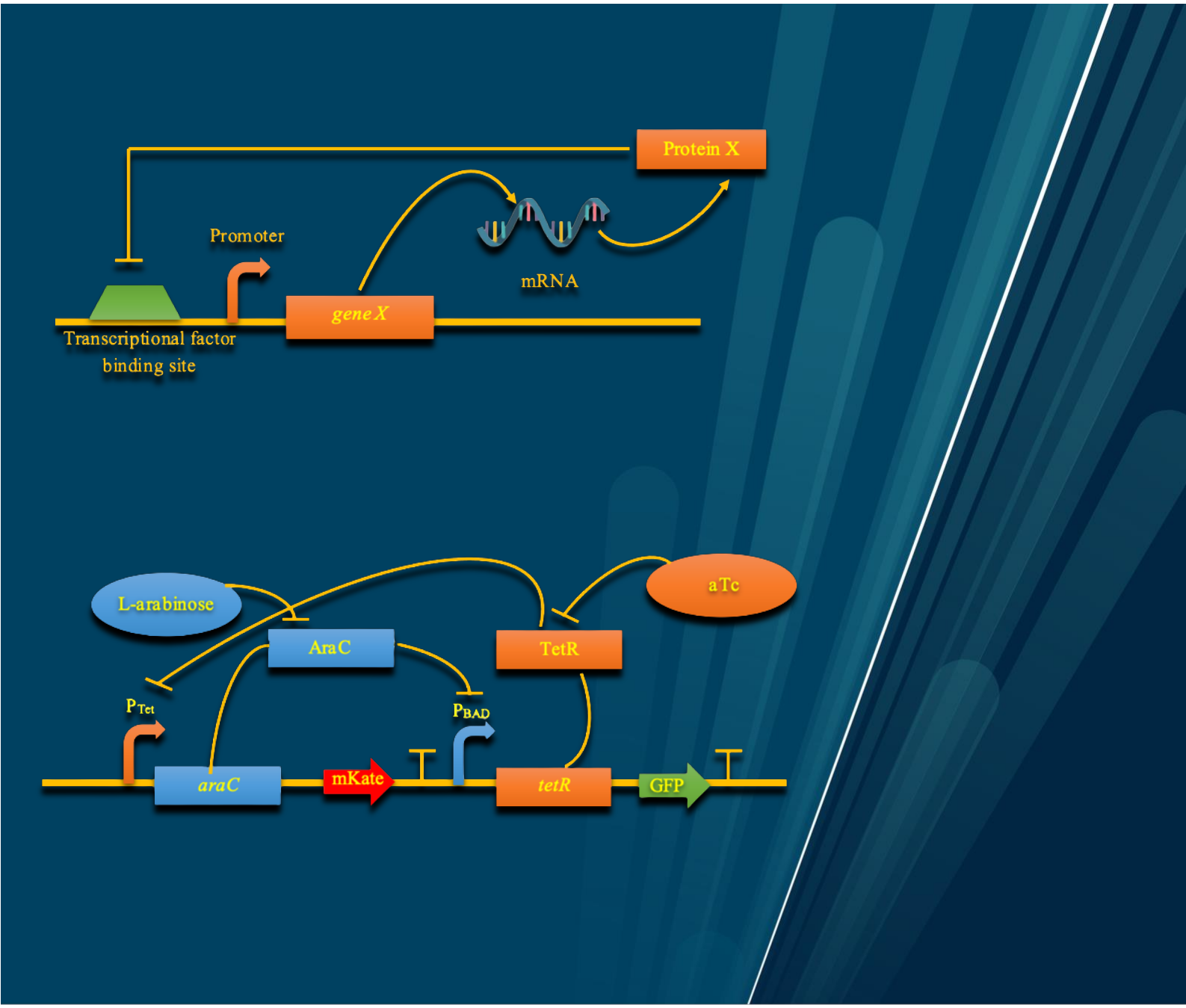
UiT The Arctic University of Norway

The Faculty of Biosciences, Fisheries and Economics

Modeling the "microbial chassis effect" on the performance of a genetic switch

Vanja Verusevski

Master's thesis in Marine biotechnology BIO-3901, March 2022



Acknowledgements

The thesis is submitted for the degree of Master of Science in Marine biotechnology at The University of Tromsø – The Arctic University of Norway. The research presented in this thesis was performed in The M2 Research Group (M2RG) at The Norwegian College of Fishery Science under the supervision of the Associate Professor Hans Christopher Bernstein and PhD candidate Dennis Tin Chat Chan. The Master's thesis constitutes 60 credits of the degree.

My studies started in August 2020 and from the early beginning I received support and guidance from my supervisor and the M2RG team. Hans, thank you for being an amazing supervisor! I would like to express my sincere gratitude to the UiT for the warm welcome and opportunity to study in beautiful Tromsø.

The past two years of living in Norway have helped me to find what my real interests are and what I want to do in my further career. I have been honored to have the opportunity to get to know Norwegian culture, tradition and to learn Norwegian language.

I am thankful to special people from my beloved Serbia that gave me support and fought together with me in this journey: Izabela, Uroš, Katarina, Bojana, Jovana and family Vojinović. Special thanks to my parents Danijela and Kristijan, I am grateful to be your daughter and to have you in my life! Thank you, my grandparents, for all the encouragement. Thanks to Nađa, Ivana, Nenad, Ljiljana and Vojin, that made my life in Norway easier. Special thanks to Dragan, my dearest, that believed in me and gave me all his love, support, and motivation.

I would also like to acknowledge the efforts of my colleagues and professors from The Faculty of Technology and Metallurgy at University of Belgrade who helped me to finish my bachelor studies in the middle of pandemic, without them I wouldn't be able to continue my education. Thank you, Jelena Lađarević and professor Dušan Mijin, for inviting me in your research and for offering your valuable advice.

UiT, Norwegian College of Fishery Science

Tromsø, March 2022

Вања Верузевски

Vanja Verusevski

Hans Christopher Bernstein

Abstract

Escherichia coli is one of the most established bacterial hosts for genetic devices, partially due to the available knowledge and tools for ease of manipulation. However, there is an incentive to increase the current number of available recipients. For instance, marine bacteria are being recognized for their potential as microbial "chassis" due to their rich genetic and metabolic diversity. With this in mind, tools for simulating the behavior of genetic devices would help synthetic biology expand its reach to untraditional hosts. Hence, here in this study, the effect of recipient on the performance of a genetic switch, the "chassis" effect, was estimated across different bacteria with an aim to indicate the possibilities of marine microorganisms as hosts.

The device considered in this study was assembled from two sub-parts 1) L-arabinose-inducible P_{BAD} promoter that expresses *tetR* and *gfp* genes encoding production of the TetR repressor and GFP fluorescent protein; and 2) anhydrotetracycline (aTc)-inducible P_{Tet} promoter that controls *araC* and *mKate* expression, which codes for the AraC repressor and mKate fluorescent protein. AraC protein is a repressor of the P_{BAD} promoter, while TeR represses P_{Tet}. The dynamic behavior and stability of this device was simulated by a mathematical model based on a system of ordinary differential equations (ODEs) that predicted possible "chassis" effect and compared its strength across selected bacterial hosts.

To further our understanding of the performance of a genetic switch, a dynamic modeling framework was established, and a behavior was simulated for a set of marine bacteria and *E. coli*. This was done by building a mathematical model that included system of already parametrized non-linear ODEs which were solved using the R programming language. The parametrization of ODEs by a non-linear model resulted in the Hill (n) and activation (K) coefficient estimates. The non-linear regression was performed on a GFP fluorescence data collected from the induction study with *E. coli*. This assay estimated GFP-, GFP/OD_{600} signals and GFP rates from the cells induced with L-arabinose.

The simulated dynamic response was quantified by a response time ($T_{1/2}$), a limiting factor for designing efficient gene circuits. The simulation estimated the fastest response of *Vibrio natriegens* and the slowest of *Pseudomonas oceani*. This outcome has indicated high potentials of *V. natriegens* for future applications in the synthetic biology. The "chassis" effect predicted by the model was estimated as a direct consequence of the specific growth rate (μ).

Keywords - synthetic biology, marine bacteria, chassis effect, genetic toggle switch, model

Table of Contents

List of Figures

List of Tables

1 Introduction

The idea of engineering microorganisms to perform predictable and programable functions has been present among biologist for more than a half century (Andrianantoandro et al., 2006). During 1990s when the genomic revolution happened, a new scientific arena was discovered, that is now called synthetic biology (Cameron et al., 2014).

This approach connects biologist and engineers with same ambition, to build unique components, pathways, and networks, that will be further used to design, build, and test biological machines (Cameron et al., 2014). Current and future applications for (re-)engineered microorganisms are immense, from drug discovery in pharmaceutical industry to food biotechnology and renewable fuel synthesis (Cameron et al., 2014). Additionally, synthetic biology is already having a major impact on modern medicine such as mRNA vaccines and cancer treatment (Ruder et al., 2011) and is expected to deliver novel therapies or cheaper commercial drugs in future (Khalil, A.S., & Collins, J.J., 2010).

A cellular host that is used as a recipient of engineered biological systems is called a “chassis”. The main role of a chassis is to propagate genetic information, express genes and perform programed biological functions (Kim et al., 2016). The most commonly used organism for these purposes is *Escherichia coli,* however, huge potential lays in marine bacteria which are considered to have broad metabolic diversity that can result in numerous biotechnological applications.

1.1 Main objectives

The aim of this work was to develop a framework that has ability to simulate and compare performance of an engineered genetic device expressed from a diverse set of marine bacteria and *E. coli*. This was done by building a mathematical model that accounts for the effect of species-specific growth rate within a system of non-linear ordinary differential equations. Seven bacterial strains were chosen for this purpose as displayed in **Table 1**.

Table 1 List of bacterial species used in this study

Species	**Reference**
Escherichia coli DH5α	Anton, B. P., & Raleigh, E. A., 2016
Pseudomonas aestusnigri VGXO14	Sanchez et al., 2014; Gomila et al., 2017a
Pseudomonas deceptionensis M1	Carrión et al., 2011; Carrión et al., 2015
Pseudomonas oceani KX20	García-Valdés et al., 2018; Wang, M.-q., & Sun, L., 2016
Pseudomonas pachastrellae KMM330	Gomila et al., 2017b
Pseudomonas taeanensis MS-3	Lee et al., 2010
Vibrio natriegens ATCC 14048	Maida et al., 2013

An additional goal was to perform a species-specific parameterization of associated Hill functions from gene induction curve data to comparatively quantify differences in device performance between selected strains.

2 Background

2.1 Synthetic biology - history and development

Synthetic biology has started to develop as unique scientific field with work of Jacob and Monod in 1960s when cellular regulation by molecular networks was discovered (Jacob et al., 1960). This event was followed with appearance of advanced molecular cloning techniques in later years. A period between 1980 and 1990 was marked with the rise of 'omics' sciences. During 1990s automated DNA sequencing become widespread which has led to complete genome sequencing of *Saccharomyces cerevisiae* and *Escherichia coli*. In January 2000, the first synthetic circuits were reported: the toggle switch and the repressilator, where green fluorescent protein (GFP) expression was used as an output for monitoring behavior of each circuit. This discovery has led to introduction of autoregulatory negative feedback circuit (Cameron et al., 2014).

In year 2004 at the Massachusetts Institute of Technology (MIT) in USA, Synthetic Biology 1.0 (SB1.0) conference was held. This event brought together scientists from biology, chemistry, physics, engineering, and computer sciences. One of the goals of this conference was to obtain whole-genome engineering of biological systems. The Registry of Standard Biological parts (RSBP) was developed as a base that will serve as digital catalogue of genetical parts in 'BioBrick' standardized format. This format was later only used for assembly to International Genetically Engineered Machine (iGEM). RSBP was later translated to the computation language called Synthetic Biology Open Language (SBOL) which has standardized way of describing synthetic parts and circuit designs. The main purpose of SBOL was to decrease level of ambiguity among scientists in this area. As a forum for sharing protocols and host laboratory websites, OpenWetWare was developed (Cameron et al., 2014).

During 2005 bacterial photography was possible in *E. coli* due to light-sensing circuit that was engineered. Progress was made in programming of ligand-controlled transcript regulation by RNA. Another conference, SB2.0 was held in 2006 at University of California, Berkeley and in years after this field spread worldwide which resulted in SB3.0 in Zurich, Switzerland and SB4.0 in Hong Kong, China. Some of the major discoveries in last decade were biofuel production using acid metabolism in *E. coli*, programmable microbial kill switch, creation of a bacterial cell with a synthetic genome, description of multiple input logic cascade, dynamic

control of metabolic flux for biodiesel production, production of artemisinin using engineered yeast strain on commercial scale and many others (Cameron et al., 2014).

In recent period, the cell-free systems (CFS) became key platforms in synthetic biology (Tinafar et al., 2019; Cho, E., & Lu, Y., 2020; Fábrega et al., 2021). Traditionally, synthetic biology was oriented to technologies that were using the whole cell. These included different biosensors that have ability to detect broad spectrum of analytes. CFS contain enzymes that are active during transcription and translation, because of that, it is possible to maintain process of the cell dogma that is autonomous compared to a cell. They can be from different origins, such as eukaryotic or prokaryotic cells (Tinafar et al., 2019).

CFS can include cell lysates, energy sources, purified proteins, amino acids, RNA, DNA which can be used for different purposes such as building of portable diagnostic devices, biomolecular manufacturing and can enhance discovery of unknown enzymes. Synthetic biology has had a major impact to a development of sensors, manufacturing of therapeutics, production of membrane proteins, macromolecular production, modification of proteins and codon tables, discovery of biomaterials and many more. It is believed that CFS will connect synthetic biology with electronics, computation, and machine learning (Tinafar et al., 2019).

2.2 Biopart Assembly Standard for Idempotent Cloning (BASIC)

One of the most important tasks in synthetic biology is to assemble DNA constructs quickly and reliably. Hence, the newly developed technologies are highly focused on obtaining a standardizable and scalable approach (Storch et al., 2015). BASIC DNA assembly represents orthogonal linker-based DNA assembly, and it is a standard that describes the design framework for linkers and a universal format for part storage. The assembly always includes formation of part-linker-part-linker bonds (Storch et al., 2017).

The work of Storch et al., 2015 showed that the BASIC assembly allows efficient, parallel assembly with high accuracy. They have determined 93% and 99.7% accuracy of four-part assemblies with single and double antibiotic selection, respectively. Additionally, double antibiotic selection in seven-part assemblies resulted in 90% accuracy. This simple and robust method brings together six key concepts: standard reusable parts, single-tier format, idempotent cloning, parallel DNA assembly, size independence, automatability (Storch et al., 2015).

The principles of the BASIC physical DNA standard lay in the integrated prefix and suffix sequences (*i*P and *i*S) which are used to ensured compatibility with the creation of fusion proteins by optimizing the amino acid coding of both the short BASIC scars and the full *i*P/*i*S sequences (Storch et al., 2015).

The integrated prefix *i*P is a 5'-TCTGGTGGGTCTCTGTCC-3' sequence and *i*S represents 5'-GGCTCGGGAGACCTATCG-3' sequence. Both, *i*P and *i*S, contain BsaI recognition site. The main role of this site is to release the parts from storage vector resulting in different four base overhangs that are present at each end. Purpose of overhangs is to ligate half linkers that are prefix- and suffix-specific (Storch et al., 2017).

To release the parts form from a storage vector, it is necessary to have two inward facing BsaI recognition sites which leaves a 4 bp scar on the prefix end and 6 bp scar on suffix. End-specific ligation is accomplished because of different digestion of 4 bp overhangs at the prefix and suffix. Ligation of BsaI and partially double-stranded oligonucleotide DNA linkers is performed at a same time. The nonligated linkers are removed in a purification step. The linker-adapted parts are then placed in an ionic buffer on higher temperature to achieve the final assembly (Storch et al., 2015).

2.3 The Genetic Toggle Switch

A genetic toggle switch is a device that can achieve bistability (Gardner et al., 2000). It can be constructed from two repressible promoters that are arranged in mutually inhibitory network. A change between two stable states is induced by presence of a specific chemical or physical signal (such as temperature). In practical manner, this device acts like a synthetic, cellular memory unit that can be used in various applications such as biotechnology, biocomputing and gene therapy (Gardner et al., 2000).

If a production of some protein is not regulated and it is produced at a constant rate under all conditions. This is known as constitutive expression, whereas a promoter that controls this expression is called constitutive promoter (Alon, U., 2019). On the other hand, inducible promoters are controlled by change of intracellular concentration of an inducer or a repressor (Alon, U., 2019).

Some of frequently used inducible promotors are lactose (lac) promoter (P_{Lac}) and tetracycline (tet) promoter (P_{Tet}). A transcription from the P_{Lac} promoter is repressed by a lac repressor

(LacI) and tet repressor (TetR) acts on the P_{Tet} promoter. The repressors bind to a promoter region and inhibit transcription. These repressors can be bound to specific inducers, LacI binds to isopropyl-β-D-1-thiogalactopyranoside (IPTG) and TetR binds to anhydrotetracycline (aTc) (Alon, U., 2019; Lutz, R., & Bujard, H., 1997).

Another widely used inducible promoter is a L-arabinose-inducible araBAD promoter (P_{BAD}). In a presence of L-arabinose (inducer) transcription from the P_{BAD} is turned on, while in its absence, transcription is present at a very low level. This promoter is repressed by AraC. The P_{BAD} promoter has showed very fast induction rate and tight control of expression. The activity of promoter is under effect of several factors such as concentration of inducer used for induction, ability of bacterial strain to degrade arabinose and the physiological characteristics of the culture which are under effect of growth media and available carbon source in it (Guzman et al., 1995; Lutz, R., & Bujard, H., 1997).

2.3.1 Design of genetic toggle switch circuits

The toggle switch circuits can be schematically represented with SBOL (The Synthetic Biology Open Language). The SBOL data standard is a data exchange representation for synthetic biology designs. This standard ensures collaborative engineering of unique biological systems (Galdzicki et al., 2014). It is a proposed standard for exchanging designs among scientists with interests in synthetic biology. This format promotes exchange of synthetic biology data between software tools, research groups and commercial service providers. (Roehner et al., 2016).

The design of simple toggle switch can be schematically described by two repressors and two inducible promoters. One promoter is inhibited by the repressor that is regulated by the other promoter, and vice versa. Hence, each promoter mutually inhibits the expression of the other. With this design, it is possible to obtain two stable states, a concept known as bistability (Gardner et al., 2000).

These two distinct states can be produced by high expression of first promoter which leads to greater concentration of second repressor which represses second promoter, or by high expression of second promoter which leads to greater concentration of repressor for first promoter which represses it (Gardner et al., 2000). Another equilibrium that can be obtained is the state in which both of genes are expressed at low levels. This state is unstable and even if cells can potentially be in this state, any perturbation would push cells to one of two stabile states (Lugagne et al., 2017).

A scheme of a toggle switch that was considered in this study and its expected behavior depending on inducer presence is given in **Figure 1**. In the presence of L-arabinose, P_{BAD} promoter is active (Guzman et al., 1995) and fluorescence of GFP reporter protein can be measured, on the other side, presence of anhydrotetracycline (aTc) initiates activity of P_{Tet} promoter (Lutz, R., & Bujard, H., 1997) which results in mKate signal. In a case that none of the inducers are present, none of the promoters show activity, neither *tetR* and *araC* genes are transcribed (Fontanarrosa et al., 2020) nor the response signals (GFP, mKate) are observable. If both of inducers are present, the behavior of presented toggle switch is unknown.

Inducer	**Promoter activity**	**Result signal**
L-arabinose	P_{BAD}	GFP
Anhydrotetracycline (aTc)	P_{Tet}	mKate
No inducer	No activity	No signal
L-arabinose + aTc	?	?

Possible outcomes of toggle switch response with different inducers

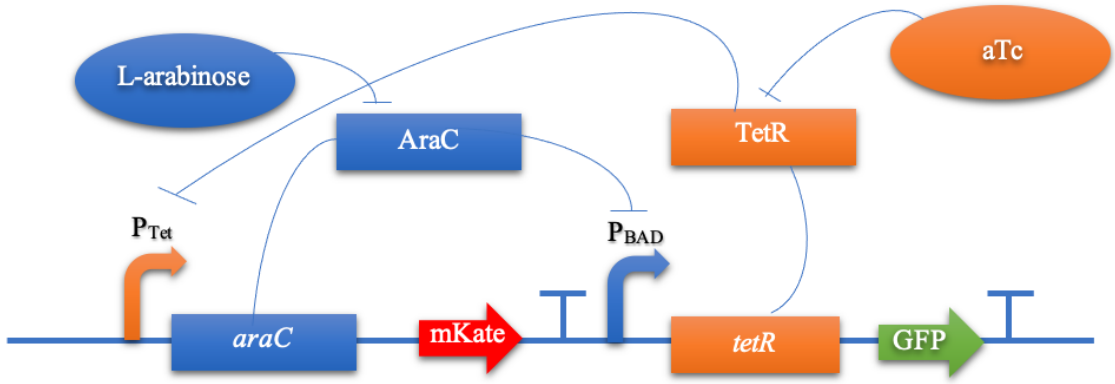


Figure 1 Schematic drawing of the genetic toggle switch *that consists of two inducible promoters (P_{Tet}, P_{BAD}), two fluorescent proteins as signals of promoter activity (mKate-RFP and GFP), two genes that are transcribed (araC and tetR), two inducers (L-arabinose and anhydrotetracycline (aTC)) and two repressors (AraC and TetR). Possible states of genetic device caused by different inducer presence are represented in the table above the drawing. The performance of a presented toggle switch was simulated in this study.*

2.4 The Hill function

Transcription networks consist of two elements: genes and transcription factors (Alon, U., 2019). According to central dogma, mRNA is translated to proteins, therefore the transcriptional factors affect the rate of protein production that has been encoded by the genes. Transcriptional factors (TF) are proteins that can activate or repress genes. Signals from the environment can activate specific TFs by altering their ability to bind with DNA or other DNA-binding proteins. Hence, extracellular molecular signals can be used by synthetic biologists to control transcription rates of target genes and consequentially the rate of mRNA and protein production (Alon, U., 2019).

If a TF increases the transcription rate, it is called activator, if opposite, in case that decreases the rate of transcription, it is called repressor. The production of all transcriptional factors, since they are proteins, is regulated by other TFs. This complex interaction between different TFs builds transcription network which describes all the regulatory transcription interactions within a cell (Alon, U., 2019).

In the case that a TF increases the transcription rate after binding to a specific promoter that is called activation or positive control. In the opposite case, where a TF decreases the rate of transcription, the term that is used is repression or negative control. An input function describes the strength of the effect of TF. For a case of production of protein Y that is controlled by single TF X, transcription network can be described as $X \rightarrow Y$. If rate of production of protein Y is equal to a concentration of X in active form (X^*), then: $rate\ of\ production\ of\ Y = f(X^*)$ (Alon, U., 2019).

The Hill function is an empirical expression used to describe an input function and for an activator it is a hyperbolic curve that has rising trend from zero to a saturation point (**Figure 2**) and it is formulated in Equation 1.

$$f(X^*) = \frac{\beta X^{*n}}{K^n + X^{*n}} \qquad \text{Equation 1}$$

Parameters for the Hill function are K, β and n. The activation coefficient (K) represents the concentration of X^* that is necessary to significantly activate expression and it is in concentration units. The maximal expression level of the promoter is represented by the β

parameter. The Hill coefficient (n) gives information about steepness of input function and usually it takes values between 1 and 4 (Alon, U., 2019).

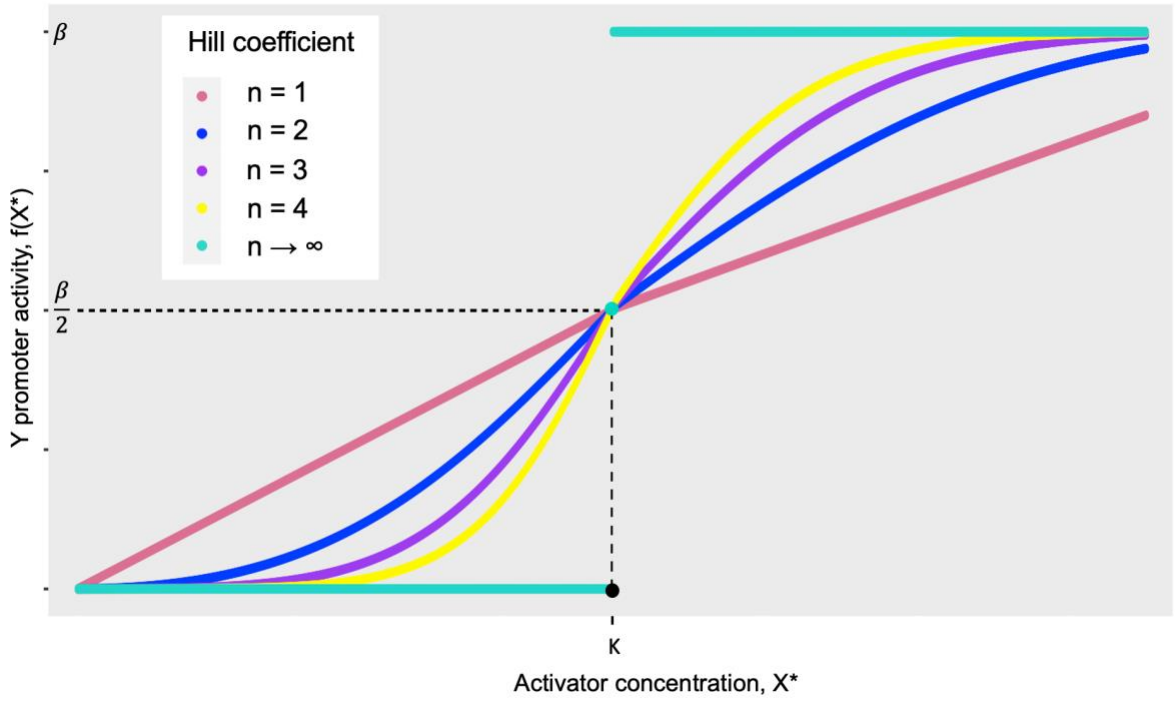


Figure 2 The Hill input function for an activator *is a curve that rises from zero and approaches a maximal saturated level. A schematic representation of the input Hill functions shows increasing monotonic trend, and the larger n, the more step-like the input function. The Hill coefficient change from 1 to 4 is represented with moderately steep input functions and for n→∞ the Hill function becomes a step function (Alon, U., 2019).*

The saturation of the Hill function at high activator concentration is achieved because the probability that the activator binds the promoter cannot exceed 1, regardless the concentration of X* (Alon, U., 2019).

The Hill function for a repressor is also a hyperbolic curve, but with decreasing trend as presented at **Figure 3**. The same parameters K, β, n act in equation for repressor (Equation 2) and final shape of the input function depends on them. In both cases, a half of the maximal expression level $\left(\frac{\beta}{2}\right)$ is accomplished when $X^* = K$ (Alon, U., 2019).

$$f(X^*) = \frac{\beta}{1 + \left(\frac{X^*}{K}\right)^n} \qquad \text{Equation 2}$$

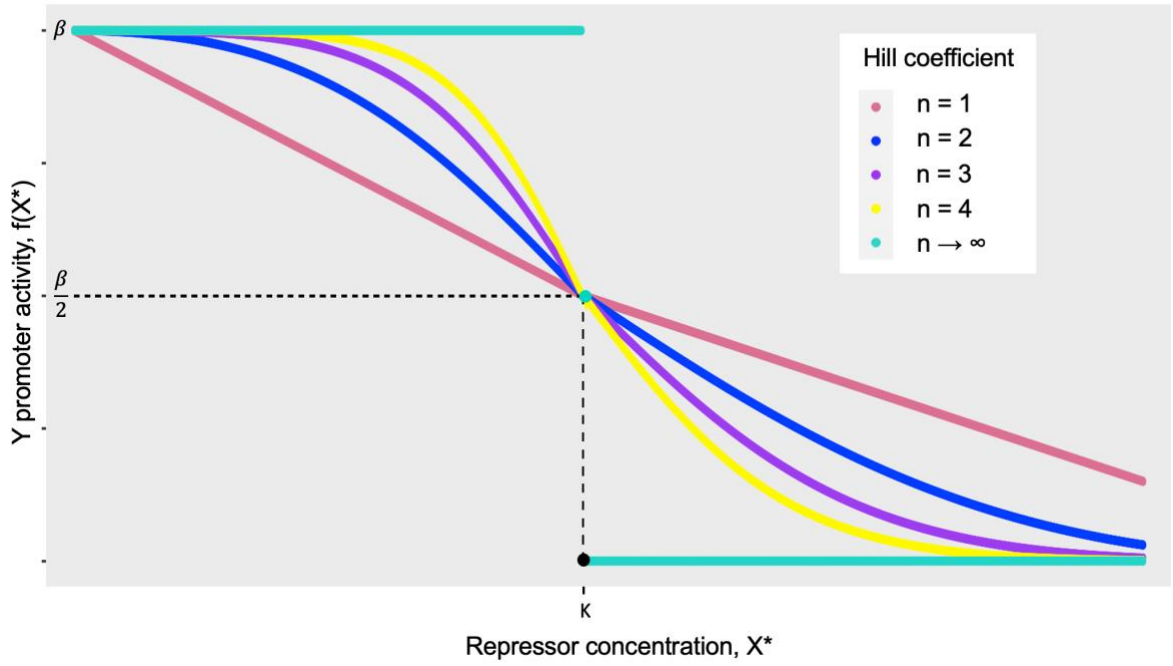


Figure 3 The Hill input function for a repressor *is a curve with decreasing trend that starts at the maximal promoter activity, β, when the repressor does not bind the promoter at all (X* = 0). Schematic representation of the input Hill functions shows decreasing monotonic trend, and the larger n, the more step-like the input function. The Hill coefficient change from 1 to 4 is represented with moderately steep input functions and for n→∞ the Hill function becomes the step function (Alon, U., 2019).*

2.5 Mathematical models in synthetic biology

Synthetic biology is in essence an engineering discipline that uses mathematical tools to program microbes to perform new functions while understanding cell mechanisms and processes. Mathematical modeling often aids in detailed design of biological circuits and in simulations that predict behavior. These simulations are often used with experiments that produce quantitative data, and further used to inspect the quality of model and its prediction. As a result of integration and quality cooperation between experimental and modeling data, it can be possible to use engineered microbes as a technological platform (Chandran et al., 2008).

The unique role of modeling has been seen a crucial link between design and realization in engineering. The model should predict the dynamics of a system in various conditions which can lead to careful selection of optimal conditions for different processes that engineer the desired behavior of microbes. The undesired conditions can be easily avoided by searching for them *in silico* instead of running numerous experiments. Mathematical models should represent a bridge that connect conceptual design ideas with downstream biological realization (Zheng, Y., & Sriram, G., 2010).

On the other hand, modeling of biological systems can be challenging even though the structures and functions of many biomolecules are known, the vast majority of molecular and cellular mechanisms are not. The complexity of biological systems is present on various scales, such as a highly complex interaction among metabolites, metabolic fluxes, proteins, RNA, and genes network which can result in obtaining feedback or feedforward loops that respond in different time scales. There is also a high sensitivity of biological systems to external, environmental factors. These factors affect the predicted behavior and lead to uncertain response of biological systems. Unlike the mechanical and electrical systems, highly precise prediction of the output of a biological system is difficult to obtain but it can be simplified to offer the insights toward synthetic circuit construction (Zheng, Y., & Sriram, G., 2010).

