GENE SWITCHES AND THEIR APPLICATIONS IN EUKARYOTIC SYSTEMS

BY

JING LIANG

THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Chemical Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2013

Urbana, Illinois

Adviser:

Professor Huimin Zhao
Abstract

Gene switches, also known as transcription factors, are important components of genetic circuits, and they have wide utility in synthetic biology, gene therapy, and developmental biology. Small molecule inducible gene switches are especially valuable as they offer an easy way to regulate target genes using an external signal. By combining DNA binding domain, ligand binding domain, and activation domain, chimeric gene switches can be engineered to turn on or off a user-defined target gene or set of genes, in response to a user-defined chemical signal.

In this thesis, we will explore some applications of gene switches in eukaryotic systems. First, we report the development of a gene switch that is able to regulate the endogenous VEGF-A expression in mammalian cell. The gene switch is specifically and reversibly controlled by 4,4’-dyhydroxybenzil, a small molecule, non-steroid synthetic ligand, which acts orthogonally in a mammalian system. After optimization of the gene switch architecture, a VEGF-A induction ratio of ~200× can be achieved in HEK293 cells at micromolar concentrations of DHB.

Second, we report the development of a system to assemble a multi-gene pathway and subsequently regulate the entire pathway in yeast using an estradiol gene switch. To demonstrate the utility of the system, we assembled the 5-gene zeaxanthin biosynthetic pathway in a single step and showed the ligand dependent, coordinated expression of all 5 genes as well as the tightly-regulated production of zeaxanthin.
Acknowledgments

I would like to thank my parents for their many years of hard work and sacrifice in bringing me up. I would like to thank my wife for being the spark in my life. I wish also to express my gratitude to my advisor, Huimin Zhao, for all the guidance he has provided. I am grateful to all the past and present members of the Zhao laboratory for their helpful discussions, suggestions, and lasting friendship.
Table of contents

Chapter 1: Introduction

1.1 Transcription factors
1.2 Estrogen receptor alpha
1.3 Engineering gene switches
   1.3.1 DNA-binding domains
   1.3.2 Ligand-binding domain
   1.3.3 Activation domain
   1.3.4 Previous studies in Zhao group
1.4 Applications of gene switches
   1.4.1 Endogenous gene regulation
   1.4.2 Exogenous gene regulation
1.5 References
1.6 Figures

Chapter 2: Orthogonal control of endogenous gene expression in mammalian cells using synthetic ligands

2.1 Introduction
2.2 Results
   2.2.1 Domain permutations
   2.2.2 Single-chain gene switch optimization
   2.2.3 Time response
   2.2.4 Ligand response
   2.2.5 Sustained induction and reversal
   2.2.6 Localization
2.3 Discussion
2.4 Conclusions and Outlook
2.5 Material and Methods
   2.5.1 Construction of gene switch
   2.5.2 ELISA assay
   2.5.3 Retroviral integration
   2.5.4 Curve fitting
2.6 References
Chapter 3:  Coordinated induction of multi-gene pathways in *S. cerevisiae* ........42

3.1 Introduction ..............................................................................................................42
3.2 Results ......................................................................................................................44
  3.2.1 Inducible promoter collection ..............................................................................44
  3.2.2 Gene switch engineering and strain development .................................................45
  3.2.3 Golden Gate pathway assembly ...........................................................................46
  3.2.4 Inducible pathway characterization ......................................................................48
3.3 Discussion .................................................................................................................50
3.4 Conclusions and outlook ..........................................................................................55
3.5 Material and methods ..............................................................................................55
  3.5.1 Strains, plasmids, media, reagents, and cell cultivation ........................................55
  3.5.2 Plasmid construction ..........................................................................................56
  3.5.3 Pathway assembly ..............................................................................................58
  3.5.4 Measurement of GFP expression .........................................................................59
  3.5.5 Measurement of zeaxanthin production ...............................................................59
  3.5.6 RNA preparation and quantitative PCR ...............................................................59
3.6 References ................................................................................................................61
3.7 Figures and tables .....................................................................................................64

2.7 Figures and tables ......................................................................................................32
Chapter 1: Introduction

1.1 Transcription factors

Over 300 distinct cell types can be found in the human body. They perform a diverse set of functions and adopt highly specialized structures, but they all share a single genome. This amazing feat is possible only through sophisticated gene regulations (1, 2). Gene regulation occurs at many levels, at the first level, epigenetics factors such as DNA and histone methylation regulate genes by controlling their accessibility (3, 4). At the second level, transcription factors regulate genes by controlling their transcription rates (5, 6). At the third level, combinations of mRNA processing and translation machineries regulate the production of gene products by controlling mRNA stability and translation rate (7, 8). Finally, gene products themselves can be regulated through processes such as allosteric interaction and phosphorylation (9, 6).

Situated early in the chain of regulators, transcription factor is the first control point that offers both high spatial and temporal resolution. As the gatekeepers of genetic information, and they are powerful modulators of everything from life (metabolism (10)), growth (developmental patterning (11)), disease (oncogenesis (12)) and death (apoptosis (13)). The ability to engineer and control transcription factors is therefore an endeavor of fundamental importance and infinite potential.

By definition, transcription factors are protein switches that control the flow of information from DNA to mRNA. Functionally, it can be divided into two classes –
activators and repressors. As their name implies, activators increase a gene’s transcription rate, whereas repressors decrease it. Structurally, most transcription factors are modular with at least two modular domains – the DNA binding domain recognizes and binds to a specific DNA sequence, and the activation domain recruits the transcriptional machineries to initiate transcription (5). Some transcription factors can also contain additional domains such as a ligand-binding domain that interacts with chemical signals (14). Gene switch is another name for transcription factor, and it makes its cellular function more apparent. The two terms are often used interchangeably.

1.2 Estrogen receptor alpha

The estrogen receptor alpha belongs to the nuclear receptor family of proteins. It is a ligand-inducible transcription factor that responds to estrogen and activates the transcription of a set of genes. The human estrogen receptor alpha (ERalpha) can be divided into 6 separate domains, classically indexed A through F (15, 16). The A/B domain is the transcriptional activation domain and is capable of ligand-independent activation. The C domain is the DNA-binding domain (DBD), which consists of two zinc-fingers recognizing the estrogen response element (ERE) (17). Since ERalpha functions as a dimer, the ERE is made up of two palindromic half-sites: AGGTCA NNN TGACCT (18). Domain D is a hinge domain that also participates in dimerization, nuclear localization, and co-activator binding (19). Domain E is the ligand-binding domain (LBD), consisting of 12 alpha helices that form a hydrophobic ligand-binding pocket (20). The natural ligand is 17β-estradiol, but it is also capable of binding to a large number of other ligands (21). Some ligands, such as 17β-estradiol, activate ERalpha, and
are known as an agonist (20), while others can inactivate ERalpha, among them, the most notable antagonistic ligand is tamoxifen, which is used in the treatment of ER positive breast cancers (22). Lastly, Domain F is responsible for modulating the activity of ERalpha, and is involved in agonist and antagonist distinction (23).

Without ligand binding, the ERalpha remains as a monomer and is bound to Hsp90 (24). Ligand binding triggers a conformational change that releases ERalpha from Hsp90, thus allowing it to dimerize (25). The dimer can now recognize the ERE and binds to ERE-containing promoters. The activation domain, together with co-activators such as SRC-1, will then recruit the rest of the transcriptional machineries and start transcription (16).

1.3 Engineering gene switches

Due to the modular nature of transcription factors, the different domains can be exchanged to achieve chimeric function. As a corollary, the different domains can also be engineered independently to achieve a totally novel function.

1.3.1 DNA-binding domains

The targeting specificity of a gene switch controls is determined by its DBD. By engineering DBDs that can bind specifically to the promoter of interest, the corresponding gene can be regulated by the gene switch. Traditionally, the Cys2-His2 zinc finger DBD is the most frequently engineered DBD, and it is also extremely common in naturally occurring mammalian transcription factors (26, 27). Each Cys2-His2 zinc finger can recognize 3bp of DNA sequence. A collection of zinc fingers that recognize any conceivable DNA triplet can theoretically be used to construct zinc finger
concatemer that recognize any DNA sequence (28, 29). In reality however, zinc fingers have strong preference for guanine, so the functional triplets are therefore mostly guanine rich, thus limiting the range of DNA sequences that can be targeted. Furthermore, zinc fingers are not entirely modular and inter-finger interaction can make zinc finger multimer functionally unpredictable (26, 30). By building a large collection of zinc finger monomers and dimers, and combine that with high-throughput screening capability, most of the difficulties, except target preference, can be alleviated. Currently, commercial custom zinc finger maker (Sigma Aldrich, St Louis, MO) can routinely produce zinc finger DBDs with an 18 bp recognition site.

1.3.2 Ligand-binding domain

The ligand-binding domain (LBD) determines the chemical signal to which an inducible gene switch responds. However, most natural occurring LBD, as intended by nature, responds to endogenous ligands. When employed in its native host, the LBD will experience interference from the host’s ligands. For instance, if ERalpha LBD is used in a gene switch for a mammalian system, endogenous estrogen may activate the gene switch independent of external control. Furthermore, when exogenous estrogen is used to control the gene switch, it will inadvertently wreak havoc with the host’s endocrine system (21). It is therefore important to have orthogonal ligand-receptor pairs, so that the gene switch and its corresponding ligand can act orthogonally to the host’s system.

Two approaches can be taken to achieve orthogonality. In one approach, we can use a LBD from a heterologous origin that has no homolog in its intended host. For example, human ERalpha can be used orthogonally in yeast, and bacterial tetracycline repressor
can be used orthogonally in human (31, 32). In the other approach, the LBD can be engineered to respond to new synthetic ligands to form novel orthogonal ligand-receptor pairs.

1.3.3 Activation domain

The activation domain determines how the gene switch interacts with the transcriptional machineries. The structural requirement for an activation domain is rather undiscriminating: (1) they often have an abundance of acidic residues, and (2) an alpha helix appears necessary (5). Given the lack of stringent criteria, activation domains are unsurprisingly promiscuous, i.e. if an activation domain works in one eukaryotic cell, it will likely work in many eukaryotic cells from any species (5). Two of the most commonly used eukaryotic activation domains are VP16 and VP64, both derived from the human herpesvirus (33). The activation domain can also be replaced by other functional domains such as a repression domain, which will create a ligand-inducible repressor instead of an activator (33).

