

University of West Bohemia
Faculty of Applied Sciences
Department of Cybernetics

BACHELOR THESIS

Analysis and implementation of regulatory mechanisms in
the bacterium *Escherichia coli*

Pilsen, 2012

Pavel Zach

Place for the original assignment

Prohlášení

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

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

V Plzni dne 17.května 2012

Pavel Zach

Declaration

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

Acknowledgements

I would like to thank my thesis supervisor, MSc. Daniel Georgiev, PhD., for bringing me to the exciting field of synthetic biology, for his patient guidance, time and valuable advices and remarks.

Other thanks belong to the Cell cybernetics lab members, in particular Tereza Puchrová, for the assistance with all the experiments performed.

Anotace

Cílem této práce je návrh a implementace zpětnovazebního mechanismu v bakterii *E. coli*, spolu s uvedením potřebných znalostí z biologie, matematického modelování a automatického řízení.

V první části se čtenář seznámí s jednotlivými fázemi genové exprese a možnostmi jejich regulace. Důraz je zde kladen na přehlednost a praktičnost výkladu. Dále následuje uvedení matematických modelů používaných pro modelování genetických regulačních sítí. Nakonec je uvedeno Nyquistovo kritérium stability jako prostředek pro analýzu sensitivity systému.

Druhá část se pak zabývá návrhem zpětnovazebního mechanismu v bakterii *E. coli* za použití biocihel, včetně uvedení potřebných konstrukčních metod. Vytvořený návrh je otestován *in silico* a jeho jednotlivé komponenty experimentálně charakterizovány, stejně jako externí poruchy na něj působící.

Klíčová slova: regulace genové exprese, matematické modely genových regulačních sítí, biocihly, syntetická biologie, buněčná kybernetika, *Escherichia coli*

Abstract

The goal of this thesis is to design and implement a feedback mechanism in the *E. coli* bacterium, as well as present all the needed information from Biology, mathematical modelling and system control theory.

In the first part, the reader is introduced to the different phases of gene expression and to their regulation mechanisms. Mathematical models of gene regulatory networks of various complexity are presented. The Nyquist stability criterion is also reviewed as a tool for determining system sensitivity.

In the second part, the reader is informed about an implementation of a feedback mechanism in *E. coli* bacterium. Autoregulatory gene networks are constructed from standardized BioBrick parts, using herein defined protocols. Implemented designs are tested *in silico* and their elementary parts are experimentally characterized.

Keywords: regulation of gene expression, mathematical models of gene regulatory networks, biobricks, synthetic biology, cell cybernetics, *Escherichia coli*

Contents

1	Introduction	1
2	Biological regulation in <i>E. coli</i>	2
2.1	Gene regulatory networks	2
2.2	Gene expression	3
2.2.1	Transcription	4
2.2.2	Translation	7
2.3	Gene regulation	8
2.3.1	Regulation of transcription and post-transcriptional modifications	9
2.3.2	Regulation of translation and post-translational modifications	11
3	Mathematical models of GRNs	13
3.1	Boolean networks	14
3.2	Continuous networks	14
3.3	Coupled ODEs	15
3.3.1	Mass action kinetics	16
3.3.2	Michaelis–Menten kinetics	16
3.4	Stochastic gene networks	18
4	Robustness analysis tools	18
4.1	Nyquist stability criterion	19
5	Implementation of negative feedback in <i>E. coli</i>	21
5.1	Problem analysis and system design	21
5.1.1	Advantages of negative autoregulation	22
5.1.2	Tuning possibilities	22
5.2	Parallel implementation of the NAR network	22
5.3	BioBrick parts	23
5.3.1	Parts selection	24
5.4	Simulation	26
5.5	Experimental methods	30
5.5.1	Transformation	31
5.5.2	Restriction digest	31
5.5.3	Ligation	33
5.5.4	Verification	33
6	Experimental results	33
6.1	Fluorescent proteins characterization	34
6.2	Characterization of external noise	34
6.3	Inherent LacI characterization	34

7 Discussion	36
7.1 Theory	36
7.2 Simulation	36
7.3 Experiment	36
List of Figures	37
References	39
A List of Abbreviations	43
B Reaction rates	44

1 Introduction

Synthetic biology (synbio) is a new, exciting field of study, interesting both for scientists in natural sciences and engineers.

The research work in this field is quite broad. For example, one of the important topics for biologist is finding how life works. As Richard P. Feynman said, "What I cannot create, I do not understand [9]." By creating life entirely from scratch, we can get deep insight into the origin of life. Chemists interested in synthetic chemistry can use cells as small factories for a production of various chemical compounds, including the ones that are very rare or very expensive to produce by conventional ways in the chemical industry. Engineers aim to design new biological systems as a platform for various technologies, using computer modelling tools and theory from system design and control or electrical engineering. Among interesting projects currently being researched are the production of a cheap anti-malarial drug, antibiotics produced by *de novo* chemical synthesis, bacteria that could soak up carbon dioxide to help reduce global warming and many others [3].

But the heart of synbio is its interdisciplinary character. In one team, engineers can collaborate with chemists and medical doctors in a development of new curing procedures or technology, using genetically modified organisms (GMO). Each specialization brings its own point of view to the problem, thus the solution contains the best that each specialization can offer.

Since the time I started with synbio, I have noticed an interesting trend in the way new systems inside the cells are designed. Engineers (which still form a minority in a scientists participated in synbio) and scientists alike are mainly interested in brute force system design without much consideration for efficiency. Compared to the control system engineering, it is the same situation when a regulator for some process is made, which ensures system stability, but doesn't much optimize the process. The goal of this thesis is to present and test a time-efficient method for tuning synthetic systems inside cells.

The first part of this work shows where and how synthetic biological systems in cells can be tuned in order to increase their efficiency, thus reducing the cost of final industrial products or reducing the time needed for patients to cure. In the second and third part, mathematical models used to describe gene regulatory networks (GRN) will be discussed together with robustness analysis tools from feedback

control theory. Last parts will be dedicated to design and implementation of a feedback mechanism in bacterium *Escherichia coli* (*E. coli*), which can serve as an effective platform for tuning gene expression.

2 Biological regulation in *E. coli*

Before we can take a look at GRNs tuning we must define what a GRN is and review some basic topics from cellular biology. Note that the description of following biological principles is shortened and simplified and is valid for a prokaryotic organisms, such as *E. coli*. More information can be found in the cited sources and molecular biology literature.

2.1 Gene regulatory networks

The behaviour of a cell is governed by its GRN. We can think of a GRN as of a complex MIMO¹ system with many feed-forward and feedback loops, which acts like a cell's "brain". The common inputs to this system are temperature, presence of chemicals in the environment, environmental pH, and many others (see Figure 1 for a schematic of a part of a GRN). The output is in most cases a specific protein. This helps cells to survive in the environment by quickly reacting to changes. For example, in a case of heat shock, the cell detects it and immediately reacts by creating heat shock-proteins, which protect the cell against high temperatures [30].

A GRN consists of deoxyribonucleic acid (DNA) segments which interact with each other and with other substances in the cell. The interactions between DNA segments are indirect, realized by their ribonucleic acid (RNA) and protein products. This governs the rates at which genes in the network are transcribed into messenger RNA (mRNA) [10].

¹Multiple-Input-Multiple-Output

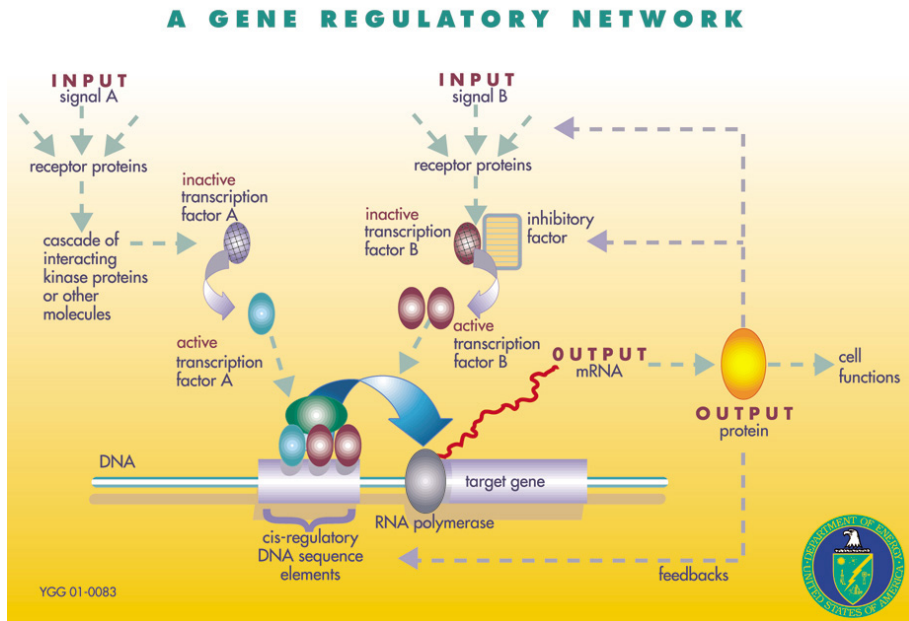


Figure 1: Schematic of a GRN. Image adapted from: Office of Biological and Environmental Research of the U.S. Department of Energy Office of Science. <http://science.energy.gov/ber/>

Engineers can think of this in terms of object-oriented computer programming: the cell is an object and DNA segments are its methods which contain some code. Receptor proteins represent user inputs which trigger the execution of the object's methods and RNAs are attributes of this object which control the object methods' parameters. After a method is executed, values of the attributes and object state are updated and some desired action (creation of a protein) is executed.

The process of this method execution is called gene expression.

2.2 Gene expression

During gene expression a functional gene product is synthesized using information encoded in a gene. The products are mainly proteins, but, in the case of non-protein coding genes, the product is a functional RNA, e.g. ribosomal RNA (rRNA) or transfer RNA (tRNA). This process, common to all cells, is so fundamental that it has been termed the *central dogma* of molecular biology [1] (see Figure 2) and can be divided into two main steps.