Models can be simplified by making different types of assumptions such as homogeneity assumption within the cell and a cell population. These systems can be modeled by ordinary differential equations (ODEs) while systems that include compartmentation, spatial segregation or intracellular gradients are commonly modeled by use of partial differential equations (PDEs). Other assumptions that are widely used in enzyme kinetics of transcriptional regulations are equilibrium, steady state, and quasi-steady state assumption. These tools are widely used for removing the time-dependency in the model, which transform ODE to simple algebraic equations that can be easily solved (Zheng, Y., & Sriram, G., 2010).

Among many models that are present now, a scientist needs to choose the one best suited for their purposes. Since model selection differs from hypothesis testing, it is important to understand that model cannot be completely true in a sense of being a completely precise and accurate description of the selected process (Kirk et al., 2013).

Two categories of mathematical models of biological systems can be described as deterministic and stochastic. A deterministic model includes a real system with analytical equations, typically ODEs and/or PDEs, which have numerical parameters and usually they are mass balances on cellular species and the model is reproducible. The stochastic models can sometimes better represent real systems that include random events. These usually examine fluctuations and noise effect on system dynamics. A deterministic model consists of interactions and parameter values that attempt to predict identical and rigid system dynamics for the same set of parameters and initial conditions. Biological systems are typically influenced by both stochastic and deterministic events and often become unpredictable with unexpected and irreproducible fluctuations (Zheng, Y., & Sriram, G., 2010).

2.6 Non-linear regression

Non-linear regression is used to relate a response (Y) to a vector of predictor variables ($x = (x_{1,\dots,}x_k)^T$). The starting assumption is that the prediction equation depends non-linearly on one or more unknown parameters. In the most cases, non-linear models rely on predictions that are between responses and their independent predictor variables that can be described as a function (Smyth, G. K., 2002). The non-linear regression models can be written in general form as:

$$f_i(y_t, x_t, \alpha_i) = u_{it}$$
$$i = 1,2, \dots, n;$$
$$t = 1,2, \dots, T$$

Equation 3

where y_t, x_t and α_i are vectors of endogenous variables, exogenous variables, and parameters; u_{it} represents unobservable error terms with zero mean (Amemiya, T., 1983). Thus, a simpler form of Equation 3 can be given as:

$$Y_i = f(x_i, \theta) + \varepsilon_i, \qquad i = 1, \dots, n$$

Equation 4

where Y_i represents responses, f is a known function of the covariate vector $x_i = (x_{i1}, \dots, x_{ik})^T$ and the parameter vector $\theta = (\theta_1, \dots, \theta_p)^T$. ε_i are random errors and usually they are assumed to be uncorrelated with mean zero and constant variance (Smyth, G. K., 2002).

In the case when both predictor and response values are known and the function f depends on some unknown parameter, the parameter estimate can be obtained by providing the best fit of the function f to observations for response variable. The best fit is provided by minimalization of the residual sums of squares (RSS) with respect to β according to Equation 5 (Smyth, G. K., 2002).

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - f(x_i, \beta))^2$$

Equation 5

The process of minimization of RSS includes minimizing the least-squares criterion or least-squares estimation. While minimizing RSS as a solution is obtained the least-squares parameter estimates, which is marked as $\hat{\beta}$. The estimates $\hat{\beta}$ are β values applicable for function f, that make the smallest possible RSS, and this can be formulated that minimum of RSS(β) is calculated when $\beta = \hat{\beta}$. The mostly used algorithm for obtaining the nonlinear regression estimates is GaussNewton method (Smyth, G. K., 2002).

The non-linear regression model fitting function that is used in R is *nls*() and it is a part of the package 'stats' (Ritz, C., & Streibig, J. C., 2008). As a parameter of fit quality, this function calculates the residual standard error (RSE) that is estimated as the square root of the residual sum of squares (RSS) divided by the residual degrees of freedom (Ritz, C., & Streibig, J. C., 2008).

2.7 Differential equations

Differential equations (DEs) are present in various forms in divergent areas of science because of their ability to describe natural phenomena. It is essential to solve these equations in an accurate and precise way to obtain an appropriate representation of examined behavior. Unique for DEs is that numerous DEs cannot have analytical solution, and they are solved numerically instead (Soetaert et al., 2012).

The simplest DEs are first order differential equations which can be represented with equation:

$$y' = \frac{dy}{dx} = f(x, y) \qquad \text{Equation 6}$$

where f represents a given function of x and y. In this case, independent variable is represented with symbol x and $y = y(x)$ is dependent variable. Presented type of equation is named a first order differential equation, since it does not contain higher derivatives than the first. This equation is also an ordinary differential equation as dependent variable, y, depends only on one independent variable, x (Soetaert et al., 2012).

A solution of Equation 6 is a differentiable function $y(x)$ for every x:

$$y'(x) = f(x, y(x)).$$ Equation 7

If it is assumed that $y(x_0)$ is known, the solution of Equation 7 in a defined interval $[x_0, x_1]$ is obtained by integrating both sides of Equation 7 with respect to x, which results in:

$$y(x) - y(x_0) = \int_{x_0}^{x} f(t, y(t))dt,\ x \in [x_0, x_1].$$ Equation 8

If the integral presented on the right side of Equation 8 can be solved to give an equation for y, then an analytic solution can be determined. In every analytic solution, the free parameter, *c*, occurs and the solution is uniquely defined with initial conditions and regulatory conditions. On the other hand, many ODEs cannot be solved analytically, and it is necessary to estimate the solution by solving them numerically instead (Soetaert et al., 2012).

ODEs can contain higher derivatives; therefore, for the second order it can be written:

$$y''(x) = f(x, y, y').$$ Equation 9

For obtaining a solution of Equation 9, it is necessary to transform it to a system of first order equations. Conversion is done by defining an extra dependent variable such as the one that is equal to the first derivative of *y*, as displayed in two following equations (Equation 10).

$$\begin{aligned} y' &= y_1 \\ {y_1}' &= f(x, y, y_1) \end{aligned}$$ Equation 10

In a special case where the first derivative is absent from Equation 9 it is advised to derive special methods for the solution of Equation 11:

$$y''(x) = f(x, y),$$ Equation 11

instead of introducing the first derivative of *y* (*y'*) in equation (Soetaert et al., 2012).

There is wide variety of numerical methods for solving ODEs. The simplest form of these methods usually includes subdivision of the domain of the independent variable, *x*, which results in several discrete points. These points have different purposes and in finite difference methods they serve to approximate values of the dependent variable, *y*, and to estimate the derivatives of *y* with respect to *x* only at these discrete points. The series of integration steps can be written as $x_0, x_1, \ldots, x_n$ and numerical method creates a sequence of values $y_0, y_1, \ldots, y_n$ that follows statement written in Equation 12 (Soetaert et al., 2012).

$$y_n \approx y(x_n), \qquad n \geq 0 \qquad \text{Equation 12}$$

where $x_n = x_0 + nh$ represents point where the estimated solution will be computed.

Moreover, the analytic solution at x_n is marked as $y(x_n)$ and as y_n is marked numerical solution that is approximated at x_n. The methods such as Euler and Implicit method are widely used numerical methods for solution of ODEs. The accuracy and convergence of numerical methods are always in question, and they are inspected with calculating the local truncation error, LTE, which is supposed to take small value if the method is accurate. Moreover, LTE can be decreased by lowering the size of the step-length of integration. Another concept that should be questioned while solving ODEs is stability. If small changes in function or in initial conditions induce large effects, the system is unstable or ill-conditioned. On the other hand, system is stable or well-conditioned only if small changes in the data initiate small changes in solution (Soetaert et al., 2012).

Partial differential equations (PDEs) represent the DEs where dependent variable is a function of more than one independent variable. A second order PDE with two independent variables x_1, x_2 can be presented with Equation 13 (Soetaert et al., 2012).

$$F(x_1, x_2, y, \frac{\partial y}{\partial x_1}, \frac{\partial y}{\partial x_2}, \frac{\partial^2 y}{\partial x_1^2}, \frac{\partial^2 y}{\partial x_2^2}, \frac{\partial^2 y}{\partial x_1 \partial x_2}) = 0 \qquad \text{Equation 13}$$

Differential algebraic equations (DAEs) occur as a specific mix of differential and algebraic equations. DAEs can be written as Equation 14.

$$F(x, y, y') = 0 \qquad \text{Equation 14}$$

Delay differential equations (DDEs) are unique type of DEs which include past values of the dependent variables and/or their derivatives. Specific characteristic of these equations is that solving these types of problems requires to acknowledge initial history which includes sequence of values instead of initial value. DDEs can be formulated as presented in Equation 15 (Soetaert et al., 2012).

$$y'(t) = f\big(t, y(t), y(t-\tau_1), y(t-\tau_2), \dots, y(t-\tau_n)\big)$$
$$y(t) = \Phi(t) \text{ for } t < t_0 \qquad \text{Equation 15}$$

Stability of systems of ODE can be determined by Liapunov and Jacobi stability analysis which are based on different concepts and in some cases, there might be points that Liapunov criteria is stable but Jacobi unstable or vice versa (Abolghasem, H., 2012).

2.7.1 Numerical methods for solving ODEs and systems of ODEs

The expected solution of differential equation is a function, not a number. In the majority of cases, it can be hard or even impossible to determine analytical solutions of differential equations, hence the numerical methods are based on simple ideas. These methods include initial condition $x(a) = x_0$ and series of approximations to the solution at a sequence points that are marked as $t_1, t_2, t_3, \dots, t_n$ which belong to interval [a,b], as a result $a = t_0 < t_1 < t_2 < t_3 \dots < t_n = b$. Because of its simplicity, the Euler's method has been widely used for solving the DEs. This method includes Geometric interpretation of differential equation: "The differential equation $x' = f(t, x)$ describes a family of functions whose tangent at the point (t, x) has slope $f(t, x)$. By adding initial condition $x(a) = x_0$, a particular solution, or solution curve, is selected from family of solution" (Mørken, K., 2017).

It is assumed that differential equation is:

$$x' = f(t, x), \qquad x(a) = x_0, \qquad \text{Equation 16}$$

and main intention is to obtain a sequence of estimates $(t_k, x_k)_{k=0}^{n}$ to the solution which follow that $t_k = a + kh$. The point of true solution is obtained by initial condition which makes (t_0, x_o) the starting point of the approximation. Solution at t_1 is derived by computing the slope

of the tangent at (t_0, x_o) which is $x_0' = f(t_0, x_o)$ and the equation of the tangent is $T_0(t) = x_0 + (t - t_0)x_0'$ (Mørken, K., 2017).

Consequently, the approximation of x_1 at t_1 is:

$$x_1 = T_0(t_1) = x_0 + hx_0' = x_0 + hf(t_0, x_0). \qquad \text{Equation 17}$$

Presented equation gives the next approximate solution point (t_1, x_1) and the same process is repeated. Generally, in Euler's method, an estimated solution (t_k, x_k) is advanced to (t_{k+1}, x_{k+1}) by following the tangent:

$$T_k(t) = x_k + (t - t_k)x_k' = x_k + (t - t_k)f(t_k, x_k) \qquad \text{Equation 18}$$

at (t_k, x_k) from t_k to $t_{k+1} = t_k + h$. Which is resulting in the approximation:

$$x_{k+1} = x_k + hf(t_k, x_k) \qquad \text{Equation 19}$$

to $x(t_{k+1})$ (Mørken, K., 2017).

Another family of numerical methods belongs to Taylor methods and the simplest case is quadratic case, but the general principle of these methods is similar. Formulation of Quadratic Taylor method: "The quadratic Taylor method advances the solution from a point (t_k, x_k) to a point (t_{k+1}, x_{k+1}) by evaluating the approximate Taylor polynomial:

$$x(t) \approx x_k + (t - t_k)x_k' + \frac{(t - t_k)^2}{2}x_k'' \qquad \text{Equation 20}$$

at $x = t_{k+1}$. The new value x_{k+1} is given by:

$$x_{k+1} = x_k + hx_k' \frac{h^2}{2}x_k'' \qquad \text{Equation 21}$$

where the values x_k, x_k' and x_k'' are obtained as derivatives and $h = t_{k+1} - t_k$" (Mørken, K., 2017). Euler's midpoint method is defined as the method where solution is achieved from (t_k, x_k) to (t_{k+1}, x_{k+1}) in two steps:

1. approximation to the solution is estimated at the midpoint $t_k + h/2$ by using Euler's method with step length $h/2$:

$$x_{k+1/2} = x_k + \frac{h}{2} f(t_k, x_k) \qquad \text{Equation 22}$$

2. solution is advanced to t_{k+1} by following the straight line from (t_k, x_k) with slope given by $f(t_k + h/2, x_{k+1/2})$ as presented in Equation 23 (Mørken, K., 2017).

$$x_{k+1} = x_k + hf(t_k + h/2, x_{k+1/2}). \qquad \text{Equation 23}$$

The Runge-Kutta methods are derived as a generalization of midpoint Euler's method. These methods include several evaluations of *f* between each step to obtain higher accuracy. The second order Runge-Kutta methods are defined as: "The numerical method which advances from (t_k, x_k) to (t_{k+1}, x_{k+1}) for the differential equation $x' = f(t, x)$ with initial condition $x(a) = x_0$, according to the formula:

$$x_{k+1} = x_k + h\left((1-\lambda) f(t_k, x_k) + \lambda f\left(t_k + \frac{h}{2\lambda}, x_k + \frac{hf(t_k, x_k)}{a\lambda}\right)\right) \qquad \text{Equation 24}$$

is 2nd order accurate for any nonzero value of the parameter λ, provided *f* and its derivatives up to order two are continuous and bounded for $t \in [a, b]$ and $x \in \mathbb{R}$" (Mørken, K., 2017).

2.7.2 Solving differential equations in R with the 'deSolve' package

R is a programming language and a software environment which can provide numerous statistical, computational, and graphical functions. The package 'deSolve' in R is widely used as a general solver for Initial Value Problems of Ordinary Differential Equations (ODE), Partial Differential Equations (PDE), Differential Algebraic Equations (DAE) and Delay Differential Equations (DDE) (Soetaert, K. & Petzoldt, T., 2011; Soetaert et al. 2010; 2012; 2015).

This package contains different functions that solve initial value problems of a system of first-order ODE, PDE, DAE and DDE. These functions have an interface to the FORTRAN functions Isoda, Isodar, Isode, Isodes of the ODEPACK (Hindmarsh, A. C., 1983) collection, to the FORTRAN functions dvode, zvode and daspk and a C-implementation of solvers of Runge-Kutta family with fixed or variable time steps. Additionally, package is designed to solve ODEs resulting from 1-D, 2-D, and 3-D PDEs that have been concerted to ODEs by numerical differencing (Soetaert, K. & Petzoldt, T., 2011).

A numerical solution of a system of differential equations is often produced under certain assumptions and approximations that can lead to numerical errors. There are several sources of these errors such as a time step, accuracy order of the solver, arithmetic of floating point, properties of the differential system itself and the stability of the solution algorithm. The R solution output is an array that contains the time value in first column, then the values of all state variables and if any the ordinary output variables (Soetaert et al., 2010).

2.8 Dynamics of inducible expression within a toggle switch

The chapter 2.3 has been dedicated to describing the design and functioning of a toggle switch. Scheme of the toggle switch design is presented at **Figure 1**. Dynamics of this system can be described by a system of two ordinary differential equations, as it is displayed in Equation 25 (Alon, U., 2019).

$$\frac{dX}{dt} = f(Y) - \alpha X$$
$$\frac{dY}{dt} = f(X) - \alpha Y$$

Equation 25

Term $\frac{dX}{dt}$ and $\frac{dY}{dt}$ are rate of change of protein *X* and *Y* concentration respectively. The $f(Y)$ and $f(X)$ are production rates and α is removal rate (the sink term). The system consists of two repressors and the production rates ($f(Y), f(X)$) can be expressed as the Hill input functions for repressors (Alon, U., 2019).

The rate of degradation (the removal rate) is a sum effect of degradation and dilution, α_{deg} and α_{dil}. The dilution term is equal to specific microbial growth rate (μ) and Equation 26 gives total degradation/dilution rate in units of $\frac{1}{time}$ (Alon, U., 2019).

$$\alpha = \alpha_{deg} + \alpha_{dil} = \alpha_{deg} + \mu$$

Equation 26

Many proteins are not actively degraded in growing cells ($\alpha_{deg} = 0$), meaning that the production of these proteins is balanced by dilution due to the increasing volume of the growing cell, $\alpha = \alpha_{dil}$ (Alon, U., 2019).

By including the Hill function and under assumption that α_{deg} is negligible, $\alpha = \mu$, the presented system (Equation 25) can be modified to a system of two ordinary differential equations and one non-differential equation that includes Monad kinetics and describes a change of specific growth rate (μ) with a single growth substrate (*S*). The system of ODE consists of following equations (Alon, U., 2019).

$$\frac{dX}{dt} = \frac{\beta}{1 + \left(\frac{Y}{K}\right)^n} - \mu X$$
$$\frac{dY}{dt} = \frac{\beta}{1 + \left(\frac{X}{K}\right)^n} - \mu Y$$

Equation 27

X and Y represent concentration of protein and parameters are β, K, n and μ. Activation coefficient (K) is concentration of X and/or Y that is necessary to significantly activate expression. Maximal expression level of the promoter is represented by β parameter and n is the Hill coefficient (Alon, U., 2019).

A non-differential equation added to the system presented in Equation 27 estimates the specific growth rate (μ) as formulated in the Equation 28 that considers the effect of a limiting substrate (S) to the specific growth rate and follows Monod kinetics (Liu, S., 2017).

$$\mu = \frac{\mu_{max} S}{k_s + S}$$

Equation 28

Maximal specific growth rate is μ_{max} and it is in units $\frac{1}{time}$, while k_s represents a saturation constant and it is in concentration units (Liu, S., 2017).

The presented system (Equation 27) has three fixed points when $n > 1$, where two of these points are stable and the whole system can be called as bistable. Each of the points present opposite gene expression states, one that results in higher production of protein X and another that is correlated to a high production of protein Y. Unstable state is the one where production of X and Y have similar values (Crawford, J. D., 1991; Strogatz, S. H., 2018; Gardner et al., 2000).

2.9 Bacterial strains

2.9.1 *Escherichia coli*

E. coli is a Gram-negative bacterium from the family Enterobacteriaceae, and it is one of the most prevalent commensal inhabitants in the human's gastrointestinal tracts as well as the warm-blooded animals. *E. coli* triggers broad spectrum of diseases in humans and it is known as a common pathogen. Pathogenic types of *E. coli* can be divided in two groups: Enteric *E. coli* and Extraintestinal *E. coli* (ExPEC) (Allocati et al., 2013).

E. coli is widely used in biotechnological purposes because its cells are easy to culture and to genetically manipulate, it has a short life cycle, well-known genetics, easy and fast transformation with endogenous DNA, fast growth kinetics and grows in inexpensive growth media. Common challenges with use of *E. coli* cells as a host organism are difficult or very low expression of a foreign gene and solubility of recombinant proteins for over expression. Expression of foreign gene in *E. coli* are under effect of mRNA's secondary structure, ability of protein to fold, its solubility, preferential codon use, toxicity of protein and need for post-translational modifications. Strategies that can enhance expression and solubility of overexpressed protein include a change of vector or the host strain, changing the culture parameters of recombinant host strain or co-expression of other genes and altering the gene sequence without any change of the functional domain (Gopal, G. J., & Kumar, A., 2013).

2.9.2 *Pseudomonas genus*

The *Pseudomonas* genus was firstly described in late 19th century (Migula, 1894; 1900) and number of species that belong to this genus has grown up to more than 200 species (Wang, M.-q., & Sun, L., 2016) and it is still growing. This group of organisms is highly heterogenous (Barbour et al., 2017), very large and widely dispersed (Nishimori et al., 2000). *Pseudomonas* has a high and unique adaptability to the different environmental conditions. The optimal growth is reached in a limited temperature range and hydrogen ion concentrations. Members of this genus are not a truly thermophilic or acidophilic (Palleroni, N. J., 2010). Most of the species that belong to *Pseudomonas* were isolated from marine environment and some were isolated from marine animals such as: *P. pachastrellae* (Romanenko et al., 2005), *P. xanthomarina* (Romanenko et al., 2005), and *P. marnicola* (Romanenko et al., 2008). Directly from seawater were isolated *P. pseudoalcaligenes* (Palleroni, N. J., 1984), *P. alcaliphilia* (Yumoto et al., 2001), and *P. litoralis* (Pascual et al., 2012). Isolates of *P. balearica* (Bennasar et al., 1996) and

P. deceptionensis (Carrion et al., 2011) have been found in marine sediment. Isolates have also been obtained from seashore environments, such as *P. aestusnigri* (Sánchez et al., 2014), *P. sabulinigri* (Kim et al., 2009), and *P. taeanensis* (Lee et al., 2010). In the study of Spröer and co-workers (Lang et al., 2010) it has been determined that presence of the hydroxylated acids such as $C_{10:0}$ 3-OH and $C_{12:0}$ 3-OH is vital for *Pseudomonas* genus. Palleroni (2015) has revealed that species of *Pseudomonas* genus have 58-69 mol% of the genomic G+C content (Wang, M.-q., & Sun, L., 2016). The cells of *Pseudomonas* species have a rod shape, they are polarly flagellated, aerobic, and Gram-negative (Palleroni, N. J., 1984).

2.9.2.1 *Pseudomonas aestusnigri*

Cells of *P. aestusnigri* are Gram-negative, rod-shaped (Palleroni, N. J., 1984), their length can vary between 1.6-2.0 μm and width from 0.6-0.7 μm. *P. aestusnigri* cells are non-pigmented with catalase-positive, oxidase-positive characteristics, strictly aerobic and non-fermentative. The characteristics of colonies include round convex shape and beige bright color with white margins of 1-3 mm diameter on Luria broth (LB) plates after incubation for 48h at 30° C. Growth of this species is present at 18-42° C with optimal temperature between 25-30° C at pH 6-10 and with 2-10% (w/v) NaCl. *P. aestusnigri* is positive for assimilation of capric acid, adipic acid, utilization of Tween 40, Tween 80, pyruvic acid methyl ester, D,L-lactic acid, sebacic acid and L-alanine (Sanchez et al., 2014).

In the study of Molitor et al. (2020) of this marine bacterium has showed a polyester degrading activity. The polyester hydrolase (PE-H) that belongs to a type IIa polyethylene terephthalate (PET) hydrolases, has been found in this species and it shows activity toward PET as a substrate (Bollinger et al., 2020).

2.9.2.2 *Pseudomonas deceptionensis*

P. deceptionensis is a Gram-negative, oxidase-positive, non-spore forming, strictly aerobic microorganism with characteristic rod-shaped cells (0.8 x 1.5-2.0 μm). This species does not produce any fluorescent pigment on King´s B medium. During incubation on TSA medium at 20° C for 72 h colonies are 1.5-2.0 mm in diameter with slightly convex shape, white colored, round and mucous. Optimal conditions for growth are between -4 and 34° C with pH value from 5 to 10. Tolerance for NaCl is up to 6% (w/v). Under anaerobic conditions on MA, growth is very poor. Cells are positive for catalase, oxidase, and hydrolysis of urea, but negative for

hydrolysis of aesculin, Tween 80, gelatin, casein, starch, and lecithin. *P. deceptionensis* neither produce indole nor hydrogen sulfite and does not reduce nitrate (Carrion et al., 2011).

2.9.2.3 *Pseudomonas oceani*

*P. oceani c*ells of strain KX 20^T are Gram-negative, strictly aerobic, rod-shaped, and non-pigmented. Dimensions of the cells are between 0.4-0.6 µm wide and 2.1-2.6 µm long. The cells are motile by a single flagellum. Cell growth can occur in temperature ranges from 4-41 °C, and pH 6.0-10.0 with 0-10 % (w/v) NaCl. The whitish colonies show circular shape, and they are not pigmented, smooth and transparent. In a King B medium there is no production of fluorescent pigments. The cells are positive for oxidase and catalase activities, moreover they show positive reaction for hydrolysis of Tween 20 and 40. On the other hand, cells are negative for hydrolysis of gelatin, starch, and DNA. The type strain KX 20^T was isolated from deep seawater and the G+C content of the genomic DNA of this strain is 62.9 mol% (Wang, M.-q., & Sun, L., 2016).

2.9.2.4 *Pseudomonas pachastrellae*

P. pachastrellae cells are Gram-negative, aerobic, non-pigmented, encapsulated, and rod-shaped. Their typical dimensions are from 1.4-1.6 µm in length and with 0.4–0.5 µm diameter. The cells are motile by a single flagellum and show positive activity to oxidase and catalase. Colonies of *P. pachastrellae* cells are not pigmented, whitish color and transparent with circular form, and smooth structure with 2–3 mm in diameter on TSA and MA. Pyocyanin is not produced on King A medium and fluorescent pigments on King B medium were not produced. During growth of strain KMM 330^T on King A medium it has been notices a slightly yellowish diffusible pigment (Romanenko et al., 2005).

The cell growth occurs at temperatures from 7-41 °C in 0-10% (w/v) NaCl. The cells are positive for hydrolysis of Tween 40 and Tween 80 but negative for hydrolysis of starch and gelatin. It can well tolerate 8% NaCl. Additionally, the strain is negative for denitrification, arginine dihydrolase, urease, lysine decarboxylase, ornithine decarboxylase, acetoin production, H_2S production, indole production, aesculin and DNA hydrolysis. The type strain, KMM 330^T, was isolated from the sponge *Pachastrella* sp., which was collected from the Philippine Sea at a water depth of 750 m (Romanenko et al., 2005).

2.9.2.5 *Pseudomonas taeanensis*

P. taeanensis is bacterium that grows in aerobic conditions, it is Gram-negative, and cells are non-pigmented. The rod-shaped cells of this species have dimensions of 1.8-2.5 μm in length and 0.8-1.0 μm width. The cells are motile with a single flagellum. *P. taeanensis* grows in colonies that are smooth in texture and they are not pigmented, mostly whitish and translucent. Colonies can be between 2-3 mm in diameter. On King A medium there is no noticeable production of pyocyanin and fluorescent pigments on King B medium were not detected. The growth conditions for this bacterium are 0-5% (w/v) NaCl, at 4-30 °C (with optimum temperature 25-28 °C) and at pH 6.0-9.5 (optimum pH 6.5-8.0). Cells show positive reaction with catalase and oxidase, but they are negative for urease and DNase. There is present reduction of nitrate to nitrite, but nitrite is not further reduced. The cells are positive for hydrolyzation of Tween 20, Tween 40, Tween 60, and Tween 80. The type strain, MS-3^T was isolated from crude oil-contaminated seawater in the Taean area of Korea (Lee et al., 2010).

Recent studies show that *P. taeanensis* MS-3^T is capable of degrading petroleum oils including gasoline, diesel, and kerosene (Lee et al., 2014). There is also a potential of using this strain for the wastewater treatment for the phosphorus (P) removal in process of activated sludge (Cai et al., 2007; Li et al., 2012). *P. taeanensis* MS-3^T was inspected for inhibitory effects of single treatments of heavy metals: Cd, Cu, Zn, Pb, and Ni and Cd showed the most toxic activity to the cell (Yoo et.al., 2018).

2.9.2.6 *Vibrio natriegens*

V. natriegens is a Gram-negative bacterium from the marine environment that doesn't show any pathogenic properties and it was firstly isolated from salt marshes (Payne et al., 1961). Independent studies of Eagon and Cenedella's group have shown that this species can be the one of the fastest growing organisms with its generation time that is less than 10 minutes (Eagon, R. G., 1962; Maida et al., 2013). The immense potential for different applications in molecular biology and biotechnology lays in this species because of its very low doubling time. Various genetic tools and methods are developed to engineer these strains for common biotechnological processes (Weinstock et al., 2016).

A study of Hoffart et al. (2017) showed that characteristics of this bacterium fulfill basic requirements for applications in biotechnology. On glucose as substrate, the growth rate is two times higher than with *E. coli*, *Bacillus subtilis*, *Corynebacterium glutamicum* and yeast. *V.*

natriegens is prototrophic and its metabolism is versatile with ability to metabolize various substrates (Hoffart et al., 2017).

V. natriegens has several mechanisms that enable high capacity for ribosome synthesis which results in high growth rates: high rRNA gene dose and stimulation of promoter activity by UP elements and Fis. Regulation of rRNA transcription might be regulated with similar mechanisms as present in *E. coli* because of similarities of rRNA promoter sequences of these two species (Aiyar et al., 2002). The cell volume of exponentially growing cells is 3.50 μm^3 $cell^{-1}$ and stationary growing cells have volume of 0.93 μm^3 $cell^{-1}$ (Thoma, F., & Blombach, B., 2021).

2.10 The Open Science Framework (OSF)

OSF is a platform established to enhance communication between scientists and researchers all over the world. The main purpose of this tool is to establish open and centralized workflows by capturing various aspects of a research lifecycle such as research idea, design of study, analyzing, collecting, and storing of research data and lastly writing and publishing research work (Foster, E. D., & Deardorff, A., 2017).

The Center for Open Science (COS), as a nonprofit organization, founded OSF to support researchers by developing tools and infrastructure for saving and managing their research. An idea of OSF is to practice open science and to preserve integrity and reproducibility of scientific research. This network has ability to create and develop projects which are available for diverse audience. OSF promotes reproducibility, transparency, and management of research data (Foster, E. D., & Deardorff, A., 2017). A detailed protocol for OSF users with functions and opportunities of the framework is summarized in work of Sullivan, I., DeHaven, A. C., & Mellor, D. T., (2019).

3 Methodology

3.1 BASIC assembly

BASIC assembly was performed through three steps: the BASIC clip reaction, BASIC clip purification and BASIC clip assembly. The BASIC clip reaction was performed in 200 µl PCR tube. The following reagents were added: 23 µl dH2O, 3 µl 10 x Promega T4 buffer, 1 µl prefix linker, 1 µl suffix linker, 200 nmol of BASIC biopart DNA, 1 µl BsaI-HF®v2 (New England Biolabs, R3733) and 0.5 µl T4 ligase (New England Biolabs, M0202L). The total volume of mixed solution for this reaction is 30 µl. The PCR tube was placed in a PCR thermal cycler, and incubated with the following program: [37 °C for 2 minutes, 20 °C for 1 minute] x 20 cycles, 60 °C for 20 minutes and stored at 4 °C.

The clipped parts were further purified by performing the BASIC clip purification, which is a magnetic bead purification method. Mag-Bind TotalPure NGS was used to purify restriction-ligation reactions (Omega Bio-Tek, M1378-01). Firstly, the 70% EtOH was prepared, and it was used 0.5 ml per BASIC reaction. The magnetic beads that were stored at 4 °C were placed into homogeneous mixture that was vortexed for 1 minute to warm to a room temperature. The volume of 54 µl of magnetic beads was mixed the 30 µl BASIC restriction-ligation solution that was obtained in previous step. The mixture was well mixed with pipetting up and down without making the bubbles. The magnetic beads were left to bind to DNA for 5 min and then placed on magnetic stand for 2 minutes. The solution was removed with a 200 µl pipette tip and wash step was performed.