1.3.4 Previous studies in Zhao group

The seminal works of previous lab members laid the foundation for my thesis. First, Karuppiah et al. established a method to switch the ligand specificity of ERalpha. Based on the crystal structure of the ER LBD, 20 sites lining the ligand-binding pocket were chosen for saturation mutagenesis. After initial rounds of iterative saturation mutagenesis, the positive mutant was then subjected to random mutagenesis to further its specificity and sensitivity. Through this method, he showed that novel orthogonal ligand-receptor pairs can be systematically developed (34). Second, McLachlan et al. introduced the synthetic gene switch architecture, where an engineered zinc finger DBD, an engineered
LBD, and a general acidic activation domain were fused to create an artificial transcription factor (35). He also demonstrated the use of orthogonal gene switches in gene circuit construction using ligand-inducible activators and repressors (36).

1.4 Applications of gene switches

In this thesis, I will address the continued development of the gene switch architecture for endogenous gene regulation and exogenous gene regulation.

1.4.1 Endogenous gene regulation

The ability to control endogenous genes has applications in gene therapy and biological research. In cases where a disease state is caused not by the total lack of a gene, but rather, by the uncoordinated or unsynchronized production of a gene product, regaining control over the endogenous gene locus could be an equally if not more effective solution than simply inserting another copy of the gene. Controlling the expression of endogenous genes can also be a valuable tool for interrogating fundamental biological questions. The high temporal control of gene switches allows researchers to turn on or off a gene of interest at a desired time point so that its effect can be studied.

Chapter 2 describes the ligand-inducible activation of endogenous vascular endothelial growth factor A (VEGF-A) in a human cell line. VEGF-A is an important signaling molecule in angiogenesis, and it plays a significant role in growth, recovery as well as disease. Using an engineered LBD and optimized single-chain gene switch architecture, endogenous VEGF-A can be induced, by an orthogonal synthetic chemical ligand, up to 250-folds, which is among the best endogenous gene activation reported in literature. We
further demonstrate that VEGF-A expression can be tightly and reversibly controlled, thus enabling an easily customizable dosage profile that is important for proper angiogenesis. This gene switch architecture can serve as the template for developing future gene switches to target any endogenous gene.

1.4.2 Exogenous gene regulation

When exogenous genes are introduced into a host, indiscriminate overexpression will often lead to counter productivity. The logic is obvious when the gene products or the biosynthetic pathway products are toxic to the host, but it is also true for non-toxic products. If the product interacts with the host’s genetic network, indiscriminate overproduction will lead to imbalance and poor overall fitness. Even when there is no direct interaction, channeling energy and metabolites aimlessly to overproduction is a futile exercise for the host. It is therefore often desirable to put any introduced genes under the control of an inducible production system.

Chapter 3 describes the development of a coordinated induction system for expressing heterologous multi-gene pathways in yeast. Using an ERalpha based gene switch in combination with a set of auxiliary strains and plasmids, multi-gene pathways can be easily assembled, and the resultant pathways are inducible by estrogen. As oppose to galactose, estrogen can be used in the presence of glucose, thus enabling the direct addition of estrogen to a culture without media change. Due to the exceptional sensitivity of ERalpha, only a minute quantity of estrogen is required to fully activate a large culture, making this a highly cost-effective method for industrial scale induction.
1.5 References


1.6  Figures

Figure 1.1  Estrogen receptor alpha domains. Amino acid position, domain index and function indicated.
Chapter 2: Orthogonal control of endogenous gene expression in mammalian cells using synthetic ligands

2.1 Introduction

Since their inception, small-molecule controlled transcription regulators, or gene switches, have been powerful tools in many areas of biological and biomedical sciences. For example, in developmental biology, signaling pathways can be dissected through temporal suppression and expression of genes within the pathways. In gene therapy, the induction of a beneficial gene product can be regulated to fall within a therapeutic window or ceased when necessary (1-4). With the advent of synthetic biology, complex gene circuitries can be built from a set of well-behaved gene switches, thus creating further demand for orthogonal gene switches (5, 6). In addition, there is a growing interest in controlling complex phenotypes, especially in mammalian cells. Controlling complex phenotypes requires the remodeling of an intricately interconnected genetic network. Due to the network’s size and complexity, many cellular phenotypes are inaccessible via the control of a single gene locus. For example, the induction of pluripotency in human somatic cells requires the simultaneous induction of at least three genes (7-9). To maximize the control over the genetic network, it is highly desirable not only to control multiple genes at the same time, but also to control them using independent orthogonal ligands.

The aim of this work is to develop a framework for the engineering of orthogonal mammalian gene switches, and demonstrate its utility through the ligand-dependent induction of an endogenous gene. Compared to typical multi-copy plasmids, genes on the
chromosome has a much lower copy number, and the induction of endogenous loci is often more challenging, therefore placing higher requirement on the efficiency of the design. Induction of endogenous genes is of particular interest in mammalian systems because its gene products have shown better therapeutic efficacy compared to those of an exogenous gene, possibly by providing the necessary splice variants (10). Such functionality will also be useful to developmental biology and synthetic biology because it sometimes takes multiple splice variants working in tandem to achieve a desired biological effect (10).

Tetracycline receptor, ecdysone receptor, chemical dimerizers, and nuclear hormone receptor are among the leading platforms for mammalian gene regulation (4). In this study, we focus on the nuclear hormone receptor platform because (a) the protein can be fully humanized, which minimizes the chance of immune response, (b) the ligand specificity can be engineered and changed to create orthogonal ligand-receptor pairs, and (c) its modular design allows convenient change of its DNA binding domain to target different genes. Under the nuclear hormone receptor platform, the ligand-binding domain from a nuclear hormone receptor, e.g. estrogen receptor and progesterone receptor, is fused with a zinc-finger DNA binding domain that binds to the promoter region of the target gene. Although estrogen receptor itself has slight activation ability, an activation domain, e.g. VP16, VP64, or p65, is typically added to enhance its induction power (11).
The ligand-induced regulation of endogenous genes has been previously demonstrated using such a gene switch platform (12, 13). However, two main shortcomings currently limit the utility of this platform. First, the ligands used for induction are not orthogonal to natural occurring nuclear hormone receptors, and have potent biological effect in mammalian systems. Second, the basal expression from the platform is considered somewhat leaky compared to the tightly controlled tetracycline receptor platform. This can be due to the natural occurrence of hormones in the mammalian system or the presence of an ectopic overexpression of an activation domain. In a recent study, Schwimmer et al. have partially addressed the first problem through the development of a benzoate X receptor gene switch (14). However, this has yet to be demonstrated in an endogenous context.

In this study, we aim to address these main limitations. To create orthogonality, we utilized a previously engineered estrogen receptor ligand binding domain (LBD) that is sensitive to 4,4-dihydroxybenzil (DHB) – a synthetic non-steroid ligand that acts orthogonally to estrogen (Figure 2.1) (15, 16). The LBD consists of amino acids 312-595 of the human estrogen receptor alpha, and the engineered version, called 4S LBD, has mutation A350M, L346I, M388Q, G521S, Y526D. DHB is synthetic ligand that is structurally similar to diethylstilbestrol, a FDA approved drug, and it has been shown to have low cytotoxicity in cell culture and low systemic toxicity in mouse model (15, 17).

The vascular endothelial growth factor A (VEGF-A) was chosen as a model target for this study. VEGF-A is an important signaling molecule in angiogenesis, and its
regulation has many therapeutic applications (18). For example, the up-regulation of VEGF-A can help new blood vessels formation as a treatment for ischemia, whereas the down-regulation of VEGF-A can restrict the growth of tumors (19, 20). To target the endogenous VEGF-A loci, we have utilized VZ-8, a zinc-finger DNA binding domain (DBD) engineered by Liu et al., which recognizes a 9 base pairs site within the VEGF-A promoter, and has been shown to activate VEGF-A expression when combined with an activation domain (21).

In this work, we have combined the VZ-8 DBD, the 4S LBD, and the p65 activation domain to make our gene switch. The human p65 protein is part of the NF-κB activation complex (22). It is 551 amino acids (aa) long and most of the activation activity is localized at the C-terminal region (23). To minimize basal expression level and maximize induction level, we performed domain length optimization of the p65 activation domain, as well as domain permutation and linker optimization of the overall architecture.

When transiently transformed into HEK293 cells, our construct in the absence of DHB showed no VEGF-A expression beyond the normal basal level. When induced with 1 μM DHB, a 170-fold increase in VEGF-A expression was observed. This induction was roughly twice as high as that achieved with a constitutive construct. When stably integrated into HEK293, >250-fold induction was observed with no detectable basal expression. This gene switch has among the highest performance characteristics as compared to previously reported endogenous gene induction studies. Furthermore, it provides the architecture to combine the zinc-finger DBD technology with the ligand
specificity engineering technology, and brings us one step closer to the ability to control multiple genes with multiple orthogonal ligands.

2.2 Results

2.2.1 Domain permutations

We attempted to use our gene switch to induce the endogenous VEGF-A expression in HEK293 cell line. HEK293 was chosen in the induction study because of its low basal VEGF-A level, and its widely reported use in endogenous VEGF-A induction studies (21, 13, 24). Our initial construct, p65-VZ8-4S, did not show any induction activity. We proceeded to try a set of different domain permutations as listed in Figure 2.2. The last two dual LBD single-chain constructs, V24P and P24V, were by far the best performing constructs, outperforming the rest by at least 5 fold (Figure 2.3).

2.2.2 Single-chain gene switch optimization

The estrogen receptor LBD dimerizes during the process of ligand activation, and it has been shown that a single-chain estrogen receptor – comprised of two LBDs fused together – can activate gene expression by intramolecular dimerization (25, 12). Since it was unclear how two chimeric estrogen receptor gene switches would dimerize, we postulated that the linker length between the LBDs could affect dimerization and thus its activation characteristics. A set of single-chain gene switches from the V24P and P24V constructs was constructed by inserting between the two 4S LBDs varying lengths of GS linkers (G₄S)ₙ ranging from 20 to 110 aa long (Figure 2.2). This set of gene switches was transiently expressed in HEK293 cells under the CMV promoter, and the VEGF-A concentration was assayed by ELISA 24 hours after DHB induction.
As shown in Figure 2.4A, it was found that in the V24P construct, induction level peaked with a 60 aa linker, which offered a 20% improvement over the construct with no linker. Furthermore, the basal expression level decreased with increasing linker length, bottoming out also at 60 aa. The induction ratio for V24P-GS60 was about 170-fold. A peak induction level of 2456 pg/mL was achieved, which was significantly higher than even the constitutive constructs VZ8-P65 and F435P. The basal level was around 13 pg/mL, lower than the EGFP negative control, but higher than the Gal-P65 negative control. VZ8-P65 consists of a VZ8 DBD fused directly to a p65 activation domain, and served as constitutive activator of VEGF-A, whereas F435P is another zinc-finger based constitutive activator of VEGF-A obtained from Bae et al. (26), and served as a benchmark for the observed induction level. Gal-P65 consists of a Gal4 DBD fused directly to a p65 activation domain, and served as a control for any non-specific induction caused by ectopic expression of p65.