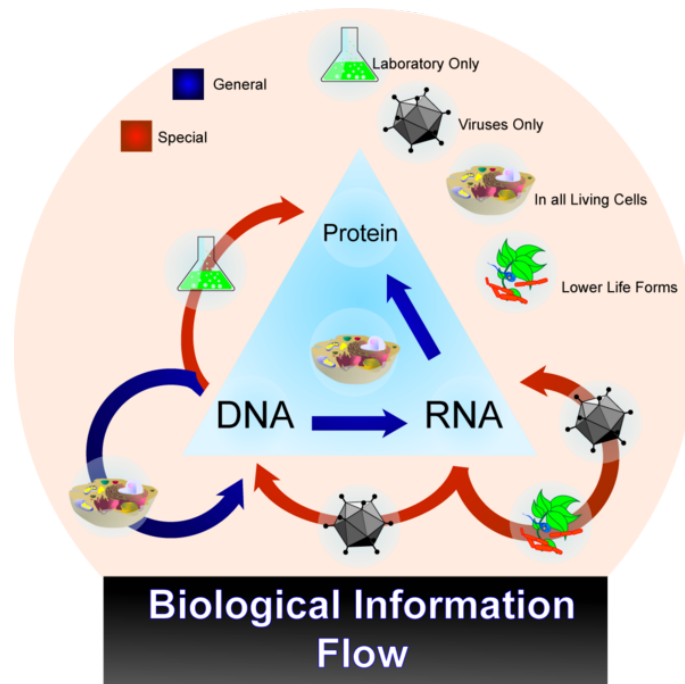


Figure 2: Central dogma of molecular biology - hereditary information is passed from DNA to RNA to proteins but not vice versa [5]. Source: <http://en.wikipedia.org/wiki/File:CDMB2.png>

2.2.1 Transcription

The first step of gene expression is transcription of a DNA sequence, which copies the particular gene into an RNA. To get deeper insight into this process it is important to know the structure both of DNA and RNA.

DNA and RNA Both DNA and RNA are nucleic acids, which consist of nucleotides which are linked together. Five types of nucleotides exist - **A**denine, **C**ytosine, **G**uanine and **T**hymine (only in DNA) or **U**racil (only in RNA). In DNA, these bases pair; adenine pairs with thymine and guanine with cytosine.

DNA consists of two DNA strands which form a double-helix. The first one is used as a template (therefore it is called the template strand, which goes in a $3' \rightarrow 5'$ direction) for RNA polymerase (RNAP) in the production of a complementary, single-stranded RNA. The second strand is called the coding strand and contains the genetic information itself (its sequence is the same as the newly created RNA

transcript, except for the substitution of uracil for thymine) - this strand goes in a 5' → 3' direction.

Thus, during transcription, the information is simply re-written in almost the same language to a different carrying medium. This copying has an important reason - DNA itself is too valuable to be tampered with.

Mechanism Process of transcription can be described in three steps:

1. Initiation

Transcription begins with finding the gene for expression (which is determined by a promoter sequence in the beginning and a terminator sequence in the end - a schematic of one transcription unit is in Figure 3). This is achieved by a protein named σ -factor, one of the RNAP subunits, which provides the RNAP with the ability to recognize specific promoter and is thus essential for the initiation of transcription [14].

Promoter tells the transcription enzymes where to start and is located 30 or so base pairs in front of the gene it controls. Terminator tells the enzymes where to stop [24].

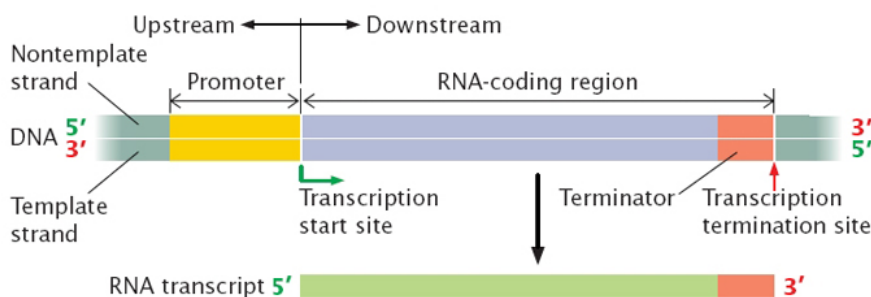


Figure 3: Scheme of a transcription unit. Credits: W. H. Freeman Pierce, Benjamin. *Genetics: A Conceptual Approach*, 2nd ed. (New York: W. H. Freeman and Company)

After the correct transcription unit is found, short stretch of DNA begins to unwind as the hydrogen bonds break and RNAP binds to the promoter. After the RNAP is activated and after the first bond is synthesized, RNAP must clear the promoter.

2. Elongation

As transcription continues, RNAP traverses the template strand from 3' → 5' and uses base pairing complementarity with the DNA template to create an RNA copy from 5' → 3'. In this copy, thymines are replaced with uracils, and the nucleotides are composed of a ribose sugar instead of deoxyribose in the sugar-phosphate backbone).

During RNA transcription more RNAPs can operate on a single DNA template and also the transcription process can be done in more rounds, so many RNA molecules can be rapidly produced from a single copy of a gene. Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases [34].

3. Termination

When the stop sequence (terminator) is found, transcription ends and the RNA molecule is released.

There are two types of terminators in bacteria: Rho-independent (also called intrinsic terminators) and Rho-dependent. The first kind causes polymerase to terminate without the involvement of other factors, the second kind requires an additional protein called Rho to induce termination [34].

Rho-independent terminators consist of two sequence elements: a short inverted repeat (of about 20 nucleotides) followed by a stretch of about eight A-T base pairs. When polymerase transcribes the inverted repeat sequence, the resulting RNA can form a stem-loop structure (also called a hairpin loop) by base pairing with itself. Hairpins are believed to cause termination by stopping the elongation complex by disrupting the A-U base pairs, which are the weakest.

Rho-dependent terminators require an additional ring-shaped protein called Rho to induce termination.

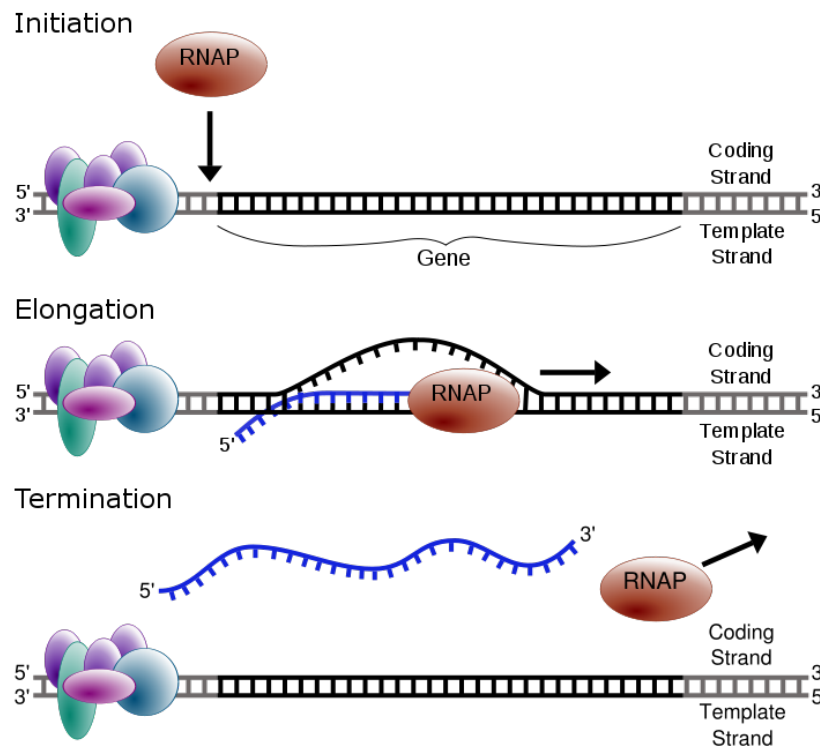


Figure 4: Three stages of a transcription. Credits: Forluvoft (Own work) [Public domain], via *Wikimedia Commons*

2.2.2 Translation

After the protein coding sequence is copied from DNA to mRNA, it is time for a second step - the creation of a protein (translation). During translation the sequence of codons (each made up of three nucleotide bases) in mRNA is converted by a ribosome into a corresponding sequence of amino acids that will later fold into an active protein [21]. In bacteria, translation occurs in the cell's cytoplasm, where all the subunits of the ribosome are located.

Mechanism Process of translation can be also divided into three steps [17]:

1. Initiation

First, small subunit of the ribosome binds to a site "up-stream" (on the 5' side) of the start of the mRNA and proceeds downstream (5' → 3') until it encounters the start codon AUG. Here it is joined by the large subunit and a special initiator tRNA, which binds to the P site on the ribosome.

2. Elongation

An aminoacyl-tRNA (a tRNA covalently bound to its amino acid) is able to base pair with the next codon on the mRNA which arrives at the second site on the ribosome - the A site. The preceding amino acid is then covalently linked to the incoming amino acid with a peptide bond. The initiator tRNA is released from the P site and the ribosome proceeds downstream (5' → 3'), repeating this process codon after codon.

3. Termination

If a codon UAA, UAG or UGA (for these STOP codons there are no tRNA molecules with anticodons - see Figure 6) is found, polypeptide is finished and proteosynthesis ends. The polypeptide is released from the ribosome and the ribosome splits into its subunits, which can be later reassembled for another round of protein synthesis.

2.3 Gene regulation

These are the basis of gene expression. We can now proceed to explore the means by which this process is and can be regulated.

The precise regulation of gene expression is crucial for the survival of a cell. The number of genes in a variety of bacteria vary from 700 to nearly 6000, but only about 600 – 800 of them are needed at any one time. Expressing all of them would be useless and, more importantly, very expensive in terms of energy (about 3000 ATP molecules per protein) [6]. So, typically genes are switched on and off in response to the need for their product.

This regulation is realized by regulatory elements, which can be divided into two groups:

- **Cis-regulatory elements** - they are present on the same molecule of DNA as the gene they regulate. These are typically located on non-coding regions of DNA.
- **Trans-regulatory elements** - they can regulate genes far away from their coding gene (usually a protein that is used in the regulation of another gene).

2.3.1 Regulation of transcription and post-transcriptional modifications

As mentioned above, gene expression is very expensive for a cell so the lower level the regulation takes place, the better.

Promoter strength (cis) The strength of a promoter is determined by how well its elements match the optimum "consensus" sequences. In the absence of regulatory proteins, these elements determine the efficiency with which polymerases bind to the promoter and, once bound, how readily they initiate transcription [34].

Activators and Repressors (trans) Most genes are controlled by extracellular signals (these signals are typically molecules present in the environment outside the cell) which are communicated to them by regulatory proteins. These regulatory proteins can work in two ways: as positive regulators, or activators; and as negative regulators, or repressors. Activators increase transcription of the regulated gene; repressors decrease or block transcription [34].

These regulators are often DNA-binding proteins that recognize specific sites at or near the genes they control. Depending on a promoter, one of the following methods takes place [34]:

- At many promoters, the RNAP binds only weakly (in the absence of regulatory proteins), therefore undergoes a transition to an open complex and starts transcription only sometimes. This low level of expression is called the basal level. In further prevent to prevent transcription, a repressor needs to bind to an operator and block RNAP elongation. To further increase transcription, an activator can help polymerases bind to the promoter. This is achieved by using one of the activator's surfaces to bind a site on the DNA and another surface to interact with the polymerase.

