynthMALA}$. Arrow indicates length and orientation of TFBS.

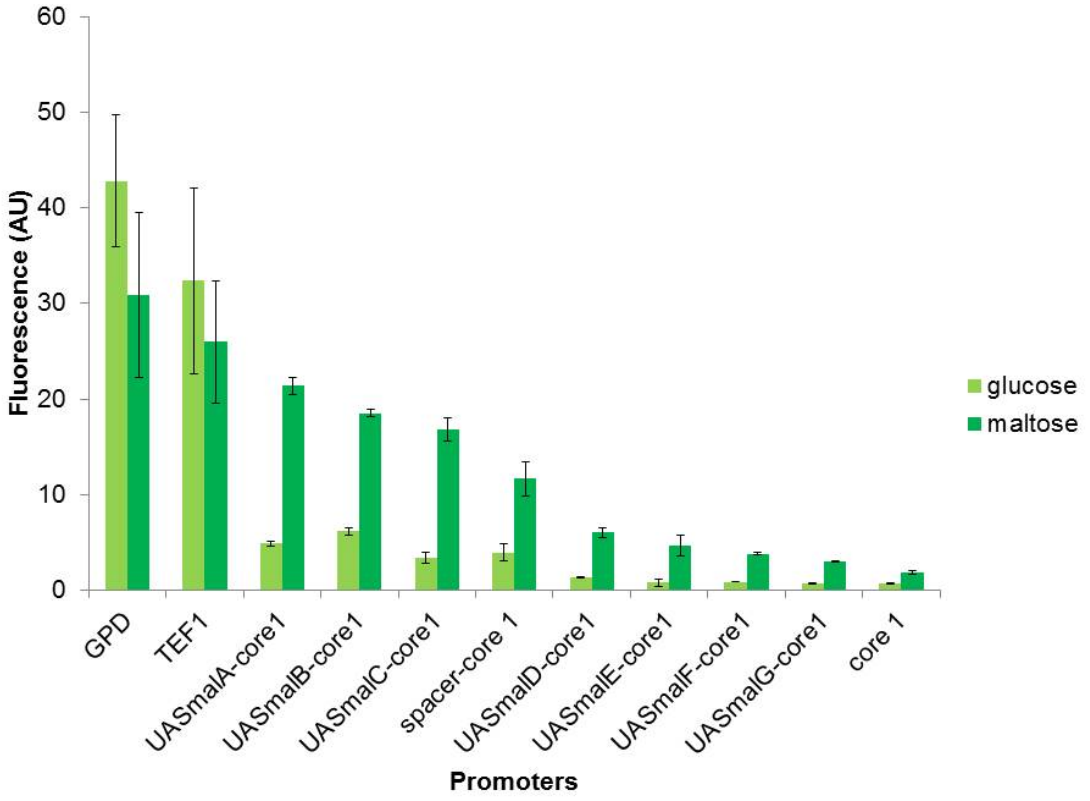


Figure 5.8: Synthetic maltose-inducible promoters possess a continuum of expression activation

Seven synthetic maltose-inducible promoters were engineering using iCAPTR. This set of seven maltose-inducible promoters exhibits a range of expression from as high as *TEF1* to as low as core 1 by itself. All synthetic maltose-inducible promoters contain the spacer present in spacer-core 1 construct. Error bars represent standard deviation among biological triplicates.

5.4 Discussions and Conclusion

Engineering efforts to increase inducible promoters thus far have relied on painstaking rational design with success highly dependent on previous research on the molecular biology of the system's parts. Furthermore, engineered promoters, like their native cousins still run several hundreds of base pairs (**Table 5.1**). To address this, I

established a high-throughput FACS-based isolation scheme, iCAPTR. Through implementation of iCAPTR, I isolated a set of galactose and maltose-inducible promoters of just ~100 bp.

The successful isolation of minimal inducible promoters demonstrates the utility of iCAPTR. I argue our isolation scheme could be adapted for other engineering applications. First, iCAPTR could be employed in any organism capable of expressing a fluorescent reporter. Though, I acknowledge high transformation efficiency would improve outcomes. Second, the inducer signal could be extended to non-molecular external stimuli such as temperature and light. Third, the sequence of the library could be tailored to suit specific sequence needs. For example, a potential GQ forming sequence library, like that used in **Chapter 4** could be employed to isolate GQs that fold and induce under GQ-stabilizing compounds. Lastly, while I focused on activating sequences in this work, repressing elements can be attained with this workflow. During iCAPTR improvements (**Figure 5.3d**), sequence elements that repress under non-inducing conditions and lose activity under inducing conditions were accidentally uncovered. In fact, these elements were able to decrease core element strength by 40%. Finally, iCAPTR requires only simple cloning techniques and a FACS instrument. Thus, this generic work flow is not only adaptable, but easy to perform.

However, the reason the other inducers failed to yield inducible promoters with iCAPTR remains unknown. From a biological standpoint, for an inducer to be an effective candidate, the inducer signal must be encoded by the cell internally, specifically

in the active states and cellular localizations of TFs, and these perturbed TFs must activate the expression of genes. To successfully isolate TFBS using iCAPTR, recruited TFs must induce detectable transcription of the fluorescent protein through interactions with the hybrid promoter library. To interact with the library, the TF ideally must be able to (i) bind to a sequence shorter than 20 bp, (ii) operate independently, and (iii) activate the synthetic core element at a distance of ~30 bp. Thus, perhaps the library scaffold was not effective TF bait, or maybe the TF perturbed by the chosen compounds cannot elicit detectable transcription. Most likely the failure to obtain promoters responsive to the set of the compounds selected is the result of a combination of these two factors. Thus, to further improve iCAPTR, a strategy for stimuli and library selection could be established.

Chapter 6: Major findings and proposal for future work

6.1 MAJOR FINDINGS

The research described in this work had one major goal: expand the promoter selection in the production hosts *S. cerevisiae* and *Y. lipolytica*. To accomplish this, the promoter was viewed as two distinct modular parts, the UAS and the core. Through hybridization of these elements using hybrid promoter engineering, full length functional promoters were assembled. In this dissertation, universal application of hybrid promoter engineering was demonstrated in the non-conventional yeast *Y. lipolytica* to establish high strength promoters. Next, drawing from these lesson, promoters in *S. cerevisiae* were engineered. Specifically, two main issues with current *S. cerevisiae* promoters were sought to be resolved: i) high sequence homology due to simple mutations or dependence on native scaffolds and ii) overall large size. To do so, UAS and the promoter core were individually minimized using *de novo* promoter libraries. First, fully synthetic, context independent, modular core elements were established. Next, a synthetic core element was employed in hybrid promoter libraries to isolate a set of synthetic constitutive minimal UAS elements. Moreover, a high throughput FACS-based isolation scheme, iCAPTR was developed to engineer short inducible promoters. With these approaches, the synthetic biology toolbox of *S. cerevisiae* was expanded with short, synthetic promoter parts.

In **Chapter 2**, the modularity of promoter parts, UAS and core were illustrated in a nonconventional yeast host. Specifically, UAS_{1B} derived from *Y. lipolytica* native

XPR2 promoter was used to amplify the minimized *S. cerevisiae* endogenous *LEU2* promoter in the oleaginous yeast *Y. lipolytica*. Linear increase in amplification with each additional UAS was revealed. Thus, this approach can be used to tune the strength of a promoter. In fact, up to 24 copies of UAS_{1B} can be used to increase the transcriptional output of the *LEU2* minimal promoter in *Y. lipolytica*. Moreover, these tandem-assembled UAS can be employed to increase expression of promoter core derivatives. In particular, we hybridized varying lengths of *Y. lipolytica* native *TEF* promoter with 8 and 16 UAS_{1B} repeats. The observed increase in promoter strength of these *TEF* derivatives highlights the utility of this approach to increase even the transcriptional output of native promoters of *Y. lipolytica*. In doing so, we overcame native expression limitations, and established the strongest promoter characterized in *Y. lipolytica*.

In **Chapters 3 through 5**, we focused on minimizing the yeast promoter by addressing each of the parts separately. In **Chapter 3**, a hybrid promoter approach was applied to library-derived core elements to establish a set of nine synthetic minimal core elements. To minimize the core, the minimal spacing required between two essential core components, the TATA box and the TSS was investigated. Specifically, the spacer sequence and length was randomized to create a collection of 27 libraries. Evaluation of the library strengths revealed a minimal spacing of ~30 bp. FACS was used to isolate 18 highly fluorescent candidates from a promoter library where a native UAS was amplifying the minimally sized core library. These candidates were subjected to five rigorous tests to ensure core elements were modular and robust. To ensure modularity,

core elements were hybridized with two native UAS, and a native TFBS to construct promoters displaying constitutive and inducible qualities. We confirmed robustness by assessing effects of DNA context. To do so, we alternated the direction of the expression cassette in the plasmid and observed expression differences by flow cytometry. Lastly, we showed that the promoters constructed with these core elements could drive expression of two genes, the initial fluorescent reporter and a β-galactosidase gene. In the end, nine minimal core elements from 18 initial candidates qualified as robust and modular. Lastly, sequence analysis revealed completely synthetic sequences, bearing no resemblance to yeast DNA sequences or to each other. The sequences span a wide range of GC content from 47 to 70% and possess a diversity of predicted TFBS, in terms of quantity, quality and directionality. Thus, this final set of nine minimal core promoters is not only robust and modular, but also sequence diverse. At just 56 bp in length, these elements are also the shortest promoter cores described to date in *S. cerevisae*.

In **Chapter 4**, we sought to synthetically minimize the UAS region of the promoter. To do so, we employed a synthetic short core element in eight hybrid promoter libraries to isolate constitutive minimal UAS elements. Specifically, we drew from four methods to establish a set of minimal UAS elements. Frist, we isolated six distinct constitutive 10 bp UAS from a N_{10}-core hybrid promoter library using FACS and tandem assembled these to create a strong triple tandem UAS to amplify a synthetic core. This triple tandem UAS was able to amplify a synthetic core to 70% of *GPD* strength. Second, we sequentially built upon three UAS in 10 bp increments to obtain a toolbox of

UAS elements that amplify a synthetic core to a continuum of strengths. To do so, we used the 10 bp UAS elements in multiple rounds of N_{10}-UAS-core hybrid promoter scaffolds. Candidates were isolated with FACS and nine promoters were confirmed with flow cytometry. Six of these UAS elements can activate a synthetic core better than native UAS elements, UAS_{CLB} and UAS_{CIT}. Third, we tapped into a G-rich library hybridized to a synthetic core to establish additional constitutive promoters. This G-rich library was comprised of UAS candidates with a high potential to form G-quadruplexes. Three constitutive promoters with strength peaking in late growth phases were derived from this library. The strongest reaches 61% of *GPD* strength. Fourth, Matthew Deaner and I successfully applied a dCAS9 promoter amplifier to augment five selected core elements. We rationally hybridized a sgRNA site to synthetic core elements and illustrated the capacity of a dCAS9-based TF to activate core elements via this site. In all, 18 library-derived UAS elements hybridized with synthetic core elements were characterized via flow cytometry. Two rationally-designed UAS elements (triple tandem UAS and sgRNA) were shown to activate core elements to strengths approaching *GPD*. This set of UAS elements hybridized to synthetic core elements collectively drive an expression range of 70-fold and reach strengths 70% of one of the strongest promoters known to yeast, *GPD*(*TDH3*). By minimizing the UAS region to several tens of base pairs, and then hybridizing these short activating elements with minimal promoter cores, the shortest promoters in *S. cerevisiae* were established.