The P24V construct offered a different profile of activation. While the induction level appeared independent of linker length, the basal expression level increased with increasing linker length. The strength of induction of the P24V construct was less than half that of V24P, but the basal level for P24V was only 3 pg/mL (outside of standard curve), which was comparable to the Gal-P65 negative control. The very low basal expression gave the P24V construct an induction ratio of 300-fold despite its lower peak induction level.
2.2.3 Time response

The speed of activation is an important gene switch parameter. Being a constitutively expressed single-step transcriptional switch, it is expected to be slower than translational switches, but comparable to that of a Tet-On system controlling a transgene (27). To characterize the response time of the gene switch, we added DHB to identical wells at one-hour intervals, and assayed their VEGF-A concentration at the end of 12 hours (Figure 2.4B). VEGF-A production was clearly detectable from 3 hours onwards, and increased steadily thereafter.

2.2.4 Ligand response

V24P-GS60 was chosen for further characterization because of its high induction power, and we hypothesized that the basal level could be reduced with lower gene switch expression level. For further characterization, the P24V-GS60 gene switch was integrated into the chromosome of HEK293 via retroviral integration and antibiotic selection. At first, single-clone isolation was not carried out because we wanted to see how the integrated switch behaved in aggregate, given that individual clones vary in their performances.

Compared to the plasmid version, the integrated version was slightly less sensitive to the ligand, possibly due to a lower gene switch expression level from the reduced copy number per cell. An induction level of 1895 pg/mL was achieved at 1 μM DHB and 2519 pg/mL was achieved at 10 μM DHB, giving an induction ratio of ~250-fold (Figure 2.5A). The induction appeared to taper off at 10 μM, but cell numbers were significantly less at high inducer concentration. Therefore, the cell number was counted at the time of
sample collection and normalized by the measured VEGF-A level. It was found that, after normalization to cell number, there was no attenuation of induction up to 10 µM DHB. However, increasing the ligand concentration beyond 10 µM was impractical due to severe impairment on cell growth. The growth impairment most likely came from the heavy metabolic burden under high induction, because no effect on cell growth could be observed in cells not expressing the gene switch, even after prolonged DHB and/or VEGF-A exposure. The induction was log-linear to ligand concentration over at least 2 orders of magnitude. This characteristic allowed the switch to be tuned by external ligand concentration and achieve induction of different strengths.

Despite its usefulness in fine-tuning gene expression, a gently sloped ligand titration curve was atypical for a 4S LBD gene switch (15). Suspecting that this observation might be an aggregated behavior, 12 clonal cell lines were isolated from the heterogeneous population, and tested for their VEGF induction level at 1 µM DHB. The induction level ranged widely from 810 to 5100 pg/mL, and the average of the 12 clonal cell line was 2740 pg/mL, which was close to that obtained from the whole heterogeneous population (Figure 2.6). When we picked the most inducible clone and performed a ligand titration, we obtained a nearly perfect sigmoidal response with a Hill’s coefficient of 2.3, and a \( K_m \) of 38 nM (Figure 2.5B). Further tests on other clones reflected that the \( K_m \) varies between clones (Figure 2.5C).

We propose that the ligand titration curve of the heterogeneous population is a weighted sum of a series of perfectly sigmoidal curves with different \( K_m \). As seen in Figure 2.5D,
using just 8 evenly spaced sigmoidal curves as the basis set, we can fit the observed heterogeneous ligand titration curve very well by assigning different weightage to them.

2.2.5 Sustained induction and reversal

In some applications, long-term sustained induction might be required, whereas in other applications, short intervals of repeated induction might be preferable. To characterize the long-term induction, induction reversal, and repeated induction behavior of the gene switch, the stably integrated HEK293 cells was subjected to three 16-day time courses (Figure 2.7). The first set of experiments was performed using the heterogeneous population of integrants. In time course 1, the cells were induced for 2 days followed by 6 days of rest; in time course 2, the cells were induced for 4 days followed by 4 days of rest; and in time course 3, the cells were maintained at 1 µM DHB through the entire duration. All values were normalized to the cell number at the time of sampling.

In time course 1 (Figure 2.7A), the VEGF-A production level increased quickly to about 2700 (pg/mL/24h) one day after ligand addition, and increased further to 3700 (pg/mL/24h) on the second day. The VEGF production decreased quickly upon ligand withdrawal, reaching basal level in two days. When the cells were re-induced on day 9, a similar behavior was observed. This showed that the induction was reversible and repeatable. In time course 2 (Figure 2.7B), the daily VEGF-A production level increased initially for 3 days, reaching up to 4800 (pg/mL/24h). However, VEGF-A production dropped on the 4th day despite continued induction. The induction sensitivity was only partially recovered after 4 days of rest, and a drop in induction level was still observable on the 4th day of re-induction. During sustained induction in time course 3 (Figure 2.7C),
a similar trend was observed for the initial 3 days of induction. The induction level then decreased steadily from the 4\textsuperscript{th} day onwards. After 15 consecutive days of induction, the VEGF-A production level dropped to near basal level.

It was found that sustained induction could only be maintained using a clonal integrant cell line, and at such a level that it does not affect cell growth. The sustained induction time course was repeated using a highly sensitive clone at 20 nM DHB, which corresponds to an induction level of about 1000 (pg/mL/24h). At this level, the induction can be sustained for the entire duration of the time course (Figure 2.7D). To demonstrate that the cells were still capable of high induction, the maximum induction of 5500 (pg/mL/24h) was elicited using 200 nM DHB on day 1 and again on day 9.

\textbf{2.2.6 Localization}

Because of the low VEGF-A background in our test cell line, we were unable to determine if the presence of the un-induced gene switch had any effect on the VEGF-A expression. It was plausible that the DNA binding domain could still bind to the promoter and affect native transcription. To test this possibility, a localization study of the gene switch was performed, i.e. if the gene switch were mostly cytosolic when un-induced, then the likelihood of it affecting native transcription would be low.

The EGFP gene was fused to the N-terminus of our gene switch and confocal microscopy was used to visualize where the gene switches were localized in the cell (Figure 2.8). To ensure the EGFP-fused gene switch would behave similarly, its induction activity was
also assayed. The activity data showed that the EGFP fusion protein was active, albeit at a lower level most likely due to a reduced expression level.

The confocal images indicated that the gene switch was mostly localized within the nucleus with and without ligand. The natural estrogen receptor is known to reside mostly in the nucleus and the presence of an engineered ER LBD alone appears sufficient to confer this localization property. For natural full length ER, the un-induced receptor is bound to heat shock protein (hsp90), which prevents dimerization and blocks the transcription regulation activity of the receptor. In contrast, it has been reported that the ER LBD alone is insufficient for binding to hsp90 (28). Whether this is true for our engineered gene switch remains to be tested.

2.3 Discussion

In this chapter, we have demonstrated that our gene switch is able to induce endogenous VEGF expression in HEK293 cells using a non-steroid orthogonal ligand. An induction ratio of over 100-fold can be achieved at µM ligand concentration. We have also shown that the induction is fast, reversible, and sustainable as long as the cellular machinery can support it.

To highlight the performance of our gene switch, it is important to make some cross comparisons with the performance of previously reported gene switches. This is, however, difficult due to differing assay conditions. We have summarized previously reported VEGF induction values, be it inducible or constitutive, and made reasonable
comparisons whenever possible (Table 2.1). Overall, our gene switch assay has among the lowest initial cell number and among the shortest accumulation time, and yet, has the highest peak VEGF induction value. Pollock et al. normalized their VEGF assay by total cellular protein content, and is therefore not directly comparable (29). Fortunately, both Pollock et al. and Liu et al. have included hypoxia positive control which suggests that Pollock’s assay conditions yield values roughly 8.5 times that of Liu’s (29, 21). Liu’s and Dent’s studies share a common construct, VZ+434-P65, and their relative value suggests that Dent’s assay yields values roughly 3.5 times that of Liu’s (21, 13). Our study shares a common construct with Bae et al., F435P, which suggests that Bae’s assay yields value roughly twice ours (26).

Taken together, our gene switch produces the highest peak induction value as well as the highest induction ratio, even when compared with constitutive constructs. It should be noted that induction ratio is highly dependent on basal level expression, and due to the low basal level expression under non-induction condition, our low-end measurement values are often too close to the detection limit of the ELISA assay to be dependable. This is especially true in the case of an integrated gene switch in a heterogeneous population, where the measured VEGF-A value is effectively zero. In that case, we have used the ELISA detection limit (around 10 pg/mL) to calculate the induction ratio. We have excluded from comparison the clonal assays value, which has a peak of 4588 pg/mL, because it varies between cell lines and we do not believe that it is a good representation of the gene switch’s performance.
In the presence of the uninduced gene switch, the basal VEGF-A level appears lower than the negative control. This is the most pronounced in the integrated HEK293 cell line. Since our localization study found that the gene switch resides mostly in the nucleus with and without induction, it is possible for the DBD to bind and disrupt regular transcription. However, due to the high uncertainty associated with low value measurements, we are unable to determine if the un-induced gene switch represses VEGF-A transcription.

We have shown that the induction is reversible and repeatable. However, we are unable to sustain the induction over a long period of time in a heterogeneous population. This is because cell growth is inhibited and cell death is triggered at a high induction level, most likely due to overwhelming metabolic burden. Since the heterogeneous population contains cell lines that have a wide range of ligand sensitivity, we are effectively enriching the non-secreting and low-secreting cells by killing off the highly inducible ones. In a clonal cell line, induction is sustainable at a level that does not affect cell growth. Unfortunately, this is achieved at a ligand concentration that is around the $K_m$, where the induction level is most sensitive to minute differences in ligand concentration, thus leading to large errors of measurement.

