University of West Bohemia
Faculty of Applied Sciences
Department of Cybernetic

MASTER'S THESIS

Plzeň 2022

Tomáš Honzík

Prohlášení

Předkládám tímto k posouzení a obhajobě diplomovou práci zpracovanou na závěr studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni.

Prohlašuji, že jsem diplomovou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů, jejichž úplný seznam je její součástí.

Plzeň, ...

vlastnoruční podpis

Declaration

I hereby declare that this master's thesis is completely my own work and that I used only the cited sources.

Acknowledgements

Writing this thesis was a real battle for me. With all its limitations I consider it a great victory over my pride and my tendency to easily give up. Here I want to express my gratitude to people without whom I wouldn't be able to finish it. I want to thank my mother for teaching me the bravery to always choose love over fear. I want to thank my father for igniting my passion for science and technology and the attitude to see every problem as an interesting puzzle to solve rather than an obstacle. And I want to thank my advisor Daniel Georgiev, thank you for your patience and willingness to go the extra mile to give me a number of important lessons.

Abstract

The design of multicellular biological yeast circuits usually relies on pheromone signaling. It is well established to use natural mating signaling as a backbone of the information transmission from one cell to another. Such designs typically use an activated form of the transcription factor STE12 (a result of the pheromone induction) as an interface for a downstream signal processing implemented within the cell. Since the STE12 is an activator, this approach only allows for the gene transcription activation and the repression must be done indirectly by expressing another transcription factor (TF). This work proposes a simple computational approach to the combination of the promoter DNA sequence and omics data for the selection of wild-type Saccharomyces Cerevisiae promoters with the desired repression of the transcription upon an alpha-factor induction. By avoiding the multistep process with intermediate TFs and building the repression mechanism directly on the promoter level, the orthogonality to cell machinery is guaranteed and time response is potentially improved.

Key words

saccharomyces cerevisiae, yeast, logic, NOT gate, multicellular networks, protein degradation, promoter, orthogonality, modularity

Abstrakt

Návrh vícebuněčných biologických kvasinkových obvodů zpravidla spoléhá na feromonovou signalizaci. Osvědčeným přístupem je použití přirozených pářících signálů jako základ pro přenos informací z jedné buňky na druhou. Takové návrhy zpravidla používají aktivovanou formu transkripčního faktoru STE12 (výsledků feromonové indukce) jako rozhraní pro další zpracování signálu vytvořeného uvnitř buňky. Jelikož je STE12 aktivátor, tento přístup umožňuje pouze aktivaci genové exprese a represe musí být provedena nepřímo za pomoci exprese pomocného transkripčního faktoru (TF). Tato práce navrhuje jednoduchý výpočetní přístup pro kombinaci DNA sekvence promotoru a omických dat pro výběr divokých kvasinkových promotorů s požadovanými vlastnostmi, tedy represí při indukci alfa-faktorem. Vyhnutím se vícekrokového přístupu a vytvořením represivního mechanismu přímo na úrovni promotoru je zajištěna ortogonalita k přirozeným buněčným procesům a umožněna lepší časová odezva.

Klíčová slova

saccharomyces cerevisiae, kvasinka, logika, NOT brána, vícebuňečné sítě, degradace bílkovin, promotor, ortogonalita, modularita

Contents

Introduction

The ultimate way to understand something is the ability to create it. Synthetic biology is a scientific field which aims to build living organisms with desired properties. Scientists already managed to build logic gates [1], time pulses generators [2], incoherent feed-forward loops [3], bistable switches [4] or oscillators [5] in single celled organisms by altering its DNA. However with the necessity of creating more and more complex functions, the synthetic biology community faces the same problem that has been faced in all engineering domains in the past. And that is how to accumulate applied effort for solving various kinds of problems and reuse it in future applications. This leads to the need for reusable components and a standardized system defining how to assemble them together. The goal of this thesis is to contribute to this area by improving current multicellular saccharomyces cerevisiae network communication by developing direct NOT logic for membrane impermeable signaling molecules, alpha-factor specifically. This approach, if successful, could enable modular, orthogonal and fast transmission of the signal from the yeast mating pheromone to production of arbitrary protein in s. cerevisiae in the NOT manner.

1 Biological Background

This chapter lays down essential background information from both biological and engineering worlds. At first, characteristics of synthetic biological circuits are described together with essential problems that are being faced while engineering complex behavior in biological systems and how different groups are trying to address these issues. Then biochemical pathways which saccharomyces cerevisiae use for detecting and reacting to presence of a pheromone of the opposite mating type are described in order to identify potential mechanisms to implement NOT operation.

1.1 Characteristics of Multicellular circuits

This section deals with two aspects of engineering biological systems. The first subsection discusses synthetic biological circuits in a more general way. It emphasizes the importance of modularity and orthogonality of biological parts and summarizes the current approaches on how to achieve these properties. It is derived why building multicellular and spatially organized networks is important. The second subsection focuses on a more specific task which is assembling logical functions in multicellular networks. It explains how a universal logic can be built, and the related importance of NOT logic operation. The disadvantages of current implementations of NOT logic on membrane impermeable molecules are then described to establish the motivation for the potentially better solution proposed in this work.

1.1.1 Synthetic Biological Circuits

Multicellular networks make biology easier to engineer. Designing information processing circuits in living organisms is a nontrivial task. In contrast to electronic engineering, in biology there is a problem with wiring. Every biological part such as folded protein, regulatory DNA or RNA is potentially informationally wired to any other within the cell and so unintentional crosstalk must be prevented. Biological parts which don't crosstalk

with the host cell machinery and each other are called orthogonal. Second side of the wiring problem is that if a signal should be transmitted from one part to another it can be complicated to set up such molecular interaction of these parts. To prevent the necessity of designing each interaction from scratch it is also important for biological parts to be modular, which among other things means that they can be mutually replaced in different configurations without losing their ability to properly transmit and process signals. This section discusses the important properties of orthogonality and modularity of biological systems and the advantage of using multicellular networks composed of multiple strains.

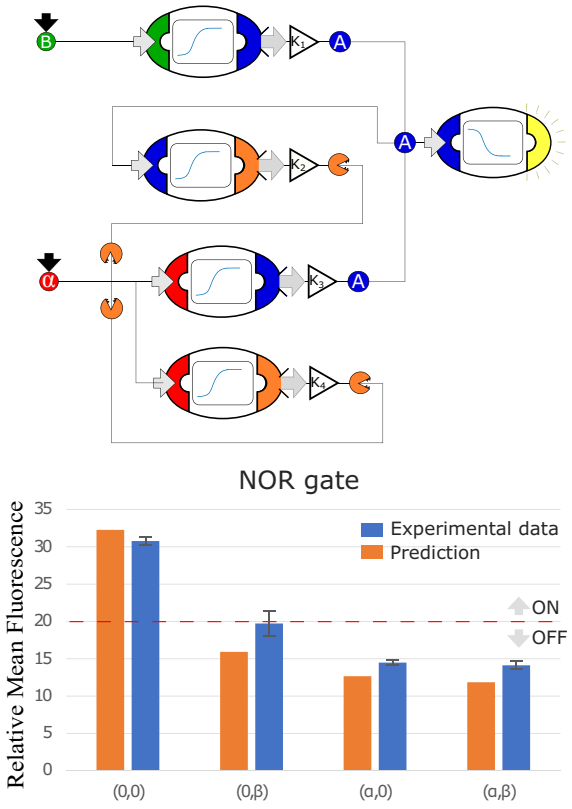


Figure 1.1: **Example of multicellular NOR gate assembled from 5 yeast strains.** Source: [6].

Choice of signaling mechanism guarantees orthogonality in biochemical reaction network. The orthogonality of a set of biological parts can be achieved in various ways. Different research groups are working on various approaches to create libraries of orthogonal biological signaling. One

group created a library of mutually orthogonal GPCR-peptide pairs which can be used for signaling between s. cerevisiae cells. These pairs were obtained by genetic mining of GPCRs and signaling peptides of fungi followed by experimental cross validation and can provide up to 32 undisturbed communication channels [7]. In another work auxin was used as a signaling molecule. Because auxin is a plant hormone it is orthogonal to the cell machinery of most of the non plant eukaryotes [8]. The auxin can enter the cell and bind to a specific degron domain merged with a transcriptional domain and mark the given transcription factor for degradation. To ensure intracellular orthogonality, a promising technique is to use dCas9-based transcription factors, where Cas9 enzyme with impaired nuclease activity is merged with transcriptionally active domain of some transcription factor. dCas9 combined with specific gRNA is able to localize the transcription factor in the desired position in the target promoter to affect the transcription rate. gRNA and its target DNA sequence pairs can be designed with great freedom to create a large library of mutually orthogonal biological parts [9]. Last but not least example approach is to implement a logic circuit outside the cell nucleus via orthogonal modular proteases [10].

Modularity is achieved by separation of intracellular circuits into different strains. The cell membrane serves as a natural encapsulation of subcircuits implemented within the cell and thus allows the reuse of a limited number of biological parts in different configurations in various strains. These strains can operate as modules communicating with each other with various kinds of signaling molecules such as α-factor, β-estradiol or auxin and report some information to the outside world with reporting molecules such as GFP [6]. To achieve modularity on the next level, it is necessary to introduce spatial organization of individual cells because the number of signaling molecules is also limited. Spatial organization of different cell types is typical for multicellular organisms, but it also occurs to some extent in case of some simple single celled organisms consortia, for example in bacterial biofilm formation [11]. In nature, spatially organized cells can use the concentration of different signaling molecules to estimate their relative position in the whole organism or consortia and based on their position execute a specific program [11, 12]. In the case of synthetic circuits, spatial organization

can be used for introduction of signal directionality. With no spatial organization each cell is broadcasting its signals to the whole population and vice versa listen to every broadcasted signal. Spatial segregation allows for sending signals just to a specific subgroup of the population, then the cells can create modules in form of these subgroups and thus even with limited number of signaling molecules or even strains, almost unlimited complexity can be achieved (e.g. homo sapiens with its approximately 22 thousands genes [13]). Thus it can be concluded that the maximal potential of modularity of synthetic biological systems lies in multicellular systems that are spatially organized.

Combination of strains in a full system can be controlled with various spatial separation or organization strategies. One of the important problems with engineering multicellular circuits is the fact that multiple strains or species living together in well-mixed liquid medium is subject to winner-takes-all phenomena, where a strain with slightly better fitness to the given environment after sufficient amount of time eventually outnumbers the other strains [14]. This problem can be addressed by introducing mutualism in between the strains or introducing spatial segregation. The conceptually simplest way to spatially segregate different strains is to manually establish spatially separated colonies on solid medium [15]. Interesting concept inspired by ecosystems naturally occuring in soil was to use porous polymer microcapsules to create so-called microbial swarmbots. In this system the winner-takes-all phenomena is prevented by spatial segregation of various strains or species into individual capsules, allowing the cell proliferation in safe environment within the capsule, but the exchange of nutrients and signal molecules remains intact due to the porous material of the capsules [14].