This mechanism is called recruitment and it is an example of cooperative binding of proteins to DNA.

- In the second case, RNAP binds efficiently unaided to the promoter and forms a stable closed complex. This closed complex does not spontaneously start elongation, an activator must stimulate this.

The activator works by triggering a conformational change in either the polymerase or the DNA. It interacts with the stable closed complex and

induces a conformational change to the open complex. This mechanism is an example of allostery.

Regulation by σ -factors (trans) As described in the transcription section, σ -factors are responsible for the recognition of the promoters. For example, σ^{70} (because it is about 70 kDA in size in E. coli) is responsible for recognition of promoters used by genes required during the exponentially growth phase (these are sometimes called "housekeeping" genes since they encode essential functions needed for the cell cycle and for normal metabolism).

But in most cases, there are several different σ -factors present in bacteria. These alternative σ -factors, which complement the primary sigma factors, allow bacterium to bring about global changes in gene expression in response to particular environmental stresses (see Figure 5).

For example, 30 heat shock genes, which express proteins that protect the cell against high temperatures, are only recognized by RNA polymerase containing factor σ^{32} , which has longer half-life at higher temperatures [6].

			Sigma factor	Regulated genes
-35 region	Spacer	-10 region		
TTGACA	N ₁₆₋₁₈	TATAAT	σ^{70}	General housekeeping genes
CTTGAAA	N ₁₃₋₁₅	CCCATNT	σ^{32}	Heat shock genes
CTGGNA	N ₆	TTGCA	σ^{54}	Nitrogen assimilation genes
		CTACACT	σ^{38}	Stationary phase/General stress genes
TAAA	N ₁₅	GCCGATAA	σ^{28}	Flagella genes

Figure 5: Alternative sigma factors and promoter recognition sequences in E. Coli. Credits: DALE, Jeremy and Simon PARK. *Molecular genetics of bacteria*. 4th ed. Hoboken

mRNA stability (cis) Most bacterial mRNA is typically being degraded with a half-life of about 2 min, which is a relatively short time. This instability of mRNA provides bacteria with the ability to respond to changes in their environment rapidly. However, some bacterial mRNA species are more stable than others, in some cases with a half-life as long as 25 min [6].

Regulatory RNA (trans) One of the posttranscriptional modifications is carried out by a regulatory, non-coding RNAs (ncRNAs). These ncRNAs don't encode proteins, but act as riboregulators, which regulate gene expression.

In many cases, these ncRNAs are "anti-sense" RNAs - aRNAs. If a region of a gene, particularly the region including the ribosome binding site and translation initiation point, is transcribed in the opposite direction, an RNA molecule will be produced that is complementary to the mRNA. This molecule can hybridize to the mRNA, and thus block the binding of ribosomes and the initiation of translation [6].

2.3.2 Regulation of translation and post-translational modifications

Ribosome binding (cis) Ribosome binding plays a comparatively minor role in the natural control of gene expression in bacteria, because it would be rather wasteful to produce large amounts of mRNA that are not required for translation. Translational control can however become important with genetically-engineered bacteria, when very high levels of transcription of a specific gene have been achieved.

The ribosome binding is controlled by the distance separating the ribosome binding site (RBS) from the initiation codon (start codon - AUG, see Figure 6). The sequence of an RBS does not seem to affect the level of translation [6].

Codon usage (cis) As shown in Figure 6, most amino acids are coded by more codons. In most cases, these codons aren't effectively equivalent, since a different tRNA species is responsible for recognition of the different codons.

Some of these tRNA species are known to be present in the cell at quite low levels, thus a gene that contains many codons that require these "rare" tRNA molecules will then be expected to suffer delays in translation that may affect the amount of the end-product formed [6].

		Second base in codon				
		U	C	A	G	
First base in codon	U	Phe	Ser	Tyr	Cys	U
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	STOP	STOP	A
		Leu	Ser	STOP	Trp	G
C	Leu	Pro	His	Arg	U	
	Leu	Pro	His	Arg	C	
	Leu	Pro	Gln	Arg	A	
	Leu	Pro	Gln	Arg	G	
A	Ile	Thr	Asn	Ser	U	
	Ile	Thr	Asn	Ser	C	
	Ile	Thr	Lys	Arg	A	
	Met	Thr	Lys	Arg	G	
G	Val	Ala	Asp	Gly	U	
	Val	Ala	Asp	Gly	C	
	Val	Ala	Glu	Gly	A	
	Val	Ala	Glu	Gly	G	

Figure 6: The standard genetic code showing amino acids for all 64 possible codons. Credits: *CLC bio*, http://www.clcbio.com/scienceimages/genetic_code.png

Protein stability (cis) Different proteins vary in their stability to a very marked degree, as might be expected from their different functions: a protein that forms part of a cellular structure is likely to be more stable than one that transmits a signal for switching on a transient cellular event [6].

Phosphorylation and dephosphorylation Function of proteins can be altered or switched on and off by an addition of a phosphate group (PO_4^{3-}). This process called phosphorylation, catalysed by enzymes called protein kinases, plays a significant role in a wide range of cellular processes. The addition of a negatively charged phosphate changes the characteristics of the protein, often by a conformational change in the protein structure. This change can increase or decrease the biological activity of an enzyme, help to move proteins between subcellular compartments, or allow interactions between proteins to occur as well as label them for degradation [27].

This process is fully reversible by a process called dephosphorylation. During this process, the phosphate is removed and the protein switches back to its original conformation. If these two conformations provide the protein with different

activities, phosphorylation of the protein will act as a molecular switch, turning the activity on or off [27]. This ability and other advantages - phosphorylation is very quick (it takes only a few seconds) and it does not require new proteins to be made or degraded, makes the phosphorylation a key player in a response to extracellular signals.

Methylation Methylation is an addition of a methyl group (or a substitution of an atom or another group by a methyl group) to a substrate. As phosphorylation, methylation is also catalysed by enzymes and effects regulation of gene expression (by inactivation of genes) and proteins function (by triggering their conformational changes).

Methylation also serves in many bacteria as a primitive immune system, allowing them to protect themselves from infection by a bacteriophage (a bacteria virus). This is achieved by the enzyme methylase which periodically methylates adenosine or cytosine in the bacterial DNA near specific sequences. Foreign DNA that are introduced into the cell are not methylated and can thereby be degraded by sequence-specific restriction enzymes.

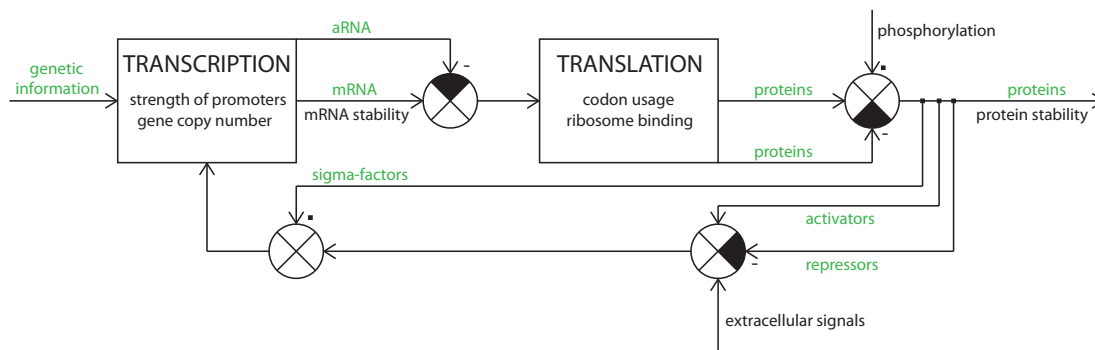


Figure 7: Block diagram of gene expression regulation

3 Mathematical models of GRNs

The processes of gene regulatory networks can now be organized into mathematical models giving further insight into cellular operations. Mathematical models of GRNs have been developed to describe both gene expression and regulation, and in some cases generate predictions that support experimental observations.

3.1 Boolean networks

One of the simplest methods for modelling GRNs is with Boolean networks. In these networks, genes are modelled using digital switches, which are characterised by the fact that they can only be in one of two states (off/on - "on" corresponds to the gene being expressed) with propagation delays (the time between the signal appearing at the input and the corresponding response at the output). Complex networks can be modelled [28] by combining switches to create logic gates (modules performing logical operations such as AND or NOT) and other functions.

Boolean networks are represented by a directed graph, where each gene, each input, and each output are denoted by a node. An arrow from one node to another is present if and only if there is a direct connection between the two nodes. Time is viewed as proceeding in discrete steps. At each step, the new state of a node is a Boolean function of the neighbouring upstream nodes.

Boolean models can provide qualitative insights but in general have lower predictive power with respect to their continuous counterparts - in the real world transcription rates may be anywhere on a continuous scale between 0 and maximal, and this can have important consequences for the rate at which other genes are transcribed, and hence for the dynamics of the network [28].

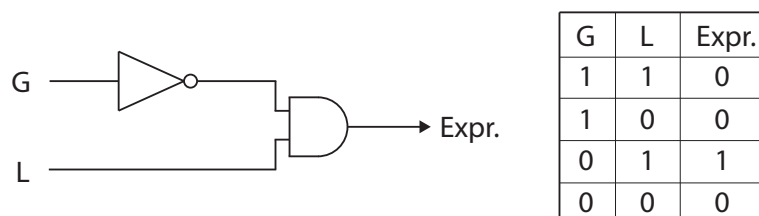


Figure 8: Example of Boolean network of lac operon - only expressed if glucose is absent and lactose is present. In fact, this scheme with one invert and one AND gate represents the Boolean inhibit function.

3.2 Continuous networks

One possible extension to Boolean networks is to let time and gene expression levels be continuous variables, while the influence among genes is still represented by switching functions. This enables us to see several properties of GRNs that cannot be captured by Boolean models.

A continuous network model was first proposed by L. Glass. This model denoted the activation of gene i with the real variable x_i [25]. We can associate a Boolean variable X_i to this x_i to get

$$X_i(t) = \begin{cases} 0 & , \quad x_i(t) < \theta_i \\ 1 & , \quad otherwise \end{cases}$$

where θ_i is the threshold value for x_i .

In a network with N nodes, each with K inputs we can define the activation rate of node i as:

$$\frac{dx_i}{dt} = -\tau_i x_i + f_i(X_{i_1}(t), X_{i_2}(t), \dots, X_{K_1}(t)), i = 1, 2, \dots, N \quad (1)$$

where τ_i is a decay parameter and f_i is a Boolean function of the inputs of node i . More information can be found in [25].