In **Chapter 5**, we developed a generic high-throughput FACS-based isolation scheme, iCAPTR. In iCAPTR, inducible minimal promoters are sourced from a N_{20}-core hybrid promoter library scaffold via tandem FACS. As a proof of concept, we isolated a completely synthetic, galactose-inducible, minimal UAS element capable of activating core element 1 as strong as the native Gal4p TF. We improved the workflow and demonstrated iCAPTR's utility through the establishment of two high strength synthetic maltose-inducible promoters. iCAPTR could be implemented into other organisms, and has a number of variables that allow it to be easily adapted in other engineering efforts.

In summary, 9 core elements and 24 constitutive and inducible UAS elements are presented in this body of work. Each part described and characterized is completely synthetic and the entire set functions together to create full length promoters of desired strengths and qualities greatly expanding *S. cerevisiae's* toolbox. As a set, the parts can be assembled into promoters spanning a range of 70-fold in terms of expression. The short size allows for a dramatic reduction in DNA burden by more than 80%. This reduction makes these promoters similar in length with *E. coli*'s. Furthermore, due to their synthetic nature, the parts have low sequence homology to the genome, and thus reduce the possibility of homologues recombination. With these elements, we increase the tools available for *S. cerevisiae* allowing scientists and engineers alike the ability to manipulative gene expression as they see fit. Lastly, the generic workflow described and developed in these experiments are capable of being applied to other organisms to increase the quantity and quality of their engineering tools.

6.2 PROPOSALS FOR FUTURE WORK

The research presented in this work significantly contributes much needed minimal promoters to the *S. cerevisiae* toolbox. I acknowledge some parts may be more suited for certain engineering efforts than others. For example, some may be more robust during low pH conditions, making them more appropriate for acid production. Therefore, a complete characterization of the synthetic promoter parts presented in this dissertation should be performed. Specifically, I recommend experiments related to DNA context, culture conditions, and TF recruitment. Lastly, I suggest this set provides an unique opportunity to learn about transcription initiation.

First, the effects of DNA context on the promoters should be further explored. Their function in other expression cassettes could be evaluated. All of the promoter characterizations were done with *SPG5* as the terminator. Given the possibility of gene loop formation[212], in which terminator and promoter make physical contact and cooperatively participate in transcription, it would be useful to know the effects of other terminators on the synthetic promoters' function. All of the experiments were done with one plasmid, p416. Thus, promoters should be evaluated using different plasmids and in different genome locations. Understanding how DNA context like these affect the promoters would provide much needed limitations with respect to where and how they can be used in a system.

Second, the synthetic promoters should be characterized in a variety of conditions by altering the carbon source(s), nutrient abundance, degree of oxygenation and pH level.

These are common factors encountered in synthetic and metabolic engineering, especially in large fermentations where they are more difficult to control[213]. Anticipating how the promoters will function in specific culture conditions is crucial to predictably engineer a metabolic pathway. Promoter strength at these conditions can be measured at the mRNA level via qRT-PCR and at the protein level via flow cytometry. mRNA abundance would be especially helpful to determine the transcription activity of the promoters in various growth phases. Elucidation of mRNA levels can help tease apart proteins inherited from the mother cell, and therefore, provide an accurate assessment of promoter activity at the transcript level. mRNA levels driven by promoters constructed with G-rich UAS sequences would be particularly interesting. In my studies, expression as measured by flow cytometry suggested this promoter remains active into stationary phase, while the commonly-used *GPD* promoter appears to not (**Figure 4.10a**). In summary, by understanding the conditions under which the synthetic promoters are transcriptionally active, synthetic biologists can choose an appropriate promoter to express their gene of choice.

Lastly, to further aid in optimal promoter use, the functionality of identified putative TFBS should be explored. By elucidating TF(s) responsible for promoter strength, the use of a particular promoter affected by certain metabolic rewiring could be avoided. Furthermore, the identification of the functional binding sites in the promoters could allow for their use in a sTF/promoter system by targeting the sTF to bind to them. The simplest approach to determine which TFs are interacting is to evaluate promoter

strength under specific TF knockout strains. More definitively, the suspected TF can be overexpressed or knocked down to avoid having to tease out compensatory effects present in knock out strains[214]. Furthermore, promoters with dynamic expression strengths with respect to culture conditions as identified by experiments outlined above may prove to have dynamic TF binding as well. For these promoters, functional TFBS should be evaluated in several conditions to ensure complete identification of binding TF. Elucidating which TF(s) bind to the synthetic promoters not only enables synthetic biologists to use these parts more appropriately, but also can shed much needed light on how promoters use their specific sequences to operate.

In this regard, I propose that this set of *de novo* minimal promoters could be used to study promoter mechanisms. Previous studies have used synthetically generated libraries of promoters to understand how promoter elements, such as TFBS, nucleosome-depleting sequences and core sequences affect promoter activity[16,69,82]. Similarly, this set of synthetic promoters could be used to understand how promoter sequences affect transcription. In fact, the set presents a unique opportunity to understand how the sequence between the TATA box and the TSS affects transcription initiation since promoter core size, TATA-box and TSS variants present, and 5` UTR contribution are all held constant. Specifically, the effects of this sequence on TSS selection at single-nucleotide resolution could be studied. This set is particularly appropriate for these experiments due to the wide variation of sequences present in spacing region between the TATA box and TSS. For example, the GC content ranges from 47 to 73%. Additionally,

unlike genomic promoters, this set has only one functional TSS, and therefore, allows for the first time investigations into how adjacent sequences affect transcription at the single nucleotide level. To elucidate how the sequence affects start site selection, SMORE-seq (simultaneous mapping of RNA ends by sequencing) can be used to identify the 5` UTR of the mRNA at single-nucleotide resolution. This experiment could provide insights on how the sequences in the core may affect transcription initiation.

In conclusion, I recommend further characterizations of the synthetic minimal promoters presented in this body of work. Specifically, I recommend experiments related to DNA context, culture conditions, and TF recruitment to ensure promoters are being employed appropriately and optimally in synthetic systems. Lastly, I also suggest that this set of promoters can be used to understand how sequences between the TATA box and TSS of the promoter core affect transcription initiation.

Chapter 7: Materials and Methods

7.1 COMMON MATERIALS AND METHODS

7.1.1 Strains and Media

p416 yeast expression vectors were propagated in *Escherichia coli* DH10β. *E. coli* strains were cultivated in LB medium[215] (Teknova) at 37 °C with 225 RPM orbital shaking. LB was supplemented with 50 μg/mL ampicillin (Sigma) for plasmid maintenance and propagation. Yeast strains (BY4741 and BY4741 Δspt3 obtained from Euroscarf and Open Biosystems respectively) were cultivated on a yeast synthetic complete medium containing 6.7 g of Yeast Nitrogen Base (Difco)/L, 20 g glucose/L and a mixture of amino acids, and nucleotides without uracil (CSM, MP Biomedicals, Solon, OH). All medium was supplemented with 1.5% agar for solid media.

For *E. coli* transformations, 50 μL of electrocompetent *E. coli* DH10β[215] were mixed with 50 ng of ligated DNA and electroporated (2 mm Electroporation Cuvettes) (Bioexpress) with Biorad Genepulser Xcell at 2.5 kV. Transformants were recovered in 1 mL SOC Medium (Cellgro), plated on LB agar, and incubated overnight at 37 °C. Single clones were amplified in 2 mL LB medium and incubated overnight at 37 °C with 225 RPM orbital shaking. Plasmids were isolated (QIAprep Spin Miniprep Kit, Qiagen) and confirmed by sequencing.

For yeast transformations, 20 μL of chemically competent *S. cerevisiae* BY4741 were transformed with 1 μg of each appropriate purified plasmid using Frozen EZ Yeast Transformation II Kit (Zymo Research, Irvine, California, United States) according to the

manufacturer's instructions. Transformations were plated on CSM -Ura plates, and incubated for two days at 30 °C. Single colonies were picked at random into 2 mL of CSM -Ura liquid media and incubated at 30 °C for 2 days in 225 RPM orbital shaking. Yeast and bacterial strains were stored at -80 °C in 15% glycerol. Plasmids from yeast were isolated using Zymoprep™ Yeast Plasmid Miniprep II kit.

7.1.2 Cloning Procedures

All p416 plasmids were assembled using restriction enzyme based cloning techniques. Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). Sequences and details can be found in **Table 7.1**. PCR and double stranding reactions were performed with Phusion DNA Polymerase from New England Biolabs (Ipswich, MA) according to manufacturer specifications. Digestions were performed according to manufacturer's (NEB) instructions. PCR products and digestions were cleaned with a QIAquick PCR Purification Kit (Qiagen). Phosphatase reactions were performed with Antarctic Phosphatase (NEB) according to manufacturer's instructions and heat-inactivated for 10 min at 65 °C. Ligations (T4 DNA Ligase, Fermentas) were performed for 3-18 hrs in 5:1-10:1 insert to backbone ratio at 16-22 °C followed by heat inactivation at 65 °C for 10 min. Cloning procedures that required a gel extraction were done so with Fermentas GeneJET Gel Extraction Kit from Thermo Fisher Scientific (Waltham, MA).

olig o	Description	Sequence
1	Fwrd primer- PCR *VSP13* term	acgcGGATCCTCACATATGAAAGTATATACCCGCTTTTGT
2	Rv primer- PCR *VSP13* term	cagGGCGCGCCCGCGCTGCGGATATTTCTAA
3	Fwrd primer- PCR *SPG5* term	gcgtcctcgagCAAAGACGTTGTTTCATCGC
4	Rv primer- PCR *SPG5* term	GAC GCG GTA CCG CTT ATT TTC TGC CGA ATT TTC AT
5	Fwrd primer- PCR *PRM9* term	gcgtcctcgagACAGAAGACGGGAGACACT
6	Rv primer- PCR *PRM9* term	ATT TTC AAC ATC GTA TTT TCC GAA GC G GTA CCG ACG C
7	Fwrd primer- PCR *CYC1* term	GCGTCaagctttcatgtaattagttatgtcacgcttaca
8	Rv primer- PCR *CYC1* term	GAC GCT CTA GAC GAG CGT CCC AAA ACC TT
9	20nt library oligonucleotide	GCGTCaagctTATAAAAGNNNNNNNNNNNNNNNNNNAAAAAAGCATCGAAAAAATCTAGActag
10	30nt library oligonucleotide	GCGTCaagctTATAAAAGNNNNNNNNNNNNNNNNNNNNNNNNNNNNAAAAAAGCATCGAAAAAATCTAGActag
11	25nt library oligonucleotide	GCGTCaagctTATAAAAGNNNNNNNNNNNNNNNNNNNNNNNNAAAAAAGCATCGAAAAAATCTAGActag