We also see important differences in the behavior of the heterogeneous and the clonal integrant populations. For example, the ligand response curve is much gentler in a heterogeneous population which allows for finer control of induction level. Theoretically, by mixing a finite set of clonal population of different ligand sensitivities,
we can tune the ligand response curve of the mixed population to achieve a sensitivity we desire.

As a ligand, DHB has shown very little toxicity in multiple cell lines, with IC50 greater than 100µM. (30) Similarly, DHB has also been found to have low systemic toxicity in mice, and intraperitoneal (IP) injection of 1 mg/kg is well tolerated. While DHB is stable in cell culture media for days, it is either rapidly cleared from or modified in mice with no detectable DHB remaining in blood 24h after IP injection. Despite the apparent short half-life, DHB has been successfully used to trigger recombination in mice through a Cre-ER fusion protein. The recombination efficiency observed using the DHB-4S ligand-receptor pair was comparable to that of the tamoxifen-ER$^T$ pair (31, 17).

2.4 Conclusions and Outlook

To conclude, we have reported the design and construction of a highly effective ligand-responsive artificial transcription factor whose performance tops previously reported designs. This study represents the missing link between gene targeting technology and ligand specificity engineering, thus making it possible to create multiple orthogonal gene switches under the control of multiple orthogonal ligands. This technology can be immediately useful to developmental biologists, and with further development, be useful in gene therapy and synthetic biology applications.

The ultimate aim of the technology developed in this study is to create a platform for the engineering of orthogonal gene switches that can independently control multiple genes in mammalian systems. The LBD used in our study is responsive to DHB and more
orthogonal LBD-ligand pairs can be created through the established method published earlier. In fact, the method has already been successfully applied in multiple studies, giving rise to multiple orthogonal LBD-ligand pairs (15, 16, 32). In order to control multiple genes, multiple DBDs also need to be engineered. Fortunately, zinc-finger DBD technology is well established and is now even commercially available (33, 26, 34, 35). Looking further, TAL effector DBD is a new class of DBD that promises even better modularity than zinc-finger DBD (36, 37). However, its incorporation into the gene switch construct will likely involve redesign and optimization.

2.5 Material and Methods

2.5.1 Construction of gene switch

The P65 gene segment was PCR amplified and cloned from human cDNA and the different truncations were subsequently generated by PCR. The engineering of the 4S ligand-binding domain was described elsewhere (15). The VZ8 DNA binding domain was constructed by overlap extension of DNA oligos as described elsewhere (21). The long GS linker was constructed by overlap extension PCR and the truncations subsequently generated by PCR. The different parts were assembled using unique restriction sites introduced in the PCR primers, and cloned into pCMV5 for transient expression and into pLNCX2 for stable integration. A list of plasmids used in this work can be found in Table 2.2.
2.5.2 ELISA assay

HEK293 cell line was obtained from Professor Jie Chen of University of Illinois and propagated in DMEM (UIUC Cell Media Facility) with 10% FBS at 37°C 5% CO₂. 48 hours before sampling, 8x10⁵ cells were seeded into each well of a 24-well plate (PureCoat Amine, BD Biosciences). 36 hours before sampling, cells for transient expression were transfected by FuGene HD (Promega, Madison WI) according to the manufacturer’s recommendation using 500 ng of the appropriate gene switch plasmids. 30 hours before sampling, ligands were added to appropriate wells. 24 hours before sampling, media was changed, and ligands re-added. After 24 hours of accumulation, the supernatant was collected and assayed in duplicate by ELISA. ELISA kit was obtained from R&D Systems (Minneapolis, MN), and performed according to the manufacturer’s recommendation.

2.5.3 Retroviral integration

Retroviral integration was performed using retrovirus packaging system from Clontech (Mountain View, CA) according to manufacturer’s recommendation. Briefly, selected gene switches were cloned into pLNCX2 retroviral vector. The retroviral vectors were co-transformed together with pVSV-G envelope vector into the GP2-293 packaging cell line to produce retrovirus. Supernatant containing retrovirus was collected after 2 days and used to infect HEK293 cells. One day after infection, HEK293 cells were put under 1000 µg/mL G418 selection. The resistant cells were gathered after 2 weeks of selection and subsequently used for characterization studies.
2.5.4 Curve fitting

Curve fitting was performed using OriginPro 8.6. The clonal ligand titration curve was fitted to the Hill equation, \( y = \text{min} + (\text{max} - \text{min}) \cdot \frac{x^n}{k^n + x^n} \) where \( n \) is the Hill coefficient and \( k \) is the \( K_m \). The heterogeneous population ligand titration curve was fitted using a weighted sum of 8 Hill equations. \( y = \frac{1}{\sum a_i} \sum_i[a_i \cdot f(\text{min}, \text{max}, n, k_i)] \), where \( f \) is a Hill equation with \( \text{min} = 0, \text{max} = 5000, n = 2.3, k_i = [3.3\times10^{-8}, 1\times10^{-7}, 3.3\times10^{-7}, 1\times10^{-6}, 3.3\times10^{-6}, 1\times10^{-5}, 3.3\times10^{-5}, 1\times10^{-4}] \), and \( a \) is the weight parameter for fitting.
2.6 References


2.7 Figures and tables

Figure 2.1  The structure of 17β-estradiol (E2), the natural ligand of estrogen receptor (ER), and the structure of 4,4-dihydroxybenzil (DHB).
Figure 2.2  Diagrammatic representations of the constructs used in this study. All constructs were cloned into pCMV5 for transient expression and pLNCX2 for stable integration. P65 is the p65 activation domain of NF-κB. VZ8 DBD is the VZ-8 zinc-finger DNA binding domain that binds within VEGF-A promoter. 4S LBD is an engineered ER ligand binding domain that is activated by DHB. (A) Construct used in p65 domain length optimization. Hatched box represents variable p65 length. (B) Variations of gene switch used in the initial screening to identify the best architecture. (C) Linker length optimization construct based on V24P.
Figure 2.3  Result from the initial architecture optimization screening. V24P and P24V clearly outperformed the rest. EGFP, EGFP+DHB, and GalP65 were used as negative controls. EGFP+CoCl$_2$ was used to emulate hypoxia and VZ8-P65 and F435P were used as positive controls. Light bar – in the absence of DHB. Dark bar – in 1 μM DHB.
**Figure 2.4**  Endogenous VEGF-A induction in transiently transfected HEK293. (A) The horizontal axis labels represent the constructs’ names, and the V24P series contain GS linkers ranging from 0 to 110 amino acids long. White bar is uninduced, and grey bar is induced with 1 µM DHB. VZ8-P65 is a constitutive positive control and F435P is a positive control obtained from Bae et al. (26) to serve as a performance yardstick. EGFP, Gal-P65, and Gal-P65 + DHB are negative controls. Error bar represents standard error, and measurements have been obtained from two separate transfections, each assayed in duplicate. The y-axis is truncated to allow clearer representation of the uninduced level. (B) Time response profile obtained by adding ligands to wells at 1-hour intervals for 12 hours.
Figure 2.5  Ligand titration curve obtained using HEK293 cells stably integrated with V24P-GS60 gene switch. Error bar represents standard error, and measurements have been obtained from two independent wells, each assayed in duplicate. (A) Ligand titration curve of a heterogeneous integrant population. Black square series represents the measured, un-normalized, VEGF-A values. The cell number at the time of sampling was used to normalize the grey circle series. (B) Ligand titration curve of a highly inducible clonal integrant population, which was fitted to a Hill curve. Min = 108, Max = 4588, Hill Coefficient = 2.3, $K_m = 38$ nM, $R^2 = 0.9996$. (C) A representative series of ligand titration curves with shifted $K_m$. (D) A fit of the observed heterogeneous ligand titration curve using the series in (C) as a basis set, $R^2 = 0.9978$. 

---

**A**

![Graph A](image)

**B**

![Graph B](image)

**C**

![Graph C](image)

**D**

![Graph D](image)
**Figure 2.6**  Range of induction levels from 12 different stable integrant clones. Induction level ranged from 810 to 5100 pg/mL, averaged 2740 pg/mL. Light bar – in the absence of DHB. Dark bar – in 1 µM DHB.
Figure 2.7  Time courses obtained using HEK293 cells stably integrated with V24P-GS60. Error bars represent standard errors, and measurements have been obtained from two wells, each assayed in duplicate. Values have been normalized by the cell number at the time of sampling. Media (and ligand) was changed 24 hours prior to each sampling.

(A) Heterogeneous time course 1, induced for 2 days at 1 µM DHB, rested for 6 days, repeat. (B) Heterogeneous time course 2, induced for 4 days at 1 µM DHB, rested for 4 days, repeat. (C) Heterogeneous time course 3, sustained induction at 1 µM DHB. (D) Clonal population time course, sustained induction at 20 nM DHB with 200 nM spikes on day 1 and day 9.
Figure 2.8  Confocal microscopy. Fluorescent gene switch constructs are created by inserting EGFP at the N-termini of the gene switch constructs used for transient expression. The EGFP constructs are transfected into HEK293 cells grown on Ibidi poly-d-lysine µ-slide (Ibidi, Verona, WI) using FuGene HD (Promega, Madison, WI). Images are taken 24 hours later using Carl Zeiss LSM 700 confocal microscope (Thornwood, NY). Nucleus stain channel shows the nuclei counter stained by Hoechst dye (Invitrogen, Carlsbad, CA). EGFP channel shows the location of the EGFP-tagged gene switches.

![Confocal microscopy image](image-url)
Table 2.1 Comparison of assay conditions and induction values in previously reported VEGF-A induction studies. When a particular condition or construct is shared between studies, the name and value is noted under Reference Values. *Not specified, assume typical working volume for the plate format.