190 µl of 70% EtOH was added and incubated at room temperature for 30 seconds and the solution was then removed. 190 µl of 70% EtOH was then added again for 30 seconds, and solution was removed from each well. The plate was left for 5 min without the lid (under the protection of a laminar flow hood) to let the magnetic beads to dry. Solution was removed from magnet and 32 µl of dH_2O. The DNA was left for 1 minute to elute and again placed on magnetic stand which allowed the magnetic beads to form an immobilized ring and the solution became clear in a minute. 30 µl of dH_2O was pipetted from the wells with eluted DNA into a fresh 1.5 ml micro-tube.

Then the BASIC clip assembly was performed and for each reaction in a 200 µl PCR tube that had 10 µl total volume of: 1 µl 10 x NEB CutSmart buffer, 1 µl each part and filled with dH_2O to 10 µl. The solution was mixed by pipetting up and down and it was incubated in thermocycler

with the BASIC assembly program that included temperature of 50 °C for 45 minutes and stored at 4°C overnight. After the reaction, the cells were transformed via heat-shock. A volume of 50 µl of previously prepared competent cells of *E. coli* DH5α was used to transform 5 µl of each BASIC assembly. The competent cells that were stored at -80 °C were thawed on ice for 5 to 10 minutes and 5 µl of BASIC assembly was cooled on ice in 1.5 ml Eppendorf tube. In the BASIC assembly was added 50 µl of competent cells and incubated on ice for 20 minutes. Then the tube was placed in water bath for 45 second at 42 °C. In tube was added 200 µl of SOC medium and it was incubated with shaking at 37 °C for 1 h. The cells were spread on agar plate with antibiotic and left at 37 °C overnight.

For the verification of successful BASIC assembly, colony PCR was performed. Firstly, in a 96-well plate, a 15 µl of mixture was made: 7.5 µl of Taq Plus DNA PolymeraseMaster Mix (VWR, 733-2597), 0.5 µl LMP_F, 0.5 µl LMP_R and 6.5 µl of dH_2O. The colony was picked with pipette tip and the tip was placed in well of 96-well plate. On well was placed lid and it was centrifuged. The plate was left in thermocycler on the program: 95°C for 7.5 minutes, 25 x [95 °C for 30 seconds, 58 °C for 1 min/1 kbp, 72 °C for 7 minutes], 72 °C for 7 minutes and stored at 4 °C. Then the electrophoresis was performed in 0.8-1% agarose gel. Solution was placed in microwave to boil, and 5 µl of ethidium-bromide was added in it and let to dry for 10 minutes. Loading dye was added in sample, 1 µl of dye to 5 µl of the sample and samples were inserted in wells. The program was set on 150 V for 30 minutes.

3.2 The induction study

3.2.1 Preparation of the *E. coli* cells and LB media

The volume of 7.5 ml of LB buffer was transferred into tube and it was added 7.5 µl of antibiotic (kanamycin). The frozen cells of *E. coli* BB23-S4 from glycerol stock were scraped on the top of the micropipette tip and used to inoculate LB with 50 µg/mL kanamycin. The tube with inoculated culture was placed in shaker at 37°C and left overnight. Inoculation of cells was performed under sterile conditions in the laminar flow hood.

The LB growth medium was made by mixing 10 g of peptone, 5 g of yeast, 10 g of NaCl and filled with distillated water to a volume of 1 l. Mixture was stirred on magnetic stirrer and autoclaved at 121°C.

3.2.2 The induction assay

The induction study was performed on *E. coli* cells (BB23-S4) using LB media supplied with antibiotic (kanamycin). L-arabinose was added in every tube to obtain desired concentrations, the tubes were shaken and the 96 well plate with black and clear bottom for BioTek Synergy H1 modular multimode microplate reader was prepared according to **Table 2**.

Column 12 on **Table 2** represents negative control, in other words, cells of *E. coli* were not inoculated these wells. A volume of 200 µl of LB media with antibiotic was added in wells marked with blank and NI (NI - no induction), then 200 µl of LB buffer with selected amount of L-arabinose and kanamycin was pipetted in other wells and lastly, the cell culture was inoculated using 96 pin replicator in every well except ones with the blanks. Four of the samples were marked with NI, which represents wells without L-arabinose.

The plate was placed in BioTek Synergy H1, and temperature was set on 30°C. Three parameters were measured: optical density at 600 nm (OD_{600}), GFP (Ex/Em 484/515 nm) and mKate (Ex/Em 585/615 nm) for 48 hours. The sfGFP was measured for further use in the non-linear models in R.

Table 2 Selected concentrations of L-arabinose *and their spatial organization in 96 well plate that was placed in BioTek Synergy H1 modular multimode microplate reader, NI represents samples without L-arabinose and blank wells are the ones without cells.*

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
A	NI	0.1	0.2	0.4	0.6	0.8	1.0	1.20	1.40	1.60	1.80	blank
B	NI	0.1	0.2	0.4	0.6	0.8	1.0	1.20	1.40	1.60	1.80	blank
C	NI	0.1	0.2	0.4	0.6	0.8	1.0	1.20	1.40	1.60	1.80	blank
D	NI	0.1	0.2	0.4	0.6	0.8	1.0	1.20	1.40	1.60	1.80	blank
E	2.0	2.5	3.0	3.5	4.0	5.0	6.0	7.0	8.0	9.0	10	blank
F	2.0	2.5	3.0	3.5	4.0	5.0	6.0	7.0	8.0	9.0	10	blank
G	2.0	2.5	3.0	3.5	4.0	5.0	6.0	7.0	8.0	9.0	10	blank
H	2.0	2.5	3.0	3.5	4.0	5.0	6.0	7.0	8.0	9.0	10	blank

The GFP was measured in relative fluorescence units and incorporated as unitless measure in Equation 29 while the parameters such as constant K, n and GFP_{max} were estimated with non-linear models. The data that included normalized GFP signal with OD_{600} was incorporated in Equation 30. $ARAconc$ represents concentration of L-arabinose and parameter K is in concentration units while n is unitless.

$$GFP(ARA) \cong \frac{GFP_{max} ARA_{conc}^{n}}{K^{n} + ARA_{conc}^{n}} \qquad \text{Equation 29}$$

$$GFP(ARA)/OD_{600} \cong \frac{GPF/OD_{600_{max}} ARA_{conc}^{n}}{K^{n} + ARA_{conc}^{n}} \qquad \text{Equation 30}$$

Non-linear model given in Equation 29 can be slightly modified to include an unitless constant K, instead of estimating the parameter in concentration units. Dividing the numerator and denominator with $ARAconc_{max}^{n}$, which represents the lowest concentration of L-arabinose that gives saturation, and with simple mathematical modifications of the Equation 30, we can obtain Equation 31. A dimensionless parameter K is calculated as $\frac{K}{ARA_{conc}max}$. In further text, as synonym for the Equation 31 will be used nondimensionalized Hill function.

$$GFP(ARA) \cong \frac{GFP_{max}\left(\frac{ARA_{conc}}{ARA_{conc}max}\right)^{n}}{K^{n}+\left(\frac{ARA_{conc}}{ARA_{conc}max}\right)^{n}} \qquad \text{Equation 31}$$

Three previous equations (Equation 29, 30 and 31) are used for building non-linear models in R. For these purposes, a *nls*() function from the 'stats' package (Ritz, C., & Streibig, J. C., 2008) was used.

Instead of a simple GFP signal, non-linear models can include rates of GFP change over time. The model that is based on GFP rates is displayed in Equation 32.

$$GFP_{rate}(ARA) \cong \frac{GFP_{rate}maxARA_{conc}^{n}}{K^{n}+ARA_{conc}^{n}} \qquad \text{Equation 32}$$

A symbol R can be used instead of $GFP_{rate}max$, the highest rate of GFP change over and Equation 32 can be modified resulting in Equation 33.

$$GFP_{rate}(ARA) \cong \frac{ARA_{conc}^{n}R}{K^{n}+ARA_{conc}^{n}} \qquad \text{Equation 33}$$

The nondimensionalized form of Equation 33 can be obtained analogously as for simple GFP signal by dividing $ARAconc$ with $ARAconc_{max}$. The parameter K estimate from this model is unitless and it is approximated by a non-linear model given in Equation 34.

$$GFP_{rate}(ARA) \cong \frac{R\left(\frac{ARA_{conc}}{ARA_{conc}max}\right)^{n}}{K^{n}+\left(\frac{ARA_{conc}}{ARA_{conc}max}\right)^{n}} \qquad \text{Equation 34}$$

3.3 Solving differential equations in R

Ordinary differential equations (ODEs) and system of ODEs can be solved with package 'deSolve' in R (Soetaert et al., 2015). ODEs and systems of ODEs presented in this study were solved in R by a code that is published as a part of a project in OSF (https://osf.io/39j5f/).

3.4 Methodology of a dynamic response of L-arabinose activated gene expression across host specific growth rates

Chapter 3.2 explained that in every well of 96 well plate OD_{600}, GFP and mKate signals were measured for 48 hours. The GFP signal that was measured during experimental time, served to estimate different GFP-time curves and slope of the linear part of the curve represents the rate of GFP change over time.

The slope was calculated by performing simple linear regression using R function *all_easylinear*() from a 'growthrates' package. The highest rate estimate is labeled as *R*, and it is in units of $\left(\frac{1}{time}\right)$. The dynamic response of L-arabinose-activated expression can be represented as simple ODE:

$$\frac{dGFP}{dt} = \frac{GFP_{max} ARA_{conc}^{n} R}{K_{AraC}^{n} + ARA_{conc}^{n}} - \alpha GFP \qquad \text{Equation 35}$$

where GFP represent GFP signal, $\frac{dGFP}{dt}$ is change of GFP signal over time in units $\left(\frac{1}{time}\right)$, GFP_{max} is maximal GFP signal from the late log phase, R is the highest rate estimate in units $\left(\frac{1}{time}\right)$, $ARAconc$ is concentration of L-arabinose, K_{AraC} is the activation coefficient that has concentration units, *n* is Hill coefficient, unitless, and α is a removal rate in units $\left(\frac{1}{time}\right)$.

If cells are actively growing, the removal rate is approximated only by dilution rate (α_{dil}) that is equal to specific growth rate (μ). The Equation 35 can be modified to include the growth rate in the sink term which is presented as:

$$\frac{dGFP}{dt} = \frac{GFP_{max} ARA_{conc}^{n} R}{K_{AraC}^{n} + ARA_{conc}^{n}} - \mu GFP \qquad \text{Equation 36}$$

where μ is specific microbial growth rate that is unique for every strain, and it is measured in $\left(\frac{1}{time}\right)$. A left side of Equation 35 and both terms on the right side of Equation 35 are in reverse time dimensions. The specific growth rates of wild type strains of species selected for this thesis were estimated in a study of PhD candidate at UiT, Dennis Tin Chat Chan.

If the term $\frac{GFP_{max}ARA_{conc}^{n}R}{K_{AraC}^{n}+ARA_{conc}^{n}}$ is named as $f(ARA)$, then the simplified Equation 35 is formulated as:

$$\frac{dGFP}{dt} = f(ARA) - \mu GFP \, . \qquad \text{Equation 37}$$

The term $f(ARA)$ can be modified to estimate $K' = \frac{K_{AraC}}{ARA_{conc}max}$ and nondimensionalized form of the Equation 35 is:

$$\frac{dGFP}{dt} = \frac{GFP_{max}\left(\frac{ARA_{conc}}{ARA_{conc}max}\right)^{n}R}{K'^{n} + \left(\frac{ARA_{conc}}{ARA_{conc}max}\right)^{n}} - \mu GFP \, . \qquad \text{Equation 38}$$

The response time ($T_{1/2}$) can be estimated from Equation 39. A symbol μ represents specific growth rate in $\left(\frac{1}{time}\right)$ dimensions and $T_{1/2}$ is expressed in time dimensions.

$$T_{1/2} = \frac{log(2)}{\mu} \qquad \text{Equation 39}$$

3.5 Mathematical model of a synthetically constructed toggle switch

The toggle switch performance was modeled for a device presented in **Figure 1**. The device consists of two promotors: P_{Tet} that can be induced with anhydrotetracycline (aTc) and P_{BAD} that can be induced with L-arabinose. The activity of P_{BAD} promoter can be monitored by measuring the GFP signal fluorescence and activity of P_{Tet} promoter can be detected and measured by a mKate fluorescence.

A production of a protein (in this case: AraC or TetR) is dependent on two processes: protein degradation and dilution. Under assumption that a degradation rate, α_{deg}, is negligible, $\alpha = \mu$, where μ represents the specific growth rate in $\left(\frac{1}{time}\right)$, a mathematical model of the toggle switch can be described with following equations (Equation 40 and 41) that rely on Equation 27 (Alon, U., 2019).

$$\frac{dC_{AraC}}{dt} = \frac{\beta_{AraC} R}{1 + \left(\frac{C_{TetR}}{K_{TetR}}\right)^{n_{AraC}}} - \mu C_{AraC} \qquad \text{Equation 40}$$

$$\frac{dC_{TetR}}{dt} = \frac{\beta_{TetR} R}{1 + \left(\frac{C_{AraC}}{K_{AraC}}\right)^{n_{TetR}}} - \mu C_{TetR} \qquad \text{Equation 41}$$

Terms that act in Equation 40 and 41 are C_{AraC}, C_{TetR} that are intracellular concentrations of AraC and TetR respectively, terms $\beta_{AraC} R$ or $\beta_{TetR} R$ represent maximum synthesis rates of proteins, and they are a product of factor β that shows maximal expression level of promoter (P_{BAD} for AraC and P_{Tet} for TetR) estimated in concentration units, and factor R that is the highest GFP rate estimate from the induction study (measured in $\left(\frac{1}{time}\right)$), therefore βR term has $\left(\frac{concentration}{time}\right)$ dimensions. K_{AraC} and K_{TetR} are the effective equilibrium dissociation constants of these repressors to their binding sites in the promoter region (the activation coefficients) in concentration units, n is effect of cooperative repression of the promoters (the Hill coefficient), unitless, and μ is specific growth rate measured in $\left(\frac{1}{time}\right)$.

The protein production can be modeled with the Hill term formulated as $\left(\frac{\beta_{AraC}R}{1+\left(\frac{C_{TetR}}{K_{TetR}}\right)^{n_{AraC}}}\right)$ and $\left(\frac{\beta_{TetR}R}{1+\left(\frac{C_{AraC}}{K_{AraC}}\right)^{n_{TetR}}}\right)$, while the dilution of the proteins can be represented with the sink term μC_{AraC} and μC_{AraC} (Equation 40 and 41).

If a concentration of AraC protein derived from *araC* gene is proportional to the intensity of mKate signal, consequentially the TetR protein concentration synthetized by the activity of *tetR* gene is proportional to intensity of GFP signal. For experimental purposes, the system can be represented as:

$$\frac{dGFP}{dt} = \frac{\beta_{AraC}R}{1+\left(\frac{mKate}{K_{TetR}}\right)^{n_{AraC}}} - \mu GFP \qquad \text{Equation 42}$$

$$\frac{dmKate}{dt} = \frac{\beta_{TetR}R}{1+\left(\frac{GFP}{K_{AraC}}\right)^{n_{TetR}}} - \mu RFP \qquad \text{Equation 43}$$

where $\mathrm{C_{AraC}} \propto \mathrm{mKate}$ and $\mathrm{C_{TetR}} \propto \mathrm{GFP}$.

If it is assumed that:

$$u = \frac{C_{AraC}}{K_{AraC}} \; and \; v = \frac{C_{TetR}}{K_{TetR}}, \qquad \text{Equation 44}$$

it is possible to nondimensionalize Equation 40 and 41 by dividing both sides with μ, the specific growth rate, and activation constant, K_{AraC} or K_{TetR}. The terms u and v are unitless. Nondimensionalized system of ODEs is represented by Equation 45. This mathematical model is based on work of Gardner, T., Cantor, C. & Collins, J. (2000).

$$\frac{du}{d\tau} = \frac{\alpha_{AraC}}{1 + (v)^n} - u$$

Equation 45

$$\frac{dv}{d\tau} = \frac{\alpha_{TetR}}{1 + (u)^n} - v$$

Unitless terms α_{AraC} and α_{TetR} are calculated as $\alpha_{AraC} = \frac{\beta_{AraC}R}{\mu K_{AraC}}$ and $\alpha_{TetR} = \frac{\beta_{TetR}R}{\mu K_{TetR}}$ where β is maximal expression level, R is the highest rate estimate from the induction study, K is activation constant, n is the Hill coefficient, μ specific growth rate and τ is respective time obtained as multiplied time variable with the specific growth rate, $\tau = t\mu$.

3.6 Open science and reproducibility

In an effort to make this thesis project open and reproducible, a code that has been written for the analysis performed in this study is supplied in a form of OSF project found on the following link: https://osf.io/39j5f/. The code is published as a RMarkdown file (.Rmd) and HTML-rendered file.

The data incorporated in the R code, including the induction study and the specific growth rate data, can be found in Data folder of the OSF project.

4 Results

4.1 Parametrization of the Hill function

The experimental data from the induction study with *E. coli* cells was used to parametrize the Hill function and to estimate specific parameters presented in Equation 29 and 30. The method of a non-linear regression was chosen for this purpose. The analysis included estimation of at least one of two parameters, *n* and/or *K*.

The GFP fluorescence, OD_{600} and mKate fluorescence were measured for 48 hours during induction assay to collect the data from all stages of bacterial cell growth. The data selected for non-linear models with GFP and GFP/OD_{600} signals was estimated from the cells that have reached the late log phase. This was done by selecting a time point for the late exponential phase by looking at resulting growth curves from every well. It has been approximated that 11 hours 50 minutes and 46 seconds after cell inoculation on 96 well plate with black and clear bottom, cell have reached desired state. The effect of L-arabinose induction was predicted by monitoring the intensity of the GFP and GFP/OD_{600} signals from the cells in the late exponential phase. The effect is graphically presented in **Figure 4** and **5**. The GFP and GFP/OD_{600} signals were used as indirect measure of promoter activity in examined cells.

The GFP_{max} value was estimated from a data collected in the induction curve study as an average GFP value of the samples in saturation part of the curve presented at **Figure 4** and the standard deviation was calculated. A similar procedure was performed to estimate $\mathrm{GPF/OD_{600}}_{\mathrm{max}}$ value (**Figure 5**). In the following models, $\mathrm{GFP_{max}}$ and $\mathrm{GPF/OD_{600}}_{\mathrm{max}}$ were used as constant parameters and **Table 3** shows their estimates with standard deviation.

Table 3 GFP_{max} and GFP/OD_{600}max estimates *and their standard deviation. Both parameters were calculated as average value from data collected from a saturation part of curves shown in* ***Figure 4*** *and* ***5****.*

	$\mathbf{GFP_{max}}$	$\mathbf{GPF/OD_{600}}_{\mathbf{max}}$
Estimate average value	2643.500	4622.374
Standard deviation	591.330	797.673

A maximal concentration of L-arabinose ($ARAconc_{max}$) used in a non-dimensional analysis was estimated as the lowest concentration of L-arabinose that results in saturation. By observing

Figure 4 and **5** concentration that gives saturation in all four replicates is 3 M ($ARAconc_{max} = 3\ M$).

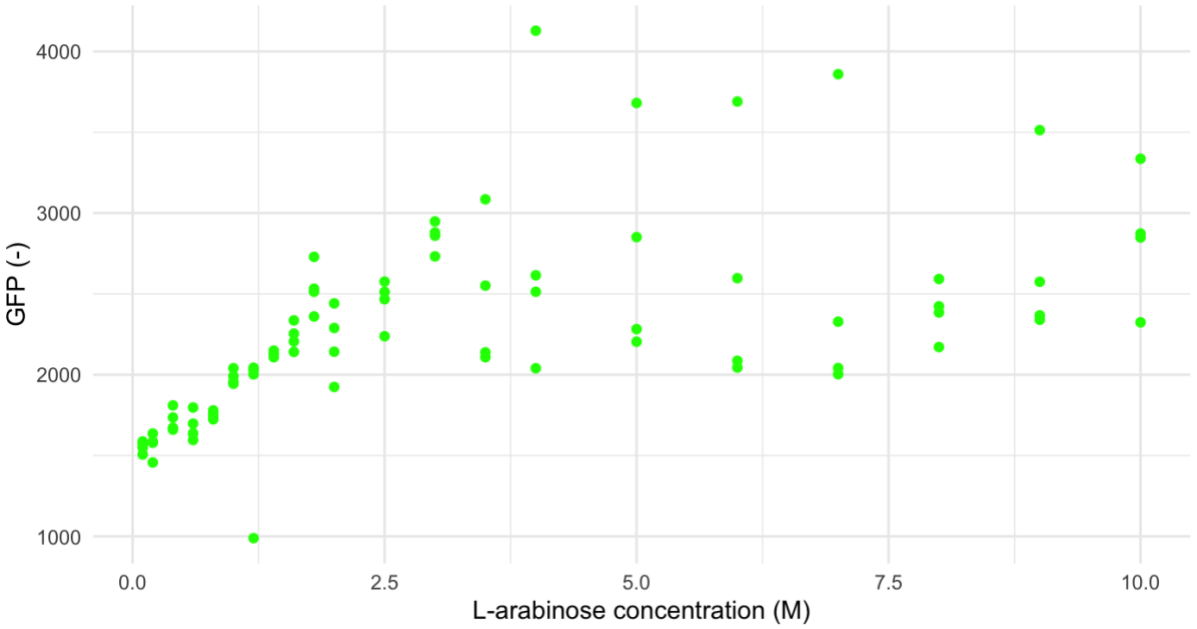


Figure 4 The effect of L-arabinose concertation on GFP signal *is manifested as increasing intensity of GFP fluorescence with inducer concentration until saturation is achieved. This effect was estimated in the late log phase of cell growth. Selected time point was 11 hours 50 min and 46 minutes from the beginning of the experiment (cell inoculation) which should represent the time when the cells have reached the late log phase. The y-axis represents GFP signal and L-arabinose concentration is displayed on x-axis.*

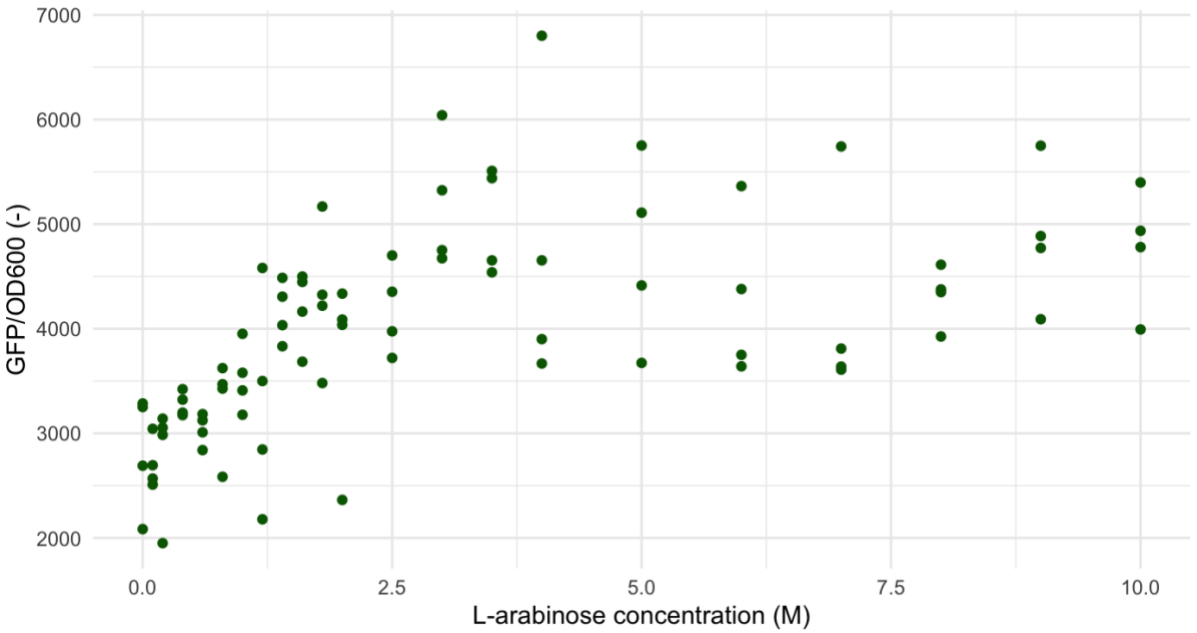


Figure 5 The effect of L-arabinose concertation on GFP/OD$_{600}$ signal *is manifested as increasing intensity of GFP/OD$_{600}$ signal with inducer concentration until saturation is achieved. This effect was estimated in a late log phase of cell growth. Selected time point was 11 hours 50 min and 46 minutes from the beginning of the experiment (cell inoculation) which should represent the time when the cells have reached the late log phase. GFP signal is divided by optical density at 600 nm and presented on y-axis while L-arabinose concentration is given on x-axis.*

4.1.1 The effect of different *n*-values on a non-linear model

4.1.1.1 The effect of different *n*-values on a non-linear model with GFP data

The estimation of the *n*-value effect on non-linear model formulated in Equation 29 with experimental GFP data was performed. The *n* parameter value was varied from 1 to 4. This study resulted in a better model fit for lower n ($n = 1$ or $n = 2$), which is graphically represented by a regression line on every diagram in **Figure 6**.

The analogous effect was predicted by the residual graphs presented in **Figure 7**. The results presented in **Figure 7** show similar shape of deviation for any selected *n*. However, the residuals of non-linear models for $n = 3$ and $n = 4$, are distributed further from zero compared with model residuals when *n* was assumed as 1 or 2.

These models (**Figure 6**) predicted an increase of a residual standard error (RSE) as a result of higher Hill coefficients implemented in the model. The lowest RSE was calculated in the simulation when *n*-value was equal to 1.

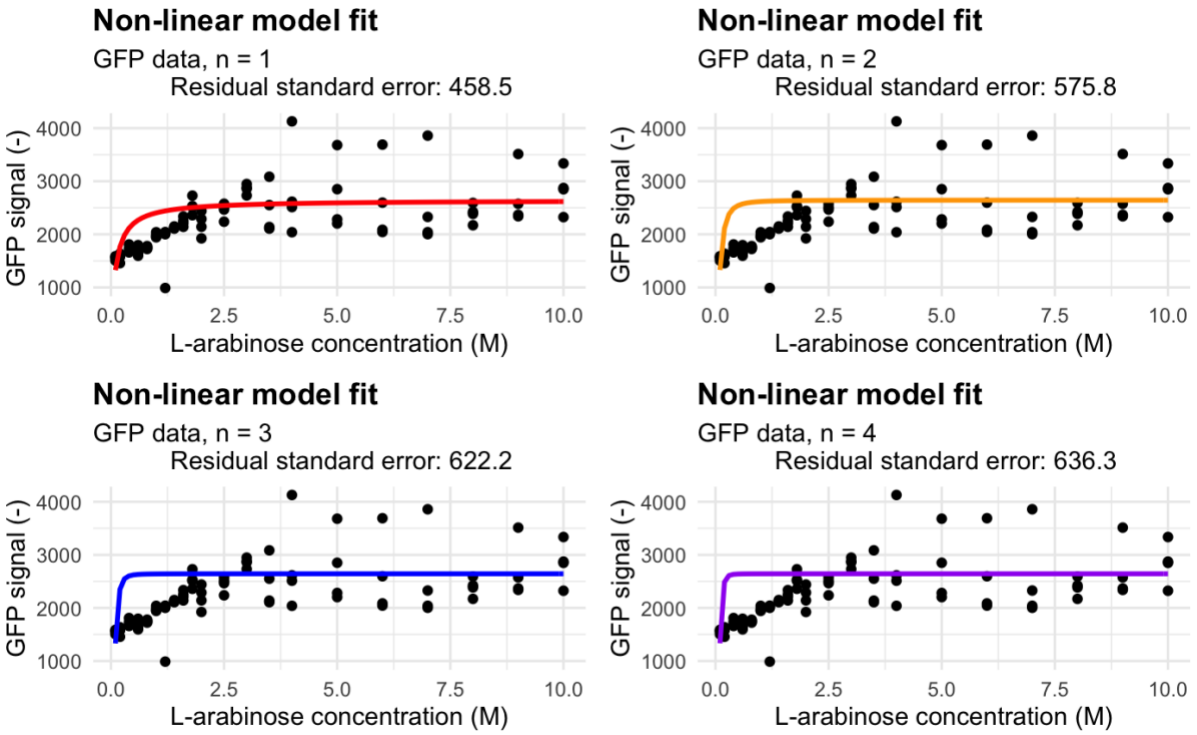


Figure 6 The effect of different Hill coefficients *is estimated as a poorer quality of a model fit and a lower model accuracy for higher n-values (as confirmed by a higher residual standard error estimate). The comparison of the effect of n-value to a fit of non-linear regression was simulated for GFP signal data. Red regression line represents non-linear model with n = 1, orange regression line is for non-linear model with n = 2, blue regression line is non-linear model with n = 3 and purple regression line is for non-linear model including n = 4. Estimated residual standard errors for the models are shown in headings in the top part of the graphs.*

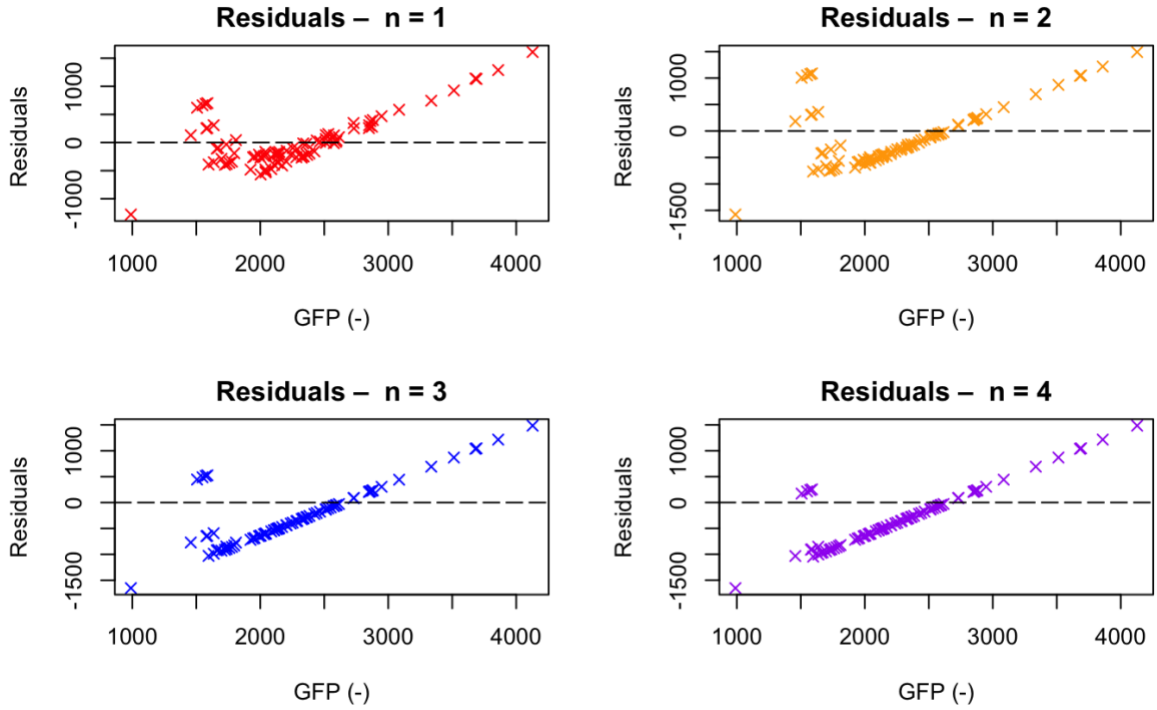


Figure 7 The effect of different Hill coefficient on residuals *from simulated non-linear models is estimated as further distributed residuals from zero for higher n estimates. Different colors are used to represent different n-values: red for n = 1, orange for n = 2, blue for n = 3 and purple for n = 4. Simulations of presented non-linear models were performed with simple GFP signals.*

4.1.1.2 The effect of different *n*-values on non-linear model with GFP/OD_{600} data

The *n*-value effect on a non-linear model (Equation 30) with experimental GFP/OD_{600} data was simulated. A better model fit was predicted when the lower values of *n* were incorporated in the equation. The graphical representation (**Figure 8**) shows that the estimated residual standard errors (RSEs) of the non-linear models containing GFP/OD_{600} data were higher than previously approximated with the simple GFP signals (**Figure 6**).