Table 7.1: Oligonucleotides used in library assemblies, cloning and qPCR

12	Primer to double strand libraries	CTA GTC TAG ATT TTT TCG ATG CTT TTT T
13	Fwrd primer- PCR expression cassettes for flip	gcgtcctcgagCAAAGACGTTGTTTCATCGC
14	Rv primer- PCR expression cassettes for flip	GAC GCG GTA CCG CTT ATT TTC TGC CGA ATT TTC AT
15	10 bpUAS library oligo	GGCGCGCCNNNNNNNNNNTTAATTAActtgtaatattctacccAAGCTTggg
16	Primer used to double strand oligo15	CCC AAG CTT GGG TAG AAT ATT ACA AGT TAA TTA A
17	Fwrd primer- PCR *CYC1* promoter	gcgtcAAGCTTatttggcgagcgttgg
18	Rv primer- PCR *CYC1* promoter	GAC GCT CTA GAT TAG TGT GTG TAT TTG TGT TTG C
19	Fwrd primer- PCR *GPD* promoter	GAC GCT CTA GAA TCC GTC GAA ACT AAG TT
20	Rv primer- PCR *GPD* promoter	GCGTCaagcttagtttatcattatcaatactcgccattt
21	Fwrd primer- PCR *TEF1* promoter	atcattGGCGCGCCatagcttcaaaatgtttctactccttttttactcttc
22	Rv primer- PCR *TEF1* promoter	Aaggtctagaaaacttagattagattgctatgctttctttctaatgagc
23	Fwrd primer- PCR *GAL1* promoter	gcgtcGGCGCGCCtagtacggattagaagccgccg
24	Rv primer-PCR *GAL1* promoter	GAC GCT TAA TTA AGT TTT TTC TCC TTG ACG TTA AAG TAT AGA GGT
25	Oligo of G4BS4 for 5 bp spacing	gcgtcttaattaaCGGAAGACTCTCCTCCGaagcttgcgtc

Table 7.1: *Oligonucleotides used in library assemblies, cloning and qPCR continued*

26	Oligo to double strand oligo 25	GAC GCA AGC TTC GGA GGA GAG TCT TCC GTT AAT TAA GAC
27	Oligo of G4BS3 for 5 bp spacing	gcgtcttaattaaCGGGCGACAGCCCTCCGaagcttgcgtc
28	Oligo to double strand oligo 27	GAC GCA AGC TTC GGA GGG CTG TCG CCC GTT AAT TAA GAC GC
29	Oligo of G4BS4 for 17 bp spacing	gcgtcGGCGCGCCCGGAAGACTCTCCTCCGTTAATTAAgcgtc
30	Oligo to double strand oligo 29	GAC GCT TAA TTA ACG GAG GAG AGT CTT CCG GGC GCG CCG ACG C
31	Oligo of G4BS3 for 17 bp spacing	gcgtcGGCGCGCCCGGGCGACAGCCCTCCGTTAATTAAgcgtc
32	Primer to double strand oligo 31	GAC GCT TAA TTA ACG GAG GGC TGT CGC CCG GGC GCG CCG ACG C
33	Oligo of AT-rich neutral spacing	gcgtcTTAATTAActtgtaatattctaatcAAGCTTgcgtc
34	Primer to double strand oligo 33	GAC GCA AGC TTG ATT AGA ATA TTA CAA GTT AAT TAA GAC GC
35	Oligo of $UAS_{F\text{-}E\text{-}C}$	atcattGGCGCGCCCCTCCTTGAAACTGAAATTTTAGCATGTGATTAATTAAggcc g

Table 7.1: *Oligonucleotides used in library assemblies, cloning and qPCR continued*

36	Primer to double strand oligo 35	CGG CCT TAA TTA ATC ACA TGC TAC ACC GCC CCC
37	Fwrd primer- qPCR of yECitrine	TTCTGTCTCCGGTGAAGGTGAA
38	Rv primer- qPCR of yECitrine	TAAGGTTGGCCATGGAACTGGCAA
39	Fwrd primer- qPCR of Alg9	ATCGTGAAATTGCAGGCAGCTTGG
40	Rv primer- qPCR of Alg9	CATGGCAACGGCAGAAGGCAATAA
41	Oligo of N_{10}-GCC-UAS_B	atcattGGCGCGCCnnnnnnnnnGCCgctcaacggcTTAATTAAggccgCGCCCGt
42	Primer to double strand oligo 41 and 47	aCGGGCGcggccTTAATTAAgccgttgagc
43	Oligo of N_{10}-CC-UAS_D	atcattGGCGCGCCnnnnnnnnnnccacagaggggcTTAATTAAggccgCGCCCGt
44	Primer to double strand oligo 43 and 48	aCGGGCGcggccTTAATTAAgcccctctgt
45	Oligo of N_{10}-CC-UAS_F	atcattGGCGCGCCnnnnnnnnnnCCCCTCCTTGAATTAATTAAggccgCGCCCGt
46	Primer to double strand oligo 45 and 49	cggccTTAATTAATTCAAGGAGG

Table 7.1: Oligonucleotides used in library assemblies, cloning and qPCR continued

47	Oligo of N_{10}-CC-UAS_{aB}	tTATGGCGCGCCNNNNNNNNNNCCCCTGTATGGCGCCGCTCAACGGCTTAA

		TTAAggccg
48	Oligo of N_{10}-CGCC-$UAS_{αD}$	AAGGCGCGCCNNNNNNNNNNCGCCGTTCAGGAGGCCACAGAGGGGCTTAATTAAGG
49	Oligo of N_{10}-UASαF	tcattGGCGCGCCNNNNNNNNNNGAGGAGGGGGCCCCTCCTTGAATTAATTAAggccg
50	Oligo of G-rich library	TATGGCGCGCCNGGGNNNGGGNNNGGGNNNGGGNTTAATTAAcTTTGGGGCGATGGGtg
51	Primer used to double strand oligo 50	caCCCATCGCCCCAAAgTTAATTAA
52	Inducible UAS library oligo	atcattGGCGCGCCnnnnnnnnnnnnnnnnnnnTTAATTAAggccgCGCCCGt
53	Primer to double strand oligo 52	AGA CGA GCG CGG CCT TAA TTA A
54	Oligo of N_{10}-GCC-UAS_A	tcattGGCGCGCCNNNNNNNNNNGCCGGGGGCGGTGTTAATTAAggccg
55	Oligo of N_{10}-GCC-UAS_C	TCATTGGCGCGCCNNNNNNNNNNGCCTAGCATGTGATTAATTAAGGCCG
56	Oligo of N_{10}-CGCC-UAS_E	TCATTGGCGCGCCNNNNNNNNNNCGCCACTG
57	Oligo of N_{10}-CC-UAS_E	TCATTGGCGCGCCNNNNNNNNNNCCACTGAAATTTTTAATTAAGGCCG

Table 7.1: Oligonucleotides used in library assemblies, cloning and qPCR continued

7.1.3 Library Preparation

Libraries assembled are listed in **Table 7.2** and **Table 7.3** by double stranding oligonucleotides listed in **Table 7.1**. Double stranding was performed using Phusion DNA Polymerase from New England Biolabs (Ipswich, MA) with a touchdown annealing step followed by a 5 min. 72 °C elongation step in thermocycler. All core element libraries and UAS libraries were cloned into the HindIII/XbaI site and AscI/PacI site respectively in p416 using cloning techniques previously mentioned unless where noted in following text. All libraries were ligated in a 3:1 ligation ratio with 2 ug of backbone in 20 ul reaction volume. Library ligations were desalted for 10 min. on nitrocellulose membrane filters (MF^{TM} 0.025µm VSWP membrane filters) after 24 hrs of ligation at 16 °C. Entire ligation mixture was transformed into freshly prepared electrocompetent *E. coli* DH10β[215] and plated onto LB plates. *E. coli* colonies were counted, scraped, and plasmids were isolated (QIAprep Spin Miniprep Kit, Qiagen) and transformed into freshly prepared BY4741[216]. *E. coli* colony counts can be found in **Table 7.2** and **Table 7.3** as library size. After 48 hrs of flask growth, aliquots of each library covering five times the size of the yeast library in terms of number of cells were stored at -80 °C in 15% glycerol.

UAS	TATA-TSS spacing (bp)	Terminator	*E. Coli* library size (10^6)	Yeast cells isolated by FACS
None	20	*CYC1*	0.3	1350
None	30	*CYC1*	0.2	574
UAS_{CIT}	20	*CYC1*	2.2	3700
UAS_{CIT}	30	*CYC1*	1.4	456
UAS_{CLB}	20	*CYC1*	3.0*	Not sorted
UAS_{CLB}	30	*CYC1*	2.4*	Not sorted
None	20	*SPG5*	0.7	9874
None	30	*SPG5*	0.7	1463
UAS_{CIT}	20	*SPG5*	0.9	1300
UAS_{CIT}	30	*SPG5*	0.1	285
UAS_{CLB}	20	*SPG5*	0.1*	Not sorted
UAS_{CLB}	30	*SPG5*	0.04*	Not sorted
None	20	*PRM9*	0.5*	Not sorted
None	30	*PRM9*	0.7	31
UAS_{CIT}	20	*PRM9*	0.04	640
UAS_{CIT}	30	*PRM9*	0.1	325
UAS_{CLB}	20	*PRM9*	2.0*	Not sorted
UAS_{CLB}	30	*PRM9*	0.6*	Not sorted
None	25	*SPG5*	1.0	2332
None	25	*PRM9*	1.0	1469

Table 7.2: Core libraries constructed

None	25	*CYC1*	0.9	1418
UAS_{CIT}	25	*SPG5*	2.4	4250
UAS_{CIT}	25	*PRM9*	0.9	3232
UAS_{CIT}	25	*CYC1*	2.3	2900

Table 7.2: Core libraries constructed (continued)

UAS library	*E. Coli* library size (10^6)	Yeast cells isolated by FACS
N_{10}-spacer	1.3	140
N_{10}-UAS_A-spacer	1.5	Not sorted
N_{10}-UAS_B-spacer	2.2	Not sorted
N_{10}-UAS_C-spacer	2.3	Not sorted
N_{10}-UAS_D-spacer	1.8	Not sorted
N_{10}-UAS_E-spacer	1.1	Not sorted
N_{10}-UAS_F-spacer	0.9	Not sorted
N_{10}-*GCC*-UAS_A-spacer	2.3	Not sorted
N_{10}-*GCC*-UAS_B-spacer	4.3	1101
N_{10}-*GCC*-UAS_C-spacer	3.4	Not sorted
N_{10}-*CC*-UAS_D-spacer	4.1	834
N_{10}-*CC*-UAS_E-spacer	1.7	Not sorted
N_{10}-*CC*-UAS_F-spacer	9.8	1218
N_{10}-CGCC-UAS_E	3.4	Not sorted
N_{10}-CGCC-UAS_{aD}-spacer	5.4	640
N_{10}-CC-UAS_{aB}-spacer	3.2	632
N10-UAS_{aF}-spacer	1.0	890
$NG_3N_3G_3N_3G_3N_3G_3N$ –spacer	2.1	123

Table 7.3: UAS-core libraries constructed

7.1.4 Flow Cytometry and FACS

Yeast cultures were started in triplicate from glycerol stock, and were grown for 2 days to stationary phase. All yeast cultures were inoculated at an OD_{600} of 0.01 and grown to an OD_{600} of 0.7-0.9 in 30 C shaker. ΔSpt3 BY4741 (Fischer Scientific) strains under galactose growth were inoculated at OD of 0.10 due to lack of consistent growth at lower OD inoculations Fluorescence was analyzed (LSRFortssa Flow Cytometer, BD Biosciences) at excitation wavelength of 488 nm and detectin wavelength of 530 nm. 10,000 events were gathered at a flow rate of 2,000 events/second. An average fluorescence and standard deviation were calculated from the mean values for the biological replicates. Flow cytometry data was analyzed using FlowJo software. Top ~0.15% of a million and two million yeast cells from each core library and UAS library respectively were sorted using BD FACS Aria Cell sorter. Cells counted by the FACS instrument are listed in **Table 7.2** and **Table 7.3**. Sorted cells were grown for 24 hrs at 30 °C in 2 mL CSM-Ura media at 225 rpm. At least ten times the amount of cells isolated were plated onto CSM-Ura as isolated from the sorting. Colonies were randomly selected from plates and grown for 2 days to stationary phase in 2 mL CSM-Ura. Yeast cultures were inoculated at an OD_{600} of 0.01 and grown to an OD_{600} of 0.7-0.9 in a 30 °C shaker. Fluorescence was analyzed (LSRFortessa Flow Cytometer, BD Biosciences) and highly fluorescent cultures were streaked onto plates, picked in triplicate and glycerol stocked. Flow analysis was performed again in triplicate as previously described to ensure robustness.