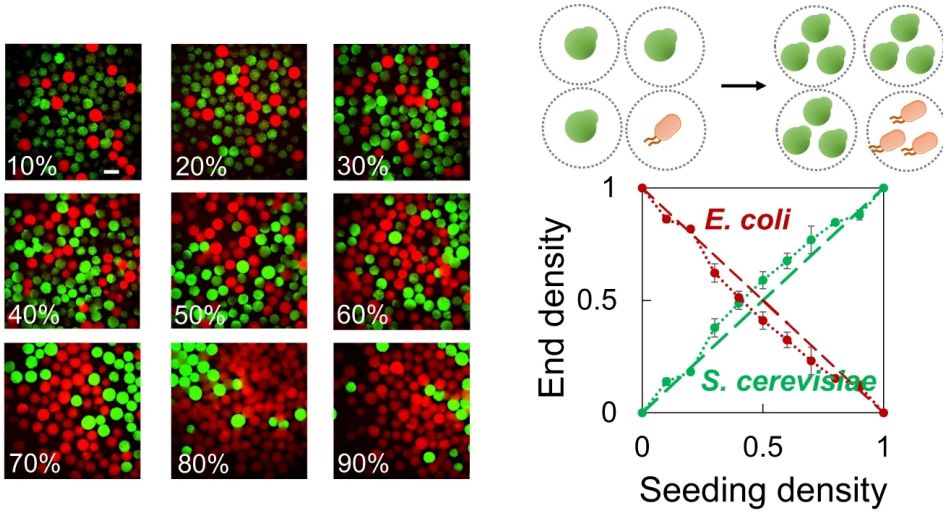
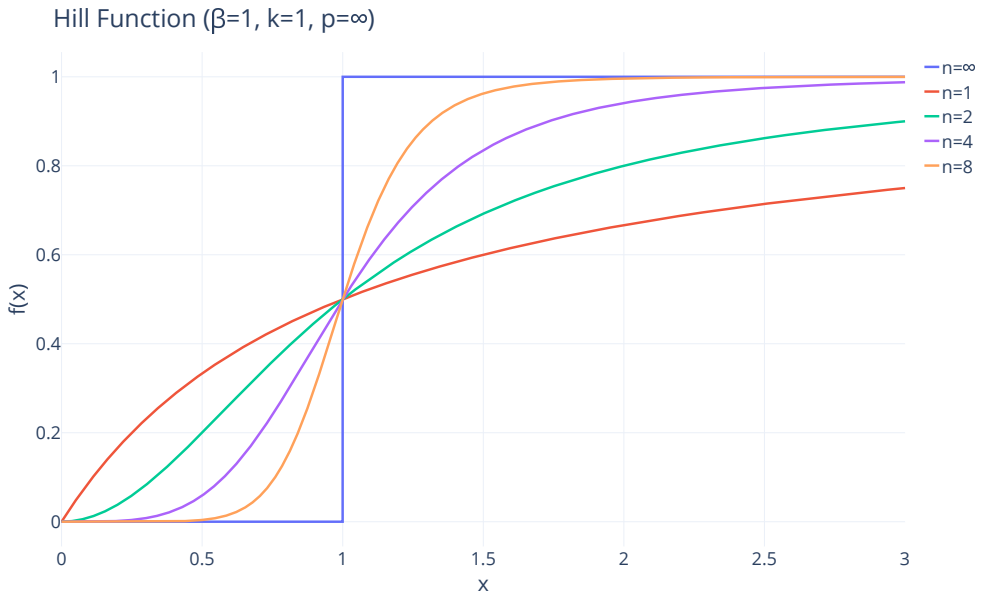


Figure 1.2: **Microbial swarmbots.** Source: [14].

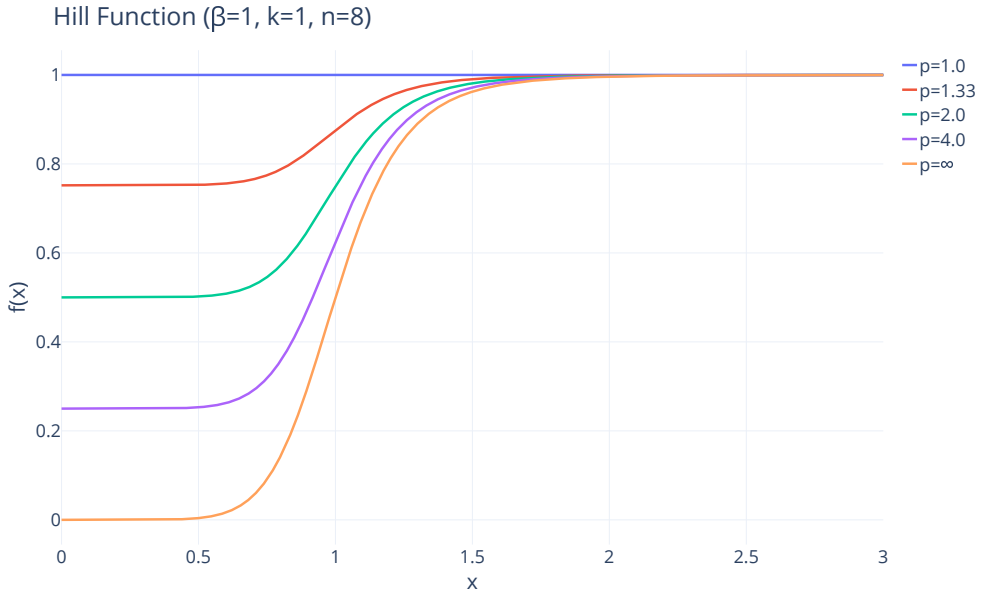
1.1.2 Multicellular Network Logic

Logical circuits are made of individual logic gates. Idealized logical gate is a function with one or more binary inputs and one binary output and can be fully determined by a $n^2 \times (n+1)$ table such as example 1.1. Magnitudes of real physical signals do not operate in the realm of binary but rather real numbers. Thus real physical devices such as transistors work with high and low values, by detecting whether the signal is below or above a certain threshold. Ideal conversion from real to a binary signal would be a step function. In biological systems, concentrations of various molecules, most of the time proteins, are typically thought of as signals and genes which produce these proteins can be designed to work as simple logical gates. Genes which are designed to work as a logical gates hardly go from low to high state in a step manner but rather follow a kinetics which can be approximated by a hill function defined by en equation 1.1, where $x \in \mathbb{R}^+$ is a magnitude of an input signal, $n \in \mathbb{N}$ is a hill coefficient, $p \in [1, \infty]$ is a fold-change, $\beta \in \mathbb{R}^+$ is maximal expression rate and $k \in \mathbb{R}^+$ is an activational threshold. To function properly, biological logical gates should maximize p and n, in the limit case, when both of them go to infinity, the step function would be achieved (see figure 1.3).

$$f(x) = \beta \cdot \frac{x^n + \frac{k^n}{p}}{x^n + k^n} \tag{1.1}$$



(a)



(b)

Figure 1.3: **Hill functions examples.** Both figures show individual instances of a hill function. The figure (a) shows how the hill function changes with respect to the hill coefficient **n** and the (b) figure shows how the hill function changes with the fold-change **p**.

INPUT		OUTPUT
0	0	0
0	1	0
1	0	0
1	1	1

Table 1.1: **Example of logical gate AND**.

Multicellular circuits can implement universal logic functions. In a multicellular network, individual cells receive signals from other cells and then process them internally. Arbitrary logic gates can be built just by reusing one functionally complete logic gate such as NOR or NAND. For that reason research groups have built libraries of universal NOR gates which can be layered on each other to perform a variety of logical functions [15, 16]. In practice NOR gates are constructed by combining NOT with OR gate and similarly NAND gates are constructed by combining NOT with AND gate [15]. This points out the importance of the NOT operation in virtually all nontrivial logical functions. As it was discussed earlier, cells themselves can be used as modules and higher order logical gates can be further constructed using multiple strains.

The membrane impermeable NOT logic is currently only possible through multistep transcriptional mechanisms. It was previously described that auxin based signalization allows for direct degradation of target protein, potentially transcription factor, and thus both activation or repression (NOT logic) can be achieved using this hormone. However in the case of membrane impermeable molecules such as signaling peptides, the same mechanism is not yet possible. Membrane impermeable signaling molecules, α-factor for example, typically bind as ligands to some cell receptor, after interaction with the receptor, a cascade of activational reactions usually follows. In current synthetic yeast biological circuits, direct NOT gate on membrane impermeable molecules doesn't exist and intermediate transcriptional steps are used to emulate the NOT behavior [6]. This typically means that an intermediate gene expressing a repressor gets activated by the active form of Ste12, and subsequently binds to the target gene and turns it off. This

approach is obviously not optimal and comes with a couple of drawbacks.

The number of translation steps strongly limits the speed of response. The translation step can be defined as a series of reactions leading to creation of a protein from its coding DNA sequence blueprint to its mature folded form. The process consists of assembly of the transcriptional machinery with the help of a transcription factor, transcription of the mRNA from the DNA, transportation of the RNA to the ribosome, translation of the mRNA to the protein (potential transcription factor), folding the protein into a mature structure and transportation of a protein to the promoter of the next gene. After accumulation of a critical amount of the produced protein, the following gene in the sequence is activated and the same cycle is repeated. The more translational steps are wired in series the worse is the speed of the response. The speed of the response can be defined as the time delay between the induction by the input molecule to detectable change in concentration of the reporter molecule. When the biological parts are layered on each other to create more complex behavior, the circuit depth increases. The imperfections of individual gates, most especially the leakiness, leads to higher and higher signal degradation [16]. Second side effect of deep circuits is the increasing transition delay between the input and output signal. From this can be easily concluded that minimizing the depth of a logical circuit has the potential to improve both signal quality and latency.

1.2 Fus3-Tec1 System

In previous sections it was described how properties like modularity and orthogonality can be achieved when building complex functions by multicellular networks. Various approaches which aim to establish the communication between individual cells were also described. A behavior that can be typically required of the synthetic circuit is to implement various logical functions. It was explained how arbitrary logical functions can be built by using only NOR or NAND operation. Because both NOR and NAND are in practice implemented by combining NOT with OR, NOT with AND respectively, in some sense the NOT operation is the most important logical operation of all. The unfortunate reality of the current implementations of NOT operations

for membrane impermeable molecules which rely on a multistep transcriptional process was revealed and critical disadvantages like the signal transition delay, the signal degradation and the waste of unnecessary amounts of components were explained. This section describes the biochemical pathway which saccharomyces cerevisiae uses for detecting and reacting to the presence of a pheromone of its opposite mating partner. The main branch of this pathway is already a well established backbone for synthetic circuits signaling, but it only works in an activational manner. After describing the pathway in general, the focus is made on a branch of the pathway which is yet overlooked by the synthetic biologist community. It is the Fus3-Tec1 system which is in wild yeast responsible for actively preventing cross activation of filamentous and invasive growth genes during the mating. This section describes the mechanism by which the Fus3-Tec1 system, a candidate for programmable NOT logic, operates.

1.2.1 Mating Pathway

The journey of the signal starts on the cell surface in the moment when the pheromone interacts with the G-protein-coupled receptor (GPCR) via Ste2 in MATa and via Ste3 in MATα cells. The GPCR alpha subunit exchanges GDP for GTP and releases the beta-gamma dimer subunit (G$\beta\gamma$, a Ste4-Ste18 heterodimer) which then binds to several substrates in the cytoplasm. The two most important G$\beta\gamma$ effectors from the perspective of this work are the Ste5-Ste11 complex and the Ste20 protein kinase, which are both brought together and attached to the inner side of the plasmatic membrane. Ste20 then phosphorylates thus activating Ste11. Besides that the Ste5 binds to the G$\beta\gamma$ it also serves as a scaffold for assembling the mitogen activated protein kinase cascade consisting of the MEKK Ste11, MEK Ste7 and two MAPKs Fus3 and Kss1. After the Ste11 is phosphorylated the signal transmission continues through the cascade by phosphorylation of Ste7 and subsequently also both Fus3 and Kss1. After being activated, Fus3 separates itself from the Ste5 and travels to the nucleus where it phosphorylates its targets on a threonine or serine residue. The key target of Fus3 is the Ste12-Dig1-Dig2 complex. Without intervention from Fus3, Dig1 and Dig2 repress the transcriptional activity of

Ste12 by binding to its functional domains with high affinity. When phosphorylated by Fus3 both Dig1 and Dig2 lose their ability to bind and repress Ste12 setting it free to promote transcription of mating genes [17–20].

1.2.2 MAPK Specificity

As mentioned earlier, the pheromone response pathway activates two MAP kinases Fus3 and Kss1. Although both of these are able to independently mediate the activation of mating transcriptomic changes, the Fus3 dependent mechanism is dominant for a pheromone induction. The Fus3 is known to be activated only in case of mating whereas Kss1 also initiates filamentous growth in response to nutrient limitations. The cross activation of the filamentos program during the mating is prevented by the Fus3 by two distinct mechanisms. The Fus3 regulates the activation level of both MAPKs by a negative feedback loop, so the actual amount of activated Kss1 is not sufficient to exceed the threshold and fully activate the downstream filamentous program after pheromone stimulation. At the same time, Fus3 promotes the proteolysis of the Tec1, the main filamentous specific transcription factor [21–23].

