Exploring Genetic Tools for Metabolic Engineering of Yeast *Yarrowia lipolytica*

Master's thesis in Biotechnology

HANNA ÖSTERMARK

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's thesis 2022

Exploring Genetic Tools for Metabolic Engineering of Yeast *Yarrowia lipolytica*

HANNA ÖSTERMARK

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Biology and Biological Engineering
Division of Systems and Synthetic Biology
Kerkhoven Lab
Chalmers University of Technology
Gothenburg, Sweden 2022

Exploring Genetic Tools for Metabolic Engineering of Yeast *Yarrowia lipolytica*
HANNA ÖSTERMARK

Supervisor: Eduard Kerkhoven, Department of Biology and Biological Engineering
Co-supervisor: Simone Zaghen, Department of Biology and Biological Engineering
Examiner: Eduard Kerkhoven, Department of Biology and Biological Engineering

Master's Thesis 2022
Department of Biology and Biological Engineering
Division of Systems and Synthetic Biology
Kerkhoven Lab
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

Exploring Genetic Tools for Metabolic Engineering of Yeast *Yarrowia lipolytica*
HANNA ÖSTERMARK
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

The oleaginous yeast *Yarrowia lipolytica* has several properties that gives it a high potential as a microbial cell factory for production of oleochemicals, proteins, and natural products. Although genetic tools have been developed for the yeast, more are needed to expand microbial fermentation using *Y. lipolytica*. The purpose of the thesis is to explore two genetic tools for metabolic engineering of *Y. lipolytica*. First, the functionality of three enzymes repurposed as colorimetric malonyl-CoA biosensors was tested. Second, a set of 18 tunable promoters with differential expression in different nitrogen conditions was investigated.

In the biosensor project, a series of shake flasks experiments were performed to explore the properties of polyketide synthases repurposed as a biosensor. The biosensor signal strength was measured through absorbance in the UV-visible spectra. The dose-dependency of the biosensor was tested by employing it in three strain backgrounds, each assumed to have a different malonyl-CoA availability. The biosensor signal varied depending on the strain background, but not always as expected. Overexpression of the biosensor enzyme gene showed that the strain with increased lipid accumulation that was expected to give the lowest signal, actually gave the highest. This could be explained by the strain oxidising lipids, thereby increasing the malonyl-CoA concentration.

In the promoter project, a strain for each promoter was constructed. GFP was used to measure the expression strength. The strains were grown in different media in a microbioreactor system which measured the biomass and fluorescence intensity. Some strains were also investigated in a fluorescence microscope. None of the tested promoter strains showed GFP expression, meaning the promoters did not initiate GFP gene transcription. The troubleshooting steps taken could not explain the unexpected results.

More research is needed to establish the tools tested in this thesis. Instead of acting as biosensors, the polyketide synthases could be used for polyketide production. The polyketide produced would then need to be identified, followed by metabolic engineering strategies for increased production. To increase the detection range for the tested promoters, a brighter fluorescent protein and a more sensitive instrument could be used.

Keywords: *Yarrowia lipolytica* | genetic engineering | promoter | biosensor | malonyl-CoA | polyketides | lipid metabolism

Acknowledgements

First of all, a big thanks to the Yarrowia group for all the help and advice, I will miss our weekly meetings. We have had interesting discussions about research, both mine and others. You have guided me in my research and have steered me towards becoming a better scientist. I am especially thankful to my supervisor and examiner Ed, and my co-supervisor Simone. You have both aided and challenged me in the production of this thesis, Ed in the big picture and Simone on a daily basis.

I would like to thank the entire Sysbio division for welcoming me as one of their own and for all the fun over this last year. Thanks to the research engineers and the admins for making it such a great workplace.

Thank you to all the students in the Dungeon for discussions about the lab work, or just listening when I complained about yet another failed transformation. This year had not been as fun without you.

Finally, a thank you to my friends and family for the support and love I have received during this thesis. A special thank you to Thomas for listening and being curious about my research, even if you did not always understand.

Hanna Östermark, Gothenburg, June 2022

Contents

List of Figures

List of Tables

1
Introduction

In a world where the population is growing and emissions are increasing, sustainable development has never been more important. To satisfy the expanding need for fuels, commodity chemicals, and pharmaceuticals, microbial fermentation can be used. This sustainable production method can utilise renewable substrates, decreasing the need for fossil fuel-derived carbon. The fermentation can be done in milder conditions, such as lower temperatures and without toxic solvents, compared to more traditional chemical production settings, which is advantageous regarding both environmental and health aspects. Furthermore, a high product specificity is possible due to genetic engineering methods [1]. Some well-known examples of microbial cell factories used today are *Saccharomyces cerevisiae* and *Escherichia coli*. *S. cerevisiae*, or baker's yeast, is known for its use in bioethanol production and in food fermentation [2]. However, it can also be used to produce more advanced products. Both *S. cerevisiae* and *E. coli* can be engineered to produce, among other products, heterologous proteins [3][4], insulin [5], and bulk chemicals [1]. These microorganisms have been the primary choice for microbial cell factories, since they are well-studied with a lot of genetic tools developed specifically for them [1]. However, they also present some limits. With recent research in systems biology and development of efficient genetic manipulation methods, more non-conventional cell factories can be employed [1]. One area where *S. cerevisiae* and *E. coli* are limited is production of oleochemicals [1]. Here it is more advantageous to use an oleaginous yeast as *Yarrowia lipolytica*. It also has a high potential in other, such as protein or biofuel production [6]. The *Y. lipolytica* genome is sequenced [7] and there are several established toolboxes to genetically manipulate it [8]. However, the knowledge and library of available tools is not as large as for *S. cerevisiae* and *E. coli* [9][10]. To further expand the use of *Y. lipolytica* as a microbial cell factory, and utilise its high potential for specific products, more molecular tools need to be developed.

1.1 Aim and scope

The aim of the thesis is to explore different genetic tools for the metabolic engineering of *Y. lipolytica*. Developing these tools would increase the possibilities of utilising this yeast as a microbial cell factory. Specifically, two areas are explored in two separate projects: 1) investigating the functionality of three enzymes repurposed as colorimetric malonyl-CoA biosensors, and 2) comparing a set of uncharacterised, tunable promoters, whose expression strength vary depending on nitrogen availability and source.

In the biosensor project, three enzymes are repurposed as biosensors for malonyl-CoA. They were selected based on an article by Yang et al. (2018) [11]. Previously, the enzymes were established and expressed in *Y. lipolytica* in two different strain backgrounds (unpublished).

In the promoter project, 27 native promoters were chosen from *Y. lipolytica* based on transcriptomics data from chemostat cultivations in a previous project (unpublished). 17 of the selected promoters showed differential expression depending on the nitrogen availability and source. The remaining 10 promoters had stable expression in the tested conditions.

1.2 Thesis outline

The thesis is divided into theory, methods, results and discussion, and finally conclusions. In the theory chapter, the yeast *Y. lipolytica* is introduced, along with its fatty acid metabolism. A selection of the current synthetic biology tools for the yeast is described. The theory also includes information about polyketides and how to produce them in *Y. lipolytica*. The chapter ends with a description of biosensors and an example of one developed for malonyl-CoA-derived products. The methods chapter contains the materials, methods, and instruments used in the thesis work to produce data. After that, the results are presented. First, the biosensor candidates are compared in different media and strain backgrounds at different cultivation time points, followed by an evaluation of the biosensor's linear dynamic range. Then, the results from testing the promoters are presented, along with troubleshooting steps taken to clarify the unexpected results. The results chapter also includes a discussion on the acquired results. Last, the conclusions chapter summarises the results of the two projects and which conclusions can be drawn from them, along with the impact of the thesis on future research on *Y. lipolytica*.

2

Theory

2.1 *Yarrowia lipolytica*

The non-conventional yeast *Yarrowia lipolytica* has quickly become more popular. The yeast has been isolated from dairy products, such as cheese, and from lipid-rich environments as sewers [12]. It is also a dimorphic yeast, forming either round budding yeast cells or pseudohyphae and septate hyphae. The morphology depends on many external factors, such as pH, nutrient sources, and aeration [13]. *Y. lipolytica* is non-pathogenic [12] and it has been classified as a Generally Recognized As Safe (GRAS) host for production of several chemicals [14]. Further, its genome was first sequenced in 2004 [7].

Y. lipolytica has several genomic and physiological properties that gives it a high potential as a microbial cell factory for production of both natural products and fatty acid-based chemicals [15]. An important property for microbial cell factories is robustness. *Y. lipolytica* can grow in varying pH conditions [15]; it has been reported to grow in pH between 2.5 and 7.5 [16]. The yeast also has high tolerance to increased salt levels and organic compounds [15][17]. Further, several different carbon sources can be utilised by the yeast [18], increasing the potential for sustainable fermentation. *Y. lipolytica* is oleaginous, meaning it has the capability to accumulate storage lipids in the form of triacylglycerols (TAGs) to more than 40% of its dry cell weight [12]. An inherent high flux through biological precursors as acetyl-CoA and malonyl-CoA [19] contributes to the high lipid accumulation.

2.1.1 Fatty acid synthesis

A simplified version of the lipid metabolism in *Y. lipolytica* is presented in figure 2.1. Under nutrient-rich conditions, glucose is taken up by the cell and undergoes glycolysis. This generates pyruvate, which in turn enters the tricarboxylic acid (TCA) cycle as acetyl-CoA. This metabolic process (along with oxidative phosphorylation) is the main way the cell generates energy in form of ATP. If nitrogen becomes limiting to the cell, but glucose is still in abundance, lipid accumulation is favoured to store the energy [18][20]. Storing energy as TAGs instead of as carbohydrates or proteins is favoured from an evolutionary perspective, likely due to the much higher energy density in TAGs since they are anhydrous and reduced [21]. The energy in the TAGs can then be accessed later, in environments where the carbon source availability is low.

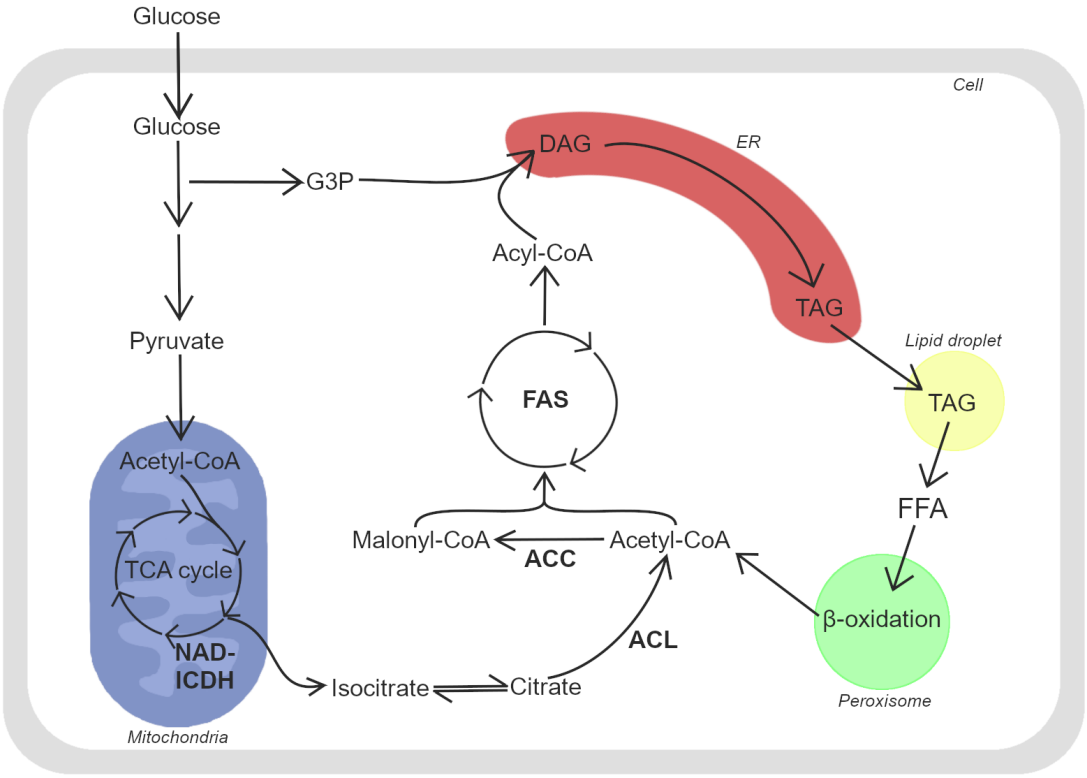


Figure 2.1: Overview of lipid metabolism in *Y. lipolytica*. Abbreviations: ACC, acetyl-CoA carboxylase; ACL, ATP citrate synthase; DAG, diacylglycerol; ER, endoplasmic reticulum; FAS, fatty acid synthase; FFA, free fatty acid; G3P, glyceraldehyde 3-phosphate; NAD-ICDH, NAD+-isocitrate dehydrogenase; TAG, triacylglycerol; TCA cycle, tricarboxylic acid cycle.

Nitrogen limitation activates the enzyme adenosine monophosphate (AMP) deaminase (AMPD) [22], which converts AMP to inosine monophosphate (IMP) and ammonium ions. This is presumed to compensate for the insufficient nitrogen levels [22]. The decreased concentration of AMP reduces the activity of the TCA cycle enzyme AMP-dependent NAD+-isocitrate dehydrogenase (NAD-ICDH), which leads to accumulation of isocitrate. Isocitrate can then isomerise to citrate by aconitase. The citrate can either be excreted by the cells (it is a common overflow metabolite in *Y. lipolytica*), or can be converted to acetyl-CoA in the cytosol by ATP citrate synthase (ACL). This enzyme linking the carbohydrate metabolism and the fatty acid biosynthesis is only present in oleaginous yeast [23]. In the cytosol the enzyme acetyl-CoA carboxylase (ACC) can convert acetyl-CoA to malonyl-CoA. The fatty acid synthase (FAS) complex catalyses the fatty acid synthesis by adding malonyl-CoA units to acetyl-CoA, elongating the fatty acid backbone with two carbons for each malonyl-CoA to produce acyl-CoA [23]. The length of the fatty acids is usually 16 or 18 carbons, which later can undergo desaturation [23]. Together with glyceraldehyde 3-phosphate (G3P) from the glycolysis, acyl-CoA can be acylated in steps in the Kennedy pathway in the endoplasmic reticulum (ER) to form diacylglycerol (DAG) [24]. DAG can then be acylated by one of two enzymes to form TAG: diacylglycerol acyltransferase (DGA1) with acyl-CoA as acyl donor, or phospholipid diacylglycerol acyltransferase (LRO1) with glycerophospholipids as acyl donor [23]. Then, the TAGs can be stored as lipid droplets [23]. If the energy stored as lipids

is needed, fatty acid degradation can occur through β-oxidation. The β-oxidation of the free fatty acids (FFA) takes place in the peroxisome, where the fatty acid backbone is reduced in a four-step cycle. As a result, acetyl-CoA is released and is free to enter the TCA cycle to produce ATP [23].

2.2 Genetic engineering of *Y. lipolytica*

To produce non-native metabolites, microorganisms need to be genetically engineered. There are many synthetic biology methods available for genetic modification of *Y. lipolytica*, but the general workflow remains the same [10]. The first step is choosing which DNA parts to use in the construction of the gene expression cassette. To achieve the desired expression in the cell, many aspects need to be considered; the promoter and terminator affect the expression level, as does the choice between plasmid expression or genomic integration. Furthermore, if the protein should be secreted or needs to be localised in the cell, an appropriate tag needs to be selected. Lastly, selection markers are necessary for selecting putative positive integrations [10].

The most important aspect to controlling gene expression is the promoter [8]. As eukaryotic promoters are diverse and can involve several elements, the strength of a promoter is determined by many factors. The core promoter, TATA box, upstream activating sequences (UASs), and proximal promoter sequences all influence the expression level [8]. For *Y. lipolytica* there are several promoters that are well-characterised and commonly used, for instance the native promoter for translation elongation factor-1 (p*TEF1*) [25]. p*TEF1* is a constitutive promoter with a high expression level in *Y. lipolytica* [8][9]. To increase expression of a specific promoter further, a hybrid promoter can be developed by fusing the core promoter with UASs [8][10]. Synthetic promoters like these are often longer than native promoters, which generally are shorter than 1000 bp [10]. Another strategy is to include an intron region from the correlating gene, which has been done for p*TEF1*. By including the start codon and 113 bp of the endogenous intron, the *TEF1* intron promoter (p*TEF1in*) gave a 17-fold stronger expression than the intronless version [26]. Gene expression can be altered by introns in several ways. The intronic region can contain enhancers, a transcriptional regulatory element increasing transcription initiation. Furthermore, the intron can facilitate mRNA export and its translation rate [27]. Seemingly, the stronger a promoter is, the higher protein expression it gives. However, this is not always true. Dulermo et al. (2017) discovered that in some cases for *Y. lipolytica*, a weaker promoter provided a more optimal protein expression [28]. Heterologous protein expression competes with normal functions, meaning it is a metabolic burden for the cell. Too high gene expression could lead to protein misfolding or forming of aggregates, or could lead to stress responses, all of which could reduce protein titres [29]. Therefore, it is vital to continue to discover promoters for *Y. lipolytica*. With a bigger pool of promoters available, it is possible to fine-tune recombinant expression in *Y. lipolytica*.

After deciding the design of the integration fragment, the different parts of DNA need to be cloned into a vector. The traditional method of using restriction enzymes to join together fragments is time-consuming. More time- and cost-efficient methods have been developed, of which many are available for *Y. lipolytica* [10]. One example is Gibson assembly, an assembly method where multiple overlapping fragments can be ligated in one step. Three enzymes are used in the process. First, a 5' exonuclease digests the double-stranded DNA fragments from the 5' end, leaving 3' complementary overhangs that anneal with each other due to the overlapping regions. Next, a DNA polymerase fills in the gaps left by the exonuclease, and the DNA nicks are finally repaired by a DNA ligase [30].

To attain the highest genetic stability in gene expression, the constructed vector containing the gene cassette should be integrated into the genome [10]. This can be achieved by utilising the yeast's repair mechanisms for double-stranded DNA breaks. By flanking the promoter-gene-terminator construct with DNA regions homologous to up- respective downstream of the integration site and breaking both strands of the DNA at the integration site, the yeast will use homologous recombination (HR) and integrate the gene construct. For acceptable rates of HR in *Y. lipolytica*, the homologous regions need to be 500-1000 bp long. The explanation is that double-stranded DNA breaks are mainly repaired using non-homologous end joining (NHEJ) instead of HR. To increase the HR use, and thus the success rate of genomic integration, the gene *KU70*, which codes for a protein responsible for double-strand break repair in NHEJ, can be deleted or disrupted [31][32]. Of all genome editing techniques, clustered regularly interspaced short palindromic repeats (CRISPR) techniques have grown to be some of the most popular. CRISPR is combined with the nuclease Cas9 to introduce double-stranded breaks in the genome at a specific location, which is defined by a short RNA fragment called single guide RNA (sgRNA). By simultaneously introducing a repair fragment (RF) with homologous ends to the break site, the cell can be coerced into repairing the DNA with a constructed gene cassette [33].

Holkenbrink et al. (2018) created one of several CRISPR-Cas9 systems for *Y. lipolytica* called EasyCloneYALI [34]. The toolbox offers two genome-editing operations mediated by CRISPR-Cas9: marker-free integration of expression constructs and knockout or mutation of genes. Specific genome sites with high expression have been selected, and a set of guide RNA (gRNA) vectors respective a set of integration vectors adapted for each site are available. The cloning site in the integration vectors is flanked by one terminator in each direction, allowing for expression of two genes in each constructed plasmid. In turn, the two terminators are flanked by DNA regions homologous to the specific genome sites, followed by restriction sites for the restriction enzyme NotI. By digesting the constructed plasmid containing the cassette to be integrated, a linear repair fragment (RF) is obtained. Both the integration vectors and the gRNA vectors contain the selection marker *AmpR* which confer resistance to the antibiotic ampicillin for plasmid amplification in *E. coli*. The gRNA vectors also contain the gene *Nat* (resistance to nourseothricin) as a selection marker in *Y. lipolytica* [34]. In this thesis, the EasyCloneYALI toolbox is used on strains derived from the strain ST6512 [35], in which the Cas9 gene is inserted in

the *Y.lipolytica* genome in the *KU70* locus, therefore disrupting its expression and increasing the efficiency of HR. A schematic figure for the entire genetic engineering process can be seen in figure 2.2.

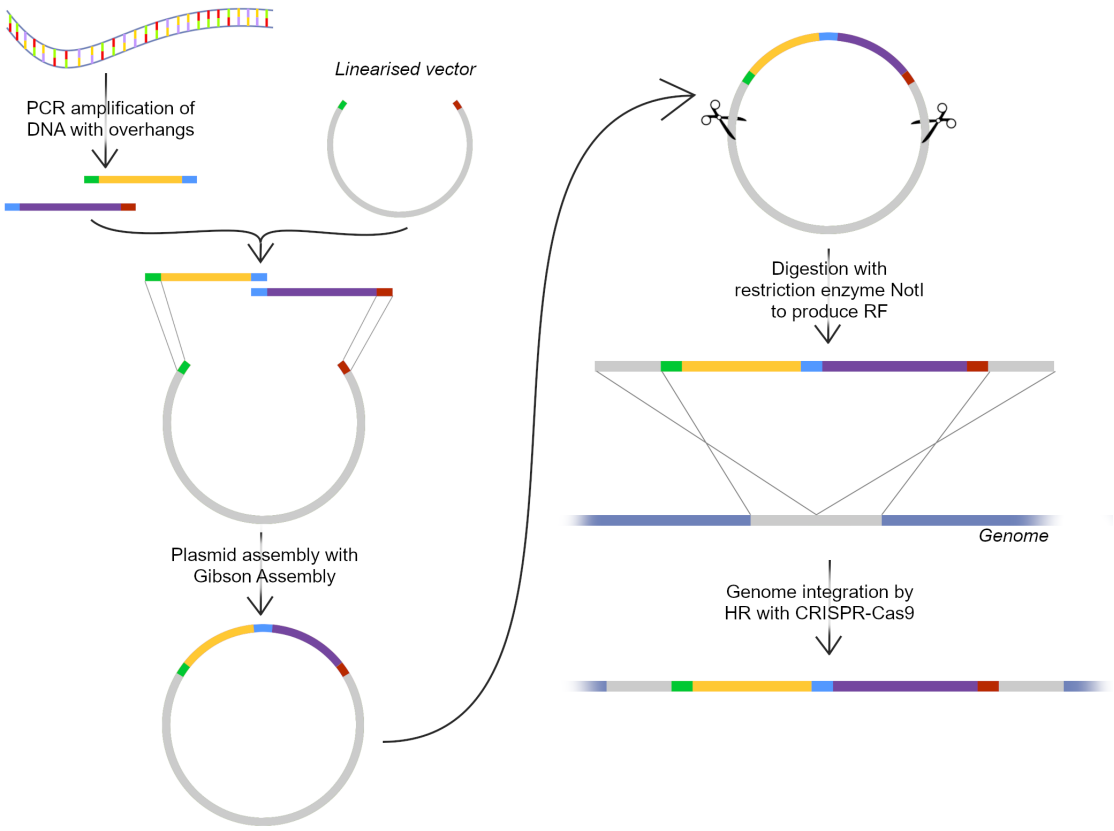


Figure 2.2: Overview of the genetic engineering process from DNA to genome integration. Abbreviations: CRISPR, clustered regularly interspaced short palindromic repeats; HR, homologous recombination; RF, repair fragment.

2.3 Polyketides

Natural products are a large family of secondary metabolites produced by a variety of organisms. Their biological activities have long been of value for humans within both medicine and agricultural areas [36]. One group of interest is polyketides, which are natural products of microbes and plants with valuable abilities such as anti-bacterial, anti-cancer, and anti-fungal [15][37]. Polyketides are biosynthesised by repeated decarboxylative condensation of extender units, such as malonyl-CoA and acetyl-CoA, onto a starting acyl substrate [38]. This is carried out by large, multifunctional enzymes called polyketide synthases (PKSs), of which there are three classes (type I-III). The different types of PKSs each produce a specific group of polyketides. Type III PKSs are small proteins that synthesise small aromatic polyketides [39].

The natural producers of polyketides only synthesise these compounds in very small amounts. Attempts have been made to produce polyketides in model host organisms, but with too low yields and titres for industrial production [40]. Moreover, chemical synthesis is difficult due to complex structures and isomer formation [41], and cannot meet the production demand [39]. These limitations could be overcome by production through heterologous expression of the PKSs in a non-conventional microbial host instead. The polyketide biosyntesis pathway resembles the fatty acid biosynthesis pathway in several aspects: PKSs and the FAS complex are similar in character, and both enzyme types use simple precursors in chain extension to form their respective product [37]. Since *Y. lipolytica* is well-suited to be a microbial cell factory and has the ability to accumulate high amounts of lipids, it makes a good candidate for producing polyketides. Several polyketides have been produced in the yeast. Markham et al. (2018) established a strain of *Y. lipolytica* that produced the simple polyketide triacetic acid lactone with an average maximum titre of 35.9 g/L [40]. Palmer et al. (2020) established biosynthesis of more advanced polyketides derived from 4-coumaroyl-CoA [42], which indicate the possibilities of producing a large variety of polyketides using *Y. lipolytica*.

2.4 Biosensors

A biosensor is a device used to detect and quantify the presence of a molecule of interest, by combining a biological component with a transducer. The biological element or bioreceptor, which is normally immobilised on a detector, can be cells or some type of organelle, enzymes, antibodies, receptors, etc. The bioreceptor interacts with the analyte in some way, which is called bio-recognition. The transducer then proportionally converts the bio-recognition to a measurable signal in one of many physicochemical ways, such as electrochemical, colorimetric, or optical [43]. Since both the biological element and the transducer can be varied in many ways, the possibilities of the structure of a biosensor are numerous, as well as the areas in which to employ it [44]. A biosensor has certain characteristics that need to be optimised for best possible performance. Usually, selectivity is the most important, so that only the analyte of interest is detected. Further, the sensitivity defines the lowest detectable concentration of the biosensor. This is vital in biosensors used in medical areas (e.g. to detect small concentrations of biomarkers). The stability of a biosensor describes its ability to withstand changes in the surrounding environment. pH and temperature varieties should not affect the performance. A biosensor should also be reproducible - have a high precision and accuracy - to give a reliable result for every use. The linearity or dose-dependency indicates how well the output signal correlates to the input. Moreover, it shows the resolution (change in input that gives a change in output) and the detectable analyte concentration range of the biosensor [43]. Finally, a biosensor should be tunable and modular for simple adaption of detection range and analyte [45].

A colorimetric biosensor for malonyl-CoA in three bacteria (*E. coli*, *Pseudomonas putida*, and *Corynebacterium glutamicum*) was developed by Yang et al. (2018) [11]. The purpose of this biosensor was to find gene targets that give an enhanced malonyl-CoA accumulation to improve production of malonyl-CoA-derived products in microorganisms. The authors repurposed a type III PKS called RppA that produces the red-coloured molecule flaviolin. A schematic description of the repurposed PKS as a biosensor is seen in figure 2.3. Five malonyl-CoA are converted to 1,3,6,8-tetra-hydroxynaphthalene by RppA, which is then spontaneously oxidised to flaviolin. Thus, the concentration of flaviolin correlates to the intracellular concentration of malonyl-CoA, which is otherwise difficult and time-consuming to quantify [11]. Due to the red colour of flaviolin, colorimetric screening was possible: the redder the colour, the more flaviolin is produced by the cell. In order to accurately compare the flaviolin level between strains, the authors tested the absorbance of the product using UV-visible spectroscopy. The λ_{max} of flaviolin is approximately 300 nm, but the absorbance background of the control strain was deemed too high for that wavelength. Therefore, the biosensor signal was defined as the absorbance at 340 nm, since the signal was sufficient enough and background noise was minimised.

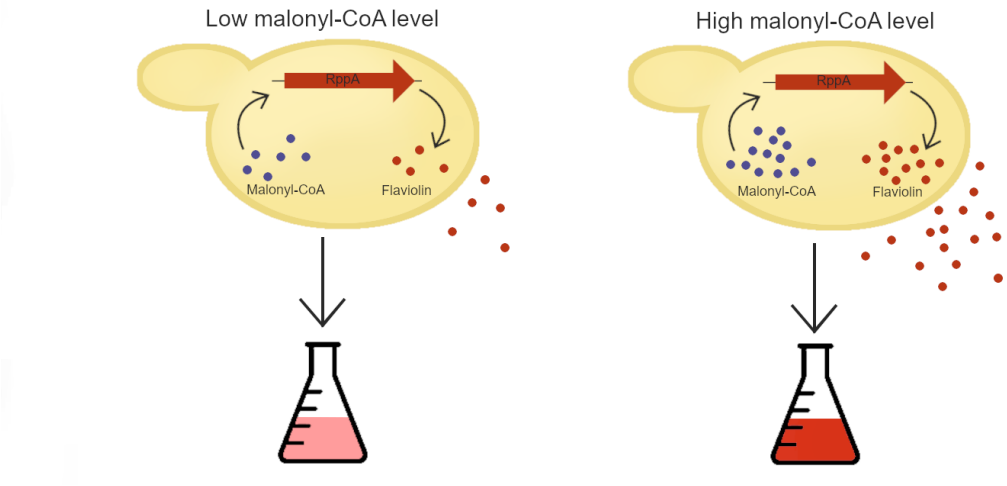


Figure 2.3: Schematic description of a colorimetric malonyl-CoA biosensor. The polyketide synthase RppA can be repurposed as a biosensor for malonyl-CoA, as it converts malonyl-CoA to red-coloured flaviolin. The biosensor can therefore be used to find strains with high flux through malonyl-CoA by screening for increased red colour.