The residual graph (**Figure 9**) indicates that increase of *n*-value affects the distribution of residuals that is seen as a further dislocating of residuals from zero. The same figure presents a better model fit for lower values of *n*, such as $n = 1$, which is additionally supported by the RSEs showed in **Figure 8**. These RSEs suggested that the lowest error is predicted for approximation of the Hill coefficient as 1. Furthermore, the residuals from this model are further distributed from zero compared to a simulation with the GFP data (**Figure 6** and **7**).

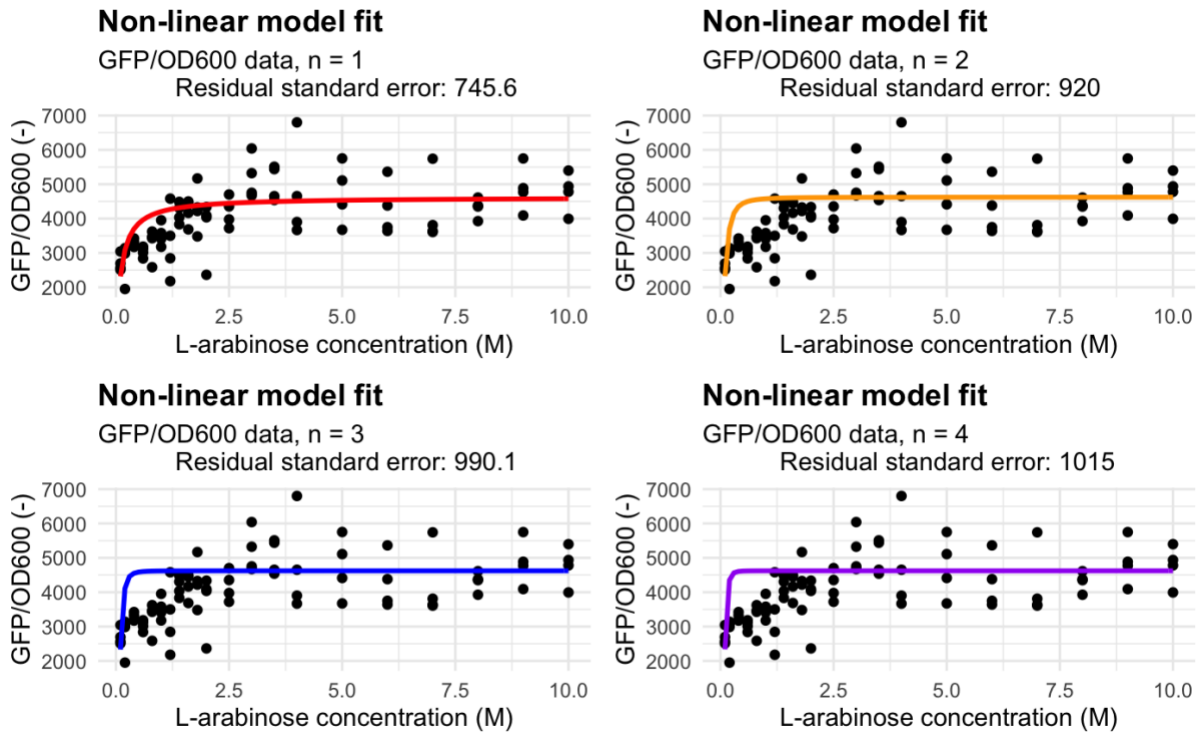


Figure 8 The effect of different Hill coefficient *is predicted as a poorer quality of a model fit and a lower model accuracy for higher n-values (as confirmed by a higher residual standard error estimate). The comparison of the effect of n value to a fit of non-linear regression was simulated for the GFP/OD₆₀₀ data. Red regression line represents non-linear model with n = 1, orange regression line is for non-linear model with n = 2, blue regression line is non-linear model with n = 3 and purple regression line is for non-linear model including n = 4. Estimated residual standard errors for the models are shown in headings in a top part of the graphs.*

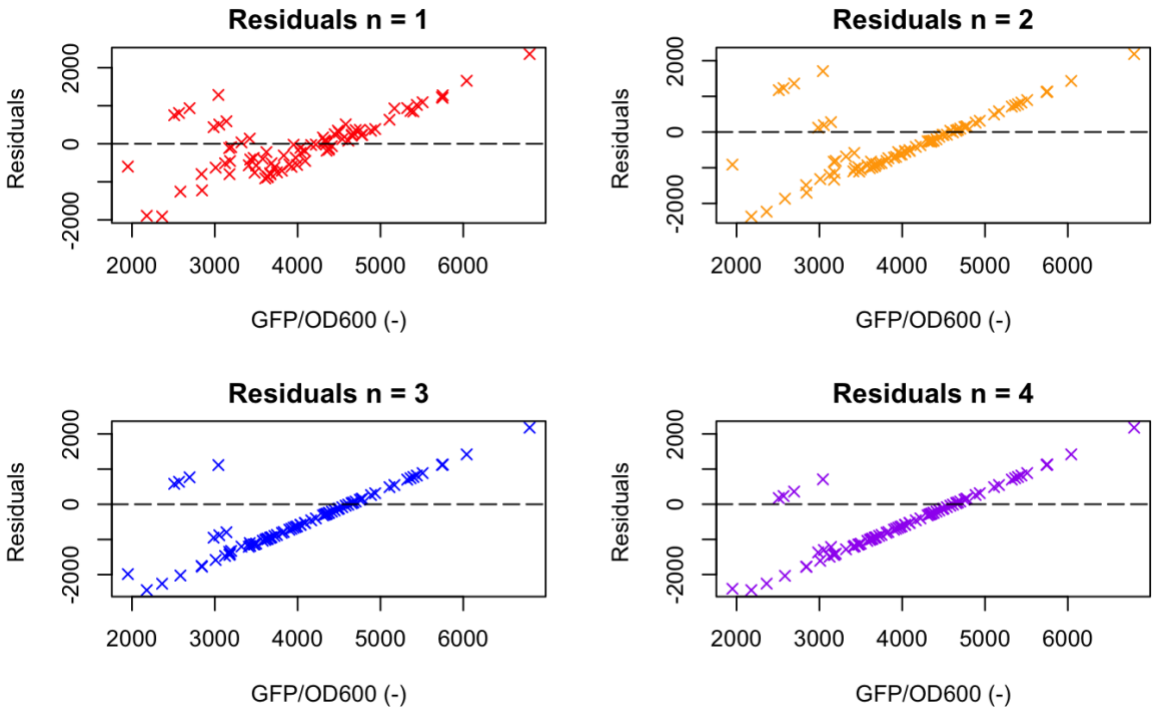


Figure 9 The effect of different Hill coefficient on residuals *is predicted as further distributed residuals from zero for higher n estimates in the simulated non-linear models. Different colors are used to represent different n-values: red for n = 1, orange for n = 2, blue for n = 3 and purple for n = 4. Simulations of presented non-linear models were performed on GFP/OD₆₀₀ data.*

4.1.2 Estimation of *n* and *K* parameters for GFP data

The non-linear model from the Equation 29 was applied to estimate *n* and *K* values. GFP_{max} was calculated previously (**Table 3**) and incorporated as a constant parameter in this model. The output of *nls*() function in R for this this model included estimated values for *n* and *K* together with statistical parameters (**Table 4**). P-values for both estimates were lower than 0.05, meaning that estimated p-values were significant. This indicates strong evidence against the null hypothesis and approximated estimates are considered significantly different from zero.

Table 4 The Hill (n) and activation coefficient (K) *estimates from the non-linear model supplied with the* ***GFP*** *data and statistical parameters: standard errors, t-value and p-value.*

	Estimate	**Standard Error**	**t-value**	**p-value**
n	0.72318	0.11719	6.171	2.46e-08
K	0.13215	0.04015	3.292	0.00147

The results of this model are graphically represented in **Figure 10**, which shows the fit of non-linear model with a red regression line and predicted residuals.

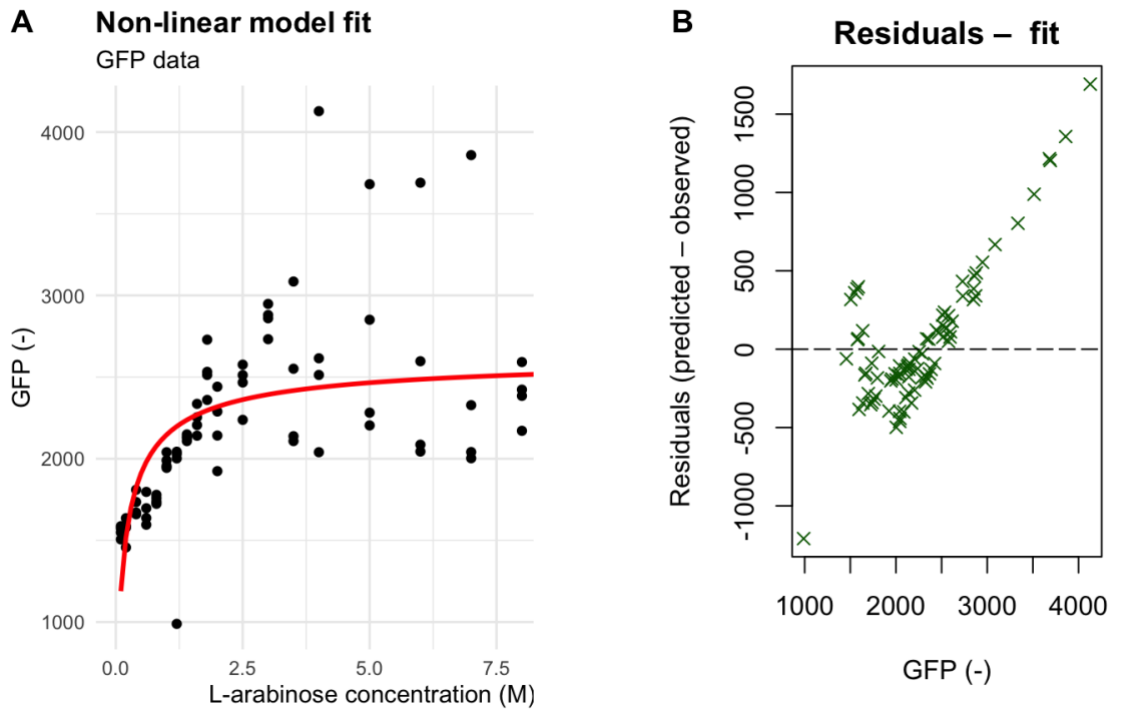


Figure 10 The non-linear model fit of the Hill function estimated with the GFP data *on the left (A) and residuals on the right (B). The model estimated values for the Hill coefficient (**n**) and activation coefficient (**K**). The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for intermediate GFP signals.*

4.1.3 Estimation of *n* and *K* parameters for GFP/OD_{600} data

The Hill function was fitted to a GFP/OD_{600} data which resulted in *n* and *K* coefficient estimates according to a model presented in Equation 30. The $GFP/OD_{600_{max}}$ was incorporated in a model as a constant parameter calculated beforehand (**Table 3**). The model showed similarly distributed residuals with a slightly higher deviation of data points compared to the ones from the model with GFP data (**Figure 10**). The graphical representation of the non-linear model fit is given in **Figure 11**. The *nls*() function output, including estimated *n* and *K* values with their statistical parameters, is presented in **Table 5**. P-values for both estimates were significant ($p < 0.05$).

Table 5 The Hill (n) and activation coefficient (K) *from the non-linear model supplied with the* ***GFP/OD₆₀₀*** *data and statistical parameters: standard errors, t-value and p-value.*

	Estimate	Standard Error	t-value	p-value
n	0.74454	0.11847	6.285	1.4e-08
K	0.11311	0.03401	3.326	0.00132

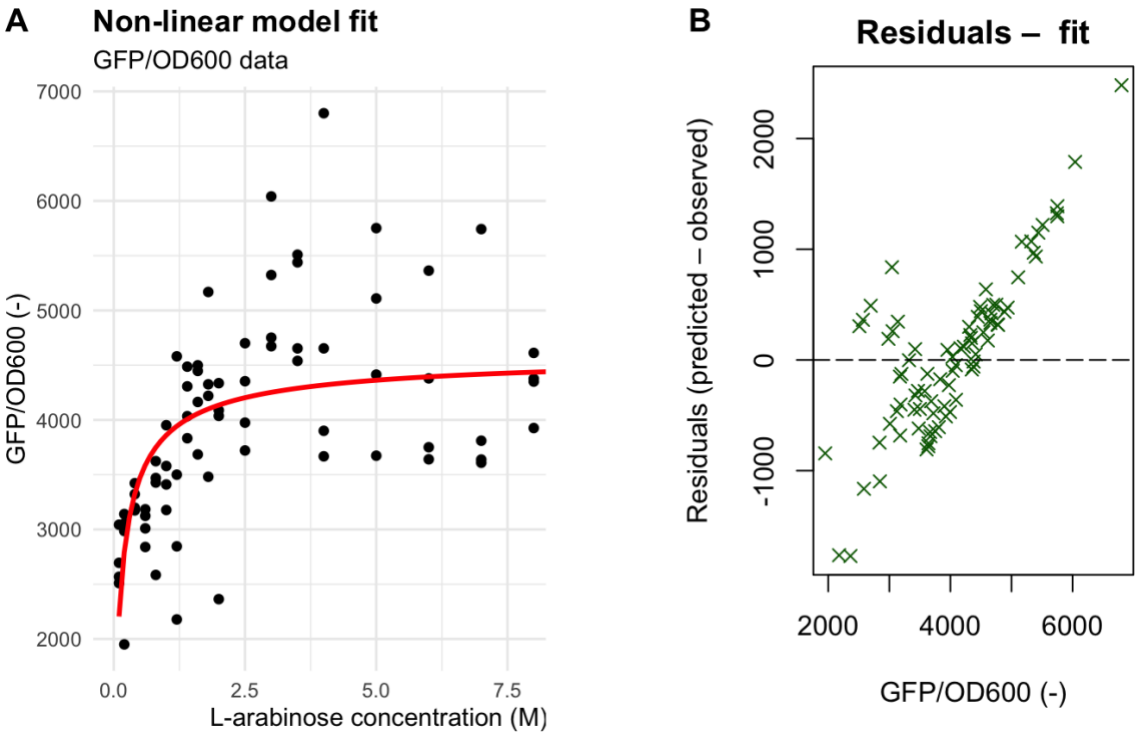


Figure 11 The non-linear model fit of the Hill function estimated with the GFP/OD₆₀₀ data *on the left (A) and residuals on the right (B). The model estimated values for the Hill coefficient (**n**) and activation coefficient (**K**). The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for intermediate GFP/OD_{600} signals.*

4.1.4 Parameter *K* estimate for GFP data

In this model, the constant *K* from the Equation 29 was estimated by the *nls*() function in R. The parameter *n* was approximated as a constant value that equals 1. GFP_{max} value was derived from the **Table 3**. The model fitted experimental GFP signal data resulting in a non-linear model output presented in **Table 6**. The p-value for *K* estimate is significant and reduced compared to the preceding model (**Table 4**). The residuals are not equally distributed around zero and show a higher deviation then in section 4.1.2. The non-linear model fit with residuals is graphically represented in **Figure 12**.

Table 6 The activation coefficient (K) *from the non-linear model supplied with the* ***GFP*** *data for* ***n = 1*** *and statistical parameters: standard errors, t-value and p-value*

	Estimate	Standard Error	t-value	p-value
K	0.19823	0.0396	6.404	8.61e-09

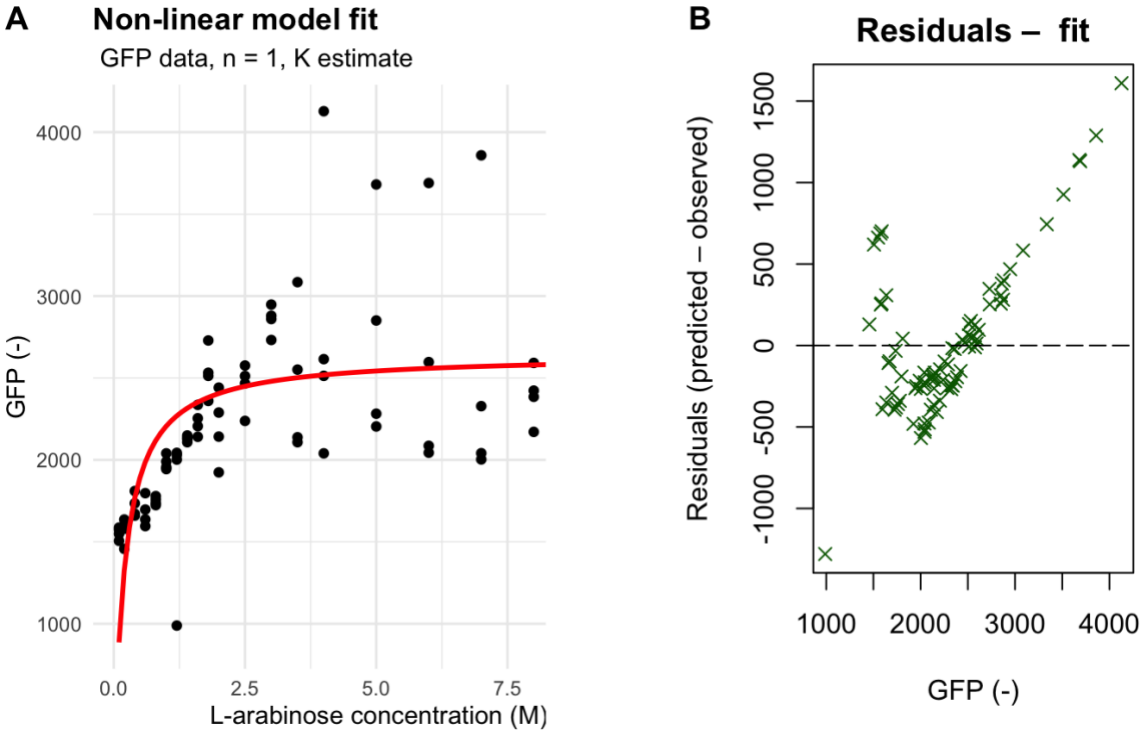


Figure 12 The non-linear model fit of the Hill function estimated with the GFP data *on the left (A) and residuals on the right (B). The model estimated value for the activation coefficient (**K**). The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for intermediate GFP signals.*

4.1.5 Parameter *K* estimate for GFP/OD_{600} data

The *K* parameter value (Equation 30) was predicted with *nls*() function in R. In this analysis, the parameter *n* was proposed to be a constant and equal to 1. $\mathrm{GPF/OD}_{600_{\max}}$ was assumed as constant (**Table 3**). Fitting the GFP/OD_{600} data to a non-linear model resulted in the estimate shown in **Table 7**. The p-value for the *K* estimate is significant and decreased compared with previous model (**Table 5**).

The residual distribution is not uniform around zero, but further dislocating from zero is not present compared with the model in section 4.1.3. The non-linear model fit and residuals are graphically represented in **Figure 13**.

Table 7 The activation coefficient (K) *estimate from the non-linear model supplied with the* ***GFP/OD₆₀₀*** *data for* ***n = 1*** *and statistical parameters: standard errors, t-value and p-value.*

	Estimate	Standard Error	t-value	p-value
K	0.16279	0.02523	6.452	6.97e-09

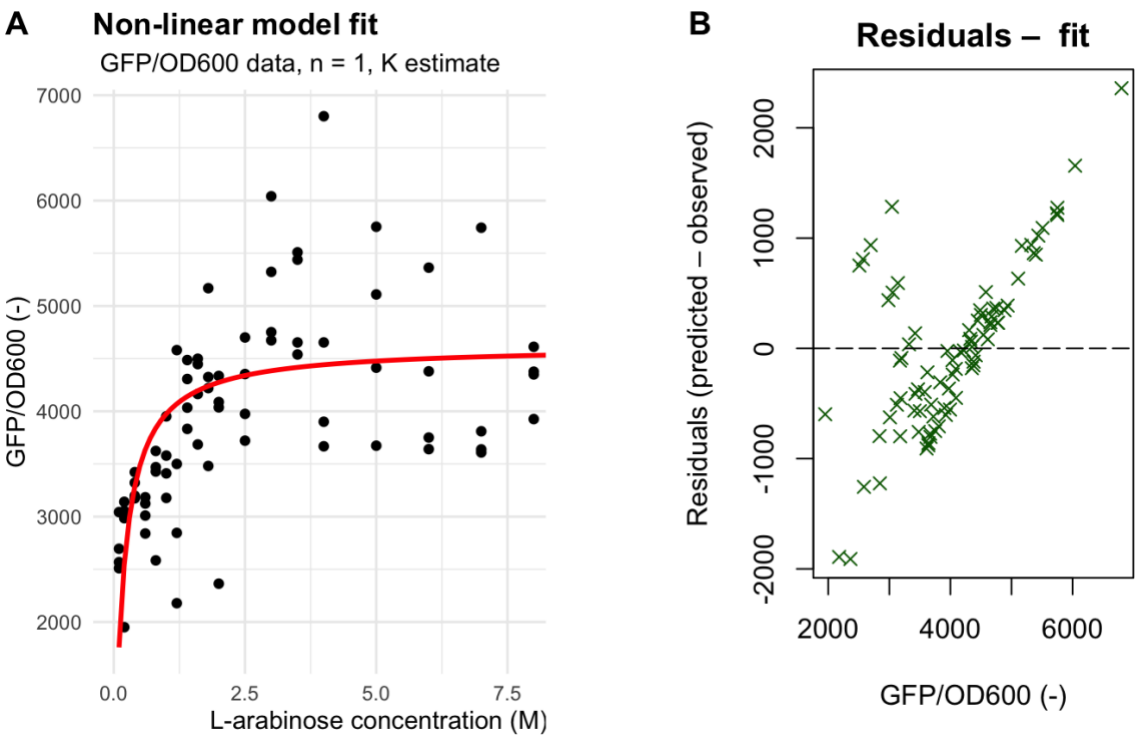


Figure 13 The non-linear model fit of the Hill function estimated with the GFP/OD₆₀₀ data *on the left (A) and residuals on the right (B). The model estimated value for the activation coefficient (**K**). The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for intermediate GFP/OD₆₀₀ signals.*

4.1.6 Parameter *K* estimate for GFP data predicted with a nondimensionalized non-linear model

In this simulation, a non-linear model for nondimensionalized Hill function from Equation 31 has been supplied with the GFP data to estimate *K* parameter. Analogous as in previous models, the GFP_{max} value was constant (**Table 3**). $ARAconc_{max}$ was selected to be 3 M and $n = 1$. The output of *nls*() function in R for current model is presented in **Table 8**. The results showed a significant p-value for *K* estimate equal the one from a dimensional analysis (**Table 6**). The graphical representation of the model fit in **Figure 14** displays unequal distribution of residuals around zero with similar trends as in foregoing model (chapter 4.1.5).

Table 8 The activation coefficient (K) estimate *from the nondimensionalized non-linear model supplied with the* ***GFP*** *data for* ***n = 1*** *with statistical parameters: standard errors, t-value, and p-value. The parameter K predicted by the simulation is unitless.*

	Estimate	Standard Error	t-value	p-value
K	0.06608	0.01032	6.404	8.61e-09

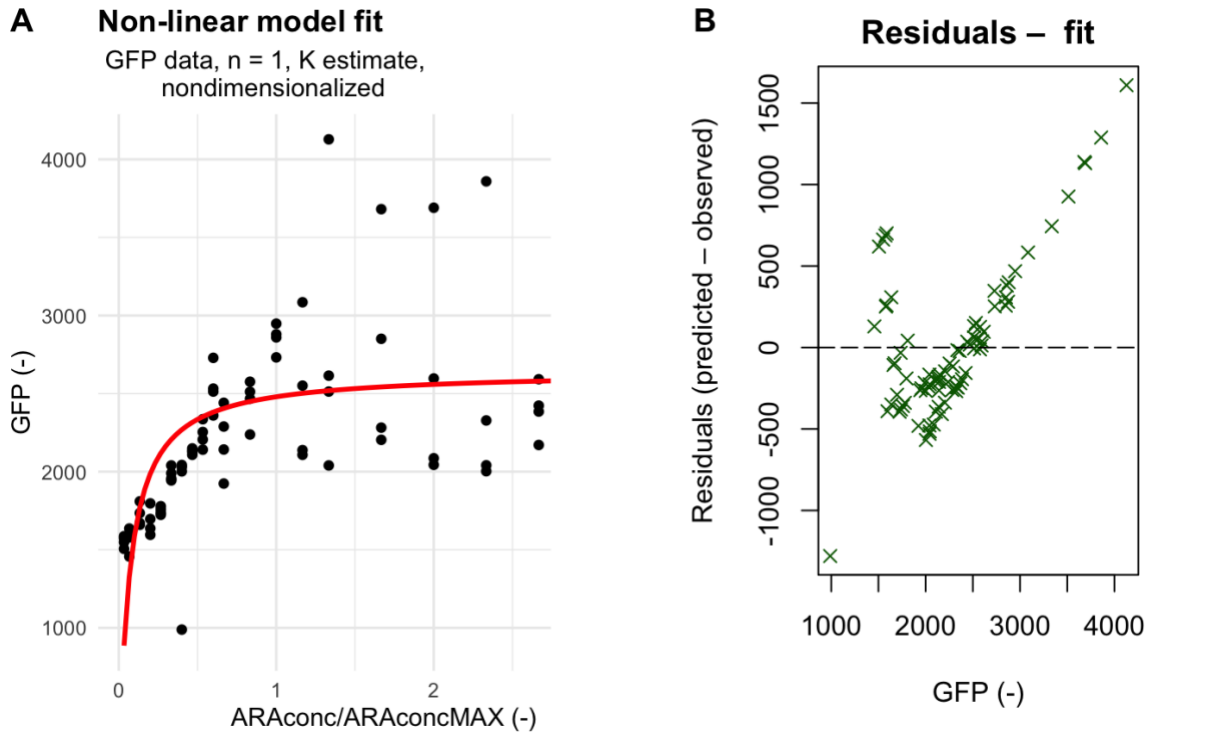


Figure 14 The non-linear model fit of the nondimensionalized Hill function estimated with the GFP data *on the left (A) and residuals on the right (B). The model estimated value for the activation coefficient (**K**) and the Hill coefficient was 1. The parameter K was considered unitless dimension and $ARAconc_{max}$ was 3 M. The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for intermediate GFP signals.*

4.1.7 Parameter *K* estimate with a nondimensionalized non-linear model for GFP/OD$_{600}$ data

The estimation of *K* parameter with the GFP/OD$_{600}$ data by the non-linear model for nondimensionalized Hill function included small changes of the Equation 31. Instead of GFP parameter from the Equation 31, it was used GFP/OD$_{600}$, while the GFP$_{max}$ value was replaced with the $\mathrm{GPF/OD}_{600\,\mathrm{max}}$(**Table 3**). The *n* value was equal to 1 and $ARAconc_{max} = 3\ M$. The data and parameters were incorporated in R (function *nls*()), which resulted in *K* parameter estimate presented in **Table 9**. A significant p-value for *K* estimate, the same as in a dimensional analysis for GFP/OD$_{600}$ data (**Table 7**), was predicted. The model fit given in **Figure 15** shows unequally distributed residuals around zero but similarly located as in the model from the section 4.1.4.

Table 9 The activation coefficient (K) estimate *from the* ***nondimensionalized*** *non-linear model supplied with the* ***GFP/OD$_{600}$*** *data for* ***n = 1*** *with statistical parameters: standard errors, t-value, and p-value. The parameter K from the simulation is considered unitless.*

	Estimate	Standard Error	t-value	p-value
K	0.05426	0.00841	6.452	6.97e-09

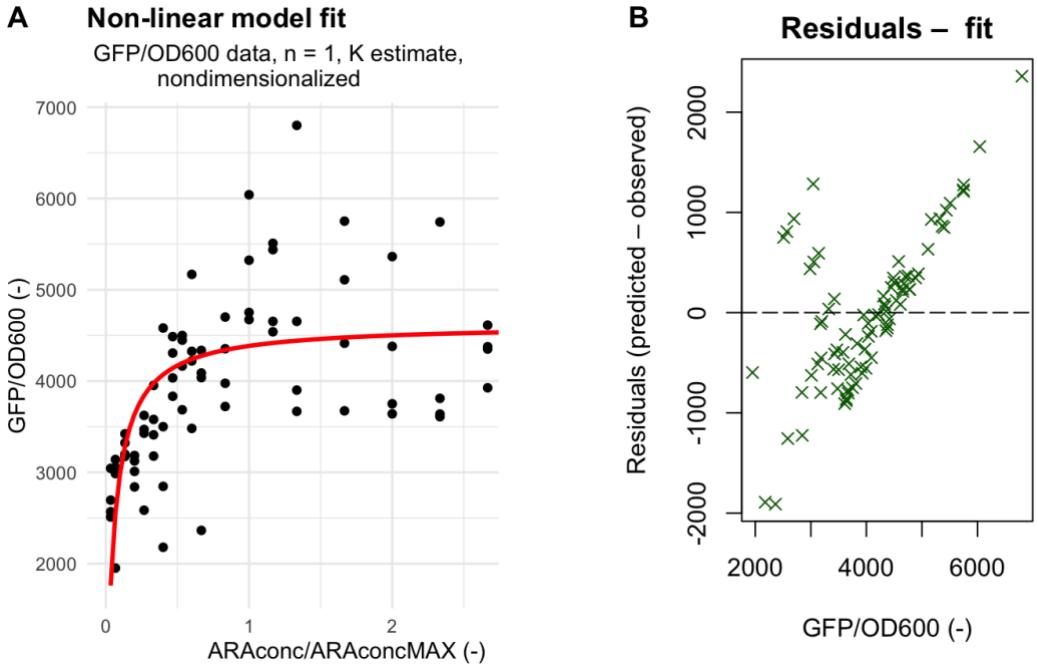


Figure 15 The non-linear model fit of the nondimensionalized Hill function estimated with the GFP/OD$_{600}$ data *on the left (A) and residuals on the right (B). The model estimated value for the activation coefficient (**K**) and the Hill coefficient was 1. The parameter K was considered unitless and $ARAconc_{max}$ was 3 M. The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for intermediate GFP/OD$_{600}$ signals.*

4.2 Parametrization of the Hill function with the rate of GFP signal change

The slope of linear part of the induction curve was calculated for every of 96 GFP-time curves by running a linear model. This was performed using the *all_easylinear*() function from 'growthrates' package in R. The estimated slope represented the rate of change of GFP signal over time.