7.2 MATERIALS AND METHODS FOR CHAPTER 2

7.2.1 Strains and Media

Yarrowia lipolytica strain PO1f (ATCC # MYA-2613), a leucine and uracil auxotroph devoid of any secreted protease activity [128] was used for all studies. *Y. lipolytica* PO1f containing plasmid was routinely cultivated at 30^{o}C with constant agitation in YSC-LEU media consisting of 20g/L glucose purchased from Fisher Scientific, 0.69g/L CSM-Leucine supplement purchased from MP Biomedicals, and 0.67g/L Yeast Nitrogen Base purchased from Becton, Dickinson, and Company. Solid media for *E.coli* and *Yarrowia lipolytica* was prepared by adding 15g/L agar (Teknova) to liquid media.

7.2.2 Cloning Procedures

All restriction enzymes were purchased from New England Biolabs and all digestions were performed according to standard protocols. PCR reactions were set up with recommended conditions using Phusion high fidelity DNA polymerase (Finnzymes). Ligation reactions were performed overnight at 16^{o}C using T4 DNA Ligase (Fermentas). Gel extractions were performed using the GeneClean gel extraction kit purchased from MP Biomedicals. Purification of small DNA fragments (<200 bp) generated during plasmid construction were performed using the MERmaid Spin Kit (Qbiogene). *E. coli* minipreps were performed using the Zyppy Plasmid Miniprep Kit (Zymo Research Corporation). *Y. lipolytica* minipreps were performed using Zymoprep Yeast Plasmid Miniprep II kit (Zymo Research Corporation). Transformation of *E. coli* strains was performed using the standard electroporator protocols [217]. Transformation of *Y. lipolytica* was performed using the Zymogen Frozen EZ Yeast Transformation Kit II (Zymo

Research Corporation). Genomic DNA was extracted from *Y. lipolytica* using the Wizard Genomic DNA Purification kit (Promega).

7.2.3 Calculation of Codon Adaptation Index

Codon Adaption Indices were calculated for the hrGFP, mStrawberry, EGFP, and yECitrine genes using the CAIcal server [218] and the codon usage table for *Y. lipolytica* available on the Codon Usage Database[219].

7.2.4 Plasmid Construction

Construction of endogenous promoter fluorescence cassettes (**Figure 7.1**): All plasmids employed for gene expression in *Y. lipolytica* were centromeric, replicative vectors based off plasmid pSl16-Cen1-1(227) [131], which was initially modified to include a new multicloning site and redubbed pMCSCen1. The cyc1 terminator (cyc1t) was amplified from p416-TEF-yECitrine [66,220] and inserted into pMCSCen1 with an EcoNI/BlpI digestion to form pMCScyc1t. Endogenous promoters as defined previously [135] EXP1 (JB096/97), GPAT (JB094/95), GPD (JB088/89), TEF (JB104/105), YAT1 (JB090/91), XPR2 (JB275/276), and XPR2fus (JB277/276) (**Table 2.1**) were amplified from *Yarrowia lipolytica* PO1f genomic DNA and ligated into pMCScyc1t using XmaI/AscI for FBA and BstBI/AscI for the rest to form pMCS-Promoter serial constructs. Reporter genes including yECitine [66] (JB083/084), mStrawberry (pmstrawberry, Clonetech) (JB153/155), EGFP [221] (JB156/158), hrGFP (pIRES-hrGFP, Agilent) (JB160/161) and lacZ [222] (JB312/313) were amplified using the indicated primers pairs and inserted into appropriate pMCS-Promoter constructs to form different pMCS-Promoter-Reporter constructs. Additionally, hrGFP and lacZ were inserted into pMCS-cyc1t to form plasmids pMCS-hrGFP and pMCS-lacZ.

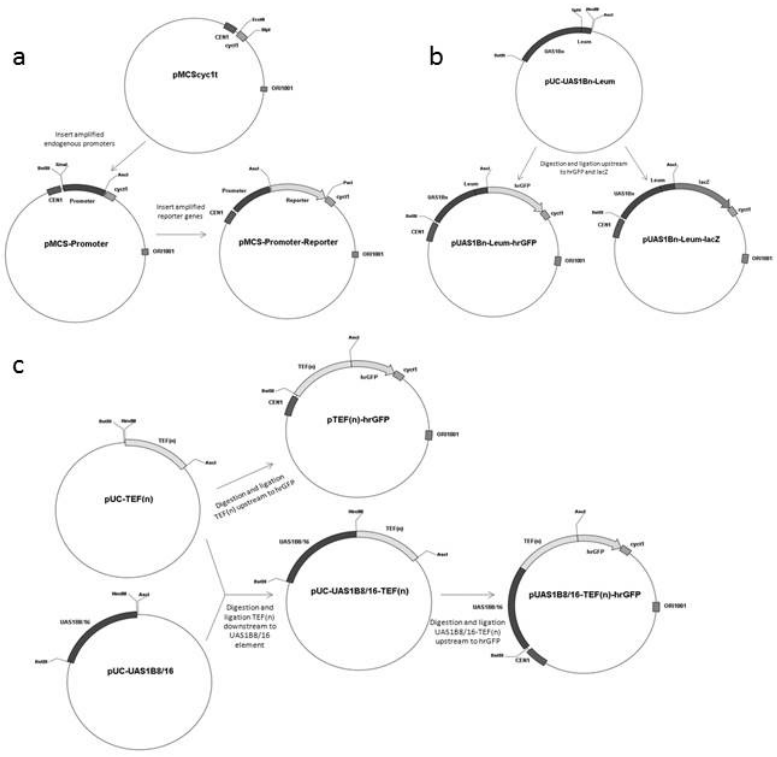


Figure 7.1: Construction of plasmids for Chapter 2

Constructions of UAS1B$_1$-Leum through UAS1B$_{32}$-Leum expression cassettes (**Figure 7.1b**): Primers JB162/163 amplified a 140bp minimal leucine promoter, Leum (**Table 2.1**) from plasmid pMCSCen1 that was inserted into pUC19 with SphI/HindIII to form pUC-Leum. Additionally, Leum was amplified using primers LQ19/20 and inserted into pMCS-HrGFP with BstBI/AscI to form plasmid pLeum-hrGFP.

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

Primers JB167/168 amplified a UAS1B oligo for BamHI-HF/SalI-HF insertion into pUC-UAS1B$_1$-Leum-No5'3' to form pUC-UAS1B$_2$-Leum-No5'3', while primers JB174/168 were used to create plasmid pUC-2dins from pUC-UAS1B$_1$-Leum-No5'3'. Primers JB169/170 amplified a third UAS1B for XbaI/BamHI-HF insertion into pUC-UAS1B$_2$-Leum-No5'3' to create pUC-UAS1B$_3$-Leum-No5'3', while primers JB174/170 were used to create plasmid pUC-3dins from pUC-UAS1B$_2$-Leum-No5'3'. Primers JB171/172 amplified a fourth UAS1B for XbaI/SacI-HF insertion into pUC-UAS1B$_3$-Leum-No5'3' to create pUC-UAS1B$_4$-Leum-No5'3', while primers JB173/172 were used to form pUC-4d5'ins from pUC-UAS1B$_3$-Leum-No5'3'. pUC-4d5'ins was edited with a SalI/SpHI mediated UAS1B replacement (from primers JB163/166) to form pUC-4dins.

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

Plasmid pUC16dblank was created by annealing together primers JB289/290, and inserting this 74bp oligo into pUC19 with NdeI/HindIII. Plasmid pUC-UAS1B$_8$-leum was digested with KpnI and then SphI-HF, and a 895bp fragment containing "KpnI-EcoRI-UAS1B$_8$-SphI" was extracted and inserted into pUC16dblank to form pUC16d8dins. pUC-UAS1B$_8$-leum was digested with EcoRI-HF, and a 901bp UAS1B$_8$ fragment was extracted and inserted into pUC16d8ins to create pUC16dins. A 1808bp UAS1B$_{16}$ fragment was BstBI/PstI gel extracted from pUC16dins and inserted into vectors pUC-UAS1B$_1$-Leum through pUC-UAS1B$_{16}$-Leum to form plasmids pUC-UAS1B$_{17}$-Leum through pUC-UAS1B$_{32}$-Leum.

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

Construction of TEF-based promoters and expression cassettes (**Figure 7.1c**): The 1004bp region upstream and including the TEF promoter was amplified from PO1f gDNA using primers LQ13 and JB105 and inserted into a pMCS-HrGFP expression cassette in which the hrGFP gene included its native ATG start site using BstBI/AscI to form plasmid pMCS-TEF(1004)-HrGFP. Promoters TEF(804) (LQ12), TEF(604) (LQ10), TEF(504) (LQ9), TEF(272) (LQ16), TEF(203) (LQ15), and TEF(136) (LQ14) were amplified from plasmid pMCS-TEF(1004)-HrGFP using the indicated primer and JB105. These seven promoters replaced TEF(1004) in pMCS-TEF(1004)-HrGFP to form the pMCS-TEF(n)-HrGFP core TEF promoter series (**Table 2.1**).

Plasmid pUC19-8d was formed by the insertion of a BstBI/HindIII digested, gel extracted UAS1B$_8$ fragment from pUC16d8dins in place of UAS1B$_{16}$-Leum in digested

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

7.2.5 Promoter Characterization with Flow Cytometry

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

7.2.6 Promoter Characterization through β–galactosidase Assay

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30^{o}C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven, and then normalized to an OD_{600} of 0.03 in 2ml fresh YSC-LEU and incubated for another 48 hours in the same conditions. The cultures were washed twice in 1ml Z buffer , resuspended in 1 ml of Z buffer, and their OD_{600} readings were recorded [137,138]. β-galactosidase assays were performed as described by Miller [137] using 10µl of chloroform-permeabilized cells, with a reaction length of 17 minutes.