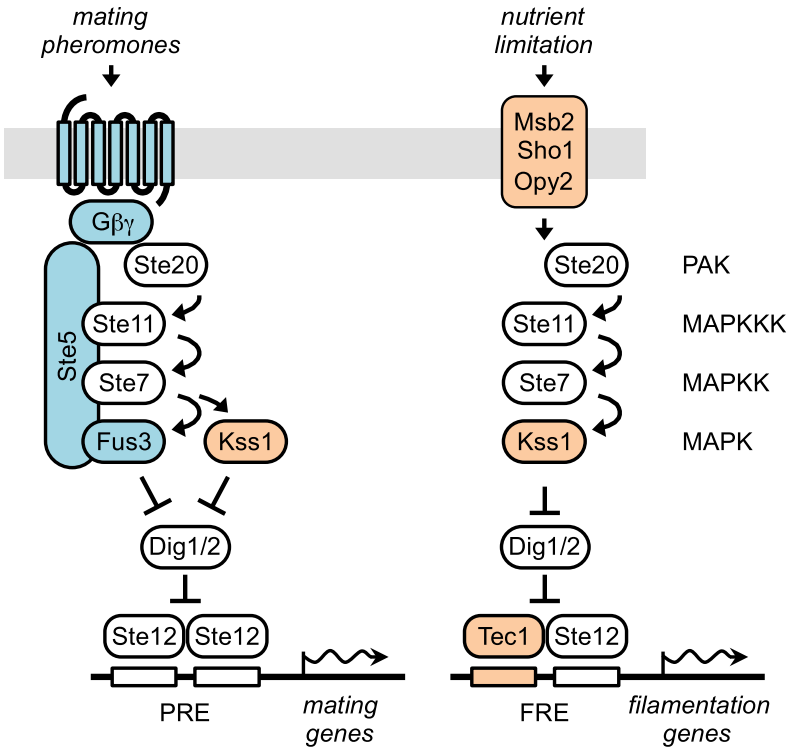


Figure 1.4: **Mating and filamentous pathway in yeast.** Source: [23].

1.2.3 Tec1

Tec1 is a transcription factor targeting filamentous genes. The activation of these genes is usually done by the Tec1/Ste12/Dig1 complex, where Dig1 disrupts the transcriptional activity in a similar manner as it does in the case of the mating pathway [24]. Both Tec1 and Ste12 have their own DNA-binding and transcriptional activation domains. It was found that Tec1 and Ste12 can cooperate in several different configurations where only a subset of the 4 domains are used to promote transcription. It was also found that Tec1 is able to activate some of the genes independently of Ste12 [25]. The Ste12 independent control is the most interesting for the potential use for induced gene repression.

1.2.4 Tec1 Degradation

The key mechanism to prevent cross activation of filamentous growth program in pheromone response is to increase the degradation rate of the Tec1.

The process of active degradation is initiated by the active form of Fus3, which phosphorylates Tec1 on T273 residue [26]. After the phosphorylation, the resulted phosphodegron is recognized by SCF-Cdc4 (Skp, Cullin, F-box containing complex, Cdc4 being the F-box protein) E3 ubiquitin ligase and the Tec1 is ubiquitinated, marked for proteolysis [27]. Without Tec1, proteins which are required for filamentous growth are not produced and thus the filamentation program is canceled.

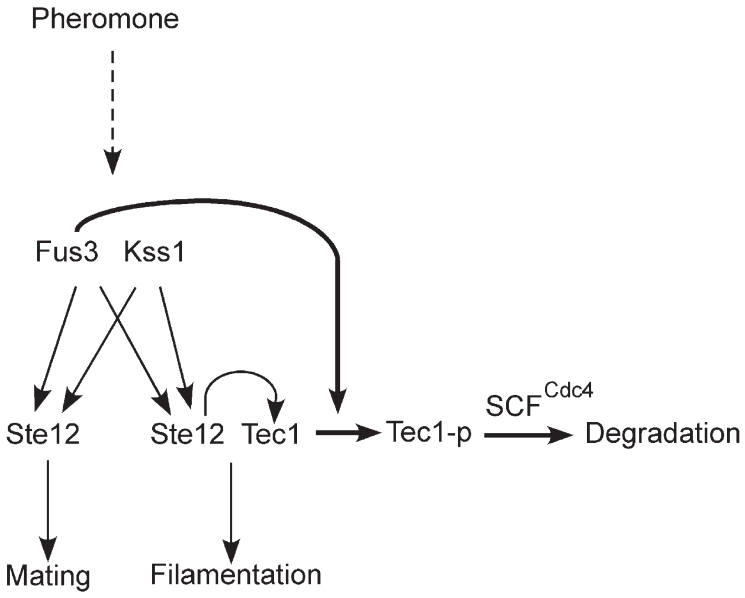


Figure 1.5: **Tec1 degradation mechanism.** Source: [27].

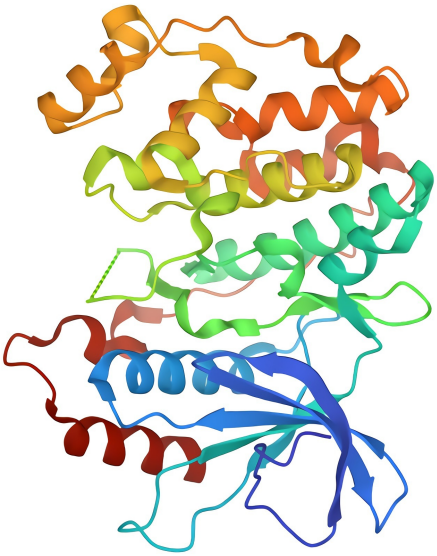


Figure 1.6: **Crystal structure of non-phosphorylated Fus3 (2B9F).** Source: [28].

2 In Silico Design

In the previous chapter the background for this work was established. The problem was formulated as the need for more direct NOT operation for membrane impermeable signaling molecules. The Fus3-Tec1 system, a system known for its ability to actively suppress invasive and filamentous growth during the mating of saccharomyces cerevisiae was proposed as the basis for the potential solution. In this chapter, it is proposed how to use the Fus3-Tec1 system to introduce programmable NOT operation for alpha-factor. After the solution proposal, it is described which data was gathered and how it was processed in order to find appropriate biological parts and how to assemble them together.

2.1 Mechanism Description

This section discusses the theoretical function of the designed system. The design is built upon the theoretical knowledge summarized in the previous chapters with some additional insights on how the transcription factor Tec1 functions in the yeast cell. In the previous chapter, the basics of synthetic biological circuit engineering was described together with the yeast mating pathway. It was also outlined how the signal from membrane impermeable mating pheromones is transmitted to the cell nucleus via mating pathway by gradual phosphorylation of intermediate kinases. It was described how the mating pathway is used in synthetic circuits to directly activate target genes or indirectly repress the target gene in multistep fashion. What is missing is to exploit analogous mechanisms to introduce direct NOT operation to the presence of alpha-factor. This section proposes the way how to exploit the Fus3-Tec1 system to introduce direct alpha-factor NOT operation.

2.1.1 Repression

The intended repression mechanism is simple: after the pheromone response gets to the point where Fus3 is activated, the transcription factor Tec1 is

marked for active degradation and so its concentration drops. The missing puzzle piece is to find target promoters whose expression is mediated by Tec1 independently. Expression of such genes should be directly repressed by the alpha-factor induction.

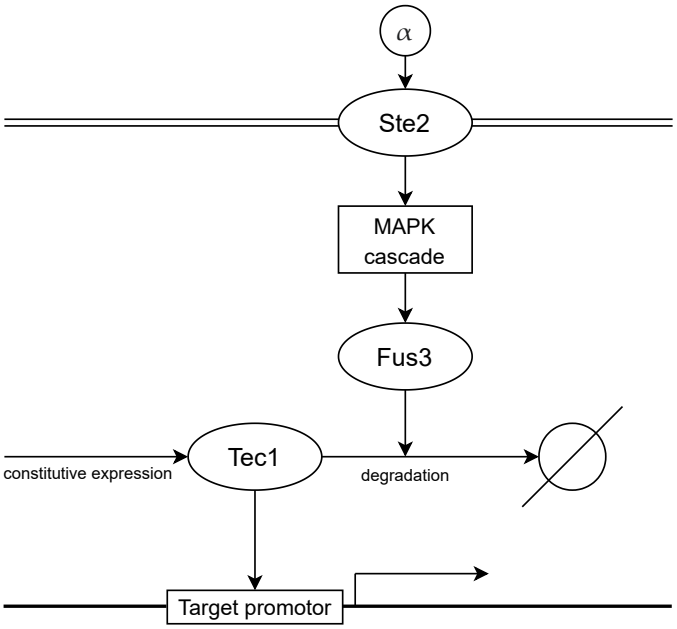


Figure 2.1: **Repression Mechanism.**

2.2 Dataset Creation

To find out which wild yeast promoters should be selected as candidates, it was necessary to gather and organize relevant data with useful information about the regulatory properties related to the alpha-factor induction and the Tec1 transcription factor interaction. The objective of this work was to computationally analyze as many wild promoters as possible, so the search for data was focused on the information that is generally applicable rather than specific studies of particular promoters. Selected data sources can be generally divided into two categories, the first group is genomics data and the second is transcriptomics data.

2.2.1 DNA Sequence Data

DNA sequences of wild promoters of the Saccharomyces cerevisiae were obtained from the NCBI [29]. Sequences of the promoters start 1000 base pairs (bp) upstream to the start codon of the relevant gene and end directly before the start of the open reading frame (ORF). In order to be able to analyze the promoter sequences with respect to the interaction with the transcription factor Tec1, a relative binding affinity matrix (RBAM) was constructed based on data measured in [24]. Values of the matrix were obtained experimentally by creating a DNA library consisting of all possible point mutations of the Tec1 consensus binding sequence (ACATTCTT) and measuring the relative binding rate of the Tec1. The actual matrix is shown in the table 2.1.

Mutation \ Consensus	A	C	A	T	T	C	T	T
A	**1.00**	0.52	**1.00**	0.04	0.10	0.05	0.11	0.29
T	0.21	0.06	0.03	**1.00**	**1.00**	0.06	**1.00**	**1.00**
G	0.56	0.11	0.08	0.03	0.04	0.05	0.15	0.32
C	0.28	**1.00**	0.03	0.03	0.04	**1.00**	0.58	1.00

Table 2.1: **Tec1 relative binding affinity matrix**. Source: [24].

Each element of the matrix corresponds to the relative binding affinity of a consensus sequence with a single substituted letter on a corresponding index.

2.2.2 Omics Data

Sequences alone are by no means not enough to select the right promoters since the presence of the Tec1 binding motif is not sufficient condition for a gene to be activated by Tec1. Considering this, the dataset was enriched for information about changes in yeast transcriptome when induced by an alpha-factor. The data was taken from the study [30]. The data obtained via DNA microarray contains relative expression rates with and without alpha-pheromone induction of more than 97 % of Saccharomyces c. genes. From

a variety of different measurement conditions analyzed in the study two instances were reused in this work. In the first setup the treated cell lines were induced with 50 nM alpha-factor and measured after 0, 15, 30, 45, 60, 90 and 120 minutes of incubation. In the second setup, treated cells were all incubated for 30 minutes but induced with different concentrations of alpha-factor ranging from 0.15 nM to 500 nM. The actual values used in this dataset represents log2 ratio between expression levels measured in alpha-pheromone treated and mock treated cell lines.

As a supplementary information to the relative expression levels, the absolute mRNA abundance under optimal environmental conditions was added to the dataset from [31]. This data was obtained via bulk mRNA sequencing and the actual values used in this dataset represents the number of mRNA molecules per pg of dry cellular weight.

2.3 Computational Method

Ater the dataset was established, a simple computational method was developed to extract the useful information from the promoter DNA sequences and reduce the transcriptomics data into a simple score by which the promoters could be filtered and sorted out.