<table>
<thead>
<tr>
<th>Author</th>
<th>Liu et al.</th>
<th>Pollock et al.</th>
<th>Bae et al.</th>
<th>Dent et al.</th>
<th>This Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Line</td>
<td>HEK293</td>
<td>HEK293</td>
<td>HEK293</td>
<td>HEK293</td>
<td>HEK293</td>
</tr>
<tr>
<td>Seed Density</td>
<td>1.60 x 10^5</td>
<td>3.00 x 10^5</td>
<td>1.00 x 10^5</td>
<td>Unknown</td>
<td>8.00 x 10^4</td>
</tr>
<tr>
<td>Format</td>
<td>24-well</td>
<td>12-well</td>
<td>12-well</td>
<td>Unknown</td>
<td>24-well</td>
</tr>
<tr>
<td>Volume</td>
<td>500μL*</td>
<td>1mL*</td>
<td>1mL</td>
<td>Unknown</td>
<td>500μL</td>
</tr>
<tr>
<td>Accumulation Time</td>
<td>40h</td>
<td>48h</td>
<td>48h</td>
<td>24h</td>
<td>24h</td>
</tr>
<tr>
<td>Peak VEGF value</td>
<td>700</td>
<td>5000</td>
<td>2200</td>
<td>1800</td>
<td>2500</td>
</tr>
<tr>
<td>VEGF unit</td>
<td>pg/mL</td>
<td>pg/mg</td>
<td>pg/mL</td>
<td>pg/mL/24h</td>
<td>pg/mL/24h</td>
</tr>
<tr>
<td>Induction Type</td>
<td>Constitutive</td>
<td>Inducible</td>
<td>Constitutive</td>
<td>Inducible</td>
<td>Inducible</td>
</tr>
<tr>
<td>ELISA Kit</td>
<td>R&amp;D Systems</td>
<td>R&amp;D Systems</td>
<td>Chemicon</td>
<td>R&amp;D Systems</td>
<td>R&amp;D Systems</td>
</tr>
<tr>
<td>Reference values</td>
<td>Hypoxia</td>
<td>Hypoxia</td>
<td>F435-P</td>
<td>VZ+434-P65</td>
<td>F435-P</td>
</tr>
<tr>
<td></td>
<td>400</td>
<td>3500</td>
<td>2200</td>
<td>2500</td>
<td>1100</td>
</tr>
<tr>
<td></td>
<td>VZ+434-P65</td>
<td>700</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plasmid</td>
<td>Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>GAL-VP64 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GAL-p65 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>p65(510-551)-VZ8-4S pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>p65(451-551)-VZ8-4S pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>p65(282-551)-VZ8-4S pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>p65(417-459)-VZ8-4S pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>p65(373-459)-VZ8-4S pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>p65(460-523)-VZ8-4S pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>V24P(451) pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>V24P-GS20 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>V24P-GS40 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>V24P-GS60 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>V24P-GS80 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>V24P-GS110 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>P24V(451) pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>P24V-GS20 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>P24V-GS40 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>P24V-GS60 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>V4 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Vg4 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>4V pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>24V pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>V24 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>4VP pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>P(451)V4 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>P(284)V4 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>VZ8-P65 pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>EGFP pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>V24P-GS60 pLNCX2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>V24P-EGFP pCMV5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3: Coordinated induction of multi-gene pathways in *S. cerevisiae*

3.1 Introduction

Coordinated gene expression is an important tool in the biosynthesis of natural products, fine chemicals, and fuels (1-3). For example, the taxol biosynthetic pathway involves the synchronized action of at least 19 genes (4). In prokaryotes, genes involved in a biosynthetic pathway are often grouped together into an operon and are transcribed as a polycistronic mRNA under the regulation of a single promoter (5, 6). The operon structure makes it easy to regulate a large number of genes (7). In eukaryotes, especially filamentous fungi and even some plants, genes in the same pathway are also often grouped together into a gene cluster; but in contrast to that in prokaryotes, the genes are not under the same promoter, and may be subjected to independent regulation (8, 9). With rare exceptions, operons generally do not exist in eukaryotes (10).

*S. cerevisiae* is an important industrial production host for heterologous pathways, and is generally suited for expressing pathways from fungi and plants, especially those involving cytochrome P450s (1). Being a eukaryote, it does not recognize operons, and each gene in a pathway will need to have its own promoters. This requirement can make construction of multi-gene pathways cumbersome. Traditionally, genes in pathways are usually broken up into multiple plasmids – at times only one gene on a plasmid (11), or two genes under a divergent promoter (12). In either format, the construction of a long pathway will require many plasmids and correspondingly many selection markers. Thanks to the recently developed large scale cloning methods, such as SLIC (13), Golden
Gate (14), Gibson (15), and DNA Assembler (16), large concatenations of promoters and genes can now be routinely made. However, large pathways constructed thus far are driven by constitutive promoters, because with the exception of Golden Gate, most of the assembly methods are homology dependent (17, 18), and non-homologous inducible promoters that respond to the same inducer are limited. Furthermore, it is commonly believed that due to the active homologous recombination machinery in yeast, homologous promoters should be avoided even if they do not interfere with DNA assembly (1).

GAL inducible promoters are commonly used for controlling gene expression in yeast (19). However, due to the high cost of the inducer, galactose, its use in industrial production is limited (20). To circumvent this problem, we have employed an estrogen receptor based gene switch to activate GAL promoters. Because of its nanomolar sensitivity, only 200 µmol 17β-estradiol is needed to fully induce a 20,000L reactor, at a cost of roughly $1.20 (Sigma Aldrich catalog 2012).

In this chapter, we present a system of constructing and activating multi-gene pathways in *S. cerevisiae*. The system consists of three parts: (a) a collection of inducible promoters, (b) an estradiol inducible yeast strain, and (c) a method of DNA assembly that allows for easy construction of inducible pathways in one step from PCR products. As a proof of concept, we assembled a five-gene zeaxanthin pathway, demonstrated the estradiol-dependent production of zeaxanthin, and characterized the pathway’s behaviour.
3.2 Results

3.2.1 Inducible promoter collection

The availability of natural GAL responsive promoters is limited to the genes involved in galactose utilization. To address this limitation, we cloned the GAL1, GAL2, and GAL10 promoters into a GFP reporter and induced the GFP expression using Gal-P65, a constitutive gene switch made from the fusion between the GAL4 DNA binding domain and the P65 activation domain. As shown in Figure 3.1A, the natural GAL promoters exhibited tight regulation – the uninduced fluorescence level was indistinguishable from the background fluorescence.

To expand the dynamic range of the promoter collection, two sets of synthetic GAL responsive promoters were created. DNA fragments containing a varying number of consensus Upstream Activation Sequences (UAS) were joined to the TATA region of GAL1 and GAL2, creating the SYN1 and SYN2 collections of promoters respectively. As shown in Figure 3.1B and 3.1C, the induction strength of the promoter can be reduced by decreasing the number of UAS. In contrast to the tightly regulated GAL1 promoter, the SYN1 collection of promoters had noticeably higher basal levels. The SYN2 collection of promoters did not suffer from increased basal level, and was generally weaker than the natural promoters. With the natural promoters and the SYN2 collection of promoters, induction strength across a wide dynamic range can be selected while the pathway of interest is maintained in tight control. Further work on the engineering of GAL inducible promoters have recently been described by Blazeck et al. (25), which provided an even wider range of GAL inducible promoters to choose from.
3.2.2 Gene switch engineering and strain development

*Domain Selection and Permutation.* To identify the optimum configuration for the gene switch, we tested different orders of functional domains and different activation domains. The gene switches were co-transformed with $P_{GAL10}$-GFP and tested for their inducibility using 100 nM estradiol. As shown in Figure 3.2A, P65-Gal-ER (6GE) had the best compromise between induction level and basal expression level, and was selected for the development of a series of estradiol-inducible yeast strains.

*Estradiol-Inducible YZE Strains.* After identifying the best gene switch configuration, we moved on to integrate the gene switch into YM954 so as to create a more convenient induction system as well as to reduce the number of plasmids required. The gene switch was integrated into the $Lys2$ locus without the use of antibiotic selection, thus leaving the most common auxotrophic markers intact and allowing the use of antibiotics for further genome engineering. Since the gene switch protein is not catalytically involved in the production of any desired compound, its over-expression will likely be detrimental to production. By placing the gene switch behind promoters of different strengths, the gene switch expression level will be varied in the different strains. Gene switch expression in YZE-19, YZE-55, YZE-100, and YZE-149 are driven by the ENO promoter and its derivatives – the numbering indicates their relative percentage strength to the wild type ENO promoter. The expression in YZE-PA is driven by the ADH1 promoter. The functionality of these strains was confirmed by observing fluorescence from the induction of $P_{GAL10}$-GFP upon estradiol addition. The flow cytometry GFP channel histograms from the strains were shown in Figure 3.2B. The mean fluorescence value shifted towards
the right at higher promoter strengths. YZE-PA was not shown, but its histogram was indistinguishable from that of YZE-55.

3.2.3 Golden Gate pathway assembly

*Pathway Assembly.* Due to the limited availability of optimal inducible promoters, we designed the pathway assembly system based on the Golden Gate cloning method so that a single good inducible promoter can be re-used to control multiple genes if necessary. Details of Golden Gate cloning have been described by Carola *et al.* (14). Briefly, each fragment in the assembly is flanked by Type IIIs restriction sites, e.g. *Bsa*I, which cleaves the DNA away from its recognition sequence, leaving a user-defined 4 bp overhang. By judiciously choosing unique 4 bp overhangs, these fragments can be ligated together in a specific order to form the pathway. Because the restriction sites are lost when the correct ligation happens, the restriction and ligation reaction can be carried out simultaneously in one tube.

To allow for modular assembly of pathways, we constructed a toolbox that contains plasmids carrying the intergenic terminator-promoters and a receiver plasmid that has the first promoter and the last terminator. The assembly scheme is shown in Figure 3.3A. For proof of concept, these plasmids were put together with the PCR products of the zeaxanthin pathway genes (Crt E, B, I, Y, and Z), and an inducible zeaxanthin pathway was constructed in a one-step Golden Gate reaction. For this pathway, \( T_{PGI1-P_{GAL2}} \) was selected for all intergenic terminator-promoters, \( P_{GAL2} \) was selected as the first promoter and \( T_{ADH1} \) was selected as the last terminator (Figure 3.3B and Figure 3.4). As all five genes are under the same promoter, it allows for easier characterization of the inducible
pathway. The assembly efficiency for the five-gene pathway was about 16% (2/12), and an alternative pathway that has TTPH1-PGAL10 in the T1-P2 position was also constructed at a similar efficiency. When the assembled pathway was co-transformed with a plasmid carrying the P65-Gal-ER gene switch, the inducible expression of zeaxanthin occurred only in the presence of estradiol (Figure 3.5).

*Plasmid Stability.* The intergenic T-P regions are approximately 800 bp long, and since TPGI1-PGAL2 was used for all T-P positions, the resulting plasmid contained four 800 bp direct repeats. It is commonly believed that plasmids with repetitive sequences are unstable in *S. cerevisiae* due to the presence of a highly active homologous recombination machinery (1, 21). We therefore performed a test of plasmid stability to see if the pathway will remain intact over multiple generations of propagations.