For a biosensor to be functional, the response range needs to be dose-dependent [43]. This is measured by adding the target substrate in different concentrations and measuring the response. This approach was however not possible for Yang et al., since malonyl-CoA is an intracellular intermediate not transported over the cell membrane. An indirect method of controlling the malonyl-CoA concentration was used instead [11]. Cerulenin is an antibiotic inhibiting biosynthesis of fatty acids and sterols, with the minimal inhibitory concentration varying for different bacterial and yeast species [46]. The specific targets are beta-ketoacyl synthases I and II of the FAS complex, which elongate the fatty acid chain with malonyl-CoA [47]. Addition of cerulenin to growth medium increases the malonyl-CoA pool indirectly, and can therefore be used to increase the production of malonyl-CoA-derived products [48][49][50][51]. In the study by Yang et al. [11], cerulenin was added to the medium

in different concentrations. Indeed, with increasing cerulenin concentration, the output signal of the biosensor also increased. The signal was then normalised with cell growth (OD_{600}) since cerulenin affects the intracellular malonyl-CoA on a single-cell basis. After confirmation that the biosensor was dose-dependent, it was applied to a genome-scale synthetic small regulatory RNA (sRNA) library transformed into *E. coli* to identify gene knockdown targets for increasing malonyl-CoA concentration. With high-throughput screening, strains containing sRNAs with stronger signal or colour than the control strain were selected for further screening. From this, certain sRNAs associated with high increase in signal were selected as knockdown gene targets. To evaluate the original goal with the biosensor (finding targets for increasing production of malonyl-CoA-derived products), the gene targets were knocked down in strains producing two different polyketides. With a simple metabolic engineering strategy, the titre of one of the polyketides increased over 50 times. Further, three additional PKSs (AaOKS, AaPKS4, and AaPKS5, all octaketide synthases from the plant *Aloe arborescens*) which use malonyl-CoA as substrate were identified as other potential biosensors for malonyl-CoA. The polyketide(s) that these enzymes produce is unknown. The product is however made from malonyl-CoA and is red-coloured as flaviolin, meaning that the principle of the biosensor remains the same [11].

3
Methods

3.1 Strains of *Y. lipolytica*

This section presents name and genotype of the strains used in both projects in the thesis. All plasmids and primers used to construct the strains can be found in tables A.1 and A.2 in Appendix A.1. The DNA sequences of promoters, genes, and terminators used in the thesis can be found in table A.3 in Appendix A.1.

3.1.1 Biosensor project

15 strains of *Y. lipolytica* were used in the biosensor project, of which six were constructed in this thesis. All strains are derived from either OKYL029, OKYL049, or JFYL007. OKYL029 and OKYL049 were derived from the strain ST6512 [35], which in turn originated from the commonly used background strain CLIB89 (W29) [52]. JFYL007 was derived from OKYL029. In OKYL029, a gene essential for hyphae formation (*MHY1*) is deleted [53]. In OKYL049, *DGA1* is overexpressed to increase lipid accumulation. Simultaneously, a gene responsible for sterol esterification (*ARE1*) is deleted [17]. In JFYL007, the gene for sterol O-acyltransferase (*ARE1*) and three genes involved in TAG production (*DGA1*, *DGA2*, and *LRO1*) are deleted to hinder lipid accumulation. The deletions in the strain are the same as in a study by Beopoulos et al. (2012) [54], but the study in which JFYL007 was created is unpublished. All strains contain the promoter *TEF1* and the terminator *LIP2* (t*LIP2*), which is a native terminator from *Y. lipolytica* from an extracellular lipase [55]. SZYL009, SZYL010, and SZYL011 were derived from JFYL007. SZYL012, SZYL013, and SZYL014 were derived from OKYL029. HOYL01, HOYL02, and HOYL03 were derived from OKYL049. HOYL04 was derived from SZYL009, HOYL05 was derived from SZYL012, and HOYL06 was derived from HOYL01. SZYL009, SZYL012, HOYL01, HOYL04, HOYL05, and HOYL06 contain *AaOKS*. SZYL010, SZYL013, and HOYL02 contain *AaPKS4*. SZYL011, SZYL014, and HOYL03 contain *AaPKS5*. The nine already existing strains and the six constructed strains are presented in table 3.1.

Table 3.1: Strains used in the biosensor project of the thesis with their respective genotype.

Strain	Genotype	Reference
OKYL029	W29 *KU70*::Cas9-DsdA Δ*mhy1*	Konzock et al. [53]
OKYL049	W29 *KU70*::Cas9-DsdA E1::p*TEF1in-DGA1*-t*PEX20* Δ*are1* Δ*mhy1*	Konzock et al. [17]
JFYL007	OKYL029 Δ*are1* Δ*lro1* Δ*dga1* Δ*dga2*	(Unpublished)
SZYL009	JFYL007 D1::p*TEF1-AaOKS*-t*LIP2*	(Unpublished)
SZYL010	JFYL007 D1::p*TEF1-AaPKS4*-t*LIP2*	(Unpublished)
SZYL011	JFYL007 D1::p*TEF1-AaPKS5*-t*LIP2*	(Unpublished)
SZYL012	OKYL029 D1::p*TEF1-AaOKS*-t*LIP2*	(Unpublished)
SZYL013	OKYL029 D1::p*TEF1-AaPKS4*-t*LIP2*	(Unpublished)
SZYL014	OKYL029 D1::p*TEF1-AaPKS5*-t*LIP2*	(Unpublished)
HOYL01	OKYL049 D1::p*TEF1-AaOKS*-t*LIP2*	This study
HOYL02	OKYL049 D1::p*TEF1-AaPKS4*-t*LIP2*	This study
HOYL03	OKYL49 D1::p*TEF1-AaPKS5*-t*LIP2*	This study
HOYL04	SZYL009 C3::p*TEF1-AaOKS*-t*LIP2*	This study
HOYL05	SZYL012 C3::p*TEF1-AaOKS*-t*LIP2*	This study
HOYL06	HOYL01 C3::p*TEF1-AaOKS*-t*LIP2*	This study

3.1.2 Promoter project

In the promoter project, the established and characterised constitutive promoter p*TEF1* was selected as positive control. Since all 27 selected promoters were uncharacterised, they were defined as the 1000 bp upstream region of the respective gene. Several promoters were bidirectional, or contained intron regions or other genes. Because of this, some of the promoters were shortened. 28 strains of *Y. lipolytica* were constructed. The first round of strain construction included 24 native, uncharacterised promoters from *Y. lipolytica* and the promoter p*TEF1* as positive control (three of the 27 selected promoters could not be established in a strain). They were all derived from OKYL029, with an inserted gene cassette in integration site D1. Each cassette consisted of a promoter from the *Y. lipolytica* genome (with the exception of strain SZYL055, which also had the terminator *PEX20* (t*PEX20*) before the promoter, in the opposite reading direction), the gene for the fluorescent protein human recombinant GFP (*hrGFP*), and the terminator *LIP2*. In the second round of strain construction, three strains containing either p*TEF1*, the native promoter for gene *PYK1* (p*PYK1*), or the native promoter for gene *GAPDH* (p*GAPDH*) were constructed in the same way as in the first round. The name of the strains and their respective genotype can be found in table 3.2.

Table 3.2: Strains used in the promoter project of the thesis with their respective genotype.

Strain	Genotype	Reference
OKYL029	W29 *KU70*::Cas9-DsdA Δ*mhy1*	Konzock et al. [53]
SZYL031	OKYL029 D1::p*YALI1_A20484g-hrGFP*-t*LIP2*	This study
SZYL032	OKYL029 D1::p*YALI1_B05639g-hrGFP*-t*LIP2*	This study
SZYL033	OKYL029 D1::p*YALI1_C01505g-hrGFP*-t*LIP2*	This study
SZYL034	OKYL029 D1::p*YALI1_C13611g-hrGFP*-t*LIP2*	This study
SZYL035	OKYL029 D1::p*YALI1_C32392g-hrGFP*-t*LIP2*	This study
SZYL036	OKYL029 D1::p*YALI1_D15557g-hrGFP*-t*LIP2*	This study
SZYL037	OKYL029 D1::p*YALI1_E01694g-hrGFP*-t*LIP2*	This study
SZYL038	OKYL029 D1::p*YALI1_E28070g-hrGFP*-t*LIP2*	This study
SZYL039	OKYL029 D1::p*YALI1_E31154g-hrGFP*-t*LIP2*	This study
SZYL040	OKYL029 D1::p*YALI1_E32241g-hrGFP*-t*LIP2*	This study
SZYL041	OKYL029 D1::p*YALI1_F29596g-hrGFP*-t*LIP2*	This study
SZYL042	OKYL029 D1::p*YALI1_A00911g-hrGFP*-t*LIP2*	This study
SZYL043	OKYL029 D1::p*YALI1_A21509g-hrGFP*-t*LIP2*	This study
SZYL044	OKYL029 D1::p*YALI1_C02160g-hrGFP*-t*LIP2*	This study
SZYL045	OKYL029 D1::p*YALI1_C24183g-hrGFP*-t*LIP2*	This study
SZYL046	OKYL029 D1::p*YALI1_C26533g-hrGFP*-t*LIP2*	This study
SZYL047	OKYL029 D1::p*YALI1_C30910g-hrGFP*-t*LIP2*	This study
SZYL048	OKYL029 D1::p*YALI1_D01497g-hrGFP*-t*LIP2*	This study
SZYL049	OKYL029 D1::p*YALI1_E02850g-hrGFP*-t*LIP2*	This study
SZYL050	OKYL029 D1::p*YALI1_F02403g-hrGFP*-t*LIP2*	This study
SZYL051	OKYL029 D1::p*YALI1_C01520g-hrGFP*-t*LIP2*	This study
SZYL052	OKYL029 D1::p*YALI1_E05779g-hrGFP*-t*LIP2*	This study
SZYL053	OKYL029 D1::p*YALI1_E09895g-hrGFP*-t*LIP2*	This study
SZYL054	OKYL029 D1::p*YALI1_E32180g-hrGFP*-t*LIP2*	This study
SZYL055	OKYL029 D1::t*PEX20*-pTEF1-*hrGFP*-t*LIP2*	This study
SZYL056	OKYL029 D1::p*TEF1-hrGFP*-t*LIP2*	This study
SZYL057	OKYL029 D1::p*PYK1-hrGFP*-t*LIP2*	This study
SZYL058	OKYL029 D1::p*GAPDH-hrGFP*-t*LIP2*	This study

3.2 Media

Delft medium for chemostats with carbon-to-nitrogen ratio (C/N) 3 contained 5.28 g/L ammonium sulphate or 2.4 g/L urea, 3.0 g/L monopotassium phosphate, 0.5 g/L magnesium sulphate heptahydrate, 1.0 mL trace metals, 1.0 mL vitamin solution, and 7.2 g/L glucose. pH was set to 5.

Delft medium for chemostats with C/N 116 contained contained 0.471 g/L am-

monium sulphate or 0.213 g/L urea, 3.0 g/L monopotassium phosphate, 0.5 g/L magnesium sulphate heptahydrate, 1.0 mL trace metals, 1.0 mL vitamin solution, and 25.0 g/L glucose. pH was set to 5.

Delft medium for shake flasks with C/N 10 contained 7.5 g/L ammonium sulphate, 14.4 g/L magnesium sulphate heptahydrate, 0.5 g/L potassium dihydrogen phosphate, 1.0 mL trace metals, 1.0 mL vitamin solution, and 34.1 g/L glucose. pH was set to 5.

Delft medium for shake flasks with C/N 100 contained 0.44 g/L ammonium sulphate, 0.5 g/L magnesium sulphate heptahydrate, 19.1 g/L potassium dihydrogen phosphate, 10.4 g/L dipotassium phosphate, 2.0 mL trace metals, 1.0 mL vitamin solution, and 20.0 g/L glucose. pH was set to 6.5.

LB medium contained 10.0 g/L peptone from casein, 10.0 g/L sodium chloride, and 5.0 g/L yeast extract. pH was set to 7.

LB+ampicillin (LB+Amp) medium was LB medium with 100 mg/L ampicillin.

LB plates contained LB medium and 16.0 g/L agar agar.

LB+Amp plates were LB plates with 100 mg/L ampicillin.

Novogy medium contained 0.5 g/L urea, 1.5 g/L yeast extract, 0.85 g/L casamino acids, 1.7 g/L yeast nitrogen base (YNB) without amino acids and ammonium sulphate, 5.1 g/L potassium hydrogen phtalate, and 100.0 g/L glucose. pH was set to 5.5.

YNB minimal medium contained 1.7 g/L YNB, 1.5 g/L ammonium chloride, 3.5 g/L disodium hydrogen phosphate, 3.5 g/L sodium phosphate monobasic monohydrate, and 20.0 g/L glucose.

YPD medium contained 20.0 g/L peptone from meat, 10.0 g/L yeast extract, and 20.0 g/L glucose.

YPD plates contained YPD medium and 20.0 g/L agar agar.

YPD+nourseothricin (YPD+Nat) plates were YPD plates with 250 mg/L nourseothricin.

Trace metal solution contained 3.0 g/L iron(II) sulphate heptahydrate, 4.5 g/L zinc sulphate heptahydrate, 4.5 g/L calcium chloride dihydrate, 0.84 g/L manganese(II) chloride dihydrate, 0.3 g/L cobalt(II) chloride hexahydrate, 0.3 g/L copper(II) sulphate pentahydrate, 0.4 g/L sodium molybdate dihydrate, 1.0 g/L boric acid, 0.1 g/L potassium iodide, and 19.0 g/L EDTA disodium salt dihydrate.

Vitamin solution contained 0.05 g/L D-biotin, 1.0 g/L D-pantothenic acid hemicalcium salt, 1.0 g/L thiamin hydrochloride, 1.0 g/L pyridoxine hydrochloride, 1.0 g/L nicotinic acid, 0.2 g/L 4-aminobenzioc acid, and 25.0 g/L myo-inositol.

3.3 PCR

Polymerase chain reaction (PCR) was used to amplify DNA. All PCR reactions were performed in S1000™ Thermal Cycler (Bio-Rad). Several different DNA polymerases were used in the thesis, each with a specific protocol presented below. Annealing temperature was decided based on the primers used. Extension time was decided based on the length of the fragment to amplify.

3.3.1 SapphireAmp® (Takara Bio)

The general reaction mixture contained 25 µL SapphireAmp® Fast PCR Master Mix (2X Premix), 0.2 µM (final concentration) forward primer, 0.2 µM (final concentration) reverse primer, 100 pg-10 ng plasmid DNA template or 50-100 ng genomic DNA template, and milliQ water up to 50 µL. The volumes were added to a PCR tube in descending order of volume.

PCR conditions were as follows: 1 cycle of initial denaturation at 98° C for 3 min, 35 cycles of denaturation at 98° C for 10 s, annealing at annealing temperature for 15 s and elongation at 72° C for 10 s/kb, and 1 cycle of final elongation at 72° C for 3 min.

3.3.2 Phusion High Fidelity (Thermo Scientific)

The general reaction mixture contained 10 µL 5X Phusion HF buffer, 1 µL 10 mM dNTPs, 0.5 µM (final concentration) forward primer, 0.5 µM (final concentration) reverse primer, 1 pg-10 ng plasmid DNA template or 50-250 ng genomic DNA template, 0.5 µL Phusion DNA polymerase, and milliQ water up to 50 µL. The volumes were added to a PCR tube in descending order of volume.

PCR conditions were as follows: 1 cycle of initial denaturation at 98° C for 30 s, 35 cycles of denaturation at 98° C for 10 s, annealing at annealing temperature for 15 s and elongation at 72° C for 15-30 s/kb, and 1 cycle of final elongation at 72° C for 5 min.

3.3.3 PrimeSTAR® HS (Takara Bio)

The general reaction mixture contained 10 µL 5X PrimeSTAR buffer (Mg^{2+} Plus), 1 µL 10 mM dNTPs, 0.2 µM (final concentration) forward primer, 0.2 µM (final concentration) reverse primer, 10 pg-1 ng plasmid DNA template or 5-200 ng genomic DNA template, 0.5 µL PrimeSTAR HS DNA polymerase, and milliQ water up to 50 µL. The volumes were added to a PCR tube in descending order of volume.

PCR conditions were as follows: 1 cycle of initial denaturation at 98° C for 1 min, 35 cycles of denaturation at 98° C for 10 s, annealing at annealing temperature for 15 s and elongation at 72° C for 1 min/kb, and 1 cycle of final elongation at 72° C for 3 min.

3.3.4 Herculase II Fusion (Agilent)

The general reaction mixture contained 10 µL 5X Herculase II reaction buffer, 0.5 µL 10 mM dNTPs, 0.25 µM (final concentration) forward primer, 0.25 µM (final concentration) reverse primer, 1-30 ng plasmid DNA template or 100-400 ng genomic DNA template, 1 µL Herculase II Fusion DNA polymerase, and milliQ water up to 50 µL. The volumes were added to a PCR tube in descending order of volume.

PCR conditions were as follows: 1 cycle of initial denaturation at 95° C for 2 min, 35 cycles of denaturation at 95° C for 10 s, annealing at annealing temperature for 20 s and elongation at 72° C for 30 s/kb, and 1 cycle of final elongation at 72° C for 3 min.

3.4 Gel electrophoresis

To visualise PCR products, gel electrophoresis was used. For pre-staining of gel, 25 mL 1% agarose solution was mixed with 1 µL GelRed Nucleic Acid Stain (Biotium) and allowed to polymerise for 30-60 min. The PCR products were loaded on the gel along with GeneRuler 1 kb DNA Ladder (Thermo Fisher) and run for 20-40 min at 80-120 V. For post-staining of gel, 25 mL 1% agarose solution was allowed to polymerise for 30-60 min. The PCR products were loaded on the gel along with GeneRuler 1 kb DNA Ladder (Thermo Fisher) and run for 20-40 min at 80-120 V. The gel was then incubated for 45 min in a solution of 50 mL milliQ water mixed with 2 µL GelGreen Nucleic Acid Stain (Biotium).

3.5 DNA purification

To purify DNA from different sources, three different GeneJET kits from Thermo Scientific were used, each with a specific protocol presented below.

3.5.1 GeneJET Plasmid Miniprep Kit

To extract and purify plasmids from recombinant *E. coli* strains, GeneJET Plasmid Miniprep Kit (Thermo Scientific) was used. The 15 mL Falcon tube with the overnight culture was centrifuged at 5100 rpm for 5 min to pellet the cells. The pelleted cells were resuspended in 250 µL Resuspension Solution and transferred to an Eppendorf tube. 250 µL Lysis Solution was added, followed mixing by inverting the tube six times. 350 µL Neutralization Solution was added, followed by mixing by inverting the tube six times. The tube was centrifuged at 12000 rpm for 5 min. The supernatant was transferred to a GeneJET spin column and centrifuged at 12000

rpm for 1 min. The flow-through in the collection tube was discarded. 500 µL Wash Solution was added to the column, followed by centrifugation at 12000 rpm for 1 min. The flow-through was discarded. The wash step was repeated. The column and collection tube was centrifuged for 1 additional minute to remove residual Wash Solution. The column was moved to an Eppendorf tube and 30-50 µL Elution Buffer was added to the center of the column membrane. The tube was incubated at room temperature for 2 min (or up to 60 min) followed by centrifugation at 12000 rpm for 2 min. The concentration of the purified plasmid DNA was measured with NanoDrop™ 2000 Spectrophotometer (Thermo Scientific).

3.5.2 GeneJET PCR Purification Kit

To purify DNA products after PCR amplification, GeneJET PCR Purification Kit (Thermo Scientific) was used. A 1:1 volume of Binding Buffer was added to the completed PCR mixture. The solution was transferred to a GeneJET purification column and centrifuged at 12000 rpm for 1 min. The flow-through in the collection tube was discarded. 700 µL Wash Buffer was added to the column, followed by centrifugation at 12000 rpm for 1 min. The flow-through was discarded. The column and collection tube was centrifuged for 1 additional minute to remove residual Wash Buffer. The column was moved to an Eppendorf tube and 30-50 µL Elution Buffer was added to the centre of the column membrane. The tube was incubated at room temperature for 2 min (or up to 60 min) followed by centrifugation at 12000 rpm for 2 min. The concentration of the purified plasmid DNA was measured with NanoDrop™ 2000 Spectrophotometer (Thermo Scientific).

3.5.3 GeneJET Gel Extraction Kit

To purify DNA products from agarose gels, GeneJET Gel Extraction Kit (Thermo Scientific) was used. The specific band was excised from the agarose gel using a scalpel and placed in an Eppendorf tube. The weight of the gel slice was measured. A 1:1 volume (volume (µL):weight (mg)) of Binding Buffer was added to the gel slice. The tube was incubated at 60° C for 10 min or until the gel was completely dissolved. The solution was transferred to a GeneJET purification column and centrifuged at 12000 rpm for 1 min. The flow-through in the collection tube was discarded. 700 µL Wash Buffer was added to the column, followed by centrifugation at 12000 rpm for 1 min. The flow-through was discarded. The column and collection tube was centrifuged for 1 additional minute to remove residual Wash Buffer. The column was moved to an Eppendorf tube and 30-50 µL Elution Buffer was added to the centre of the column membrane. The tube was incubated at room temperature for 2 min (or up to 60 min) followed by centrifugation at 12000 rpm for 2 min. The concentration of the purified plasmid DNA was measured with NanoDrop™ 2000 Spectrophotometer (Thermo Scientific).

3.6 Gibson assembly

Plasmids were constructed using the Gibson assembly method. The vector backbone was PCR amplified using Phusion High Fidelity DNA polymerase (Thermo Scientific) and primers with overhangs matching the insert. The vector backbone was then purified from agarose gel using GeneJET Gel Extraction Kit. The insert was PCR amplified using either Phusion High Fidelity (Thermo Scientific), PrimeSTAR (Takara), or Herculase (Agilent) and primers with overhangs matching the vector backbone. The insert was then purified using GeneJET PCR Purification Kit. The length of the insert was verified using gel electrophoresis. 5 µL Gibson Assembly Master Mix (2X) (NEB), 50-100 ng vector backbone, 3-5 fold molar excess insert, and milliQ water to 5 µL were mixed and incubated at 50°C for 30 min before transforming into *E. coli* competent cells.

3.7 *E. coli* transformation

Chemically competent *E. coli* DH5α were completely thawed on ice. 2 µL Gibson assembly mix was added to the cells and the contents were gently mixed, after which they were incubated on ice for 30 min. Next, the cells were heat-shocked for 90 s at 42°C and then incubated on ice for 2 min. 500 µL LB medium was added to the cells, which were incubated for 60 min at 37° while shaking at 180 rpm. 75 µL cell mixture was plated on an LB+Amp plate, which was incubated at 37°C for 16 h.

To screen for correctly constructed plasmids, single colonies were picked from the transformation plate using sterile pipette tips. The tips were placed in 5 µL of MilliQ water, of which 1 µL was used as template in a SapphireAmp PCR with vector-specific primers. Clones with a PCR product of expected length were used to inoculate 5 mL LB+Amp medium in a 15 mL Falcon tube and incubated at 37°C while shaking at 180 rpm for 16 h. Plasmid extraction and purification was done with GeneJET Plasmid Miniprep Kit (Thermo Scientific), followed by sequencing at Eurofins Genomics. Correctly constructed transformants were used to inoculate 5 mL LB+Amp medium in a 15 mL Falcon tube, which was incubated at 37°C while shaking at 180 rpm for 16 h. 900 µL cell culture and 300 µL 80% glycerol were mixed and stored at -80°C to preserve the constructed strain.

3.8 *Y. lipolytica* transformation

Selection marker-free transformation of *Y. lipolytica* was performed using CRISPR/-Cas9, with NotI digested repair fragment and gRNA plasmid chosen according to the EasyCloneYALI toolbox [34]. For D1 locus, the integration vector was pCfB6684 and the gRNA vector was pCfB6631. For C3 locus, the integration vector was pCfB6371 and the gRNA vector was pCfB6630. The *Y. lipolytica* parent strain was streaked on a YPD plate, followed by incubation in 30°C for 18 h. The cells were detached from the plate by washing it with 2 mL milliQ water. The suspension was collected in an Eppendorf tube and centrifuged for 5 min at 3000 rpm. The supernatant was

discarded and the cell pellet was resuspended in 1 mL milliQ water, followed by centrifugation for 5 min at 3000 rpm. Again, the supernatant was discarded and the cell pellet was resuspended in 1 mL milliQ water. The OD_{600} was measured. 1 OD unit was centrifuged for 5 min at 3000 rpm and the supernatant was discarded. 500 ng repair fragment and 300 ng gRNA plasmid were used to resuspend the cell pellet, followed by addition of 100 µL transformation mix (43.8% sterile-filtrated PEG, 0.1 M sterile-filtrated lithium acetate, 0.25 g/L salmon sperm ssDNA, and 100 mM sterile-filtrated dithiothreitol). Next, the cell mixture was incubated at 39°C for 60 min, followed by centrifugation for 5 min at 3000 rpm. The supernatant was discarded and the cell pellet was resuspended in 500 µL YPD medium and incubated at 30°C while shaking at 200 rpm for 2 h. Next, the cells were centrifuged for 5 min at 3000 rpm and the supernatant was discarded. The cell pellet was resuspended in 100 µL milliQ water and plated on a YPD+Nat plate.

To screen for correctly constructed plasmids, single colonies were picked from the transformation plate using sterile pipette tips. The tips were placed in 10 µL of 20 mM sodium hydroxide. The solution was boiled at 98°C for 10 min, followed by 1 min centrifugation. 1 µL of the supernatant was used as template in a SapphireAmp PCR with integration site-specific primers. Clones with a PCR product of expected length were used to inoculate 8 mL YPD medium in a 50 mL Falcon tube and incubated at 30°C while shaking at 200 rpm for 16 h. One loopful of cell culture was streaked on a YPD plate and incubated at 30°C for 24 h. Single colonies were picked with sterile pipette tips and placed on first a YPD plate, then a YPD+Nat plate. Colonies that survived on YPD but not YPD+Nat had lost the gRNA plasmid and were screened using colony PCR again. Positive transformants were used to inoculate 8 mL YPD medium in a 50 mL Falcon tube, which was incubated at 30°C while shaking at 200 rpm for 16 h. 900 µL cell culture and 300 µL 80% glycerol were mixed and stored at -80°C to preserve the constructed strain.

3.9 Shake flasks

The strains to be used were streaked on YPD plates. One loopful of yeast cells were used to inoculate 5 mL YPD medium in a 50 mL Falcon tube, which was incubated at 30°C while shaking at 200 rpm for 16 h. The OD_{600} was measured. 10 mL medium was added to 100 mL shake flasks, followed by inoculation with overnight preculture at starting OD_{600} 0.1. Shake flasks were incubated at 30°C while shaking at 220 rpm for 40-144 h. During sampling, the flasks were quickly removed from the cultivation conditions. This brief interruption was considered to have no effect on the cultivation. To sample, the OD_{600} was measured and 300-1000 µL of cultivation liquid was pipetted into a 1.5 mL Eppendorf tube, followed by centrifugation for 5 min at 3000 rpm. To measure the absorbance, 50-200 µL of the supernatant was transferred to a 96-well microplate (Sarstedt). Dilutions were made with milliQ water. The absorbance in the UV-visible spectrum was measured using FLUOstar® Omega microplate reader (BMG LABTECH) and the absorbance was normalised to OD_{600}. Blank was milliQ water. For shake flask experiments with cerulenin, supplementation for a final concentration of 1 mg/L from a stock (10 mM cerulenin

in ethanol) occurred after 40 h.

The absorbance spectra were analysed using RStudio. Negative absorbance values were set to 0. Normalisation was done by dividing the absorbance with the measured OD_{600} to acquire a single-cell-based signal.

3.10 Biolector

The microbioreactor system BioLector® I (m2p-labs) allows for high-throughput fermentation with online monitoring of several parameters. The strains to be tested in the BioLector were streaked on YPD plates. One loopful of yeast cells were used to inoculate 8 mL YPD medium in a 50 mL Falcon tube, which was incubated at 30°C while shaking at 200 rpm for 16 h. The OD_{600} was measured. 5 OD units were centrifuged for 5 min at 3000 rpm and the supernatant was discarded. The cell pellet was washed with 1 mL milliQ water and centrifuged for 5 min at 3000 rpm. The supernatant was discarded and the cell pellet was resuspended in 1 mL milliQ water. OD_{600} was remeasured and the resuspended cell solution was used to inoculate 1 mL of medium with starting OD_{600} 0.1. 200 µL was added to each well in a 96-well CELLSTAR microplate (Greiner Bio-one), followed by the placement of a sterile AeraSeal™ film (Excel Scientific) on the plate. The biolector shaking speed was set to 995 rpm, the temperature was set to 28°C, and humidity control was on. Biomass (scattered light) was measured with gain 20 and GFP (fluorescence intensity, $\lambda_{excitation} = 488$ nm, $\lambda_{emission} = 520$ nm) was measured at gains 20, 40, 60, 80, and 100. Measurements were taken every 30 min. Runs were ended after 48-72 h.

The data from the BioLector was analysed using RStudio. The cell auto-fluorescence was removed by subtracting the fluorescence intensity from the negative control (OKYL029) from each strain's fluorescence intensity. Normalisation was done by dividing the fluorescence intensity with the the biomass (scattered light) to acquire a single-cell-based fluorescence intensity.

3.11 pH experiment

The pH of the media to be tested (Novogy, Delft for chemostats with urea or ammonium sulphate, C/N 3 or 116) was measured. The strains to be cultivated were streaked on YPD plates. One loopful of yeast cells were used to inoculate 8 mL YPD medium in a 50 mL Falcon tube, which was incubated at 30 °C while shaking at 200 rpm for 16 h. The OD_{600} was measured. 5 mL of medium was added to 100 mL shake flasks, followed by inoculation with overnight preculture at OD_{600} 0.1. Shake flasks were incubated at 30°C while shaking at 220 rpm for 48 h. 2 mL of culture was placed into a 2 mL Eppendorf tube and centrifuged for 5 min at 5000 rpm. The pH of the supernatant was measured.

3.12 Fluorescence microscopy

Fluorescence microscopy was used to investigate the presence of GFP in cells. Cell cultures were diluted 10 times. 10 µL was placed on a microscope slide and covered with a cover glass. Photos were taken of the cells with 100X magnification using brightfield and GFP channel in microscope Leica DMI4000 B (Leica Microsystems) with camera Leica DFC360 FX (Leica microsystems). Settings were adjusted for each set of images to best accommodate the specific experiment. For time courses the same settings were used at each time point. The images were edited in the microscope software platform LAS X Life Science (Leica Microsystems).

For investigating the strains SZYL055-58 in the fluorescence microscope, the strains were cultivated in shake flasks (see section 3.9). At time points 24 h, 48 h, 72 h, and 96 h, 20 µL was taken from each shake flask. The microscope slides with the samples were then prepared as described earlier.

4

Results and discussion

4.1 Biosensor project

Previously, the three PKSs AaOKS, AaPKS4, and AaPKS5 had been transformed into *Y. lipolytica* strain backgrounds OKYL029 and JFYL007 to form strains SZYL009-014 (unpublished). The strains, along with the OKYL029 and JFYL007 as negative controls, were cultivated in the complex medium Novogy medium for 120 h. The absorbance in the UV-visible spectrum of the culture supernatant was measured and normalised to OD_{600}. For the OKYL029-derived strains, a difference in absorbance, especially around the wavelengths 280-300 nm, could be seen between the control and the test strains. In the JFYL007-derived strains, there was absorbance in the same wavelength region, but no difference could be seen between the control and test strains. Further, the culture supernatant from the test strains had a yellow-red colour, which could not be seen for the controls (unpublished). The colour change and the absorbance difference between control and test strains showed a proof-of-concept of the enzymes working as biosensors in the yeast. The theory for the large amount of absorbance seen for the control strains was that the growth medium had too much background absorbance. A possible solution would be to blank the spectrophotometer with medium instead of water.