2.3.1 Sequence Analysis

The constructed RBAM contains information about how the relative binding affinity changes with single point mutations of consensus binding sequence. Let $L(x)$ be a utility function which maps a 8 bp long DNA sequence to a scalar utility corresponding to a relative binding affinity to the Tec1, then each element of the RBAM corresponds to the utility score for a consensus sequence with the single given substitution. From all the possible 4^8 (65k) input sequences, only for $4 \cdot 8$ (32) of them the utility score is known. To estimate the utility for the rest of the possible sequences it was assumed that all of the point mutations impair the binding affinity independently of each other and so that the score can be computed as a scalar product of all partial scores of individual letters in the sequence. To formulate this mathematically, the

encoding of a given sequence needs to be established first. The encoding was chosen to be one-hot encoded individual letters stacked next to each other, the encoding of individual letters is following

$$\begin{aligned} \mathrm{A} &\rightarrow (1,0,0,0)^T \\ \mathrm{T} &\rightarrow (0,1,0,0)^T \\ \mathrm{G} &\rightarrow (0,0,1,0)^T \\ \mathrm{C} &\rightarrow (0,0,0,1)^T. \end{aligned} \tag{2.1}$$

and the encoding of a whole sequence is demonstrated in the following example. Let's mark S a randomly generated DNA sequence

$$\mathrm{ATCTGGAT} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = S_{example}. \tag{2.2}$$

And let M be the relative binding affinity matrix

$$M = \begin{bmatrix} 1.00 & 0.52 & 1.00 & 0.04 & 0.10 & 0.05 & 0.11 & 0.29 \\ 0.21 & 0.06 & 0.03 & 1.00 & 1.00 & 0.06 & 1.00 & 1.00 \\ 0.56 & 0.11 & 0.08 & 0.03 & 0.04 & 0.05 & 0.15 & 0.32 \\ 0.28 & 1.00 & 0.03 & 0.03 & 0.04 & 1.00 & 0.58 & 1.00 \end{bmatrix}, \tag{2.3}$$

then the resulting utility function has the following form

$$L(S) = \mathrm{mtr}\left(S^T \cdot M\right), \tag{2.4}$$

where operator mtr stands for multiplicative trace which transform a $n \times n$ square matrix to a scalar by a formula

$$\mathrm{mtr}\,(A) = \prod_{i=1}^{n} a_{ii}, \tag{2.5}$$

where a_{ii} denotes the element on the ith row and ith column of A. The utility of the example sequence is

$$L(\mathrm{ATCTGGAT}) = 3.96 \cdot 10^{-7}. \tag{2.6}$$

In order to evaluate the whole 1000 bp long promoter sequence, a sliding window of the size of 8 bp was used to gradually evaluate each of the 992 possible subsequences. For each segment the score was computed as well as for its reverse DNA complement. From the resulting array of 992×2 scores, additional three values were extracted for the filtering purposes, the number of ≥ 0.99 scores, the number of ≥ 0.5 scores and the number of ≥ 0.2 scores.

The graph on figure 2.2 shows the evaluation of a randomly selected promoter from the database. The positive values correspond to the relative affinity binding to a forward DNA strand and the negative values stand for the binding affinity to the reverse DNA strand.

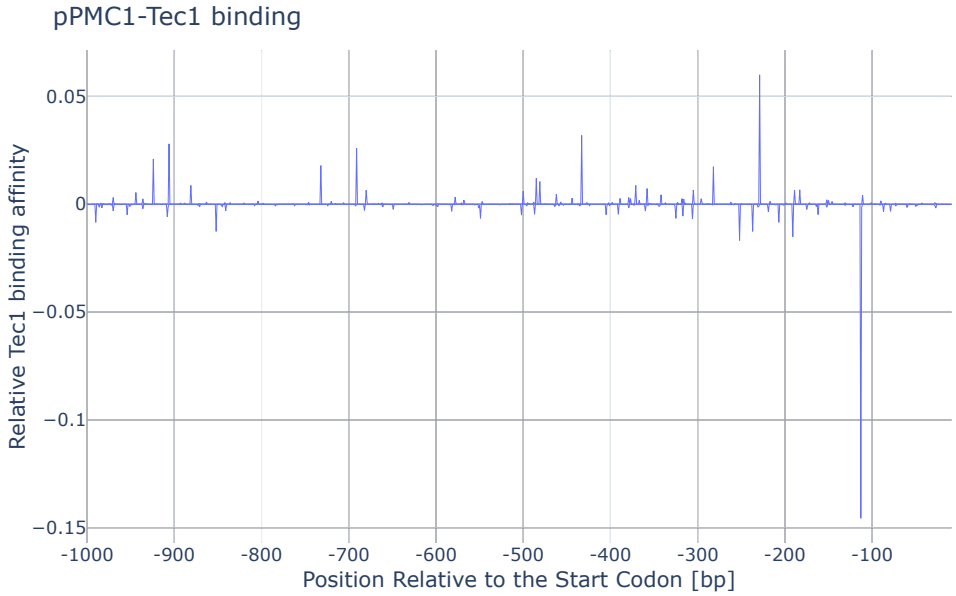


Figure 2.2: **Tec1 relative binding affinity to a randomly selected promoter.** The curve represents the relative Tec1 binding affinity of the promoter on a given position. Positive values represent the score of the forward DNA strand whereas the negative values stand for the reverse DNA strand score.

2.4 Selection process

The data was gathered for a total number of 5690 wild Saccharomyces c. promoters. The first step to select the ones of interest was to filter out the obviously undesirable ones. To address this, a couple of metrics were computed from the original dataset.

In case of the transcriptomics data, the goal was to filtered out genes which either are not repressed at all or the relative expression rates are inconsistent throughout the time, ie. oscillating between high and low rates. From the ratios measured 30, 45, 60, 90 and 120 minutes after 50 nM alpha-factor induction, the mean value and the mean absolute difference between each time step was computed. Genes with mean log2 ratio higher than -1 (less than 2-fold repression) were excluded as well as the genes with mean absolute differences between time steps higher than the actual absolute mean ratio. Motivation for this filter is shown in an example in the figure 2.3.

As previously stated, from the 2×992 binding scores relating to each of the promoter sequences, values exceeding 0.99, 0.5 and 0.2 were counted in order to capture the key information about the promoter sequence. Next step was to filter out all genes whose promoters don't contain even one ≥ 0.5 Tec1 binding site.

After applying these three filters, 49 potential candidates remained from the original 5690 promoters (less than 1 %), the potential candidates are listed in table 2.2. Remaining promoters were then sorted by the mean log2 expression ratio and analyzed manually.

In the first phase of the manual selection, all of the 49 candidate genes were briefly analyzed and a reasoning was applied jointly on all of the collected information. The simple metrics like the mean repression rate and number of Tec1 binding motifs was taken into consideration as well as the raw values of omics data and relative positioning between individual binding sites and transcription start site (TSS). From this step about 10 most interesting genes were chosen. To complete the analysis, the information about the gene's function was studied in the literature. In the end two promoters (GAS1 and CLN1) were finally chosen to be experimentally validated.

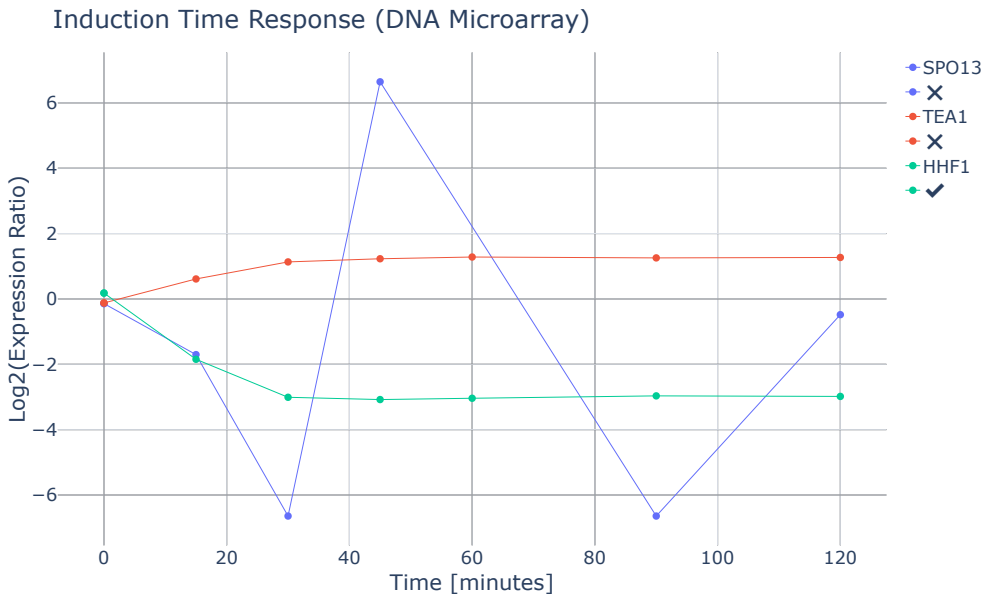


Figure 2.3: **Transcriptomics filter example.** In the figure there are two examples of the excluded genes. The SPO13 is repressed on average, but its expression is inconsistent. The TEA1 reaction to the alpha-factor is consistent, but not in a repressive manner. The only consistently repressed gene is the HHF1 and so it passes the transcriptomics filter.

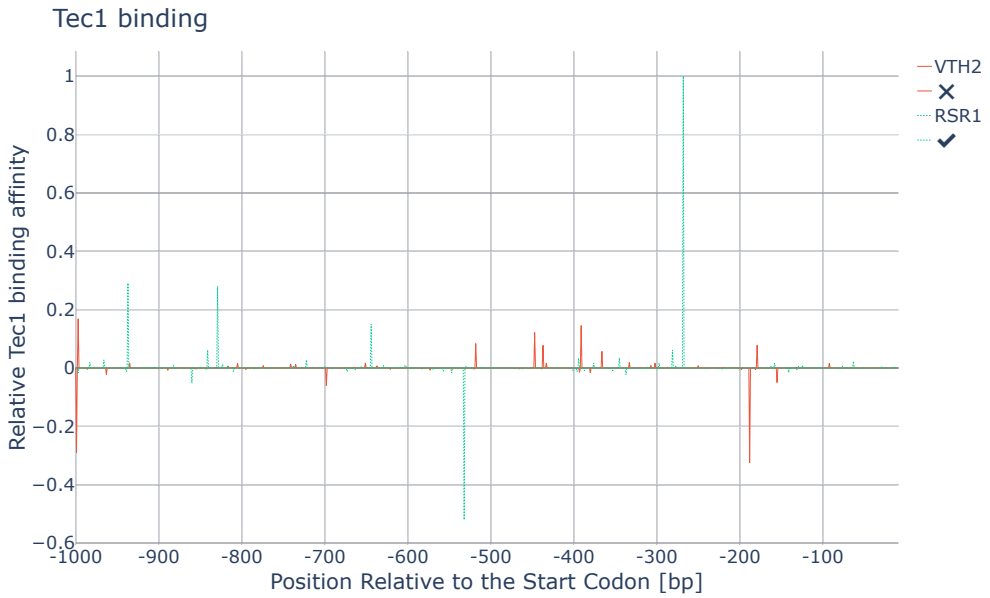


Figure 2.4: **Sequence filter example.** In this example the first promoter VTH2 doesn't contain any Tec1 binding motif with more than 50% binding affinity relative to consensus sequence and thus is filtered out. The RSR1 promoter contains one 100% and one 50% binding motif so it's not excluded.

2.4.1 Candidate Promoters

The potential promoter candidates are listed in table 2.2. Thanks to the computational filters applied before, the list was small enough for all the genes to be briefly manually investigated. This section contains examples of five proteins and their annotation together with hypotheses how it could be related to the Fus3-Tec1 system. The CLN1 and GAS1 are in terms of their gene ontology most likely to be in direct relation to the Fus3-Tec1 system and thus were selected for further experimental validation.

RTT109

Rtt109 (YLL002W) is required for acetylation of histone H3 on two residues, K9 and K56 and plays an important various kinds of stresses related to DNA damage [32, 33]. However the mechanism by which the Rtt109 could be regulated by Tec1 is not obvious.

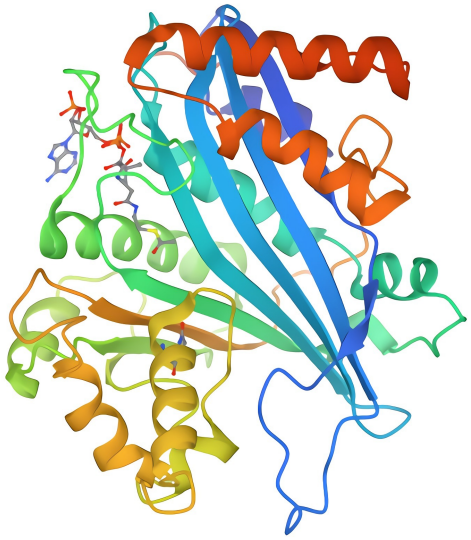


Figure 2.5: **Crystal structure of RTT109-AC-CoA complex (3QM0).** Source: [34]

MMR1

MMR1 (YLR190W) is a phospholipid binding protein, which interacts with mitochondria and with Myo2p, functioning as an adaptor that recruits Myo2p and facilitates actin-based transport of mitochondria to the bud [35, 36]. However the mechanism by which the Mmr1 could be regulated by Tec1 is not obvious.