Since the complete zeaxanthin pathway produces a yellow compound (and thus yellow colonies) readily detectable by visual inspection, we assessed the plasmid stability by monitoring the number of white versus yellow colonies over continuous culture. The above zeaxanthin pathway was separately transformed into YZE-55 and YZE-PA, then plated onto selective plates. The initial colonies were collected, diluted, and re-plated with estradiol. The number of yellow and white colonies in this first re-plate was labelled as day 0 in Table 3.1. A yellow colony was then picked for uninduced continuous culture, diluting daily at 1/100, and an aliquot of the culture was plated with estradiol every two days (Figure 3.6). As shown in Table 3.1, there was no significant increase in the number of white colonies after 8 days of continuous culture. With the exception of the day 0 re-
plate, there were hardly any white colonies. A cell growing from the day 0 re-plate to day 8 saturated culture had gone through about 63 generations of divisions, more than enough to saturate a 20,000L batch reactor, and yet, only about 0.5% of the cells had lost the pathway. This showed that the zeaxanthin pathway, despite multiple direct repeat sequences, could be stably propagated in *S. cerevisiae*.

**3.2.4 Inducible pathway characterization**

*Protein Level Induction.* While the promoters were characterized individually before the pathway assembly, putting multiple promoters together in the same plasmid may result in position dependency. Therefore, we sought to investigate whether the protein induction levels depend on where the gene is situated within the long concatenation of gene cassettes. Using the same toolbox plasmids as above, five more pathways were assembled, each with a different zeaxanthin pathway gene replaced with the GFP gene. The GFP gene in the five plasmids are thus all under P<sub>GAL2</sub> promoters. Assuming that neighbouring structural genes do not interfere with promoter activity, any difference in induction level can be attributed to positional effect. As shown in Figure 3.7A, GFP at all five positions were estradiol inducible. The basal level expressions were indistinguishable from background fluorescence. The protein induction level appeared to dip at the 3<sup>rd</sup> and 4<sup>th</sup> position, but the expression level at all five positions varied by no more than two-fold. This showed that the genes at all five positions can be effectively regulated, even though the exact induction strength has some position dependency.

*mRNA Level Induction.* Next, we investigated whether all five pathway genes can be induced simultaneously. Because the proteins were not tagged, we instead measured the
mRNA level induction of the zeaxanthin pathway genes. As shown in Figure 3.7B, all five zeaxanthin pathway genes can be induced, and the trend in relative induction strength is the same as what we have observed from the GFP fluorescence measurement. This gave further evidence of the positional dependency of promoter strength. As mRNA and protein abundance are not linearly correlated, a big change in mRNA level doesn’t correspond to a big change in protein expression level, only the trend is consistent. Interestingly, when mRNA induction ratio and protein abundance as reported by GFP fluorescence are plotted against each other, a power-law trend emerged (Figure 3.8) – which coincides with the trend observed by Beyer et al. in their genome scale study of transcription-translation relationships in *S. cerevisiae* (22).

**Inducible Strain Comparison.** Production of zeaxanthin was compared in the five YZE strains that differed in their promoters driving the gene switch expression. The inducible zeaxanthin pathway was transformed into the five strains and tested for their zeaxanthin production at 24 and 48 hours after induction. After sampling at 24h, the culture was re-inoculated 1/100 into fresh media. As shown in Figure 3.5B, high gene switch promoter strengths, e.g. YZE-100 and YZE-149, are bad for production. YZE-19, 55, and YZE-PA showed similar production level at 24h but YZE-PA’s production decreased after re-inoculation. YZE-19 and YZE-PA were picked for further comparison.

**Production Time Curve.** YZE-19 and YZE-PA, each harbouring an inducible pathway plasmid, were compared to YM954 harbouring a constitutive pathway. The constitutive pathway was obtained from a previous publication and used as a benchmark in our
comparison (16). To make the result more comparable, the constitutive pathway was re-cloned into pRS416K2 which is the backbone of the inducible pathway, and YZE’s parent strain, YM954, was used to host the pathway. As shown in Figure 3.7C, YZE-19 is the best producer. At 72h post induction, YZE-19 accumulated roughly twice the zeaxanthin compared to YZE-PA and about 50-fold more compared to the constitutive pathway. The measured concentration by HPLC was 1.5 µg/mL, which corresponds to 15 ng/OD600 cells or about 75 µg/g dry cell weight.

**Ligand Titration.** To identify an optimal inducer concentration, we performed a ligand titration in YZE-19 and YZE-PA. The strains harbouring the inducible pathway were subjected to different estradiol concentration, and the zeaxanthin production was measured after 48h. In agreement with the time course experiment, YZE-19 is a better producer. As shown in Figure 3.7D, it is in fact more sensitive to estradiol than YZE-PA. Peak induction in YZE-19 can be obtained at 10⁻⁸ M estradiol.

### 3.3 Discussion

In this study, we have presented a system for the construction and regulation of multi-gene pathways in *S. cerevisiae*. Genes of an entire pathway can be concatenated together, each driven by an inducible promoter, forming a transcriptional unit that can be regulated as a whole – similar to what can be achieved by a bacterial operon. By using inducible promoters of different strengths for different genes, it is possible to balance the flux through the pathway, which is the main challenge in metabolic engineering and synthetic biology (23).
We currently have a set of six promoters that are tightly regulated, with induction strength that spans an 8-fold difference from the weakest to the strongest. More inducible promoters have recently been developed by Blazeck et al. (24), and can easily be incorporated into our pathway construction system. As indicated in Figure 3.7D, induction strength can depend on the gene switch’s expression level, and can therefore differ in the different YZE strains. When using promoter strength to tune pathway gene expressions, it is also possible to pick induction strength through varying the ligand concentration. However, ligand titration curves need to be obtained in the host strain to find out what the induction level is at any given ligand concentration. Furthermore, in this non-saturation range, small differences in ligand concentration can cause large differences in induction strength, making it difficult to control the induction level. For the above reasons, we find that it is best to vary induction strength by varying maximum induction levels.

Based on the results of gene switch configuration assessment, we integrated the best gene switch into the chromosome to create five inducible yeast strains differing in gene switch expression levels. When used in conjunction with the inducible pathways, no additional plasmid is necessary. As shown in Figure 3.5B, high gene switch expression can be detrimental to the production of the desired product. Due to high clonal variation in adaptation, when a pathway is transformed into any given strain and induced on a plate, colonies with different degrees of yellow colouration (indicating zeaxanthin production) will occur. We picked both light- and dark-yellow colonies, assayed their product, and confirmed that they were all producing zeaxanthin, just at different levels (data not
shown). High producing colonies occur at a frequency of roughly 1 in 10, and the adaptation can be propagated. When comparing the zeaxanthin production level in the different strains, we selected the darkest yellow colonies from each plate.

Using the Golden Gate assembly method and our plasmid toolbox, the 5-gene zeaxanthin pathway can be assembled from their PCR products in a single step. This corresponds to a simultaneous ligation of 10 fragments forming a 13.3 kb plasmid with an efficiency of 16%. Similar efficiency was observed for the assembly of 2 zeaxanthin pathways and 5 GFP-replaced pathways. As the correct pathways could be picked from a reasonable number of colonies, no optimization was performed. For longer pathways, optimization of the Golden Gate reaction condition will probably be necessary, and should optimization fail, a two-stage hierarchical assembly can instead be used. In that case, the toolbox will be expanded to include two to three ampicillin-resistant intermediate plasmids. PCR products of the genes and the T-P fragments from T-P plasmids will first be assembled into the intermediate plasmids, up to 5 genes at a time. The intermediate plasmids will then be used as the substrate for a second Golden Gate reaction, which puts the intermediate assemblies together into the final receiver plasmid. If a target gene contains the BsaI restriction site, it is possible to PCR-amplify the gene in multiple segments, breaking and replacing each natural BsaI site through primers. These fragments can then be used in the 1-step Golden Gate assembly system as per normal. Another way to work around the restriction site is to use another Type IIIs restriction enzyme that gives a 4 bp 5’ overhang for the problematic target gene, and break the 1-step assembly to 2-step assembly involving first digesting and then ligating.
The initial transformations of the assembled zeaxanthin pathway have a significant fraction of white colonies (Table 3.1), but when a yellow colony is picked and propagated, only 0.5% white colony can be observed after 8 days of continuous cultivation, with no increase in the percentage of white colonies over the 8 days. This suggests that the plasmid may be less stable in *E. coli* such that the initial plasmid preparation contained a small portion of truncated recombinant plasmids, and the proportion was magnified through the preferential transformation of smaller plasmids. Interestingly, when the plasmid for yeast transformation was prepared in the Stbl3 strain instead of the TOP10 or DH5α strains, the fraction of white colonies decreased drastically. Since Stbl3 is a strain better suited for amplifying unstable plasmids, it gives further evidence for plasmid instability in *E. coli*.

Ideally, promoter strength should be independent of its position in the pathway, but this was not the case, at least for the GAL2 promoter. We observed that the induction strength tended to dip towards the middle of the pathway, and this has been supported at both protein (Figure 3.7A) and mRNA (Figure 3.7B) levels. One possible cause of this observation could be the 1-D search mechanism of DNA binding proteins. DNA binding proteins look for their cognate sequences in many ways, and one of the major search mechanisms is by first binding non-specifically to DNA, then performing a 1-D search along the DNA (25-27). The promoters at the flanks have access to a large ring of non-cognate DNA, i.e. the plasmid backbone, and the gene switch will therefore have a higher chance of finding these promoters. In contrast, the promoters in the middle will have a
much smaller binding area, and will have to rely more on a 3-D diffusion mechanism. More experiments are needed to verify this explanation and come up with a way to predict the modification of induction strength due to positional effect.

Despite some positional dependency of induction strength, the pathway is tightly regulated – with no measurable production in the absence of estradiol. To benchmark the induced production, we used a constitutive zeaxanthin pathway previously reported in literature (16). Under the condition of our production time curve experiment, the inducible pathway resulted in about 50-fold higher production level than the constitutive pathway. In making the comparison, we note that the genes on the constitutive pathway have five different constitutive promoters. Since constitutive promoters can sometimes be repressed under certain culture conditions, we may have just picked a poor production condition for the constitutive pathway. While we have not demonstrated the superior production capacity of an inducible pathway due to the above reason, we have at least demonstrated the greater reliability and predictability of an inducible system. Indeed, in another previous previously reported zeaxanthin production using S. cerevisiae, the authors achieved a similar production level as our inducible pathway using a mixture of inducible and constitutive promoters (2, 28).