4.1.1 Evaluation of biosensor in different media

To investigate the signal strength of the eight strains SZYL009-014, OKYL029, and JFYL007 in different media, a new experiment was set up. The strains were grown in the two defined media Delft for shake flasks C/N 10 and Delft for shake flasks C/N 100. For comparison, Novogy medium was also included in the experiment. The hypothesis was that the Delft media would have less background absorbance than the Novogy medium. Further, by comparing two different C/N ratios for the Delft medium, the dose-dependency of the biosensor enzymes could be tested. It was expected that the nitrogen limitation in Delft C/N 100 would activate the fatty acid synthesis pathway, thus increasing the flux through malonyl-CoA and resulting in an increased product formation by the biosensor enzymes. As JFYL007 cannot accumulate lipids, it was also hypothesised that the JFYL007 strains would have a higher signal than the OKYL029 strains. After 144 h of cultivation in shake flasks, the absorbance in the UV-visible spectrum of the culture supernatants was measured. The spectra are shown in figure 4.1.

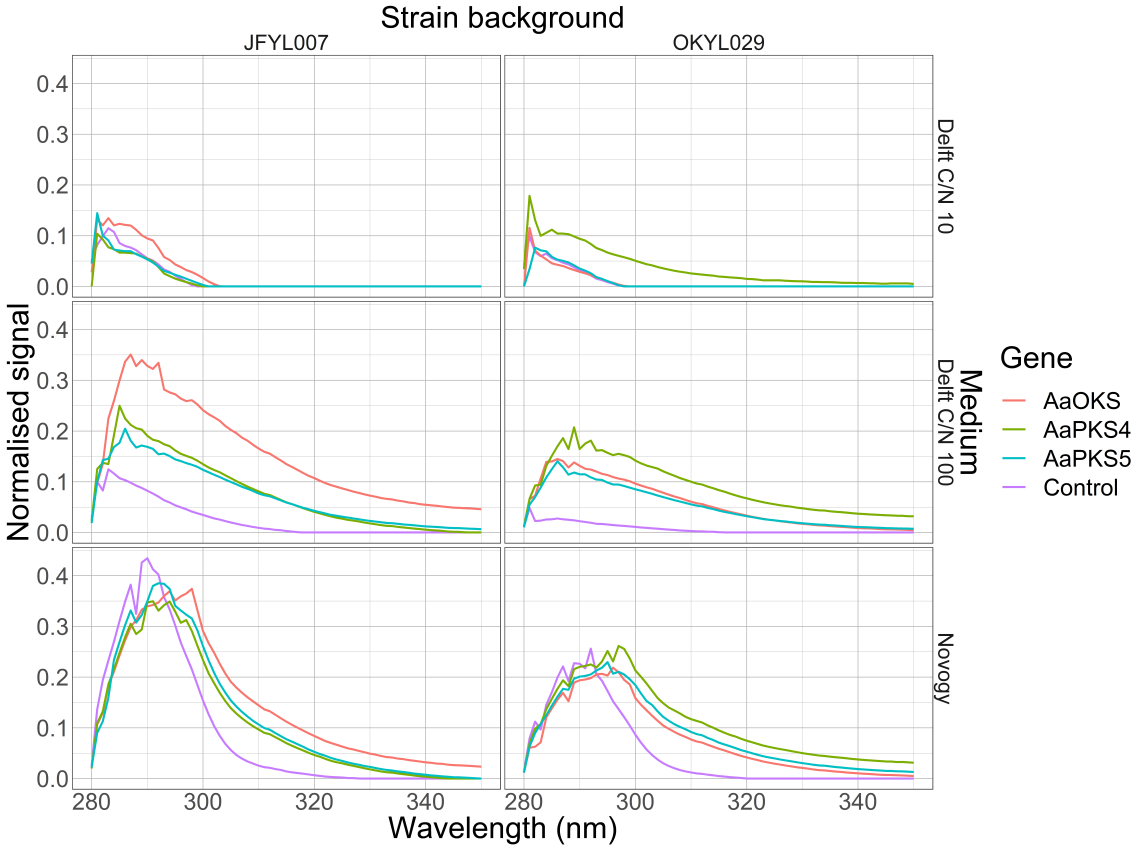


Figure 4.1: Absorbance of culture supernatant of strains SZYL009-014, OKYL029, and JFYL007. Strains SZYL009 and SZYL012 express AaOKS, strains SZYL010 and SZYL013 express AaPKS4, and strains SZYL011 and SZYL014 express AaPKS5. SZYL009-011 are derived from JFYL007 and SZYL012-14 are derived from OKYL029. JFYL007 and OKYL029 are controls. The strains were grown in either Delft C/N 10, Delft C/N 100, or Novogy medium for 144 h. The graphs are divided by medium and strain background. The line colour represents inserted gene.

The absorbance was lowest in Delft C/N 10 for both strain backgrounds. The control strains had approximately the same absorbance as the test strains in Delft C/N 10 (figure 4.1). This is probably explained by a very low product formation due to nutrient limitation, which would make it hard to compare signal strength between strains. Therefore, this medium was excluded from future experiments.

In Delft C/N 100, there was a clear difference between the absorbance in the control versus the test sample. The AaOKS enzyme strain (red) gave the highest signal among the JFYL007 strains, while the highest signal among the OKYL029 strains was seen in the strain with the AaPKS4 enzyme (green) (figure 4.1).

In the Novogy medium, the absorbance for the control strains was as high as for the strains with polyketide synthases in the range 280-290 nm (figure 4.1). This had only been observed for JFYL007-derived strains in the earlier experiment before the thesis (unpublished). To test if the background fluorescence could be minimised, an additional measurement with fresh media used as blank for all samples was taken. For Delft C/N 10, all strains had negative absorbance after blank correction, but the general appearance of the spectrum was the same as when the blank was water. For

Delft C/N 100 and Novogy, the control strains and the strains derived from JFYL007 with AaPKS4 and AaPKS5 (SZYL010 and SZYL011) had negative absorbance after blank correction (spectra not shown). The negative absorbance could be due to that the media used as blank was fresh, therefore containing all nutrients. At the time of the sampling, the cells had consumed these nutrients. The tested supernatants would therefore absorb less light than the blank and after blank correction, this would appear as negative absorbance. A true blank would therefore be the supernatant of the corresponding control strain, which is equivalent to comparing the net absorbance between the controls and the test samples. Therefore, regardless of the high background absorbance, the strains grown in the Novogy medium did not have a high net absorbance. Thus, it was also excluded from further experiments, which only included Delft C/N 100 medium.

When comparing the OKYL029-based strains to the JFYL007-based strains, it was clear that the JFYL007-based strains generally had a higher signal in all media, as seen in figure 4.1. This was expected since the JFYL007 strains should have more malonyl-CoA available for product formation, especially in the nitrogen-limited Delft C/N 100 medium. Moreover, a clear difference could be seen in signal strength between C/N ratios, which could indicate that the biosensor is dose-dependent for malonyl-CoA.

4.1.2 Variation in biosensor signal over time

The pathway for the product is shared with the lipid biosynthesis pathway. Since the lipids accumulate over time, it is reasonable to believe that the reporter product concentration also increases after a certain amount of days. To investigate the production rate of the product, a time course experiment was set up with technical duplicates of the strains SZYL009-014, OKYL029, and JFYL007. Samples were taken at 72 h, 96 h, 120 h, and 144 h. The spectra for the experiment is presented in figure 4.2. After 72 h, the peak at around 280-320 nm was low for all strains (except control, for which it was very low). For the remaining time points (96 h, 120 h, and 144 h), the peak in the spectrum was both higher and sharper for all strains. The spectra for the controls had the same shape but lower absorbance. Compared to 96 h and 120 h, the absorbance at 144 h was decreased (not observed in control). This could indicate that the cells are metabolising the products, or it self-degrades. This was not investigated further, but is an interesting note about the product molecule itself. From this experiment, it was decided to evaluate future experiments at 96 h. Since the results from 96 h and onward were very similar, it is unnecessary and time inefficient to continue the cultivation from a biosensor perspective. The time difference might be small in low-throughput screening conditions like this experiment, but will have a greater impact in high-throughput screening.

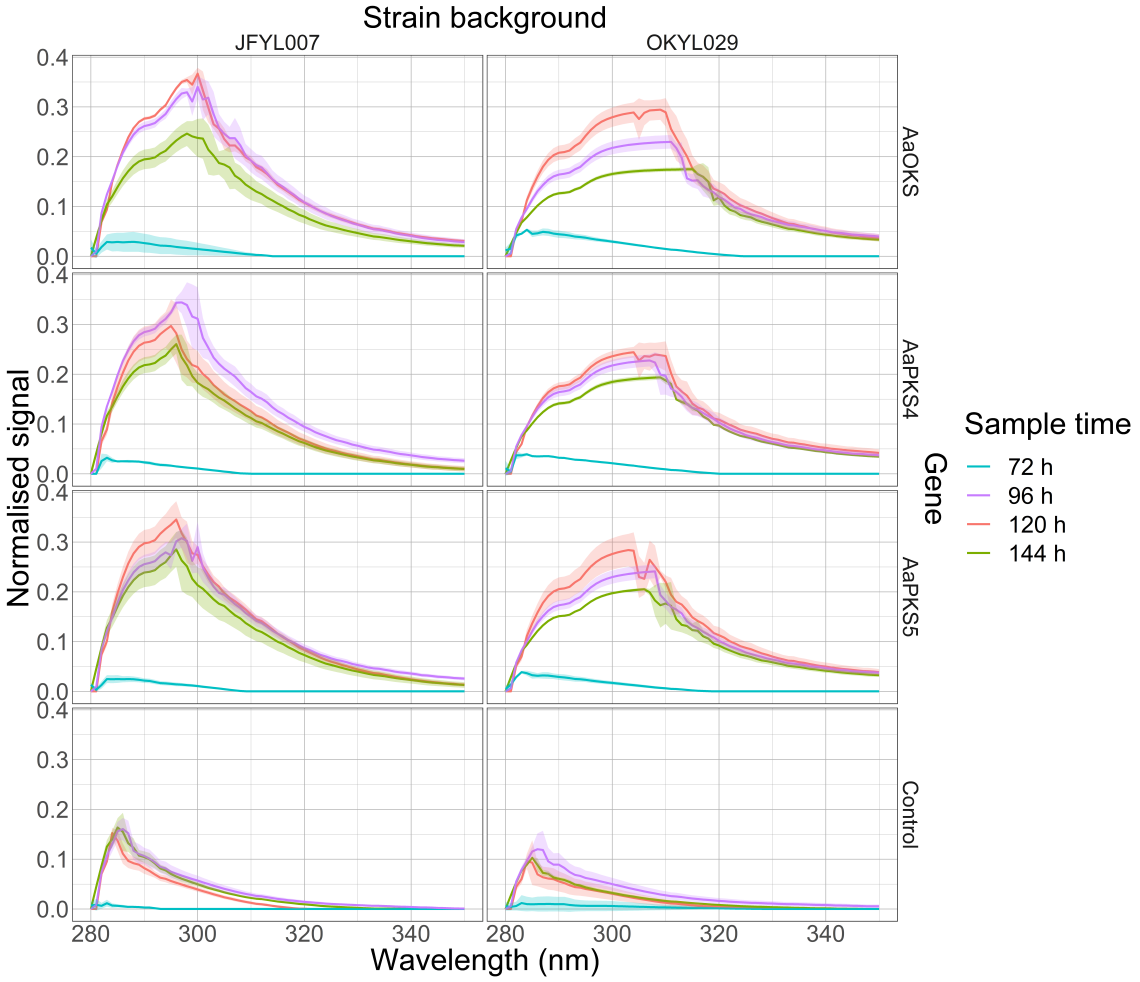


Figure 4.2: Absorbance of culture supernatant of strains SZYL009-014, OKYL029, and JFYL007 sampled at different time points. Strains SZYL009 and SZYL012 express AaOKS, strains SZYL010 and SZYL013 express AaPKS4, and strains SZYL011 and SZYL014 express AaPKS5. SZYL009-011 are derived from JFYL007 and SZYL012-14 are derived from OKYL029. JFYL007 and OKYL029 are controls. Strains were grown in Delft C/N 100 and sampled at 72 h, 96 h, 120 h, and 144 h. The graphs are divided by inserted gene and strain background. The line colour represents sample time. The spectra represent average of biological duplicates and the shadow represents standard deviation.

4.1.3 Testing the biosensor in different strain backgrounds

So far, the difference between the JFYL007- and OKYL029-based strains had not been big. To further investigate the biosensor, another strain (OKYL049) was transformed with the three PKSs to generate the strains HOYL01-03. OKYL049 overexpresses genes that are implied in lipid accumulation. Theoretically, this means that a larger share of the flux through malonyl-CoA would be directed to fatty acid synthesis, compared to in JFYL007. Less malonyl-CoA flux is therefore directed to the biosensor product. This would be seen as a lowered biosensor signal in OKYL049-derived strains compared to JFYL007-derived strains. An experiment to compare the three strain backgrounds was set up, see figure 4.3. The signal in the OKYL049 strains was comparable to the OKYL029 strains, with a clear difference between control and strains with a PKS. The expected result that there would be less signal in these strains can be seen in the normalised spectra (figure 4.3).

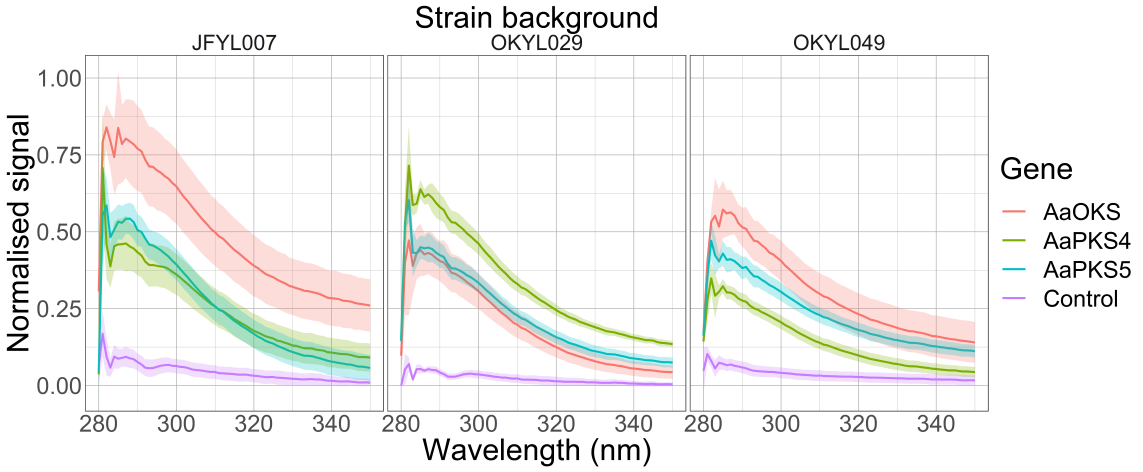


Figure 4.3: Absorbance of culture supernatant of strains SZYL009-014, HOYL01-03, OKYL029, JFYL007, and OKYL049. Strains SZYL009, SZYL012, and HOYL01 express AaOKS, strains SZYL010, SZYL013, and HOYL02 express AaPKS4, and strains SZYL011, SZYL014, and HOYL03 express AaPKS5. SZYL009-011 are derived from JFYL007, SZYL012-14 are derived from OKYL029, and HOYL01-03 are derived from OKYL049. JFYL007, OKYL029, and OKYL049 are controls. Strain were grown in Delft C/N 100 for 96 h. The graphs are divided by strain background. The line colour represents inserted gene. The spectra represent average of biological triplicates and the shadow represents standard deviation.

Of the three PKSs tested as biosensor, the AaOKS enzyme was deemed the most suitable candidate. The hypothesis was that the JFYL007 strain would have the highest malonyl-CoA concentration, followed by OKYL029 and OKYL049. AaOKS gave a high signal in the strain with JFYL007 background, while the signal was lower in the other two strain backgrounds (figure 4.3). This pattern coincides with the hypothesis. The AaPKS5 enzyme seemed to produce a similar amount of product regardless of strain background, indicating no linear dose-dependency. Therefore, AaPKS5 was excluded. In the strains with the AaPKS4 enzyme, the strain background OKYL029 had the highest signal, followed by JFYL007 and OKYL049. As this result did not align with the hypothesis, the enzyme was excluded.

4.1.4 Testing biosensor dose-dependency with cerulenin

To further investigate the dose-dependency of the enzyme, an experiment to increase the concentration of malonyl-CoA was executed. The intracellular malonyl-CoA concentration was indirectly increased by supplementation of cerulenin, an antibiotic that inhibits the fatty acid synthase complex in the fatty acid biosynthesis. The expected outcome was that the biggest signal difference could be seen in the two strains with lipid accumulation (SZYL012 and HOYL01), since the fatty acid synthesis pathway would be blocked, redirecting malonyl-CoA flux towards the biosensor product. Samples were taken at 24 h, 40 h, 48 h, 72 h, and 96 h, with cerulenin supplemented at 40 h. The strains tested were two sets of the three strains that expressed the OKS enzyme (SZYL009, SZYL012, and HOYL01). One set were supplemented with cerulenin. The spectra acquired in the experiment can be seen in figure 4.4.

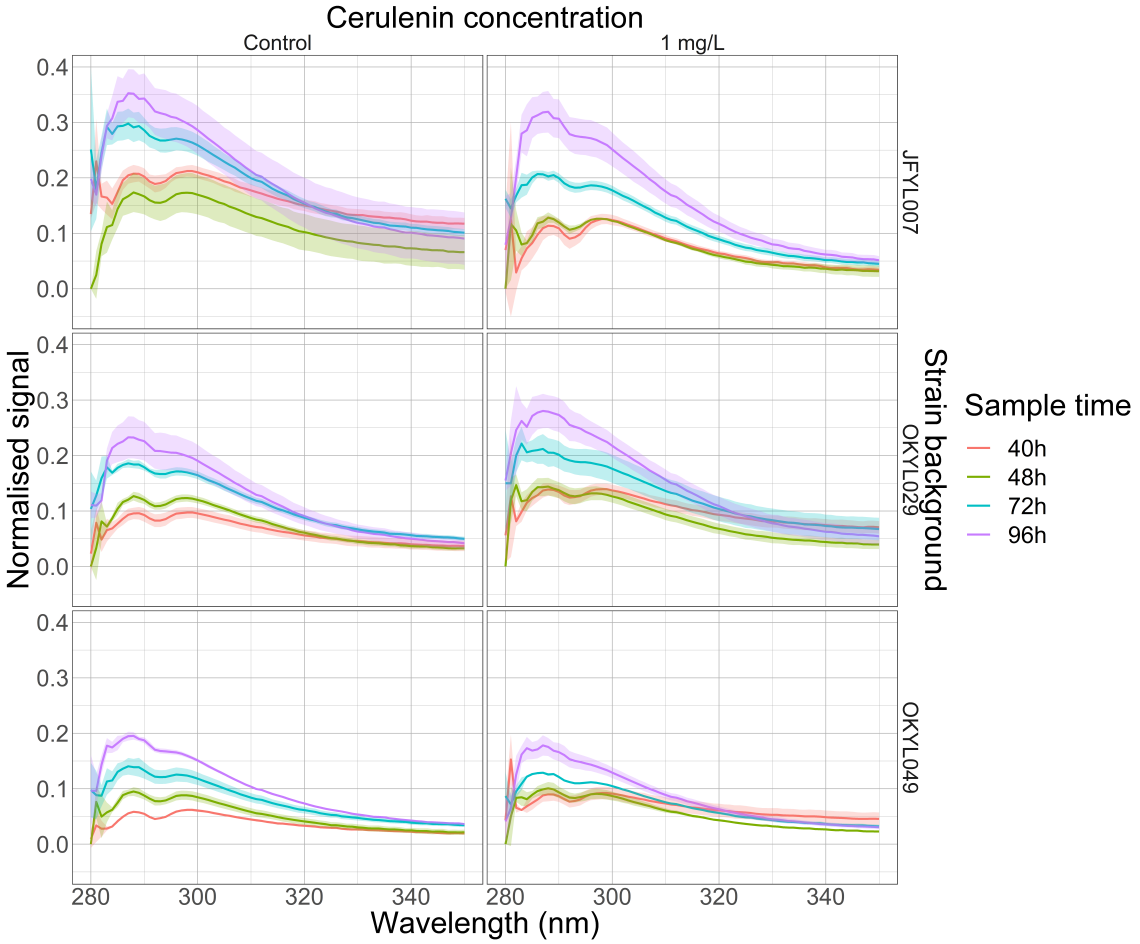


Figure 4.4: Absorbance of culture supernatant of strains SZYL009, SZYL012, and HOYL01, with or without supplementation of cerulenin. Strains SZYL009, SZYL012, and HOYL01 express AaOKS. SZYL009 is derived from JFYL007, SZYL012 is derived from OKYL029, and HOYL01 is derived from OKYL049. Two sets of the strains were grown in Delft C/N 100 and sampled at 40 h, 48 h, 72 h, and 96 h. 1 mg/L cerulenin were supplemented to one set at 40 h. The graphs are divided by strain background and cerulenin concentration. The line colour represents sample time. The spectra represent average of biological triplicates and the shadow represents standard deviation.

Generally, the spectra for the strains grown with or without cerulenin, respectively, are very similar at all time points. This indicates that cerulenin had little to no effect, since the signal strength did not increase after the antibiotic supplementation. For SZYL009 (derived from JFYL007), the spectra are slightly different. The SZYL009 replicates grown in the control condition had a higher overall absorbance compared to the replicates supplemented with cerulenin, mostly noticeable in the earlier time points. Therefore, the net absorbance increase from 40 h to 96 h was bigger when cerulenin was supplemented. However, no difference could be seen in the two strains where a change was expected (figure 4.4). This spurred two hypotheses. The first was that the PKS was saturated. The increased malonyl-CoA concentration following the cerulenin supplementation could therefore not be utilised by the enzyme, since it was already working at its maximum rate. The second was if the cerulenin actually had an effect on the intracellular malonyl-CoA concentration. The first hypothesis could be investigated by overexpressing the enzyme AaOKS and comparing the signal to the signal from strains with no or one copy respectively.

4.1.5 Overexpression of biosensor enzyme AaOKS

Three more strains were constructed (HOYL04, HOYL05, and HOYL06) by integrating another copy of the *AaOKS* gene into the strains already expressing one copy (SZYL009, SZYL012, and HOYL01). Another shake flask experiment was set up, where the tested strains were JFYL007, OKYL029, and OKYL049 with no, one, or two copies of the gene *AaOKS*. Samples were taken in similar fashion to earlier at time points 40 h, 72 h, and 96 h. At 40 h, cerulenin was supplemented to one of two sets of shake flasks containing the strains with two *AaOKS* copies. The effect of the gene copy number can be seen in the spectra presented in figure 4.5.

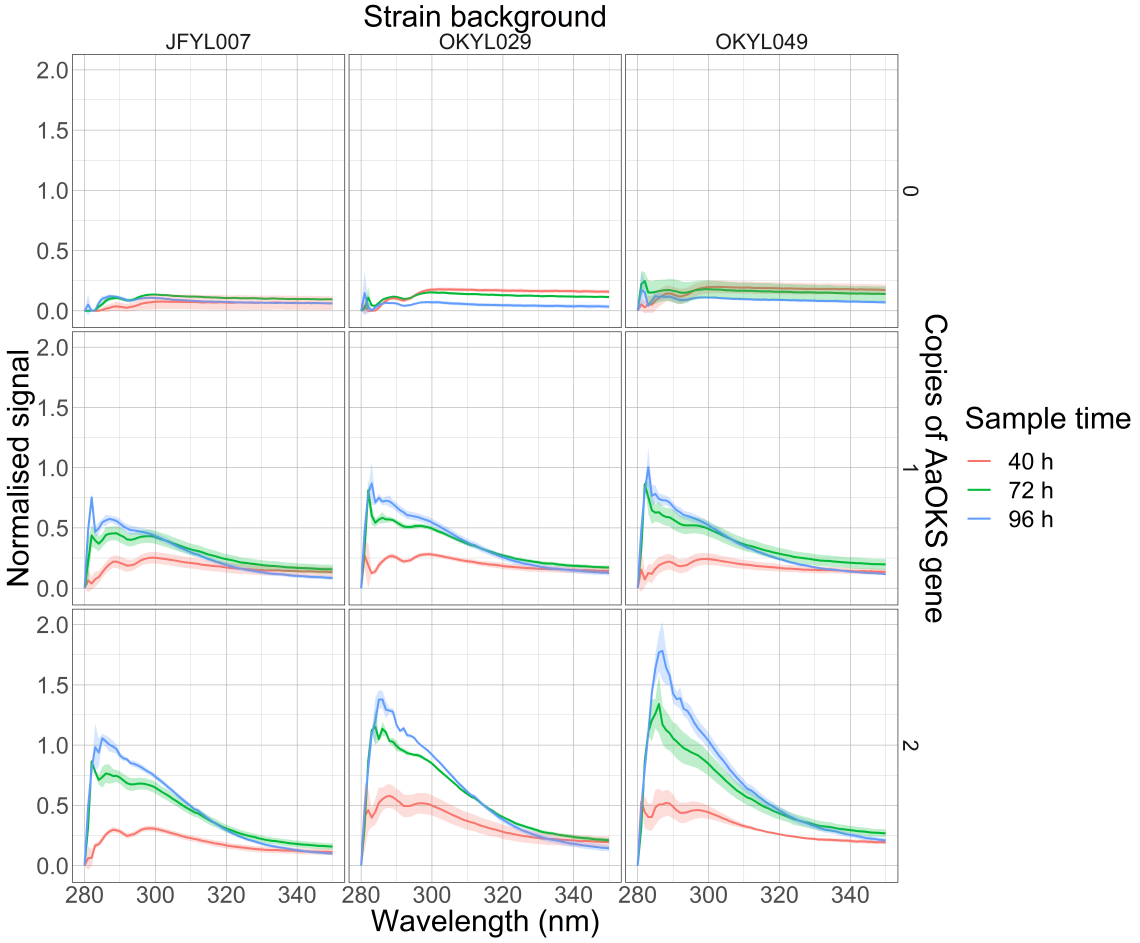


Figure 4.5: Absorbance of culture supernatant of strains SZYL009, SZYL012, HOYL01, HOYL04-06, JFYL007, OKYL029, and OKYL049, sampled at different time points. Strains SZYL009, SZYL012, and HOYL01 express AaOKS. SZYL009 is derived from JFYL007, SZYL012 is derived from OKYL029, HOYL01 are derived from OKYL049, HOYL04 is derived from SZYL009, HOYL05 is derived from SZYL012, and HOYL06 is derived from HOYL01. JFYL007, OKYL029, and OKYL049 are controls. Strains were grown in Delft C/N 100 and sampled at 40, 72 h, and 96 h. The graphs are divided by copies of AaOKS gene and strain background. The line colour represents sample time. The spectra represent average of biological triplicates and the shadow represents standard deviation.

For the control strains with no copies of the gene (JFYL007, OKYL029, and OKYL049), the absorbance was very low at all time points, which was expected. The strains containing one copy of *AaOKS* (SZYL009, SZYL012, and HOYL01) had a similar absorbance pattern as seen in previous experiments, with the difference that the JFYL007-based strain had lower absorbance than the OKYL029- and OKYL049-

based strains, which in turn had a similar absorbance. Previously, the absorbance has been highest for the JFYL007-based strain, or similar for all three strains (see figures 4.1, 4.2, and 4.3). The difference between the three strain backgrounds became even more discernible when the *AaOKS* gene is overexpressed (two copies), especially at the last time point (96 h). The normalised signal was then unexpectedly almost double in the OKYL049-based strain than in the JFYL007-based strain (see figure 4.5). The explanation could be related to the strains' respective genotype, as mentioned earlier. OKYL049 and OKYL029 can accumulate lipids (OKYL049 in a higher capacity) in the form of TAGs, while JFYL007 cannot. Therefore, it was assumed that JFYL007 strains could direct a higher flux of malonyl-CoA towards the biosensor product. However, TAGs can be degraded in β-oxidation to produce acetyl-CoA which can be converted to malonyl-CoA using the enzyme ACC. Therefore, both OKYL029 and OKYL049 can increase the malonyl-CoA flux by oxidising accumulated TAGs. This is not possible in JFYL007 due to its inability to accumulate TAGs. With two copies of the *AaOKS* using malonyl-CoA, the effect of the β-oxidation contributing with flux of malonyl-CoA becomes more noticeable. This could be investigated by knockdown or knockout of the genes corresponding to the six acyl-CoA oxidases carrying out the first step of β-oxidation [23] simultaneously as the *AaOKS* gene is overexpressed. A decrease in signal in this strain compared to the strain with solely overexpression of *AaOKS* could indicate that β-oxidation is responsible for increasing the flux through malonyl-CoA, and therefore also production of the signal molecule.

The effect of cerulenin on the strains with two copies of *AaOKS* (HOYL04-06) can be seen in figure 4.6. The cerulenin supplementation did not make a difference in the signal, except for a small increase for the OKYL49 strain at 96 h. It is possible that the enzyme was still saturated, and even more copies are needed to see the full effect of the cerulenin. An alternative is to place the gene for AaOKS under control of a stronger promoter, such as p*TEF1in*, to increase the expression of the gene and thus the concentration of enzyme. This approach would be more logical for an enzyme biosensor, since the genetic modification would be simpler and less time-consuming. Another theory to the low effect of the cerulenin is that the concentration is too low. Yang et al. (2018) performed a cerulenin titration, which showed that a signal increase could be seen for AaOKS in the same concentration used here (4.5 µM or 1 mg/L), but the biggest difference was visible at 20 µM (4 mg/L). An important note here is that the experiment was done in *E. coli*, which might be affected at a different concentration than *Y. lipolytica*. In the few articles found where cerulenin was used in *Y. lipolytica*, the most common cerulenin concentration was 1 mg/L (4.5 uM) [48][50][51][56], but one study used a higher concentration of 4 mg/L (20 µM) [49]. In this thesis, the intention of supplementing cerulenin was to see if the signal of the biosensor would increase, not to maximise the biosensor molecule production. If a clear difference could be seen when 1 mg/L cerulenin was supplemented, it could be of interest to perform a titration to better measure the dose-dependency of the biosensor. However, since the difference in signal was very small, the choice of cerulenin concentration can be supported.