A distribution of the corresponding rates to L-arabinose induction is presented in **Figure 16**. The effect of L-arabinose concentration on the cells from the induction study is considered as an increasing rate trend with increasing L-arabinose concertation until a saturation point. The curve itself has a hyperbolic or saturation shape.

A maximal rate (R) estimate from the induction assay with *E. coli* cells was further used to approximate coefficients for the models in following chapters. These parameters were applied for simulations of a dynamic response of L-arabinose activated gene expression across host-specific growth rates and for predictions of the genetic toggle switch performance. R was considered as a maximal estimated rate from all of the samples ($R = 0.3237\ hour^{-1}$).

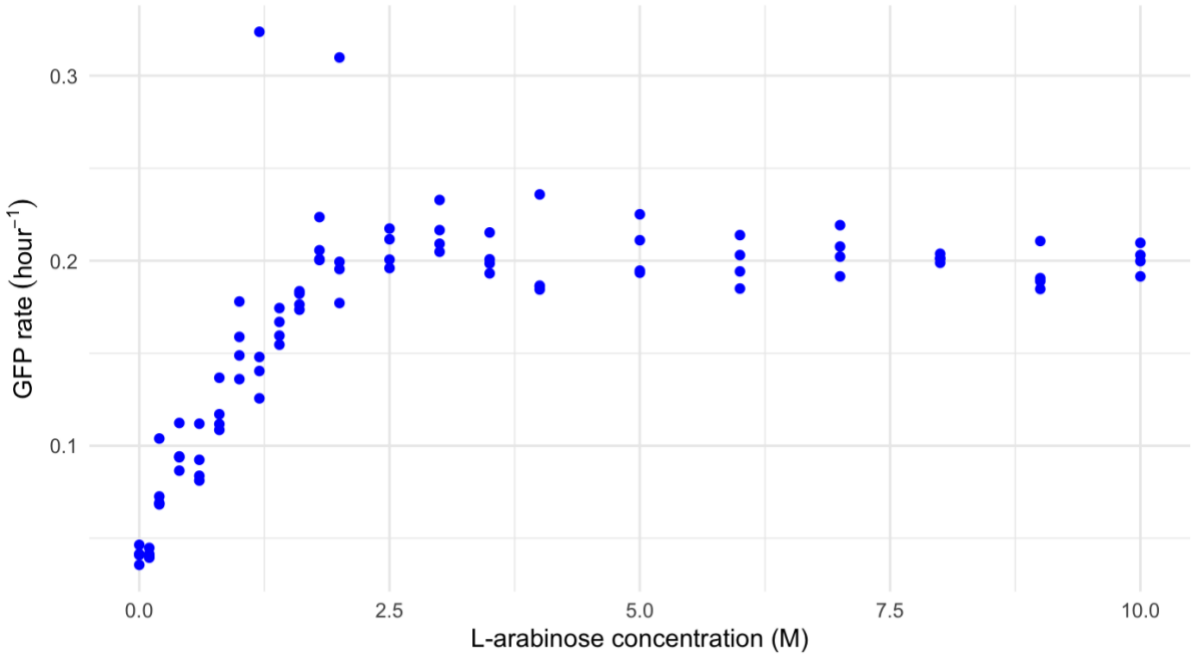


Figure 16 Estimated GFP rates for corresponding L-arabinose concentration. *The increase of L-arabinose concentration is followed by a higher resulting GFP rate estimate until saturation is reached.*

The effect of L-arabinose concentration ($ARAconc$) on cell growth was simulated by estimating a maximal specific growth rate (μ_{max}) for every sample. The μ_{max} was approximated as a slope of linear part of every OD_{600}-L-arabinose curve. The calculated μ_{max} values were fitted to a regression line from a linear model which resulted in parameters from **Table 10** and it is graphically presented in **Figure 17**. P-value for the slope is much greater than 0.05 ($p = 0.51$) which suggest that slope is not significantly different from zero, consequentially, the growth is not responsive to L-arabinose.

Table 10 The linear model output *with statistical parameters: standard errors, t-value, and p-value; correlated variables are concentration of L-arabinose and μ_{max}, the maximal specific growth rate.*

	Estimate	**Standard Error**	**t-value**	**p-value**
Intercept	0.2950	0.0054	54.514	< 2e-16
ARAconc	0.0008	0.0012	0.661	0.51

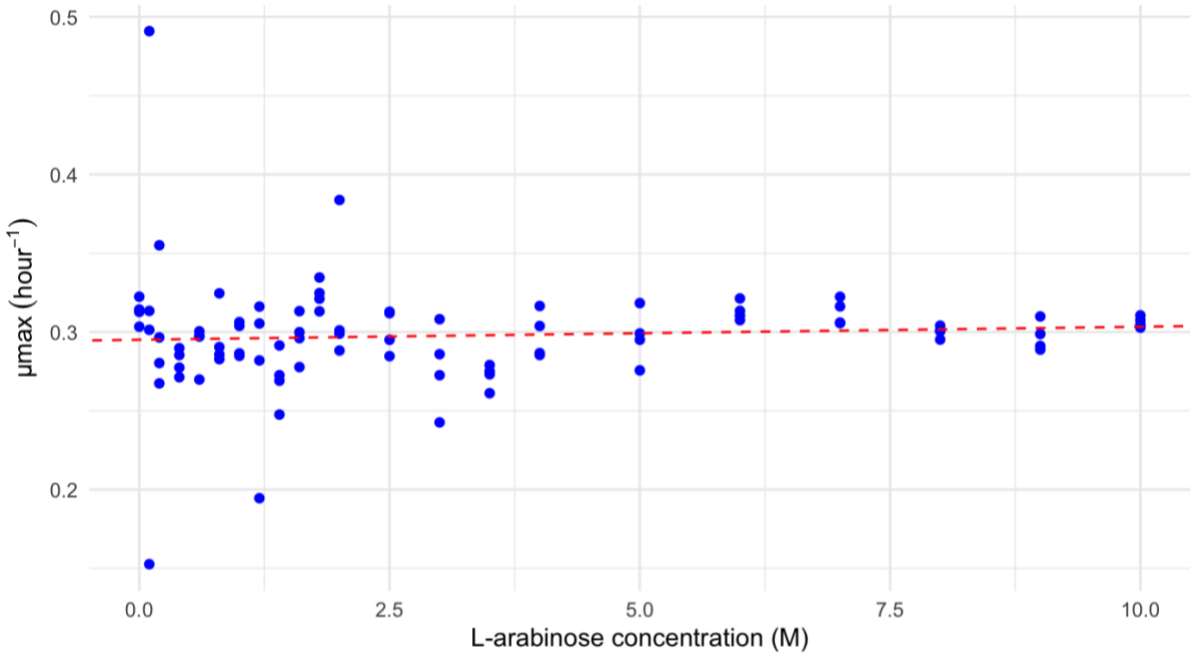


Figure 17 Graphical representation of the linear regression model *between L-arabinose concentration and μ_{max} shows that the cell growth is not responsive to L-arabinose induction. The model predicted that estimated maximal specific growth rates (μ_{max}) are not correlated with the L-arabinose concentration. A dashed red line represents a linear regression line and blue dots show estimated μ_{max} values. The slope was approximated to 0.0008 hour^{-1} and corresponding p-value was insignificant, which suggest that the slope is most likely to be zero and no correlation between μ_{max} and L-arabinose was predicted.*

As there was no effect of L-arabinose on the cell growth estimated (**Table 10**, **Figure 17**), it was unnecessary to consider GFP/OD_{600}-change-over-time (estimated as a rate of GFP signal normalized with OD_{600}). Hence, only the rates of GFP-change-over-time were taken into account in following non-linear models.

4.2.1 The effect of different *n*-values on non-linear model (GFP rates)

The calculated GFP rates were used to parametrize new non-linear models according to Equation 33. A simulation of the effect of *n*-values on non-linear models supplied with GFP rates data predicted better model fit for lower *n*-values (**Figure 18** and **19**). The better fit for $n = 1$ and $n = 2$ can be visually noticed by comparing different regression lines and how close data points are to the line. The model predicted a higher residual standard error (RSE) for elevated Hill coefficients as displayed in **Figure 18**.

Use of GFP rates over GFP fluorescence data resulted in better model fits, whereas the residuals were distributed closer around zero. This outcome can be visualized by comparing the GFP rate model results displayed in **Figure 18** and **19** to the GFP signal model results showed in **Figure 6** and **7**. The RSEs of the non-linear models with GFP rates data (**Figure 18**) were remarkably lower than in a simulation with simple GFP signal (**Figure 6**). Residual graphs given in **Figure 19** show similar form of deviation. The residuals of the models that include low *n*-values are distributed closer around zero than the ones estimated for higher *n*-values.

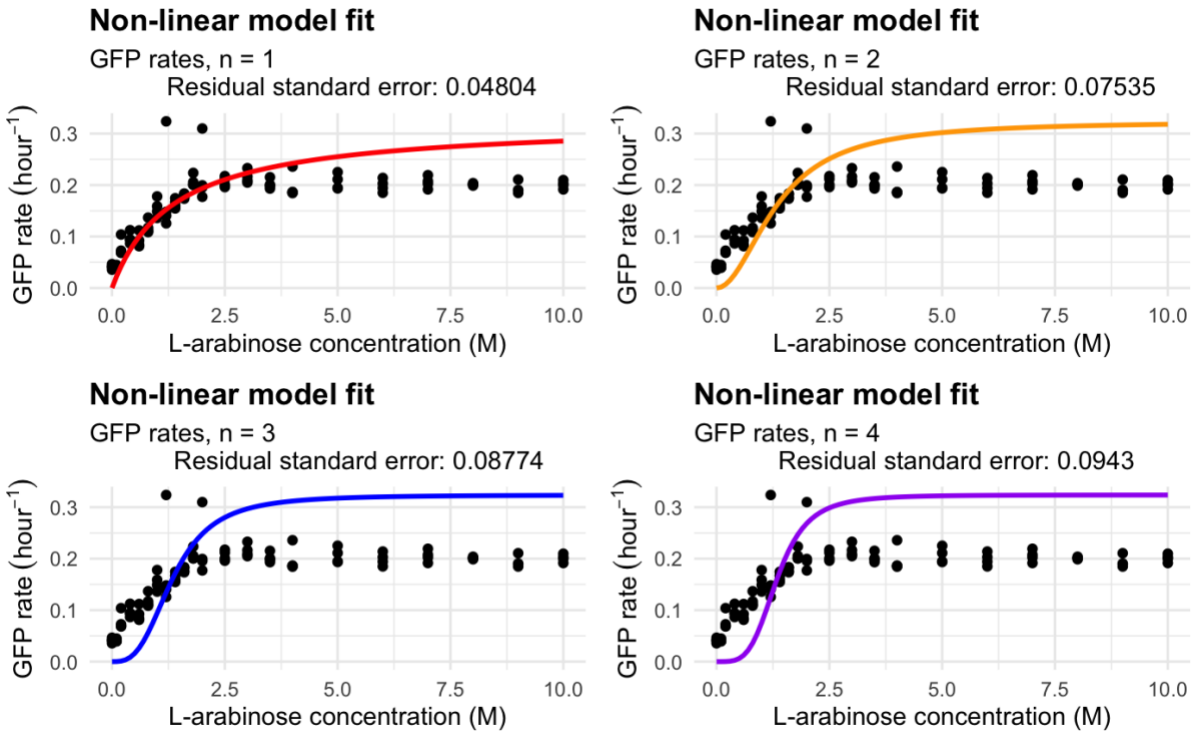


Figure 18 The effect of different Hill coefficients *is estimated as a poorer quality of a model fit and a lower model accuracy for elevated Hill coefficients (as confirmed by a higher residual standard error estimate). The comparison of the effect of n value to a fit of non-linear regression was simulated for GFP rates data. Estimated residual standard errors with GFP rates are significantly lower than with the simple GFP signal. Red regression line represents non-linear model with n = 1, orange regression line is for non-linear model with n = 2, blue regression line is non-linear model with n = 3 and purple regression line is for non-linear model including n = 4. Estimated residual standard errors for the models are shown in headings in a top part of the graphs.*

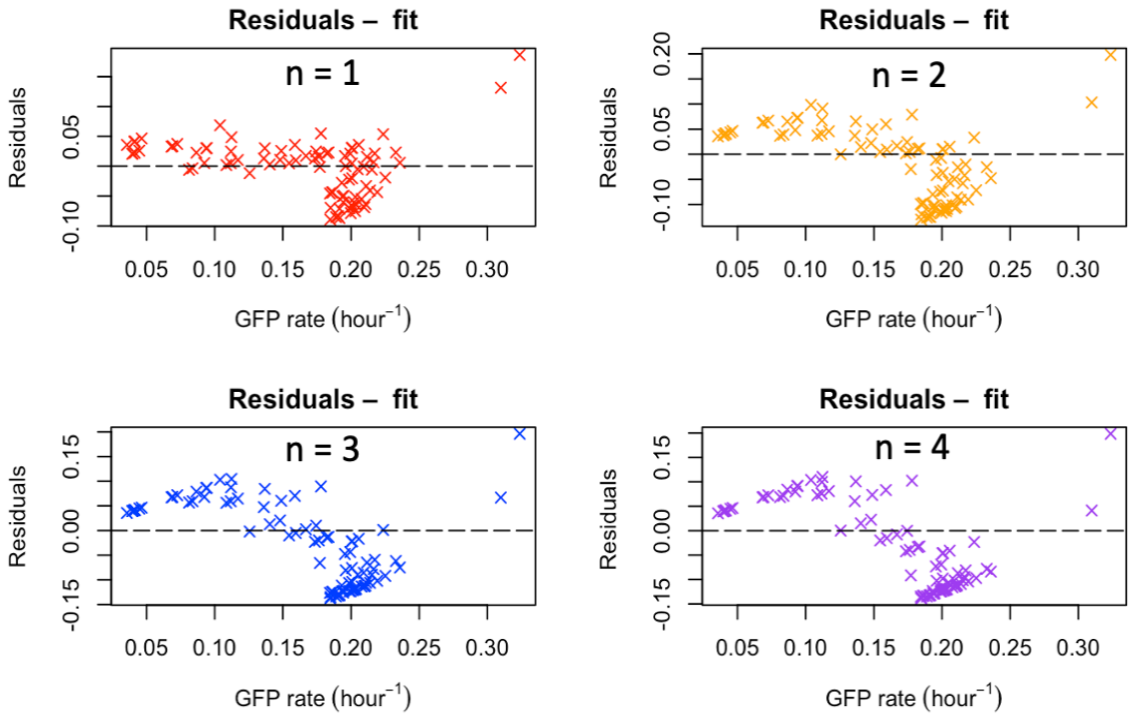


Figure 19 The effect of different Hill coefficients on residuals *is estimated as further distributed residuals from zero for higher n estimates in simulated non-linear models. Different colors are used to represent different n-values: red for n = 1, orange for n = 2, blue for n = 3 and purple for n = 4. Simulations of presented non-linear models were performed with the GFP rates data.*

4.2.2 Estimation of *n* and *K* parameters with GFP rates

The GFP rate data was incorporated in the model presented in the Equation 33 and the constants *n* and *K* were estimated. The maximal rate (*R*) value was approximated as the maximum of the GFP rates from the induction study (**Figure 16**). The estimated parameters from the non-linear model given in **Table 11** represent the output of the *nls*() function in R. P-values for both estimates were significant, and they were noticeably decreased than those from the previous model with the GFP fluorescence input (**Table 4**).

Table 11 The Hill (n) and activation coefficient (K) estimates *from the non-linear model supplied with the* ***GFP rates*** *data and statistical parameters: standard errors, t-value and p-value.*

	Estimate	Standard Error	t-value	p-value
n	0.48324	0.04619	10.46	< 2e-16
K	1.47833	0.15777	9.37	8.68e-15

A regression line from this model (**Figure 20**, red line) was predicted to lay closer to data points and residuals were distributed closer around zero (**Figure 20**) compared to the earlier model (**Figure 10**).

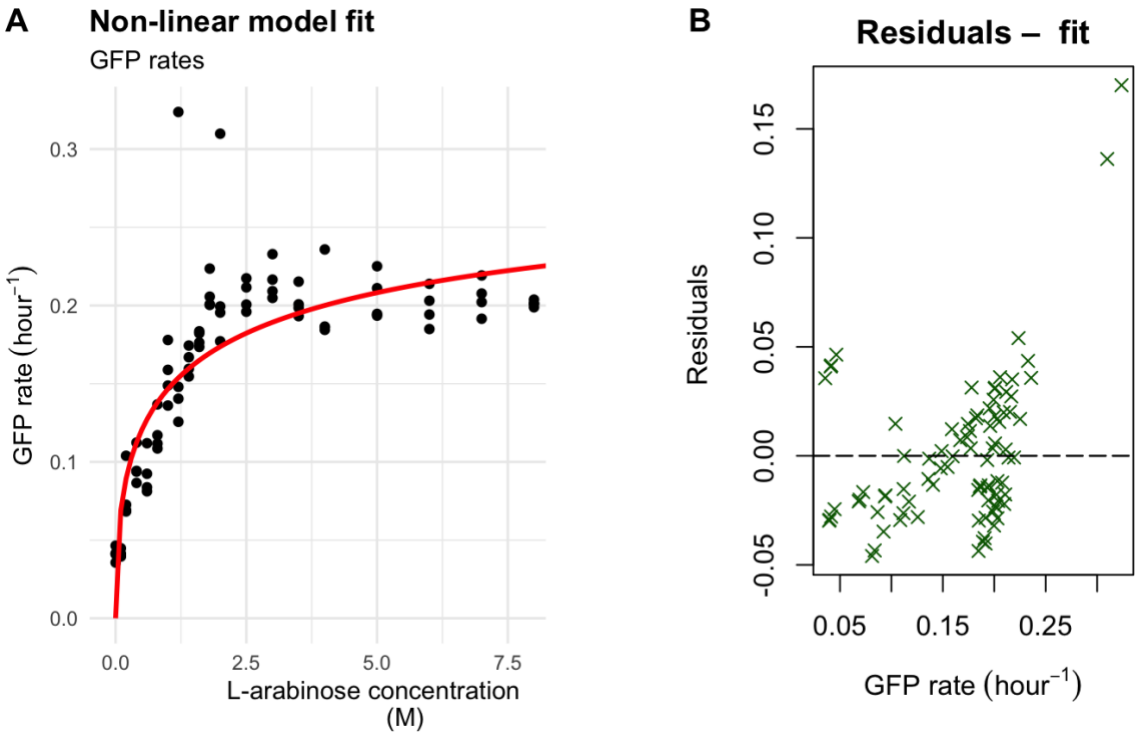


Figure 20 The non-linear model fit of the Hill function supplied with the GFP rates data *on the left (A) and residuals on the right (B). The model estimated value for the Hill coefficient (**n**) and activation coefficient (**K**). The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals compared with the model supplied with GFP data.*

4.2.3 Parameter *K* estimate for non-linear model with rate of GFP

The activation coefficient, *K*, (Equation 33) was estimated by a non-linear model supplied with the GFP rates data. The Hill coefficient, *n*, was 1. The output, including a *K* estimate and statistical parameters (**Table 12**), is a result of the *nls*() function in R. A p-value for the *K* estimate was significant. Compared to previous models (**Table 6**), that had the GFP and GFP/OD_{600} as inputs, a RSE of this model was extremely lower and the p-value was also decreased. The model fit and residuals are graphically represented in **Figure 21**.

Table 12 The activation coefficient (K) estimate *from the non-linear model supplied with the **GFP rates** data for **n = 1** and statistical parameters: standard errors, t-value and p-value.*

	Estimate	**Standard Error**	**t-value**	**p-value**
K	1.6285	0.1332	12.23	< 2e-16

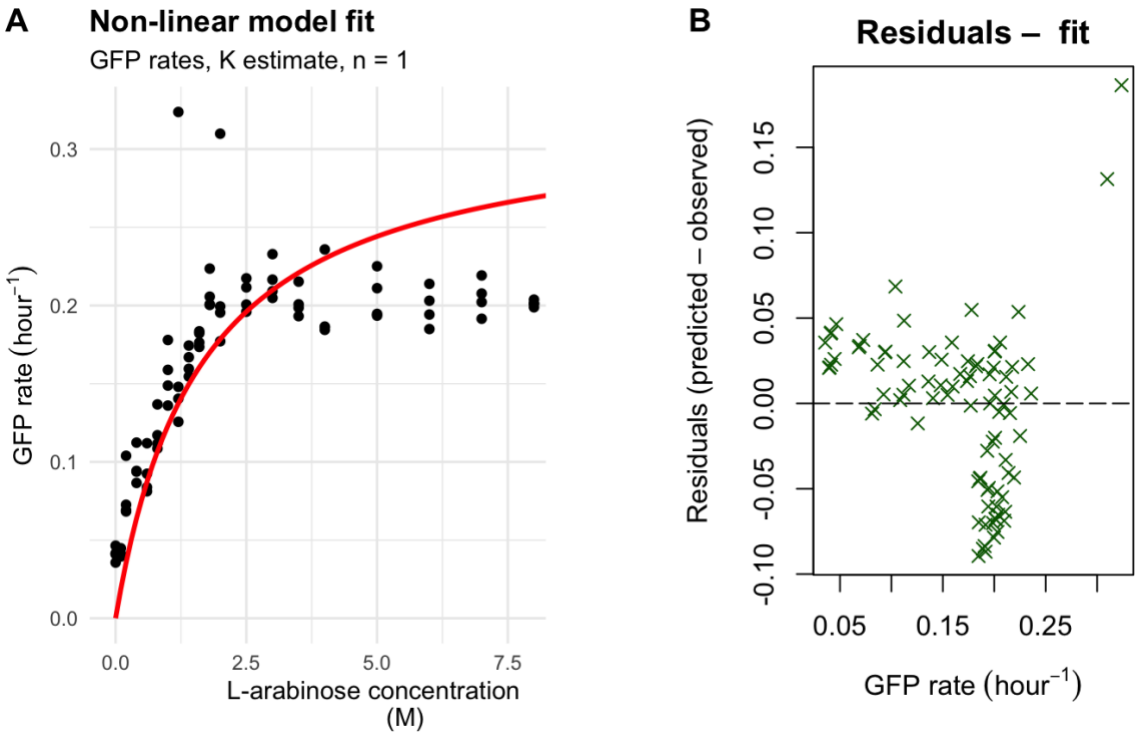


Figure 21 The non-linear model fit of the Hill function supplied with the GFP rates data *on the left (A) and residuals on the right (B). The model estimated value for activation coefficient (**K**). The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals compared with the model supplied with GFP data. A lower deviation of residuals is estimated for lower and intermediate GFP rate estimates.*

4.2.4 Parameter *K* estimate with nondimensionalized non-linear model for the GFP rates data

The nondimensionalized model presented in the Equation 34 has been used for estimation of K parameter. The n was considered as $n = 1$. Previously calculated GFP rates and Equation 34 were incorporated in the *nls*() function in R. The K estimate and statistical parameters from the *nls*() output are shown in **Table 13**. A significant and decreased p-value was predicted whereas the RSE ($RSE = 0.0480$) of this model was extremely reduced compared to previous model estimates (the GFP signal data) given in **Table 8**.

Table 13 The activation coefficient (K) estimate *from the **nondimensionalized** non-linear model supplied with the **GFP rates** data for **n = 1** with statistical parameters: standard errors, t-value, and p-value. The parameter K from the simulation is unitless.*

	Estimate	Standard Error	t-value	p-value
K	0.5428	0.0444	12.23	< 2e-16

The model fit is graphically displayed in **Figure 22** that shows a regression line of the model (red line) on the left and the residual graph on the right. Residuals are distributed closer around zero compared with previous model containing the simple GFP signal (**Figure 14**).

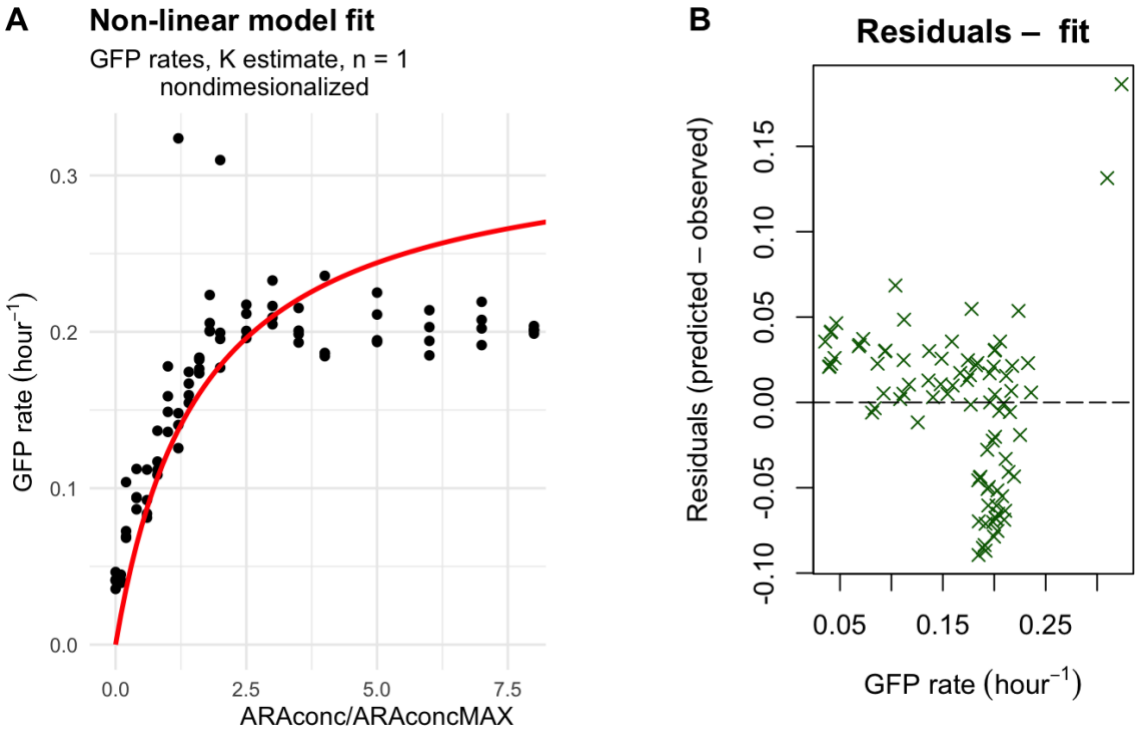


Figure 22 The non-linear model fit of the nondimensional Hill function estimated with the GFP rates data *on the left (A) and residuals on the right (B). The model estimated value for the activation coefficient (**K**) and the Hill coefficient was 1. The parameter K was considered unitless dimension and $ARAconc_{max}$ was 3 M. The regression line from the model is marked in red (A). The residual graph (B) shows uneven distribution of residuals around zero and a lower deviation of residuals is estimated for the lower and intermediate GFP rate estimates.*

Herein, we can conclude that the non-linear models with the GFP rates as inputs resulted in a better distribution of residuals, meaning that residuals were distributed closer around zero. Moreover, the models with GFP rates showed lower residual standard error and lower p-values compared to the models that have had GFP signal instead.

A better model fit was estimated with the rates data, resulting in more data points closely distributed around a regression line. The effect of different *n*-values had weaker predicted impact on overall quality of model with GFP rates.

4.3 Dynamic response of L-arabinose-activated gene expression across host-specific growth rates

The dynamic analysis was initially performed only for the cells of *E. coli*, as the maximal rate ($R = 0.3237\ hour^{-1}$) and the activation constant ($K = 1.6285\ M$, **Table 12**) were estimated through the non-linear models containing the induction study data collected for this bacterial strain. The activation coefficient was approximated through different non-linear models with the GFP, GFP/OD_{600} signal and GFP rates inputs from the induction assay performed with *E. coli*. The selected *K* parameter estimate for the dynamic response simulation was predicted by a non-linear model supplied with the GFP rates data. This model showed very low residual standard error ($RSE = 0.0480$) and high overall quality regarding the statistical parameters (**Table 12**) compared to all other models performed in this study.

The specific growth rate (μ) for *E. coli* was previously experimentally estimated to be μ = 0.0185 $hour^{-1}$. The Hill coefficient was assumed as 1. A prediction of dynamic response of the one side of the switch was considered in the concentration range from 0.1 M to 10 M of L-arabinose ($ARAconc$).

A dynamic response of L-arabinose activated gene expression was simulated by the Equation 35 for the one side of the toggle switch. In this simulation, the activity of the P_{BAD} promoter that controls expression of the *tetR* gene and a transcriptionally fused GFP as the reporter protein (signal) was modeled (**Figure 1**).

Equation 35 was solved with the *ode*() function that is part of the 'deSolve' package in R, and resulted in **Figure 23** presenting the change of GFP signal over time as a result of cell induction with 0.1 (red), 1 (orange), 5 (blue) and 10 M (purple) of L-arabinose. The response of gene expression in *E. coli* showed saturation trend of GFP signal change for every examined inducer concentration. The higher GFP response signals during time were estimated for the elevated $ARAconc$ values.

The result of performed simulation in the selected range of $ARAconc$ given in **Figure 23** was that induced cells reach the saturation point after a certain time regardless the initial inducer concentration. The effect of different inducer concentrations affected the corresponding GFP signal intensity for the saturation state. The saturation point was estimated to be reached at higher GFP signals in the cases when the cells were induced with raised $ARAconc$. Estimated dynamic response curves displayed in **Figure 23** have similar shape and trend, these include

initial uprise of GFP signal followed with further increase of signal and final stagnation of signal meaning that the saturation is achieved.