3.3 Coupled ODEs

Dynamical properties of reaction networks can be described by coupled ODEs, considering the next two assumptions [31]:

1. Component concentrations are homogeneous in the reaction space. This assumption holds for simpler organisms at longer time scales but may fail for more complex organisms possessing cytoskeletal compartments.
2. Variables representing chemical concentrations are continuous functions of time. This is achieved if the number of molecules of each species are sufficiently large.

The GRN is assigned a time-invariant, ordinary state-space model. The concentration of a chemical species i is denoted by the state X_i and a reaction j is modelled by a reaction rate $r_j(X)$, which can be a general non-linear function of X . Each reaction j is also described by its stoichiometric coefficients c_{ij} that state how many units of X_i are produced ($c_{ij} > 0$) or annihilated ($c_{ij} < 0$) each time the reaction occurs.

In general, each equation is in form of

$$\begin{aligned} \frac{dX_i}{dt} = & \text{(rates of } X_i \text{ synthesis)} - \\ & \text{(rates of } X_i \text{ annihilation)} = \sum c_{ij}r_j(X) \end{aligned} \quad (2)$$

The right expressions and the rates of all reactions are often derived from the Law of mass action or from the Michaelis-Menten kinetics.

3.3.1 Mass action kinetics

The law of mass action, first proposed by C. M. Guldberg and P. Waage in 1864, states that the rate of a chemical reaction is directly proportional to the molecular concentrations of the reacting substances [19]. Thus, for the reaction



where the rate constant k expresses the probability that the molecules are well oriented and have sufficient energy to react [13], the rate law is

$$v = kAB \quad (4)$$

The change of the concentrations in time is given by

$$\frac{dA}{dt} = \frac{dB}{dt} = -kAB \quad \text{and} \quad \frac{dC}{dt} = \frac{dD}{dt} = kAB \quad (5)$$

3.3.2 Michaelis–Menten kinetics

Assuming we have an enzymatic transformation (with enzyme E) of substrate X into a product P :



In 1913, Michaelis and Menten have proposed a new mechanism for this process. First, enzyme E binds to a substrate X to form a complex C . In this complex, E converts X to P , dissociates from P and continues to a beginning of the reaction.

This reaction scheme can be written as follows:



In their analysis, Michaelis and Menten made one important assumption - the substrate S is in instantaneous equilibrium with the complex C (i.e. $k_1, k_{-1} \gg k_2$) [13]. With this assumption, we can write

$$k_1 ES = k_{-1}C \quad (8)$$

and since the total enzyme concentration $E_T = E + C$, we can express C as

$$C = \frac{E_T S}{\frac{k_{-1}}{k_1} + S} \quad (9)$$

The rate of a production of P follows the mass action law, thus

$$\frac{dP}{dt} = k_2 C = V_{max} \frac{S}{K_S + S} \quad (10)$$

But in 1925, Briggs and Haldane suggested another assumption - the total enzyme concentration is much less than the initial substrate concentration ($E_T \ll X_0$) [31]. With this assumption, a steady state in which the concentration of ES is essentially constant is made very shortly after mixing E and S [13].

This quasi-steady state approximation implies that

$$k_1 ES - k_{-1}C - k_2 C = 0 \quad (11)$$

We can express C (with the use of $E_T = E + C$):

$$C = \frac{k_1 E_T S}{k_1 S + k_{-1} + k_2} = \frac{E_T S}{S + \frac{k_{-1} + k_2}{k_1}} \quad (12)$$

Now we can write again the expression of the production rate for P in (10)

$$\frac{dP}{dt} = k_2 C = \frac{k_2 E_T S}{S + \frac{k_{-1} + k_2}{k_1}} = V_{max} \frac{S}{S + K_M} \quad (13)$$

where $K_M = \frac{k_{-1} + k_2}{k_1}$ is a Michaelis constant and $V_{max} = k_2 E_T$.

The use of the Michaelis-Menten kinetics is for reducing the number of variables which describe an enzymatic conversion process, such as phosphorylation or dephosphorylation [31].

3.4 Stochastic gene networks

All events in a cell, including gene expression [8], depend directly or indirectly on probabilistic collisions between molecules [22], thus exhibiting stochastic behaviour.

Since the magnitude of stochastic fluctuations (for a single reaction) scales with $1/\sqrt{N}$, the accuracy of deterministic description depends on the number of molecules N [4]. Generally, deterministic models give us good approximation of reactions having $> 10^2$ – 10^3 molecules per reactant.

The main disadvantage of this approach is high computational complexity.

4 Robustness analysis tools

Robustness is an important properties of both engineered and biological systems. While robustness is often used synonymously with system stability, in cybernetics it is defined as insensitivity to specific disturbances. Microscopic biological systems face many sources of uncertainty and hence robustness plays a key role in their design.

Various methods for determining systems stability exist. Herein we will describe one of the methods referred as the Nyquist stability criterion.

As an examples, we will consider this simple linear, time-invariant feedback-loop system:

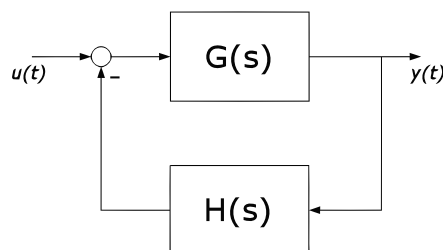


Figure 9: Block diagram of simple linear, time-invariant feedback-loop system

This system has an input $u(t)$, an output $y(t)$, an open loop transfer function $G(s)$ and a feedback transfer function $H(s)$.

The closed loop transfer function of this system is

$$F(s) = \frac{G(s)}{1 + G(s)H(s)}. \quad (14)$$

We can also compute the sensitivity transfer function with respect to changes in plant G :

$$S(s) = \frac{\partial F(s)/F(s)}{\partial G(s)/G(s)} = \frac{1}{1 + G(s)H(s)}. \quad (15)$$

For both of these transfer functions, the case $G(s)H(s) = -1$ represents instability and high sensitivity.

4.1 Nyquist stability criterion

Nyquist stability criterion is based on an analysis of the Nyquist plot of the open loop system. From this plot we can tell if the closed loop system will be stable or not.

Advantage of this approach is that we do not need to explicitly compute the poles and zeros of either the closed-loop or the open-loop system, so it can be easily applied to systems defined by non-rational functions, such as systems with delays.

The transfer function for our example system without the feedback loop is $G(s)$, which can be defined as

$$G(s) = \frac{Q(s)}{P(s)} \quad (16)$$

where $Q(s)$ and $P(s)$ are polynomials of degree m and n (for real $m \leq n$).

Now we can create a function $D(s)$ which is equal to the denominator of a transfer function of the closed-loop system (take $H(s) = 1$)

$$D(s) = 1 + G(s) = \frac{P(s) + Q(s)}{P(s)} \quad (17)$$

The denominator of this function is equal to the characteristic polynomial of the open-loop system $G(s)$ and the numerator is equal to the characteristic polynomial of the closed-loop system.

Stable system must have all roots of the characteristic polynomial of the closed-loop system ($P(s) + Q(s) = 0$) in the left complex half-plane - we can write it with use of the argument principle as

$$\Delta \arg_{-\infty < \omega < \infty} [1 + G(j\omega)] = 2\pi \cdot p \quad (18)$$

where p is the number of poles of $G(s)$ in the right half-plane of the complex plane.

Now we can define the Nyquist stability criterion:

Nyquist stability criterion. *The closed-loop system is stable if the frequency characteristic of the open-loop system $G(j\omega)$ in a complex plane travels in a positive direction around point $[-1; 0j]$ as many times as how many poles have the transfer function $G(s)$ in a right complex half-plane.*

We can create the Nyquist plot by construction of a vector $[1 + G(j\omega)]$ which has the beginning in the point $[-1; 0j]$ and the end is moving along frequency characteristic $G(j\omega)$.

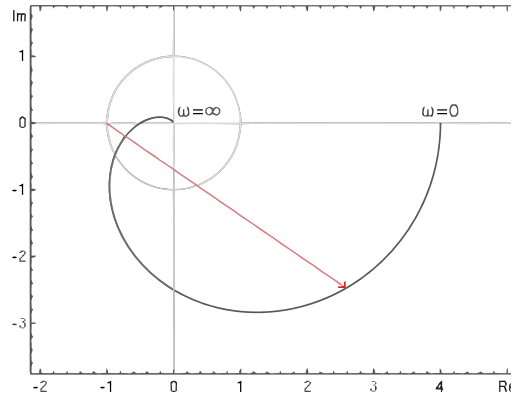


Figure 10: Example of the Nyquist plot, vector $[1 + G(j\omega)]$ is red. If the open-loop system $G(s)$ is stable, this system's closed-loop is stable too.

In practice, we can often simplify this criterion. If the open-loop system $G(s)$ is stable, than the closed-loop system is stable if the argument change in the (18) is equal to zero. In this case, we do not need to count the circles around the point $[-1; 0i]$. For stability it is sufficient if this point is located to the left of the frequency characteristic $G(j\omega)$.

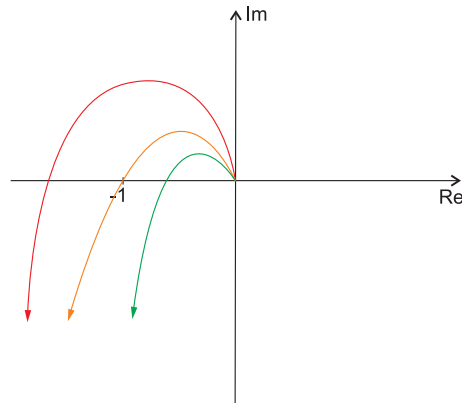


Figure 11: Frequency characteristics $G(j\omega)$ for different gains. The red is unstable, the orange is marginally stable and the green is stable

This approach gives us a detailed view of the system stability and sensitivity for all values of ω .

5 Implementation of negative feedback in *E. coli*

In this section, mathematical modelling of the biological elements introduced in Section 3 is used to guide the implementation of a synthetic gene regulatory network in *E. coli*.

This work is done as one of my research project in the Cell Cybernetics Lab² concerned with the construction of iterative algorithms and experimental procedures for tuning a negative autoregulatory (NAR) transcription network to yield desired protein levels with minimal sensitivities to likely perturbations.

5.1 Problem analysis and system design

Various regulatory elements of gene expression (e.g., promoters, regulatory proteins) are well characterized. Changes to these elements that influence the expression rates have also been identified and quantified. Every new design, however, requires specific adjustments. Assaying all possible designs is difficult due to the number of available combinations. Hence, what is still missing is a platform for systematic tuning of transcription networks [12].