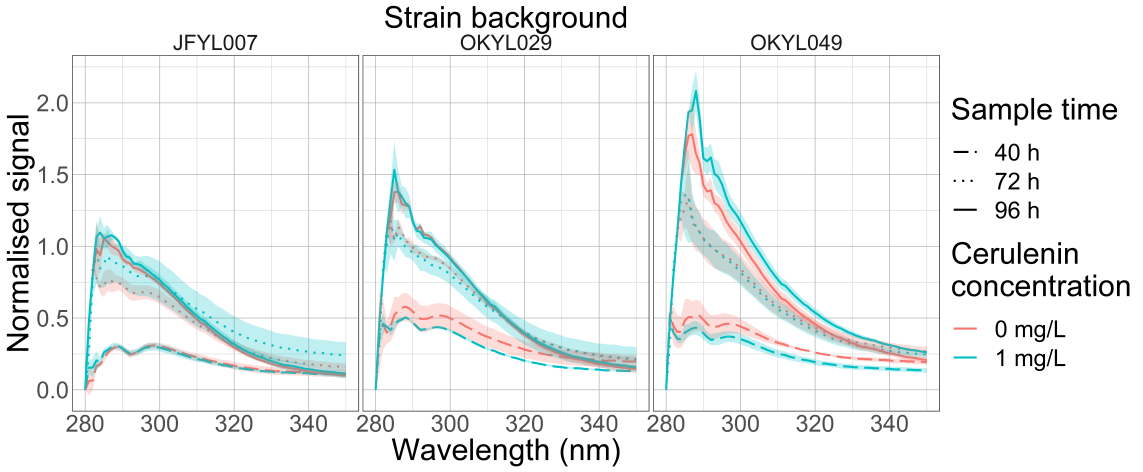


Figure 4.6: Absorbance of culture supernatant of strains HOYL04-06, with or without supplementation of cerulenin, sampled at different time points. Strains HOYL04-06 express AaOKS. HOYL04 is derived from SZYL009, HOYL05 is derived from SZYL012, and HOYL06 is derived from HOYL01. Two sets of the strains were grown in Delft C/N 100 and sampled at 40 h, 72 h, and 96 h. 1 mg/L cerulenin were supplemented to one set at 40 h. The graphs are divided by strain background. The line colour represents cerulenin concentration and the line type represents sample time. The spectra represent average of biological triplicates and the shadow represents standard deviation.

From the results presented here, it cannot be stated that the biosensor was dose-dependent for malonyl-CoA. The intracellular malonyl-CoA concentration could not be measured, and the indirect method of increasing it by adding cerulenin did not have a large effect. Therefore, no clear correlation between malonyl-CoA and biosensor signal was presented. However, an assumed correlation could be made based on the acquired results. Nitrogen limitation, which triggers lipid accumulation and thus increases the flux through malonyl-CoA in *Y. lipolytica*, seemingly also increased the signal (figure 4.1). Further, the malonyl-CoA availability was assumed to be different in the three strain backgrounds. This had an effect on the biosensor signal, but not always as expected.

4.1.6 Future perspectives

Further studies to show a correlation between a high flux through malonyl-CoA and a high biosensor signal (high production of polyketide reporter molecule) are needed to establish the polyketide enzyme as a biosensor in *Y. lipolytica*. However, because the biosensor did not work as expected according to its purpose in this thesis, it would be interesting to try an alternative strategy. Instead of only using the red polyketide as a reporter molecule for the amount of malonyl-CoA in the cell, the focus could be redirected to maximising production of that potentially valuable polyketide. By overexpressing the protein gene, the signal increased, showing a potential for increasing the production further through metabolic engineering. However, the product is unknown, so is the specific pathway towards the product. Yang et al. (2018) give suggestions to which polyketide the three PKSs produce, but does not fully identify it [11]. Therefore, computational tools applied to a genome-scale model (GEM) of *Y. lipolytica* cannot be used to identify gene targets for overexpression or knockouts. To identify the polyketide, a cooperation with a lab at Danmarks

Tekniske Universitet (DTU) was initiated in the thesis, but no results are available yet. If the product is known, the normalised signal can be compared to a standard curve of the product and thus be quantified. This would indicate the current titre and if the project should proceed further to increase production or if it simply would not be economically viable.

4.2 Promoter project

In a previous experiment (unpublished), *Y. lipolytica* was cultivated in chemostats in four different Delft media, where the nitrogen source and availability (through C/N ratio) were varied. Transcriptomics data were acquired using RNAseq, which indicated that there is differential expression of genes based on both nitrogen source and availability. Since this most likely could be attributed to the promoters of the genes, it was of interest to investigate the expression strength of the promoters in different conditions and over time to acquire the full regulation profile.

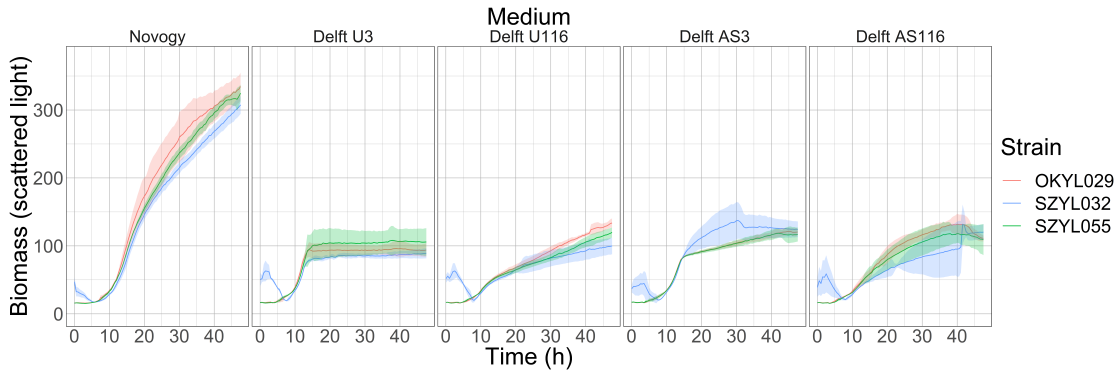
4.2.1 Measuring growth and GFP expression of promoter strains

From 27 chosen promoters from the *Y. lipolytica* genome, 24 plasmids were created based on JFEC133. Three promoter plasmids could not be constructed. Each plasmid contained a selected promoter, followed by *hrGFP* and the terminator *LIP2*. The gene cassettes were transformed into *Y. lipolytica* to generate the 24 strains SZYL031-54. An additional *Y. lipolytica* strain (SZYL055) was constructed, which contained the *Tef1* promoter as a positive control. 18 promoter strains (SZYL031-45, SZYL047, SZYL049, and SZYL054), the positive control strain (SZYL055), and the negative control strain (OKYL029) were then tested in the BioLector in the media Delft medium for chemostats with urea and C/N 3 or 116, and Delft medium for chemostats with ammonium sulphate and C/N 3 or 116. Some of the strains were also tested in YPD medium, YNB minimal medium, or Novogy medium. The four Delft media had been tested in the previous experiment in the chemostats, thus it was interesting to see if the results from that study could be replicated in the BioLector. The three other media were used to test the promoters in more conditions.

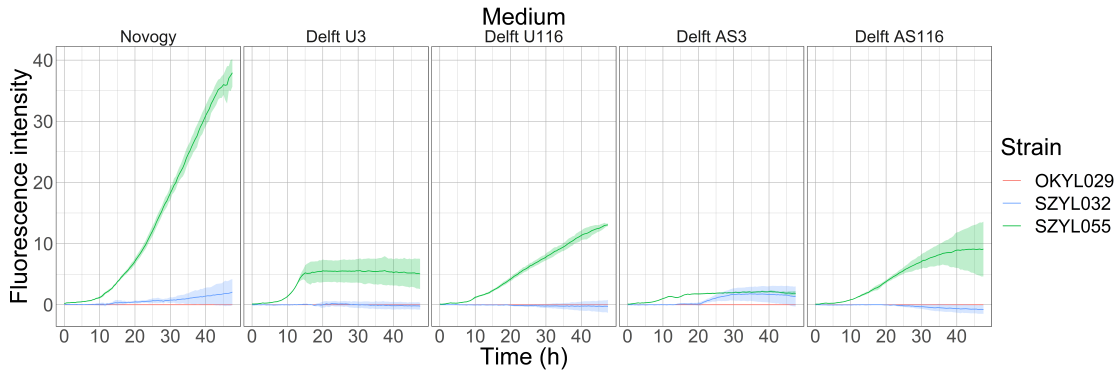
The first run contained YPD medium, which was discovered to have too much background fluorescence to use (results not shown). The fluorescence intensity was much higher than for the Novogy and Delft media at the start of the run. In the next run, it was replaced with YNB minimal medium, which had the same problem (results not shown). Both of the media were excluded from the experiment. Novogy medium had more background fluorescence than the Delft media, but it was low enough to keep (results not shown). In addition to the background fluorescence in the media, the cells were auto-fluorescent. This was determined since the negative control strain (OKYL029, the strain that all other strains are derived from) had a measured fluorescence intensity above zero, which increased over time as the cells grew. Therefore, the fluorescence intensity for all strains was normalised by sub-

tracting the fluorescence intensity for the negative control for all time points.

The biomass (scattered light) and fluorescence intensity measured in the BioLector for three strains (negative control OKYL029, positive control SZYL055, and representative test strain SZYL032) can be seen in figure 4.7. For the results from all strains, see figures A.1 and A.2 in Appendix A.1. The growth was very similar for all strains, but some variety was seen towards the end of the cultivations (figure 4.7a), likely due to evaporation in the wells. Since all promoter strains were derived from OKYL029, the growth for all strains was expected to be the same. Unexpectedly, none of the strains, except the positive control (SZYL055), expressed GFP.



(a) Biomass over time of OKYL029, SZYL032, and SZYL055 grown in five different media.



(b) Fluorescence intensity over time of OKYL029, SZYL032, and OKYL049 grown in five different media.

Figure 4.7: Biomass and fluorescence intensity over time of OKYL029, SZYL032, and SZYL055 grown in five different media. OKYL029 is negative control, SZYL055 is positive control and has *pTEF1*, and SZYL032 is test strain and has an uncharacterised *Y. lipolytica* promoter. Strains were cultivated in a 96-well microplate in BioLector for 48 h. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The graphs are divided by medium. The line colour represents strain. The data shown represent average of technical quadruplicates and the shadow represents standard deviation. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

The fluorescence intensity for the 18 tested promoter strains was measured at the same level as the negative control. The probability that all of the 18 tested promoters would be false positives and not have expression in any medium is very improbable. Further, the transcriptomics data (unpublished) showed that the *TEF1* gene had approximately the same counts as many of the genes connected to promoters tested here. The expression of the tested promoters should therefore theoretically be similar in strength to the *TEF1* promoter, and thus have a similar fluorescence

intensity, which cannot be seen. One difference to note here is that the transcriptomics data is from chemostat bioreactors, which differs from the cultivation in the BioLector. However, it is unlikely that in the chemostats, the expression for all 18 tested promoters is similar, while when growing in conditions more similar to a batch cultivation, only one promoter (p*TEF1*) gives expression.

To verify that there was no GFP expression in the promoter strains, cultures were examined in a fluorescence microscope after a 48 h run in the BioLector. For the promoter strains tested, images were taken of one strain (SZYL032) in all media, and all strains from the BioLector run (SZYL032, SZYL033, SZYL040, and SZYL042) in one medium (Delft for chemostats with urea, C/N 3). The microscope images of strains SZYL055, OKYL029, and SZYL032 can be seen in figure 4.8.

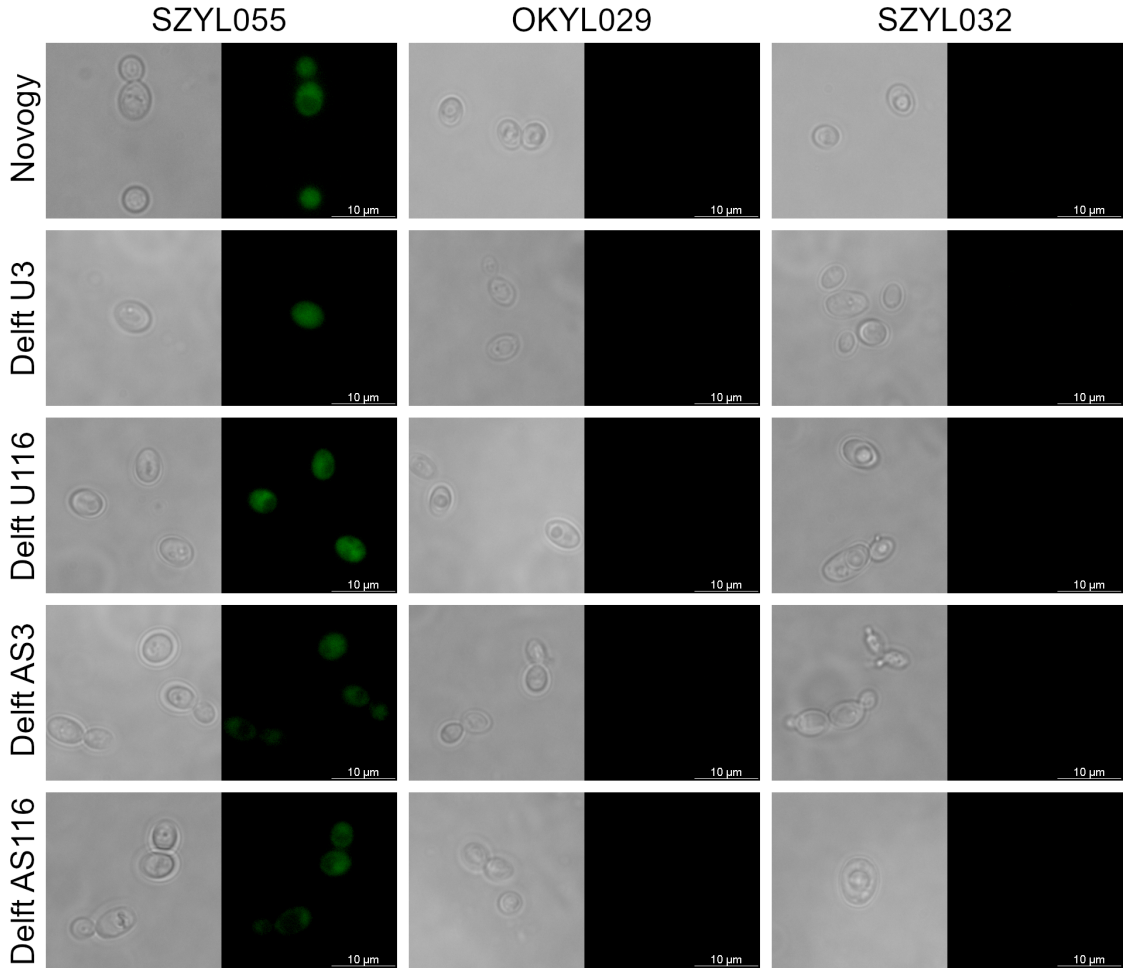


Figure 4.8: Fluorescence microscopy images of SZYL055, OKYL029, and SZYL032 grown in five different media. SZYL055 is positive control and has *pTEF1*, OKYL029 is negative control, and SZYL032 is test strain and has an uncharacterised *Y. lipolytica* promoter. Strains were cultivated in a 96-well microplate in BioLector for 48 h, then photographed in fluorescence microscope in brightfield and GFP channel. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The images are divided by medium and strain. Scale bar is 10 µm. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

The negative control expressed no GFP in any media as expected. The images were edited in the same manner so that no green could be seen in the background, mean-

ing that if any green was visible, it would be in the cells. None of the promoter strains tested showed any fluorescence, regardless of media. These results support the results from the BioLector. For the positive control SZYL055, the GFP was clearly visible in all media but with varying intensity, similarly to the data from the BioLector. p*TEF1* is classified as a constitutive promoter and should therefore have the same expression strength in all conditions, but the variance in fluorescence intensity in different media indicates otherwise. The number and size of the cells could explain the variation; more and bigger cells can overall express more GFP. Nonetheless, the normalised-to-biomass fluorescence intensity measured by the Biolector still varied between media (figure 4.9), but the differences were less prominent. Up until 10 h of incubation in the Biolector, the normalised fluorescence intensity was very similar in all media, with a slightly lower intensity for the cells grown in Delft media with ammonium sulphate. After that, the fluorescence intensity increased for cells grown in Novogy medium or both Delft media with C/N 116. For the Delft media with C/N 3, the intensity remained the same or even decreased (figure 4.9). Although these results indicate that p*TEF1* is not constitutive, such a conclusion cannot be drawn from data on one strain grown in five different media.

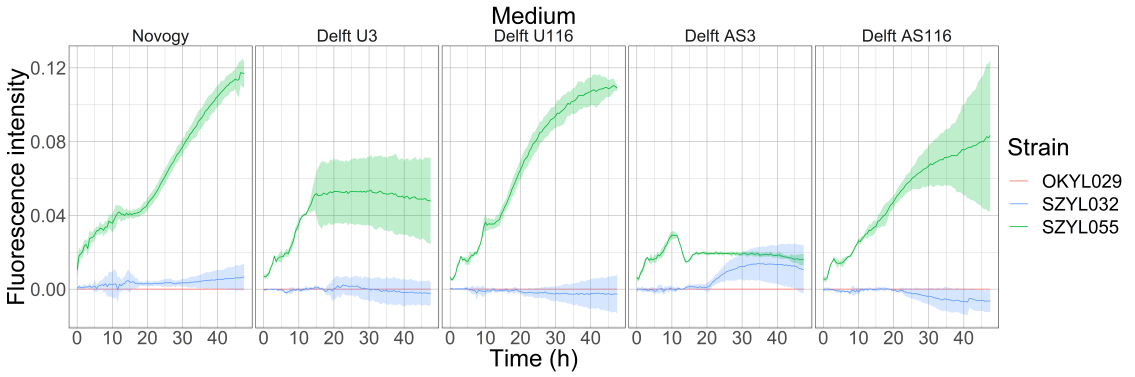


Figure 4.9: Normalised-to-biomass fluorescence intensity over time of OKYL029, SZYL032, and SZYL055 grown in five different media. OKYL029 is negative control, Strains were cultivated in a 96-well microplate in BioLector for 48 h. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The graphs are divided by medium. The line colour represents strain. The data shown represent average of technical quadruplicates normalised to biomass and the shadow represents standard deviation. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

Novogy and Delft with urea C/N 116 seemed to give approximately the same expression, followed by Delft with ammonium sulphate C/N 116. Cells grown in Novogy medium seemed to accumulate large amounts of lipids, which can be seen in the growth curves (figure 4.7a). After the exponential phase (approximately 18 h), there was not a clear plateau marking beginning of the stationary phase. Instead, the biomass increased steadily, indicating that the cell size was growing due to accumulation of lipids. Both Delft media with C/N 116 stimulated lipid production due to nitrogen limitation. For the two Delft media with C/N 3, the normalised fluorescence intensity was much lower than for the other three media, which could be due to glucose limitation. This leads to inhibited growth and slower metabolism, refocusing resources to more essential operations in the cell than to express GFP. Interestingly, the nitrogen source also seemed to have an effect, with lower normalised fluorescence intensity in the two media with ammonium sulphate compared to the

respective media with urea. This could be explained by how GFP is affected by pH.

4.2.2 Investigating effect of pH on GFP fluorescence

Most GFP variants are sensitive to decreases in pH due to a phenol group on the chromophore [57], which could also be true for the variant used in this experiment (*hrGFP*) although no information on this has been found. For the pH sensitive fluorescent proteins, a lower pH results in a decrease in fluorescence (absorbance in specific wavelength) [57]. Since the GFP in the tested cells was expressed in the cytosol, the intracellular pH would have to have been lowered for the protein to be affected by it. Generally, *Y. lipolytica* is stable during many different cultivation methods including pH; it can grow in pH varying between 2.5 and 7.5 [16]. This should indicate that the intracellular pH is kept at a fairly constant level. Orij et al. tested the variation in intracellular pH when the external pH was changed in *Saccharomyces cerevisiae* [58]. The intracellular pH was measured in cells growing in either pH 3.0 or 6.0, followed by a change in external pH to 6.0 or 3.0. The pH change in the cytosol was considered negligent (it did not drop below pH 7.0) and the growth rate remained the same for pH between 3.0-7.5 [58]. Although these data come from *S. cerevisiae*, *Y. lipolytica* most likely behaves in a similar way.

To test if the GFP in the promoter strains could have been affected by too low intracellular pH, two promoter strains (SZYL032 and SZYL035) were cultivated in shake flasks in the Novogy medium and the four Delft for chemostats media. The pH of the media was measured before and after cultivation (supernatant of culture), see table 4.1 for the values.

Table 4.1: pH of five different media before and after 48 h cultivation in shake flasks. The pH value after cultivation is the mean with standard deviation of the pH of in total six samples (biological triplicates of strains SZYL032 and SZYL035). Abbreviations: AS, ammonium sulphate.

Medium	pH before cultivation	pH after cultivation
Novogy	5.51	5.46 ± 0.02
Delft with urea C/N 3	5.66	8.41 ± 0.16
Delft with urea C/N 116	5.60	3.46 ± 0.06
Delft with AS C/N 3	5.32	2.37 ± 0.03
Delft with AS C/N 116	5.38	2.47 ± 0.16

The Novogy medium is the only medium that is buffered, therefore it is reasonable that the pH was stable and did not drop or rise significantly during cultivation. For the Delft media, the results are varied. The pH in the two media with ammonium sulphate decreased approximately 2.5 points, from 5 to 2.37 ± 0.03 and 2.47 ± 0.16, respectively. For the two Delft media with urea, the end pH was very different: 8.41 ± 0.16 for C/N 3 and 3.46 ± 0.06 for C/N 116. A potential explanation for the increase in pH for Delft with urea C/N 3 could be due to urea degradation. As the carbon source is limiting in the medium, there is an abundance of nitrogen in the

form of urea. In aqueous solutions, urea degrades to ammonium, which increases the pH, and cyanate. The cyanate is further degraded to carbon dioxide and ammonia [59]. The high pH should not have affected the GFP expression [57], but an external pH above 8.0 can result in cell lysis, due to a reversed pH gradient [58]. This could not been seen in the tested strains, as the growth in Delft wth urea C/N 3 was equivalent to the growth in the other media (figures 4.7a and A.1).

For the positive control, the fluorescence intensity was lower in cultivations with ammonium sulphate than with urea. The study by Orij et al. only estimated the intracellular pH in an external pH as low as 3.0. The pH in the Delft media with ammonium sulphate after cultivation was below this value. Therefore, it is possible that the intracellular pH of the cells grown in this media did indeed drop, and thus affected the GFP expression in all of the strains. If the expression was very low from the beginning, the small amounts of GFP could have been negatively affected by a drop in intracellular pH, resulting in no fluorescence. However, this does not explain why the promoter strains had no fluorescence in the other media. The pH in the Novogy medium was most likely stable throughout the cultivation, yet there was no GFP expression in the promoter strains. For the Delft media with urea, the pH never reached below 3. Therefore, the lack of GFP expression in all promoter strains cannot be fully attributed to a decrease in intracellular pH.

4.2.3 Troubleshooting steps for unexpected results

Since the only yeast strain that expressed GFP as expected (SZYL055) was constructed in a different way than the promoter strains, something could be wrong with the cloning strategy. During construction, only the promoter region was sequenced in the constructed plasmids before transforming *Y. lipolytica*. If the GFP was in some way incorrect, for instance if the starting codon (ATG) was mutated, the protein could not have been expressed. Again, the probability that such a mutation occurred in 18 separate plasmids is very unlikely. To double-check that no mutations had occurred, the GFP part of the plasmid for strain SZYL036 was sequenced. No mutations were present in the plasmid. To verify that the yeast strains were constructed as expected, the entire inserted cassette (promoter+*hrGFP*+t*LIP2*) was sequenced for strains SZYL031, SZYL032, SZYL035, and SZYL036. In SZYL032, there were no mutations in the promoter region. In SZYL031, SZYL035 and SZYL036, there were a few mismatches, insertions, or deletions in the promoters. SZYL031 had deletions at bases 525 (A) and 535 (A). SZYL035 had a deletion at base 808 (A). SZYL036 had an insertion at base 566 (A). However, the quality score of these readings was very low compared to the rest of the sequencing, since they were close to the end of the sequencing fragment. It is also possible that the reference genome was not identical to the template genome used to amplify the promoters. Since there is no previous characterisation of the sequenced promoters, it is not possible to know if the discrepancies found would affect the expression strength. Furthermore, there were no mutations in the *hrGFP*, including the starting codon, or in the junction between promoter and *hrGFP*. From these results, it was assumed that all, or at least the majority, of the *Y. lipolytica* strains were constructed correctly.

Apart from not being constructed in the same way as the other promoter strains, SZYL055 also has the terminator t*PEX20* on the reverse strand upstream of p*TEF1*. During primer design, it was assumed that this terminator had no relevance in the expression of a gene on the other DNA strand, and was therefore kept for convenience reasons. If this statement is true, a strain with p*TEF1* and without t*PEX20* should have the same expression of GFP as SZYL055. To investigate this, three new strains were constructed, containing one of three promoters previously used in *Y. lipolytica*: p*TEF1* (same as SZYL055 but without t*PEX20*, called SZYL056), p*PYK1* (SZYL057), and p*GAPDH* (SZYL058) [17]. p*PYK1* and p*GAPDH* were included to compare the promoter strength of p*TEF1*, since it is possible that all tested promoters were too weak in comparison to p*TEF1*. The strains SZYL056-58 were cultivated in shake flasks in Novogy and the four Delft media for 96 h. Samples were taken every 24 h to control the fluorescence of GFP in a fluorescence microscope. The images taken at 48 h can be seen in figure 4.10. For the pictures taken at 24 h, 72 h, and 96 h, see figures A.3, A.4, and A.5 in Appendix A.1.

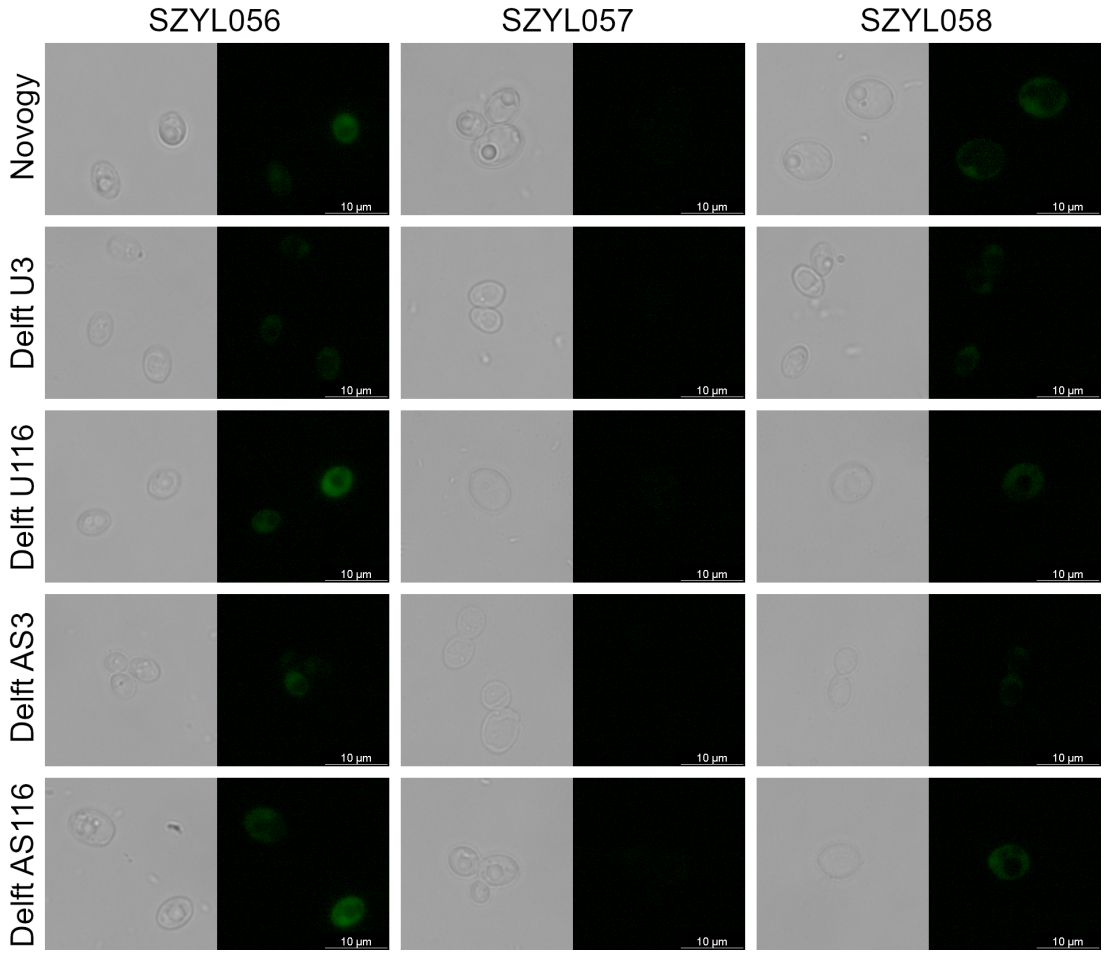


Figure 4.10: Fluorescence microscopy images of SZYL056, SZYL057, and SZYL058 grown for 48 h in five different media. SZYL056 has p*TEF1*, SZYL057 has p*PYK1*, and SZYL058 has p*GAPDH*. Strains were cultivated in shake flasks for 48 h, then photographed in fluorescence microscope in brightfield and GFP channel. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The images are divided by medium and strain. Scale bar is 10 µm. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

For the strains SZYL056 and SZYL058, fluorescence can be seen at all time points except at 24 h. For SZYL056, the fluorescence was equally strong at 48 h, 72 h, and 96 h. For SZYL058, the fluorescence was most intense at 48 h. In SZYL057, no fluorescence could be detected at any of the time points. The most intense fluorescence was seen in SZYL056 (*TEF1* promoter), followed by SZYL058 (*GAPDH* promoter), and lastly SZYL057 (*PYK1* promoter), indicating the strength of each promoter. The order also corresponds to the number of gene counts found for the gene associated with each respective promoter in the transcriptomics data from the project (unpublished) leading up to this experiment. Possibly, the expression level of GFP under the *PYK1* promoter could be too low to be detected, since the promoter have worked previously [17]. This could indicate that there was GFP expression in the promoter strains, only it was too low for detection in either the BioLector or the fluorescence microscope. Similarly to the fluorescence intensity seen in both the BioLector and the fluorescence microscope for SZYL055, the fluorescence varied for strains SZYL056 and SZYL058 depending on which media the cells were cultivated in. Also seen in this experiment was that the fluorescence increased over time for SZYL056, while it decreased for SZYL058. The promoters tested are classified as constitutive, the expression of GFP should therefore be stable through all phases of cell growth. The accumulation of GFP in SZYL056 could also be seen in the BioLector for strain SZYL055, since the measured fluorescence intensity increased throughout the run, even when it was normalised to the biomass.