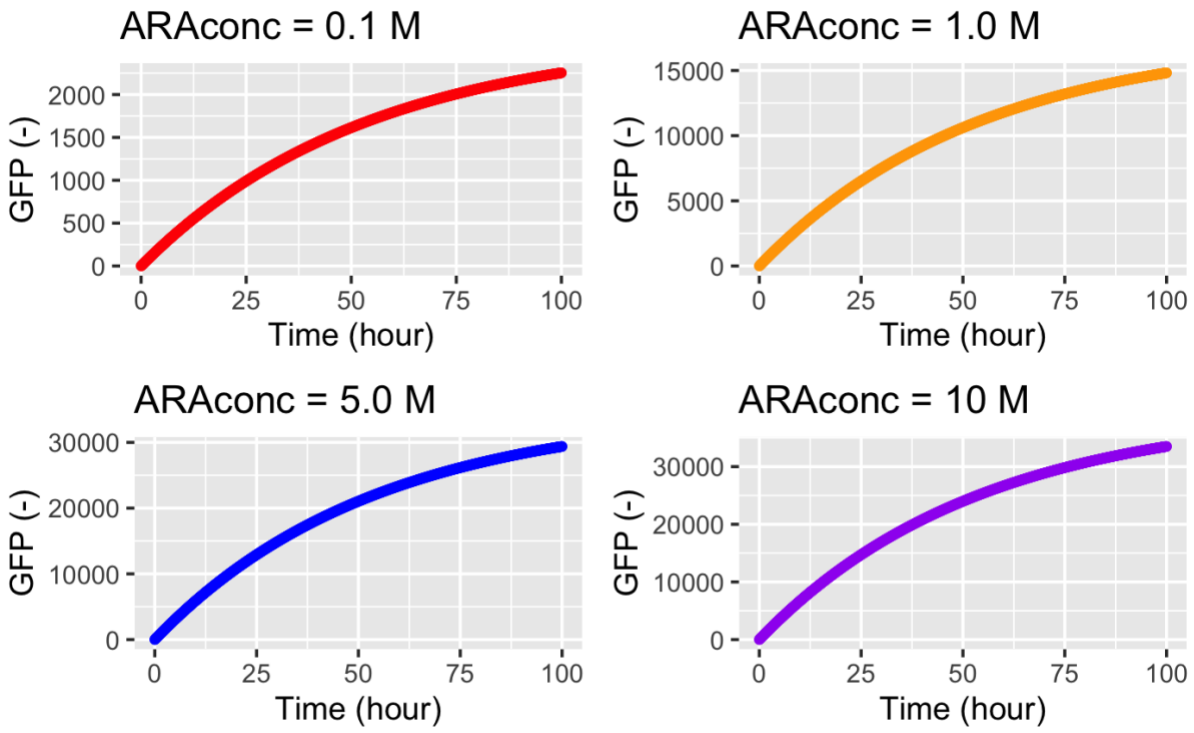


Figure 23 The dynamic response simulation of the one side of the toggle switch in E. coli *shows saturation trend of GFP signal over time regardless the inducer concentration (L-arabinose). The higher GFP response signals were estimated for induction with the higher concentrations of L-arabinose. Red line represents simulated dynamic response for* $ARAconc$ *= 0.1 M, orange line shows response with* $ARAconc$ *= 1 M, blue line is with* $ARAconc$ *= 5 M, and purple line is for induction with 10 M of L-arabinose.*

Under the assumption that the maximal rate ($R = 0.3237\ hour^{-1}$) and estimated activation coefficient ($K = 1.6285\ M$) approximated for *E. coli* are equivalent for all bacterial strains, it was possible to predict a dynamic response of L-arabinose-activated gene expression across host-specific growth rates that resulted in **Figure 24**. The assumptions were made to ensure that $f(ARA)$ from the Equation 37 is equal in every model and that precited performance of model is only under effect of different microbial growth rates (μ).

The incorporation of different μ-values resulted in a different sink term being considered within the model which led to a change in the response curve appearance. The initial inducer concentration for the performed analysis was 1 M of L-arabinose ($ARAconc = 1\ M$).

The primary result of the simulation indicated that every curve has a saturation tendency which includes initial uprise of GFP signal until the saturation point where it remains constant and under effect of dilution after some time it begins to decrease (**Figure 24**). For the simulated time frame (1 - 100 hours) this effect can be noticed completely for the dynamic response

simulated for *V. natriegens* (μ = 0.0336 hour^{-1}) that is presented by a pink line. The lowest growth rate, μ, was approximated for the cells of *P. oceani* (μ = 0.0052 hour^{-1}), consequentially the response simulation for chosen time frame did not indicate reaching a saturation (red line).

To question if all simulated curves achieve the saturation, the simulation of a dynamic response of bacterial strain with the lowest estimated μ, *P. oceani*, was modeled for a longer period of time (t = 1-1000 hours). The performed analysis predicted a saturation point approximately 750 hours after induction with 1 M L-arabinose. The GFP signal change over time in *P. oceani* was approximated with a dynamic saturation curve (red line) that is presented in the top left corner of **Figure 24**.

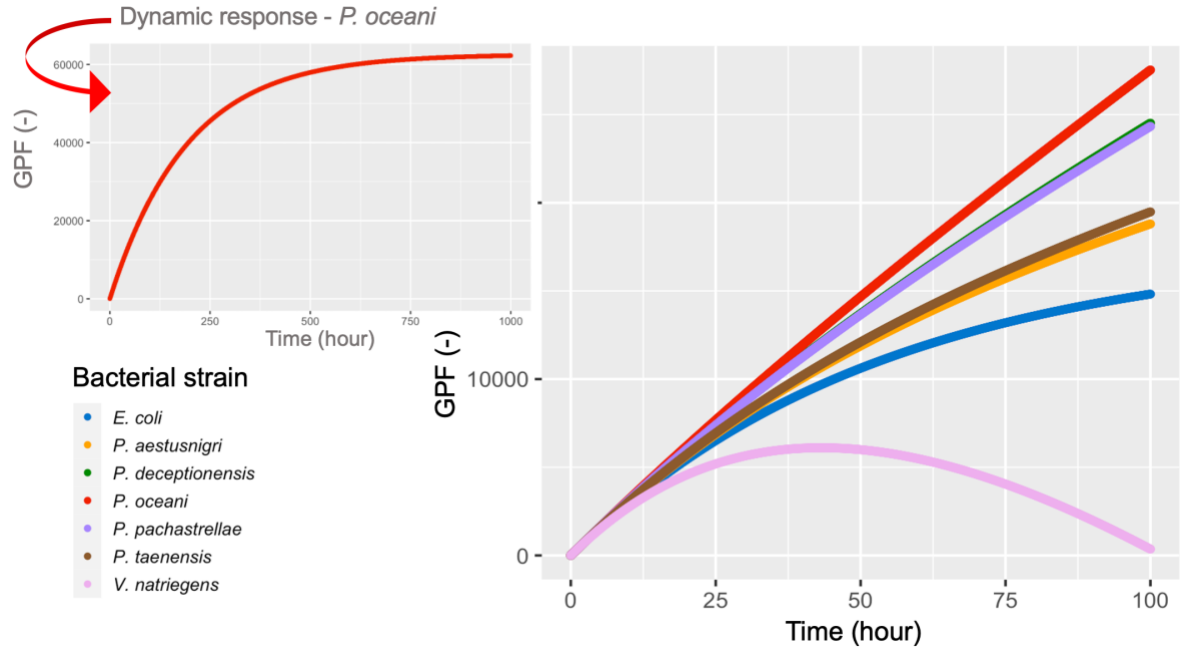


Figure 24 The dynamic response simulation *of L-arabinose-activated expression across host-specific growth rates shows a saturation tendency including initial uprise of GFP signal, saturation and final decrease under the dilution effect. GFP signal is presented on Y-axis, and it is unitless. Time is estimated in hours and presented on X-axis. Approximated response for E. coli is represented with blue line, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color. The dynamic response simulation of L-arabinose-activated expression in P. oceani is presented in the top left corner and shows saturation tendency including initial uprise of GFP signal followed with later saturation that is achieved approximately 750 hours after induction with 1 M L-arabinose.*

Simulated dynamic curves show almost identical tendency for the bacterial strains with similar growth rate estimates, therefore the lines that represent response of *P. deceptionensis* (green line), μ = 0.0084 hour^{-1}, and *P. pachastrellae* (purple line), μ = 0.0086 hour^{-1}, are nearly indistinguishable in **Figure 24**. This is also supported with similar estimated responses for *P. aestusnigri* (orange curve), μ = 0.0144 hour^{-1}, and *P. taenensis* (brown curve), μ = 0.0136 hour^{-1}.

A response time ($T_{1/2}$) was calculated according to Equation 39, and distribution of $T_{1/2}$ among selected bacterial strains is displayed in **Figure 25**. The response time was considered as a time to reach half of the steady-state protein concentration. The $T_{1/2}$ was assumed to be governed only by the removal rate hidden in the dilution term formulated as $\alpha = \alpha_{dil} = \mu$. The fastest response was estimated from *V. natriegens,* and the slowest from *P. oceani*.

The result of $T_{1/2}$ prediction was that a higher response time was approximated for a slower estimated growth, and opposite, a lower response time was estimated for an elevated μ-value. Therefore, predicted $T_{1/2}$ for *V. natriegens*, bacterial strain with the fastest predicted growth rate (μ = 0.0336 hour^{-1}), was 20.61 hour (marked in pink in **Figure 25**). The highest response time, $T_{1/2}$ = 132.2 hour (marked in red), was approximated for *P. oceani* as a result of estimated μ that was 0.0052 hour^{-1}, the lowest predicted growth rate among all selected strains.

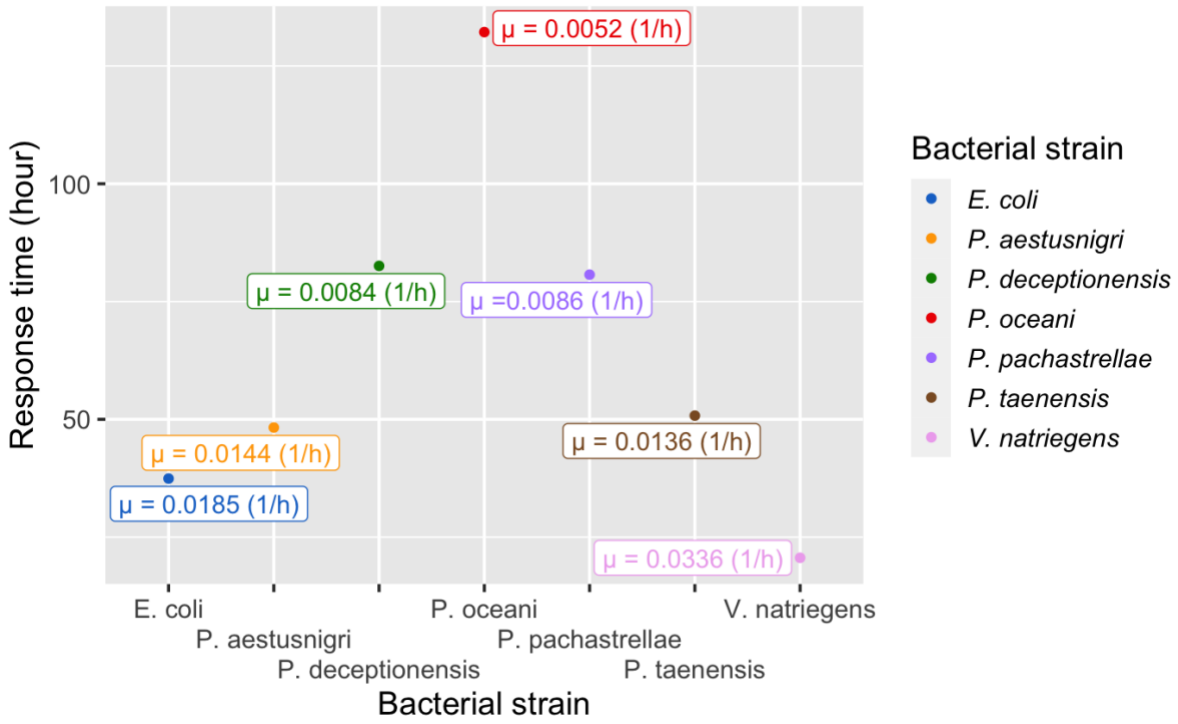


Figure 25 Distribution of the estimated response times among bacterial strains *is governed by the species-specific growth rate (presented in rectangles). The fastest response was estimated from V. natriegens, and the slowest from P. oceani. Approximated response time was higher for the bacterial strains that show lower growth abilities (lower μ estimates). The estimated $T_{1/2}$ and μ-values (in rectangles) are represented in blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

A graphical representation of specific growth rate (μ) distribution for calculated response times ($T_{1/2}$) is given in **Figure 26**. The higher the μ value, the faster response, and a lower approximated response time. The estimated $T_{1/2}$ were predicted for the μ-range that differed almost in order of magnitude between the lowest and highest μ estimate.

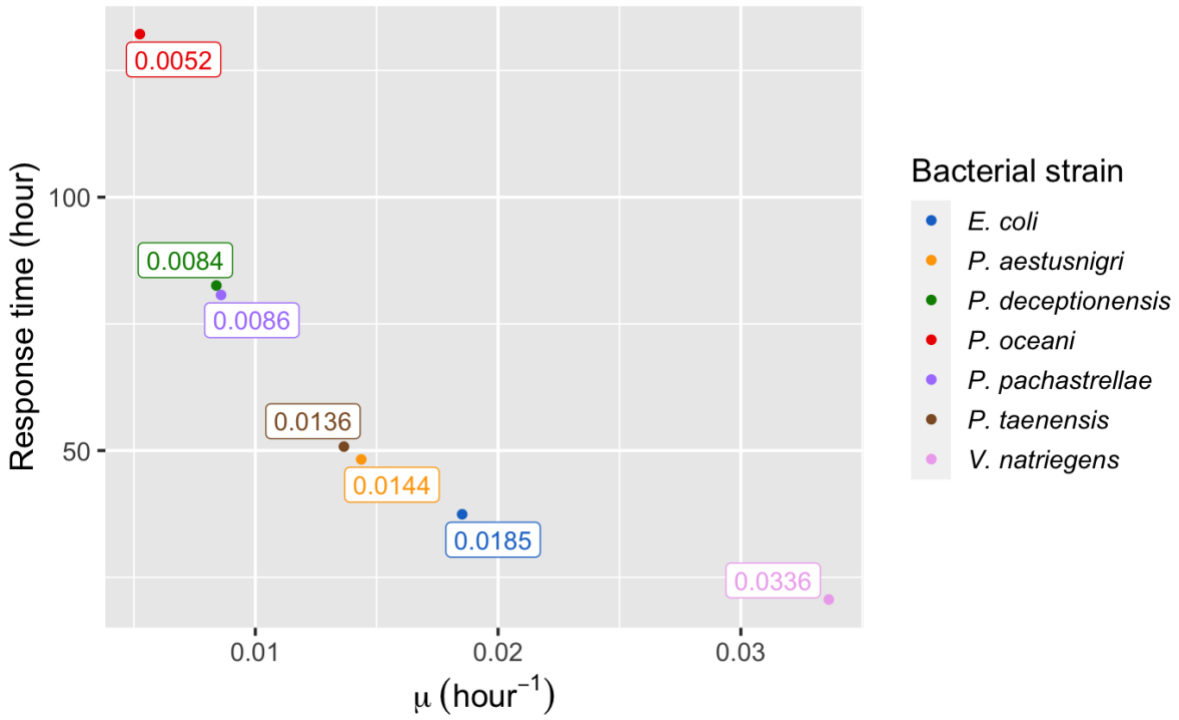


Figure 26 Distribution of the estimated response times among the growth rates (μ) *show the fastest response from V. natriegens, and the slowest from P. oceani. Approximated response time was higher for the bacterial strains that show lower growth abilities (lower μ estimates). The estimated $T_{1/2}$ and μ-values (in rectangles) are represented in blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

4.4 Simulation of a toggle switch performance

In this chapter, a synthetically build toggle switch was modeled as a dynamic system. This network was described by a system of ODEs presented with Equation 40 and 41 for dimensional analysis and in Equation 44 for nondimensional analysis. The solution of the ODE system was estimated by a function *ode*() from package 'deSolve' in R.

It was assumed that activation coefficients K_{AraC} and K_{TetR} are the same for every examined strain and that K_{AraC} is equal to K_{TetR}. The transcriptional factors have very different modes of action, and it is not expected to obtain equal activation coefficient values for different promoters (P_{BAD} and P_{Tet}), neither for the different hosts. The assumptions were made exceptionally for the simulation purposes in order to predict the behavior of the simplified model.

In the previous analysis (chapter 4.1 and 4.2) the activation coefficient (K_{AraC}) was estimated through different non-linear models supplied with the GFP, GFP/OD_{600} signal and GFP rates collected from the induction assay performed with *E. coli*. The selected *K* parameter for the simulation of a toggle switch performance was derived from the model with the GFP rates data ($K = 1.6285\,M$). A very low residual standard error ($RSE = 0.0480$), excellent quality regarding the statistical parameters (**Table 12**) and predicted non-linear fit together with distribution of residuals (**Figure 21**), supported selection of the *K* estimate from this model in further analysis.

The simulation was performed under assumption that the Hill coefficient (*n*) is 1 and that the maximal expression level of promoter was obtained ($\beta = 1$). The rate of degradation was considered negligible $\alpha_{deg} \approx 0$ with the intention to approximate the sink term (Equation 27) as equal to $\alpha = \alpha_{dil} = \mu$.

The performance of the toggle switch (**Figure 1**) was approximated by a model that simulated behavior of the device through the approximations of the repressors concentrations (AraC, TetR) that were considered as results of the promoter activity. The initial conditions for the simulation included that none of the repressors were present. Previously determined model parameters ensured approximation of the effect of the species-specific growth rates on a device behavior which resulted in curves given in **Figure 27** (2D) and **28** (3D). By observing a time variable in the results graph, it is possible to visualize that in every case response of the device was predicted, but it varied in favor of production of AraC or TetR as displayed in **Figure 28**.

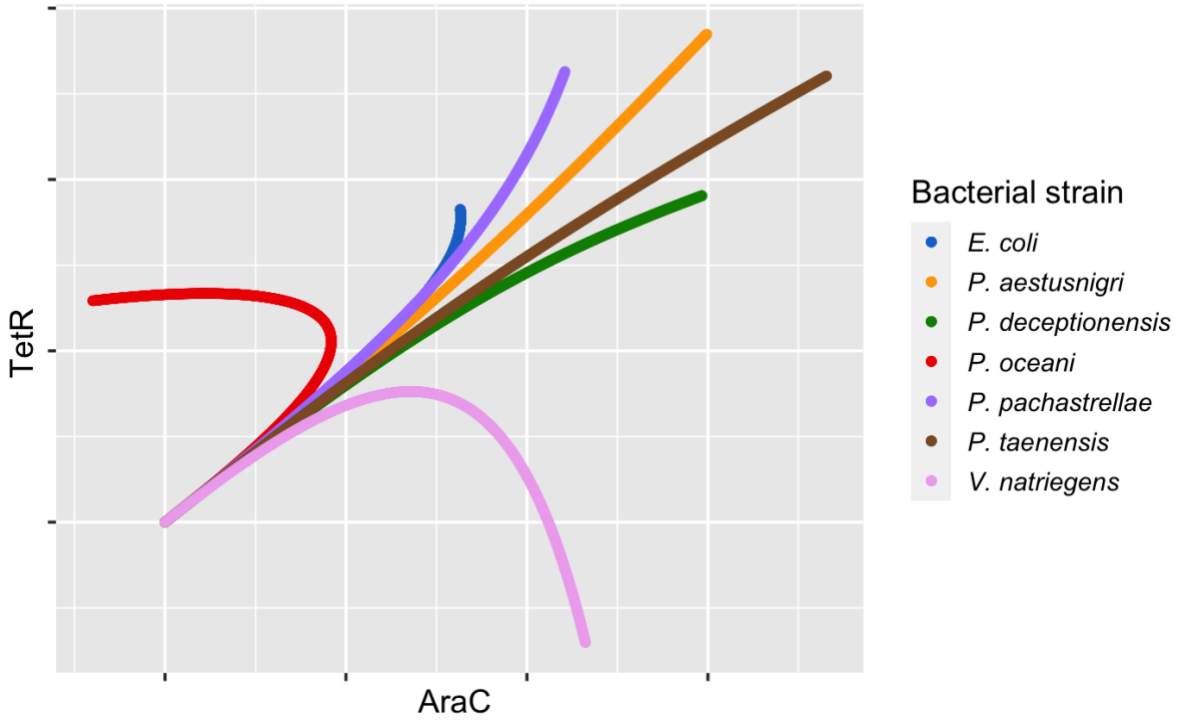


Figure 27 Simulated performance of the toggle switch in different species *displays the effect of the species-specific microbial growth rate on expression ability of bacterial strain that is approximated with the change of TetR and AraC. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that* $K_{AraC} = K_{TetR} = K$*. Production of AraC and TetR, proteins that are synthetized by the toggle switch, is presented on X and Y axis respectively and the dimensions are considered arbitrary. Bacterial strains are marked in different colors: blue for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

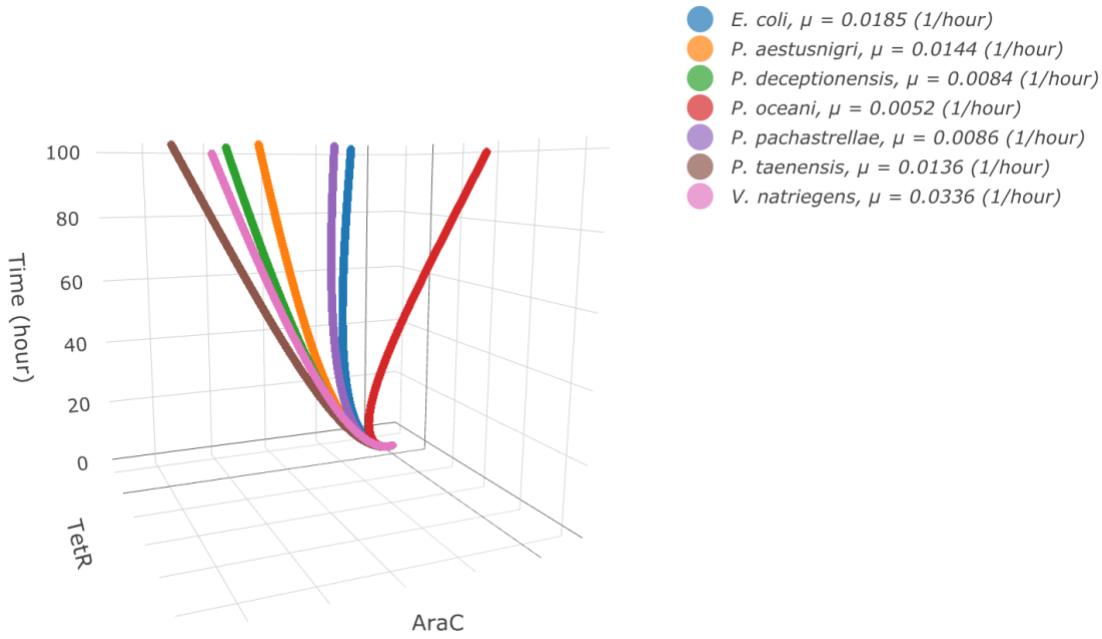


Figure 28 Simulated performance of the toggle switch in different hosts *approximates the effect of species-specific growth rate on the expression ability of bacterial strain by estimating the production of AraC and TetR. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that* $K_{AraC} = K_{TetR} = K$*. Production of AraC and TetR, proteins that are synthetized by the toggle switch, is presented on X and Y axis respectively and the dimensions are considered arbitrary. The Z-axis shows time. Bacterial strains are marked in different colors: blue for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

The result of simulated performance of the device in *P. pachastreallae*, *P. deceptionensis, P. aestusnigri* and *P. taenensis* was recognized as a continuous increase of AraC and TetR concentration. As an outcome of the toggle switch behavior in *V. natriegens*, an initial uprise of AraC and TetR followed with further increase of AraC was predicted, on the other hand, the simulated trend of the TetR concentration was decreasing. The model presented initial synthesis of both repressors in *E. coli* and a further production of TetR, in contrast, AraC was consumed. The simulation of the toggle switch behavior in *P. oceani* showed an initial TetR and AraC production and later AraC consumption, whereas TetR concentration was estimated almost constant during the simulation time.

Another approach for the simulation of the toggle switch behavior across different bacteria included a nondimensional analysis that relies on the Equation 45. Presented system of ODEs was implemented through R function *ode*() under the identical assumptions as beforehand. This analysis allows modeling the dynamic system in much more convenient way. The activation coefficient (*K*) used in this model was estimated in chapter 4.2.3 (**Table 12**). The activity of the promoter is represented by unitless quantities, such as *u* and *v*, that have a concentration of synthetized protein incorporated in them as Equation 44 shows.

The primary result of the non-dimensional model was that the activity of the toggle switch was achieved in every host. This was approximated as a change of indirect measures, *u* and *v* (**Figure 29** and **30**). Due to the limitation of presenting the results in two dimensions, such as not visible simulated behavior for *E. coli*, the time variable needs to be considered in analysis (**Figure 30**). The unitless time variable in the 3D graph is a respective time, τ, estimated as product of a time and the specific growth rate, $\tau = t\mu$. The activity of the toggle switch was predicted by a model with a respect to *u* and *v*, regardless the change of μ.

The simulated performance among different hosts presented in **Figure 29** and **30** is a direct result of the growth rate that affects expression ability of the bacterial strain. The higher the μ is, the more rapid cell division is estimated. This results in higher dilution rates of proteins (TetR and AraC) and plasmid DNA per cell in the fast-growing hosts. The predicted decrease in protein concentration (TetR and AraC) in dimensional analysis, and analogously decrease of *u* and *v* in non-dimensional model were the results of the estimated dilution effect approximated by the specific growth rate. The chassis effect was predicted among selected bacterial strains, and it was simulated in the model environment which ensured that the parameters for the each host were kept constant and equal, except the growth rate.

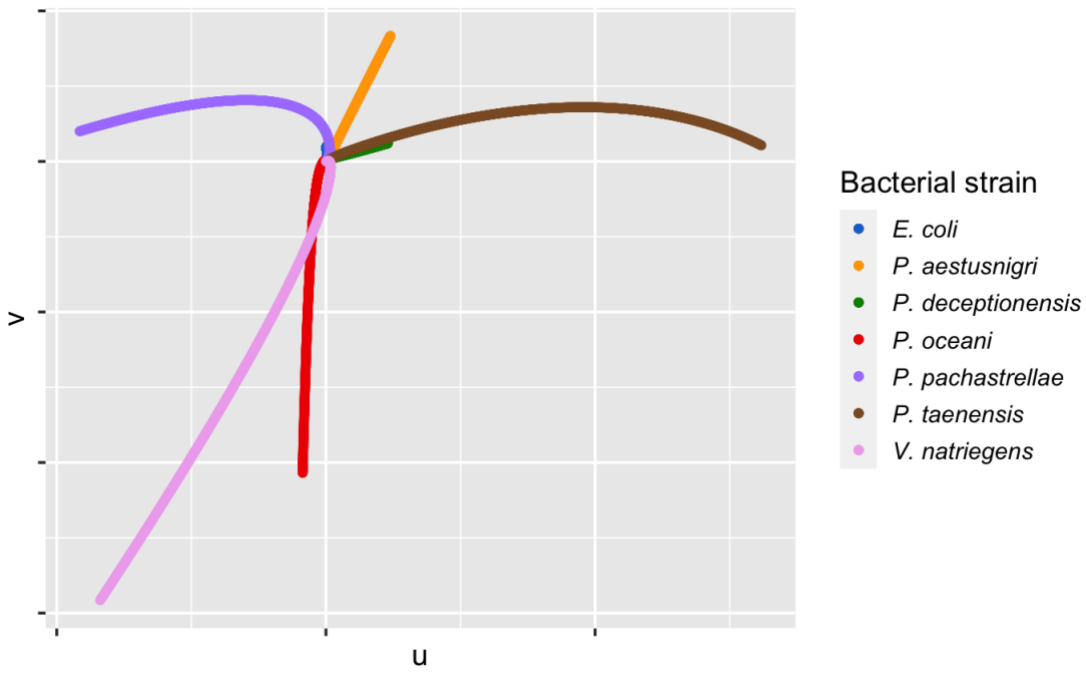


Figure 29 Simulated performance of the toggle switch in different species with a nondimensional analysis *shows the effect of the species-specific microbial growth rate on expression ability of bacterial strain approximated with u and v change. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that* K_{AraC} = K_{TetR} = K. u ($u = C_{AraC}/K_{AraC}$) *and* v ($v = C_{TetR}/K_{TetR}$) *are quantities that represent activity of the toggle switch, and they are positioned on X and Y axis respectively and their dimensions are considered arbitrary. Bacterial strains are marked in different colors: blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

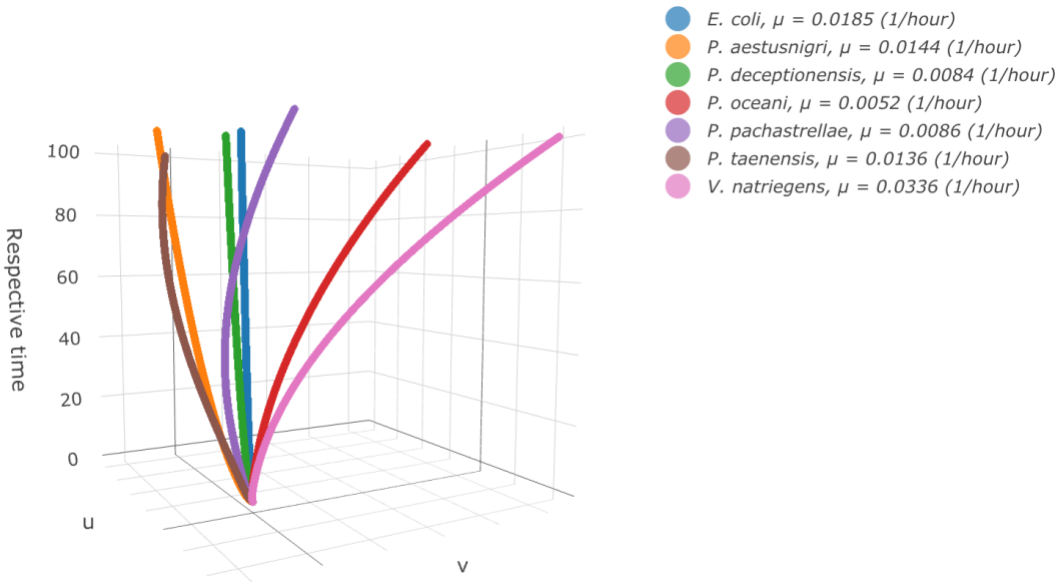


Figure 30 Simulated performance of the toggle switch in different hosts by a nondimensional analysis *in different species is under the effect of species-specific growth rate. The resulting change in expression ability of bacterial strain is estimated by the change of* u *and* v. *The predicted behavior was modeled for n = 1, K = 1.6285 M, and* $\beta = 1$, *that the rate of degradation is negligible and that* K_{AraC} = K_{TetR} = K. u ($u = C_{AraC}/K_{AraC}$) *and* v ($v = C_{TetR}/K_{TetR}$) *are quantities that represent activity of the toggle switch, and they are positioned on X and Y axis respectively and their dimensions are considered arbitrary. The Z-axis shows respective time,* τ, *estimated as product of time and specific growth rate,* $\tau = t\mu$. *Bacterial strains are marked in different colors: blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

The result of nondimensional simulation was that changes of *u* and *v* with respective time were estimated as uprise of both quantities in *P. taenensis*, *P. aestusnigri* and *P. deceptionensis*. The approximation of the *u* and *v* change in *P. pachastrellae* and *V. natriegens* indicated initial increase of measures, that was followed with a constant decrease of *u* and *v*. The decreasing trend of both quantities is present in *P. oceani.* The simulation performed for *E. coli* as a host showed no change of *u* and *v*.

The microbial chassis effect was estimated in dimensional (**Figure 27** and **28**) and nondimensional (**Figure 29** and **30)** analysis resulting in a different simulated actions of the modeled genetic toggle switches across hosts which were the consequence of the different species-specific growth rates considered by a model. The simulation resulted in unique behavior of every bacterial strain that can be presented by a different trend of the result function of ODE system that was modeled (**Figure 28 - 30**).