7.2.7 Promoter Characterization through qRT-PCR

Y. lipolytica PO1f strains, transformed with different plasmids, were inoculated directly from glycerol stock (in biological triplicate) in YSC-LEU media for 48 hours at 30^{o}C with shaking, and normalized to an OD_{600} = 0.03 in 2ml fresh YSC-LEU media and incubated for another 48 hours at 30^{o}C in a rotary drum (CT-7, New Brunswick Scientific) at speed seven. The cells were pelleted and total RNA was extracted using the RiboPureTM-Yeast Kit (Ambion). 1000ng of RNA from each sample was used for a reverse transcription reaction with the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). A 1.2µl sample from each reaction was used to set up a qPCR reaction (in triplicate) with FastStart SYBR Green Master (Roche) using primers 5'-TCAGCGACTTCTTCATCCAGAGCTTC-3' and 5'-ACACGAACATCTCCTCGATCAGGTTG-3' as described in the manual with a non-template control. The reactions were run with Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems) using fast 96 well plates (Applied Biosystems).

The data was analyzed with ABI 7900HTsequence detection systems (version 2.4; Applied Biosystems).

7.2.8 Plasmid Stability Test

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

7.3 Materials and methods for Chapter 3

7.3.2 qPCR Assay

Yeast cultures were grown to OD_{600} of 0.7 to 0.9. Fluorescence was measured simultaneous with RNA extraction of 500 uL of culture (Quick-RNA Miniprep, Zymo Research Corporation). RNA was reverse-transcribed (High Capacity cDNA Reverse Transcription Kit, Applied Biosystems) and quantified in triplicate (SYBR Green PCR Master Mix, Life Technologies) after RNA extraction. Transcript levels were measured relative to that of a housekeeping gene (ALG9) (Viia 7 Real Time PCR Instrument, Life Technologies). Primers used for quantification are listed in **Table 7.1**.

7.3.3 LacZ assay

Yeast cultures were grown from triplicate glycerol stock for 2 days. Cultures were inoculated at 0.01 OD and grown overnight to OD_{600} of 0.7 to 0.9. Cells were

mixed with appropriate reagents and incubated according to instructions (AB Gal-Screen® System). Chemiluminescent signal was measured with Biotek Cytation 3 imaging reader.

7.4 MATERIALS AND METHODS FOR CHAPTER 4

7.4.1 Construction of dCAS9 expression vectors

All plasmids for expression of dCas9 were derived from the pJED103 vector series acquired from AddGene catalog #46921 (Weissman 2015). A general sgRNA expression cassette was cloned into the linearized pJED103-TDH3-dCas9-ADH1t plasmid (primers via Gibson Assembly. Next, a dCas9-VPR construct was PCR amplified from AddGene catalog #63801 **(**Chavez 2015) and then cloned via Gibson Assembly into pJED103-sgRNA-TDH3-ADH1t digested with NcoI and XhoI in order to linearize between TDH3p and ADH1t.. The synthetic guide sequence (SGS1) was constructed in this vector by inverse PCR using primers MD007 and MD008 followed by blunt-end ligation. To construct synthetic core promoters with the 20 base pair sgRNA site upstream, 5 core promoters were digested using AscI and PacI (NEB) and then ligated with annealed oligo pairs MD009/MD0010 and MD0011/MD0012 with complementary overhangs. Sequence of primers used are listed in **Table 7.4**.

Primer	Template	Sequence
MD007	pJED103	GTCACAGACTCAGGGTTCTAGTTTTAGAGCTAGAAATAGCAAGTTAAAA
MD008	pJED103	Gatcatttatctttcactgcgg
MD009	None	CGCGCCGTCACAGACTCAGGGTTCTACGGTTAAT
MD010	None	TAACCGTAGAACCCTGAGTCTGTGACGG
MD011	None	CGCGCCGTCACAGACTCAGGGTTCTACGGATACTATTTTCTGAGACGAATTAAT
MD012	None	TAATTCGTCTCAGAAAATAGTATCCGTAGAACCCTGAGTCTGTGACGG

Table 7.4: Primers used to construct dCAS9 expression vectors

7.5 Materials and methods for Chapter 5

7.5.1 iCAPTR FACS

Unless noted otherwise below, FACS and flow cytometry was run under same specifications as outlined in section 7.1.4. N_{20}-core 1 hybrid promoter library was sorted via FACS for weakly fluorescent cells (20 to 40 percentile of fluorescence). 197,000 weakly fluorescent cells were isolated from 2.0 million recorded fluorescent events. Weakly fluorescent cells were recovered overnight in 2 mL CSM-Ura liquid minimal media in 30 °C shaker. After recovery, cells were diluted to 0.015 OD_{600} in minimal media with inducer at concentrations ($C_{5\text{-}10}$) listed in **Table 5.2**, and in minimal media with no inducers. Cell culture under inducer concentrations of $C_{5\text{-}10}$ were grown to exponential phase (OD_{600} of 0.7-0.9) and sorted via FACS. Cell cultures with no inducers were allowed to grow to stationary phase (OD_{600} of ~3.0) before being exposed to inducer concentrations of $C_{40\text{-}50}$ as listed in **Table 5.2** for 4 hours. After 4 hours, cells

were sorted via FACS. Top 10% fluorescent cells out of 2.0 million fluorescent events were sorted from both cultures exposed to inducers $C_{5\text{-}10}$ and $C_{4\text{-}50}$ (**Table 7.5**). Pooled cells were allowed to recover overnight in 30 °C shaker with 2 mL CSM-Ura media before having plasmids isolated (Zymoprep™ Yeast Plasmid Miniprep II kit). Prepped plasmids were transformed into *E. Coli* DH10β retransformed for plasmid amplification. Amplified plasmids were transformed into freshly prepared BY4741[216] and cells were plated. 96 highly fluorescent colonies from plated retransformed BY4741 were picked qualitatively using a blue light, grown to stationary phase in 1 mL CSM-Ura liquid media and subjected to a -80 freeze/thaw cycle in 15% glycerol for at least 24 hours. Fluorescence was measured under inducing and non-inducing conditions at growth phase which candidate was initially sorted in with inducer, either exponential or stationary. Plasmids were isolated and retransformed from any candidates displaying more than a 2-fold induction. Strength of these candidates was confirmed in triplicate with flow cytometry.

Inducer	Cells isolated from FACS under exponential growth phase sort	Cells isolated from FACS under stationary growth phase sort
Caffeine	173,183	183,333
Itaconic acid	176,617	159,321
Aspirin	180,002	180,000
Eugenol	181,306	179,993
Vanillin	193,583	Not performed
Calcium	177,789	188,261
Copper	199,559	193,900
Maltose	185,382	Not performed

Table 7.5: Cells isolated from FACS. Cell count is based on FACS cell counting

7.5.2 Survival assays

Survival assays were performed with BY4741 with a p416 plasmid. Seed culture was inoculated from glycerol stock and grown for 2 days in 2 ml of liquid minimal media in a 14 mL culture tube. Seed culture was diluted to 0.01OD. In a 96-well plate, 100 µL diluted culture was grown under various inducer concentrations. OD_{600} was measured every 15 minutes over 24 hours using Bioscreen C (Growth- CurvesUSA, Piscataway, NJ, USA) with continuous shaking at 30 °C.

7.5.3 Preparation of inducer solutions

Inducers were prepped as follows: aqueous solution of 200 g/L D+ Maltose monohydrate (S-A), aqueous solution of 125 mg/mL Cupric sulfate pentahydrate (Amresco), aqueous solution of 111 mg/mL Calcium chloride, minimum 93%, granular, anhydrous (S-A), 330 mg/mL Eugenol, 99% (ACROS ORGANICS) in ethanol, aqueous solution of 45 g/L

itaconic acid, >99% (Aldrich), aqueous solution of 19.4 mg/mL, aqueous solution of 6.08 g/L 0-vanillin, >99% (ACROS ORGANICS), aqueous solution of 19.4 mg/mL caffeine, 99.9%, anhydrous (NuSci), 99.9%, 67.5 mg/ml aspirin in ethanol.

References

1. Liu, L., Redden, H. & Alper, H.S. Frontiers of yeast metabolic engineering: diversifying beyond ethanol and Saccharomyces. *Current Opinion in Biotechnology* **24**, 1023-1030 (2013).
2. Curran, K.A. & Alper, H.S. Expanding the chemical palate of cells by combining systems biology and metabolic engineering. *Metabolic Engineering* **14**, 289-297 (2012).
3. Nevoigt, E. Progress in metabolic engineering of *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* **72**, 379-412 (2008).
4. Miura, F. *et al.* A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proceedings of the National Academy of Sciences* **103**, 17846-17851 (2006).
5. Park, D., Morris, A.R., Battenhouse, A. & Iyer, V.R. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Research* (2014).
6. Hahn, S. & Young, E.T. Transcriptional Regulation in *Saccharomyces cerevisiae*: Transcription Factor Regulation and Function, Mechanisms of Initiation, and Roles of Activators and Coactivators. *Genetics* **189**, 705-736 (2011).
7. Basehoar, A.D., Zanton, S.J. & Pugh, B.F. Identification and Distinct Regulation of Yeast TATA Box-Containing Genes. *Cell* **116**, 699-709 (2004).
8. Rhee, H.S. & Pugh, B.F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295-301 (2012).
9. Bae, S.H., Han, H.W. & Moon, J. Functional analysis of the molecular interactions of TATA box-containing genes and essential genes. *PLoS One* **10**, e0120848 (2015).
10. Lubliner, S., Keren, L. & Segal, E. Sequence features of yeast and human core promoters that are predictive of maximal promoter activity. *Nucleic Acids Research* **41**, 5569-5581 (2013).
11. Struhl, K. Promoters, Activator proteins, and the mechanism of transcription initiation in yeast. *Cell* **49**, 295-297 (1987).
12. Zhang, Z. & Dietrich, F.S. Mapping of transcription start sites in Saccharomyces cerevisiae using 5′ SAGE. *Nucleic Acids Research* **33**, 2838-2851 (2005).
13. 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*. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 8562-8566 (1985).
14. Furter-Graves, E.M. & Hall, B.D. DNA sequence elements required for transcription initiation of the Schizosaccharomyces pombe ADH gene in Saccharomyces cerevisiae. *Molecular and General Genetics MGG* **223**, 407-416 (1990).
15. Rojas-Duran, M.F. & Gilbert, W.V. Alternative transcription start site selection leads to large differences in translation activity in yeast. *Rna-a Publication of the Rna Society* **18**, 2299-2305 (2012).
16. Lubliner, S. *et al.* Core promoter sequence in yeast is a major determinant of expression level. *Genome Research* **25**, 1008-17 (2015).