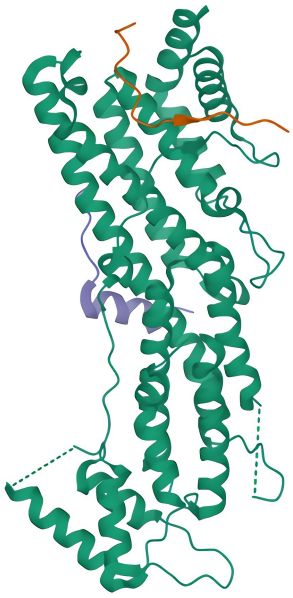


Figure 2.6: **Structure of Myo2-GTD in complex with Mmr1 (6IXP).** Source: [37]

PDS5

Pds5 (YMR076C) is a cohesion maintenance factor which is involved in sister chromatid condensation and cohesion. It regulates homolog pairing and facilitates synaptonemal complex formation, axis formation, inter homolog recombination and synapsis during meiosis [38, 39]. However the mechanism by which the Pds5 could be regulated by Tec1 is not obvious.

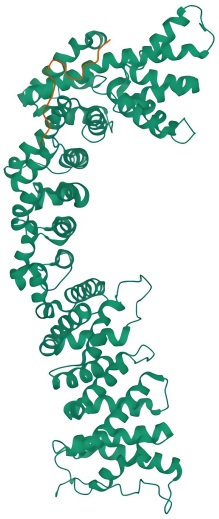


Figure 2.7: **Structure of the Pds5-Scc1 complex and implications for cohesin function (5FRS)** Source: [40]

GAS1

The Gas1 (YMR307W), a Beta-1,3-glucanosyltransferase participates in the cell wall assembly. Gas1Δ mutants suffer from several cell wall morphogenetic defects like abnormal cell shape and size and generally lower cell wall integrity. Gas1Δ cells are not capable of proper filamentous and pseudohyphal growth but conversely are oversensitive to mating pheromones [41–44]. The fact that the GAS1 is required for filamentous growth is consistent with the possibility that it is activated by Tec1 and thus repressed during the mating by the Fus3-Tec1 system.

CLN1

Cln1 (YMR199W) gene is known for its regulatory function in the yeast cell cycle. Cln1 promotes the transition from G1 phase to S phase by activating Cdc28 kinase in cooperation with other G1 cyclins (Cln2, Cln3) [45–49]. Since the Cln1 plays a key role in the transition from G1 to S phase, it makes sense that its expression is shut down during the pheromone response which causes arrest in the G1 phase. Thus the Fus3-Tec1 system could theoretically be the mechanism by which this expression shutdown is taking place.

ORF Name	Gene Name	mRNA per pgDW	Mean Expression Ratio	>99% bindings	>50% bindings	>20% bindings
YMR199W	**CLN1**	**0.7219**	**-3.8820**	**0**	**1**	**4**
YCL014W	BUD3	0.3836	-3.2050	0	1	1
YBL002W	HTB2	1.9514	-2.8947	0	2	2
YMR144W	FDO1	0.3462	-2.8868	0	1	1
YOR373W	NUD1	0.3998	-2.8323	0	1	2
YBL003C	HTA2	1.3130	-2.8137	0	1	1
YDR224C	HTB1	3.5547	-2.7831	0	1	1
YHR154W	RTT107	0.4358	-2.6064	0	1	3
YOR247W	SRL1	1.0607	-2.3606	0	1	2
YJR092W	BUD4	0.4772	-2.3234	0	1	1
YDR451C	YHP1	0.3258	-2.2862	0	1	1
YPR149W	NCE102	8.5865	-2.0696	0	1	2
YMR215W	GAS3	0.6472	-1.9739	0	1	1
YOR074C	CDC21	0.3902	-1.8031	0	1	1
YPL163C	SVS1	0.6646	-1.7732	0	4	7
YNL102W	POL1	0.3868	-1.5852	0	2	2
YBR093C	PHO5	NULL	-1.5653	0	2	2
YER145C	FTR1	1.5586	-1.5480	0	1	2
YNL327W	EGT2	1.1462	-1.5155	1	1	1
YPL014W	CIP1	0.2473	-1.4789	1	2	2
YLR212C	TUB4	0.4209	-1.4769	1	2	3
YDL179W	PCL9	0.3315	-1.4583	0	1	1
YLR190W	MMR1	0.8005	-1.4484	0	1	1
YJL187C	SWE1	0.2281	-1.4371	1	2	2
YOR315W	SFG1	0.5559	-1.4098	1	2	2
YDR488C	PAC11	0.1719	-1.4012	0	1	2
YMR307W	**GAS1**	**4.0689**	**-1.3567**	**0**	**1**	**2**
YLR049C	MLO50	0.4219	-1.3421	0	4	4
YMR006C	PLB2	1.0479	-1.2863	0	1	1
YEL061C	CIN8	0.4111	-1.2823	0	1	1
YJL158C	CIS3	1.6226	-1.2504	0	1	2
YMR274C	RCE1	0.1443	-1.2424	0	3	3
YMR078C	CTF18	0.4303	-1.2125	1	1	2
YBL031W	SHE1	0.1599	-1.2105	1	3	3
YNL160W	YGP1	4.7896	-1.2105	0	2	2
YAR008W	SEN34	0.2690	-1.2072	0	1	1
YPL018W	CTF19	NULL	-1.1766	1	1	1
YPL124W	SPC29	0.1823	-1.1594	0	1	1
YGR152C	RSR1	0.4758	-1.1281	1	2	4
YOL007C	CSI2	0.3941	-1.1095	0	1	1
YHR149C	SKG6	0.4532	-1.1002	0	1	2
YJR048W	CYC1	2.3193	-1.0663	0	1	2
YLL021W	SPA2	0.7344	-1.0537	1	1	1
YIL140W	AXL2	0.7143	-1.0531	1	1	1
YMR076C	PDS5	0.5549	-1.0511	0	1	1
YCL061C	MRC1	0.3502	-1.0497	0	2	3
YBL009W	ALK2	0.4587	-1.0464	0	2	4
YPL155C	KIP2	0.2837	-1.0112	0	2	2
YLL002W	RTT109	0.2170	-1.0059	0	2	2

Table 2.2: **Potential promoter candidates.**

3 Experimental Validation

In the previous chapters the background knowledge of the given topic was established. The computational tool was used to select candidate genes with potentially desired promoters. In silico design based on genomics and experimental transcriptional data cannot be trusted completely. Next logical step was to validate the proposed mechanism by laboratory experiments. This chapter describes the methods which were used to confirm the NOT operation programmed into selected promoters and the process of how the experiments were conducted.

3.1 Experimental Setup

Selected candidate promoters were laboratory tested for the ability to repress genetic transcription upon alpha-pheromone induction. Wild promoters were extracted from yeast strain with no genetic modifications in the target locus. They were assembled into the genes consisting of the promoter itself, ORF of the reporter protein and a terminator. Constructed genes were transformed into MATa mutants with knocked out standard selection auxotrophic markers, adhesion, pheromone secretion and pheromone proteolysis genes (*his3*Δ, *leu2*Δ, *met15*Δ, *ura3*Δ, *aga2*Δ, *mfa1*Δ, *mfa2*Δ, *bar1*Δ). Expression levels were then measured in conditions with and without the presence of the alpha-pheromone.

The reporter protein was chosen to be β-lactamase (BLA), but after the first trials, the pCLN1 showed almost no signal even when the culture was diluted to double concentration and incubated for a longer time, so the CNL1 promoter was assembled with Luciferase (NanoLuc) as a reporter in the end. After this change more than sufficient signal levels were measured with pCLN1. The pGAS1 was measured via BLA with no problems.

3.2 Materials and Methods

This section describes the methods which were used to laboratory extract and assemble individual biological parts and strains.

3.2.1 Yeast Strains

Strain	Relevant genotype	Reference
BY4741	MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	[50]
S0	BY4741 *bar1Δ mfa1Δ mfa2Δ aga2Δ*	XENO
S001	S0 *his3*::*pCLN1-NanoLuc*	This work
S002	S0 *his3*::*pCLN1-BLA*	This work
S003	S0 *his3*::*pGAS1-BLA*	This work

Table 3.1: **Yeast strains used in this work.**

All yeast strains used in this work are listed in table 3.1. BY4741 mutant with bar1Δ, mfa1Δ, mfa2Δ and aga2Δ knock-outs provided by XENO Cell Innovations s.r.o. (XENO) was used as a background strain for the created library of yeast strains. Yeast strains carrying relevant integrated genes were obtained using High efficient yeast transformation and verified by PCR on agarose gel and by functional test for presence of reporter molecule. All of the integrated genes were targeted to the locus his3.

3.2.2 Plasmid Construction

Plasmid	Alias	Description	Reference
P0		domestication cassete	XENO
P1		his3-based integrative vector	XENO
P2		BLA	XENO
P3		NanoLuc	XENO, [51]
P4		terminator	XENO
P005		P0 *pCLN1-part1*	This work
P006		P0 *pCLN1-part2*	This work
P007		P0 *pCLN1-part3*	This work
P008		P0 *pCLN1-complete*	This work
P009		P0 *pGAS1*	This work
P010	*pCLN1-BLA*	P1 P008 P2 P4	This work
P011	*pCLN1-NanoLuc*	P1 P008 P3 P4	This work
P012	*pGAS1-BLA*	P1 P009 P2 P4	This work

Table 3.2: **Plasmids used in this work.**

All plasmids used in this work are listed in table 3.2. All plasmids were constructed using Golden Gate assembly and MoClo cloning method [52, 53]. At first, promoters were extracted from the S0 strain genome via PCR and integrated into the P0 domestication cassette plasmid. Since the wild pCLN1 promoter DNA sequence contains two unwanted BsmBI restriction recognition sites, the promoter was extracted as three parts and then assembled into a single sequence containing two point mutations in the positions -518 and -608 relative to the start codon disrupting the recognition sites. Obtained promoters were then built into the gene plasmid consisting of his3-based integrative vector, reporter protein BLA or NanoLuc and a terminator.

3.2.3 Dose-Response Measurement

BLA Essay

Yeast strains carrying constructed genes were grown to exponential phase in YPD medium. Before dilution to 0.05 OD, cells were washed two times

in 1X PBS 0.1% gel and resuspended in YPD again. Half of the replicates were then treated with 1 μM alpha-factor. Pheromone treated as well as the mock cell lines were incubated in 30°C with mixing for 2 hours. Incubated cultures were then centrifuged to isolate supernatant with dissolved reporter molecules. Supernatant was then mixed with 1.25X nitrocefin in ratio 1:4 to make 100 μL of final measurement solution. Light absorption in wavelengths 486 nm (cyan) and 700 nm (red) was measured 20-times with one minute period. 486 nm is an absorption wavelength of a degraded nitrocefin created after reaction with BLA whereas 700 nm is used to subtract the absorbance of the background. After the subtraction of the 700 nm absorbance, the first value of the series is subtracted as well so the curve was located in the origin of the coordinate system. Slope of the linearly approximable early phase of the curve was computed to estimate the relative expression level [54].

NanoLuc Essay

The initial treatment up to the supernatant isolation was the same as in the case of BLA dose-response essay. 20 μL of supernatant was then mixed with 50 μL luciferase buffer, 29 μL distilled water and 1 μL of 1X furimazine. The 470 nm (light blue) luminescence of the resulting mixture was measured 10-times with one minute period and averaged to get the final values.