After establishing that at least two of the three strains with characterised promoters could express GFP, it was investigated how the expression level compared with the one seen in SZYL055. It would be specifically interesting to compare the two strains that contain the promoter *TEF1*: SZYL055 and SZYL056. The difference, as mentioned earlier, is that SZYL055 has the terminator *PEX20* on the opposite strand upstream of p*TEF1*. The comparison could indicate if this terminator would have an effect on the promoter activity or not. The four strains SZYL055-58 were cultivated in shake flasks in Delft with ammonium sulphate C/N 116 for 72 h, followed by fluorescence imaging in the fluorescence microscope. The images can be seen in figure 4.11.

New settings (for instance longer exposure time) were used for these images to see if any fluorescence could be detected in SZYL057. All four strains showed GFP fluorescence, with the two strains containing p*TEF1* being the strongest, and the strain with p*PYK1* being the weakest. Interestingly, SZYL056 seemed to be more fluorescent than SZYL055, indicating that t*PEX20* had a negative influence on the promoter strength. Since fluorescence for SZYL057 could only be seen with longer exposure time, the theory that this could be the case for the promoter strains too was suggested. The settings that were used when capturing images of the promoter strains were therefore used to capture SZYL057. If no fluorescence was detected in SZYL057 while using these settings, then it is possible that the GFP expression in the promoter strains was at the same level or lower as SZYL057. The comparison of SZYL057 and SZYL032 in Delft with ammonium sulphate C/N 116, with the same settings for the fluorescence microscope, can be seen in figure 4.12.

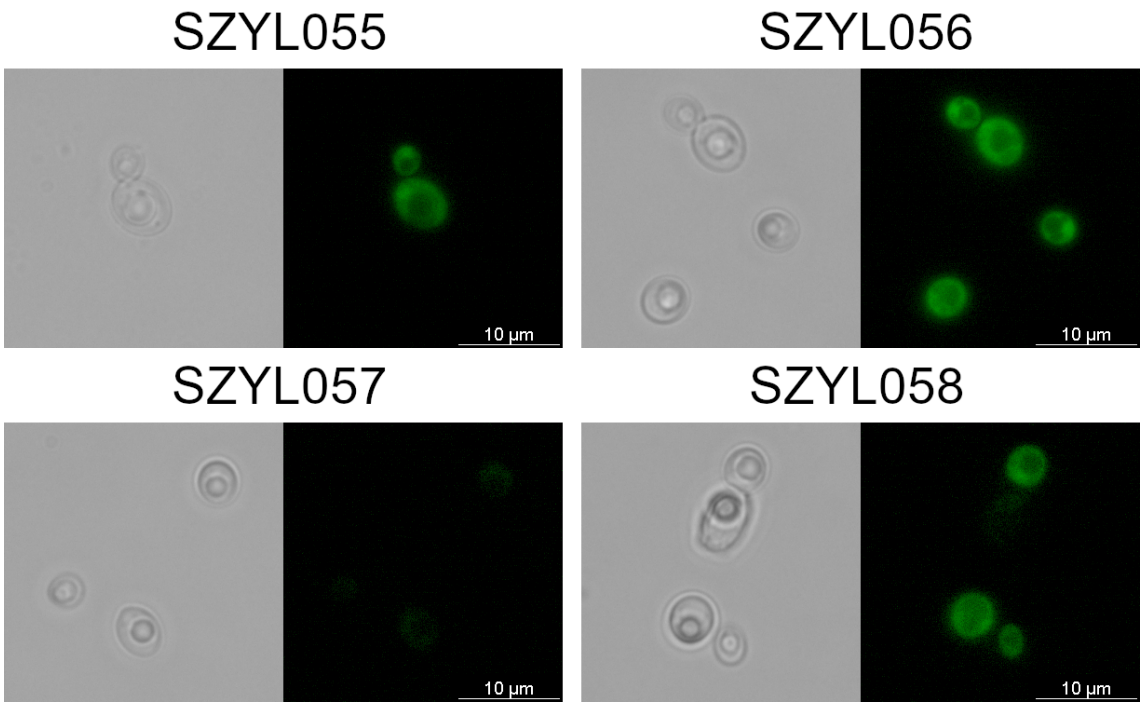


Figure 4.11: Fluorescence microscopy images of SZYL055, SZYL056, SZYL057, and SZYL058. SZYL055 and SZYL056 has p*TEF1*, SZYL057 has p*PYK1*, and SZYL058 has p*GAPDH*. Strains were cultivated in Delft with ammonium sulphate C/N 116 in shake flasks for 72 h, then photographed in fluorescence microscope in brightfield and GFP channel. The images are divided by strain. Scale bar is 10 µm.

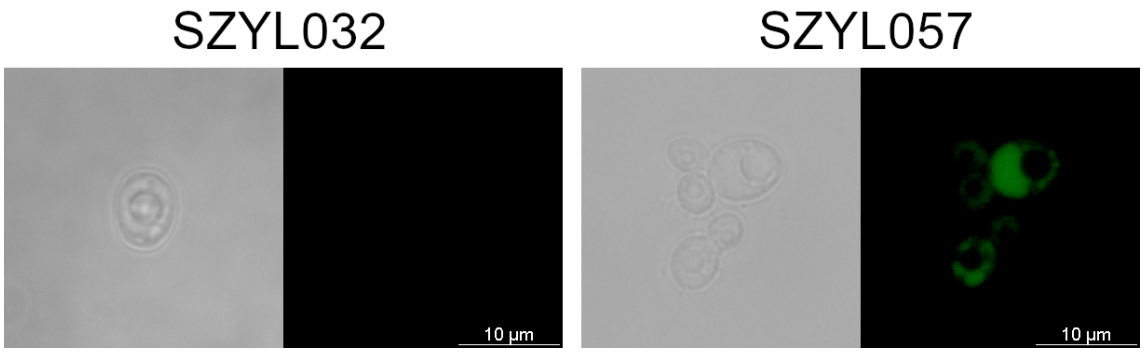


Figure 4.12: Fluorescence microscopy images of SZYL032 and SZYL057. SZYL032 has an uncharacterised *Y. lipolytica* promoter and SZYL057 has p*PYK1*. SZYL032 was cultivated in a 96-well microplate in BioLector for 48 h. SZYL057 was cultivated in a shake flask for 72 h. The images are divided by strain. Scale bar is 10 µm.

One difference to note here is that SZYL057 was cultivated in shake flasks for 72 h, while SZYL032 was taken from the 96-well microplate after a BioLector run that ran for 48 h. The comparison is therefore not based on equal conditions. However, there was a high fluorescence for SZYL057, higher than in figure 4.11. The settings used to capture the promoter strains were therefore more sensitive, meaning that if any fluorescence would have been expressed in SZYL032, it would most likely have been visible. Moreover, the gene counts for *PYK1* (promoter in SZYL057) was approximately 20 times lower than for the gene controlled by the promoter in SZYL032 (unpublished), further strengthening that the promoter strains had no GFP expression.

Taking all results from this project into account, it can be stated quite clearly that the promoters used in the promoter strains did not initiate expression. The reason for this remains unknown. One thing to note is that the experiment was based on data from chemostats, while the results presented here come from shake flasks or BioLector runs (both equivalent to batch fermentations). In chemostats, the cells are constantly in nutrient limitation, while this occurs only after a certain time point in batch fermentations. However, it is unlikely that this is the explanation for the lack of GFP expression. GFP is seen in the positive control after cultivation in both shake flasks and BioLector. Further, it was determined that the cloning strategy used was not faulty, since GFP expression was seen in strains SZYL056-58. It was also concluded that it was unlikely that the intracellular pH dropped to a level where the fluorescence of GFP was affected. Another possible explanation, that was not explored in the thesis, could be that a transcriptional regulatory element was missing in the cells to initiate transcription of the *hrGFP* gene. The initiation process is complicated in eukaryotes and could rely on more that just the promoter sequence [8]. During strain construction, the gene cassette was integrated in the D1 locus, meaning that any elements in the original genomic location were not present. For the three characterised promoters tested (p*TEF1*, p*PYK1*, and p*GAPDH*), the core promoter sequence itself was enough for expression. However, by adding a part of the first intragenic intron of *TEF1* to p*TEF1* (creating p*TEF1in*), the expression has previously been described to increase 17 times [26]. It is therefore possible that such transcriptional regulatory elements could be found in the original genomic locations for the tested uncharacterised promoters. Without them, transcription initiation could have been lowered or even non-existent, leading to very low or no GFP expression.

4.2.4 Future perspectives

There are several possibilities in continuing this research. First, it would be interesting to find current, or produce new, transcriptomics data from shake flasks to determine the expression strength of the tested promoters in batch cultivation. This data would be more representative for the experimental setup used in this thesis, rather than the chemostat data used for comparison now. Furthermore, although the fluorescent protein used (*hrGFP*) worked well for p*TEF1*, p*PYK1*, and p*GAPDH*, it might not have been the best reporter gene for the other tested promoters. There are many available fluorescent proteins to choose between, each with properties more suitable for specific purposes [60]. A selection of fluorescent proteins could be tested under the *TEF1* promoter to compare the fluorescence intensity and determine which has the strongest signal. By using a fluorescent protein with an overall stronger signal, the detection range would increase so very low expression could be detected. Similarly, it is possible that the instrument chosen in the experimental setup is not sensitive enough. No data on the sensitivity of the BioLector could be found, meaning it cannot be compared to other instruments. Potentially, flow cytometry could be used instead to offer more sensitivity in measuring the fluorescence.

5
Conclusions

This research explores two genetic tools for metabolic engineering of *Y. lipolytica*. First, a potential colorimetric biosensor for malonyl-CoA was evaluated. Its dose-dependency was tested by employing the biosensor in three strains, each with a different theoretical flux through malonyl-CoA. An indirect method to further increase the malonyl-CoA concentration in each of the strains was tested as well, through the supplementation of cerulenin. When only one copy of the enzyme biosensor was used, the highest signal could be seen in JFYL007-derived strains. This was expected since the strains cannot accumulate lipids due to gene deletions. Therefore, the malonyl-CoA available will be directed towards producing the biosensor molecule instead of fatty acids. Interestingly, when two copies of the biosensor enzyme were introduced, the strain based on OKYL049 showed the highest signal. This is probably due to an increased flux through malonyl-CoA from β-oxidation of stored TAGs. Addition of cerulenin had seemingly no to little effect on increasing the biosensor signal, which indicates that the biosensor did not work as expected. The continued research could instead focus on attempting to maximise the production of the biosensor reporter molecule. As the enzyme repurposed as the biosensor is a PKS, the product is a polyketide and could therefore be of interest to produce. The first step towards developing a production strategy will be identification of the product itself and the pathway towards it. This will open up for finding gene targets for metabolic engineering using computational tools.

Second, the expression strength of a set of native, uncharacterised promoters was compared in different conditions. From a previous, unpublished project, the genes under the aforementioned promoters were picked specifically due to differential expression depending on nitrogen availability and source. To test the full regulation profile of the promoters in the different conditions, a new experiment was set up. A gene for the fluorescent protein GFP was placed under control of the promoters, as the fluorescence intensity of each strain was defined as the expression strength of each promoter. No fluorescence was measured in any of the strains containing the uncharacterised promoters, only in the positive control. Several troubleshooting steps were taken, such as verifying the gene cassette in the yeast and evaluating the cloning strategy. None of them resulted in an explanation to the unexpected results. It is possible that vital transcriptional regulatory elements for the promoters were present in the original genomic locations, as only the core promoters were tested in the project. Further research should focus on finding a more suitable and brighter fluorescent protein to increase the lower detection range. A more sensitive instrument as a flow cytometer could also be used.

Bibliography

[1] Ko YS, Kim JW, Lee JA, Han T, Kim GB, Park JE, et al. Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. Chemical Society Reviews. 2020;49(14):4615–4636.

[2] Kavšček M, Stražar M, Curk T, Natter K, Petrovič U. Yeast as a cell factory: current state and perspectives. Microbial cell factories. 2015;14(1):1–10.

[3] Mattanovich D, Sauer M, Gasser B. Yeast biotechnology: teaching the old dog new tricks. Microbial cell factories. 2014;13(1):1–5.

[4] Ferrer-Miralles N, Villaverde A. Bacterial cell factories for recombinant protein production; expanding the catalogue. Microbial cell factories. 2013;12(1):1–4.

[5] Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MMM, Ramadan HA, et al. Cell factories for insulin production. Microbial cell factories. 2014;13(1):1–9.

[6] Abghari A, Chen S. Yarrowia lipolytica as an oleaginous cell factory platform for production of fatty acid-based biofuel and bioproducts. Frontiers in Energy Research. 2014;2:21.

[7] Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. Nature. 2004;430(6995):35–44.

[8] Darvishi F, Ariana M, Marella ER, Borodina I. Advances in synthetic biology of oleaginous yeast Yarrowia lipolytica for producing non-native chemicals. Applied microbiology and biotechnology. 2018;102(14):5925–5938.

[9] Wong L, Engel J, Jin E, Holdridge B, Xu P. YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in Yarrowia lipolytica. Metabolic engineering communications. 2017;5:68–77.

[10] Larroude M, Rossignol T, Nicaud JM, Ledesma-Amaro R. Synthetic biology tools for engineering Yarrowia lipolytica. Biotechnology Advances. 2018 dec;36(8):2150–2164.

[11] Yang D, Kim WJ, Yoo SM, Choi JH, Ha SH, Lee MH, et al. Repurposing type III polyketide synthase as a malonyl- CoA biosensor for metabolic engineering in bacteria. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(40):9835–9844.

[12] Nicaud JM. Yarrowia lipolytica. Yeast. 2012 oct;29(10):409–418. Available from: `https://onlinelibrary.wiley.com/doi/full/10.1002/yea.2921`.

[13] Barth G, Gaillardin C. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS microbiology reviews. 1997;19(4):219–237.

[14] CDC, NIH. Biosafety in Microbiological and Biomedical Laboratories 6th Edition Centers for Disease Control and Prevention National Institutes of Health; 2020.

[15] Miller KK, Alper HS. Yarrowia lipolytica: more than an oleaginous workhorse. Applied Microbiology and Biotechnology. 2019;103(23-24):9251–9262.

[16] Egermeier M, Russmayer H, Sauer M, Marx H. Metabolic flexibility of Yarrowia lipolytica growing on glycerol. Frontiers in microbiology. 2017;8:49.

[17] Konzock O, Zaghen S, Norbeck J. Tolerance of Yarrowia lipolytica to inhibitors commonly found in lignocellulosic hydrolysates. BMC microbiology. 2021;21(1):1–10.

[18] Bao W, Li Z, Wang X, Gao R, Zhou X, Cheng S, et al.. Approaches to improve the lipid synthesis of oleaginous yeast Yarrowia lipolytica: A review. Pergamon; 2021.

[19] Markham KA, Alper HS. Synthetic biology expands the industrial potential of Yarrowia lipolytica. Trends in biotechnology. 2018;36(10):1085–1095.

[20] Kerkhoven EJ, Pomraning KR, Baker SE, Nielsen J. Regulation of amino-acid metabolism controls flux to lipid accumulation in yarrowia lipolytica. npj Systems Biology and Applications. 2016 mar;2(1):1–7. Available from: `https://www.nature.com/articles/npjsba20165`.

[21] Berg JM, Tymoczko JL, Gatto, Jr L Gregory J Stryer. Biochemistry. 8th ed. New York: W.H. Freeman And Company; 2015.

[22] Zhang H, Zhang L, Chen H, Chen YQ, Chen W, Song Y, et al. Enhanced lipid accumulation in the yeast Yarrowia lipolytica by over-expression of ATP: citrate lyase from Mus musculus. Journal of biotechnology. 2014;192:78–84.

[23] Ledesma-Amaro R, Nicaud JM. Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. Progress in lipid research. 2016;61:40–50.

[24] Beopoulos A, Chardot T, Nicaud JM. Yarrowia lipolytica: A model and a tool to understand the mechanisms implicated in lipid accumulation. Biochimie. 2009;91(6):692–696.

[25] Müller S, Sandal T, Kamp-Hansen P, Dalbøge H. Comparison of expression systems in the yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of two novel promoters from Yarrowia lipolytica. Yeast. 1998;14(14):1267–1283.

[26] Tai M, Stephanopoulos G. Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. Metabolic engineering. 2013;15:1–9.

[27] Le Hir H, Nott A, Moore MJ. How introns influence and enhance eukaryotic gene expression. Trends in biochemical sciences. 2003;28(4):215–220.

[28] Dulermo R, Brunel F, Dulermo T, Ledesma-Amaro R, Vion J, Trassaert M, et al. Using a vector pool containing variable-strength promoters to optimize protein production in Yarrowia lipolytica. Microbial Cell Factories. 2017;16(1):1–11.

[29] Kastberg LLB, Ard R, Jensen MK, Workman CT. Burden Imposed by Heterologous Protein Production in Two Major Industrial Yeast Cell Factories: Identifying Sources and Mitigation Strategies. Front Fungal Biol. 2022;3(1).

[30] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature methods. 2009;6(5):343–345.

[31] Kretzschmar A, Otto C, Holz M, Werner S, Hübner L, Barth G. Increased homologous integration frequency in Yarrowia lipolytica strains defective in non-homologous end-joining. Current genetics. 2013;59(1):63–72.

[32] Verbeke J, Beopoulos A, Nicaud JM. Efficient homologous recombination with short length flanking fragments in Ku70 deficient Yarrowia lipolytica strains. Biotechnology letters. 2013;35(4):571–576.

[33] Schwartz CM, Hussain MS, Blenner M, Wheeldon I. Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR–Cas9-mediated genome editing in Yarrowia lipolytica. ACS synthetic biology. 2016;5(4):356–359.

[34] Holkenbrink C, Dam MI, Kildegaard KR, Beder J, Dahlin J, Doménech Belda D, et al. EasyCloneYALI: CRISPR/Cas9-Based Synthetic Toolbox for Engineering of the Yeast Yarrowia lipolytica. Biotechnology Journal. 2018 sep;13(9):1700543. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/biot.201700543.

[35] Marella ER, Dahlin J, Dam MI, Ter Horst J, Christensen HB, Sudarsan S, et al. A single-host fermentation process for the production of flavor lactones from non-hydroxylated fatty acids. Metabolic engineering. 2020;61:427–436.

[36] Katz L, Baltz RH. Natural product discovery: past, present, and future; 2016. Available from: https://academic.oup.com/jimb/article/43/2-3/155/5995722.

[37] Staunton J, Weissman KJ. Polyketide biosynthesis: A millennium review. Natural Product Reports. 2001;18(4):380–416.

[38] Miyanaga A. Structure and function of polyketide biosynthetic enzymes: various strategies for production of structurally diverse polyketides. Bioscience, Biotechnology, and Biochemistry. 2017;81(12):2227–2236.

[39] Muhammad A, Feng X, Rasool A, Sun W, Li C. Production of plant natural products through engineered Yarrowia lipolytica. Biotechnology Advances. 2020;43(May):107555. Available from: https://doi.org/10.1016/j.biotechadv.2020.107555.

[40] Markham KA, Palmer CM, Chwatko M, Wagner JM, Murray C, Vazquez S, et al. Rewiring yarrowia lipolytica toward triacetic acid lactone for materials generation. Proceedings of the National Academy of Sciences of the United States of America. 2018;115(9):2096–2101.

[41] De Luca V, Salim V, Atsumi SM, Yu F. Mining the biodiversity of plants: a revolution in the making. Science. 2012;336(6089):1658–1661.

[42] Palmer CM, Miller KK, Nguyen A, Alper HS. Engineering 4-coumaroyl-CoA derived polyketide production in Yarrowia lipolytica through a β-oxidation mediated strategy. Metabolic Engineering. 2020;57(August 2019):174–181. Available from: https://doi.org/10.1016/j.ymben.2019.11.006.

[43] Nikhil B, Pawan J, Nello F, Pedro E. Introduction to biosensors. Essays Biochem. 2016;60(1):1–8.

[44] Turner A, Karube I, Wilson GS. Biosensors: fundamentals and applications. Oxford university press; 1987.

[45] Quijano-Rubio A, Yeh HW, Park J, Lee H, Langan RA, Boyken SE, et al. De novo design of modular and tunable protein biosensors. Nature. 2021;591(7850):482–487.

[46] Omura S. The antibiotic cerulenin, a novel tool for biochemistry as an inhibitor of fatty acid synthesis. Bacteriological Reviews. 1976 sep;40(3):681–697. Available from: https://journals.asm.org/doi/abs/10.1128/br.40.3.681-697.1976.

[47] Price AC, Choi KH, Heath RJ, Li Z, White SW, Rock CO. Inhibition of β-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin: structure and mechanism. Journal of Biological Chemistry. 2001;276(9):6551–6559.

[48] Lv Y, Marsafari M, Koffas M, Zhou J, Xu P. Optimizing Oleaginous Yeast Cell Factories for Flavonoids and Hydroxylated Flavonoids Biosynthesis. ACS Synthetic Biology. 2019;8(11):2514–2523. Available from: https://doi.org/10.1016/j.mec.2019.e00121.

[49] Yu J, Landberg J, Shavarebi F, Bilanchone V, Okerlund A, Wanninayake U, et al. Bioengineering triacetic acid lactone production in Yarrowia lipolytica for pogostone synthesis. Biotechnology and Bioengineering. 2018;115(9):2383–2388.

[50] Liu H, Marsafari M, Wang F, Deng L, Xu P. Engineering acetyl-CoA metabolic shortcut for eco-friendly production of polyketides triacetic acid lactone in Yarrowia lipolytica. Metabolic Engineering. 2019 dec;56:60–68.

[51] Liu H, Wang F, Deng L, Xu P. Genetic and bioprocess engineering to improve squalene production in Yarrowia lipolytica. Bioresource technology. 2020;317:123991.

[52] Gaillardin CM, Charoy V, Heslot H. A study of copulation, sporulation and meiotic segregation in Candida lipolytica. Archiv für Mikrobiologie. 1973;92(1):69–83.

[53] Konzock O, Norbeck J. Deletion of MHY1 abolishes hyphae formation in Yarrowia lipolytica without negative effects on stress tolerance. PloS one. 2020;15(4):e0231161.

[54] Beopoulos A, Haddouche R, Kabran P, Dulermo T, Chardot T, Nicaud JM. Identification and characterization of DGA2, an acyltransferase of the DGAT1 acyl-CoA: diacylglycerol acyltransferase family in the oleaginous yeast Yarrowia lipolytica. New insights into the storage lipid metabolism of oleaginous yeasts. Applied microbiology and biotechnology. 2012;93(4):1523–1537.

[55] Pignède G, Wang H, Fudalej F, Gaillardin C, Seman M, Nicaud JM. Characterization of an extracellular lipase encoded by LIP2 in Yarrowia lipolytica. Journal of bacteriology. 2000;182(10):2802–2810.

[56] Marsafari M, Xu P. Debottlenecking mevalonate pathway for antimalarial drug precursor amorphadiene biosynthesis in Yarrowia lipolytica. Metabolic Engineering Communications. 2020;10. Available from: https://doi.org/10.1016/j.mec.2019.e00121.

[57] Bizzarri R, Serresi M, Luin S, Beltram F. Green fluorescent protein based pH indicators for in vivo use: A review. Springer; 2009. Available from: https://link.springer.com/article/10.1007/s00216-008-2515-9.

[58] Orij R, Postmus J, Beek AT, Brul S, Smits GJ. In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in Saccharomyces cerevisiae reveals a relation between intracellular pH and growth. Microbiology. 2009 jan;155(1):268–278. Available from: https://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.022038-0.

[59] Alexandrova AN, Jorgensen WL. Why urea eliminates ammonia rather than hydrolyzes in aqueous solution. The Journal of Physical Chemistry B. 2007;111(4):720–730.

[60] Lambert TJ. FPbase: a community-editable fluorescent protein database. Nature methods. 2019;16(4):277–278.

A

Appendix

A.1 Appendix figures

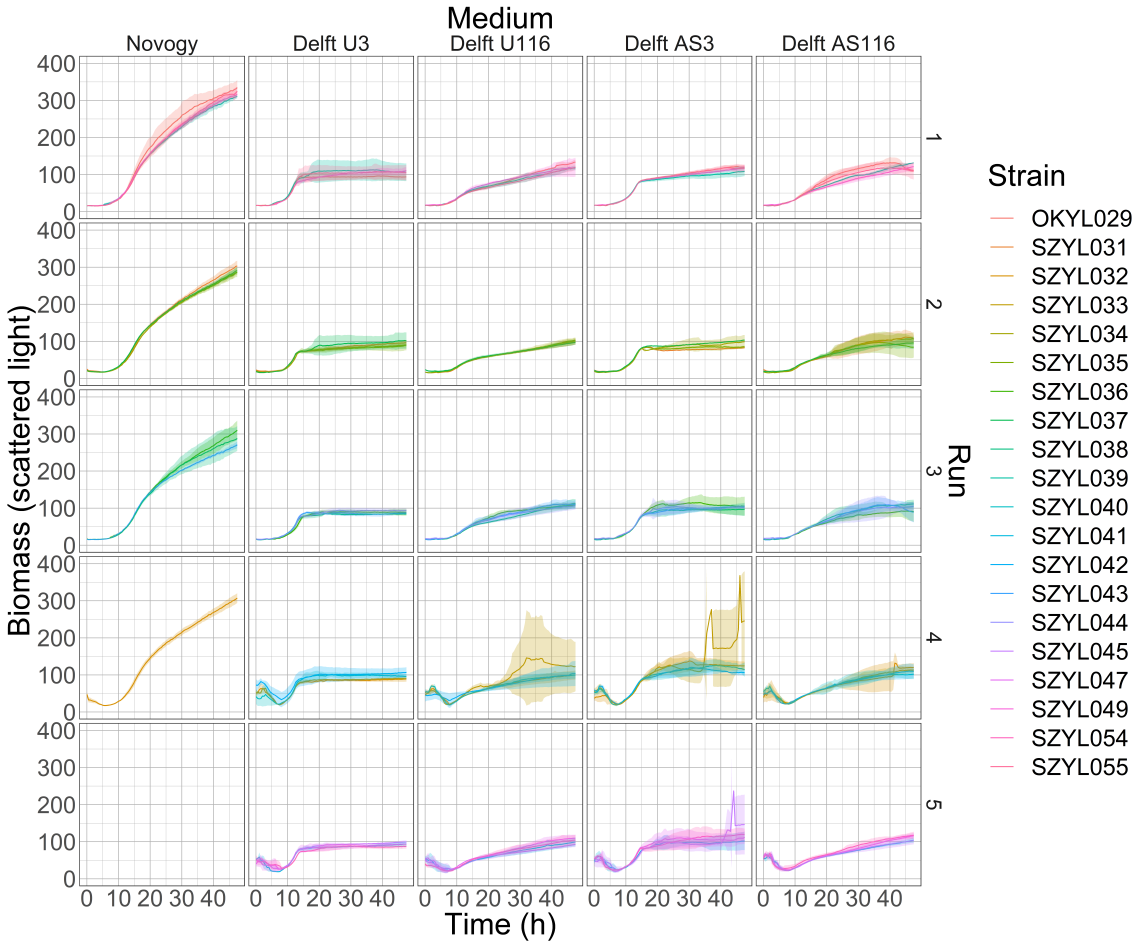


Figure A.1: Biomass over time of OKYL029, SZYL055, and 18 promoter strains grown in five different media. OKYL029 is negative control, SZYL055 is positive control, and SZYL031-SZYL045, SZYL047, SZYL048, and SZYL054 are test strains. Strains were cultivated in a 96-well microplate in BioLector for 48 h. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The graphs are divided by run and medium. The line colour represents strain. The data shown represent average of technical quadruplicates and the shadow represents standard deviation. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

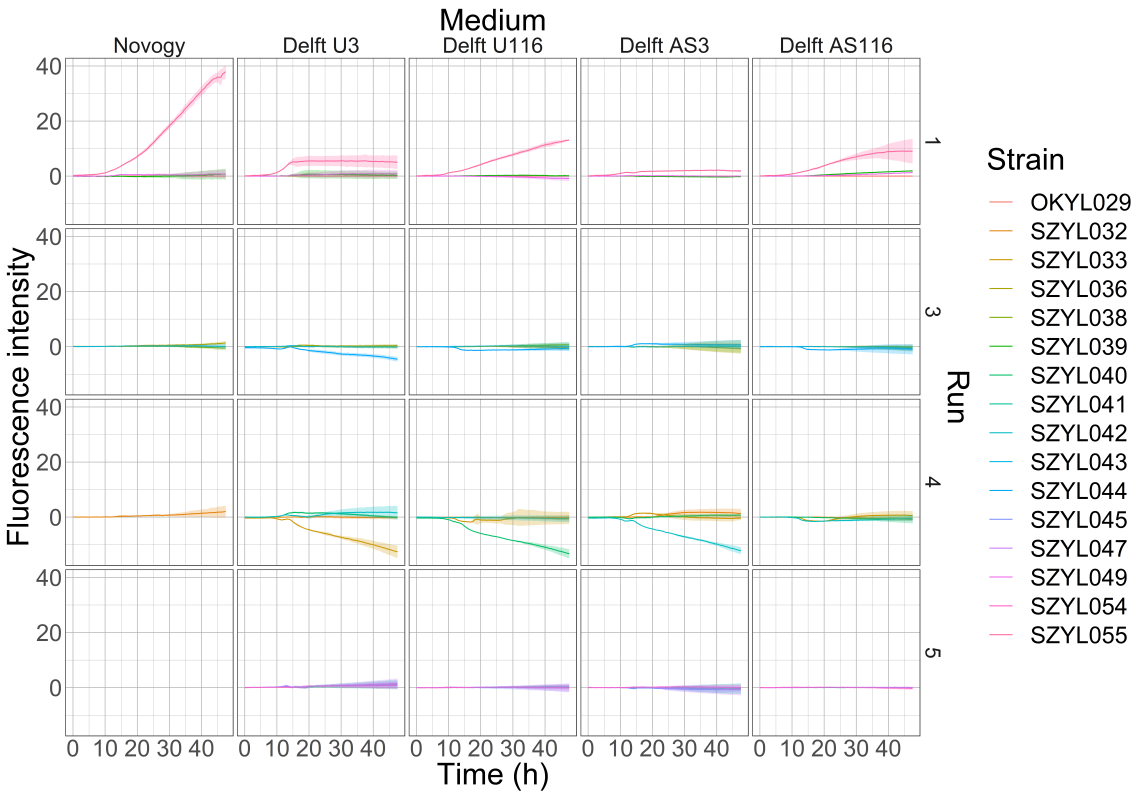


Figure A.2: Fluorescence intensity over time of OKYL029, SZYL055, and 18 promoter strains grown in five different media. OKYL029 is negative control, SZYL055 is positive control, and SZYL031-SZYL045, SZYL047, SZYL048, and SZYL054 are test strains. Strains were cultivated in a 96-well microplate in BioLector for 48 h. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The graphs are divided by run and medium. The data from run 2 is not shown since the negative control strain OKYL029 was not present in that run, meaning that the auto-fluorescence of the cells could not be removed for the promoter strains in that run. The line colour represents strain. The data shown represent average of technical quadruplicates and the shadow represents standard deviation. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