²Cell Cybernetics Lab, Department of Cybernetics, University of West Bohemia, Pilsen, Czech Republic

The underlying project idea is to use multiple implementations of a given promoter design in a given cell. By realizing competing designs at a cell, we can practically eliminate the effects caused by the environmental noise. This approach is used to tune the negative autoregulatory network.

5.1.1 Advantages of negative autoregulation

As was shown in [26], negative autoregulation serves in transcription networks to speed up the response. In an experiment with two designs, which were set to achieve an equal quasi-steady state (i.e. weak promoter to the unrepressed circuit and strong promoter to the design with NAR), the rise-time of the design with NAR was about one fifth of the design without NAR.

Negative autoregulation also increases robustness to many perturbations and for this ability it is very frequent motif found to provide robustness, for example, in gene regulatory networks [11].

5.1.2 Tuning possibilities

There are several ways how to tune the parameters of this system. The first and most important is to regulate the strength of promoters. As was mentioned in section 2.3.1, the promoter strength can be strongly influenced by changes in its coding sequence, mainly by a changes of a gaps between certain sequences.

Other possibility is to use plasmid vector with different replication origin. This influences the gene copy number, and indirectly the speed with which the proteins are produced.

On a translational level, changing the distance between RBS from the initiation codon can influence the speed of a translation.

5.2 Parallel implementation of the NAR network

The proposed system consists of one, auto inhibiting gene which produces a certain repressor protein. This protein also represses the transcription of a different gene which encodes a certain reporter protein and is transcribed from an alternate promoter. In such a way, other alternate designs can be placed inside the same cell. Note, all genes are on the same plasmid³, they have the same copy numbers

³Plasmid is double-stranded DNA molecule which is separated from the chromosomal DNA, thus can be easily transferred between cells

and are transcribed together. The schematic of this system is in Figure 12.

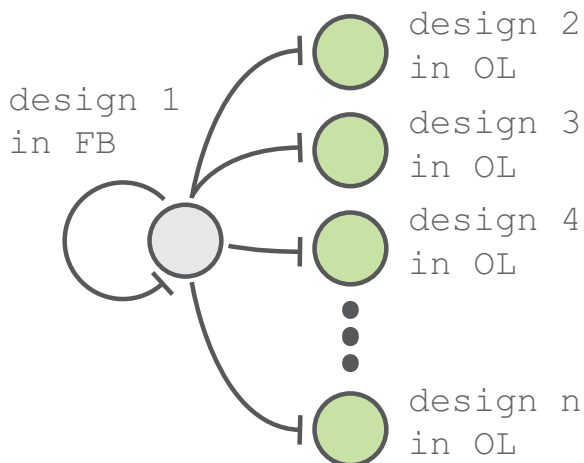


Figure 12: Illustration of our system. All parallel design are under control of one autoinhibiting gene. This parallel design testing ensures that external noise effects influence all designs similarly and that internal noise is averaged out [12]

5.3 BioBrick parts

For the implementation of this system, standard biological parts (BioBrick parts) were chosen. BioBrick parts are DNA sequences of a precisely defined structure and function, with defined prefix and suffix. This allows their users to combine them together on a plasmid and then incorporate the newly created composite part into living cells (mainly into *E. coli*).

All the parts are organized in a catalog, which is available on-line at the Registry of Standard Biological Parts website [23]. Parts in the catalog are divided into groups by their type and function (Figure 13). Each categorized part has its own unique identification number, description and well-characterized DNA sequence. Since the catalog is maintained by a open Synthetic Biology community, users share their experience with all the parts they used and everyone can contribute by submitting his own part.

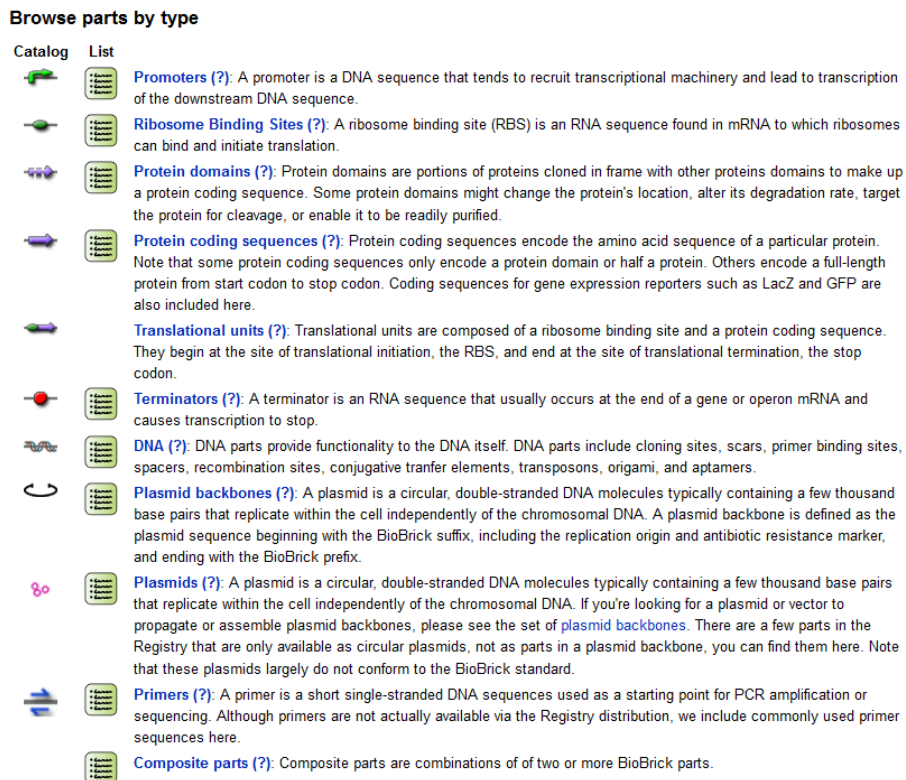


Figure 13: Main page of the catalog, allowing users to browse part by type

Parts from the catalog are distributed for free in DNA Distribution Kit Plates. The main purpose of this kit is to serve as a starting platform for the annual iGEM competition [16].

5.3.1 Parts selection

For our system, we needed to find the following parts:

- Repressor DNA coding sequence
- Promoter which is negatively regulated by this repressor protein
- Different promoter which is negatively regulated by the same repressor protein
- DNA coding sequence of a target protein
- DNA coding sequence of a reporter protein

We have chosen Lac repressor as our repressor protein. The Lac repressor is one of the best characterized repressors in *E. coli*. It is the product of the naturally occurring LacI gene. In a wild-type *E. coli*, this protein inhibits the transcription of the Lac operon by binding to a DNA sequence known as the Lac operator. By fusing the Lac operator with the LacI coding sequence downstream of a constitutive promoter, we can construct an autoinhibiting gene (see Figure 12). Both parts, the LacI gene and a promoter with the LacI operator can be found in the Spring 2011 DNA distribution kit plates⁴.

Fluorescence is a common method for determining the amount of a given protein in a cell. Each of the alternate designs corresponds to a different fluorescence protein. Intact promoter and coding sequences with the respective LacI operators are available in the catalog.

With regards to the specifications of our fluorescence spectrophotometer (the wavelength ranges of the emission and excitation optical filters), three parts were chosen:

1. **Green fluorescent protein (GFP)** generator regulated by the same promoter as we have chosen for LacI
2. **Red fluorescent protein (RFP)** generator again regulated by the same promoter as we have chosen for LacI
3. **RFP** generator regulated by a different promoter which can be also inhibited by Lac repressor

In this initial work available parts were used and therefore the number of parts to be constructed was reduced. Further tuning, however, would require specific synthesis of the alternate promoter designs.

Last part, the autoinhibitive gene was also used for calibration measurements by comparing the fluorescence intensity of GFP and RFP with identical promoters to normalize their values. This allows us to compare the amounts of both of the fluorescent proteins.

The final parts were:

- Lac repressor inhibited promoter (**p1**): BBa_R0010

⁴The LacI gene was already equipped with a RBS site at the beginning and two terminator sequences at the end.

- LacI coding sequence with RBS and terminators (**LacI**): BBa_I732820
- LacI regulated GFP generator (**p1GFP**): BBa_K082034
- LacI regulated RFP generator (**p1RFP**): BBa_J04450
- LacI regulated RFP generator with a different promoter (**p2RFP**): BBa_J5526

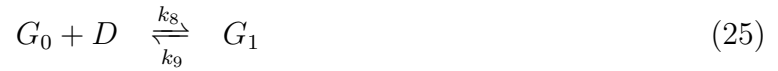
5.4 Simulation

System behaviour was first tested *in silico* using the final model of the underlying chemical reaction network.

First, a set of chemical reactions describing our systems was created. Then, following the Mass-action kinetics rules, these reactions were transformed into system ODEs. Used reaction rates can be found in Appendix B.

Chemical reactions for p1+LacI

$$G_T = G_0 + G_1 + G_2 \quad (19)$$



$$\dot{R} = k_1(G_T - G_1 - G_2) - k_2R \quad (27)$$

$$\dot{L} = k_3R - k_4L - 2k_5L^2 + k_6D \quad (28)$$

$$\begin{aligned} \dot{D} = & 2k_5L^2 - k_6D - k_8(G_T - G_1 - G_2)D - k_{10}G_1D - k_7D + k_9G_1 + \\ & + k_{11}G_2 \end{aligned} \quad (29)$$

$$\dot{G}_1 = k_8(G_T - G_1 - G_2)D - k_9G_1 - k_{10}G_1D + k_{11}G_2 \quad (30)$$

$$\dot{G}_2 = k_{10}G_1D - k_{11}G_2 \quad (31)$$

Chemical reactions for p1GFP



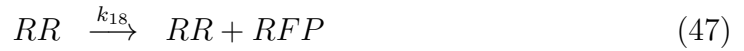
$$\dot{GF}_1 = k_8(G_T - GF_1 - GF_2)D - k_9GF_1 - k_{10}GF_1D + k_{11}GF_2 \quad (39)$$

$$\dot{GF}_2 = k_{10}GF_1D - k_{11}GF_2 \quad (40)$$

$$\dot{RF} = k_{12}(G_T - GF_1 - GF_2) - k_{13}RF \quad (41)$$

$$\dot{GFP} = k_{16}RF - k_{17}GFP \quad (42)$$

Chemical reactions for p1RFP



$$G\dot{R}_1 = k_8(G_T - GR_1 - GR_2)D - k_9GR_1 - k_{10}GR_1D + k_{11}GR_2 \quad (50)$$

$$G\dot{R}_2 = k_{10}GR_1D - k_{11}GR_2 \quad (51)$$

$$R\dot{R} = k_{14}(G_T - GR_1 - GR_2) - k_{15}RR \quad (52)$$

$$R\dot{P}F = k_{18}RR - k_{19}RFP \quad (53)$$

By combining these three models, we can simulate the behaviour of the physical composite parts. First, we tested the effects of IPTG induction on **p1Lp1Gp1R**. IPTG binds to the Lac repressor and inactivates it, hence the Lac repressor can't inhibit expression of genes under control of the **p1** promoter. Expectation from the IPTG induction simulation should be an increase in fluorescent protein concentrations with an increasing amount of IPTG. This was confirmed: as we can see in Figure 14, fluorescent proteins level were increasing with the increasing IPTG concentration. Also, since both genes were on the same plasmid, both GFP and RFP had the same steady-state concentration.