3.2.4 Source Code

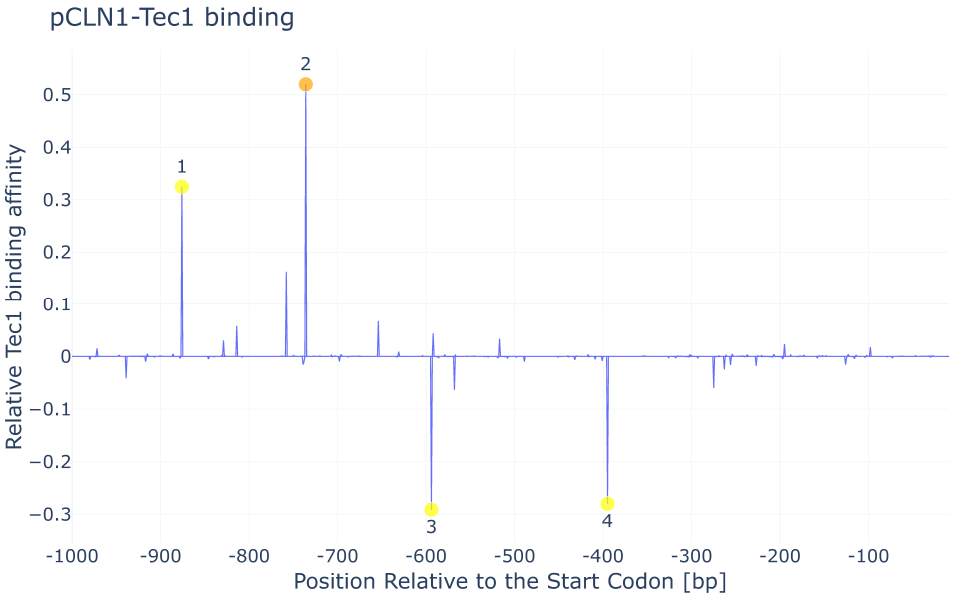
The source code is available on `https://github.com/thonzyk/Master-Thesis`.

4 Results

This chapter summarizes both computational and experimental results. From the final list of candidate promoters two were chosen for the experimental validation, one of them GAS1 from a gene coding a Beta-1,3 glucanosyltransferase required for cell wall assembly and second CLN1 a G1 cyclin involved in regulation of the cell cycle. Both of the selected promoters have several TEC1 binding motifs and promising negative expression ratio measured by microarrays.

4.1 Computational analysis

Following figures 4.1 and 4.2 show the analysis of the promoter sequence. The (a) subfigure shows the relative binding affinity to the of the Tec1 to the promoter sequence, here the x axis represents the coordinate in the promoter sequence in terms of number of base pairs from the start codon, the absolute value of y axis represents the relative binding affinity of the Tec1 to the given region and the sign of the value represents whether the binding site is presented on the forward (positive) or reverse (negative) strand of the DNA. The (b) subfigure shows the actual DNA sequence of the given promoter, where binding sites of Tec1 are highlighted together with other basic annotations of the promoter sequence such as TATA-box, Transcription start site (TSS) and start codon of the gene.



(a)

GAATGTTACGGGACTAACAGGCGGATGTAAATTACTCACTTAAACGC
AGCCAAACATCATCGAGAACTTAGGGTAGCGTGCCACAAAATTTGCA
TGAATAAACTTTTGTTTTCCTAATTCGACAGCATTCCC(1)TTGTTC
GCAACACTTCACTGATAGGAAATCGAATAGCGCACACTCTCTTCTGG
GACATACCCCAATGCGGTAAAGCCACGAAAACACCGCGCGTAAAGGG
GTAAACAAGTCCATTCCTACAACCTCTTGGAGAAATTCTT(2)TACC
TACTACAACCCCGCGCCTGATACTTTCAGTATTCATGACAACTCGAG
CCAGATCCCGCTCGTGGGCGTGTTCATTCTGTGACGATCCACTAGCG
ACTTCTTTGTTCAGCCTGCAAGAGACGCGTTCAAGGAAGAATTC(3)
GCGATTTTACTTCTTCGAGGGAATCTCGCACCGCGTTAGTTAGTTTC
CAACCTTGAAAGCATCGGAGACGCATTTTTGGCGATTTTGCTGGATT
GAGCTGAATGGTGCCAGGTCGAGGCTGGGAGGGAGACTAACTCGAAA
GTGACGAAGACTCGAAAATTAAAAAAAAAGATACTGCAGAAGGCAAG
ATTGAGAATGG(4)AGTAAAGGCAGCGTGGGTCCCCTGTGGAAACCG
CAGTTTTCCTGCGCCAAGTGGTACCGGTGCGAGTGCAGCAATTAATC
TCTCGATATTTTCTTAGTATCTCTTTTTATATAAGAATATATTTTGG
AATTGGTAATGCTTATCTTCAATAGTTTCTTAGTTGAATGCACACTT
AAGAGCAAATTGGCCAAGGAGTTCTTCGTTCGCTTTAATTTATTTCC
TGGTTATTGTCAATTTATTCATCCCATCTCCCCAGGATAGAAGAAAT
TAGTGTAATTTTGCTGACAATACATTTTAACGACGATAACAATAATA
GCAATTAAATAAAATAGCACTACCACCACTCCACTGCTCGTTAGCTA
TTTCTGTAAAATAAATAAAAAGATCATG

XXX >25% TEC1 binding
XXX >50% TEC1 binding
XXX >99% TEC1 binding
XXX TATA-box
XXX Transcription Start Site
XXX Translation Start Codon

(b)

Figure 4.1: **CLN1 promoter annotation.** The TATA-box and the TSS annotations were taken from [55].

(a)

ACATTTGGCATGGTGGAATGGTAAATGGTACAATACGGGCATGGGCT
GGTCTATCATTTTACAGCCCATACGAGAATACTTCGTGAAGATTATG
GAATCTAGTTATTTTGCAGCAGATTTCTTTCTTGGTCACTTCCTCTT
ATTTATACAAACGCCCATAATATTGTTACCATTTATCGATTACTGGC
ATACAATGGTGCTATTTTGGATGAATCCAAGAAGTATTATCGCGCAT
AAAAGAATTT(1)TGACTCGAAAGCAGAGAGCATTAAGAAGTAGGAT
AGTTTCAAAATATTTCTCTCTATATTTTGTGATGCTAGGTGTGCTCC
TTTTTATGCTCATCGCGCCATTCTT(2)TGCTGGTGATTTCGTATCT
AGTCCGCAAGAATTGCTGGAAGGCACCCTTTTCGAAGGAATCTTCCA
ACCTAACAACCAAAACAACAACGATACTGGTCCAAATGCTCCATCGA
CAATCCTGACTACAACACCAACTTTACCTACCTTTAGGACTGTCGCA
TAGGGATAAAAAAAAAAAAGTTATTTTGAGTAGCTGATAAAGCGAGCT
GGTGCCTATCATAGCCGGCTCAGACTTTTTATGAATTCACAGGCCAG
CCCTGGCTATTCTTTTGCGTACTTTTAGTTCGATATATTTTCGCGGC
TCGCGTTTTGTTTGCTTCTTATTTTACACTGAGTTTTCGTGCCGCAA
ACGTGGAGATGGGAAAAAGAAAAGTCGGGAAAATAATGAGAAATTTC
TACTTTTGGTATTCCTCATACAGCCTGCGCGGTTTATTAGTAAAATA
CCCGATAATCCTCGAGGTTTGAAAAACTTTTCCCTCTACTACTGTTG
ACACGGATTTTTTTATTTAAGAGGAAAAGTCGTGGTTGTTTTCCTCG
AACAAATTAGATATCCATAAATAGTTGTGTCGTTTTATTAAGCTATT
TCAAAATCAGTTTTTATTTTTAAAGTCTGATAAAACAAAAACAACAA
ACACAGCTAAATCTCAACAATG

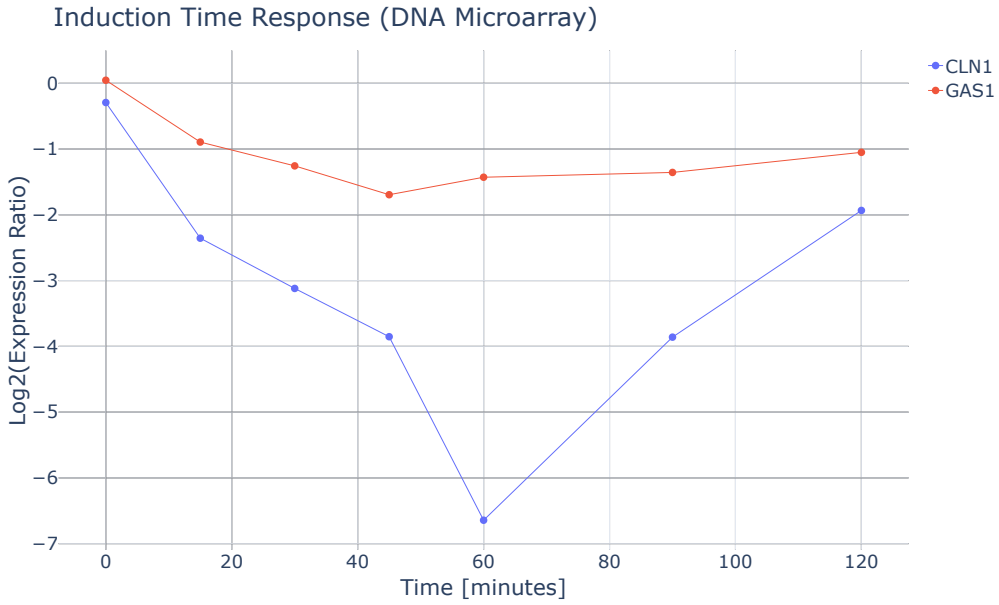
XXX >25% TEC1 binding
XXX >50% TEC1 binding
XXX >99% TEC1 binding
XXX TATA-box
XXX Transcription Start Site
XXX Translation Start Codon

(b)

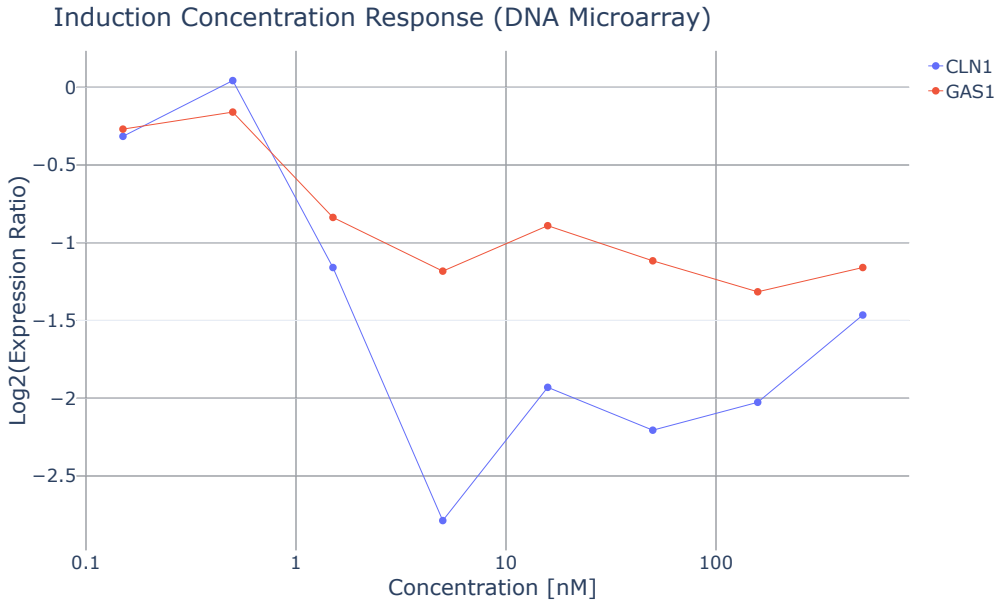
Figure 4.2: **GAS1 promoter annotation.** The TATA-box and the TSS annotations were taken from [55].

The figure 4.3a shows the DNA Microarray analysis of the time response of selected promoters. The x axis is time in minutes and y axis is log2 expression ratio of the measured expression of cells treated and not treated with alpha-factor, in other words negative values represents log2 fold repression

rate. The figure 4.3b is analogous, with the difference that the x axis represents concentration of alpha-factor instead of time.