We chose GAL4 DBD and its corresponding UAS in our study because the GAL4 DBD based induction system has been well characterized and it is commonly employed for inducible expression in yeast. There is, however, a drawback in the system – being a native yeast protein, there are multiple regulatory and interaction targets for GAL4 DBD
and some of these interactions can lead to unexpected phenotypes. For example, the YZE strains will not flocculate when induced, but if un-induced, will flocculate when the carbon source is exhausted. This could be due to interactions of GAL4 with other cellular signalling pathways, with the overall effect of tricking the cell into acting as if there is still galactose around.

3.4 Conclusions and outlook

With the tools introduced in this chapter, biosynthetic pathways can be easily refactored and studied in *S. cerevisiae*. More importantly, entire pathways can now be predictably and reliably induced using estradiol, a commonly available chemical. If desired, the ER LBD can also be engineered to generate new orthogonal ligand-receptor pairs, giving us the ability to regulate multiple pathways simultaneously and orthogonally (29-31). The ability to coordinate the expression of multiple genes in *S. cerevisiae* will be a useful addition to the toolbox of microbiologists, metabolic engineers, and synthetic biologists.

3.5 Material and methods

3.5.1 Strains, plasmids, media, reagents, and cell cultivation

TOP10 (*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG Δ-*) and Stbl3 (*F- mcrB mrr hsdS20 (rB Δ, mB Δ) recA13 supE44 ara-14 galK2 lacY1 proA2 rpsL20 (Str) xyl-5 Δ leu mtl-1r Δ*) (Life Technologies, Carlsbad, CA) were used for routine plasmid cloning and amplification in E. coli. YM954 (*MATa; ura3-52 his3-200 ade2-100 lys2-801 trp1-901 leu2-3,112 gal4-542 gal80-538*) (kindly provided by Florence Vignols and Stanley Fields) was used for
yeast experiments. Yeast strains were cultivated in either synthetic dropout medium (0.17% Difco yeast nitrogen base without amino acids and ammonium sulfate, 0.5% ammonium sulfate, and 0.083% amino acid dropout mix) or YPA medium (1% yeast extract, 2% peptone, and 0.01% adenine hemisulfate) supplemented with 2% glucose as a carbon source. E. coli strains were cultured in Luria broth (LB; Fisher Scientific, Pittsburgh, PA, USA).

YZE-19 through 149 and YZE-PA were created in this study by integrating the P65-GalER (6GE) gene switch into YM954 at the LYS2 locus, and are lys2+. pLys2-P<sub>ENO19,55,100,149</sub>-6GE and pLys2-P<sub>ADH1</sub>-6GE were linearized with DraIII and transformed into YM954 by chemical transformation as described elsewhere (32). Integrants were selected using SC-Lys plates and confirmed by PCR as well as functional assay.

All restriction and DNA processing enzymes were obtained from New England Biolabs (Ipswich, MA) and all chemicals were obtained from Sigma Aldrich (St Louis, MO) unless otherwise specified. PCR reactions were carried out using Phusion DNA Polymerase (Dharmacon, Lafayette, CO) and primers from Integrated DNA Technologies (Coralville, IA).

3.5.2 Plasmid construction

The natural GAL promoters were PCR amplified directly from the genomic DNA of S. cerevisiae. The resulting PCR products were overlap extended with another PCR fragment encoding the green fluorescent protein (GFP) and the ADH1 terminator. The promoter-gene-terminator cassette was subsequently ligated into pRS416 centromeric
vector (ATCC 87521) via BamHI and SacI to create pRS416-Gal1-GFP, pRS416-Gal2-GFP, and pRS416-Gal10-GFP. The synthetic GAL UASes were synthesized from overlap extension of oligonucleotides. These were overlap extended with the TATA box region of Gal1 and Gal2 to create the SYN1 and SYN2 series of promoters respectively, and subsequently cloned into pRS416 to create pRS416-SYN1-GFP and pRS416-SYN2-GFP. Sequences of SYN1 and SYN2 promoter can be found in Table 3.4.

The gene switch constructs were generated by ligation of either a P65 or GAL4 activation domain (AD) with an estrogen receptor ligand binding domain (ER LBD) and a GAL4 DNA binding domain (DBD). The P65 and ER LBD were PCR amplified from human genomic DNA whereas the GAL4 AD and DBD were amplified from S. cerevisiae genomic DNA. The fragments were cloned directly into pRS414-PMT, which is a pRS414 vector (ATCC 87519) modified to contain an ADH1 promoter and ADH1 terminator interspaced by a multiple cloning site.

The gene switch integration plasmids, pLys2-PENO19, 55, 100, 149-6GE and pLys2-PADH1-6GE were created by ligating the promoter fragments (EcoRI, KpnI) and 6GE-TADH1 (KpnI, MluI) into pLys2 (EcoRI, MluI), which is pNEB193 (New England Biolabs, Ipswich, MA) with the LYS2 gene inserted into the AatII site. PENO19, 55, 100, 149 promoters were obtained by random mutagenesis of the wild type ENO promoter as described elsewhere (33).
The T-P plasmids were constructed by the ligation of the PGI1 terminator (AflII, EcoRI) and the GAL2 promoter (EcoRI, XbaI) into pFUS_A (AflII, XbaI) with appropriate BsaI sites introduced at the two ends of the T-P cassette. pFUS_A is an *E. coli* plasmid with a spectinomycin resistance marker as described elsewhere (34). The PGI1 terminator was PCR amplified from pRS416-Zeax which is a constitutive zeaxanthin pathway as described elsewhere (16), whereas the GAL2 promoter was PCR amplified from the above-described pRS416-Gal2-GFP.

The Golden Gate receiver plasmid, containing the first promoter and the last terminator interspaced by a BsaI-flanked LacZα cassette, was constructed by the overlap extension of the GAL2 promoter (from pRS416-Gal2-GFP), LacZα cassette (from pFUS_A), and ADH1 terminator (from pRS416-Gal2-GFP). The overlap extended fragment was cloned into pRS416K2 via HindIII and NotI. pRS416K2 was modified from pRS416 by the removal of all BsaI sites, the LacZα remnant, the f1 origin, and by the replacement of AmpR with KanR.

3.5.3 Pathway assembly

The Golden Gate reaction for pathway assembly was carried out in 20 µL using 100 ng of receiver plasmid, 100 ng of each T-P plasmid, 10 ng/kb of the PCR gene fragments, 1 µL of BsaI-HF, and 1 µL High Concentration T4 DNA ligase, in 1X T4 ligase buffer. The reaction was thermocycled as follows: 37 ºC for 10 min, (37 ºC for 5 min, 16 ºC for 10 min) repeated 10 times, 37 ºC for 10 min, 75 ºC for 5 min. The ligation product was transformed into TOP10 competent cells (Life Technologies, Carlsbad, CA) by heat-shock, and white colonies on X-GAL/IPTG plates were screened for the correctly
assembled pathway. Primer sequences used to obtain the PCR gene fragments are listed in Table 3.2.

3.5.4 Measurement of GFP expression

GFP expression level was measured by flow cytometry using LSRII (BD Biosciences, Franklin Lakes, NJ). Single colonies were picked into liquid synthetic dropout media and grown overnight with or without induction. Small aliquots of the cultures were then washed and resuspended in phosphate buffered saline for flow cytometry analysis.

3.5.5 Measurement of zeaxanthin production

Cells from 6 mL of \textit{S. cerevisiae} dropout media culture were collected by centrifugation, resuspended in 1 M sorbitol, 10 mM Tris Buffer pH 7.4, 5 U/mL Zymolyase (Zymo Research, Irvine, CA), and incubated at 37 °C for 45 min. Cells were then collected again by centrifugation and 600 µL methanol was used to extract the zeaxanthin directly. For quantification, 80 µL of methanol extract was loaded onto an Agilent ZORBAX SB-C18 column and monitored at 450 nm on an Agilent 1100 series HPLC (Agilent Technologies, Palo Alto, CA). The pump program was 0.6 mL/min, 100% methanol for 10 min. Elution was monitored at 450 nm with reference set at 360 nm. Authentic zeaxanthin from Sigma (St Louis, MO) was used as a standard.

3.5.6 RNA preparation and quantitative PCR

Yeast colonies were picked from synthetic dropout plates into 3 mL synthetic dropout medium and grown overnight at 30 °C with shaking at 250 RPM to saturation. 100 µL of overnight culture was then used to inoculate 3 mL of fresh media, and estradiol was added to the appropriate samples. After growing for 18 hours, cells from 0.5 mL of each
culture were harvested by centrifugation. RNA was then isolated using the RNeasy Mini Kit (QIAGEN, Valencia, CA) according to the manufacturer’s protocol. 1 µg of the resulting total RNA was treated by 10 units of recombinant DNaseI (Takara Bio, Otsu, Shiga, Japan) in 20 µL for 40 minutes according to the manufacturer’s protocol. Complementary DNA (cDNA) was then synthesized using the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA) according to the manufacturer’s protocol and stored at -20 °C until use. Quantitative PCR primers were designed using the Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Table 3.3). Reactions were performed using Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems, Austin, TX) according to the manufacturer’s protocol. Expression was quantified relative to the ALG9 gene using the standard curve method (35).
3.6 References


3.7 Figures and tables

Figure 3.1 The behaviour of GAL inducible promoters, monitored via GFP expression and induced using Gal-P65. The GFP channel histograms obtained from flow cytometry are shown. Black indicates uninduced sample and red indicates induced sample. The sample identity and the percentage of positives are shown within each histogram. (A) The behaviour of natural GAL inducible promoters cloned from *S. cerevisiae*. (B) The behaviour of synthetic GAL inducible promoters, which were generated by fusing the synthetic UAS sequences with the TATA box of the GAL1 promoter. (C) Similar to (B), but fused with the TATA box of the GAL2 promoter.