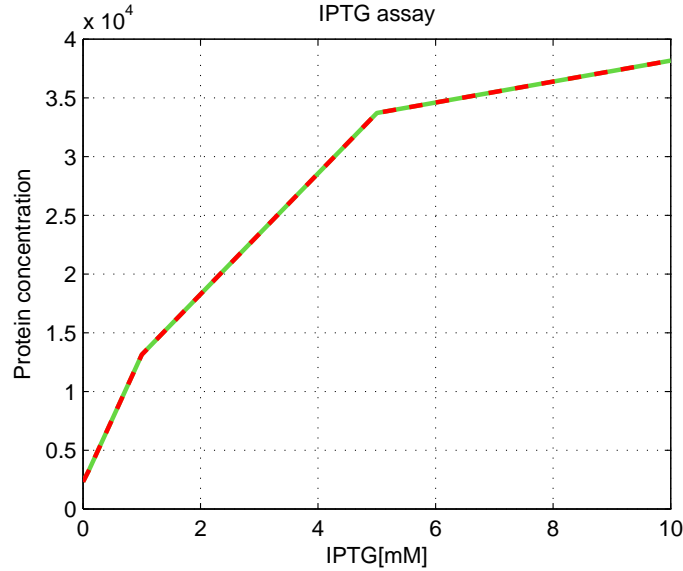


Figure 14: IPTG assay simulation results for **p1Lp1Gp1R**. With increasing level of IPTG, increase in fluorescent proteins concentrations can be seen (p1R is the dashed red line, p1G is the solid green line).

The reduction of external noise in parallel designs was simulated using **p1Lp1Gp1R** and **p1Lp1G** with **p1Lp1R** plasmids. Transcription rates of the corresponding genes were subjects to random perturbations, corresponding to temperature changes, enzyme fluctuations, etc.

As shown in Figure 15, parallel design (i.e. **p1Lp1Gp1R**) didn't show any protein variations in fluorescent proteins concentrations, while the serial design (**p1Lp1G** and **p1Lp1R** on a different plasmid) showed large differences in steady-state protein concentrations. This proves that in the parallel design, an influence of external noise is efficiently removed.

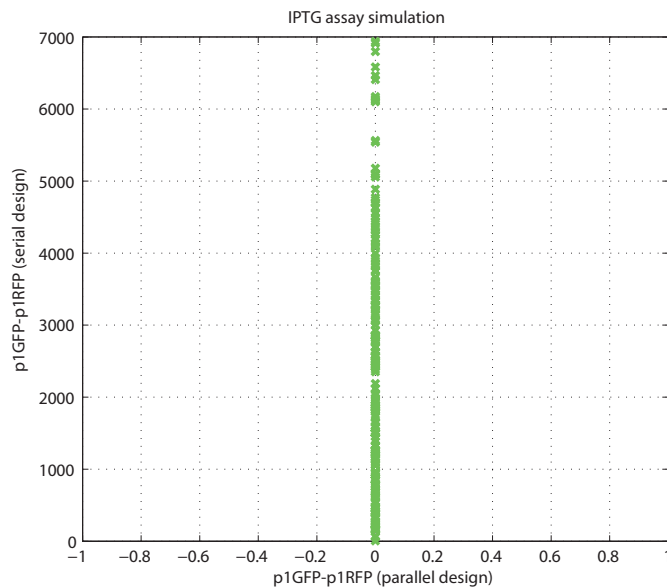


Figure 15: Effects of different design methods on external noise. Parallel design didn't show any differences in fluorescent protein concentrations (x-axis), while concentrations of proteins in the serial design show large variation (y-axis).

To verify better sensitivity to perturbations in NAR, we plotted the Nyquist diagram of linearized the **p1L** system. Linearized model was obtained from equations (27) - (31), using the total gene number G_T as the system input and uninhibited gene number (G_0) as the system output.

The obtained Nyquist diagram shows lowest sensitivity to perturbations at low frequencies, and increased sensitivity at higher frequencies. The sensitivity is the greatest at a given finite frequency.

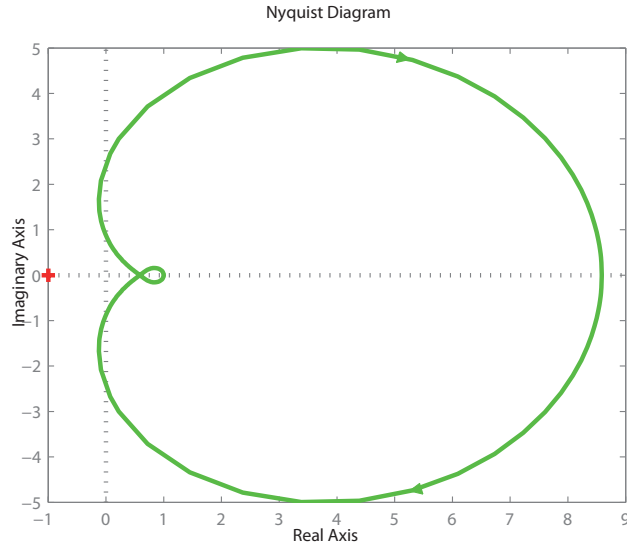


Figure 16: Nyquist plot of the linearized **p1L** system

5.5 Experimental methods

Recombination was used to construct the following four plasmids

1. Autoinhibiting gene **p1LacI** with **p1GFP** and **p1RFP** generators⁵
2. Autoinhibiting gene **p1LacI** with **p1GFP** and **p2RFP** generators
3. Autoinhibiting gene **p1LacI** with **p1GFP** generator
4. Autoinhibiting gene **p1LacI** with **p2RFP** generator

All these systems can be assembled from the parts we have selected in Section 5.3.1. Since we can join only two parts at the same time, this process will consist of two steps. For the first one, we need to join together the following parts:

1. **p1 + LacI** - to get the autoinhibiting gene
2. **p1GFP + p1RFP** - the purpose of this ligation is to place both parts on the same plasmid (see Section 5.1)
3. **p1GFP + p2RFP**

For the construction in 1, we can describe the DNA synthesis process.

⁵Generator is a part that contains assembled promoter, RBS, gene coding sequence and terminators

5.5.1 Transformation

Since the amount of DNA that comes with the DNA Distribution kit isn't enough for assembly, the first thing we need to do is to transform this DNA into cells and then make our own stocks with sufficient amount of DNA. The principle of this step is to put a small amount of DNA into competent cells, let them grow (replicate) and then harvest back the amplified DNA. This can be achieved by the following protocols from the Registry of Standard Biological parts website [15].

5.5.2 Restriction digest

When we have sufficient amount of purified⁶ DNA, we can proceed to restriction digest both parts.

All BioBrick parts come on a plasmid vector with the form shown in Figure 17. These plasmids contain the replication origin, which is responsible for the replication of plasmids during cell growth and division and influences plasmid copy numbers per cell. The antibiotic resistance marker is used as the selective agent for cells that contain this plasmid.

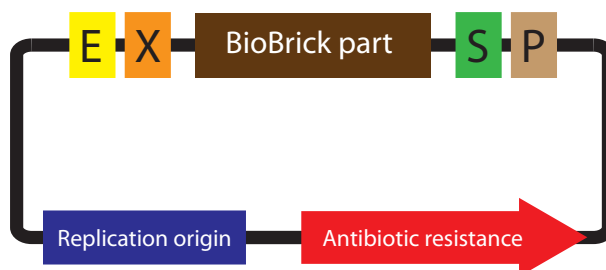


Figure 17: Schematic of a BioBrick plasmid

The cloning site, labelled by the letters E, X, S and P, is recognized by appropriate restriction enzymes which are able to cut the DNA. Following the restriction digest [18], both sides of the cleaved DNA contain "sticky" ends, as shown in Figure 18.

⁶After extraction from the cells, DNA needs to be purified from salt and enzymatic residues with the use of some commercial purification kit



Figure 18: "Sticky" ends after restriction digest by restriction enzyme E. Specific sequence GAATTC is found and cut on both DNA strands.

In our example, **p1+LacI**, the p1 insert is placed in front of the **LacI** coding sequence on the LacI vector (vector is the portion of a plasmid outside of the insert). Therefore we need to cut **p1** at the sites E & S and **LacI** at the sites E & X.

After this procedure, we have **p1** with a sticky end created by E, which is complementary only to other sticky ends created by E, and sticky ends created by S, which is complementary to sticky ends created by either S or X (other sticky ends are not complementary - see 2.2.1 for information about DNA). Since **LacI** has sticky ends created by E and X, **p1** and **LacI** can be combined as desired (see Figure 19).

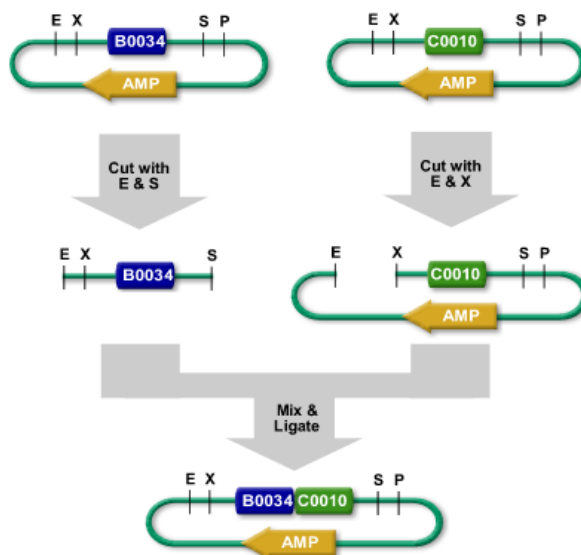


Figure 19: Schematic of a restriction digest of two parts and their ligation. Source: Registry of Standard Biological Parts website, http://partsregistry.org/Assembly:Standard_assembly

5.5.3 Ligation

After the restriction digest and the subsequent combination, the cuts in the DNA must be repaired. This is achieved by the activity of DNA ligase, which creates the phosphodiester bond between the neighbouring nucleotides. The specific protocol can be found at [7]. After successful ligation, the plasmids are again transfected into competent cells.