(a)



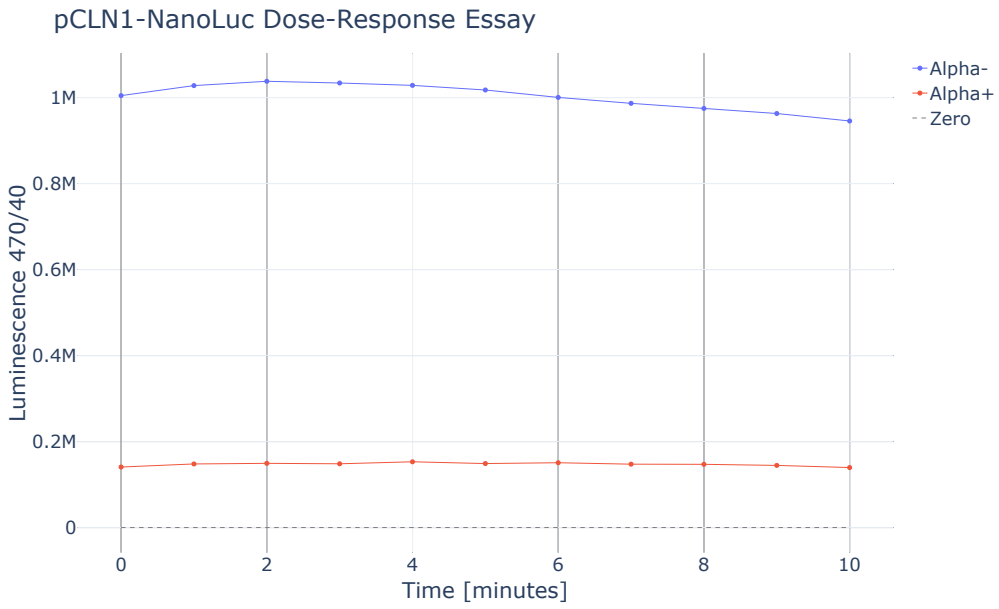
(b)

Figure 4.3: **Selected Promoters DNA Microarray Analysis.**

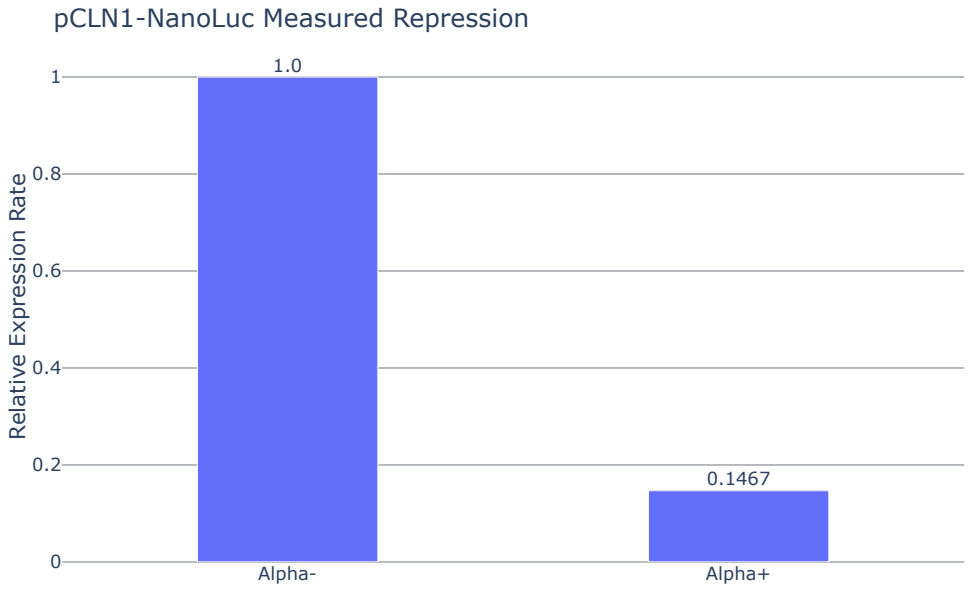
4.2 Experimental Analysis

The experimental validation was executed by measuring the difference in concentrations of reporter molecules in case with and without alpha-factor induction. In case of the lucipherase, the concentration of the molecule is proportional to the luminescence, thus the expectation for measured values is to be constant and the expression rate can be estimated as a mean value of all measured values. In case of BLA what is actually measured is the concentration of the molecules resulting from reaction of nitrocefin and BLA. The concentration of BLA is thus proportional to the speed of this reaction and can be estimated from the slope of the concentration of degraded BLA.

Both promoters showed significant levels of pheromone induced transcription repression. The pCLN1 gene was repressed almost 7-times while the pGAS1 gene was repressed nearly 2-times. Figure 4.4 shows the experimental result of measuring the alpha-factor dose response of the CLN1 promoter. Subfigure (a) shows time series of the raw measured values of luminescence 470/40 and the (b) subfigure visualizes to fold-induction of measured reporter protein, where alpha- column corresponds to control well without alpha factor whereas the alpha+ columns corresponds to the well with 1μM alpha-factor. The same applies to the figure 4.5. The only difference is the way how the raw data is used to estimate the repression rate. In the case of measuring luciferase, the expression level was simply computed as a mean of all the measured values throughout the time. In case of the BLA the expression rate is estimated as the slope of the curve in the initial linear phase, which is in the case of this measurement 20 minutes.

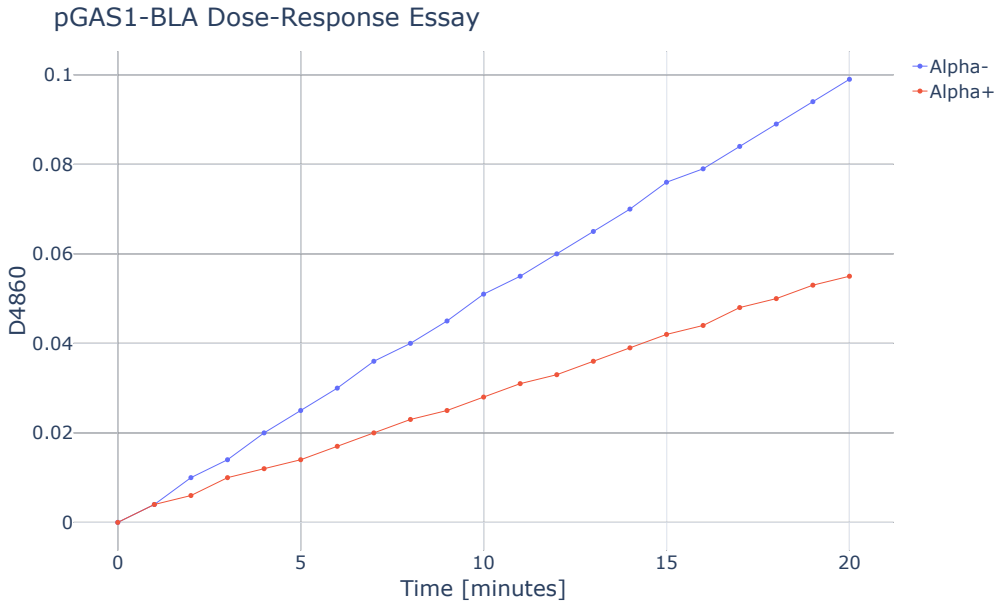


(a)

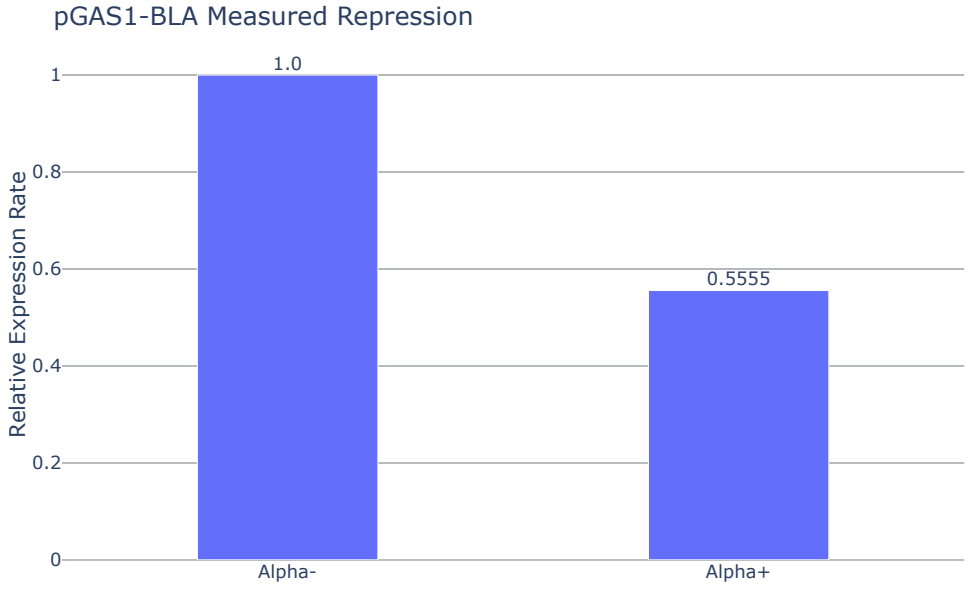


(b)

Figure 4.4: **pCLN1 Experimental Validation.**



(a)



(b)

Figure 4.5: **pGAS1 Experimental Validation.**

Discussion

This work introduced a biological mechanism which could be used for engineering a pheromone inducible repression which can be thought of as a NOT gate with respect to alpha-pheromone presence. Based on theoretical knowledge a dataset containing useful information about promoter sequences and omics was created. After the creation of the dataset a simple computational tool for selecting wild Saccharomyces cerevisiae promoters with desired repression upon alpha-factor induction was formulated and implemented. The tool was able to filter out more than 99 % original promoters and allowed them to sort out and show summarizing information about the remaining ones. The candidate promoters were manually inspected in order to select a subset of the candidates for experimental validation. The experiments were done on two yeast promoters, one stronger GAS1 and one weaker CLN1. Selected promoters were extracted from the yeast strain which had wild type sequences in the target loci. Extracted promoters were then assembled into functional reporter genes and transformed into yeast. The constructed genes integrated in the resulting strains were experimentally tested for signs of pheromone induced transcriptional repression. All of the experimentally tested promoters showed significant expression repression when treated with alpha-factor. The stronger promoter was repressed approximately 2-fold and the weaker one 7-fold. This means that the promoters are with some limitations usable in synthetic circuits.

Future work should confirm or disprove that the repression mechanism is really happening according to the theoretical assumptions. If the promoters are truly repressed by the proposed mechanism, their sequences could be further optimized for higher repression rates. The established tool could be also used to create a larger promoter library, given the fact that the set of potential candidates proposed by the tool is unexhausted. Another useful thing to research in the future would be to analyze the promoter related expression dynamics with respect to time and pheromone concentration.

Bibliography

[1] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. A. Voigt, "Genetic programs constructed from layered logic gates in single cells," *Nature*, vol. 491, no. 7423, p. 249–253, 2012.

[2] X. J. Gao, L. S. Chong, M. S. Kim, and M. B. Elowitz, "Programmable protein circuits in living cells," *Science*, vol. 361, no. 6408, p. 1252–1258, 2018.

[3] R. Entus, B. Aufderheide, and H. M. Sauro, "Design and implementation of three incoherent feed-forward motif based biological concentration sensors," *Systems and Synthetic Biology*, vol. 1, no. 3, p. 119–128, 2007.

[4] S. Chen et al., "Automated design of genetic toggle switches with predetermined bistability," *ACS Synthetic Biology*, vol. 1, no. 7, p. 284–290, 2012.

[5] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, p. 335–338, 2000.

[6] A. Carignano, D. H. Chen, C. Mallory, R. C. Wright, G. Seelig, and E. Klavins, "Modular, robust, and extendible multicellular circuit design in yeast," *eLife*, vol. 11, p. e74540, 2022.

[7] V. W. Cornish et al., "A scalable peptide-gpcr language for engineering multicellular communication," *Nature Communications*, vol. 9, no. 1, 2018.

[8] A. Khakhar, N. J. Bolten, J. Nemhauser, and E. Klavins, "Cell–cell communication in yeast using auxin biosynthesis and auxin responsive crispr transcription factors," *ACS Synthetic Biology*, vol. 5, no. 4, p. 279–286, 2015.

[9] M. W. Gander, J. D. Vrana, W. E. Voje, J. M. Carothers, and E. Klavins, "Digital logic circuits in yeast with crispr-dcas9 nor gates," *Nature Communications*, vol. 8, no. 1, 2017.