17. Stewart, A.J., Hannenhalli, S. & Plotkin, J.B. Why Transcription Factor Binding Sites Are Ten Nucleotides Long. *Genetics* **192**, 973-985 (2012).
18. Erb, I. & van Nimwegen, E. Transcription Factor Binding Site Positioning in Yeast: Proximal Promoter Motifs Characterize TATA-Less Promoters. *PLoS ONE* **6**, e24279 (2011).
19. Bilu, Y. & Barkai, N. The design of transcription-factor binding sites is affected by combinatorial regulation. *Genome Biology* **6**, R103 (2005).
20. Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).
21. Schaefke, B., Wang, T.Y., Wang, C.Y. & Li, W.H. Gains and Losses of Transcription Factor Binding Sites in *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *Genome Biology Evolution* **7**, 2245-57 (2015).
22. Vinces, M.D., Legendre, M., Caldara, M., Hagihara, M. & Verstrepen, K.J. Unstable Tandem Repeats in Promoters Confer Transcriptional Evolvability. *Science* **324**, 1213-1216 (2009).
23. Brukner, I., Sánchez, R., Suck, D. & Pongor, S. Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides. *The EMBO Journal* **14**, 1812-1818 (1995).
24. Tirosh, I., Berman, J. & Barkai, N. The pattern and evolution of yeast promoter bendability. *Trends in Genetics* **23**, 318-21 (2007).
25. Hershman, S.G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **36**, 144-156 (2008).
26. Capra, J.A., Paeschke, K., Singh, M. & Zakian, V.A. G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in *Saccharomyces cerevisiae*. *Plos Computational Biology* **6**(2010).
27. Wong, B., Chen, S., Kwon, J.-A. & Rich, A. Characterization of Z-DNA as a nucleosome-boundary element in yeast *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2229-2234 (2007).
28. Oh, D.B., Kim, Y.G. & Rich, A. Z-DNA-binding proteins can act as potent effectors of gene expression in vivo. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16666-71 (2002).
29. Hahn, S. Structure and mechanism of the RNA polymerase II transcription machinery. *Nature Structural Molecular Biology* **11**, 394-403 (2004).
30. Thomas, M.C. & Chiang, C.-M. The General Transcription Machinery and General Cofactors. *Critical Reviews in Biochemistry and Molecular Biology* **41**, 105-178 (2006).
31. Cheung, Alan C.M. & Cramer, P. A Movie of RNA Polymerase II Transcription. *Cell* **149**, 1431-1437.
32. Grunberg, S., Warfield, L. & Hahn, S. Architecture of the RNA polymerase II preinitiation complex and mechanism of ATP-dependent promoter opening. *Nature Structural and Molecular Biology* **19**, 788+ (2012).
33. Mogno, I., Vallania, F., Mitra, R.D. & Cohen, B.A. TATA is a modular component of synthetic promoters. *Genome Research* **20**, 1391-7 (2010).

34. Parvin, J.D. & Sharp, P.A. DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* **73**, 533-540 (1993).
35. Hausner, W., Wettach, J., Hethke, C. & Thomm, M. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *Journal of Biological Chemistry* **271**, 30144-8 (1996).
36. Qureshi, S.A., Bell, S.D. & Jackson, S.P. Factor requirements for transcription in the Archaeon *Sulfolobus shibatae*. *The EMBO Journal* **16**, 2927-2936 (1997).
37. Struhl, K. Duality of TBP, the universal transcription factor. *Science* **263**, 1103-4 (1994).
38. Hernandez, N. TBP, a universal eukaryotic transcription factor? *Genes and Development* **7**, 1291-308 (1993).
39. Faiger, H., Ivanchenko, M., Cohen, I. & Haran, T.E. TBP flanking sequences: asymmetry of binding, long-range effects and consensus sequences. *Nucleic Acids Research* **34**, 104-119 (2006).
40. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512-520 (1993).
41. Kim, J.L., Nikolov, D.B. & Burley, S.K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520-527 (1993).
42. Field, Y. *et al.* Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Computational Biology* **4**, e1000216 (2008).
43. Lee, T.I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* **405**, 701-4 (2000).
44. Bonnet, J. *et al.* The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. *Genes and Development* **28**, 1999-2012 (2014).
45. Richmond, T.J. & Davey, C.A. The structure of DNA in the nucleosome core. *Nature* **423**, 145+ (2003).
46. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genetics* **39**, 1235-1244 (2007).
47. Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Natural and Structural Molecular Biology* **20**, 267-73 (2013).
48. Fan, X. *et al.* Nucleosome depletion at yeast terminators is not intrinsic and can occur by a transcriptional mechanism linked to 3'-end formation. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 17945-17950 (2010).
49. Lee, C., Shibata, Y., Rao, B., Strahl, B. & Lieb, J. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nature Genetics* **36**, 900 - 905 (2004).
50. Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O. & Schreiber, S.L. Global nucleosome occupancy in yeast. *Genome Biology* **5**, R62-R62 (2004).
51. Iyer, V. & Struhl, K. Mechanism of differential utilization of the his3 TR and TC TATA elements. *Mol Cell Biol* **15**, 7059 - 7066 (1995).
52. Liu, X., Lee, C.K., Granek, J.A., Clarke, N.D. & Lieb, J.D. Whole-genome comparison of Leu3 binding in vitro and in vivo reveals the importance of nucleosome occupancy in target site selection. *Genome Research* **16**, 1517-28 (2006).

53. Lee, T.I. *et al.* Transcriptional regulatory networks in *Sacch aromyces cerevisiae*. *Science* **298**, 799-804 (2002).
54. Ozonov, E.A. & van Nimwegen, E. Nucleosome free regions in yeast promoters result from competitive binding of transcription factors that interact with chromatin modifiers. *PLoS Computational Biology* **9**, e1003181 (2013).
55. Tirosh, I., Barkai, N. & Verstrepen, K. Promoter architecture and the evolvability of gene expression. *Journal of Biology* **8**, 95 (2009).
56. Cui, F., Cole, H.A., Clark, D.J. & Zhurkin, V.B. Transcriptional activation of yeast genes disrupts intragenic nucleosome phasing. *Nucleic Acids Research* **40**, 10753-64 (2012).
57. Svaren, J. & Horz, W. Transcription factors vs nucleosomes: regulation of the PHO5 promoter in yeast. *Trends in Biochemical Sciences* **22**, 93-7 (1997).
58. Reinke, H. & Horz, W. Histones are first hyperacetylated and then lose contact with the activated PHO5 promoter. *Mol Cell* **11**, 1599-607 (2003).
59. Fedor, M.J. & Kornberg, R.D. Upstream activation sequence-dependent alteration of chromatin structure and transcription activation of the yeast GAL1-GAL10 genes. *Molecular Cell Biology* **9**, 1721-32 (1989).
60. Da Silva, N.A. & Srikrishnan, S. Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Research* **12**, 197-214 (2012).
61. Koivistoinen, O.M. *et al.* Glycolic acid production in the engineered yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. *Microbial Cell Factories* **12**(2013).
62. Hector, R.E., Dien, B.S., Cotta, M.A. & Mertens, J.A. Growth and fermentation of D-xylose by *Saccharomyces cerevisiae* expressing a novel D-xylose isomerase originating from the bacterium *Prevotella ruminicola* TC2-24. *Biotechnology for Biofuels* **6**(2013).
63. Du, J., Yuan, Y., Si, T., Lian, J. & Zhao, H. Customized optimization of metabolic pathways by combinatorial transcriptional engineering. *Nucleic Acids Research* **40**, e142 (2012).
64. Jensen, P.R. & Hammer, K. Artificial promoters for metabolic optimization. *Biotechnology and Bioengineering* **58**, 191-195 (1998).
65. Hammer, K., Mijakovic, I. & Jensen, P.R. Synthetic promoter libraries – tuning of gene expression. *Trends in Biotechnology* **24**, 53-55 (2006).
66. Alper, H., Fischer, C., Nevoigt, E. & Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 12678-12683 (2005).
67. Nevoigt, E. *et al.* Engineering of Promoter Replacement Cassettes for Fine-Tuning of Gene Expression in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology* **72**, 5266-5273 (2006).
68. Ligr, M., Siddharthan, R., Cross, F.R. & Siggia, E.D. Gene Expression From Random Libraries of Yeast Promoters. *Genetics* **172**, 2113-2122 (2006).
69. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology* **30**, 521-530 (2012).
70. Ellis, T., Wang, X. & Collins, J.J. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nature Biotechnology* **27**, 465-471 (2009).

71. Jeppsson, M., Johansson, B., Jensen, P.R., Hahn-Hägerdal, B. & Gorwa-Grauslund, M.F. The level of glucose-6-phosphate dehydrogenase activity strongly influences xylose fermentation and inhibitor sensitivity in recombinant *Saccharomyces cerevisiae* strains. *Yeast* **20**, 1263-1272 (2003).
72. Blount, B.A., Weenink, T., Vasylechko, S. & Ellis, T. Rational Diversification of a Promoter Providing Fine-Tuned Expression and Orthogonal Regulation for Synthetic Biology. *PLoS ONE* **7**, e33279 (2012).
73. Mazumder, M. & McMillen, D.R. Design and characterization of a dual-mode promoter with activation and repression capability for tuning gene expression in yeast. *Nucleic Acids Research* **42**, 9514-9522 (2014).
74. Murphy, K.F., Balázsi, G. & Collins, J.J. Combinatorial promoter design for engineering noisy gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12726-12731 (2007).
75. McIsaac, R.S., Gibney, P.A., Chandran, S.S., Benjamin, K.R. & Botstein, D. Synthetic biology tools for programming gene expression without nutritional perturbations in *Saccharomyces cerevisiae*. *Nucleic Acids Research* (2014).
76. Rantasalo, A. *et al.* Synthetic Transcription Amplifier System for Orthogonal Control of Gene Expression in Saccharomyces cerevisiae. *PLoS ONE* **11**, e0148320 (2016).
77. Blazeck, J., Garg, R., Reed, B. & Alper, H.S. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering* **109**(2012).
78. Liang, J., Ning, J.C. & Zhao, H. Coordinated induction of multi-gene pathways in *Saccharomyces cerevisiae*. *Nucleic Acids Research* (2012).
79. Blazeck, J., Garg, R., Reed, B. & Alper, H. Controlling promoter strength and regulation in *Saccharomyces cerevisiae* using synthetic hybrid promoters. *Biotechnology and Bioengineering* **109**, 2884 - 2895 (2012).
80. Yarrington, R.M., Goodrum, J.M. & Stillman, D.J. Nucleosomes Are Essential for Proper Regulation of a Multigated Promoter in *Saccharomyces cerevisiae*. *Genetics* **202**, 551-563 (2016).
81. Kim, S., Lee, K., Bae, S.-J. & Hahn, J.-S. Promoters inducible by aromatic amino acids and γ-aminobutyrate (GABA) for metabolic engineering applications in *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology* **99**, 2705-2714 (2015).
82. Raveh-Sadka, T. *et al.* Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nature Genetics* **44**, 743-750 (2012).
83. Ichikawa, Y. *et al.* Sequence-directed nucleosome-depletion is sufficient to activate transcription from a yeast core promoter in vivo. *Biochemical and Biophysical Research Communications* **476**, 57-62 (2016).
84. Rajkumar, A.S. *et al.* Engineering of synthetic, stress-responsive yeast promoters. *Nucleic Acids Research* (2016).
85. McIsaac, R.S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research* **41**, e57-e57 (2013).