5.5.4 Verification

The success of the above process must be carefully verified. First, transformed cells are placed on agar plates⁷ supplemented with appropriate antibiotic. As shown in Figure 19, transferred plasmids contain an antibiotic resistance marker, therefore only cells which have this plasmid will survive.

Since the first method can verify only the presence of plasmids (success of the transformation process), a second method for verification of the ligation must be performed.

After the plasmids are extracted from the transformed cells, another restriction digest is performed to separate inserts from their vectors. Then, digested samples are placed in agarose gel (supplemented with ethidium bromide) together with reference DNA ladders⁸. Next, a DC electric field is applied across the agarose gel forcing the negatively charged DNA molecules to move from the negative electrode to the positive electrode. Once DNA fragments of different lengths are sufficiently separated, the agarose gel is placed under ultra-violet (UV) light. The location of the DNA fragments is revealed by the ethidium bromide in the gel, which fluoresces upon intercalating into the DNA double helix. By this method, DNA lengths in base-pair units can be measured and verified against the part lengths listed in the catalog.

6 Experimental results

One of the most important experiment is characterization of the elementary parts. These parts need to be well-characterized in order to correctly interpret any further experimental results with composite parts. Following results were obtained from

⁷Petri dish which contains a growth medium

⁸Mix of DNA molecules of different known lengths

measurements in a 5-hour IPTG assay. In this assay, *E. coli* bacteria with our fluorescent proteins (**p1G** and **p1R**) were induced at 3 different IPTG levels and their fluorescence was measured every 20 minutes.

6.1 Fluorescent proteins characterization

First, development of proteins fluorescence in time was observed. During the 5-hour period, fluorescence is most likely steadily increasing at an accelerated rate after 3 hours. This is most likely caused by the onset of the exponential growth phase. As we can see in Figure 20 and Figure 21, GFP exhibits significantly larger fluorescence intensity than RFP.

6.2 Characterization of external noise

Since we had at least four samples of each cell culture for every IPTG level, we could compute a confidence interval for each measurement as the difference between maximal and minimal measured value (excluding a single outlier).

These differences became more significant with higher protein levels. This showed us that gene expression is subject to external noise.

6.3 Inherent LacI characterization

Lastly we characterized the cell background, namely the inherent Lac repressor effects. To see how this repressor influences expression of our fluorescent proteins, cells were induced with different IPTG levels (for the effects of IPTG see Section 5.4).

The effects of IPTG induction were more visible with increasing time. After 240 min statistically significant differences in fluorescence values for different IPTG levels were observed. This means that the inherent Lac repressor slightly influences the expression rates of our fluorescence proteins. To increase the range to which these rates can be regulated by IPTG, Lac repressor generator, **p1L**, must be added.

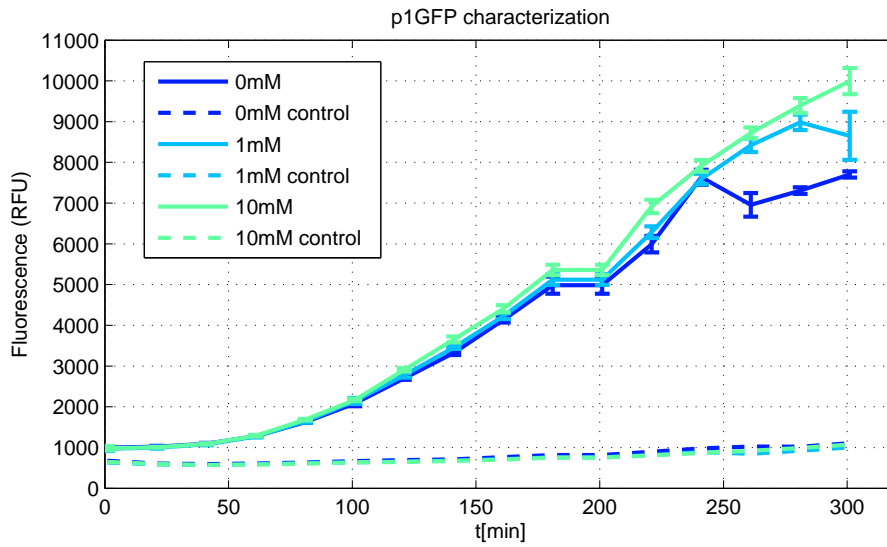


Figure 20: GFP characterization plot. Cells with **p1GFP** were induced at three different IPTG levels and their fluorescence was measured for a period of 5 hours. Dashed lines are controls - measurements with red light filter (ex: 590/10nm, em: 620/10nm).

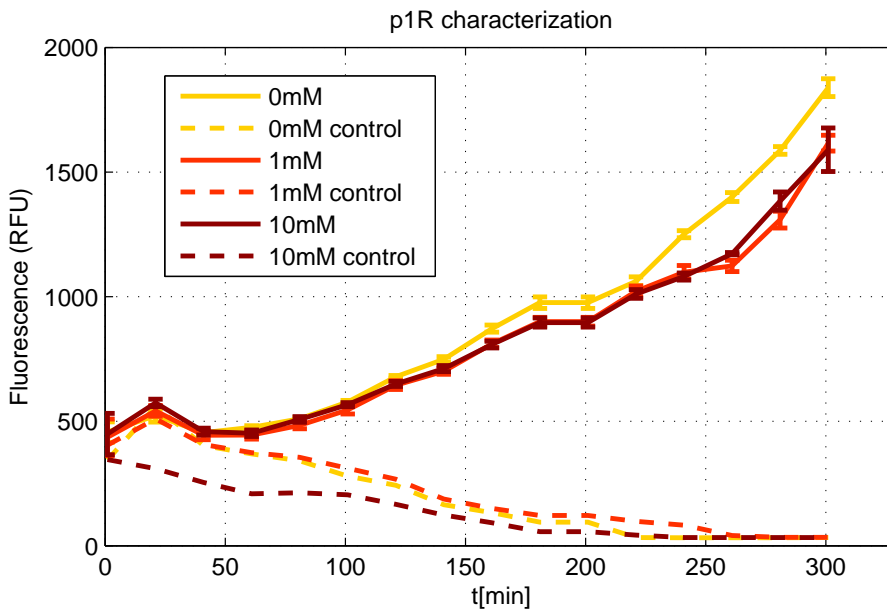


Figure 21: RFP characterization plot. Cells with **p1RFP** were induced at three different IPTG levels and their fluorescence was measured for a period of 5 hours. Dashed lines are controls - measurements with green light filter (ex: 485/20nm, em: 530/25nm).

7 Discussion

7.1 Theory

As discussed in the Biology section, the majority of parts and processes involved in gene expression and its regulation are well characterized and described. Together with appropriate methods of mathematical modelling and with certain robustness analysis tools, this information can be used for building new synthetic biological systems in cells or for their better regulation.

7.2 Simulation

All performed simulations gave us expected results. The accuracy of acquired results is strongly influenced by used parameters (reaction rates). This is a common problem in modelling of biological systems - models are over-parametrized and parameters themselves are often unknown and must be somehow estimated, or are known but vary in very large range for different individuals in different environments.

Also, as was mentioned in Section 3.4, cellular events exhibit stochastic behaviour, thus more appropriate model would also model stochastic events.

7.3 Experiment

Performed experiments gave us important information about the used fluorescent proteins and the inherent LacI characteristics. With these results, we can begin with the construction of the proposed composite parts. With these parts, other project assumptions, such as that insensitivity to the number of parallel designs can be tested. Also, the reduction of external noise (Figure 15) achieved through parallel design, can be verified.

Once these assumptions are verified, a set of different promoters will be made. With these promoters, the proposed iterative algorithm will be implemented *in vivo*.

List of Figures

1	Schematic of a GRN. Image adapted from: Office of Biological and Environmental Research of the U.S. Department of Energy Office of Science. http://science.energy.gov/ber/	3
2	Central dogma of molecular biology - hereditary information is passed from DNA to RNA to proteins but not vice versa [5]. Source: http://en.wikipedia.org/wiki/File:CDMB2.png	4
3	Scheme of a transcription unit. Credits: W. H. Freeman Pierce, Benjamin. <i>Genetics: A Conceptual Approach</i> , 2nd ed. (New York: W. H. Freeman and Company)	5
4	Three stages of a transcription. Credits: Forluvoft (Own work) [Public domain], via <i>Wikimedia Commons</i>	7
5	Alternative sigma factors and promoter recognition sequences in E. Coli. Credits: DALE, Jeremy and Simon PARK. <i>Molecular genetics of bacteria</i> . 4th ed. Hoboken	10
6	The standard genetic code showing amino acids for all 64 possible codons. Credits: <i>CLC bio</i> , http://www.clcbio.com/scienceimages/genetic_code.png	12
7	Block diagram of gene expression regulation	13
8	Example of Boolean network of lac operon - only expressed if glucose is absent and lactose is present. In fact, this scheme with one invert and one AND gate represents the Boolean inhibit function.	14
9	Block diagram of simple linear, time-invariant feedback-loop system	18
10	Example of the Nyquist plot, vector $[1 + G(j\omega)]$ is red. If the open-loop system $G(s)$ is stable, this system's closed-loop is stable too.	20
11	Frequency characteristics $G(j\omega)$ for different gains. The red is unstable, the orange is marginally stable and the green is stable	21
12	Illustration of our system. All parallel design are under control of one autoinhibiting gene. This parallel design testing ensures that external noise effects influence all designs similarly and that internal noise is averaged out [12]	23
13	Main page of the catalog, allowing users to browse part by type	24

14	IPTG assay simulation results for p1Lp1Gp1R . With increasing level of IPTG, increase in fluorescent proteins concentrations can be seen (p1R is the dashed red line, p1G is the solid green line).	28
15	Effects of different design methods on external noise. Parallel design didn't show any differences in fluorescent protein concentrations (x-axis), while concentrations of proteins in the serial design show large variation (y-axis).	29
16	Nyquist plot of the linearized p1L system	30
17	Schematic of a BioBrick plasmid	31
18	"Sticky" ends after restriction digest by restriction enzyme E. Specific sequence GAATTC is found and cut on both DNA strands.	32
19	Schematic of a restriction digest of two parts and their ligation. Source: Registry of Standard Biological Parts website, http://partsregistry.org/Assembly:Standard_assembly	32
20	GFP characterization plot. Cells with p1GFP were induced at three different IPTG levels and their fluorescence was measured for a period of 5 hours. Dashed lines are controls - measurements with red light filter (ex: 590/10nm, em: 620/10nm).	35
21	RFP characterization plot. Cells with p1RFP were induced at three different IPTG levels and their fluorescence was measured for a period of 5 hours. Dashed lines are controls - measurements with green light filter (ex: 485/20nm, em: 530/25nm).	35