![Histograms of GFP expression](image)

Figure 3.2 (A) Gene switch activity monitored via GFP expression, and measured using flow cytometry. GFP expression levels are represented by the arithmetic means of GFP fluorescence (arbitrary unit). The mean of negative control has been subtracted from that of the samples. Blue columns represent uninduced samples, and red columns
represent induced samples. Values are the average of two independent samples and error bar indicates the standard error. Labels: CTR is negative control with empty plasmids, and Y2H is a yeast two hybrid system positive control from (29). GE, EG, AGE, AEG, 6GE, and 6EG are different combinations and permutations of Gal4 DBD(G), ER LBD(E), P65(6), and Gal4 AD(A) ordered from the N-terminus of the resulting gene switch protein. (B) Inducibility of the YZE strains monitored via GFP expression. The GFP channel histograms are shown. Black indicates negative control without gene switch, blue indicates uninduced sample, and red indicates induced sample.
Figure 3.3  (A) The inducible pathway assembly scheme. The T-P plasmids have BsaI-excisable terminator and promoter fragments, and the receiver plasmid has BsaI-excisable LacZα marker for blue-white screening. The toolbox plasmids together with the PCR products of the pathway genes can be used to assemble an inducible pathway in a Golden Gate one pot reaction. BsaI sites are in light blue, Kan is the kanamycin resistance gene, Spe is the spectinomycin resistance gene. (B) A schematic of the assembled inducible zeaxanthin pathway. All 4 intergenic T-P fragments are identical except for the 4bp overhang region.
**Figure 3.4** Unique 4 bp overhang assignment in pathway assembly. By judiciously choosing the 4 bp overhang created by *Bsa*I, pathways can be assembled in a user-specified order. The TP fragments and the receiver fragment (pRS416K2) are supplied by toolbox plasmids, whereas the Crt fragments are PCR products.
Figure 3.5  (A) HPLC trace showing the inducible production of zeaxanthin. (B) Zeaxanthin production in different YZE hosts. Blue columns represent data from the first 24-hour samples. At the end of the 24-hour period, cells were re-inoculated into fresh media for another 24-hour. Red columns represent data from the second 24-hour samples. The values represent the average of 4 independent samples, and the error bar represents the standard error.
Figure 3.6  Inducible zeaxanthin pathway stability assessment. The plasmids harboring the assembled zeaxanthin pathway were purified form E. coli and used to transform YZE-PA and YZE-55. After two days incubation, the initial colonies were collected and re-plated onto an SC-URA plate with 0.1 µM estradiol. When colonies reappeared, the numbers of white and yellow colonies were counted. A yellow colony was subsequently picked for continuous cultivation in SC-URA media, diluted daily at 1/100. Aliquots of the continuous culture were plated on SC-URA with estradiol on day 2, 4, 6, and 8.
Figure 3.7  (A) Protein level induction at the different positions using EGFP as the reporter. The genes along the pathway are arranged from 5’ to 3’ in the order Crt E-B-I-Y-Z. Zea is the negative control with the 5-gene inducible zeaxanthin pathway without any EGFP gene in it. E, B, I, Y, Z:EGFP have their respective zeaxanthin pathway gene replaced by EGFP. The uninduced EGFP-replaced pathways have identical basal expression levels, and they have been collapsed under crt:EGFP. The values represent the average of 4 independent samples, and the error bar represents the standard error. (B) mRNA level induction at the different positions as measured by quantitative RT-PCR. The value represents the average of 2 independent samples, and the error bar represents the standard error. (C) Production time course that benchmarks the inducible pathway in two different YZE strains to a constitutive pathway in YM954. The values represent the average of 3 independent samples, and the error bar represents standard error. (D) Ligand titration curve of the zeaxanthin production against the estradiol concentration. The values represent the average of 2 independent samples, and the error bar represents standard error.
Figure 3.8  
mRNA induction vs EGFP expression.

\[ y = 163.58x^{0.3095} \]

\[ R^2 = 0.99538 \]
Table 3.1  Assessment of plasmid stability. White colony is an indication that the yeast cell has lost at least part of the zeaxanthin pathway. The colony counts are the average of two independent experiments.

<table>
<thead>
<tr>
<th>Integrant</th>
<th>Time</th>
<th>Total Colony</th>
<th>White Colony</th>
<th>% White</th>
</tr>
</thead>
<tbody>
<tr>
<td>YZE-55</td>
<td>Day 0</td>
<td>60</td>
<td>12</td>
<td>20.00</td>
</tr>
<tr>
<td></td>
<td>Day 2</td>
<td>243.5</td>
<td>0.5</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>Day 4</td>
<td>195</td>
<td>1</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Day 6</td>
<td>360</td>
<td>1</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>Day 8</td>
<td>267.5</td>
<td>1.5</td>
<td>0.56</td>
</tr>
<tr>
<td>YZE-PA</td>
<td>Day 0</td>
<td>110</td>
<td>39</td>
<td>35.45</td>
</tr>
<tr>
<td></td>
<td>Day 2</td>
<td>295</td>
<td>1.5</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Day 4</td>
<td>195.5</td>
<td>0.5</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>Day 6</td>
<td>404</td>
<td>2</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Day 8</td>
<td>267</td>
<td>1</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Table 3.2  Primer sequences used in zeaxanthin pathway cloning. The black portion is gene-specific, and the blue portion is position-specific. The 4 bp overhangs are underlined, and they determine the order in which the genes are assembled.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequences (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>BsaI&gt;-CrtE For</td>
<td><strong>CTTGAAGGTCTCC</strong>CATATA  ATGACGGTCTGCGCAAAAAAACAC</td>
</tr>
<tr>
<td>CrtE&lt;-BsaI Rev</td>
<td><strong>CAATAAGGTCTCG</strong>CTACCT GAACTGACGGCAGCGAGTTTTTTGTC</td>
</tr>
<tr>
<td>BsaI&gt;-CrtB For</td>
<td><strong>CTTGAAGGTCTCCC</strong>AAAAATGAATAATCCGTCGTTACTCAATCATGCG</td>
</tr>
<tr>
<td>CrtB&lt;-BsaI Rev</td>
<td><strong>CAATAAGGTCTCG</strong>GATGCTAGAGCGGGCGCTGCCAG</td>
</tr>
<tr>
<td>BsaI&gt;-CrtI For</td>
<td><strong>CTTGAAGGTCTCC</strong>AACAATGAAACCAACTACGGTAATTGGTGCAGG</td>
</tr>
<tr>
<td>CrtI&lt;-BsaI Rev</td>
<td><strong>CAATAAGGTCTCG</strong>GTCA TCATATCAGATCCTCCAGCATCAAACCTG</td>
</tr>
<tr>
<td>BsaI&gt;-CrtY For</td>
<td><strong>CTTGAAGGTCTCC</strong>CACAATGCAACCGCATTATGATCTGATTCTCG</td>
</tr>
<tr>
<td>CrtY&lt;-BsaI Rev</td>
<td><strong>CAATAAGGTCTCG</strong>GGA TTAACGATGAGTCGTCATAATGGCTTTGC</td>
</tr>
<tr>
<td>BsaI&gt;-CrtZ For</td>
<td><strong>CTTGAAGGTCTCC</strong>TACAATGTTGTGGATTTGGAATGCCCTG</td>
</tr>
<tr>
<td>CrtZ&lt;-BsaI Rev</td>
<td><strong>CAATAAGGTCTCG</strong>GGATTTACTTCCCGGATGCGGGC</td>
</tr>
</tbody>
</table>

Table 3.3  Primer sequences used in quantitative RT-PCR.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG9 Forward</td>
<td><strong>CTACCATCAGAACCGCATTC</strong></td>
</tr>
<tr>
<td>ALG9 Reverse</td>
<td>TCCATGATACAGGAGCAAGC</td>
</tr>
<tr>
<td>crtB Forward</td>
<td>TTGCGGCTTTTTCAGGAAG</td>
</tr>
<tr>
<td>crtB Reverse</td>
<td>AATAGCGCAGGTATCATCC</td>
</tr>
<tr>
<td>crtE Forward</td>
<td>AACTGCTGGACGATTCTGAC</td>
</tr>
<tr>
<td>crtE Reverse</td>
<td>TCATCGACGAGCTGAAGATG</td>
</tr>
<tr>
<td>ctrl Forward</td>
<td>AAGCCGTGCATTAGAGGAC</td>
</tr>
<tr>
<td>ctrl Reverse</td>
<td>TCTGCAAGTTTGTGACTG</td>
</tr>
<tr>
<td>crtY Forward</td>
<td>TTGACTGAGAGCCAACATCG</td>
</tr>
<tr>
<td>crtY Reverse</td>
<td>TGTCGCTGTAAAACCTCAGC</td>
</tr>
<tr>
<td>crtZ Forward</td>
<td>GCCATTCGGCTATATCCAC</td>
</tr>
<tr>
<td>crtZ Reverse</td>
<td>CGCATAGAGGAAGCAAAAACAG</td>
</tr>
</tbody>
</table>
Table 3.4 SYN1 and SYN2 promoter sequences.

<table>
<thead>
<tr>
<th>SYN1A (4 x UAS)</th>
<th>SYN1B (3 x UAS)</th>
<th>SYN1C (2 x UAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TATTGAAAGTACGGAGTGTAGGCTCGAGATGC</td>
<td>TATTGAAAGTACGGAGTGTAGGCTCGAGATGC</td>
<td>TATTGAAAGTACGGAGTGTAGGCTCGAGATGC</td>
</tr>
<tr>
<td>GCTCCGTACACGGAGTGTAGGCTCGAGATGC</td>
<td>GCTCCGTACACGGAGTGTAGGCTCGAGATGC</td>
<td>GCTCCGTACACGGAGTGTAGGCTCGAGATGC</td>
</tr>
<tr>
<td>TATA GAL1</td>
<td>TATA GAL1</td>
<td>TATA GAL1</td>
</tr>
<tr>
<td>AAAAGATTCTACACACACTACGGAGGTGCACG</td>
<td>AAAAGATTCTACACACACTACGGAGGTGCACG</td>
<td>AAAAGATTCTACACACACTACGGAGGTGCACG</td>
</tr>
<tr>
<td>TATA GAL2</td>
<td>TATA GAL2</td>
<td>TATA GAL2</td>
</tr>
<tr>
<td>CGAGATTAGTTAAGGCTCTCGAGATGCGATG</td>
<td>CGAGATTAGTTAAGGCTCTCGAGATGCGATG</td>
<td>CGAGATTAGTTAAGGCTCTCGAGATGCGATG</td>
</tr>
<tr>
<td>TATA ATG</td>
<td>TATA ATG</td>
<td>TATA ATG</td>
</tr>
</tbody>
</table>

**TATA GAL1**

AAAAGATTCTACACACACTACGGAGGTGCACG

**TATA GAL2**

CGAGATTAGTTAAGGCTCTCGAGATGCGATG

---

74