86. Ikushima, S., Zhao, Y. & Boeke, J.D. Development of a Tightly Controlled Off Switch for *Saccharomyces cerevisiae* Regulated by Camphor, a Low-Cost Natural Product. *G3: Genes|Genomes|Genetics* **5**, 1983-1990 (2015).
87. Ottoz, D.S.M., Rudolf, F. & Stelling, J. Inducible, tightly regulated and growth condition-independent transcription factor in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **42**, e130-e130 (2014).
88. Cuperus, J.T., Lo, R.S., Shumaker, L., Proctor, J. & Fields, S. A tetO toolkit to alter expression of genes in *Saccharomyces cerevisiae*. *ACS synthetic biology* **4**, 842-852 (2015).
89. Gertz, J., Siggia, E.D. & Cohen, B.A. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature* **457**, 215-218 (2009).
90. 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*. *Biotechnology Journal* **11**, 866-876 (2016).
91. Farzadfard, F., Perli, S.D. & Lu, T.K. Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synthetic Biology* **2**, 604-613 (2013).
92. Curran, K.A. *et al.* Design of synthetic yeast promoters via tuning of nucleosome architecture. *Nature Communications* **5**(2014).
93. Blazeck, J. & Alper, H.S. Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnology Journal* **8**, 46-58 (2013).
94. Anderson, J.D. & Widom, J. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol Cell Biol* **21**, 3830-9 (2001).
95. Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr Opin Struct Biol* **19**, 65-71 (2009).
96. Louvion, J.-F., Havaux-Copf, B. & Picard, D. Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene* **131**, 129-134 (1993).
97. Gao, C.Y. & Pinkham, J.L. Tightly regulated, beta-estradiol dose-dependent expression system for yeast. *Biotechniques* **29**, 1226-31 (2000).
98. Quintero, M.J., Maya, D., Arevalo-Rodriguez, M., Cebolla, A. & Chavez, S. An improved system for estradiol-dependent regulation of gene expression in yeast. *Microbology Cell Factories* **6**, 10 (2007).
99. McIsaac, R.S. *et al.* Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell* **22**, 4447-4459 (2011).
100. Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. Zif268 protein–DNA complex refined at 1.6å: a model system for understanding zinc finger–DNA interactions. *Structure* **4**, 1171-1180 (1996).
101. Beerli, R.R. & Barbas, C.F. Engineering polydactyl zinc-finger transcription factors. *Nature Biotechnology* **20**, 135-141 (2002).
102. Jamieson, A.C., Wang, H. & Kim, S.-H. A zinc finger directory for high-affinity DNA recognition. *Proceedings of the National Academy of Sciences* **93**, 12834-12839 (1996).

103. Beerli, R.R., Segal, D.J., Dreier, B. & Barbas, C.F. Toward controlling gene expression at will: Specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14628-14633 (1998).
104. Dreier, B., Segal, D.J. & Barbas, C.F. Insights into the molecular recognition of the 5 '-GNN-3 ' family of DNA sequences by zinc finger domains. *Journal of Molecular Biology* **303**, 489-502 (2000).
105. McIsaac, R.S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research* **41**, e57 (2013).
106. Blazeck, J., Liu, L., Redden, H. & Alper, H. Tuning Gene Expression in *Yarrowia lipolytica* by a Hybrid Promoter Approach. *Applied and Environmental Microbiology* **77**, 7905-7914 (2011).
107. Beopoulos, A. *et al.* Control of Lipid Accumulation in the Yeast *Yarrowia lipolytica. Applied and Environmental Microbiology* **74**, 7779-7789 (2008).
108. Beopoulos, A., Chardot, T. & Nicaud, J.M. *Yarrowia lipolytica*: A model and a tool to understand the mechanisms implicated in lipid accumulation. *Biochimie* **91**, 692-6 (2009).
109. Beopoulos, A. *et al. Yarrowia lipolytica* as a model for bio-oil production. *Progress in Lipid Research* **48**, 375-387 (2009).
110. Andre, A. *et al.* Biotechnological conversions of bio-diesel-derived crude glycerol by *Yarrowia lipolytica* strains. *Engineering in Life Sciences* **9**, 468-478 (2009).
111. Barth, G. & Gaillardin, C. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. *Fems Microbiology Reviews* **19**, 219-237 (1997).
112. Fickers, P. *et al.* Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *Fems Yeast Research* **5**, 527-543 (2005).
113. Muller, S., Sandal, T., Kamp-Hansen, P. & Dalboge, H. Comparison of expression systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. *Yeast* **14**, 1267-1283 (1998).
114. Madzak, C., Gaillardin, C. & Beckerich, J.M. Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *Journal of Biotechnology* **109**, 63-81 (2004).
115. Dominguez, A. *et al.* Non-conventional yeasts as hosts for heterologous protein production. *International Microbiology* **1**, 131-42 (1998).
116. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35-44 (2004).
117. Sherman, D. *et al.* Genolevures complete genomes provide data and tools for comparative genomics of hemiascomycetous yeasts. *Nucleic Acids Research* **34**, D432-D435 (2006).
118. Davidow, L.S. *et al.* Integrative transformation of the yeast *Yarrowia lipolytica*. *Current Genetics* **10**, 39-48 (1985).
119. Chen, D.C., Yang, B.C. & Kuo, T.T. One-step transformation of yeast in stationary phase. *Current Genetics* **21**, 83-84 (1992).

120. Fournier, P. *et al.* Colocalization of centromeric and replcative functions on autonomously replicating sequences isolated from the yeast *Yarrowia liplytica*. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4912-4916 (1993).
121. Fickers, P., Le Dall, M.T., Gaillardin, C., Thonart, P. & Nicaud, J.M. New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *Journal of Microbiological Methods* **55**, 727-737 (2003).
122. Vernis, L. *et al.* An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology* **17**, 1995-2004 (1997).
123. Vernis, L. *et al.* Only centromeres can supply the partition system required for ARS function in the yeast *Yarrowia lipolytica*. *Journal of Molecular Biology* **305**, 203-217 (2001).
124. Matsuoka, M. *et al.* Analysis of regions essential for the function of chromosomal replicator sequences from *Yarrowia lipolytica*. *Molecular & General Genetics* **237**, 327-333 (1993).
125. Juretzek, T. *et al.* Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast* **18**, 97-113 (2001).
126. Ledall, M.T., Nicaud, J.M. & Gaillardin, C. Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Current Genetics* **26**, 38-44 (1994).
127. Vanheerikhuizen, H. *et al.* Heterogeneity in the ribosomal-RNA genes of the yeast *Yarrowia lipolytica* - cloning and analysis of 2 size classes of repeats. *Gene* **39**, 213-222 (1985).
128. Madzak, C., Treton, B. & Blanchin-Roland, S. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. *Journal of Molecular Microbiology and Biotechnology* **2**, 207-216 (2000).
129. Madzak, C., Blanchin-Roland, S., Otero, R.R.C. & Gaillardin, C. Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. *Microbiology-Uk* **145**, 75-87 (1999).
130. Blanchinroland, S., Otero, R.R.C. & Gaillardin, C. 2 Upstream Activation Sequences Control the Expression of the Xpr2 Gene in the Yeast *Yarrowia lipolytica*. *Molecular and Cellular Biology* **14**, 327-338 (1994).
131. 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. *Journal of Biosciences and Bioengineering* **105**, 571-8 (2008).
132. Gasmi, N., Fudalej, F., Kallel, H. & Nicaud, J.M. A molecular approach to optimize hIFN alpha 2b expression and secretion in *Yarrowia lipolytica*. *Applied Microbiology and Biotechnology* **89**, 109-119.
133. Haas, J., Park, E.C. & Seed, B. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Current Biology* **6**, 315-324 (1996).

134. Nicaud, J.M., Fabre, E., Becherich, J.M., Fournier, P. & Gaillardin, C. Cloning, sequencing and amplification of the alkaline extracellular protease (XPR2) gene of the yeast *Yarrowia lipolytica*. *Journal of Biotechnology* **12**, 285-297 (1989).
135. Damude, H.G.H., Gillies, Peter John, Macool, Daniel Joseph, Picataggio, Stephen K., Pollak, Dana Walters M. , Ragghianti, James John, Xue, Zhixiong, Yadav, Narendra S., Zhang, Hongxiang , Zhu, Quinn Qun. High eicosapentaenoic acid producing strains of *Yarrowia lipolytica*. (United States, 2006).
136. Damude, H.G. *et al.* High eicosapentaenoic acid producing strains of *Yarrowia lipolytica*. (Google Patents, 2011).
137. Miller, J.H. *Experiments in molecular genetics*, xvi, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1972).
138. Gaillardin, C. & Ribet, A.M. LEU2 directed expression of beta-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica*. *Current Genetics* **11**, 369-75 (1987).
139. Young, E. & Alper, H. Synthetic biology: tools to design, build, and optimize cellular processes. *Journal of Biomedicine and Biotechnology* **2010**, 130781 (2010).
140. Liu, L., Reed, B. & Alper, H. From Pathways to Genomes and Beyond: The Metabolic Engineering Toolbox and Its Place in Biofuels Production. *Green* **1**, 81-95 (2011).
141. Juretzek, T., Wang, H., Nicaud, J.M., Mauersberger, S. & Barth, G. Comparison of Promoters Suitable for Regulated Overexpression of β-Galactosidase in the Alkane-Utilizing Yeast *Yarroiwa lipolytica*. *Biotechnoogy and Bioprocess Engineering* **5**, 320-326 (2000).
142. Redden, H. & Alper, H.S. The development and characterization of synthetic minimal yeast promoters. *Nature Communication* **6**(2015).
143. Smanski, M.J. *et al.* Functional optimization of gene clusters by combinatorial design and assembly. *Nature Biotechnology* **32**, 1241-1249 (2014).
144. Singh, V. Recent advancements in synthetic biology: Current status and challenges. *Gene* **535**, 1-11 (2014).
145. Redden, H., Morse, N. & Alper, H.S. The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Research*, (2014).
146. Zhang, G. & Darst, S.A. Structure of the *Escherichia coli* RNA Polymerase α Subunit Amino-Terminal Domain. *Science* **281**, 262-266 (1998).
147. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. *Science* **292**, 1863-1876 (2001).
148. Blount, B.A., Weenink, T., Vasylechko, S. & Ellis, T. Rational Diversification of a Promoter Providing Fine-Tuned Expression and Orthogonal Regulation for Synthetic Biology. *Plos One* **7**(2012).
149. Iyer, V. & Struhl, K. Poly(Da-Dt), a Ubiquitous Promoter Element That Stimulates Transcription Via Its Intrinsic DNA-Structure. *Embo Journal* **14**, 2570-2579 (1995).
150. Blazeck, J., Garg, R., Reed, B. & Alper, H.S. Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnology and Bioengineering* **109**, 2884-2895 (2012).
151. Khalil, Ahmad S. *et al.* A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions. *Cell* **150**, 647-658 (2012).

152. Afonso, B., Silver, P.A. & Ajo-Franklin, C.M. A synthetic circuit for selectively arresting daughter cells to create aging populations. *Nucleic Acids Research* **38**, 2727-2735 (2010).
153. Leuther, K.K., Bushnell, D.A. & Kornberg, R.D. Two-Dimensional Crystallography of TFIIB–and IIE–RNA Polymerase II Complexes: Implications for Start Site Selection and Initiation Complex Formation. *Cell* **85**, 773-779 (1996).
154. Carninci, P. *et al.* Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* **38**, 626-635 (2006).
155. Rosenkrantz, M., Kell, C.S., Pennell, E.A., Webster, M. & Devenish, L.J. Distinct Upstream Activation Regions for Glucose-Repressed and Derepressed Expression of the Yeast Citrate Synthase Gene Cit1. *Current Genetics* **25**, 185-195 (1994).
156. Curran, K.A., Karim, A.S., Gupta, A. & Alper, H.S. Use of expression-enhancing terminators in *Saccharomyces cerevisiae* to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metabolic Engineering* **19**, 88-97 (2013).
157. Van Slyke, C. & Grayhack, E.J. The essential transcription factor Reb1p interacts with the CLB2 UAS outside of the G2/M control region. *Nucleic Acids Research* **31**, 4597-4607 (2003).
158. Teixeira, M.C. *et al.* The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **42**, D161-D166 (2014).
159. Mohibullah, N. & Hahn, S. Site-specific cross-linking of TBP in vivo and in vitro reveals a direct functional interaction with the SAGA subunit Spt3. *Genes & Development* **22**, 2994-3006 (2008).
160. Bhaumik, S.R. & Green, M.R. Differential Requirement of SAGA Components for Recruitment of TATA-Box-Binding Protein to Promoters In Vivo. *Molecular and Cellular Biology* **22**, 7365-7371 (2002).
161. Dudley, A.M., Rougeulle, C. & Winston, F. The Spt components of SAGA facilitate TBP binding to a promoter at a post-activator-binding step in vivo. *Genes & Development* **13**, 2940-2945 (1999).
162. Leavitt, J. & Alper, H. Advances and Current Limitations in Transcript-level Control of Gene Expression. *Current Opinion in Biotechnology* (2015).
163. Harbison, C.T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).
164. Bochman, M.L., Paeschke, K. & Zakian, V.A. DNA secondary structures: stability and function of G-quadruplex structures. *Nature reviews. Genetics* **13**, 770-780 (2012).
165. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *Journal of the American Chemical Society* **126**, 16405-16415 (2004).
166. Guédin, A., Gros, J., Alberti, P. & Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Research* **38**, 7858-7868 (2010).
167. Jinek, M. *et al.* A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816-821 (2012).
168. Garneau, J.E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010).

169. Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Meth* **12**, 326-328 (2015).
170. O'Connell, M.R. *et al.* Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**, 263-266 (2014).
171. Dhakal, S. *et al.* Structural and mechanical properties of individual human telomeric G-quadruplexes in molecularly crowded solutions. *Nucleic Acids Research* **41**, 3915-3923 (2013).
172. Rakhit, R., Navarro, R. & Wandless, Thomas J. Chemical Biology Strategies for Posttranslational Control of Protein Function. *Chemistry & Biology* **21**, 1238-1252 (2014).
173. Babiskin, A.H. & Smolke, C.D. A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Molecular Systems Biology* **7**(2011).
174. Hanson, S., Berthelot, K., Fink, B., McCarthy, J. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular microbiology* **49**, 1627-1637 (2003).
175. Suess, B. *et al.* Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic acids research* **31**, 1853-1858 (2003).
176. Hanson, S., Bauer, G., Fink, B. & Suess, B. Molecular analysis of a synthetic tetracycline-binding riboswitch. *RNA* **11**, 503-511 (2005).
177. Werstuck, G. & Green, M. Controlling gene expression in living cells through small molecule-RNA interactions. *Science* **282**, 296-298 (1998).
178. Fegan, A., White, B., Carlson, J.C.T. & Wagner, C.R. Chemically Controlled Protein Assembly: Techniques and Applications. *Chemical Reviews* **110**, 3315-3336 (2010).
179. Weinhandl, K., Winkler, M., Glieder, A. & Camattari, A. Carbon source dependent promoters in yeasts. *Microbial Cell Factories* **13**, 5 (2014).
180. Adams, B.G. Induction of Galactokinase in *Saccharomyces cerevisiae*: Kinetics of Induction and Glucose Effects. *Journal of Bacteriology* **111**, 308-315 (1972).
181. Ideker, T. *et al.* Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science* **292**, 929-934 (2001).
182. Yin, Z., Hatton, L. & Brown, A.J.P. Differential post-transcriptional regulation of yeast mRNAs in response to high and low glucose concentrations. *Molecular Microbiology* **35**, 553-565 (2000).
183. Etcheverry, T. [26] Induced expression using yeast copper metallothionein promoter. in *Methods in Enzymology*, Vol. Volume 185 (ed. David, V.G.) 319-329 (Academic Press, 1990).
184. Murakami, K., Tsubouchi, R., Fukayama, M. & Yoshino, M. Copper-dependent inhibition and oxidative inactivation with affinity cleavage of yeast glutathione reductase. *BioMetals* **27**, 551-558 (2014).
185. Shen, M.W.Y., Fang, F., Sandmeyer, S. & Da Silva, N.A. Development and characterization of a vector set with regulated promoters for systematic metabolic engineering in Saccharomyces cerevisiae. *Yeast* **29**(2012).

186. Price, V.L., Taylor, W.E., Clevenger, W., Worthington, M. & Young, E.T. [25] Expression of heterologous proteins in *Saccharomyces cerevisiae* using the ADH2 promoter. in *Methods in Enzymology*, Vol. Volume 185 308-318 (Academic Press, 1990).
187. Butt, T.R. *et al.* Copper metallothionein of yeast, structure of the gene, and regulation of expression. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 3332-3336 (1984).
188. Karin, M. *et al.* Primary structure and transcription of an amplified genetic locus: the CUP1 locus of yeast. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 337-341 (1984).
189. Peng, B., 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. *Microbial Cell Factories* **14**, 91 (2015).
190. Kapteyn, J.C. *et al.* Low external pH induces HOG1-dependent changes in the organization of the *Saccharomyces cerevisiae* cell wall. *Molecular Microbiology* **39**, 469-480 (2001).
191. Jelinsky, S.A., Estep, P., Church, G.M. & Samson, L.D. Regulatory Networks Revealed by Transcriptional Profiling of Damaged *Saccharomyces cerevisiae* Cells: Rpn4 Links Base Excision Repair with Proteasomes. *Molecular and Cellular Biology* **20**, 8157-8167 (2000).
192. Marmorstein, R., Carey, M., Ptashne, M. & Harrison, S.C. DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* **356**, 408-414 (1992).
193. Kuranda, K., Leberre, V., Sokol, S., Palamarczyk, G. & François, J. Investigating the caffeine effects in the yeast *Saccharomyces cerevisiae* brings new insights into the connection between TOR, PKC and Ras/cAMP signalling pathways. *Molecular Microbiology* **61**, 1147-1166 (2006).
194. Farrugia, G., Bannister, W.H., Vassallo, N. & Balzan, R. Aspirin-induced apoptosis of yeast cells is associated with mitochondrial superoxide radical accumulation and NAD(P)H oxidation. *FEMS Yeast Research* **13**, 755-768 (2013).
195. Roberts, S.K., McAinsh, M., Cantopher, H. & Sandison, S. Calcium Dependence of Eugenol Tolerance and Toxicity in *Saccharomyces cerevisiae*. *PLoS ONE* **9**, e102712 (2014).
196. Darvishi, E., Omidi, M., Bushehri, A.A.S., Golshani, A. & Smith, M.L. The Antifungal Eugenol Perturbs Dual Aromatic and Branched-Chain Amino Acid Permeases in the Cytoplasmic Membrane of Yeast. *PLoS ONE* **8**, e76028 (2013).
197. Nguyen, T.T.M., Iwaki, A. & Izawa, S. The ADH7 Promoter of *Saccharomyces cerevisiae* is Vanillin-Inducible and Enables mRNA Translation Under Severe Vanillin Stress. *Frontiers in Microbiology* **6**, 1390 (2015).
198. Iwaki, A., Kawai, T., Yamamoto, Y. & Izawa, S. Biomass Conversion Inhibitors Furfural and 5-Hydroxymethylfurfural Induce Formation of Messenger RNP Granules and Attenuate Translation Activity in Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **79**, 1661-1667 (2013).

199. Nguyen, T.T.M., Iwaki, A., Ohya, Y. & Izawa, S. Vanillin causes the activation of Yap1 and mitochondrial fragmentation in *Saccharomyces cerevisiae*. *Journal of Bioscience and Bioengineering* **117**, 33-38 (2014).
200. Nguyen, T.T.M., Kitajima, S. & Izawa, S. Importance of glucose-6-phosphate dehydrogenase (G6PDH) for vanillin tolerance in *Saccharomyces cerevisiae*. *Journal of Bioscience and Bioengineering* **118**, 263-269 (2014).
201. Stathopoulos, A.M. & Cyert, M.S. Calcineurin acts through the CRZ1/TCN1-encoded transcription factor to regulate gene expression in yeast. *Genes & Development* **11**, 3432-3444 (1997).
202. Cai, L., Dalal, C.K. & Elowitz, M.B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485-490 (2008).
203. Nevitt, T., Öhrvik, H. & Thiele, D.J. Charting the travels of copper in eukaryotes from yeast to mammals. *Biochimica et biophysica acta* **1823**, 1580-1593 (2012).
204. Dueñas-Sánchez, R., Gutiérrez, G., Rincón, A.M., Codón, A.C. & Benítez, T. Transcriptional regulation of fermentative and respiratory metabolism in *Saccharomyces cerevisiae* industrial bakers' strains. *FEMS Yeast Research* **12**, 625-636 (2012).
205. Chow, T.H.C., Sollitti, P. & Marmur, J. Structure of the multigene family of MAL loci in Saccharomyces. *Molecular and General Genetics MGG* **217**, 60-69 (1989).
206. Hill, Sandra M. *et al.* Asymmetric Inheritance of Aggregated Proteins and Age Reset in Yeast Are Regulated by Vac17-Dependent Vacuolar Functions. *Cell Reports* **16**, 826-838.
207. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The Cost of Protein Production. *Cell Reports* **14**, 22-31.
208. Kamei, Y., Tamada, Y., Nakayama, Y., Fukusaki, E. & Mukai, Y. Changes in transcription and metabolism during the early stage of replicative cellular senescence in budding yeast. *Journal of Biological Chemistry* **289**, 32081-93 (2014).
209. Sadhu, M.J. *et al.* Nutritional Control of Epigenetic Processes in Yeast and Human Cells. *Genetics* **195**, 831-844 (2013).
210. Falcon, A.A. & Aris, J.P. Plasmid accumulation reduces life span in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **278**, 41607-17 (2003).
211. Walther, K. & Schuller, H.J. Adr1 and Cat8 synergistically activate the glucose-regulated alcohol dehydrogenase gene ADH2 of the yeast *Saccharomyces cerevisiae*. *Microbiology* **147**, 2037-44 (2001).
212. O'Sullivan, J.M. *et al.* Gene loops juxtapose promoters and terminators in yeast. *Nature Genetics* **36**, 1014-8 (2004).
213. Hollinshead, W., He, L. & Tang, Y.J. Biofuel production: an odyssey from metabolic engineering to fermentation scale-up. *Frontiers in Microbiology* **5**, 344 (2014).
214. Teng, X. *et al.* Genome-wide Consequences of Deleting Any Single Gene. *Molecular Cell* **52**, 485-494 (2013).
215. Sambrook, J. & Russell, D.W. *Molecular cloning: a laboratory manual*, (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 2001).
216. Gietz, R.D. & Schiestl, R.H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protocols* **2**, 31-34 (2007).

217. Sambrook, J. & Russell, D.W. *Molecular cloning : a laboratory manual*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001).
218. Puigbo, P., Bravo, I.G. & Garcia-Vallve, S. CAIcal: A combined set of tools to assess codon usage adaptation. *Biology Direct* **3**(2008).
219. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Research* **28**, 292-292 (2000).
220. Mumberg, D., Muller, R. & Funk, M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene* **156**, 119-22 (1995).
221. Sheff, M.A. & Thorn, K.S. Optimized cassettes for fluorescent protein tagging in Saccharomyces cerevisiae. *Yeast* **21**, 661-670 (2004).
222. Kalnins, A., Otto, K., Ruther, U. & Muller-Hill, B. Sequence of the lacZ gene of *Escherichia coli. EMBO J* **2**, 593-7 (1983).