References

- [1] ALBERTS, Bruce. *Essential cell biology*. 2nd ed. New York: Garland Science, c2004, 740 p. ISBN 08-153-3481-8.
- [2] BALÁTĚ, Jaroslav. *Automatické řízení*. 2nd ed., rev., Prague: BEN, 2004, 663 p. ISBN 80-730-0148-9.
- [3] BUNTING, Madeleine. Scientists have a new way to reshape nature, but none can predict the cost. *The Guardian* [online]. 2007-10-22, 2008-01-12 [cit. 2012-03-13]. Available at:
<http://www.guardian.co.uk/commentisfree/2007/oct/22/comment.comment>
- [4] CHEN, William W., Mario NIEPEL and Peter K. SORGER. Classic and contemporary approaches to modeling biochemical reactions. *GENES & DEVELOPMENT*. 2010, v. 24, 1861–1875. ISSN 0890-9369/10
- [5] CRICK, Francis. *On protein synthesis*. London, 1956. Available at:
<http://profiles.nlm.nih.gov/ps/access/SCBBFT.pdf>
- [6] DALE, Jeremy and Simon PARK. *Molecular genetics of bacteria*. 4th ed. Hoboken, N.J.: John Wiley, c2004, 346 p. ISBN 04-708-5085-X
- [7] DNA ligation. In: *OpenWetWare: Share your science*. [online]. 2012 [cit. 2012-05-10]. Available at:
http://openwetware.org/wiki/DNA_ligation
- [8] ELOWITZ, M. B. Stochastic Gene Expression in a Single Cell. *Science* [online]. 2002, vol. 297, n. 5584, p. 1183-1186 [cit. 2012-04-10]. ISSN 00368075. DOI: 10.1126/science.1070919. Available at:
<http://www.sciencemag.org/cgi/doi/10.1126/science.1070919>
- [9] FEYNMAN, Richard P. : Quotes. In: *Goodreads* [online]. 2012 [cit. 2012-03-13]. Available at:
http://www.goodreads.com/author/quotes/1429989.Richard_P_Feynman
- [10] Gene regulatory network. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001-2012 [cit. 2012-03-14]. Available at: http://en.wikipedia.org/wiki/Gene_regulatory_network

- [11] FRITSCHÉ-GUENTHER, Raphaela, Franziska WITZEL, Anja SIEBER, Ricarda HERR, Nadine SCHMIDT, Sandra BRAUN, Tilman BRUMMER, Christine SERS a Nils BLÜTHGEN. Strong negative feedback from Erk to Raf confers robustness to MAPK signalling. *Molecular Systems Biology*. 2011-5-24, vol. 7, n. 489, o. -. ISSN 1744-4292. DOI: 10.1038/msb.2011.27. Available at:
<http://www.nature.com/doi/finder/10.1038/msb.2011.27>
- [12] GEORGIEV, Daniel and Pavel ZACH. Department of Cybernetics, University of West Bohemia, Pilsen, Czech Republic. *Systematic tuning of a negative autoregulatory transcription network* [poster]. Presented at Applied Synthetic Biology in Europe, Barcelona, 6-8 February. 2012 [cit. 2012-05-06].
- [13] GONZE, D. and M. KAUFMAN. *Chemical and enzyme kinetics*. Université Libre de Bruxelles, 2011. Available at:
<http://homepages.ulb.ac.be/~dgonze/TEACHING/kinetics.pdf>
- [14] GOPAL, B. Sigma Factors. *SciTopics* [online]. 2010-06-29 [cit. 2012-03-21]. Available at:
<http://www.scitopics.com/Sigma.Factors.html>
- [15] Help:Spring 2011 DNA distribution. *Registry of Standard Biological Parts* [online]. 2011 [cit. 2012-05-10]. Available at:
http://partsregistry.org/Help:Spring_2011_DNA_distribution
- [16] *IGEM* [online]. 2007 [cit. 2012-05-06]. Available at: <http://www.igem.org>
- [17] KIMBALL, John W. Gene Translation. *Kimball's Biology Pages* [online]. 2011-05-16 [cit. 2012-03-15]. Available at:
<http://users.rcn.com/~jkimball.ma.ultranet/BiologyPages/T/Translation.html>
- [18] Knight:Restriction Digest. In: *OpenWetWare: Share your science*. [online]. 2008 [cit. 2012-05-10]. Available at:
http://openwetware.org/wiki/Knight:Restriction_Digest
- [19] Law Of Mass Action - definition. In: *Biology online* [online]. 2005-10-03 [cit. 2012-04-22]. Available at:
http://www.biology-online.org/dictionary/Law_Of_Mass_Action

- [20] MASTASCUSA, E. J. Integral Control Systems. *Control Systems Lessons* [online]. 2004 [cit. 2012-04-29]. Available at:
<http://www.facstaff.bucknell.edu/mastascu/econtrolhtml/Intro/Intro3.html>
- [21] PASSARGE, Eberhard. *Color atlas of genetics*. 2nd ed., enl. and rev. New York: Thieme, 2001, 457 p. ISBN 08-657-7958-9.
- [22] PAULSSON, Johan. Models of stochastic gene expression. *Physics of Life Reviews*. 2005, vol. 2, issue 2, p. 157-175. ISSN 15710645. DOI: 10.1016/j.plrev.2005.03.003. Available at:
<http://www.sciencedirect.com/science/article/pii/S1571064505000138>
- [23] *Registry of Standard Biological Parts* [online]. 2006 [cit. 2012-05-06]. Available at: <http://partsregistry.org>
- [24] ROBINSON, T. *Genetics for dummies*. Hoboken, N.J.: Wiley, c2005, 364 p. ISBN 07-645-9554-7.
- [25] ROLI, A., F. VERNOCCHI and R. SERRA. *Continuous network models of gene expression in knock-out experiments: A preliminary study*. In Artificial life and evolutionary computation - Proceedings of WIVACE 2008, 2008. Available at:
<http://www.lia.deis.unibo.it/aro/pubs/gnpaper-final.pdf>
- [26] ROSENFELD, Nitzan, Michael B ELOWITZ a Uri ALON. Negative Autoregulation Speeds the Response Times of Transcription Networks. *Journal of Molecular Biology*. 2002, vol. 323, n. 5, p. 785-793. ISSN 00222836. DOI: 10.1016/S0022-2836(02)00994-4. Available at:
<http://www.sciencedirect.com/science/article/pii/S0022283602009944>
- [27] SECKO, David. Protein phosphorylation: A global regulator of cellular activity. *The Science Creative Quarterly* [online]. 2003-08 [cit. 2012-03-26]. Available at:
<http://www.scq.ubc.ca/protein-phosphorylation-a-global-regulator-of-cellular-activity/>
- [28] SCHILSTRA, Maria J. and Hamid BOLOURI. Modelling the Regulation of Gene Expression in Genetic Regulatory Networks. *The NetBuilder Home page: Theory* [online]. 2003-08-21 [cit. 2012-04-02]. Available at:

- <http://strc.herts.ac.uk/bio/maria/NetBuilder/Theory/NetBuilderTheoryDownload.pdf>
- [29] STAMATAKIS, Michail and Kyriacos ZYGOURAKIS. Deterministic and stochastic population-level simulations of an artificial lac operon genetic network. *BMC Bioinformatics*. 2011, vol. 12, n. 1, p. 301-. ISSN 1471-2105. DOI: 10.1186/1471-2105-12-301. Available at: <http://www.biomedcentral.com/1471-2105/12/301>
- [30] STRAUS, David B., William A. WALTER and Carol A. GROSS. The heat shock response of *E. coli* is regulated by changes in the concentration of σ^{32} . *Nature* [online]. 1987-9-24, vol. 329, n. 6137, p. 348-351 [cit. 2012-03-14]. ISSN 0028-0836. DOI: 10.1038/329348a0. Available at: <http://www.nature.com/doifinder/10.1038/329348a0>
- [31] SZALLASI, Zoltan, Jörg STELLING and Vipul PERIWAL. *System modeling in cell biology: from concepts to nuts and bolts*. Cambridge, Mass.: MIT Press, c2006, 448 p. ISBN 978-026-2195-485.
- [32] TŮMA, František. *Teorie řízení*. 5th ed., enl. In Pilsen: University of West Bohemia, 2009, 215 p. ISBN 978-80-7043-819-0.
- [33] VODRÁŽKA, Zdeněk. *Biochemie*. 2nd ed., rev. Praha: Academia, 2002, 191 p. ISBN 80-200-0600-1
- [34] WATSON, James Dewey. *Molecular biology of the gene*. 5th ed. San Francisco: Benjamin Cummings, c2004, 732 p. ISBN 03-212-2368-3.

A List of Abbreviations

DNA deoxyribonucleic acid

GFP green fluorescent protein

GMO genetically modified organism

GRN gene regulatory network

NAR negative autoregulation

ODE ordinary differential equations

pH potential of hydrogen

RBS ribosome binding site

RFP red fluorescent protein

RNA ribonucleic acid

RNAP RNA polymerase

B Reaction rates

$k_1 = 0.23$ (LacI transcription rate [$nM.min^{-1}$])

$k_2 = 0.462$ (LacI mRNA degradation constant [min^{-1}])

$k_3 = 15$ (LacI monomer translation constant [min^{-1}])

$k_4 = 0.2$ (LacI monomer degradation constant [min^{-1}])

$k_5 = 50$ (LacI dimerization rate constant [$nM.min^{-1}$])

$k_6 = 10e-3$ (LacI dimer dissociation rate constant [min^{-1}])

$k_7 = 0.2$ (LacI dimer degradation constant [min^{-1}])

$k_8 = 960$ (Association rate constant for repression [$nM.min^{-1}$])

$k_9 = 2.4$ (Dissociation rate constant for repression [min^{-1}])

$k_{10} = k_8$

$k_{11} = k_9$

$k_{12} = 0.23$ (GFP transcription rate [$nM.min^{-1}$])

$k_{13} = 0.462$ (GFP mRNA degradation constant [min^{-1}])

$k_{14} = 0.23$ (RFP transcription rate [$nM.min^{-1}$])

$k_{15} = 0.462$ (RFP mRNA degradation constant [min^{-1}])

$k_{16} = 15$ (GFP translation constant [min^{-1}])

$k_{17} = 0.2$ (GFP degradation constant [min^{-1}])

$k_{18} = 15$ (RFP translation constant [min^{-1}])

$k_{19} = 0.2$ (RFP degradation constant [min^{-1}])

Note: All parameters were taken from [29], Table 3: Parameters of the lac operon models.