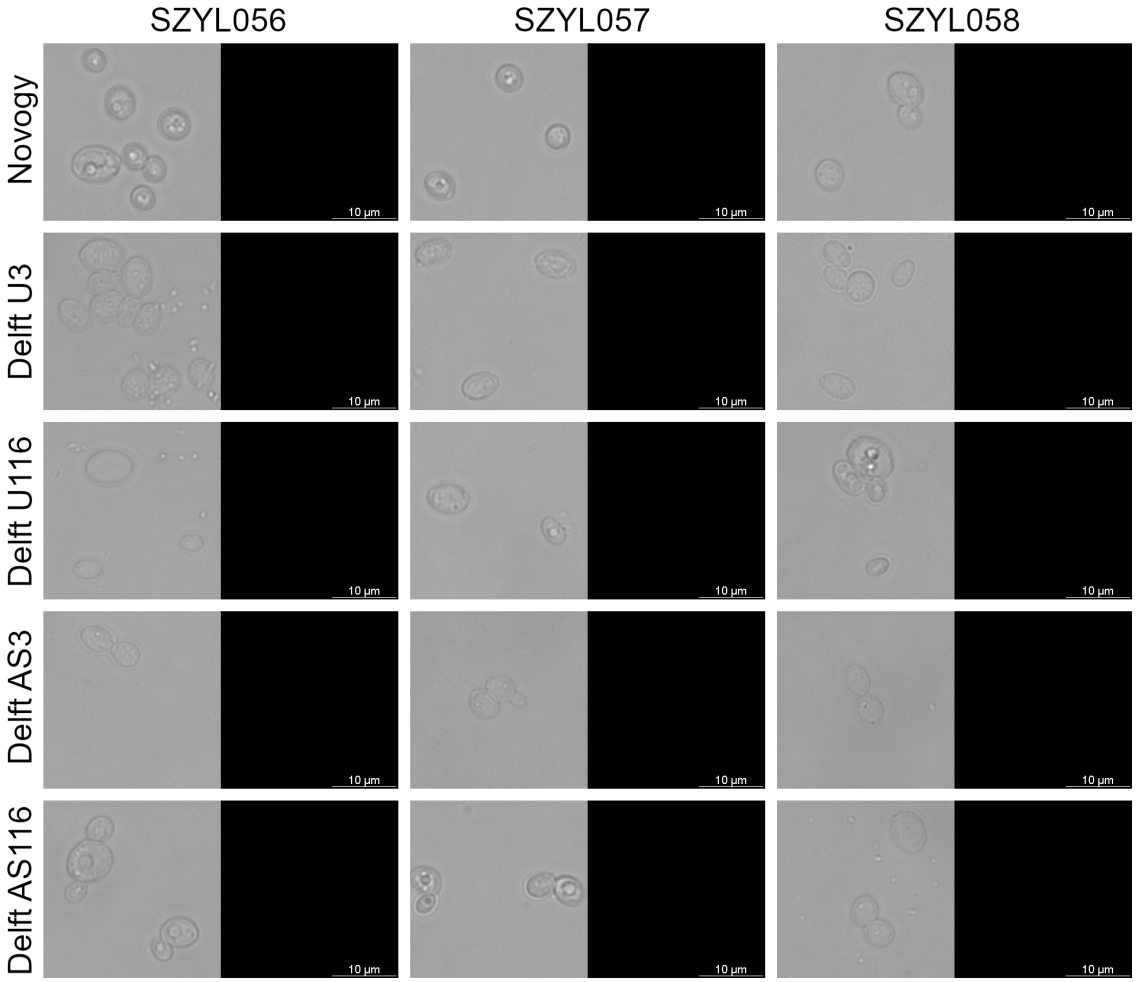


Figure A.3: Fluorescence microscopy images of SZYL056, SZYL057, and SZYL058 grown for 24 h in five different media. SZYL056 has p*TEF1*, SZYL057 has p*PYK1*, and SZYL058 has p*GAPDH*. Strains were cultivated in shake flasks for 24 h, then photographed in fluorescence microscope in brightfield and GFP channel. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The images are divided by medium and strain. Scale bar is 10 µm. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

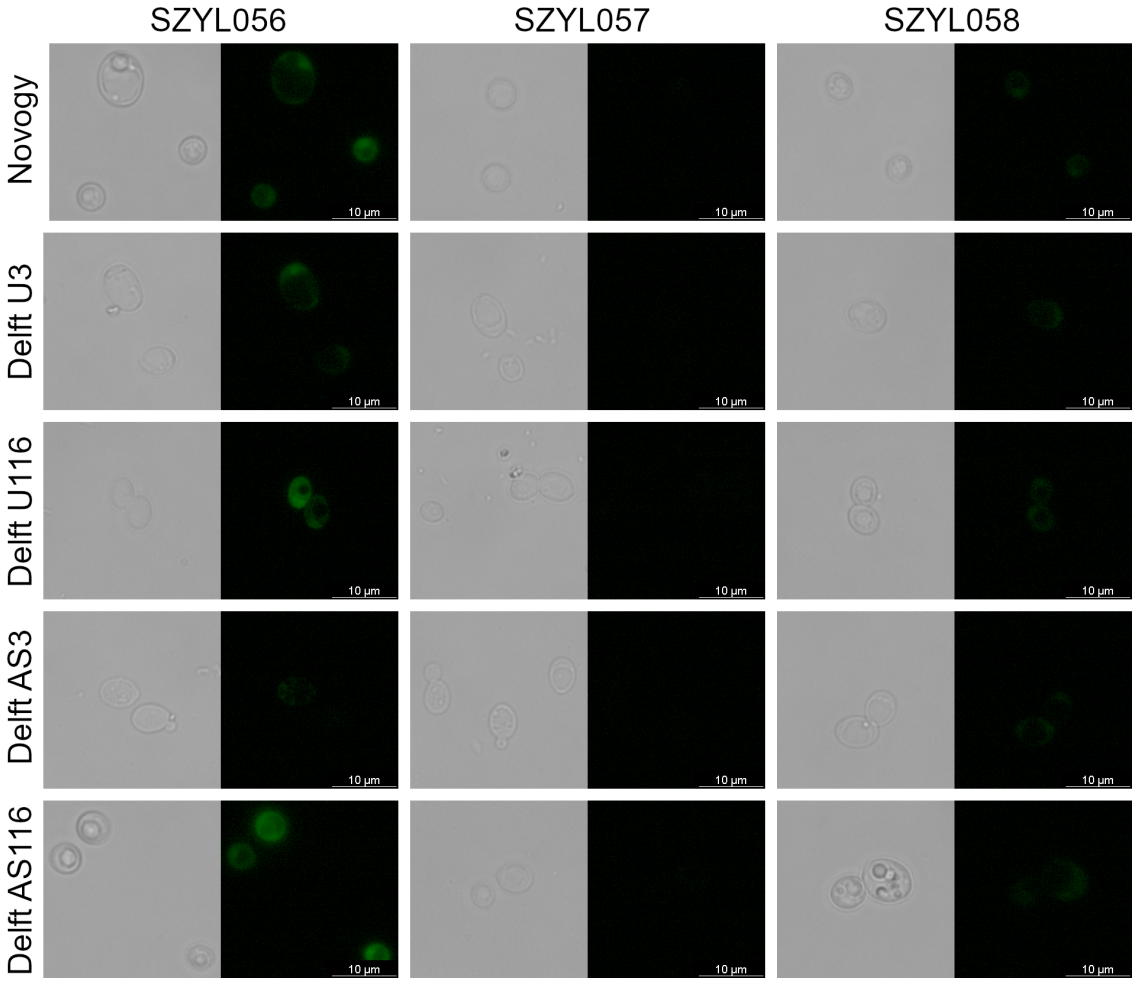


Figure A.4: Fluorescence microscopy images of SZYL056, SZYL057, and SZYL058 grown for 72 h in five different media. SZYL056 has p*TEF1*, SZYL057 has p*PYK1*, and SZYL058 has p*GAPDH*. Strains were cultivated in shake flasks for 72 h, then photographed in fluorescence microscope in brightfield and GFP channel. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The images are divided by medium and strain. Scale bar is 10 µm. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

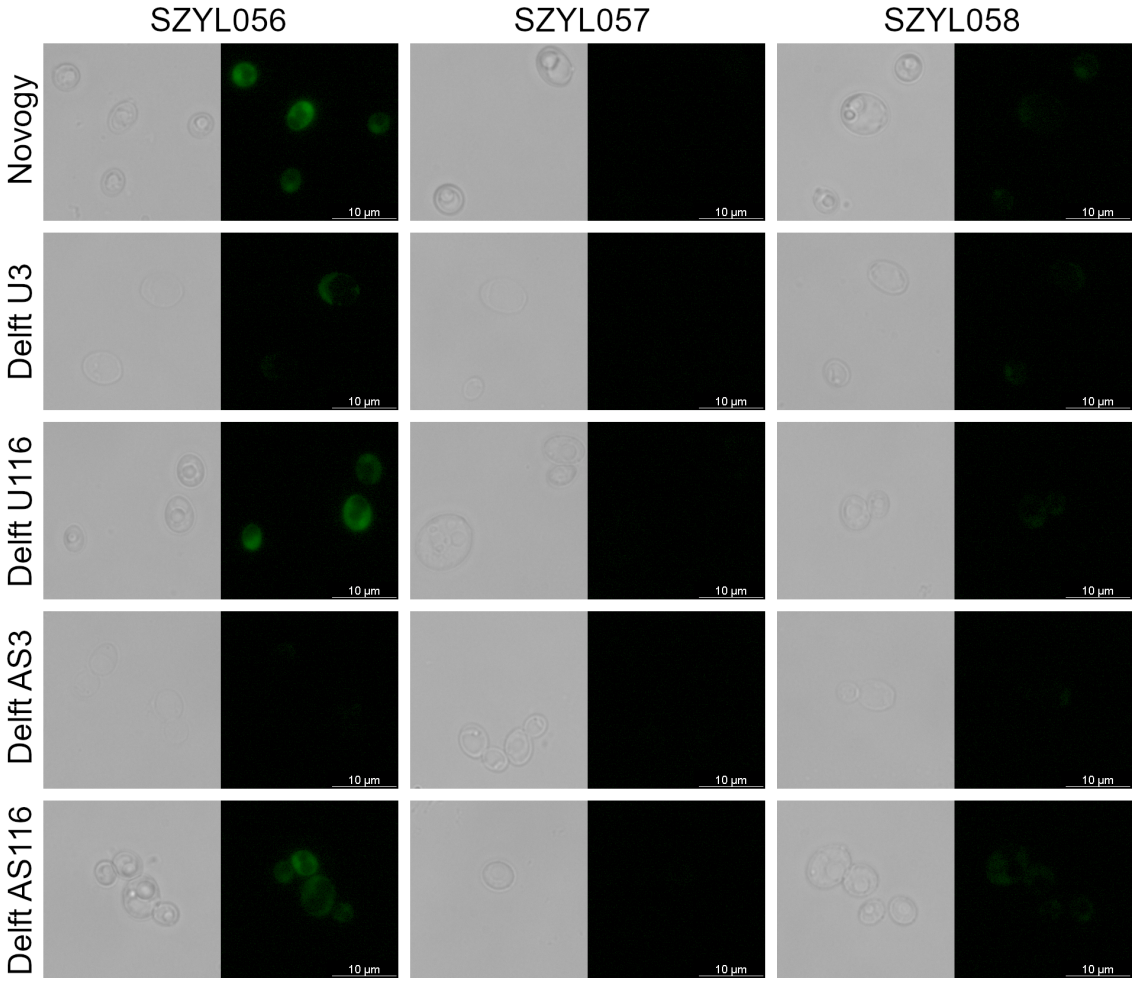


Figure A.5: Fluorescence microscopy images of SZYL056, SZYL057, and SZYL058 grown for 96 h in five different media. SZYL056 has p*TEF1*, SZYL057 has p*PYK1*, and SZYL058 has p*GAPDH*. Strains were cultivated in shake flasks for 96 h, then photographed in fluorescence microscope in brightfield and GFP channel. The cultivation media were Novogy, Delft with urea C/N 3 or 116, and Delft with ammonium sulphate C/N 3 or 116. The images are divided by medium and strain. Scale bar is 10 µm. Abbreviations: U3, urea C/N 3; U116, urea C/N 116; AS3, ammonium sulphate C/N 3; AS116, ammonium sulphate C/N 116.

A.2 Appendix tables

Table A.1: Plasmids used in the thesis projects with description and reference.

Plasmid	Description	Reference
pCfB6630	gRNA vector for *Y. lipolytica* genome integration site C3	Holkenbrink et al. [34]
pCfB6371	Integration vector for *Y. lipolytica* genome integration site C3	Holkenbrink et al. [34]
pCfB6631	gRNA vector for *Y. lipolytica* genome integration site D1	Holkenbrink et al. [34]
pCfB6684	Integration vector for *Y. lipolytica* genome integration site D1	Holkenbrink et al. [34]
SZEC27	pCfB6684 p*TEF1-AaOKS*	(Unpublished)
SZEC28	pCfB6684 p*TEF1-AaPKS4*	(Unpublished)
SZEC29	pCfB6684 p*TEF1-AaPKS5*	(Unpublished)
HOEC03	pCfB6371 p*TEF1-AaOKS*	This study
JFEC133	pCfB6684 p*TEF1-hrGFP*	(Unpublished)
SZEC31	JFEC133Δp*tef1* p*YALI1_A20484g*	This study
SZEC32	JFEC133Δp*tef1* p*YALI1_B05639g*	This study
SZEC33	JFEC133Δp*tef1* p*YALI1_C01505g*	This study
SZEC34	JFEC133Δp*tef1* p*YALI1_C13611g*	This study
SZEC35	JFEC133Δp*tef1* p*YALI1_C32392g*	This study
SZEC36	JFEC133Δp*tef1* p*YALI1_D15557g*	This study
SZEC37	JFEC133Δp*tef1* p*YALI1_E01694g*	This study
SZEC38	JFEC133Δp*tef1* p*YALI1_E28070g*	This study
SZEC39	JFEC133Δp*tef1* p*YALI1_E31154g*	This study
SZEC40	JFEC133Δp*tef1* p*YALI1_E32241g*	This study
SZEC41	JFEC133Δp*tef1* p*YALI1_F29596g*	This study
SZEC42	JFEC133Δp*tef1* p*YALI1_A00911g*	This study
SZEC43	JFEC133Δp*tef1* p*YALI1_A21509g*	This study
SZEC44	JFEC133Δp*tef1* p*YALI1_C02160g*	This study
SZEC45	JFEC133Δp*tef1* p*YALI1_C24183g*	This study
SZEC46	JFEC133Δp*tef1* p*YALI1_C26533g*	This study
SZEC47	JFEC133Δp*tef1* p*YALI1_C30910g*	This study
SZEC48	JFEC133Δp*tef1* p*YALI1_D01497g*	This study
SZEC49	JFEC133Δp*tef1* p*YALI1_E02850g*	This study
SZEC50	JFEC133Δp*tef1* p*YALI1_F02403g*	This study
SZEC51	JFEC133Δp*tef1* p*YALI1_C01520g*	This study
SZEC52	JFEC133Δp*tef1* p*YALI1_E05779g*	This study
SZEC53	JFEC133Δp*tef1* p*YALI1_E09895g*	This study
SZEC54	JFEC133Δp*tef1* p*YALI1_E32180g*	This study
SZEC55	JFEC133Δp*tef1* p*TEF1*	This study
SZEC56	JFEC133Δp*tef1* p*PYK1*	This study
SZEC57	JFEC133Δp*tef1* p*GADPH*	This study

Table A.2: Primers used in the thesis projects for PCR amplification of DNA fragments.

Primer	Sequence
GFP_FW	ATGGTGTCCAAGCAGATTCTC
Int_D1_RV	GACATTGCTCTAACACTCTGCATTG
YALI1_A00911g_long_FW	AATGCAGAGTGTTAGAGCAATGTCGAGATTATGACTGAA ATATCCGCAGG
YALI1_A00911g_long_RV	GTGTTCTTGAGAATCTGCTTGGACACCATCCTAAGTCTG CTTCACTCAAG
YALI1_A20484g_FW	CAATGCAGAGTGTTAGAGCAATGTCACGCAATTCCGAGA AGTCCC
YALI1_A20484g_RV	TCTTGAGAATCTGCTTGGACACCATTTTTAGAGAGGATG AGAGATG
YALI1_A21509g_FW	CAATGCAGAGTGTTAGAGCAATGTCCTTCAGGCCGAGAT AACCAAG
YALI1_A21509g_RV	TCTTGAGAATCTGCTTGGACACCATGGTTGGCGTTGTGG TGGTATG
YALI1_B05639g_FW	CAATGCAGAGTGTTAGAGCAATGTCTTGCCACCAAAGTG TTGTCG
YALI1_B05639g_RV	TCTTGAGAATCTGCTTGGACACCATATGTAAGCAGTAGA ATTAGC
YALI1_B12489g_FW	CAATGCAGAGTGTTAGAGCAATGTCGCCAATGGGGGTTT CGACATG
YALI1_B12489g_RV	TCTTGAGAATCTGCTTGGACACCATTGGGAGAGAGAAG GGTGGAAAG
YALI1_B13614g_FW	CAATGCAGAGTGTTAGAGCAATGTCGCGGTGCATTCCAA GGAGGG
YALI1_B13614g_RV	TCTTGAGAATCTGCTTGGACACCATTAGTGACAGAAGTG GATGGTG
YALI1_C01505g_FW	CAATGCAGAGTGTTAGAGCAATGTCGCTGTGGATGGCAC CTCTCTG
YALI1_C01505g_RV	TCTTGAGAATCTGCTTGGACACCATGTCTAGTTGTAACG GAGTTG
YALI1_C01520g_FW	CAATGCAGAGTGTTAGAGCAATGTCGGAGAAGAAAGAA TCTACAG
YALI1_C01520g_RV	TCTTGAGAATCTGCTTGGACACCATTTTGTGGTGGCGGT TTTTCAC
YALI1_C02160g_FW	CAATGCAGAGTGTTAGAGCAATGTCCATGTTGGAATCGA ATTTAG
YALI1_C02160g_RV	TCTTGAGAATCTGCTTGGACACCATTGATTAAAAAACGA TTTTCACCTTTCC
YALI1_C13611g_FW	CAATGCAGAGTGTTAGAGCAATGTCGTACGCCAACATGG ATATCATC
YALI1_C13611g_RV	TCTTGAGAATCTGCTTGGACACCATCGTTCTCAGTGTCG TTACTTATATC
YALI1_C24183g_FW	CAATGCAGAGTGTTAGAGCAATGTCTCTCTTGCCCCACT TAATCC
YALI1_C24183g_RV	TCTTGAGAATCTGCTTGGACACCATCAGCAAACAGCCTC AGCGTTG
YALI1_C26533g_FW	CAATGCAGAGTGTTAGAGCAATGTCACGAATGAGGTTAA GCGGCG
YALI1_C26533g_RV	TCTTGAGAATCTGCTTGGACACCATTGGCGTGCTTGAAG AACCAG
YALI1_C30910g_FW	CAATGCAGAGTGTTAGAGCAATGTCAGTCTAGGTGGAAT ATAGCATAATAG
YALI1_C30910g_RV	TCTTGAGAATCTGCTTGGACACCATCTAGTTACTTATCT ATATATCTAAGTATTGTTATC
YALI1_C32392g_FW	CAATGCAGAGTGTTAGAGCAATGTCCTATACAACTTGCC TCAGGC
YALI1_C32392g_RV	TCTTGAGAATCTGCTTGGACACCATCTGGGTTAGTGGGG AAGGATTTTG
YALI1_D01497g_FW	CAATGCAGAGTGTTAGAGCAATGTCGTACAATATTACAG ATAACCGTGC
YALI1_D01497g_RV	TCTTGAGAATCTGCTTGGACACCATCACCTCGATTATCA GAGCTC

Continued on following page

YALI1_D15557g_FW	CAATGCAGAGTGTTAGAGCAATGTCCAATGGTCTACTGTGTGCAGACC
YALI1_D15557g_RV	TCTTGAGAATCTGCTTGGACACCATCGTCTAATTATCTATACCAG
YALI1_E01694g_FW	CAATGCAGAGTGTTAGAGCAATGTCTACCACCTCGAGAAGCAAATAC
YALI1_E01694g_RV	TCTTGAGAATCTGCTTGGACACCATCGTGTTATGACGTAGAACTC
YALI1_E02850g_FW	CAATGCAGAGTGTTAGAGCAATGTCCCGAAAAGAGTGTGCCCCTTTG
YALI1_E02850g_RV	TCTTGAGAATCTGCTTGGACACCATTTGTGTAATGTGCGAAAGCTATG
YALI1_E05779g_FW	CAATGCAGAGTGTTAGAGCAATGTCTGCTATTGAGGTCATGGGAATG
YALI1_E05779g_RV	TCTTGAGAATCTGCTTGGACACCATGTACCACAGTTGAACCAGAATC
YALI1_E09895g_FW	CAATGCAGAGTGTTAGAGCAATGTCTGAGGTTCGCCGTAGATCATC
YALI1_E09895g_RV	TCTTGAGAATCTGCTTGGACACCATCTCTCGACTCGGTTATTACTTC
YALI1_E28070g_FW	CAATGCAGAGTGTTAGAGCAATGTCGATAAGGACAGATGTGGTCATAG
YALI1_E28070g_RV	TCTTGAGAATCTGCTTGGACACCATGACTGAGTTGGAGACGGTCTG
YALI1_E31154g_FW	CAATGCAGAGTGTTAGAGCAATGTCTTGTGTAAGCACATCTCGTTC
YALI1_E31154g_RV	TCTTGAGAATCTGCTTGGACACCATCTTGAGATGGCTATTGTGGC
YALI1_E32180g_FW	CAATGCAGAGTGTTAGAGCAATGTCCATCCCTATCTCCTGTTGCTAC
YALI1_E32180g_RV	TCTTGAGAATCTGCTTGGACACCATTTGTGAATTAGGGTGGTGAG
YALI1_E32241g_FW	CAATGCAGAGTGTTAGAGCAATGTCGCGTCTCGGGATCTGCATGTTG
YALI1_E32241g_RV	TCTTGAGAATCTGCTTGGACACCATAGATGATTGTTTAATGTGTTGGATGG
YALI1_F02403g_FW	CAATGCAGAGTGTTAGAGCAATGTCATTGTCGGCCAGACGTACGC
YALI1_F02403g_RV	TCTTGAGAATCTGCTTGGACACCATGTTGGGTTGTCTGGAAGGTC
YALI1_F23616g_FW	CAATGCAGAGTGTTAGAGCAATGTCGTACGCATCAGATACACACAG
YALI1_F23616g_RV	TCTTGAGAATCTGCTTGGACACCATTGTGTGAGGTGGTGTTCGGAG
YALI1_F29596g_FW	CAATGCAGAGTGTTAGAGCAATGTCGGATTATGGTGATGTAATCTG
YALI1_F29596g_RV	TCTTGAGAATCTGCTTGGACACCATGATTGTTGAGGCGGTGGTGAC
pTEF1_amp_FW	CAATGCAGAGTGTTAGAGCAATGTCAGAGACCGGGTTGGCGGCGCATTTG
pTEF1_amp_RV	TCTTGAGAATCTGCTTGGACACCATTTTGAATGATTCTTATACTCAGAAG
pPYK1_FW	CAATGCAGAGTGTTAGAGCAATGTCTGCGCGCTTCTGTTGTTTGCAACC
pPYK1_RV	TCTTGAGAATCTGCTTGGACACCATTGTAACTGTGGTGTGAATTTCTCCG
pGAPDH_FW	CAATGCAGAGTGTTAGAGCAATGTCGGTTGAAATGAATCGGCCGACGCTC
pGAPDH_RV	TCTTGAGAATCTGCTTGGACACCATTGTTGATGTGTGTTTAATTCAAG
TLip_2_RV	AGGTTGATTCCGAACAGAAG
TPex20_2_FW	CTCACTTCCCCATCCACAC

Table A.3: Sequences for the promoters, terminators, and genes used in the thesis projects.

Name	Sequence
p*TEF1*	AGAGACCGGGTTGGCGGCGCATTTGTGTCCCAAAAAACAGCCCCA ATTGCCCCAATTGACCCCAAATTGACCCAGTAGCGGGCCCAACCC CGGCGAGAGCCCCCTTCTCCCCACATATCAAACCTCCCCCGGTTCC CACACTTGCCGTTAAGGGCGTAGGGTACTGCAGTCTGGAATCTAC GCTTGTTCAGACTTTGTACTAGTTTCTTTGTCTGGCCATCCGGGT AACCCATGCCGGACGCAAAATAGACTACTGAAAATTTTTTTGCTT TGTGGTTGGGACTTTAGCCAAGGGTATAAAAGACCACCGTCCCCG AATTACCTTTCCTCTTCTTTTCTCTCTCTCCTTGTCAACTCACACC CGAAATCGTTAAGCATTTCCTTCTGAGTATAAGAATCATTCAAA
p*PYK1*	TGCGCGCTTCTGTTGTTTGCAACCAGTCTCCGTGGTGTGCAGCAT ACATTGTTCCCGCCTCTCCTTGTCTTGTTGGAAGGCCGATGTCGC TGACTGTATGTACCGTTTTTTTTGTACCGTAGTACATGCAGGGCT TGGTATTTTCCAACTACAGTACATACAGGTCTTAGAGTGCTGATT GGAGATAGATATGAATGGAGTGTACGAGTGGAAACAAAGCGGGTT AGATATGTGTACTTGTACATCTGTGATATTGGTAGTATTGACAAG CGGTAGTCATTTCAGTGCATCGCCGTGCCCTTTCTACTATCCCCTT GCGCCATCAATCTCCCCCTTCATCAATCCACCTCTGGCAGCTCTTC TAGAAGACCTTTTTACAGTCTCCCAATTTTATCGTCTAGTGACGG CAGACCTTGTAAGCAGATATGTATCATGAGTCACGATAGCTGGAC AGACCAATGGCATGCGGGCAAATAACTCCCACAGACGCTCTCCCT CCGGCGCACAAAGCCTCGTGCTCTGAACACGCCCCAGTTGATTTG ACAGCTCTCAACATTCGTGTGAACTTTTTTAGCGGGAAAAAGTAA CATGACGTTGACCGTGCGGGGCTACATGTAGCAGCTGGGTGTGCT AACTACGGATACATGCCTACAACCCCCACAAGTCAAGACCATTGC GACGCGGAAACAGGAGCCCGCAAAAGAGGAGAAAAACAACGGCGA GACTCGGGGGCGGAGTGGGTCACGTGACTTTCCTTTTTCCCCTCA CCTGGCCCGCTCCGTCCATATCTCTGTCGTACAAGACAATATTGTC GCAACGCAAAAGGTCCATAAATTACTGGGTAGACGCAACTCTATT TGAAGGCAACCTACCGTTTGCTTTTAGTGTTTTGGTTTTGTTACC ATATCCAAAAAAAAACCATATATCCAAAAATTCCGCTGCACCATCT CTTCTTCTCTCCATCAACTACCCCTGCGGAGAAATTCACACCACAG TTACA
p*GAPDH*	GGTTGAAATGAATCGGCCGACGCTCGGTAGTCGGAAAGAGCCGGG ACCGGCCGGCGAGCATAAACCGGACGCAGTAGGATGTCCTGCACG GGTCTTTTTGTGGGGTGTGGAGAAAGGGGTGCTTGGAGATGGAA GCCGGTAGAACCGGGCTGCTTGGGGGGATTTGGGGCCGCTGGGC TCCAAAGAGGGGTAGGCATTTCGTTGGGGTTACGTAATTGCGGCA TTTGGGTCCTGCGCGCATGTCCCATTGGTCAGAATTAGTCCGGAT AGGAGACTTATCAGCCAATCACAGCGCCGGATCCACCTGTAGGTT GGGTTGGGTGGGAGCACCCCTCCACAGAGTAGAGTCAAACAGCAG CAGCAACATGATAGTTGGGGGGTGTGCGTGTTAAAGGAAAAAAAAA GAAGCTTGGGTTATATTCCCGCTCTATTTAGAGGTTGCGGGATAG ACGCCGACGGAGGGCAATGGCGCCATGGAACCTTGCGGATATCGA TACGCCGCGGCGGACTGCGTCCGAACCAGCTCCAGCAGCGTTTTT TCCGGGCCATTGAGCCGACTGCGACCCCGCCAACGTGTCTTGGCC CACGCACTCATGTCATGTTGGTGTTGGGAGGCCACTTTTTAAGTA GCACAAGGCACCTAGCTCGCAGCAAGGTGTCCGAACCAAAGAAGC GGCTGCAGTGGTGCAAACGGGGCGGAAACGGCGGGAAAAAGCCA CGGGGGCACGAATTGAGGCACGCCCTCGAATTTGAGACGAGTCAC GGCCCCATTCGCCCGCGCAATGGCTCGCCAACGCCCGGTCTTTTG CACCACATCAGGTTACCCCAAGCCAAACCTTTGTGTTAAAAAGCTT AACATATTATACCGAACGTAGGTTTGGGCGGGCTTGCTCCGTCTG TCCAAGGCAACATTTATATAAGGGTCTGCATCGCCGGCTCAATTG AATCTTTTTTCTTCTTCTCTTCTCTATATTCATTCTTGAATTAAAC ACACATCAACA