4.4.1 Simulation of a toggle switch performance for initial aTc presence

The preceding simulation included initial conditions which assumed absence of both inducers, anhydrotetracycline (aTc) and L-arabinose. As supposed for a toggle switch circuit (**Figure 1**), in presence of excess aTc, all cells are in a red-mKate-dominant (AraC) state, while the cells switch to a green-GFP-dominant (TetR) state in the presence of L-arabinose. To ensure that one of the sides of the device is repressed and one is activated, the initial conditions were changed to estimate AraC-dominant state by simulating the aTc induction.

The same assumptions were implemented as before: the activation coefficients were estimated as equal and derived from **Table 12**, $n = 1$, the maximal expression level of promoter was assumed ($\beta = 1$) and the rate of degradation was considered negligible $\alpha_{deg} \approx 0$. The change of AraC and TetR predicted by the model is given in **Figure 31** (2D). The three-dimensional representation of the performed analysis is displayed in **Figure 32**.

The activity of the P_{BAD} promoter was present in every strain and the difference in intensity of the promoter expression in different hosts was perceived. The chassis effect was simulated in selected conditions and clearer visual description of this effect may be given with a respect to time variable as it is shown in **Figure 32** (dimensional analysis) and **Figure 34** (nondimensional analysis).

The protein production was considered by the Hill term $\left(\frac{\beta_{AraC}R}{1+\left(\frac{C_{TetR}}{K_{TetR}}\right)^{n_{AraC}}};\ \frac{\beta_{TetR}R}{1+\left(\frac{C_{AraC}}{K_{AraC}}\right)^{n_{TetR}}}\right)$ and the dilution ($\alpha = \alpha_{dil} = \mu$) was modeled with the sink term (μC_{AraC} and μC_{TetR}) as formulated in Equation 40 and 41. The effect of μ in performed simulations was estimated through the approximation that the more rapid cell division, the higher estimated protein dilution is (**Figure 27 - 34**).

Almost identical trends that were predicted in the previous analysis (chapter 4.4) were estimated from the model of the toggle switch circuit behavior in presence of excess aTc in the initial phase of simulation. Visually, the functions are more deviating (**Figure 31**) compared with previously predicted (**Figure 27**).

The simulation resulted in prediction of the increase of both repressor concertation, AraC and TetR, in *P. pachastreallae*, *P. deceptionensis, P. aestusnigri* and *P. taenensis.* The initial uprise of the two proteins concentration was present in *V. natriegens* which was followed with the further increase of AraC and decrease of TetR. The similar initial arise of repressor was estimated in *E. coli*, while in the later simulation production of TetR and consumption of AraC were predicted. The simulated performance of *P. oceani* showed the initial synthesis the pair of proteins and as the simulation progressed AraC concentration started to decrease, while TetR concentration was estimated to have almost linear increase over observed time period.

The result of the simulation of *u* and *v* quantities change in the aTc-induced toggle switch was almost identical as the outcome of the analysis without induction of the system. This can be expressed as: increase of *u* and *v* in *P. aestusnigri*, *P. deceptionensis* and *P. taenensis*; decrease of both quantities in *P. oceani*; initial uprise of *u* and further decline of *u* and *v* in *V. natriegens*; initial increase of *v* and later decrease of the quantities in *P. pachastreallae;* and lastly, almost constant *u* and *v* values approximated in *E. coli*.

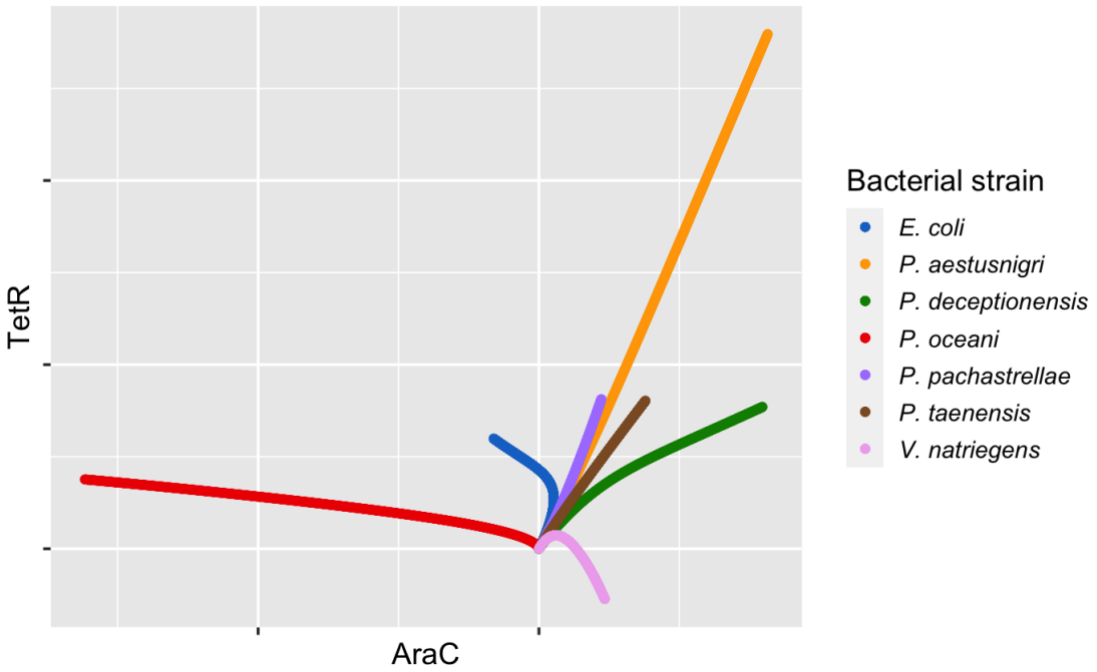


Figure 31 Simulated performance of the toggle switch induced with aTc *in different species shows the effect of the species-specific growth rate on expression ability of bacterial strain by approximating the change of TetR and AraC. The more rapid cell division (higher μ-value), the higher estimated protein dilution. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that K_{AraC} = K_{TetR} = K. Production of AraC and TetR, proteins that are synthetized by the toggle switch, is presented on X and Y axis respectively and the dimensions are considered arbitrary. Bacterial strains are marked in different colors: blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

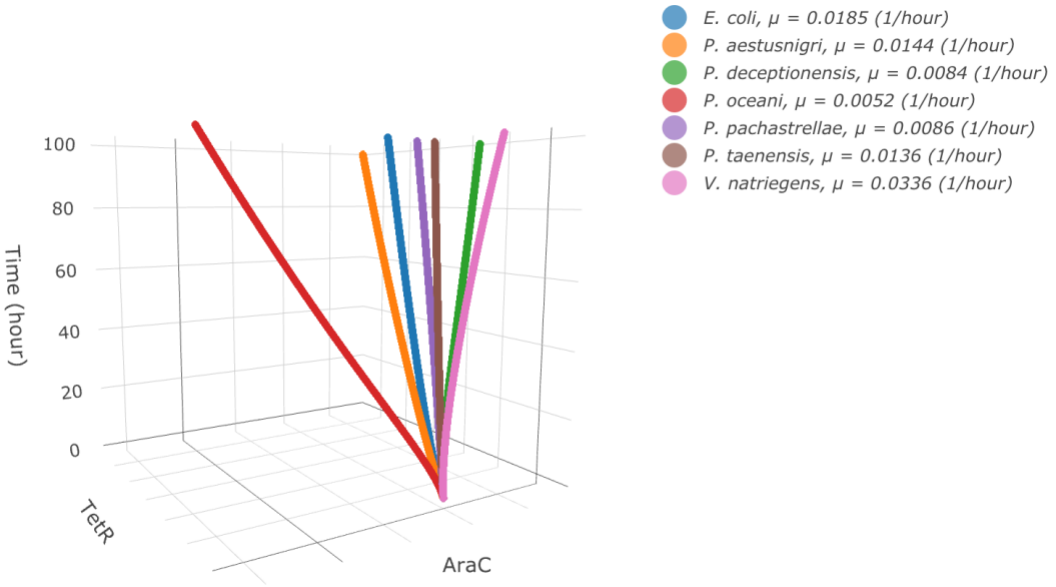


Figure 32 Simulated performance of the toggle switch induced with aTc *in different hosts approximates the effect of species-specific growth rate on the expression ability of bacterial strain by estimating the production of AraC and TetR. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that K_{AraC} = K_{TetR} = K. Production of AraC and TetR, proteins that are synthetized by the toggle switch, is presented on X and Y axis respectively and the dimensions are considered arbitrary. The Z-axis shows time. Bacterial strains are marked in different colors: blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

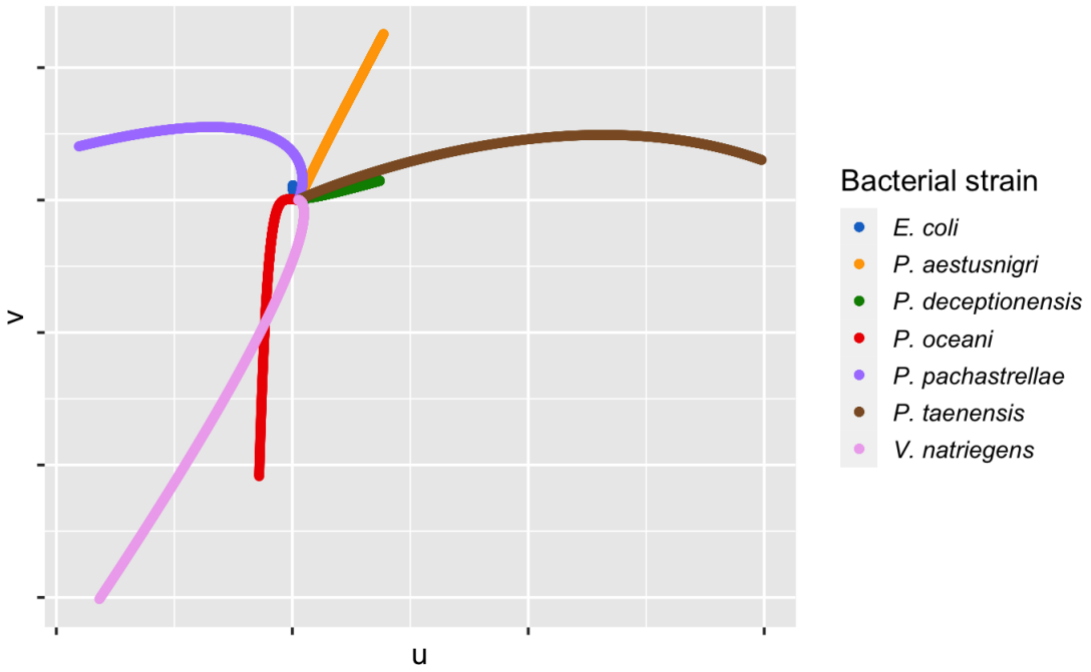


Figure 33 Simulated performance of the aTc-induced toggle switch in different species with a nondimensional analysis *shows the effect of the species-specific growth rate on expression ability of bacterial strain by approximating the u and v change. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that* K_{AraC} = K_{TetR} = K. u ($u = C_{AraC}/K_{AraC}$) *and* v ($v = C_{TetR}/K_{TetR}$) *are quantities that represent activity of the toggle switch, and they are positioned on X and Y axis respectively, and their dimensions are considered arbitrary. Bacterial strains are marked in different colors: blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

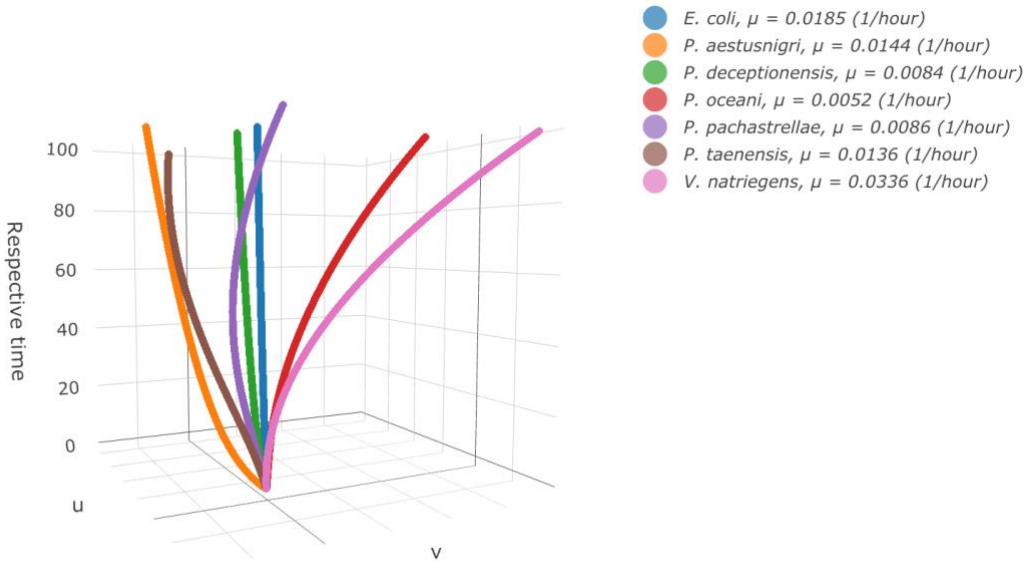


Figure 34 Simulated performance of the aTc-induced toggle switch in different species with a nondimensional analysis *in different species is under effect of the growth rate which is represented in different expression ability of bacterial strain that is approximated with u and v change over time. The predicted behavior was modeled for n = 1, K = 1.6285 M, and under the assumption that the maximal level of promoter is achieved, that the rate of degradation is negligible and that* K_{AraC} = K_{TetR} = K. u ($u = C_{AraC}/K_{AraC}$) *and* v ($v = C_{TetR}/K_{TetR}$) *are quantities that represent activity of the toggle switch, and they are positioned on X and Y axis respectively and their dimensions are considered arbitrary. The Z-axis shows respective time,* τ, *estimated as product of time and specific growth rate,* $\tau = t\mu$. *Bacterial strains are marked in different colors: blue color for E. coli, orange for P. aestusnigri, green for P. deceptionensis, red for P. oceani, purple for P. pachastrellae, brown for P. taenensis and V. natriegens in pink color.*

The comparison of the simulated device behavior in presence and absence of the inducer (aTc) resulted in similar trends of the protein concentration change (AraC, TetR). The induction effect on the toggle switch performance was modeled by a dimensional and non-dimensional analysis. The predicted effect was estimated as a change of the ODEs system solution function linearity, but the trends of simulated functions have remained the same.

The overall results of the previously performed simulations of the toggle switch performance across different microbial hosts indicate that the estimated concentration of synthetized proteins by the device are the consequences of the species-specific growth rates that affect the expression ability of the bacterial strain. The μ-value effect was estimated through the sink term, an expression that considered the dilution rate of the proteins, which is formulated as μC_{AraC} and μC_{TetR} in Equation 40 and 41. The formulated outcome of presented models was predicted in dimensional and non-dimensional analysis for the system without initial induction and for the system simulated in the AraC-dominant state which was modeled with the presence of excess aTc.

5 Discussion

The aim of the performed study was to enhance the understanding of the genetic toggle system and to predict and approximate its functioning in various microbial hosts including the marine bacteria and *E. coli* (**Table 1**). This was done by developing a framework which simulated and compared the performance of engineered genetic devices across different hosts. For this purpose, a mathematical model that accounted the effect of species-specific growth rate within a system of non-linear ordinary differential equations was created. The system of ODEs (Equation 40 and 41) was solved to estimate the protein production rate, predicted by the Hill function, and the dilution rate, approximated with the sink term.

The parametrization of the Hill function applied in the simulation was accomplished by creating a non-linear model in R that predicted the Hill (n) and the activation (K) coefficient. The model was supplied with the data collected from the induction study performed with *E. coli* cells induced with L-arabinose. In this assay, GFP and mKate fluorescence were measured together with OD_{600}, while GFP, GFP/OD_{600} signals from the late log phase of cell growth and GFP rates were further used for the estimation of parameters. Moreover, the activation coefficient approximated by the non-linear regression was included in the model based on one ODE (Equation 36) that simulated a dynamic response across different hosts. In addition to this analysis, the response time ($T_{1/2}$) for every bacterial species was estimated.

5.1 Parametrization of the Hill function with non-linear models

The overall goal of parametrizing the Hill function is to estimate specific constants from relevant data, such as the activation (K) and Hill coefficient (n) presented in Equation 1 and 2 (Alon, U., 2019). In this specific case, monitoring the GFP signal, that is a result of promoter activity can lead to prediction of mentioned specific parameters. The approximation of K and n value was performed by incorporating estimated GFP signal from induction study to a non-linear model (generally presented in Equation 29).

The direct measure of promoter activity relies on determination of the presence of the repressors in the system, in this case AraC and TetR, or the number of messenger RNAs (mRNAs) from promoter in individual cells (Sepúlveda et al., 2016; Bintu et al., 2005). The operon activity is under control of a regulator gene that is transcribed (*araC*, *tetR*) which results in production of a small protein molecule. If the operon is off, no presence of the repressor is determined (Lutz,

R., & Bujard, H., 1997; Fontanarrosa et al., 2020). The precise calculation of the protein concentration can be challenging due to their modifications, different locations in the cell and properties of the environment in which they are found (Knight, M. I., & Chambers, P. J., 2003).

An accurate estimation of the promoter activity by the bulk measurements is often infeasible due to unique number of copies of the promoter for each cell. The difference is a result of an individual copy number of the plasmid or the genome in which the promoter is carried. The plasmid copy number is determined by the origin of replication and dependent on the growth conditions and genetics of the host strain while the genome copy number is under the effect of growth rate and the distance to the origin (Shao et al., 2021).

To overcome the difficulties of quantifying the promoter activity, the reporter proteins such as GFP and mKate can be included in the system (Stark et al., 2018). In other words, in the system from this study P_{Tet} controls the expression of *araC* and additionally of *mKate* gene, while P_{BAD} controls *mKate* and *tetR* gene expression. By measuring the fluorescence intensity of mKate and GFP produced by the device, the levels of the two repressors can be easily approximated (Lugagne et al., 2017).

The GFP requires a longer time to reach a stabile fluorescence emission because of low folding rate of the protein into active form and it continues to fluoresce even after the cell death (Gutiérrez et al., 2015). The limitation of GFP signal use is that it cannot be amplified, hence the detection of low expression levels may not be monitored. Because of relatively slow folding properties of GFP, the study of the processes that include fast transcriptional activation may be challenging (Salehi Jouzani, G. R., & Vasilievna Goldenkova, I., 2005).

There are different GFP variants with improved characteristics, but a lag between synthesis and fluorescence is still present (Southward, C. M., & Surette, M. G., 2002). Moreover, previous studies showed that the formation of a fluorophore is couped to the folding of the fusion protein (Waldo et al., 1999; Cluzel et al., 2000). The fluorophore formation requires oxygen, hence the use of GFP as reporter requires aerobic conditions. GFP is sensitive to pH perturbations, and it is recommended to use GFP variants with lower sensitivity to pH when changes of the intracellular pH are expected (Southward, C. M., & Surette, M. G., 2002).

A promoter activity can be indirectly estimated by the intensity of GFP fluorescence (DeLisa et al., 1999) including the simple GFP and GFP/OD_{600} signals or the rate of GFP change over time. The selection of input signal directly affects model estimates and model characteristics,

such as model fit and distribution of the residuals. Therefore, in this study, every output from these three approaches was compared, whereas the coefficient estimates with the highest quality of statistical parameters were selected for the further analysis.

Estimating the rates of GFP-change-over-time from the induced cells and supplying the non-linear models with them resulted in better statistical parameters reported for the models. This outcome was a consequence of estimating GFP production trend for the complete selected phase, the log phase, instead of approximating the signal in only one time point.

GFP-, GFP/OD_{600} signal and GFP rates estimates from the cells that have reached the late log phase showed saturation behavior to increasing inducer concentration (**Figure 4, 5** and **16**). This includes both signal and rate escalation with L-arabinose concentration increase until the saturation point is reached. Therefore, this response can be described by the non-linear model with saturation trend as the Hill function.

In the later simulations (4.3 and 4.4) the dilution term was modeled by the specific growth rate only. Therefore, the Hill function was parametrized with GFP data estimated from the cells that have reached a log phase. A prerequisite for estimation of the growth rate is that cells are in a growing phase. If the simulation considered the cells with arrested growth, the ones in a stationary phase, the sink term, α, should have accounted both degradation and the dilution rate. Hence, the assumption that α_{deg} is negligible should be rejected and the removal rate can be modeled as $\alpha = \alpha_{deg} + \alpha_{dil} = \alpha_{deg} + \mu$.

The first analysis included estimation of the effect of different *n*-values to a model. A change of this constant has significantly affected the non-linear model fit and its residuals regardless the input signal data. The change of *n*-value resulted in a better model fit with lower estimated *n*, such as for $n = 1$. On the other hand, increase of coefficient up to 4 is followed by poorer quality of model fit. The *n*-value had weaker effect on models supplied with GFP rates input compared to other data, which is remarked as better statistical parameters estimated for the GFP rate models with higher *n*.

Activation coefficient (*K*) was estimated under assumption that $n = 1$ since it was the closest whole number to a non-linear model output estimate which was $n = 0.72318$, $n = 0.74454$, $n = 0.48324$ for GFP, GFP/OD_{600} and GFP rates input respectively (**Table 4**, **5** and **11**). Additional reason was that *n* is defined as constant that takes values between 1 and 4 (Alon, U., 2019) and estimating the *n* value as lower than 1 would not be theoretically supported.

This decision, $n = 1$, disables further bistability analysis of the toggle switch since it is achievable only for the elevated Hill coefficients (Osella et al., 2014). The bistability of the toggle is achieved due to the mutually inhibitory arrangement of the repressor genes. It is reached when one of the inhibitors repress gene expression with cooperativity that is greater than one ($n > 1$). The increase of *n* affects the robustness of the system which results in achieving the bistability even in a case of weaker promoters due to greater bistabile region (Gardner et al., 2000).

The parametrization of the Hill function was performed for the *E. coli* and these coefficients, *n* and *K*, were used for the later simulations across all other selected organisms. By making the assumption that the constants are equal among bacterial host, the uncertainty of predicted Hill term parameters in other bacterial strains increases. The quality of parametrization can be improved by performing the induction study for every host and estimating the specific activation (*K*) and Hill (*n*) coefficient for every bacterial strain. Moreover, in an advanced approach collection of the aTc-induction assay data together with L-arabinose induction study could result in estimation of both activation coefficients, K_{AraC} and K_{TetR}.

The general characteristic of all presented non-linear models is that residuals are not equally distributed around zero, with a trend of further dislocating for extremely- low or high L-arabinose concentrations. This trend is especially perceptible in residual graphs with high n-value, such as 3 or 4. A careful selection of the inducer concentration in narrower range as a part of the experimental design, might result in a better residual distribution which should allow a higher certainty of the model.

The importance of parametrization of the Hill function in nondimensionalized form with selected input data (GFP, GFP/OD_{600} and GFP rates) is hidden in constant *K* that is unitless. This could be beneficial for analysis of the dynamical system which may be accomplished in a much convenient manner.

The difference that implementing various input signal produce, is seen as better non-linear model fit and closer distribution of residuals around zero with models containing GFP rates instead of GFP or GFP/OD_{600} signals. The improved quality of model fit was estimated by the statistical parameters and by monitoring how well regression line visually describes the data. P-values in the models with GFP rates were significant and lower than the ones approximated with other model inputs. This suggests that the non-linear model supplied with the rates data simulates the Hill function with a higher precision.

The cell growth is not a result of induction with L-arabinose as specific growth rates were not responsive to L-arabinose. This was determined by calculating μ_{max} as a slope of the linear part of OD_{600}-time curves and running linear model between estimated growth rates and different inducer concentrations. There was no correlation between μ_{max} and L-arabinose concentration which was proven by p-value for the linear model slope estimate that was not significant (**Table 10**). Because of that, there was no intention to examine the effect of inducer concentration to a growth such as estimating GFP rates that are normalized with OD_{600}. Therefore, all non-linear models that parametrized the Hill function from a rate approach, were supplied only with the GFP data, as it was considered responsive to L-arabinose.

5.2 Dynamic response of L-arabinose-activated gene expression across host-specific growth rates

The dynamic response is a parameter that characterizes the promoter expression ability in a presence of the inducer. A transcription factor binding influences the attributes of the gene expression which can be monitored by the green fluorescent protein (GFP) reporter gene. The activity of the promoter can be approximated indirectly with the GFP expression levels. Use of GFP as a transcriptional reporter allows measurement of a gene expression in real time in living cells and at the single-cell level (Southward, C. M., & Surette, M. G., 2002). GFP is usually included in non-disruptive studies that observe living cells since it does not interfere with the host growth (Andersen et al., 1998).

The dynamic response of L-arabinose-activated expression can be simulated by the simple ODE presented in Equation 35. The model accounts GFP signal as indirect measure of the promoter activity. The change of GFP signal over time is approximated with the Hill term, the protein production rate, that accounts the strength of the effect of a transcription factor on a target gene and the dilution/degradation term ($-\alpha GFP$). The selected induction study data was collected from the cells that were in the exponential phase of the growth.

The choice to estimate the reporter signals only from the growing cells, allowed us to assume that degradation rate is negligible and that the reduction in protein concentration is due to the increase of cell volume during growth only ($\alpha = \alpha_{dil} = \mu$). In the specific case that the stationary phase of the cell growth is selected, the degradation rate should be accounted in the sink term ($\alpha = \alpha_{deg} + \mu$). From the simulation perspective, the estimation of the specific growth rate is relatively experimentally easy, while α_{deg} approximation might be challenging

as mechanisms of intracellular protein degradation are still being investigated and direct monitoring of intracellular protein concentration may be difficult.

The dynamic response can be quantified by estimating the response time ($T_{1/2}$), which represents the time to reach halfway between the initial and final levels of dynamic processes that can be estimated as a half of the steady-state protein concentration. This measure is determined only by the removal rate (α), which is approximated as $\alpha = \alpha_{dil} = \mu$ for actively growing cells. The fast removal allows rapid changes in the concentration.

The estimation of $T_{1/2}$ by Equation 39 indicates that the higher specific growth rate, the more rapid $T_{1/2}$ is. The response time is considered as limiting factor for designing efficient gene circuits (Alon, U., 2019). Preferred fast-dynamics systems, that reach the steady state faster, are characterized by a low $T_{1/2}$. This suggests a rapid transition of TF in active state when signal appears in environment, followed with a fast binding of TF to the promoter of a gene which is further transcribed, and the mRNA is translated resulting in protein production (Alon, U., 2019).

The study from Rosenfeld et al. (2002) indicates that response times can be increased by two strategies. Firstly, by a negative autoregulation, that represents a network motif in which a transcription factor inhibits its own expression (Stewart et al., 2013). This approach speeds up the rise-time, while not affecting the turn-off time (Rosenfeld et. al, 2002). The alternative way would be a degradation of the gene product and then $T_{1/2}$ is estimated by the degradation rate. Therefore, both the rise-time and the turn-off time of protein levels are decreased (Rosenfeld et. al, 2002).

In the steady-state conditions, concentration of the protein is constant and equal to a ratio of the production and removal rate. The higher protein concentrations are obtained with higher production rates. On the other hand, higher removal rate lowers the steady-state protein concentration. The response time is not affected by the production rate, but it influences the steady-state level. To maintain the steady state when the rapid degradation rates of the proteins are estimated, it is necessary to balance it with the high production rate (Alon, U., 2019). From engineering perspective, it is beneficial to closely match the time scales of production and dilution/degradation to quickly obtain the steady state.

The initial analysis of dynamic response was performed for *E. coli* cells for only one-half of the toggle switch as the induction data was collected from the part of device activated with one

molecule, L-arabinose. A selected side of the toggle switch contains the P_{BAD} promoter that expresses *tetR* (encode the TetR production) and *gfp* genes (encode the GFP production). The collected induction assay data was a result of the activity of the one part of the device due to induction of only one promoter. L-arabinose was expected to trigger the activity of P_{BAD} promoter (Lee et al., 1981) in *E. coli* and that was quantitively measured with GFP signal intensity. The mKate signal was monitored during the experiment and did not indicate any activity of the P_{Tet} promoter during the investigation, which suggested that this promoter was repressed with TetR protein synthetized by the P_{BAD} promoter.

As a result of GFP signal record, it was possible to display different GFP-time curves which led to visualization and assessment of correlation between inducer concentration and GFP signal. This interaction was explained by non-linear models formulated for *E. coli* cells and quantitively determined through estimated coefficients by *nls*() function in R (Ritz, C., & Streibig, J. C., 2008). Approximation of these parameters was necessary for dynamic response analysis that included incorporation of predicted parameters in an ODE that describes a behavior of one side of toggle switch (Equation 35) with *E. coli* as a host. The ODE was solved in R, and it was expected to be a function (Soetaert et al., 2012), which indeed was, and it showed saturation trend as presented in **Figure 23**.

The estimated GFP signals showed a saturation trend with the increase of L-arabinose concentration, consequentially synthesis of TetR protein is under the same influence of the inducer. The activity of P_{BAD} promoter was triggered with L-arabinose which induced transcription of *tetR* gene and synthesis of TetR protein which was reported by GFP. P_{BAD} promoter has showed activity when induced with L-arabinose which was possible to describe with the model based on Hill's kinetics given in Equation 35 (Lee et al., 1981; Alon, U., 2019).

The dynamics simulation of the toggle switch performed in this study was limited to only one side of the device given the data that was accumulated from the induction assay. By collecting additional induction data from the system induced with aTc, it is possible to predict a performance of the entire device. In that case, the P_{Tet} promoter activity can be indirectly measured by fluorescence of mKate reporter protein. mKate signal data should be then used for estimation of activation coefficient for P_{Tet} promoter, K_{TetR}, and for a maximal rate of mKate signal, ($mKate_{rate}max = R_{mKate}$). These parameters can be approximated by a non-linear model in R as it was previously done for K_{AraC} and R.

To enable comparison of the dynamic response between bacterial strains, it was convenient to make some assumptions. Hence, $f(ARA)$ from Equation 37 is assumed to be the same for all bacterial strains allowing us to compare the dynamic response among strains. Therefore, the changes of coefficient K_{TetR} and maximal rate (R) values across bacterial strains were not considered but they are expected in a real system. The sink term in simulation was different for every host and included dilution rate term which was approximated as the specific growth rate.

The estimated dynamic response across different bacteria in **Figure 24** has showed the fastest response of *V. natriegens* that had the highest μ, the result was additionally confirmed by response time calculation and their distribution among strains presented in **Figure 26**. Currently, *V. natriegens* is considered as the one of the fastest growing organisms (Eagon R. G., 1962; Maida et al., 2013). Its Vmax Express strain shows a doubling two times faster than E. coli and has the ability to generate high quantities of protein per cell volume (Des Soye et al., 2018). Moreover, *V. natriegens* fulfills the requirements for applications in biotechnology while being a marine bacterium (Hoffart et al., 2017) and it might be considered as a model organism due to its unique attributes. The slowest response was simulated for *P. oceani* which was considered as a consequence of very low μ estimated for this organism.