[10] X. J. Gao, L. S. Chong, M. S. Kim, and M. B. Elowitz, "Programmable protein circuits in living cells," *Science*, vol. 361, no. 6408, p. 1252–1258, 2018.

[11] B. L. Bassler, "Small talk: Cell-to-cell communication in bacteria," *Cell*, vol. 109, no. 4, p. 421–424, 2002.

[12] T. Tabata, "Morphogens, their identification and regulation," *Development*, vol. 131, no. 4, p. 703–712, 2004.

[13] M. Pertea and S. L. Salzberg, "Between a chicken and a grape: estimating the number of human genes," *Genome Biology*, vol. 11, no. 5, p. 206, 2010.

[14] L. Wang et al., "Engineering consortia by polymeric microbial swarmbots," *Nature Communications*, vol. 13, no. 1, 2022.

[15] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded nor gates and chemical 'wires'," *Nature*, vol. 469, no. 7329, p. 212–215, 2011.

[16] M. W. Gander, J. D. Vrana, W. E. Voje, J. M. Carothers, and E. Klavins, "Digital logic circuits in yeast with crispr-dcas9 nor gates," *Nature Communications*, vol. 8, no. 1, 2017.

[17] L. Bardwell, "A walk-through of the yeast mating pheromone response pathway," *Peptides*, vol. 25, no. 9, p. 1465–1476, 2004.

[18] B. Kofahl and E. Klipp, "Modelling the dynamics of the yeast pheromone pathway," *Yeast*, vol. 21, no. 10, p. 831–850, 2004.

[19] H. Pi, C. T. Chien, and S. Fields, "Transcriptional activation upon pheromone stimulation mediated by a small domain of saccharomyces cerevisiae ste12p," *Molecular and Cellular Biology*, vol. 17, no. 11, p. 6410–6418, 1997.

[20] I. Herskowitz, "Map kinase pathways in yeast: For mating and more," *Cell*, vol. 80, no. 2, p. 187–197, 1995.

[21] N. Hao et al., "Combined computational and experimental analysis reveals mitogen-activated protein kinase–mediated feedback phosphorylation as a mechanism for signaling specificity," *Molecular Biology of the Cell*, vol. 23, no. 19, p. 3899–3910, 2012.

[22] R. E. Chen and J. Thorner, "Function and regulation in mapk signaling pathways: Lessons learned from the yeast saccharomyces cerevisiae," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, vol. 1773, no. 8, p. 1311–1340, 2007.

[23] M. J. Winters and P. M. Pryciak, "Analysis of the thresholds for transcriptional activation by the yeast map kinases fus3 and kss1," *Molecular Biology of the Cell*, vol. 29, no. 5, p. 669–682, 2018.

[24] B. Heise, J. van der Felden, S. Kern, M. Malcher, S. Brückner, and H.-U. Mösch, "The tea transcription factor tec1 confers promoter-specific gene regulation by ste12-dependent and -independent mechanisms," *Eukaryotic Cell*, vol. 9, no. 4, p. 514–531, 2010.

[25] KohlerT., S. Wesche, N. Taheri, G. H. Braus, and MoschH.-U., "Dual role of the saccharomyces cerevisiae tea/atts family transcription factor tec1p in regulation of gene expression and cellular development," *Eukaryotic Cell*, vol. 1, no. 5, p. 673–686, 2002.

[26] M. Z. Bao, T. R. Shock, and H. D. Madhani, "Multisite phosphorylation of the saccharomyces cerevisiae filamentous growth regulator tec1 is required for its recognition by the e3 ubiquitin ligase adaptor cdc4 and its subsequent destruction in vivo," *Eukaryotic Cell*, vol. 9, no. 1, p. 31–36, 2010.

[27] S. Chou, L. Huang, and H. Liu, "Fus3-regulated tec1 degradation through scfcdc4 determines mapk signaling specificity during mating in yeast," *Cell*, vol. 119, no. 7, p. 981–990, 2004.

[28] A. Reményi, M. C. Good, R. P. Bhattacharyya, and W. A. Lim, "The role of docking interactions in mediating signaling input, output, and discrimination in the yeast mapk network," *Molecular Cell*, vol. 20, no. 6, p. 951–962, 2005.

[29] E. W. Sayers et al., "Database resources of the national center for biotechnology information," *Nucleic Acids Research*, vol. 50, no. D1, p. D20–D26, 2022.

[30] C. J. Roberts et al., "Signaling and circuitry of multiple mapk pathways revealed by a matrix of global gene expression profiles," *Science*, vol. 287, no. 5454, p. 873–880, 2000.

[31] P.-J. Lahtvee et al., "Absolute quantification of protein and mrna abundances demonstrate variability in gene-specific translation efficiency in yeast," *Cell Systems*, vol. 4, no. 5, pp. 495–504, 2017.

[32] R. Driscoll, A. Hudson, and S. P. Jackson, "Yeast rtt109 promotes genome stability by acetylating histone h3 on lysine 56," *Science*, vol. 315, no. 5812, p. 649–652, 2007.

[33] Y. Voichek, R. Bar-Ziv, and N. Barkai, "Expression homeostasis during dna replication," *Science*, vol. 351, no. 6277, p. 1087–1090, 2016.

[34] Y. Tang et al., "Erratum: Fungal rtt109 histone acetyltransferase is an unexpected structural homolog of metazoan p300/cbp," *Nature Structural Molecular Biology*, vol. 15, no. 9, p. 998–998, 2008.

[35] W. Chen, H. A. Ping, and L. L. Lackner, "Direct membrane binding and self-interaction contribute to mmr1 function in mitochondrial inheritance," *Molecular Biology of the Cell*, vol. 29, no. 19, p. 2346–2357, 2018.

[36] T. Itoh, A. Toh-e, and Y. Matsui, "Mmr1p is a mitochondrial factor for myo2p-dependent inheritance of mitochondria in the budding yeast," *The EMBO Journal*, vol. 23, no. 13, p. 2520–2530, 2004.

[37] K. Tang, Y. Li, C. Yu, and Z. Wei, "Structural mechanism for versatile cargo recognition by the yeast class v myosin myo2," *Journal of Biological Chemistry*, vol. 294, no. 15, p. 5896–5906, 2019.

[38] T. Hartman, K. Stead, D. Koshland, and V. Guacci, "Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in saccharomyces cerevisiae," *Journal of Cell Biology*, vol. 103, no. 3, p. 613–626, 2000.

[39] S. Panizza, T. Tanaka, A. Hochwagen, F. Eisenhaber, and K. Nasmyth, "Pds5 cooperates with cohesin in maintaining sister chromatid cohesion," *Current Biology*, vol. 10, no. 24, p. 1557–1564, 2000.

[40] Kyle W. Muir, M. Kschonsak, Y. Li, J. Metz, Christian H. Haering, and D. Panne, "Structure of the pds5-scc1 complex and implications for cohesin function," *Cell Reports*, vol. 14, no. 9, p. 2116–2126, 2016.

[41] S. A. Chasse et al., "Genome-scale analysis reveals sst2 as the principal regulator of mating pheromone signaling in the yeast saccharomyces cerevisiae," *Eukaryotic Cell*, vol. 5, no. 2, p. 330–346, 2006.

[42] L. Popolo and M. Vai, "The gas1 glycoprotein, a putative wall polymer cross-linker," *Biochimica Et Biophysica Acta*, vol. 1426, no. 2, p. 385–400, 1999.

[43] L. Popolo, T. Gualtieri, and E. Ragni, "The yeast cell-wall salvage pathway," *Medical Mycology*, vol. 39, no. 1, p. 111–121, 2001.

[44] C. M. Kang and Y. W. Jiang, "Genome-wide survey of non-essential genes required for slowed dna synthesis-induced filamentous growth in yeast," *Yeast*, vol. 22, no. 2, p. 79–90, 2005.

[45] J. A. Hadwiger, C. Wittenberg, H. E. Richardson, M. de Barros Lopes, and S. I. Reed, "A family of cyclin homologs that control the g1 phase in yeast," *Proceedings of the National Academy of Sciences*, vol. 86, no. 16, p. 6255–6259, 1989.

[46] F. R. Cross and A. H. Tinkelenberg, "A potential positive feedback loop controlling cln1 and cln2 gene expression at the start of the yeast cell cycle," *Cell*, vol. 65, no. 5, p. 875–883, 1991.

[47] M. Tyers, G. Tokiwa, and B. Futcher, "Comparison of the saccharomyces cerevisiae g1 cyclins: Cln3 may be an upstream activator of cln1, cln2 and other cyclins," *The EMBO Journal*, vol. 12, no. 5, p. 1955–1968, 1993.

[48] F. Cvrcková and K. Nasmyth, "Yeast g1 cyclins cln1 and cln2 and a gap-like protein have a role in bud formation," *The EMBO journal*, vol. 12, no. 13, p. 5277–5286, 1993.

[49] F. R. Cross, M. Hoek, J. D. McKinney, and A. H. Tinkelenberg, "Role of swi4 in cell cycle regulation of cln2 expression," *Molecular and Cellular Biology*, vol. 14, no. 7, p. 4779–4787, 1994.

[50] C. B. Brachmann et al., "Designer deletion strains derived from saccharomyces cerevisiae s288c: a useful set of strains and plasmids for pcr-mediated gene disruption and other applications," *Yeast*, vol. 14, no. 2, pp. 115–132, 1998.

[51] C. G. England, E. B. Ehlerding, and W. Cai, "Nanoluc: A small luciferase is brightening up the field of bioluminescence," *Bioconjugate Chemistry*, vol. 27, no. 5, p. 1175–1187, 2016.

[52] C. Engler, R. Kandzia, and S. Marillonnet, "A one pot, one step, precision cloning method with high throughput capability," *PLoS ONE*, vol. 3, no. 11, p. 3647, 2008.

[53] M. E. Lee, W. C. DeLoache, B. Cervantes, and J. E. Dueber, "A highly characterized yeast toolkit for modular, multipart assembly," *ACS Synthetic Biology*, vol. 4, no. 9, p. 975–986, 2015.

[54] D. Georgiev, M. Cienciala, H. Kasl, L. Berne and T. Puchrova, "Biological computing systems and methods for multivariate surface analysis and object detection," Oct. 14, 2021.

[55] J. McMillan, Z. Lu, J. S. Rodriguez, T.-H. Ahn, and Z. Lin, "Yeastss: an integrative web database of yeast transcription start sites," *Database (Oxford)*, vol. 2019, 2019.

[56] R. Dreos, G. Ambrosini, R. Groux, R. Cavin Périer, and P. Bucher, "The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms," *Nucleic Acids Research*, vol. 45, no. D1, p. D51–D55, 2016.

[57] J. M. Cherry et al., "Saccharomyces genome database: the genomics resource of budding yeast," *Nucleic Acids Research*, vol. 40, no. D1, p. D700–D705, 2012.

[58] J. Usher, "The mechanisms of mating in pathogenic fungi—a plastic trait," *Genes*, vol. 10, no. 10, p. 831, 2019.

[59] B. Chapa-y-Lazo, S. Lee, H. Regan, and P. Sudbery, "The mating projections of saccharomyces cerevisiae and candida albicans show key characteristics of hyphal growth," *Fungal Biology*, vol. 115, no. 6, p. 547–556, 2011.

[60] P. K. Grant et al., "Interpretation of morphogen gradients by a synthetic bistable circuit," *Nature Communications*, vol. 11, no. 1, 2020.

[61] M. B. Elowitz and S. Leibler, "A synthetic oscillatory network of transcriptional regulators," *Nature*, vol. 403, no. 6767, p. 335–338, 2000.

[62] E. T. Wurtzel et al., "Revolutionizing agriculture with synthetic biology," *Nature Plants*, vol. 5, no. 12, p. 1207–1210, 2019.

[63] I. Herskowitz, "Life cycle of the budding yeast saccharomyces cerevisiae," *Microbiological Reviews*, vol. 52, no. 4, p. 536–553, 1988.

List of Abbreviations

bp	base pairs
BLA	β-lactamase
OD	optical density
ORF	open reading frame
RBAM	relative binding affinity matrix
TEAD	TEA DNA-binding domain
MAPK	mitogen-activated protein kinase
DNA	deoxyribonucleic acid
TF	transcription factor
GPCR	G protein-coupled receptor

List of Figures

List of Tables