Continued on following page

p*YALI1_A20484g*	ACGCAATTCCGAGAAGTCCCTAGGGTTGACCGAGTTATCGGGCCA GTGCTGGACCAACCTCATAGTCGACTACACTACAGTACGGACTTG CAGCCTGAGAATGAAACTCTGATTGCGATTGGATTCGGCTGTGGG GAGTAGATTATTTGCCACAGTGTGGGGTGTAGACTAATCAGAAAC GAGATGCCGAGTAGACTAAAGGAAATGGTTCTAATTTTGGCGTGT AATGGCGGATATCTGAAGTAAAGGAGTGGGTGGCTCCAACGGTGT GAAAATCTCGCTGACAGGTGCGGTTCTGGGCACCCAAACTGCTCC CAGTCTATCACTCAACTGCAAGGTACCGCTTAGGTTGGAGAGGTC TTGTGGGGGGAAAGTAAAAGGGTAAAAGGGTAAAAGGGTAAGCG ATGTGCTCTAAGCAAGGAAAAATTAACATAGCAAGTGATGGAATC GGAGGGCGTTGTAGAAGGGGTGAAAATAGACGGCGACAGTGCTT GAAATCATCCCAATAATCAGTGCAAGTCTTCAATTGACGACAAAA CTGATCGAGAAAGTTGCCTACTGAATTTATATACGTTCGAGTGTA CCTCCACCGACACATGTTTCACATGCCTTCCTGTCTCCAACAAACG GGCTTCCCCACAAGCTGAACTTTCATGCGGAAGAGAGTCAGATAC GGAAGCTGAGCAAGCGTGTTCGGCAGAAGAAGCGAAAATGTCCAC GTGCAAGGCCCAATTCTTCGCACAACCGCACTTTCACGTGAATAT GAGCCCATTTTGCCCTTCAATATCTCTCCAGCAACTGCGTCAATCA TGTACACGCCGCTGGGTTTATTCCTCCGGAGTGGCATGATAATCT GCACCTCTCACATTCACATGGGGAAAGAGTTTAAGCCTGAAAGTC ATAAGTGTCTAATAGTGCGGGGTATATATAGGTGTTGCCAGCTCG TGTTTCCAAGTTCCTCCCTTCTGTCACCTTATCACATCTCTCATCC TCTCTAAAA
p*YALI1_B05639g*	TTGCCACCAAAGTGTTGTCGAGACGGTAAAAGTAAAAAGAGTCCT GCATGAGACCCATTGGTATGCTAGAGTTTGCCAGTGTGACGGGCG TCCGCGTTGGGACAAGTCACGTGAGGTAGGGTTAGCCCTAAAACA TTAGGGACAAGTATGAATAGAGGAGAAATGAGATTGAAAAAAAGC TGGTTTGGTGGAAATAGGGAAATTTTGTGTATTAGTTTGAAGACA TGATTGTTGCGACTATTGTGTATTTTTTAGTTATTTTTCGACTATT ATTACATAATCGGGGGGGTTTAAAATTAATGGATTTTTGAATTTT TTACAAATGTTAGTAAAGAGTTTTGATTTTGATTTTGATTTTGAT TTTGATTTATTTTGATTTTATTTTTAATTTATTTTCAATTTATTTT TAAAAAGACACACCCAATCGGAAGAAATAACATTACCCGAATTGC CAACATTACAGGCAATGATTCGATTACTCATACTCGTGGAGAGTC ACATGACCTAAACGAAGACGTTAAGACTGACTGGCTAATTCTACT GCTTACAT
p*YALI1_C01505g*	GCTGTGGATGGCACCTCTCTGACTGTGACTGCAGTTACCGAGGAC ACCTTCTCGATCATGATGGTGGCTTACACTCAAGAGAAGGTAGTT CTGGCTAAGAAGGAGGTGGGAGATTTCGTCAACGTGGAGGTGGA CTTTATGGGAAAGCTGCTGGAGAAGCAGATCGATGGGGCGGTGGA GAAGAAGCTTGGAGGCCTGGAGGCGTTGATTGAGAAGATTGTCGA CAAGAAGCTGAAGCAGTAAGTGCGCAGGAGACACAGGAGACGAGC TGAGTGGCGATAGTTCAATGGTCGCTGTCGTGTTACCTTCTAAGC TAGTTGAGTGATAACGCTCTACTCTACTCTACTCTACTTGTAGGAC TAAAGACGACGGACTGCGAATGAGACTTCTTGAGATATCAAAAGC TATGCACGAATGATAAATGAATAATATGATGTGGTAATGGTTTGT AGGCATCTAACAATGTGTGCTCATGTGGTCAGCAGGATTACAGTA ACGTCAATGAGTGGCCATGACAAGAGTTACAGTAAGGGACGAAGC AGACTGGAGATAGGAATATCAACGTATTAAGCAAGCATGAAACAG TCATTCTAAACTGGATTTCTGTCTTACAAGTATTCTATTAGCCCGT TCAAGGCCAACACGAAGATTCTGAAAAGTGTGATTGACTCTGCTC TGTCCGATTGTACATATACGTTGCTTTTGAAGATGTTTATTAGTT TTTGGCGATTTTTCACGATTTTAGCGATTTTATTTTAACGATTATT TTAACGATTTTAATGATATAACGGTTTCAACGATTTCAACGATTTC AACGATTTAAACGATTTTAATGAATATCGCAATTTATATGTTTGG CTTCAAATTAATTTCCAAACTCTCTTTGTCAGAGTTATGACCCGGA ACCTTAAACGCACTTCCCCACATAACTTAGTCCTCTCTCTTTGTTC CTTTCTGCTGCTGCCCTGTGACGGCCCGCTCAACTCCGTTACAACT AGAC

Continued on following page

p*YALI1_C13611g*	GTACGCCAACATGGATATCATCTCGCACATTTATCTCAATTGCAA
	GTTAGATCTGAGAGATGATTGGTTGGTGGGCACCAGTGTCACCTC
	CGAGGTCGATGATGCCTACCCACAGGAGGTGGCGCTACGTGCTTT
	GGTACAGTTTTTCAATACGCTGAGGTATCCCAAGCAGATGGAAAA
	GTTGGGGTTTGTGAGGAAGAATGCTGACTTTTTTGCTGCCGAGGA
	CGACTAGCATTGATTTGGCGAACAGCATCGAGTTAGCGATCTTCA
	TTGTGTTGAAGGATGGCGCTGTTCTCTCCATTTTCAAGAATTCGC
	CATTCACATGTTGGGTCAGAGTAGGACATCGCATAATCAAATGTA
	TTTTATAGCATGATACTAAACATGTAGCACCTACTTGTAACTACTC
	TACCAGTTACTCACGTGAGCTTCACGGCCTGGCACAACAGAGACA
	CACTGCTTTCGACACTCAATTGCGACCAGTGAATCTTCTATGCCCG
	AGAGGTGTTGGGCTCGGTATCTGGCGGACATAATTCCTGCACCTG
	ACCGTCCGTACTCTCGAACCCTCTTCTCCATCCTCTCATTGCGATT
	CAATTCCGGAATCTCGTAGATGGACGACACCACCAGCAGATTTCC
	ACGTCTCTGCCTAAAGCCAAGATTTGTTTGCCGTAGAAGTCCACA
	AAACCCGTCCCGCTGTTTCAGTCTTTGACTTCTTTCCACGAAGGGT
	AACAGGCCGTTGAGCATATACAAGTGGACCGGAAACCAGCCGTCC
	TCGTCGCTCGACATAAGAAAGTCCTCTTCCTTGTCGTTGAAGAGT
	AGTCTCAATCCGCGGATATGGGACAGAAAATGACCATGGTCTGAT
	CGTGGTTTCTCAGGGAGCACTTTGTTCCTCTCTTTGGTGACGGTG
	TGGGCGCGAGTTCCGCCTTCAAGAGAAGAAATGTGGAGCGGTTAC
	ATGTTGCAAATTCCAGCTTACTCTTTGATATAAGTAACGACACTG
	AGAACG
p*YALI1_C32392g*	CTATACAACTTGCCTCAGGCGAAAAAACATGTGGAACAGACTGGC
	CGACTACCCCAATGGCTTCAGGAATGCAGATCACTTCAATATCTC
	ATGGCGCCAAAAATGGGTCTCAAGGTGATCGACGCATGGGTCTCT
	TCCTCCCTGGTGAACCTCAAGGTGCTCGATATCCGCTCCAACATCA
	TCACCATTTTCCCAGATCATCTTGCCAGACTCTCCGATAGTCTGCT
	GGTTGTCAATTTGGACGATAATCCGTGTCTGGACAAGTGGATGAA
	ACGAAACAAGCAATTTGCAGAAAGATTTAATGCAGCAATTGCTCG
	AGCTTTGGCCCTACCCAAGAAGAAGGCTCCTAACTCCTACTCTAC
	GCGACTTGTTGGCACTATTTCGTCTTCCTCAGAGTCAGAGCAGCA
	CCAACAGCAGTCCAGAACCAGCAAGAACATCTTCAAGTCGTTTCG
	AAAAAAGCGACATTCGACCTTTAATCTTGAGACGCTCAGTCCACC
	ATCTTCGGACGATGATGACGATTTCCTCACCTCTCTGACACCAGAT
	CCACCAATTGGAAACACCTTCTCTTCTGCATCCATGGCCTCTGCCG
	GGTCGTCTCAGGGCTCGAGAACATACTGTGACTCCGTGGAGTCAG
	CAGGGTCCTCTGCTTCGATGGTCAAGCCTGTCGACATTGCTAAGA
	CCAACGTTCTGATAGATCTGTTTCGAGATATTTGGGAATCGACAA
	CTTTGGATCTGCTACCTGACATCAGAGAGCCCATTTCAGTTTATA
	GCCAGGAGCTAAATCCCACAAACACCTCTCTGATCCAATTTGGAG
	CGGCAGAAATCTCACGACAAAAAGTGCTGGACCTCTATGAACAAC
	TGGTGAACCTCGAGTTCACCTACGTGCAACACATGACCGAACTAT
	GTGATGTGAGTATTCTTGATGAGAACTGAAGAGCAGTATTTCTTA
	GTCTTACAGAATGGAAAGGAAGGACTGCAAAATCCTTCCCCACTA
	ACCCAG
p*YALI1_D15557g*	CAATGGTCTACTGTGTGCAGACCTCGGTCACACCCCCGGCGTCGC
	AGACCCCTCGACTCGACGACAAGCTGGTCAAGAAGTTTGGCACCG
	TGTGGAAGCAGTTTGAGAGCTCGGAAACGGGGTGGAAAAAACAGA
	TTGTTTTCTGGACCAACAAGATGCTCGAGCAAATACCCTATGAGG
	AGTGGGGCCTCAAGTCGATCCCCCGACAGTCGGCGGTTTTGCGTC
	GTCTCAAGCAGATCGAGGCCAAGGACGAAAAGGAGCTGTTTGAAA
	AGAACGAAAAGCTACCTGTGGCAGAACAGCTGCCCGTGGTGCAAC
	TGGAGGACATTGCTAAGAAGCACCACGAGGAGCTGCTCAAGGTGC
	CGCTGGTGTTCCCCAACCACCTGTCGACCCCGGATCAGGCCAAAG
	CCCAGCTGCTCAAGCTTGCTATCGACGGAGCCAAACATCACAGAA
	AGTACATGATCTGGTGCGCCATTGGAGCTCCACTGACGCTGCCCG
	TGGCGCTGCTGCCCGTGCTACCCAACCTGCCCGGATTCTACCTTG
	TTTTCCGATGCTGGTCGCACTGGAAGGCCCTGGAGGGTGCCAACC
	ACCTCAAACATCTCATCGAGGACAATCATTTGGAGTTTGTGTCCT
	CCGAACAGGTCTTTGACGCCTACAACAAGGCTGAAGACACTCACG
	TGCTTAAGAACGTGGCAAAACTGGCCGAGGAAGACGGGGTTGCCC
	CTGCCACTCCGACCTCCAACGAGACGGTGGAGAACTCTCTGCTCG
	TGGACAACACCAGCATTGACCATATCGTGGAGACGGTACAGCTTC
	CTGATCTCGCGGCAGAACTCAAACGGGCCCGACGACAGGTCGTCT
	ACAACATTGACAAGGAGATCAAGAAGCGGGAGCAGAAATAGACAT
	GTTAGATACTGGTATAGATAATTAGACG

Continued on following page

p*YALI1_E01694g*	TACCACCTCGAGAAGCAAATACAGAAAGCCCGTACTGAGGCTTCC ACTCCTCTACCTCTTCGTTTTTTGGTGGGTTCTCTTCCTCTTGAGC CAGATCAGCTTCTCGCCTCTCGACCAGATCCTGTCGCCAGTTCTTC TCCTCCAGATAAATGATCTTCTTGGTCTTGGCCTTCTCTTCTCGGA GACGGGTCTCCTCTTCTAGCTTGAGTCGTTCTTTTTCGGGAAGCT CTCCGGCAACAAATTCGGACACTTGCTTGACATTTGAGGTTTCCT TCTCGATCTCGAAAAGCGATCCATGCGATTCCTCCGGCCGCTGTT CCTTCTTGGGTCTTGTATCCGCTTTGCTACCCTTGAGATTGAACG AGATCATGGGGAATAGTATCTTGTGTTGTACCGTATGTGTTGAGA CGGTATAAGACTTTCGGGAAGTGTAAATACTTAGTGCAGTCTTGT GCAGCTGATATGTGCACCACGAGTACATACAGTATTCTGCTGGTA ATGGCAGTGACGTATGACTACTGTGCTGTATTGTAATCGTAGGTG TTACTCGTTCAGCTGAACGATATGATATTTGATCGCGTGTTTTTC AGTCATCTTTTATCGAAGAGTGATCGGTGTCTATACGCAATTACA CAACCTCCATTGGTGTTTAACAAGATCTCCCAGCAGCTCAACTGT CTGCAAAGTGATTACTTGAACGTACAGGTAAAGTGTAGCACAAAC AATGCCCTATTGGGACATAGAGTCAACTGTAGTATATATACAACA CCGTTCTCACGTTTGATCCATTTTTTCCCCCTTGACACTTATCTCA GCGTAATAGTATCGTACAGGGACCAGTATCAAAATTCTTATCCCC CAACCACTAATCTACTCCCAACCTTATCGCCATGCAGACGAGCAGA ATAAAATTAAAGGGCGTATTGGGCCTCTAGATTTTGGGAAAAAAG AAATCACCGTGTTTTGGACGTGAGTAAGTGGAGTTCTACGTCATA ACACG
p*YALI1_E28070g*	GATAAGGACAGATGTGGTCATAGACCATAGAAACGTGTAGTGGGC AGTAGCGGTGGTGGGGACTTGAAAAATATGAATAAACCGGCTGTT AAATGGCCCTCCTCCCCGTTTTTCCCCGCATTTTCCTCCATATCTA CCAAAATGGTCACTGCTTTAGACCCTCAACCTCCCACAGGAAGAT CGTCGCCCATCGAACCTAACGGTGGAGATTGCCCACTGTGACCCC GGTGATAGACACGGGGAAAACACGCAGCTTGTGGAGATTAGATAC ACGGCTTGATTGGCTGAAAGGAGGCCGGCGAGAGATAGGATCAGA AAGTGGATCAAAGATAGGGCGTGAGCTGTGATAGGAAAGATAGGA TATGATAGGGAGATTGAGCTGATATGATCTAAACGTCATCTGTAG AGACACTGACAAACAGACACGTGACCAGGGAAACACGTTACTGGT AAACACGTGATCTAGAACCGATCACCACTTCATTTTCTGGAAGAG GAAGTTCAATATCAGTCACTTCAGAGCCCATCTGTAGAGTAATAC AAGTACCCATATCAACAGTATCAACAGTATCAACAGTATCAACAG TATCAATAATCATCTGACACCAGGCTTTTGGTATTGTCTGATCCGT CTGATCACTCTGATCCGCCTGGTATCTCTGATCTGCCTTACACCTC TCCCTACCCTATCATCTCCCTACCCTATCATCTCCCTACCATATCA TCTCCTACCCTATCATCTCCTACCCTATCATTCCTCTGCTCATCCA GATCTACTATCTCTTACTCGAGCTCGGATTTCGACGGTGATTTTTC TGCCGGAATCGAACCCTTTGACAACCCCATCCCACCACTCCATGTC TCATTGCTCATTCCGATCCTCTTCATGATCATAAGGCCATTCCAAC CGAGCGATATGAGACACGCCATCACCCCATAAACATCCCACACCG GCTACCCACGGCTATTTGTGCATTATCAGACCGTCTCCAACTCAGT C

Continued on following page

p*YALI1_E31154g*	TTGTGTAAGCACATCTCGTTCTGGACATCATTACAGTACAACCGG TACTTGTACACCTGCCGCATCCATGCAACAAACGCGTGGAGATAT CTGTATGGAGATTGGCTGGGAGACAAACGTAGTTGGAGGGTCTGT GCAGCTCAGGAGAGTCTAAGGGGTCTTGAGCGAGCGTTAAGGAGA GTGCAAGACTTCCCCACAAATGTCTGCGACGCTGTATCATGCTGA TTTGTGTGTTTCAAGCGTATTTGTGTGACCTACCATGGTTTCTAC TCTCCTCCGATGGAACACTCCATTTAAGATTGCTCAACAATACACA ATATTAGAGCTGATAAAGAGCAGCCGTATCTGCTGCCATGCATCG TTCCTGACGTGGAAAAGAAAACAGTAGGCGGGAGTAGTACGGTGG GACAGGTGTAGGAGAGGTGGACAAGGGTGTAATTTTTTTCCCGGT TGATTGTGCAACGAAAACAAGCAGTATTTCGCATCGGAGTCCGAC TTGCTACCACCTTCTTTGAAATCAAACGCTCTCATCCTCATTTTAC ACATTTCAGTTTGTACCTGTCAGCACGCCTTGCGCTTCACGTGGT TAGAGTATCCACTCATGGCGGTCTCTCTGCACATGTGTGAACTTT ATCAGGAGGTAGAAAAACGGAGTTGGGGTGATTATCTGGTTATTT GGGCAACTTCTGCCAGAAAAATTTCGATTGAATAGATTAATTGAC ATCCTCTCCAAATTAATCGATATCCGACCGTCCTACGACCCCTATA ACCCCGCCAAATGCTCTCCATGGACCCCCCATCTGTAATCTGCAGT GTGCTTAAACGTGCCTAAGCAAAAAAGGGAAAAAAGCCAAACAAA TCCACCCTAAGACGTTTGGCGCTAAAAACATGTGCTCGCCTCCTG AACAAGCACGTTACTTTTGTTGGGCCACCCCGGTGCTATTAAAGT CCAGACTTGCCCTCCTCCGCGCCCCATCTCCGCCACAATAGCCATC TCAAG
p*YALI1_E32241g*	GCGTCTCGGGATCTGCATGTTGTGTGGCAAAGCCGGCCTTTGAGC CCTAAGCTTGTCTATCCGGGTAAACTTTTGACACTATTCACCCGTG GATTTCTTCATAATATCAGTTGTTGGTTGATGTTGTTCCGTATCT GATGTGGTGCCGTGGGTCATCGCCAGAGGTGTTGAAAGGTGGTGT GAAACGGGCTCCAGAGCCTTGCAACTCGGGCTGGCAGATCCACTC AACCGATTTCCCTTCCCCACACGGCCGCTGCGTCCAATCTGCCCAT CCCAAGTGTTAATAAACCACAAGCCCCCGTTAGTGTCGAAAAATT GGCTACAAGCCCTATATAAGAAAAGTTGCGGAGGCTTAAAAAAGG AGGCGTGCATCAAACAAGACCTCCTGCAGCTTGGCAATCGCCTTG ATAGGCCCTATCGGTGTGCATACTGGCAAGTCGAGCCGAGTAACC ACCGATATGATGCCGGCTCAGACACCGAACTCGATGAGCTATGAT CAGCAACAGTGACTCGACGTATGACACCCCACAAATCAGCAAGAT ACCCCACATTGCCCTTATATAGGCCTTAGGAGACGTTGTCCTGAT CATCAAACCATCCAACACATTAAACAATCATCT
p*YALI1_F29596g*	GGATTATGGTGATGTAATCTGTCCTGTACGAAGTGGTCCGTGTGG GTGCGATGGTTAGTGCTGGCAAGGTTCACTGCTGAGTAAGCTTGG GTGGCAAACGTGGAAGAGGTATTTGGGGAGTAACTTGAGTGTGTG AGACTGTGGATAGTTTATTGTGGGCAATTTACGGAGTTCCATTGG TTGTTAATATTGTTATCGATGACTTGCACAGTGGTTTGAGGGATA CTCTCACTGAGTCTGCTTACAGGTGGCGTTGTGTAATTGTCCCAG TAAGAGCTTAAATTAGTTGGTGTTCTGGTTAGCCAAAAAGAGGAG TTTGAGAGTGATTAAAAAAAAAATTAAAAAAATCGAATATAATAC ACAATACATGGTAATCATCGCTGATCTAGAGAAAAGATATACTCA ATTGATGTCAATCTTTTTGCAGATAAGAGTCTGTTCATTCCATATA TTCCTCTACATGTATTACAATACATACTCTACATTTACGAATAAAT CTCCTCATATTGTGGGCCCACTTAACCTTTCTCACGCAATCTTTTT ACCCCTGTATGAGTACGATACTACTTGTAGTATAGACAGAGAGGT ACGATACTCACCATGATGTCACTTTGTTGTTCGTAACCCACTCCAT TGAGTCTCAGCTTGTTTGTTTCTATGATAAGCAAACCTTGTCGCA GTCACCACCGCCTCAACAATC

Continued on following page

p*YALI1_A00911g*	GAGATTATGACTGAAATATCCGCAGGTCATTCATGGAAGTGAGGT
	TGACCCTGTGTTCGTACTTGTATTACAGATATTCCGAGTGAATTA
	TAGGCAATTAAGGACTACGGTTCCATTCAGACACAAATACACCAT
	TGGAATGGGATATTTCTGTCGCTTATGAGACATATTATCAATCTC
	TATTTCGTATCAAAATCTTGGCTACTGCAGCCAAGAAACCTACTT
	GTAGTGCGGTTGTACGGTTCAAAAAGGGCCCAATATTAATTAATG
	ATGGACCGAAACATTTTCAGAAAGTACGCTAAAAGTCACTTCTTC
	ACTGTAGCTGTCCCTGAACTAGCAGTAGATTCGAATCAGGCTGTG
	GTATGTACTGTACAGCCCATTAGCTCATTTTTGGGCCCTGCTACC
	AAGATTGTGTTCTAAGATAGATTGTCCGGGGTCTCGGCAAGTGGA
	GAAAAGTCTTGGTGATTATCACGGAAGTGACCCATTAGTTTTGTC
	GTAGCCCCTTTTTGAATTTCTTTCCCTTCATCTATACTTGTCTCAC
	GTTTCCACTTCAACGAGCGGCGACCGGGCATCACTAGTCTCATTT
	CGACATCGCTCGTATCTTCGGAGTCAATATCAAACATACTCCGATC
	TAAATATGACAAGATACCCGTTTGGTCAGCATCATAATATCTGTT
	TCTTGTTCTGACAGTAGCAATATCAGACGACTGCCTCTCAACGAC
	AAAGAACCATTGCCGACCGAGTCTTTGAACGGGATACAGGTTGAC
	GTCAATGTAAGATTCATTTTGTCTCATATTATTGAGTATTCATCTT
	CGGAGGACTTATTTTCCGGTCTTGAGTGAAGCAGACTTAGG
p*YALI1_A21509g*	CTTCAGGCCGAGATAACCAAGAAATACAACTCTCAAAAGACAGGA
	CGGCTAGGCTACCTTCCTGTTGCCAAGTGGACATGGTACGCCATT
	CGGTACTGTGAGCTCTCTAGTCCCATGAATCTCCTTGCCGGTAAA
	GAGCTCTGCTACCGCAAGCCCCCCAACAATGAAAACGTTGCCTTA
	GCCTGACCTACTGTTTCCTTGAAACGCAGCAAGCGGGTCAAGCGA
	AACAATCGACCTGGCCGTCTGAGAGCCATTGCTGTTATTATGGCC
	GAAAACAAGCAGCTGGAACGGGACTACACACTCTTTTCTCTGCCA
	GTGAACCTCTATCATTTAGTCGAGCTGGCCAGTGCACATGAAGAG
	ACAATCACCAAAACTGCGGCTCAGCGAAAAGTTATCAAAAGAGAA
	AAAGAATTGCAACTCGAGCGAAAACAAATGGAACACCTCCGGTCC
	TCTAAAAAGGAGTTCCTTGTTGAAGGTATGGCCTAAGCCTAAGCC
	CAAAAAGTTGGGATCTACCCGTCTGTGTATTATAGTAGATATTGT
	ATTATAGTGAAGAATAAGTGAAAACGTAGTCATATCAAGTGTTCT
	CCGGTGCAAAAAAGTCATGGACGACCACACACATACCATTCGCTT
	CGCCCATGATGTATGCTCATGTTTAAGTTGAACCATGCTCCGTTA
	TCTCCGAACGGCCAATAAAGTTTCGGGTACACACCATGATGCAGT
	TCACGTGACGAGACGCGCATCTTTGGAATTACTTACGGCCAAAGA
	CACACCAGTCGTAACATACCACCACAACGCCAACC
p*YALI1_C02160g*	CATGTTGGAATCGAATTTAGAGATGCTCTTCTTGAGCATCTTAAA
	AAGTTTACCGTCTCTGTGGCTGGAGGCTCGATTCTGAGCATGGAT
	ATGGCCAACTACACCAAGATGGTCAACTCGTGGGGCATCTCTGAG
	TTGTCTGAGGCGTTTGTGATTGTGAACGACCTGGCTGGTATTTTC
	GCAGCTGAGCTGGCCATGGTTAACTCACTGGTCAAGGAGGGTCCT
	CTGAGTAAGACCAAGCCGTTTATTCTGAAGGAGTATCTGAGTAAA
	CGAAGTGATTATCAGGGAGGAACGGTGAACAAGATGTTTGCAGCA
	ACGAATAAGATCCAGGCGTTGGGTATGATGTAGACATGACCCCTG
	AGCACATCGTATGGGTCTTAATTCGCATCTTGAGCTGAGCGATGA
	TGTCGCATACTGTGGTCAGCGCAAGTGCTTGGCTGGTATATATTG
	ACAATCAACATTAGATATTGATGATACTTGTAGTTTGTTTAGCTG
	TGGTGTACCTCTGAGGCAGTCTCATCATCCTCATTTTTTCAAAGTA
	CAAACACTGGCACGAAAACATTTATATCCATCCTATTCCCCATGAA
	CCAAACTTTCATCCTAACACAAGATCCGGGGATAGCCGACTATCA
	GCGTCGAAAGGGACAATTACCTTACCTCATGTCTCTAGTGATCAA
	GATGTACACAATCTGATTGTCGTACCGAGTCGGGAACCCGAGATA
	TGTTTGCACAATGAGTGTTGTTCTTCCCAACTATCATTTCAACTAA
	ATCTCAACTAAATCTGTTAAGCCGCTAATATCGTCAAGTAAATGG
	TTTCTCTGGTTGGAGCATTGGCTGTTATATTATTTCGGCATAATA
	CAGCGCGCGCCTGACTACACGGTCTCACTGCTCGATCTTAGTTCT
	CTTACTCTATGTATCAGCCTGGGTTAGAATATATCCTGAGAAAGC
	ACAAAGACTTTGCAAGAAGATCAGTGGAAAGGTGAAAATCGTTTT
	TTAATCA

Continued on following page

p*YALI1_C24183g*	TCTCTTGCCCCACTTAATCCTGTTCTCGGCTCTAGCTCAAGATTCG
	AAACACACCAGCGTTTGGCACGTGACGTGGAGGTTTTTTGGTGAC
	AAGCGAGGTCGTCTGCTGGGCACGTGACGTGGTTTTTTTTTTGGT
	GACGTGCAAGCGAGGTCGTCTGTCTTTTTGGCGACTTTGGCGGCT
	TTGGCAGTTTGGCAGTTTGGCGGCTTCGGTGAGTCAGACAAGCCA
	CTGTCTCACTTGGAGAACTGTTAAGACAGACAGACAGACAGACAG
	AGACAGAGACAGACGTGCTGAGACAAACAGAACATGTACTTGTGC
	TGATACTCGTACTGTACTGAGATTTCGTTGCTGCGACCCCTGCAA
	TGGTGCGGGGGAAGAGAACGGGCAATTGTAGTAATTGTTGCGATG
	GCATCCAAGGGAGAAATTTCTCAATTGATCAATTGCAATTAATAT
	TCTGCGGTGGACACTCTTGACTGCAAGCCATTGTGGTACTATCGT
	ACCCCGGCACCCCAATCAATGCCATCTACTCGATCACTAGATGACC
	ATCGACACGCATCGGCCTGAGTTAATGCCGACTCCGAGACCCGAC
	AACCCAATCCGAGGCTCCGACCGTCTCCTCCCCGCTGCGACTGCCT
	ACTCGTACGCCATCCCTCCATATCACCTCACTGGAACCCTGTTGAA
	GACAGTCTTAGGCCAGGCTTTGCGCGAAATAATGGCCCAGAGTAA
	CCGCCAGCACCTTTCCACAACAAGAGCCAAAGCCAGAAGGTCTCC
	AACCGTGCGTCTACAATGCGATATCCGATACGAGCAGAGGACGCT
	AACTTGTGTCTCCAGCAGTCGAGCACCGAGGGTTTGTCTTGGCTC
	CAACGCTGAGGCTGTTTGCTG
p*YALI1_C26533g*	ACGAATGAGGTTAAGCGGCGCGGAAGTCGAGTGGTTGACTCTACC
	ACTGAGAAAATCACTGCTCCTCTTCCAAACCCTGTCTCTGCTGCTC
	CTCTGCCTATCCCGGTCAAGGAGGAGGTTCTCTTCAAGGATGAAG
	GACTTTTCAGTGATATCCACGTCAAGATAGAGAACTCGCCTCCAG
	AGCCTCACGATTCGCTGGAGCACTTCCTTGGCACGAATGCCATGG
	TCACTCCTGCTCACACGGCCTCTGCTCCTTCTTCTACCAACGGCGA
	GGACGACTTCAAGGGTGCCACCACCTTTAGTCAGGCCAATTCTGC
	GTGCGGACAGCTTGAGCAAACTCAGGAGTCGTGGGATGTCATTGA
	CGACTCGTTCAAGACCCTGATGATGCCAAGTATGTCTACTGGAGG
	AGCTGCTCACTGCAACATGCGTTCTCTGAGTGTTGGCAATCATAG
	CATGGCTCTGGATTACACTCACAACAATCAGCACCAGGGGCATAA
	CCACCACGGCCAGAACCACAATCAGCACCACCACAACAGTATTCCC
	GCGGACCTGTTTGCTCCTCCTCCTCTGCACCCTCAACAGTTGACTG
	CTACTGGGTTCACGGACATTGACAGGGTGCGGCCAGTTACTGGAG
	ACAATTTTGACTGGGGAGTGAACGGTATTTTGGACTGGGATTGGA
	AGTAAGGTACTGAAACACTGTACTGTAAAGGAGTGAAGATGACGA
	AGGGGTGTTTGTACATAAGTGCATGTTTATTTATTTTTATTTTAT
	TGTTTGATTCTACAAAGTGGATATTTTGTTATATTCATATTCAGT
	GTGCAATTGTATGGAGATGATGACCCAGCCAATTTGGGGCATATC
	CAGAAGTATATGACGTTGGTAGCAGTGTATGGCAGTGATGATGTA
	TATCATAAGACAAAGACCCGCCACATTCTTCAGTCAAGTATATAG
	ACATGTAGTGTTTCTATCAGCAAACTTGGTGCTGGTTCTTCAAGC
	ACGCCA
p*YALI1_C30910g*	AGTCTAGGTGGAATATAGCATAATAGAAATAGAGCAACACCCAGT
	GCCGATGTAGATTGGAAATGCCGCTTCTCTGAGCGCTGTCAAGTT
	CCTTGGAAGTTGTATGTAAATACATTGATGGAAACAGGGCAAGGA
	AACGAGGGAAACAAGGAAGCGAGGGAAACAGGGAAACGAGGGAA
	ACATCAGATGACTCTTTAAAACAAAGCCCATACGTGAACTACAGA
	TACGTGATCTCACGTGACCTCTCAAACACCTCTTGCACTGTTCCTG
	TCAACTGCTGTACTGAATGGCATATAGTCCATATCGCCGAGCTGC
	TGAGAGCATTAAATATATATCATTGAGAAGTTCAGATTGTTATGA
	TCACAATGGTGAAGTCCTGATGGCAGTATCTTTTCTCCTTCAGAA
	GTTGTCTCTTCTATTCTCATGAACTATACATGTAGAAGTATGTAC
	GGAGTTTGGTACAAGAATGGGCGGAAAAGGAGGATTAATTTACGT
	GATATTCTTTGGAAGGCATATTTGATTGGATTTTCTTTGTTTTTTC
	AAGTGTTTCCAGGTTCTTTCCGAGATTCCGTCGTTTTCTGACTCTC
	CAAAGCTGGACGGTTGAGTGCTGGTTGAACATGTCGATTGGGGTG
	TCTATAGATAGGAGATTTTTAATGGAAAAGGACTTTTTCAGTACA
	AGTACAACCACTCAGACTGCACTTGTACTTGTACCTATATTGTACA
	AGCGATGAGTTTCGTGGTGATTACGTCTGTCGTTTCTGGTCCTGG
	ATGTCGAGATAGGGATCTCCCCAAAAATGATCTGAAGGAATGTGT
	CTGGAGATGTTGGTAAGGCGTAAAACAGTTTGAGGAGCCTACAAG
	AGCAATATGCACTTTGAATGGTGTTTATCGGGTACATATTCATCG
	ATTGGTAGACACATGTCAGCATTTCTTGACCGTACGGTGGTAAGT
	ATAATCTCAAGTAATATGATAACAATACTTAGATATATAGATAAG
	TAACTAG