5.3 Simulation of a toggle switch performance across different bacterial hosts

The genetic toggle switch modeled as a dynamic system was described with a system of ODEs (Equation 40 and 41) and its performance was estimated by the solution computed with *ode*() function in R. The intention of modeling the toggle switch was to understand and explain its behavior with additional goal to compare how a performance of the device is impacted by the change of organism. Numerous strategies that can be conducted to model a dynamic network has been developed recently (Zheng, Y., & Sriram, G., 2010; Kirk et al., 2013).

By the term “chassis” we refer to an organism that is a recipient of engineered biological systems. A propagation of genetic information together with a gene expression and programable biological functions are activities that a host should perform (Kim et al., 2016). An identical engineered genetic device shows unique performance across different chassis due to differences in host physiology and interactions between genetic parts and the host cell (Bartoli et al., 2020). Hence, the effect that a recipient has on the device is called a "chassis" effect.

This thesis approach included assumption that model parameters for the Hill term (Equation 40 and 41), the activation (K) and the Hill coefficient (n), are constant and equal for every bacterial strain. K was derived from **Table 12** and n was estimated as $n = 1$. It was supposed that this system reaches the maximal expression level of promoter, and that dilution process is approximated with a specific growth rate in a sink term. Under these conditions the "chassis" effect was simulated in controlled manner where the only impact was from the μ.

A limitation of presented approach is that activation coefficients of different promoters are considered equal ($K_{AraC} = K_{TetR}$). To overcome this drawback, an additional induction assay with aTc-induction may be performed to estimate both coefficients. Moreover, K_{AraC} and K_{TetR} are considered equal among different hosts and their influence on the Hill term is not considered within a model. Further studies might estimate the "chassis" effect by modeling species-specific Hill term which should be supplied with coefficients estimated from induction study performed in every bacterial strain for both sides of the toggle switch.

To estimate possible effects that presence of inducer might cause, the simulation that accounted the initial concentration of aTc was performed. It was expected that presence of one inducer should keep one side of the toggle switch inactive and that model may predict the activity of the other side (Lee et al., 1981; Alon, U., 2019; Guzman et al., 1995). The performed analysis, given in chapter 4.4.1, indicated presence of the "chassis" effect regardless to initial concentration which was proven with dimensional and nondimensional analysis.

The simulated performance of the toggle switch was unique for every examined strain due to different specific growth rates. Both dimensional and nondimensional analysis have indicated a presence of the "chassis" effect with a respect to different inducer's concentrations. The effect was considered through several perspectives, such as ability of the device to synthetize proteins, the synthesis rate estimate and monitoring if device stops to produce proteins after some time.

As shown in simulation figures (**Figure 27-34**) activity of toggle switch is present regardless μ change and initial inducer concentration. The modification of the growth rate affected the solution function from the ODEs system that describes the network. The activity of the device was determined as a change in concentration of at least one of the proteins (AraC and/or TetR).

A performance of the toggle switch was simulated under specific assumptions including $n = 1$. This selection of Hill coefficient value was made through considering statistical parameters and a quality of a non-linear model, that is described in chapters 4.1 and 4.2. For

bistability analysis, *n* should be greater than 1 (Alon, U., 2019). Hence, it was not possible to examine two stabile states of the toggle switch in these conditions.

The bistabile systems switch into the corresponding stable state that is triggered by the input signal and remain in that state even after removal of the signal (Lebar et al., 2014). The bistability of simulated toggle switch system given in **Figure 1** is presented with two stabile states: red-mKate-dominant (AraC) state and green-GFP-dominant (TetR) state. In presence of L-arabinose the device switches to TetR-state and in presence of excess aTc to AraC-dominant state (Lebar et al., 2014; Lugagne et al., 2017).

In further research, for examination of two stabile states of the device, a nullcline analysis may be performed. Nullclines represent curves at which one of the proteins has constant concentration, in other words, rate of concentration change of this protein is zero. The crossing points of the two nullclines are places where neither protein changes nor there is a fixed point (Alon, U., 2019). Another tool that can be used for determination of stability of the system of ODEs that are describing the genetic toggle switch is Jacobi stability analysis (Abolghasem, H., 2012). Moreover, the stability criteria can be defined by examination different initial conditions for the toggle switch performance (e.g., the Hill coefficient, the activation coefficient, initial concentration of repressor/inducer etc.).

It is important to highlight, that mathematical models are useful tools to predict and approximate the nature when experimental approach is too challenging to perform. A simulation of dynamic process should support better understanding of its mechanism and predict all factors that affect its performance. From the synthetic biology perspective, a mathematical modeling can be used to estimate potentials of microbes to perform new functions by predicting their performance *in silico*. By merging experimental quantitative data and formulated models, it is possible to achieve application of engineered microbes as technological platform (Chandran et al., 2008). A unique role of a scientist is to understand model limitations, such as difficulty to describe a process of importance completely accurately, and to select the best possible option for their research.

Synthetic biology as a scientific field is constantly seeking for new bacterial hosts for various engineering purposes. The vast majority of the dynamic studies of the toggle switch is focused on *E. coli* while neglecting the unique potentials of the other species. Hence, the intention of this study was to predict the toggle switch performance across different bacterial hosts including the marine bacteria and to point out the possibilities of marine bacteria as recipients in future.

6 Conclusion

Finally, I would like to summarize the main findings of this thesis.

- The microbial growth is not responsive to L-arabinose concentration.
- Parametrization of Hill function with non-linear model showed the best quality of the estimates with GFP rates compared to the GFP and GFP/OD_{600} signals with the respect to the statistical parameters of *nls*() function output.
- GFP, GFP/OD_{600} signals and GFP rates estimated from *E. coli* cells in the late log phase show saturation behavior with increasing L-arabinose concentration.
- The Hill coefficient, *n*, has significant impact on non-linear models and a poorer model quality with a respect to statistical parameters, is a consequence of elevated *n*-value.
- The effect of the Hill coefficient on a non-linear model with the GFP rates data is not as intense as with GFP signals or GFP/OD_{600}.
- None of the estimated non-linear models resulted in equal distributions of residuals.
- The model residuals were further dislocated from zero in areas of extremely low- or high L-arabinose concentration.
- A dynamic response of L-arabinose-activated gene expression in *E. coli* can be represented with a hyperbolic saturation curve.
- The simulation of the dynamic response of L-arabinose-activated gene expression across host-specific growth rates, under assumption that the maximal rate (R) and activation coefficient (K) do not change with the strain, indicated a hyperbolic saturation trend for every host.
- The fastest response of the one side of the toggle switch, when assumed that $f(ARA)$ is equal for every bacterial strain, was estimated for *V. natriegens* as a host organism.
- The highest response time ($T_{1/2}$) and the slowest response was simulated for *P. oceani.*
- The "chassis" effect was approximated by a simulation of the toggle switch performance that estimated a high impact of the specific growth rate (μ) on a behavior of the genetic device which was concluded in dimensional and nondimensional analysis.
- The presence of the "chassis" effect within a toggle switch across different bacterial strains was predicted in conditions when none of the inducers were present in the environment and for the initial aTc-induction.
- The simulations suggested that the "chassis" effect manifests itself as an influence of the species-specific growth rate (μ) on the performance of the device.

7 References

Abolghasem, H. (2012). Liapunov Stability Versus Jacobi Stability. *Journal of Dynamical Systems and Geometric Theories*, 10(1), 13-32. doi:10.1080/1726037X.2012.10698604

Aiyar, S. E., Gaal, T., & Gourse, R. L. (2002). rRNA Promoter Activity in the Fast-Growing Bacterium *Vibrio natriegens*. *Journal of Bacteriology, 184*(5), 1349-1358. doi:10.1128/JB.184.5.1349-1358.2002

Allocati, N., Masulli, M., Alexeyev, M. F., & Di Ilio, C. (2013). Escherichia coli in Europe: An Overview. *International Journal of Environmental Research and Public Health, 10*(12), 6235-6254. https://doi.org/10.3390/ijerph10126235

Alon, U. (2019). *An introduction to systems biology: design principles of biological circuits*. Chapman and Hall/CRC.

Amemiya, T. (1983). Non-linear regression models. *Handbook of econometrics, 1*, 333-389.

Andersen, J.B., Sternberg, C., Poulsen, L.K., Bjorn, S.P., Givskov, M., and Molin, S. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and environmental microbiology*: 2240–2246. doi: 10.1128/AEM.64.6.2240-2246.1998

Andrianantoandro, E., Basu, S., Karig, D. K., & Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology, 2*(1), 2006.0028. https://doi.org/10.1038/msb4100073

Anton, B. P., & Raleigh, E. A. (2016). Complete Genome Sequence of NEB 5-alpha, a Derivative of Escherichia coli K-12 DH5α. *Genome Announcements*, 4(6). doi:10.1128/genomeA.01245-16

Ascher, U. M., Mattheij, R. M. M., & Russell, R. D. (1995). *Numerical solution of boundary value problems for ordinary differential equations*. (Unabridged, corr. republication. ed.) (Classics in applied mathematics; Vol. 13). Society for Industrial and Applied Mathematics (SIAM).

Barbour, A. G., Adeolu, M., & Gupta, R. S. (2017). Division of the genus Borrelia into two genera (corresponding to Lyme disease and relapsing fever groups) reflects their genetic and phenotypic distinctiveness and will lead to a better understanding of these two groups of microbes. *International journal of systematic and evolutionary microbiology*, *67*(6), 2058-2067. doi: 10.1099/ijsem.0.001815

Bartoli, V., Meaker, G. A., di Bernardo, M., & Gorochowski, T. E. (2020). Tunable genetic devices through simultaneous control of transcription and translation. *Nature Communications, 11*(1), 2095. doi:10.1038/s41467-020-15653-7

Bennasar, A., Rossello-Mora, R., Lalucat, J., & Moore, E. R. (1996). 16S rRNA Gene Sequence Analysis Relative to Genomovars of Pseudomonas stutzeri and Proposal of Pseudomonas balearica sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, *46*(1), 200-205. https://doi.org/10.1099/00207713-46-1-200

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Phillips, R. (2005). Transcriptional regulation by the numbers: applications. *Current Opinion in Genetics & Development*, 15(2), 125-135. https://doi.org/10.1016/j.gde.2005.02.006

Bollinger, A., Thies, S., Knieps-Grunhagen, E., Gertzen, C., Kobus, S., Hoppner, A., Jaeger, K. E. (2020). A Novel Polyester Hydrolase From the Marine Bacterium Pseudomonas aestusnigri - Structural and Functional Insights. *Frontiers in Microbiology, 11*, 114. doi:10.3389/fmicb.2020.00114

Cai, T.M., L.B. Guan, L.W. Chen, S. Cai, X.D. Li, Z.L. Cuiand, and S.P. Li. 2007. Enhanced biological phosphorus removal with Pseudomonas putida GM6 from activated sludge. *Pedosphere*. 17(5):624-629. https://doi.org/10.1016/S1002-0160(07)60074-5

Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nature reviews. Microbiology, 12*(5), 381-390. doi:10.1038/nrmicro3239

Carrión, O., Curson, A. R. J., Kumaresan, D., Fu, Y., Lang, A. S., Mercadé, E., & Todd, J. D. (2015). A novel pathway producing dimethylsulphide in bacteria is widespread in soil environments. *Nature Communications*, 6(1), 6579. doi:10.1038/ncomms7579

Carrión, O., Miñana-Galbis, D., Montes, M. J., & Mercadé, E. (2011). Pseudomonas deceptionensis sp. nov., a psychrotolerant bacterium from the Antarctic. International Journal of Systematic and Evolutionary Microbiology, 61(Pt 10), 2401-2405. doi:10.1099/

Chandran, D., Copeland, W. B., Sleight, S. C., & Sauro, H. M. (2008). Mathematical modeling and synthetic biology. *Drug Discovery Today: Disease Models, 5*(4), 299-309. https://doi.org/10.1016/j.ddmod.2009.07.002

Cho, E., & Lu, Y. (2020). Compartmentalizing Cell-Free Systems: Toward Creating Life-Like Artificial Cells and Beyond. *ACS Synthetic Biology*, 9(11), 2881-2901. doi:10.1021/acssynbio.0c00433

Cluzel, P., Surette, M., and Leibler, S. (2000). An ultrasensitive bacterial motor revealed by monitoring signaling proteins in single cells. *Science*, 287: 1652–1655.

Cook, R. D., Tsai, C.-L., & Wei, B. (1986). Bias in nonlinear regression. *Biometrika, 73*(3), 615-623. https://doi.org/10.2307/2336526

Crawford, J. D. (1991). Introduction to bifurcation theory. *Reviews of Modern Physics*, 63(4), 991-1037. doi:10.1103/RevModPhys.63.991

DeLisa, M. P., Li, J., Rao, G., Weigand, W. A., & Bentley, W. E. (1999). Monitoring GFP-operon fusion protein expression during high cell density cultivation of Escherichia coli using an on-line optical sensor. *Biotechnology and Bioengineering*, 65(1), 54-64. https://doi.org/10.1002/(SICI)1097-0290(19991005)65:1<54::AID-BIT7>3.0.CO;2-R

Des Soye, B. J., Davidson, S. R., Weinstock, M. T., Gibson, D. G., & Jewett, M. C. (2018). Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. *ACS Synthetic Biology, 7*(9), 2245-2255. doi:10.1021/acssynbio.8b00252

Eagon, R. G. (1962). Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. *Journal of bacteriology*, 83(4), 736-737. https://doi.org/10.1128/jb.83.4.736-737.1962

Fábrega, M.-J., Knödlseder, N., Nevot, G., Sanvicente, M., Toloza, L., Santos-Moreno, J., & Güell, M. (2021). Establishing a Cell-Free Transcription–Translation Platform for Cutibacterium acnes to Prototype Engineered Metabolic and Synthetic Biology. *ACS Biomaterials Science & Engineering*. doi:10.1021/acsbiomaterials.1c00894

Fontanarrosa, P., Doosthosseini, H., Borujeni, A. E., Dorfan, Y., Voigt, C. A., & Myers, C. (2020). Genetic Circuit Dynamics: Hazard and Glitch Analysis. *ACS Synthetic Biology*, 9(9), 2324-2338. doi:10.1021/acssynbio.0c00055

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association* : 105(2), 203-206. doi:10.5195/jmla.2017.88

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Sauro, H. M. (2014). The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology, 32*(6), 545-550. doi:10.1038/nbt.2891

Gallant. A.R. (1975). Nonlinear regression. *The American Statistician, 29*(2), 73-81. https://doi.org/10.2307/2683268

García-Valdés, E., Gomila, M., Mulet, M., & Lalucat, J. (2018). Draft Genome Sequence of Pseudomonas oceani DSM 100277T, a Deep-Sea Bacterium. *Genome Announcements,* 6(15), e00254-00218. doi:10.1128/genomeA.00254-18

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature, 403*(6767), 339-342. doi:10.1038/35002131

Gomila, M., Mulet, M., Lalucat, J., & Garcia-Valdes, E. (2017a). Draft Genome Sequence of the Marine Bacterium Pseudomonas aestusnigri VGXO14(T). *Genome announcements, 5*(32). doi:10.1128/genomeA.00765-17

Gomila, M., Mulet, M., Lalucat, J., & García-Valdés, E. (2017b). Draft Genome Sequence of Pseudomonas pachastrellae Strain CCUG 46540T, a Deep-Sea Bacterium. *Genome Announcements*, 5(14), e00136-00117. doi:10.1128/genomeA.00136-17

Gopal, G. J., & Kumar, A. (2013). Strategies for the production of recombinant protein in Escherichia coli. *Protein Journal, 32*(6), 419-425. doi:10.1007/s10930-013-9502-5

Gutiérrez, J., Amaro, F., & Martín-González, A. (2015). Heavy metal whole-cell biosensors using eukaryotic microorganisms: An updated critical review. *Frontiers in Microbiology, 6*, 1-8. doi:10.3389/fmicb.2015.00048

Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology, 177*(14), 4121-4130. doi:10.1128/jb.177.14.4121-4130.1995

Hairer, E., & Wanner, G. (1996). *Solving ordinary differential equations II: Stiff and differential- algebraic problems*. Heidelberg: Springer.

Hindmarsh, A. C. (1983). ODEPACK, a systematized collection of ODE solvers. *Scientific computing*, 55-64.

Hoffart, E., Grenz, S., Lange, J., Nitschel, R., Müller, F., Schwentner, A., Drake, H. L. (2017). High Substrate Uptake Rates Empower Vibrio natriegens as Production Host for Industrial Biotechnology. *Applied and Environmental Microbiology, 83*(22), e01614-01617. doi:10.1128/AEM.01614-17

Jacob, F., Perrin, D., Sánchez, C., & Monod, J. (1960). The operon: a group of genes whose expression is co-ordinated by an operator. *Compte Rendu de l'Academie des Sciences, 250*, 1727-1729. doi: 10.1016/j.crvi.2005.04.005

Khalil, A. S., & Collins, J. J. (2010). Synthetic biology: applications come of age. *Nature reviews. Genetics, 11*(5), 367-379. doi:10.1038/nrg2775

Kim, J., Salvador, M., Saunders, E., Gonzalez, J., Avignone-Rossa, C., & Jimenez, J. I. (2016). Properties of alternative microbial hosts used in synthetic biology: towards the design of a modular chassis. *Essays in Biochemistry, 60*(4), 303-313. doi:10.1042/EBC20160015

Kim, K. H., Roh, S. W., Chang, H. W., Nam, Y. D., Yoon, J. H., Jeon, C. O., & Bae, J. W. (2009). Pseudomonas sabulinigri sp. nov., isolated from black beach sand. *International journal of systematic and evolutionary microbiology*, *59*(1), 38-41. doi: 10.1099/ijs.0.65866-0.

Kirk, P., Thorne, T., & Stumpf, M. P. H. (2013). Model selection in systems and synthetic biology. *Current Opinion in Biotechnology, 24*(4), 767-774. https://doi.org/10.1016/j.copbio.2013.03.012

Knight, M. I., & Chambers, P. J. (2003). Problems associated with determining protein concentration: a comparison of techniques for protein estimations. *Molecular Biotechnology, 23*(1), 19-28. doi:10.1385/mb:23:1:19

Lang, E., Burghartz, M., Spring, S., Swiderski, J., & Spröer, C. (2010). Pseudomonas benzenivorans sp. nov. and Pseudomonas saponiphila sp. nov., represented by xenobiotics degrading type strains. *Current microbiology*, *60*(2), 85-91. doi: 10.1007/s00284-009-9507-7

Lebar, T., Bezeljak, U., Golob, A., Jerala, M., Kadunc, L., Pirš, B., Jerala, R. (2014). A bistable genetic switch based on designable DNA-binding domains. *Nature Communications, 5*(1), 5007. doi:10.1038/ncomms6007

Lee, D. H., Moon, S. R., Park, Y. H., Kim, J. H., Kim, H., Parales, R. E., & Kahng, H. Y. (2010). Pseudomonas taeanensis sp. nov., isolated from a crude oil-contaminated seashore. *International journal of systematic and evolutionary microbiology*, *60*(12), 2719-2723. doi: 10.1099/ijs.0.018093-0.

Lee, H. H., Ostrov, N., Wong, B. G., Gold, M. A., Khalil, A. S., & Church, G. M. (2016). *Vibrio natriegens*, a new genomic powerhouse. *bioRxiv*, 058487. doi:10.1101/058487

Lee, H.H., Ostrov, N., Wong, B.G. *et al.* Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nature microbiology* **4,** 1105–1113 (2019). https://doi.org/10.1038/s41564-019-0423-8

Lee, N. L., Gielow, W. O., & Wallace, R. G. (1981). Mechanism of araC autoregulation and the domains of two overlapping promoters, Pc and PBAD, in the L-arabinose regulatory region of Escherichia coli. *Proceedings of the National Academy of Sciences*, 78(2), 752-756. doi:10.1073/pnas.78.2.752

Lee, S.-Y., Kim, S. H., Lee, D.-G., Shin, S., Yun, S. H., Choi, C.-W., Kim, S. I. (2014). Draft Genome Sequence of Petroleum Oil-Degrading Marine Bacterium Pseudomonas taeanensis Strain MS-3, Isolated from a Crude Oil-Contaminated Seashore. *Genome Announcements, 2*(1), e00818-00813. doi:10.1128/genomeA.00818-13

Li, H.F., B.Z. Li, E.T. Wang, J.S. Yang, & H.L. Yuan. (2012). Removal of low concentration of phosphorus from solution by free and immobilized cells of Pseudomonas stutzeri YG-24. *Desalination*. 286:242-247. https://doi.org/10.1016/j.desal.2011.11.029

Liu, S. (2017). Chapter 11 - How Cells Grow. In S. Liu (Ed.), Bioprocess Engineering (Second Edition) (pp. 629-697): Elsevier.

Lugagne, J. B., Sosa Carrillo, S., Kirch, M., Kohler, A., Batt, G., & Hersen, P. (2017). Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature communications*, 8(1), 1671. doi:10.1038/s41467-017-01498-0

Lutz, R., & Bujard, H. (1997). Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Research*, 25(6), 1203-1210. doi:10.1093/nar/25.6.1203

Maida, I., E. Bosi, E. Perrin, M. C. Papaleo, V. Orlandini, M. Fondi, R. Fani, J. Wiegel, G. Bianconi, & F. Canganella. 2013. "Draft Genome Sequence of the Fast Growing Bacterium Vibrio Natriegens Strain DSMZ759. "*Genome Announcements* 1(4): e00648–13– e00648–13. doi: 10.1128/genomeA.00648-13

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., & Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology*, 17: 969–973. doi: 10.1038/13657

Migula, W. (1894). Über ein neues System der Bakterien, Arbeiten aus dem Bakteriologischen Institut der Technischen Hochschule zu Karlsruhe.

Migula, W. (1900). *System der bakterien: bd. Specielle systematik der bakterien* (Vol. 2). G. Fischer.

Molitor, R., Bollinger, A., Kubicki, S., Loeschcke, A., Jaeger, K. E., & Thies, S. (2020). Agar plate-based screening methods for the identification of polyester hydrolysis by Pseudomonas species. *Microbial biotechnology, 13*(1), 274-284. doi: 10.1111/1751-7915.13418.

Mørken, K. (2017). *Numerical Algorithms and Digital Representation*. Department of Mathematics University of Oslo.

Nishimori, E., Kita-Tsukamoto, K., & Wakabayashi, H. (2000). Pseudomonas plecoglossicida sp. nov., the causative agent of bacterial haemorrhagic ascites of ayu, Plecoglossus altivelis. *International Journal of Systematic and Evolutionary Microbiology*, 50(1), 83-89. https://doi.org/10.1099/00207713-50-1-83

Osella, M., Riba, A., Testori, A., Corà, D., & Caselle, M. (2014). Interplay of microRNA and epigenetic regulation in the human regulatory network. *Frontiers in genetics, 5*, 345. doi:10.3389/fgene.2014.00345

Palleroni, N. J. (1984). Genus I. Pseudomonas. *Bergey's manual of systematic bacteriology*, *1*, 141-199.

Palleroni, N. J. (1993). Pseudomonas classification. *Antonie van Leeuwenhoek, 64*(3), 231-251. doi:10.1007/BF00873084

Palleroni, N. J. (2010). The Pseudomonas Story. *Environmental Microbiology*, 12(6), 1377-1383. https://doi.org/10.1111/

Palleroni, N. J. (2015). Pseudomonas. Bergey's manual of systematics of archaea and bacteria. *John Wiley & Sons, Inc*, *10*, 9781118960608.

Pascual, J., Lucena, T., Ruvira, M. A., Giordano, A., Gambacorta, A., Garay, E., & Macián, M. C. (2012). Pseudomonas litoralis sp. nov., isolated from Mediterranean seawater. *International journal of systematic and evolutionary microbiology*, *62*(2), 438-444. doi: 10.1099/ijs.0.029447-0

Payne, W. J., R. G. Eagon, & A. K. Williams. (1961). "Some Observations on the Physiology of Pseudomonas Natriegens Nov. Spec." *Antonie van Leeuwenhoek*27: 121–28. doi: 10.1007/BF02538432

Ritz, C., & Streibig, J. C. (2008). *Nonlinear regression with R* (Vol. 31): Springer.

Roehner, N., Beal, J., Clancy, K., Bartley, B., Misirli, G., Grünberg, R., Myers, C. J. (2016). Sharing Structure and Function in Biological Design with SBOL 2.0. *ACS Synthetic Biology, 5*(6), 498-506. doi:10.1021/acssynbio.5b00215

Romanenko, L. A., Uchino, M., Falsen, E., Frolova, G. M., Zhukova, N. V., & Mikhailov, V. V. (2005). Pseudomonas pachastrellae sp. nov., isolated from a marine sponge. *International journal of systematic and evolutionary microbiology, 55*(Pt 2), 919-924. doi:10.1099/ijs.0.63176-0

Romanenko, L. A., Uchino, M., Tebo, B. M., Tanaka, N., Frolova, G. M., & Mikhailov, V. V. (2008). Pseudomonas marincola sp. nov., isolated from marine environments. *International journal of systematic and evolutionary microbiology*, *58*(3), 706-710. doi: 10.1099/ijs.0.65406-0.

Rosenfeld, N., Elowitz, M. B., & Alon, U. (2002). Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology, 323*(5), 785-793. https://doi.org/10.1016/S0022-2836(02)00994-4

Ruder, W. C., Lu, T., & Collins, J. J. (2011). Synthetic biology moving into the clinic. *Science, 333*(6047), 1248-1252. doi: 10.1126/science.1206843

Salehi Jouzani, G. R., & Vasilievna Goldenkova, I. (2005). A New Reporter Gene Technology: Opportunities and Perspectives. *Iranian Journal of Biotechnology, 3*(1), 1-15. http://www.ijbiotech.com/article_6961.html

Sanchez, D., Mulet, M., Rodriguez, A. C., David, Z., Lalucat, J., & Garcia-Valdes, E. (2014). Pseudomonas aestusnigri sp. nov., isolated from crude oil-contaminated intertidal sand samples after the Prestige oil spill. *Systematic and Applied Microbiology, 37*(2), 89-94. doi:10.1016/j.syapm.2013.09.004

Sepúlveda, L. A., Xu, H., Zhang, J., Wang, M., & Golding, I. (2016). Measurement of gene regulation in individual cells reveals rapid switching between promoter states. *Science, 351*(6278), 1218-1222. doi:10.1126/science.aad0635

Shampine, L. F., Gladwell, I., & Thompson, S. (2003). *Solving ODEs with MATLAB*. Cambridge: Cambridge University Press

Shao, B., Rammohan, J., Anderson, D. A., Alperovich, N., Ross, D., & Voigt, C. A. (2021). Single-cell measurement of plasmid copy number and promoter activity. *Nature Communications, 12*(1), 1475. doi:10.1038/s41467-021-21734-y

Smyth, G. K. (2002). Nonlinear regression. *Encyclopedia of environmetrics, 3*, 1405-1411.

Soetaert, K., Cash, J., & Mazzia, F. (2012). *Solving differential equations in R*. Berlin: Springer.

Soetaert, K., Petzoldt, T. (2011). *Solving ODEs, DAEs, DDEs and PDEs in R*. JNAIAM.

Soetaert, K., Petzoldt, T., & R, W. (2010). Solving Differential Equations in R. *The R Journal, 2*, 5. doi:10.32614/RJ-2010-013

Soetaert, K., Petzoldt, T., Setzer, R. W., & Petzoldt, M. T. (2015). Package 'deSolve'. *Solving Initial Value Differential Equations, 2010d. deSolve vignette-R package version, 1*.

Southward, C. M., & Surette, M. G. (2002). The dynamic microbe: green fluorescent protein brings bacteria to light. *Molecular Microbiology*, 45(5), 1191-1196. https://doi.org/10.1046/j.1365-2958.2002.03089.x

Stark, J. C., Huang, A., Nguyen, P. Q., Dubner, R. S., Hsu, K. J., Ferrante, T. C., Jewett, M. C. (2018). BioBits Bright: A fluorescent synthetic biology education kit. *Science Advances, 4*(8), eaat5107. doi:10.1126/sciadv.aat5107

Stewart, A. J., Seymour, R. M., Pomiankowski, A., & Reuter, M. (2013). Under-Dominance Constrains the Evolution of Negative Autoregulation in Diploids. *PLOS Computational Biology, 9*(3), e1002992. doi:10.1371/journal.pcbi.1002992

Storch, M., Casini, A., Mackrow, B., Ellis, T., & Baldwin, G. S. (2017). BASIC: A Simple and Accurate Modular DNA Assembly Method. *Methods in molecular biology, 1472*, 79-91. doi:10.1007/978-1-4939-6343-0_6

Storch, M., Casini, A., Mackrow, B., Fleming, T., Trewhitt, H., Ellis, T., & Baldwin, G. S. (2015). BASIC: A New Biopart Assembly Standard for Idempotent Cloning Provides

Accurate, Single-Tier DNA Assembly for Synthetic Biology. *ACS synthetic biology, 4*(7), 781-787. doi:10.1021/sb500356d

Strogatz, S. H. (2018). Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering: CRC press.

Sullivan, I., DeHaven, A., & Mellor, D. (2019). Open and Reproducible Research on Open Science Framework. Current Protocols Essential Laboratory Techniques, 18(1), e32. https://doi.org/10.1002/cpet.32

Thoma, F., & Blombach, B. (2021). Metabolic engineering of Vibrio natriegens. *Essays in Biochemistry*, 65(2), 381-392. doi:10.1042/ebc20200135

Tinafar, A., Jaenes, K., & Pardee, K. (2019). Synthetic Biology Goes Cell-Free. *BMC Biology*, 17(1), 64. doi:10.1186/s12915-019-0685-x

Waldo, G.S., Standish, B.M., Berendzen, J., & Terwilliger, T.C. (1999). Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology*, 17: 691–695. https://doi.org/10.1038/10904

Wang, M.-q., & Sun, L. (2016). Pseudomonas oceani sp. nov., isolated from deep seawater. *International Journal of Systematic and Evolutionary Microbiology, 66*(10), 4250-4255. https://doi.org/10.1099/ijsem.0.001343

Weinstock, M. T., Hesek, E. D., Wilson, C. M., & Gibson, D. G. (2016). Vibrio natriegens as a fast-growing host for molecular biology. *Nature Methods, 13*(10), 849-851. doi:10.1038/nmeth.3970

Yoo, J., Kim, D. H., Oh, E. J., & Chung, K. Y. (2018). Quantifying the Interactive Inhibitory Effect of Heavy Metals on the Growth and Phosphorus Removal of Pseudomonas taeanensis. *Korean Journal of Soil Science and Fertilizer*, 51(1), 35-49. https://doi.org/10.7745/KJSSF.2018.51.1.035

Yumoto, I., Yamazaki, K., Hishinuma, M., Nodasaka, Y., Suemori, A., Nakajima, K., & Kawasaki, K. (2001). Pseudomonas alcaliphila sp. nov., a novel facultatively psychrophilic alkaliphile isolated from seawater. *International Journal of Systematic and Evolutionary Microbiology*, *51*(2), 349-355. doi: 10.1099/00207713-51-2-349

Zheng, Y., & Sriram, G. (2010). Mathematical Modeling: Bridging the Gap between Concept and Realization in Synthetic Biology. *Journal of Biomedicine and Biotechnology, 2010*, 541609. doi:10.1155/2010/541609