Continued on following page

p*YALI1_D01497g*	TTGTACAATATTACAGATAACCGTGCATATTACTTATTTTAGTCAT AATCACTGCACCACTTCAATGTCGAAACCTGCAAAACAACAACTAC GAAAGCTGATAGCCCAGAGGGTCGCTGATGTTAGCAAAGACAGCA TCCAGAAACAGTCTCTCGCCACAACAGCGGCTCTCAGAACGCTCA ACGAGTACAAATCGGCCCAGAACATTGCCTTCTACATGAACATGG ACAGTGGAGAACTGGAGACCATGGATATGATCCGAAACGCGTTTG AGGACAAAAAACGGGTGTTTCTGCCACGCATTGAGAAACTAGCCG ACACAGCCGACAAGATCCATCCGTGCCAAAAATCCGAGCTCAGAA TGCTCGAAATTGTCGATTTCAAAGACGTCCAGAACTTGCAGCCAA GAGGACCATATAAACTGAAGGAACCGACCAAGGATGCCAAGGATG TTCTTGAGTCAGGTGGACTGGGGCTCATTGTGCTTCCAGGGGTAG CCATGACCCCTCAATGTGCCAGATTGGGCCACGGGGTGGGCTTCT ATGATGCATACATTCATAAGCACGAGAAGCTGCTGGGAGTGCCTA AGTTGGTAGGTGTTTGTCTGAGAGAGCAACTAGTGGACGAATTGC CACTAGAAGAACATGACCGGCTGCTGGATGTCGTTGTTTCAGGGG CCAGCGTGTATAGAAAGGTTGAATAGCAACTTAGTGTATATAAGT AAGTAATAGCTAACTAGTTTTCAAATGTTTGTTGATTTATATAAT CCTCCCAACAGTTTTTGACAAATTTAATTAGAATATCTACTTAAGT AGCAGATGCAAGCACAAAACGCAACCCCGAGCTCTGATAATCGAG GTG
p*YALI1_E02850g*	CCGAAAAGAGTGTGCCCCTTTGTATGCCTGCAGAAAAAATATGGA GGCTTTCACATAGCAAGGTGCACCACTGATTGGTTCCAGTTTCTC GCGTTTTAACAGGTAGTGTGTTATTGTAGCTGCATCTGGAGTACA CTTACACGGCGGTGGGCGGGGATGTTTCAAAACAACAGTCTTATC TGGCCAATTGAATCTGCGTTCTACCCGTCTTCTCAACTTAGCGATC GATTTTTTTCTTTCCGCGCCTACCGTAATTTCCTTGTAATTCTTTC TCACTCGTTTTGTCTTTTTCGGTTGTCAATCGCATCATTGTGCATC CCACACGAGAAATCAACACAACCGCATCTGATAATCATTCCCAGG AAGGAAAAATAACGCATTCCAGGAGAAACAGCACTGTTTGTGGGT GTAATTGGTATTCGTAATGAAGGTATCGCGTCATCCTTCTATCTA CAACAACTACAGAGTCGACATATATACCTACAACCCCAACTTTCGC AGCTGTCTTGTATGGTTTTGGATACACGGACCCACCGACACTTTT TACCCCGATGCCTTTCTTTCTCTGCTTTCACGTTTCTTTAGACTTT CAATCAACGCATGCAGATGTGAGGCTACACATACATTTCGGTTCA TCTGGTGGTAATCGATGTTCCCACACAAGACAGACTATTATCAAA AACACATTCATTGAGATCTGCTAGTAAATACGATGAAAATACCAA CACCCTGAAGTATCGTGACTATTTTACTTACTCAGATCTGTTCGCC CCTGAATATAACCATGTAAGCTCCTCCGTGTACAGCATCCCATGCA CGGTACCTTAGCCAAGGCTAGACGGAAGTCTAATAATACGCCCAT GCCTAATAAGGGAACACAGACATTCATCTTGGGAAGATCGATCGT CCTTTTGTGTTCTGTCGCTGTCTACTTAAATACCTACTTTGCCCCC CCCATCTTAGTTTTTCTTTCCTTACATAGCTTTCGCACATTACACA A
p*YALI1_F02403g*	ATTGTCGGCCAGACGTACGCTGGCAGTTGGATTCGAGGTGGGGCC CCGTTCTGCAATGCGACAGTACAGTACTGTATTCTTGGCAGCTGG TTCTCCACTGTATGCTCCTCCTCCACCCTACGCGTCTGTCGCCAAC GGTTAATGGCAGTAAGAACGCAGACAAGAAACTTAGACGAACAGA ATCGAACAAAAACGAACTCAGACCAAGTTCAGCACTTTGCTTTTG ACGCTTCTTGTGAACCTTTCTTCATTTGAAACGTTCACTCCTCTGC TGTATGTTGTCGTACTCGCTCTCTCTTCTACAAGTAGCCTCTCTTT AAACATACAGTACGTTAACACTGTGCAGCTGTTTGAGCCGGAACT GGTATATGTGCTCTTCATAGATCGGGGAAAACAGGTGCGGTACTT GTATAACGCCGCCTTGAGCGCACAACAAGTCTAGCATCAGTGCTT CGGGATAGATTGCCCATCATGTGCTTCCATGCGGAACCCCAAACG GGACTTCCGTGTCCCCTCCGTATAAATAAAGCCGTGCTAAAAACT AGCACCTTAAACTAACAAGAGAAAAATGGGGGGACAAGCGCGGAT GAAGACAAGCCAAAACTCCCTACTGTAACAGGGTGGTTATACCCG TTAAATTAGTCATCCATTATGTATTAGTCAGCAATGTCTCGGCTA GCTACTGTTTTTGACTGCCCCGTTTATATGTACTTAAAGCTTAAAG TACTGTACGCGTACGATACACAGAAGACTCCTCCCCAAATGTGGG TTTAATTAATGAAACATTGGGTTCATGCGAGATTGTCAGACGTGA TTCTTTCAGATTTCCAAGTTACGGGATTCTCAACTTTCTGAAGGT GCTCGTAATCTGACCATCGTTTCAGCAACATGTCTCTGTAATCTCT CCGATTCCACCCCACTGCTCCGTTAACTCATAACCTATGCTACCTT TCAAACTAACTGATGGTGCATAATCTAAAGACCTTCCAGACAACC CAAC

Continued on following page

p*YALI1_C01520g*	GGAGAAGAAAGAATCTACAGGCGGTGATTCGTGTGAGAAGACTCA GTAGCGTTTATGGGGACAGGGGAGCGTGGAGGGAGAGCGTGGAG GGGAACTGTGGAAGGGGGGTCGCTATAATATGTACTGTACTGTAC GTGAGATTGAACAGCTACCGAATGGTACCGGCGGCAAATAGTTTT AGTACGGGAACCAATGATATATATAGAGATAGGTACTTTAGACGA GATAAGAGTCATTTGGCAAAGATTGTAGGACAGCTACATCGTCTG TCATTTTCTATCGAGATTCAATATGTTTACTTAACTTGAAGTTTAG ATTAGTTGAAGATTAGACGAAGAGAAAGTTGTGCTGGGGGATACA TTTCGCTGTTACAAAGCAGGCAGGGGTTTGAAAAGTATTCCAATC GATAGAGACATACGTAGTTACTAACTATTTGGTGTTTAGGTAAGG CCTGTTTAATATAATATACTGAAGAATTTGCATTTCAATGTTTGTC AGCAAGTAACAATTTCTGGAGGATTAGTTGGTACTTTAGTCGAGT TAAAAGCTAACTCCGAGCATCCCTGACAGAAGGGGAAAGGAGTCA TTTCTGAGAGATTACCCATATCTCTCTTCAAGTTACGACTCTGAAA CTATTAGTAGATTAAGACTTGTATTATTCCTCTATGTCGTATATCT GGTCTTCGATTTTACTCGGCTTTTTTATTTTCTCGAATTTTATTTT TTCGAATTTTTTATTTCAGAAGAGAACTGCGTCAACTCTTGGAAA ATTGTACCACGTGTCCAACTGGAGCACCTTAACCAACGGCAGACT TATTATAAATACTCGAGTGAATTCAAAACTTGGTATTCTCCAGCA GTTCACTTGTGCTAACTCGCCCCCCACCAACTACCCCCACAGTAAG CCATTACCAACAAGGTTTGAGATGCATGACTCTGAATCAGATGAT CAGATCATCTTTCTCAACCACCCGACACCGTGAAAAACCGCCACCA CAAA
p*YALI1_E05779g*	TGCTATTGAGGTCATGGGAATGATGACTGGAAAGATCTTGCCCAA CACATTTGTTGTCATGGACTGCTATCCTCTTCCAGTGGAGGGTAC AGAGACAAGGGTCAATGCCCAGCAGGAGGGTATCGAGTTCATGGT GGAGTATCTGCAGGGCCTCAAGGACGTGGGACGCCGAGAAAACAT TGTCGGATGGTACCATTCGCATCCCGGCTACGGTTGTTGGCTCAG TGGTATTGACGTGGACACACAGTTTCAGAACCAGCAGTTCCAGGA ACCGTTCCTGGCGGTGGTGGTGGACCCCAACCGAACCATCTCAGC TGGAAAGGTCGAGATCGGTGCTTTCCGAACCTATCCCAAGGATTA CAAGCCCCCCAAGAAGGCTACTAAACAAAACCAGGACCAGAGCGT GCCTTTATCCAAGGCCAAGGACTACGGAGCGCACTCTGAGCGATA CTACGAGCTGGACGTGTCATTTTTCAAGTCCTCGCTGGATGAAAA CCTGCTGCAATTACTCTGGAACAAAAACTGGGCCGCCACCTTATC TCAGTCGACTATTCAGCTAAACCACGACTACACATCCAAGCTGAT GCTTGATCTGAGTGAGAAGAATGCACAGCTAGCGATCGGGCTCGG CGAGAAAACTCCCCAGTCTCAGGGTCGAGGTTTCCGAGAGGCTAT GTCAAAGGCTGACAACGAGCCTCACACTAACCTGCTCAACTACAG CACAAAGGGCCAGTGGGAGGCTGTGAACCGGTCTGTAAAGGATGG AGTTCAGATCGGGTCCGACGAGCTGCAGGGTCTCATGTCTCTTGA GATCCAGAGACGGCTGTTTGGTCGAGCGAAGTAGGAGGTATAGCA TATTTAATTTAACGAATTCATCAATACAAGGAGGGACAGAAGTGA CAGCACGACGACATACATTTTATCTTAGCAAGGAGTCACTGTAGA TATTAGTCACTGTATACTACAGTATGTACAGTAGATTCTGGTTCA ACTGTGGTAC

Continued on following page

p*YALI1_E09895g*	TGAGGTTCGCCGTAGATCATCTTATCCATCGTCTTTATATCTGTTT
	GCATGGACTGATGCCAAGAAGGGAAAACCGGGCGAGACCGCTTGA
	AATTTCACCTCTGCCAAGAAAAGGGCCTCCGATATGACTTCCGTC
	AAGCGCAATGTGGACAAAGACCCTGGCCGGTGCCAATCATGGAAT
	CGCTCCGAGATGCAGATGTATGCAAGCTCCTAATCAGATCCAGTC
	GGACATTTCACATAGAAAACGAACAAAGAAGATGAAAGGAAAACA
	GAGGATATGAATAGAATCCTCTGATCGAGATCTTTCCATCGGTAC
	CAATAACTAAGTCGAAAAGTGGCCAGCTCGAAAGATGATCCGACT
	TGGTCTTGAACTCAAAACTTACCAACATTTGAAAAAAATTATATA
	AAAAAATACTGATTGACAGATTAACACCTCCTTAGATTGGCGATT
	AGTGCAATGATGGAAGAGCTTGGCACACTAATGGCTTGATTACAC
	GATACAGTGACTTTCCTTACCCAGCAACTTTTCAACCGTTGTCTTT
	GTTCAGATCAGACACGGTCCACTACTGCGCACCGAGCAGGTAATT
	AACAACAGTATATATCCACCAACCGGCACGCCAAGAAGCTCAGAT
	GTCGCACCAAACCCGCCCAGTCTGGTTACAACTTCCCTGGTAGGT
	GCTGGTTTGAGTTGATTGGACGGGGGTCATCTGTGCATGTGCCGA
	GGCAGGTAGTGTAGGTAGTGTAGTGTCTTGTAGATCTCTATTCCA
	TCTGATAAGCTGAAGAAGTTGTATGAGCTGGATTTCAAGTACAGT
	AGAGATGGAGTCAGATGGAATAACCAGCACAATGAGTTCGATTCA
	ATGCTCTCACTTTGCTGCAACAAGTTTTCCATCTTTTCCATCTACA
	AAAGATTTGATTGATAACTTTTCATCCAAACTTTAGGACGACTGA
	AACAAGCCCGCGCAAACAATGCTTGAAGCAGAAGTAATAACCGAG
	TCGAGAG
p*YALI1_E32180g*	CATCCCTATCTCCTGTTGCTACAGCACCCATTGCTAGCGTAGGACC
	TTCATCTGTGGCCTCCGCCAGCGAATCCGAGTCTGCTGCTGTTTT
	CCAGCCCAGCACCGGTTCTGCAGGATCTACGCCCGCACTAGCCAA
	TGTCGCCACTCTGGCGCGAATCAGCCCCATGGTGATGGGCGTTGC
	TCTTGTGGCGGTCCCCGCCTTCGCCATGCAATTTGTCTAAATAAG
	TTTGCAAAAAGATCGTATTATAGTTGGAGCAAGGGAGAAATGTAG
	AGTGTGAAAGACTCACTATGGTCCGGGCTTATCTCGACCAATAGC
	CAAAGTCTGGAGTTTCTGAGAGAAAAAGGCAAGATACGTATGTAA
	CAAAGCGACGCATGGTACAATAATACCGGAGGCATGTATCATAGA
	GAGTTAGTGGTTCGATGATGGCACTGGTGCCTGGTATGACTTTAT
	ACGGCTGACTACATATTTGTCCTCAGACATACAATTACAGTCAAG
	CACTTACCCTTGGACATCTGTAGGTACCCCCCGGCCAAGACGATC
	TCAGCGTGTCGTATGTCGGATTGGCGTAGCTCCCTCGCTCGTCAA
	TTGGCTCCCATCTACTTTCTTCTGCTTGGCTACACCCAGCATGTCT
	GCCATGGCTCGTTTTCGTGCCTTATCTATCCTCCCAGTATTACCAA
	CTCTAAATGACATGATGTGATTGGGTCTACACTTTCATATCAGAG
	ATAAGGAGTAGCACAGTTGCATAAAAAGCCCAACTCTAATCAGCT
	TCTTCCTTTCTTGTAATTAGTACAAAGGTGATTAGCGAAATCTGG
	AAGCTTAGTTGGCCCTAAAAAAATCAAAAAAAGCAAAAAACGAAA
	AACGAAAAACCACAGTTTTGAGAACAGGGGAGGTAACGAAAGGAA
	TCGATATATCTGAACCGCTCTTCCGATCTCCACGGATCCCGAGACC
	GGCCTTTGATTCTTCCCTACAACCAACCATTCTCACCACCCTAATT
	CACAA
t*LIP2*	CTTCTGTTCGGAATCAACCTCAAGGTTAACGGCCACGATCCCCTC
	GTTGTTACTCTTGGTCAGCCCATTGTCGGTAACGCTGGCTTTGCT
	AACTGGGTCGATAAACTCTTCTTTGGCCAGGAGAACCCCGATGTC
	TCCAAGGTGTCCAAAGACCGAAAGCTCTACCGAATCACCCACCGA
	GGAGATATCGTCCCTCAAGTGCCCTTCTGGGACGGTTACCAGCAC
	TGCTCTGGTGAGGTCTTTATTGACTGGCCCCTGATCCACCCTCCTC
	TCTCCAACGTTGTCATGTGCCAGGGCCAGAGCAATAAACAGTGCT
	CTGCCGGTAACACTCTGCTCCAGCAGGTCAATGTGATTGGAAACC
	ATCTGCAGTACTTCGTCACCGAGGGTGTCTGTGGTATCTAAGCTA
	TTTATCACTCTTTACAACTTCTACCTCAACTATCTACTTTAATAAA
	TGAATATCGTTTATTCTCTATGATTACTGTATATGCGTTCCTCTAA
	GACAAATCGAAACCAGCATGCGATCGAATGGCATACAAAAGTTTC
	TTCCGAAGTTGATCAATGTCCTGATAGTCAGGCAGCTTGAGAAGA
	TTGACACAGGTGGAGGCCGTAGGGAACCGATCAACCTGTCTACCA
	GCGTTACGAATGGCAAATGACGGGTTCAAAGCCTTGAATCCTTGC
	AATGGTGCCTTGGATACTGATGTCACAAACTTAAGAAGCAGCCGC
	TTGTCCTCTTCCTCGAAACTCTCAAACACAGTCCAGAAGTCCTTTA
	TAGTTTGATCTGTATCCAGATAGCCTCCGTAATTGGTGTGTGTCT
	TCAAATCCCAGACGTCCACATTGGCATGTCCTCCACTGATAAGCAT
	TTGAAGTTCATCTGCGTTGAACATTGAGACCCACGAAGGGTCAAT
	GAGCTGGTATAGACCGCCCAAGAATGCATCTG

Continued on following page

t*PEX20*	AAGTGTGGATGGGGAAGTGAGTGCCCGGTTCTGTGTGCACAATTG
	GCAATCCAAGATGGATGGATTCAACACAGGGATATAGCGAGCTAC
	GTGGTGGTGCGAGGATATAGCAACGGATATTTATGTTTGACACTT
	GAGAATGTACGATACAAGCACTGTCCAAGTACAATACTAAACATA
	CTGTACATACTCATACTCGTACCCGGGCAACGGTTTCACTTGAGT
	GCAGTGGCTAGTGCTCTTACTCGTACAGTGTGCAATACTGCGTAT
	CATAGTCTTTGATGTATATCGTATTCATTCATGTTAGTTGCGT
hrGFP	ATGGTGTCCAAGCAGATTCTCAAGAACACTTGTCTGCAGGAGGTC
	ATGTCTTATAAGGTGAATCTCGAGGGTATCGTCAACAACCACGTC
	TTCACCATGGAGGGCTGTGGAAAGGGCAACATTCTCTTTGGAAAC
	CAGCTGGTGCAAATCCGAGTGACCAAGGGTGCACCTCTTCCCTTT
	GCCTTTGACATTGTCTCGCCAGCGTTCCAGTACGGCAATCGTACG
	TTCACAAAGTACCCCAACGACATCTCCGACTACTTCATCCAGTCTT
	TCCCTGCTGGATTCATGTACGAGCGAACCCTACGCTATGAAGATG
	GAGGGTTGGTAGAAATTCGGAGCGATATCAACCTGATCGAGGACA
	AGTTCGTTTACCGAGTTGAGTACAAAGGTTCCAACTTCCCCGACG
	ACGGTCCCGTGATGCAGAAGACCATCTTGGGTATTGAGCCCTCGT
	TTGAAGCCATGTACATGAACAATGGGGTGCTGGTTGGCGAGGTTA
	TACTGGTCTACAAGCTCAACTCAGGAAAATACTACTCTTGCCACAT
	GAAAACACTCATGAAGTCCAAGGGAGTGGTCAAGGAGTTTCCGTC
	TTACCATTTCATTCAGCACAGACTGGAGAAGACTTATGTTGAAGA
	TGGCGGTTTTGTGGAGCAGCATGAAACTGCCATCGCTCAAATGAC
	CAGCATTGGCAAGCCTCTGGGCAGTCTTCACGAGTGGGTCTGA
AaOKS	ATGTCTTCGTTGTCCAACGCTTCTCACCTTATGGAGGATGTCCAA
	GGCATCCGAAAGGCCCAGCGAGCCGATGGAACCGCCACTGTTATG
	GCCATTGGCACCGCTCATCCTCCTCACATTTTCCCCCAGGATACGT
	ACGCTGACTTTTACTTTCGTGCCACCAACTCTGAGCACAAGGTCG
	AGCTCAAGAAAAAGTTTGACAGAATCTGCAAGAAGACCATGATTG
	GAAAGCGATACTTCAACTACGACGAGGAATTTCTGAAAAAATATC
	CCAACATCACGTCTTTCGATGAGCCTAGTCTCAACGACCGACAGG
	ACATTTGTGTGCCCGGAGTGCCAGCGCTCGGTGCCGAAGCTGCTG
	TCAAGGCTATTGCAGAATGGGGCCGCCCCAAGTCCGAAATCACCC
	ATTTGGTCTTCTGCACTTCGTGTGGTGTGGACATGCCCAGTGCCG
	ATTTCCAGTGCGCAAAGCTCCTAGGTTTGAGGACAAATGTCAACA
	AGTACTGCGTCTACATGCAGGGATGCTACGCAGGGGGAACAGTCA
	TGCGGTATGCCAAAGATCTGGCCGAGAACAACAGAGGAGCTCGAG
	TCCTGGTGGTGTGTGCTGAGCTGACCATCATTGGACTTAGAGGTC
	CCAACGAGAGCCATCTCGACAACGCTATCGGGAACTCCCTGTTTG
	GAGACGGTGCTGCCGCGTTGATCGTAGGCTCGGACCCCATCATCG
	GCGTGGAGAAGCCCATGTTTGAGATTGTCTGTGCCAAGCAAACGG
	TGATTCCGAATTCCGAGGACGTTATCCACCTGCACATGCGAGAGG
	CGGGCCTCATGTTCTACATGTCCAAGGACTCTCCTGAGACTATCT
	CCAATAATGTGGAAGCCTGTCTGGTTGATGTTTTCAAGTCTGTAG
	GCATGACTCCTCCGGAAGACTGGAACTCGCTCTTCTGGATCCCCC
	ACCCAGGAGGCCGGGCCATACTTGACCAGGTGGAGGCCAAGCTCA
	AGCTACGACCTGAAAAGTTCCGAGCTACTCGAACTGTGCTGTGGG
	ACTGCGGCAACATGGTGTCAGCATGTGTTCTGTACATTCTCGATG
	AGATGCGTCGGAAAAGCGCGGACGAGGGACTGGAAACATATGGA
	GAGGGCCTGGAGTGGGGTGTTCTTCTCGGCTTCGGTCCAGGTATG
	ACCGTTGAGACCATTCTGCTGCATTCACTTCCTCTGATGTGA

Continued on following page

AaPKS4	ATGGGCTCACTGTCCAATTATTCGCCGGTGATGGAGGATGTGCAG
	GCTATCCGCAAGGCCCAAAAGGCGGACGGAACTGCAACTGTCATG
	GCCATTGGTACTGCCCATCCTCCTCACATCTTCCCCCAGGACACCT
	ACGCCGATTTCTACTTTCGAGCCACAAATAGCGAGCACAAGGTCG
	AGCTGAAAAAGAAGTTCGACCGAATATGCAAGAAGACCATGATTG
	GCAAGCGATACTTCAACTACGACGAGGAGTTTCTCAAGAAGTACC
	CCAACATCACATCGTTTGACGAGCCCTCTCTCAACGACCGACAAG
	ATATCTGTGTGCCGGGCGTACCTGCTCTAGGCGCCGAAGCAGCAG
	TCAAGGCTATTGCCGAGTGGGGTCGGCCCAAGTCTGAAATCACCC
	ACTTGGTATTCTGCACCTCCTGTGGTGTGGACATGCCCTCAGCTG
	ACTTCCAGTGCGCGAAACTGTTAGGTCTGCGAACCAACGTCAACA
	AGTACTGTGTCTACATGCAGGGCTGCTACGCTGGCGGGACAGTCA
	TGCGGTATGCCAAAGATCTGGCCGAGAACAACAGAGGGGCAAGAG
	TTCTCGTTGTGTGCGCTGAGCTCACCATCATCGGTCTTCGTGGTC
	CTAACGAGTCTCATCTGGACAACGCGATCGGCAACAGCCTTTTTG
	GCGACGGCGCTGCTGCCCTGATTGTGGGATCTGATCCCATCATTG
	GCGTTGAGCGACCCATGTTTGAGATTGTCTGTGCCAAGCAGACGG
	TCATTCCCAATTCGGAAGATGTTATCCACCTCCACATGCGGGAGG
	CTGGACTCATGTTCTACATGTCCAAGGACTCTCCAGAAACCATTTC
	CAACAATGTGGAGGCCTGTCTGGTTGACGTGTTCAAGTCCGTCGG
	CATGACCCCTCCCGAGGACTGGAACTCGCTGTTCTGGATCCCCCA
	TCCGGGAGGACGCGCGATTTTGGACCAGGTGGAAGCTCGACTCAA
	GCTCAGACCAGAAAAGTTTGGGGCGACCCGAACTGTTCTGTGGGA
	CTGTGGAAACATGGTCAGTGCCTGCGTGTTGTACATTCTCGATGA
	GATGCGTCGAAAATCTGTTGCTGATGGATTGGCTACGTATGGAGA
	GGGACTTGAATGGGGAGTGCTGCTAGGTTTTGGTCCTGGTATGAC
	GGTGGAGACTATTCTGCTTCATAGCCTGCCTCCAGTGTGA
AaPKS5	ATGGGCTCCATTGCCGAAAGTTCTCCTCTGATGTCCAGGGAGAAC
	GTCGAGGGAATCCGAAAGGCCCAGCGAGCTGAGGGCACAGCCACG
	GTCATGGCTATTGGCACCGCGCATCCCCCCCACATTTTCCCGCAAG
	ACACCTATGCCGACTTTTACTTTCGAGCCACCAACTCTGAGCACAA
	GGTGGAGCTCAAAAAAAAGTTTGACCGAATCTGCAAGAAAACCAT
	GATCGGCAAACGCTACTTCAACTACGACGAGGAGTTTCTGAAGAA
	ATACCCCAACATTACATCTTTCGACGAGCCCTCTCTGAACGACCGG
	CAGGACATTTGCGTGCCAGGAGTACCTGCTCTTGGCAAGGAGGCG
	GCACTCAAGGCTATCGAGGAATGGGGCCAACCCCTAAGCAAGATC
	ACCCACCTCGTTTTCTGCACCTCGTGTGGTGTCGACATGCCCTCCG
	CAGATTTCCAGCTGGCCAAGCTCCTGGGTCTCAACACCAATGTCA
	ATAAGTACTGCGTCTACATGCAGGGCTGTTATGCTGGAGGAACTG
	TGCTCAGATACGCCAAGGATCTGGCTGAAAACAACCGAGGAAGTC
	GAGTCCTTGTCGTTTGTGCGGAGCTGACCATCATTGGGTTGAGAG
	GACCAAACGAGTCTCATCTCGACAACGCGATCGGCAATAGCCTTT
	TTGGTGATGGTGCGGCTGCCCTCATTGTTGGAGCCGATCCTATAG
	TTGGCATCGAAAAGCCCATCTTTGAGATTGTGTGTGCAAAGCAGA
	CGGTGATTCCTGATAGCGAGGACGTGATCCACTTGCATCTGAGAG
	AGGCTGGTTTGATGTTCTACATGTCAAAGGACTCGCCCGAAACAA
	TTTCCAACAATGTGGAAGGTTGTCTGGTGGACATCTTCAAGTCCG
	TAGGGATGACTCCTCCTGCCGACTGGAACTCCCTCTTCTGGATCC
	CACACCCCGGAGGACGTGCTATCTTAGATGAAGTAGAGGCCCGGC
	TCAAGCTGCGGCCCGAGAAGTTCCGTGCAACGCGACACGTTCTGT
	GGGAGTACGGCAACATGGTCTCAGCCTGCGTTCTTTACATTTTGG
	ATGAGATGCGAAACAAGTCGGCAGCTGACGGGCTAGGTACTTATG
	GCGAGGGTCTGGAATGGGGAGTCTTGCTGGGCTTTGGACCTGGTA
	TGACTGTGGAGACTATTCTGCTTCATTCGCTGCCTCCGGTGTGA

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY