Hill Kinetics Meets P Systems: A Case Study on Gene Regulatory Networks as Computing Agents *in silico* and *in vivo*

Thomas Hinze[1], Sikander Hayat[2], Thorsten Lenser[1], Naoki Matsumaru[1], and Peter Dittrich[1]

[1] Friedrich-Schiller-Universität Jena, Bio Systems Analysis Group
Ernst-Abbe-Platz 1–4, D-07743 Jena, Germany
{hinze,thlenser,naoki,dittrich}@minet.uni-jena.de

[2] Universität des Saarlandes, Computational Biology Group
Center for Bioinformatics, P.O. Box 15 11 50, D-66041 Saarbrücken, Germany
s.hayat@bioinformatik.uni-saarland.de

Abstract. Modelling and simulation of biological reaction networks is an essential task in systems biology aiming at formalisation, understanding, and prediction of processes in living organisms. Currently, a variety of modelling approaches for specific purposes coexists. P systems form such an approach which owing to its algebraic nature opens growing fields of application. Here, emulating the dynamical system behaviour based on reaction kinetics is of particular interest to explore network functions. We demonstrate a transformation of Hill kinetics for gene regulatory networks (GRNs) into the P systems framework. Examples address the switching dynamics of GRNs acting as inverter, NAND gate, and RS flip-flop. An adapted study *in vivo* experimentally verifies both practicability for computational units and validity of the system model.

1 Introduction

Along with the development of systems biology, a variety of modelling techniques for biological reaction networks have been established during the last years [1]. Inspired by different methodologies, three fundamental concepts emerged mostly independent of each other: *analytic*, *stochastic*, and *algebraic* approaches. Each paradigm specifically emphasises certain modelling aspects. Analytic approaches, primarily adopted from chemical reaction kinetics, enable a macroscopic view on species concentrations in many-body systems. Based on differential equations considering generation and consumption rates of species, deterministic monitoring and prediction of temporal or spatial system behaviour is efficiently expressed by continuous average concentration gradients. In contrast, stochastic approaches reflect aspects of uncertainty in biological reaction networks by incorporating randomness and probabilities. So, ranges of possible scenarios and their statistical distribution can be studied facilitating a direct comparison with

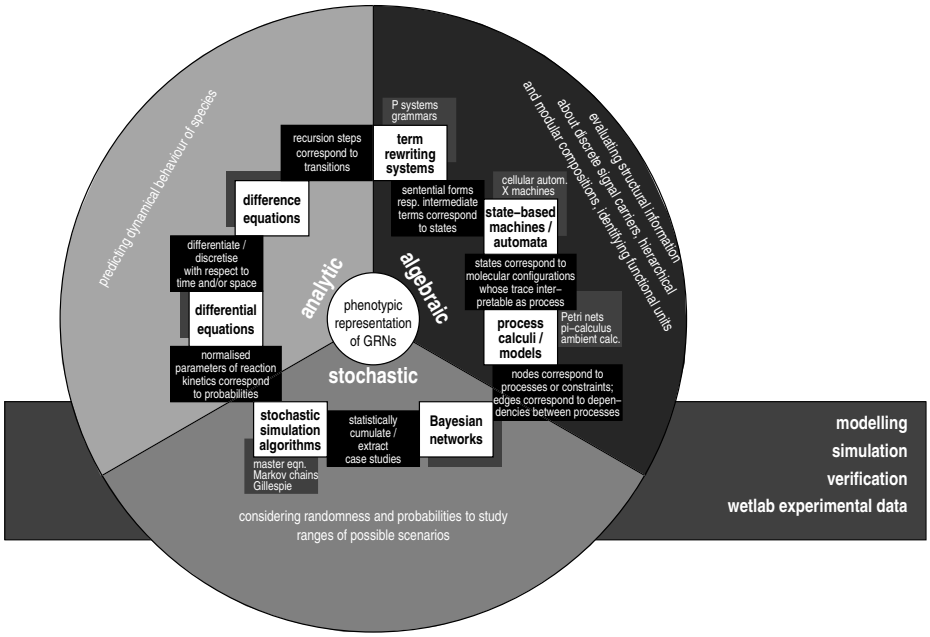


Fig. 1. Modelling approaches for biological reaction networks and bridges between them. Algorithmic strategies behind these bridges allow model transformations. Stochastic, analytic, and algebraic approaches form fundamental paradigms, categorisable into subclasses (white highlighted) and transformational concepts (black highlighted)

wetlab experimental data. Statistical tools help in discovering correlations between network components. Furthermore, algebraic approaches appear as flexible instruments regarding the level of abstraction for system description. Due to their discrete principle of operation, they work by embedding as well as evaluating structural information, modularisation, molecular tracing, and hierarchical graduation of provided system information.

Combining advantages of several paradigms comes more and more into the focus of research. On the one hand, heterogeneous models subsume elements from different approaches into an extended framework. On the other hand, transformation strategies aim to model shifting between approaches, see Figure 1. Thus, specific analysis tools as well as advanced techniques for classification, simplification, comparison, and unification can become applicable more easily. This is additionally motivated by the fact that all three paradigms are independently known to be capable of constructing Turing complete models for computation [14].

In general, P systems represent term rewriting mechanisms, hence algebraic constructs [19, 20]. Substantiated by the progress in proteomics, investigating

the dynamical behaviour of biological reaction networks is essential to understand their function. Although P systems containing appropriate kinetics are useful, reaction kinetics is mostly defined for analytic models. In this paper, we contribute to bridging this gap for GRNs.

Related work addresses corresponding P systems for phenotypic representations of some biological network classes. While metabolic P systems [15] and P systems for cell signalling [11, 18] have already been equipped with mass-action kinetics derived from underlying reaction mechanism [7], P systems for GRNs [2, 5] and for quorum sensing [3] are restricted to formulate inhibiting or activating effects qualitatively. In order to introduce a homogeneous quantitative model, we decided to incorporate Hill kinetics [16] to the P systems framework by describing the cooperativity in GRNs dynamically using sigmoid-shaped transfer functions that are more precise than two-stage on/off switching.

The paper is organised as follows: Based on the definition of Hill kinetics, we present a method for discretisation that leads to P systems Π_{Hill} whose properties are discussed briefly. A case study includes GRNs acting as inverter, NAND gate, and RS flip-flop. For each logic gate, its GRN in concert with ODEs derived from Hill kinetics, corresponding P system, and simulation results are shown. Finally, we verify that a reporter gene encoding the green fluorescent protein (*gfp*) with transcription factors N-acyl homoserine lactone (AHL) and isopropyl-βD-thiogalactopyranoside (IPTG) can mimic the aforementioned RS flip-flop *in vivo*. Here, *gfp* expression is quantified using flow cytometry.

2 Transforming Hill Kinetics to P System

Hill Kinetics

Hill kinetics [16] represents a homogeneous analytic approach to model cooperative and competitive aspects of interacting biochemical reaction networks dynamically. It formulates the relative intensity of gene regulations by sigmoid-shaped threshold functions h of degree $m \in \mathbb{N}_+$ and threshold $\Theta > 0$ such that $x \geq 0$ specifies the concentration level of a transcription factor that activates resp. inhibits gene expression. Function value h then returns the normalised change in concentration level of the corresponding gene product:

activation (upregulation) $\rightarrow$: $\quad \mathrm{h}^+(x, \Theta, m) = \frac{x^m}{x^m + \Theta^m}$

inhibition (downregulation) $\perp$: $\mathrm{h}^-(x, \Theta, m) = 1 - \mathrm{h}^+(x, \Theta, m)$

Functions h^+ and h^- together with a proportional factor c_1 quantify the production rate of a certain gene product *GeneProduct*. Here we assume a linear spontaneous decay with rate $c_2[\mathit{GeneProduct}]$ such that the differential equation takes the form $\frac{\mathrm{d}\,[\mathit{GeneProduct}]}{\mathrm{d}\,t} = \mathit{ProductionRate} - c_2[\mathit{GeneProduct}]$. Different activation and inhibition rates are simply multiplied as in the following example illustrated in Figure 2 ($c_1, c_2 \in \mathbb{R}_+$):

$$\frac{\mathrm{d}\,[\mathit{GeneProduct}]}{\mathrm{d}\,t} = c_1 \cdot \mathrm{h}^+(A_1, \Theta_{A_1}, m) \cdot \ldots \cdot \mathrm{h}^+(A_n, \Theta_{A_n}, m) \cdot \big(1 - \mathrm{h}^+(I_1, \Theta_{I_1}, m) \cdot \ldots \cdot \mathrm{h}^+(I_p, \Theta_{I_p}, m)\big) - c_2[\mathit{GeneProduct}]$$

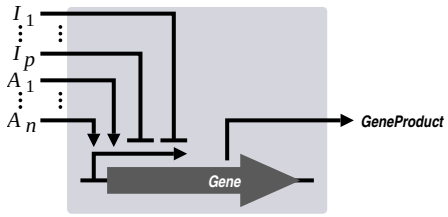


Fig. 2. Gene regulatory unit. Repetitive expression of a *Gene* leads to generation of a specific *GeneProduct*, a protein whose amino acid sequence is encoded by the DNA sequence of the *Gene*. Transcription factors (specific single proteins or complexes) quantitatively control the expression rate by their present concentration. Two types of transcription factors can be distinguished: Inhibitors, here symbolised by $I_1, \ldots, I_p$, repress *Gene* expression by downregulation while activators $A_1, \ldots, A_n$ cause the opposite amplifying effect by upregulation.

For simplicity, each differential coefficient $\frac{\mathrm{d}\,y}{\mathrm{d}\,t}$ is subsequently denoted as $\dot{y}$.

By coupling gene regulatory units we obtain GRNs. Here, gene products can act as transcription factors for other genes within the network. Additional complex formation among gene products allows conjunctive composition of transcription factors and the introduction of further nonlinearities. Thus, an effective signal transduction and combination between different network elements becomes feasible.

Discretisation

The analytic nature of Hill kinetics based on continuous concentrations requires a discretisation with respect to value and time in order to derive a homologous term rewriting mechanism. Following the intention to approximate continuous concentrations by absolute particle numbers, we assume a large but finite pool of molecules. The application of a reaction rule in terms of a rewriting process removes a number of reactant particles from this pool and simultaneously adds all products. Therefore, selection and priorisation of reaction rules to apply are controlled by an underlying iteration scheme with temporally stepwise operation.

Since Hill kinetics is characterised by variable reaction rates due to the sigmoid-shaped functions h, this variability should also be reflected in the term rewriting mechanism. For this reason, we introduce dynamic stoichiometric factors resulting in time dependent reaction rules. Let $\Delta\tau > 0$ be the constant time discretisation interval (step length), the gene regulatory unit depicted in Figure 2 consists of two reaction rules with variable stoichiometric factors s and u:

$$s\ \mathit{Gene} \longrightarrow s\ \mathit{GeneProduct} + s\ \mathit{Gene}\ \big|_{A_1,\ldots,A_n,\neg I_1,\ldots,\neg I_p} \quad \text{where}$$

$$s = \lfloor \Delta\tau \cdot c_1 \cdot [\mathit{Gene}] \cdot \mathrm{h}^+(A_1, \Theta_{A_1}, m) \cdot \ldots \cdot \mathrm{h}^+(A_n, \Theta_{A_n}, m) \cdot \big(1 - \mathrm{h}^+(I_1, \Theta_{I_1}, m) \cdot \ldots \cdot \mathrm{h}^+(I_p, \Theta_{I_p}, m)\big) \rfloor$$

$$u\ \mathit{GeneProduct} \longrightarrow \emptyset \quad \text{where} \quad u = \lfloor \Delta\tau \cdot c_2 \cdot [\mathit{GeneProduct}] \rfloor$$

Here, the upper reaction formulates the generation of *GeneProduct* particles with regard to the limiting resource of available *Gene* objects. Reaction conditions coming from the presence of activators $A_1, \ldots, A_n$ and absence ($\neg$) of inhibitors $I_1, \ldots, I_p$ affect the stoichiometric factor s. The notation of indexes after the vertical bar declares the elements which occur in the h-components $(\mathrm{h}^+, \mathrm{h}^-)$ of the function regulating the rule. In order to map normalised concentrations from Hill kinetics into absolute particle numbers, we introduce the factor term $[Gene]$ which represents the total number of *Gene* objects present in the reaction system. Accordingly, the decay (consumption) of *GeneProduct* is expressed by the lower transition rule.

The rounding regulation ($\lfloor\ \rfloor$) provides for integer numbers as stoichiometric factors. This is necessary for handling the discrete manner of term rewriting. Nevertheless, a discretisation error can occur and propagate over the time course. The higher the total number of particles in the reaction system is initially set, the more this inaccuracy can be reduced.

Now, we incorporate the reaction system obtained by discretisation into the P systems framework. Therefore, we firstly define some syntactical conventions with respect to formal languages and multisets.

Formal Language and Multiset Prerequisites

We denote the empty word by ε. Let A be an arbitrary set and $\mathbb{N}$ the set of natural numbers including zero. A multiset over A is a mapping $F : A \longrightarrow \mathbb{N} \cup \{\infty\}$. $F(a)$, also denoted as $[a]_F$, specifies the multiplicity of $a \in A$ in F. Multisets can be written as an elementwise enumeration of the form $\{(a_1, F(a_1)), (a_2, F(a_2)), \ldots\}$ since $\forall (a, b_1), (a, b_2) \in F : b_1 = b_2$. The support $\mathrm{supp}(F) \subseteq A$ of F is defined by $\mathrm{supp}(F) = \{a \in A \mid F(a) > 0\}$. A multiset F over A is said to be empty iff $\forall a \in A : F(a) = 0$. The cardinality $|F|$ of F over A is $|F| = \sum_{a \in A} F(a)$. Let F_1 and F_2 be multisets over A. F_1 is a subset of F_2, denoted as $F_1 \subseteq F_2$, iff $\forall a \in A : (F_1(a) \leq F_2(a))$. Multisets F_1 and F_2 are equal iff $F_1 \subseteq F_2 \wedge F_2 \subseteq F_1$. The intersection $F_1 \cap F_2 = \{(a, F(a)) \mid a \in A \wedge F(a) = \min(F_1(a), F_2(a))\}$, the multiset sum $F_1 \uplus F_2 = \{(a, F(a)) \mid a \in A \wedge F(a) = F_1(a) + F_2(a)\}$, and the multiset difference $F_1 \ominus F_2 = \{(a, F(a)) \mid a \in A \wedge F(a) = \max(F_1(a) - F_2(a), 0)\}$ form multiset operations. The term $\langle A \rangle = \{F : A \longrightarrow \mathbb{N} \cup \{\infty\}\}$ describes the set of all multisets over A while $\mathcal{P}(A)$ denotes the power set of A.

Transformation: Definition of the Corresponding P System

The general form of a P system Π_{Hill} emulating the dynamical behaviour of GRNs using Hill kinetics is a construct

$$\Pi_{\mathrm{Hill}} = (V_{\mathrm{Genes}}, V_{\mathrm{GeneProducts}}, \Sigma, [_1]_1, L_0, r_1, \ldots, r_k, \mathrm{f}_1, \ldots, \mathrm{f}_k, \Delta\tau, m)$$

where V_{Genes} denotes the alphabet of genes, $V_{\mathrm{GeneProducts}}$ the alphabet of gene products (without loss of generality $V_{\mathrm{Genes}} \cap V_{\mathrm{GeneProducts}} = \emptyset$), and $\Sigma \subseteq V_{\mathrm{GeneProducts}}$ represents the output alphabet. Π_{Hill} does not incorporate

inner membranes, so the only membrane is the skin membrane $[_1]_1$. The single membrane property results from the spatial globality of GRNs within an organism: Gene expression is located in the cell nuclei flanked by a receptor-controlled intercellular transduction and combination of transcription factors. Resulting GRNs form independent network structures of high stability within living organisms.

Let $V = V_{\text{Genes}} \cup V_{\text{GeneProducts}}$. The multiset $L_0 \in \langle V \rangle$ over V holds the initial configuration of the system.

Initial reaction rules[1] $r_i \in \langle E_{i,0} \rangle \times \langle P_{i,0} \rangle \times \mathcal{P}(TF_i)$ with multiset of reactants $E_{i,0} \subseteq V \times \mathbb{N}$, multiset of products $P_{i,0} \subseteq V \times \mathbb{N}$ and set of involved transcription factors $TF_i \in V_{\text{GeneProducts}}$, $i = 1, \ldots, k$, define the potential system activity at time point 0. A function $\mathrm{f}_i : \mathbb{R}_+ \times \langle V \rangle \times \mathbb{N}_+ \to \mathbb{N}$ is associated with each initial reaction rule r_i. This function adapts the stoichiometric factors according to the discretised Hill kinetics as described above.

Furthermore, we introduce two global parameters. The time discretisation interval $\Delta\tau \in \mathbb{R}_+$ corresponds to the length of a time step between discrete time points t and $t+1$. The degree $m \in \mathbb{N}_+$ is used for all embedded sigmoid-shaped functions.

Finally, the dynamical behaviour of the P system is specified by an iteration scheme updating both the system configuration L_t and the stoichiometric factors of reaction rules r_i starting from L_0 where $i = 1, \ldots, k$:

$$L_{t+1} = L_t \ominus \mathit{Reactants}_t \uplus \mathit{Products}_t \quad \text{with}$$

$$\mathit{Reactants}_t = \biguplus_{i=1}^{k} (E_{i,t+1} \cap L_t)$$

$$\mathit{Products}_t = \biguplus_{i=1}^{k} (P_{i,t+1} \cap L_t)$$

$$E_{i,t+1} = \{(e, a') \mid (e, a) \in E_{i,t} \ \wedge a' = \mathrm{f}_i(\Delta\tau, L_t, m)\} \tag{1}$$

$$P_{i,t+1} = \{(q, b') \mid (q, b) \in P_{i,t} \ \wedge b' = \mathrm{f}_i(\Delta\tau, L_t, m)\} \tag{2}$$

Informally, the specification of $E_{i,t+1}$ and $P_{i,t+1}$ means that all reactants e and products q remain unchanged over the time course. Just their stoichiometric factors are updated from value a to a' (reactants) and from b to b' (products) according to functions f_i. These functions may utilise the numbers of copies for all $|V|$ types of particles recently present in the system. The cardinality $|L_t \cap \{(w_j, \infty)\}|$ then identifies this amount for any $w_j \in V$.

In terms of computational devices, P systems Π_{Hill} carry an output providing the outcome of a calculation. For this purpose, the multiplicity of those gene products listed in the output alphabet is suitable. We define an output function $\text{output} : \mathbb{N} \to \mathbb{N}$ by

$$\text{output}(t) = |L_t \cap \{(w, \infty) \mid w \in \Sigma\}|.$$

[1] Note that in our case the stoichiometry of reaction rules changes over time which is used to implement time-varying reaction rates.

For better readability, we subsequently write a reaction rule $r_i = \Big(\big\{(e_1,a_1),\ldots,(e_h,a_h)\big\},\big\{(q_1,b_1),\ldots,(q_v,b_v)\big\},\big\{tf_1,\ldots,tf_c\big\}\Big)$ with $\mathrm{supp}(E_{i,t}) = \{e_1,\ldots,e_h\}$ and $\mathrm{supp}(P_{i,t}) = \{q_1,\ldots,q_v\}$ as well as $TF_i = \{tf_1,\ldots,tf_c\}$ by using the chemical denotation $r_i \;:\; a_1\ e_1 + \ldots + a_h\ e_h \longrightarrow b_1\ q_1 + \ldots + b_v\ q_v\big|_{tf_1,\ldots,tf_c}$. As a first example, Π_{Hill} of the gene regulatory unit shown in Figure 2 reads:

$$
\begin{aligned}
\Pi_{\mathrm{Hill,GRNunit}} &= (V_{\mathrm{Genes}}, V_{\mathrm{GeneProducts}}, \Sigma, [_1]_1, L_0, r_1, r_2, \mathrm{f}_1, \mathrm{f}_2, \Delta\tau, m)\\
V_{\mathrm{Genes}} &= \{Gene\}\\
V_{\mathrm{GeneProducts}} &= \{A_1,\ldots,A_n,\neg I_1,\ldots,\neg I_p, GeneProduct\}\\
\Sigma &= \{GeneProduct\}\\
L_0 &= \{(Gene,g),(A_1,a_1),\ldots,(A_n,a_n),(\neg I_1,i_1),\ldots,(\neg I_p,i_p)\}\\
r_1 &\;:\; s_1\ Gene \longrightarrow s_1\ GeneProduct + s_1\ Gene\big|_{A_1,\ldots,A_n,\neg I_1,\ldots,\neg I_p}\\
r_2 &\;:\; s_2\ GeneProduct \longrightarrow \emptyset\\
\mathrm{f}_1(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(Gene,\infty)\}| \cdot\\
&\qquad \frac{|L_t \cap \{(A_1,\infty)\}|^m}{|L_t \cap \{(A_1,\infty)\}|^m + \Theta_{A_1}^m} \cdot \ldots \cdot \frac{|L_t \cap \{(A_n,\infty)\}|^m}{|L_t \cap \{(A_n,\infty)\}|^m + \Theta_{A_n}^m} \cdot\\
&\qquad \left(1 - \frac{|L_t \cap \{(\neg I_1,\infty)\}|^m}{|L_t \cap \{(\neg I_1,\infty)\}|^m + \Theta_{\neg I_1}^m} \cdot \ldots \cdot \frac{|L_t \cap \{(\neg I_p,\infty)\}|^m}{|L_t \cap \{(\neg I_p,\infty)\}|^m + \Theta_{\neg I_p}^m}\right) \rfloor\\
\mathrm{f}_2(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(GeneProduct,\infty)\}| \rfloor\\
\Delta\tau &\in \mathbb{R}_+\\
m &\in \mathbb{N}_+
\end{aligned}
$$

Note that s_1 at time point $t+1$ is equal to $\mathrm{f}_1(\Delta\tau, L_t, m)$ at time point t or holds its initialisation value at time point 0. Respectively, s_2 at time point $t+1$ is equal to $\mathrm{f}_2(\Delta\tau, L_t, m)$ at time point t or holds its initialisation value at time point 0, see equations (1) and (2).

At low molecular concentrations, deterministic application of Hill functions can conflict between different functions which want to update the system configuration. This is the case if the amount of reactants is too small to satisfy the needs of all functions. Since the number of multiset elements always remains nonnegative (see definition of $\ominus$), the system can violate mass conservation by satisfying these needs. A system extension based on stochastic rewriting mechanisms can overcome this insufficiency.

System Classification, Properties and Universality

Π_{Hill} belongs to P systems with symbol objects and time varying transition rules whose evolution is based on conditional rewriting by quantitative usage of promoters and inhibitors. Thus, the dynamical behaviour formulated in Hill kinetics is time- and value-discretely approximated by a stepwise adaptation. This leads to a deterministic principle of operation.

From the view on computational completeness, there are several indicators for Turing universality. On the one hand, we will demonstrate within the next section how NAND gates and compositions of NAND gates can be emulated by P systems of the form Π_{Hill}. Arbitrarily extendable circuits consisting of coupled NAND gates can be seen as computational complete [14]. On the other hand, the multiplicity of each symbol object within the system may range through the whole recursively enumerable set of natural numbers. So, copies of a gene product expressed by a dedicated gene are able to represent the register value of a random access machine. Autoactivation loops keep a register at a certain value while external activation increases the amount of gene product (increment operation) and external inhibition decreases respectively (decrement operation). Incrementing and decrementing transcription factors always form complexes with program counter objects. The interplay of those specific transcription factors manages the program control.

3 Case Study: Computational Units and Circuits

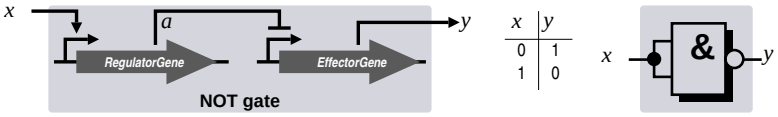
Artificial GRNs have been instrumental in elucidating basic principles that govern the dynamics and consequences of stochasticity in the gene expression of naturally occurring GRNs. The realisation as computational circuits infers inherent evolutionary fault tolerance and robustness to these modular units.

In a case study, we introduce three artificial GRNs for logic gates (inverter, NAND gate, RS flip-flop) and describe their dynamical behaviour quantitatively by an ordinary differential equation model using Hill kinetics and by corresponding P systems Π_{Hill}.

A variety of distinguishable transcription factors given by their concentration over the time course enables communication between as well as coupling of computational units. Thus, circuit engineering becomes feasible.

Inverter

input: concentration level of transcription factor x
output: concentration level of gene product y

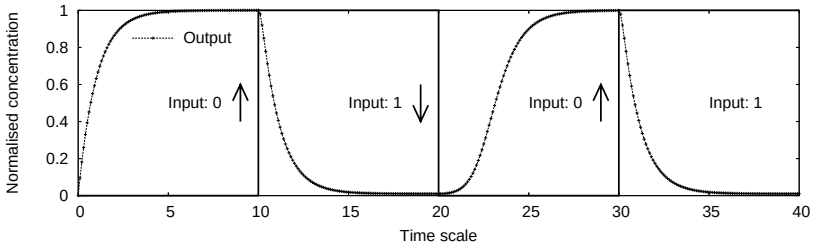


Ordinary Differential Equations (ODE)

$$\dot{a} = \mathrm{h}^{+}(x, \Theta_x, m) - a$$
$$\dot{y} = \mathrm{h}^{-}(a, \Theta_a, m) - y$$

Simulation Result (Copasi, ODE solver [12])

dynamical behaviour depicted for $m = 2$, $\Theta_j = 0.1$, $j \in \{x, a\}$,
$a(0) = 0$, $y(0) = 0$, $x(t) = \begin{cases} 0 \text{ for } 0 \leq t < 10;\ 20 \leq t < 30 \\ 1 \text{ for } 10 \leq t < 20;\ 30 \leq t < 40 \end{cases}$

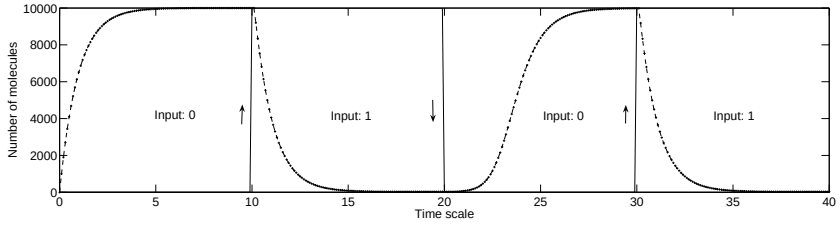


Corresponding P System

$$
\begin{aligned}
\Pi_{\text{Hill,GRNinv.}} &= (V_{\text{Genes}}, V_{\text{GeneProducts}}, \Sigma, [\,_1]_1, L_0, r_1, \ldots, r_4, \mathrm{f}_1, \ldots, \mathrm{f}_4, \Delta\tau, m) \\
V_{\text{Genes}} &= \{\mathit{RegulatorGene}, \mathit{EffectorGene}\} \\
V_{\text{GeneProducts}} &= \{x, y, \neg a\} \\
\Sigma &= \{y\} \\
L_0 &= \{(\mathit{RegulatorGene}, rg), (\mathit{EffectorGene}, eg), (x, x_0), (y, y_0), (\neg a, a_0)\} \\
r_1 &: s_1\ \mathit{RegulatorGene} \longrightarrow s_1\ \neg a + s_1\ \mathit{RegulatorGene}\ \big|_x \\
r_2 &: s_2\ \neg a \longrightarrow \emptyset \\
r_3 &: s_3\ \mathit{EffectorGene} \longrightarrow s_3\ y + s_3\ \mathit{EffectorGene}\ \big|_{\neg a} \\
r_4 &: s_4\ y \longrightarrow \emptyset \\
\mathrm{f}_1(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{RegulatorGene}, \infty)\}| \cdot \frac{|L_t \cap \{(x,\infty)\}|^m}{|L_t \cap \{(x,\infty)\}|^m + \Theta_x^m} \right\rfloor \\
\mathrm{f}_2(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(\neg a, \infty)\}| \rfloor \\
\mathrm{f}_3(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{EffectorGene}, \infty)\}| \cdot \left(1 - \frac{|L_t \cap \{(\neg a,\infty)\}|^m}{|L_t \cap \{(\neg a,\infty)\}|^m + \Theta_{\neg a}^m}\right) \right\rfloor \\
\mathrm{f}_4(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(y, \infty)\}| \rfloor \\
\Delta\tau &\in \mathbb{R}_+ \\
m &\in \mathbb{N}_+
\end{aligned}
$$

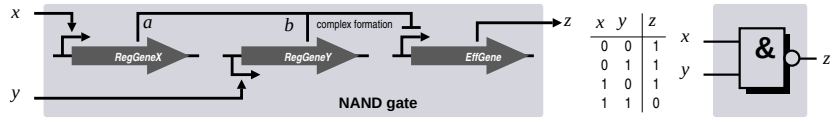
Simulation Result (MATLAB, P system iteration scheme)

dynamical behaviour depicted for $m = 2$, $\Delta\tau = 0.1$, $\Theta_j = 500$, $j \in \{x, \neg a\}$
$rg = 10,000$, $eg = 10,000$, $x_0 = 0$, $y_0 = 0$, $a_0 = 0$



NAND gate

input: concentration levels of transcription factors x (input1), y (input2)
output: concentration level of gene product z

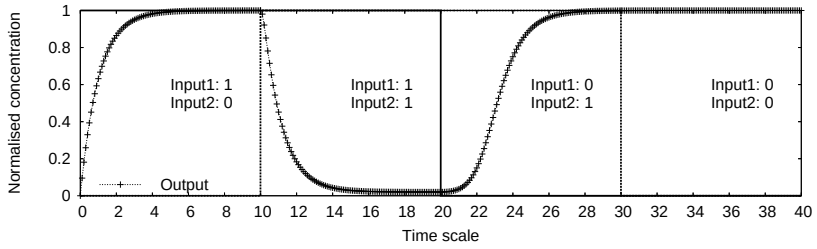


Ordinary Differential Equations

$$\begin{aligned}\dot{a} &= \mathrm{h}^+(x, \Theta_x, m) - a \\ \dot{b} &= \mathrm{h}^+(y, \Theta_y, m) - b \\ \dot{z} &= 1 - \mathrm{h}^+(a, \Theta_a, m) \cdot \mathrm{h}^+(b, \Theta_b, m) - z\end{aligned}$$

Simulation Result (Copasi, ODE solver)

dynamical behaviour depicted for $m = 2$, $\Theta_j = 0.1$, $j \in \{x, y, a, b\}$

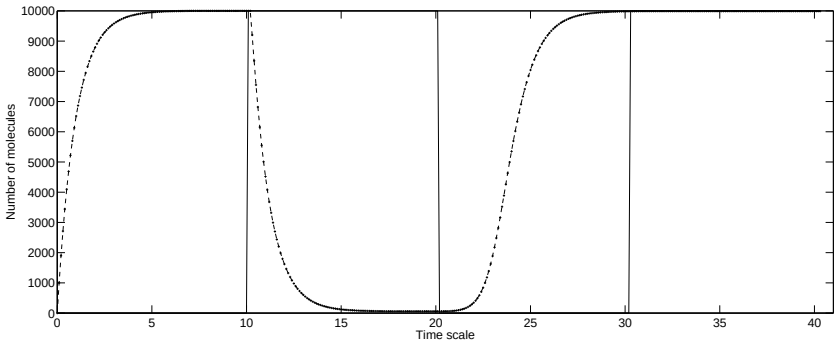


Corresponding P System

$$\begin{aligned}
\Pi_{\mathrm{Hill,GRNnand}} &= (V_{\mathrm{Genes}}, V_{\mathrm{GeneProducts}}, \Sigma, [\,_1]_1, L_0, r_1, \ldots, r_6, \mathrm{f}_1, \ldots, \mathrm{f}_6, \Delta\tau, m) \\
V_{\mathrm{Genes}} &= \{\mathit{RegGeneX}, \mathit{RegGeneY}, \mathit{EffGene}\} \\
V_{\mathrm{GeneProducts}} &= \{x, y, z, \neg a, \neg b\} \\
\Sigma &= \{z\} \\
L_0 &= \{(\mathit{RegGeneX}, rgx), (\mathit{RegGeneY}, rgy), (\mathit{EffGene}, eg), \\
&\qquad (x, x_0), (y, y_0), (z, z_0), (\neg a, a_0), (\neg b, b_0)\} \\
r_1 &: s_1\ \mathit{RegGeneX} \longrightarrow s_1\ \neg a + s_1\ \mathit{RegGeneX}\ \big|_x \\
r_2 &: s_2\ \neg a \longrightarrow \emptyset \\
r_3 &: s_3\ \mathit{RegGeneY} \longrightarrow s_3\ \neg b + s_3\ \mathit{RegGeneY}\ \big|_y \\
r_4 &: s_4\ \neg b \longrightarrow \emptyset \\
r_5 &: s_5\ \mathit{EffGene} \longrightarrow s_5\ z + s_5\ \mathit{EffGene}\ \big|_{\neg a, \neg b} \\
r_6 &: s_6\ z \longrightarrow \emptyset \\
\mathrm{f}_1(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{RegGeneX}, \infty)\}| \cdot \frac{|L_t \cap \{(x,\infty)\}|^m}{|L_t \cap \{(x,\infty)\}|^m + \Theta_x^m} \right\rfloor \\
\mathrm{f}_2(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(\neg a, \infty)\}| \rfloor \\
\mathrm{f}_3(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{RegGeneY}, \infty)\}| \cdot \frac{|L_t \cap \{(y,\infty)\}|^m}{|L_t \cap \{(y,\infty)\}|^m + \Theta_y^m} \right\rfloor \\
\mathrm{f}_4(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(\neg b, \infty)\}| \rfloor \\
\mathrm{f}_5(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{EffGene}, \infty)\}| \cdot \right. \\
&\qquad \left. \left(1 - \frac{|L_t \cap \{(\neg a,\infty)\}|^m}{|L_t \cap \{(\neg a,\infty)\}|^m + \Theta_{\neg a}^m} \cdot \frac{|L_t \cap \{(\neg b,\infty)\}|^m}{|L_t \cap \{(\neg b,\infty)\}|^m + \Theta_{\neg b}^m}\right) \right\rfloor \\
\mathrm{f}_6(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(z, \infty)\}| \rfloor \\
\Delta\tau &\in \mathbb{R}_+ \\
m &\in \mathbb{N}_+
\end{aligned}$$

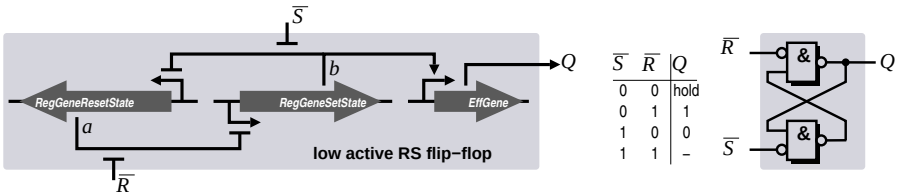
Simulation Result (MATLAB, P system iteration scheme)

dynamical behaviour depicted for $m = 2$, $\Delta\tau = 0.1$, $\Theta_j = 500$, $j \in \{x, y, \neg a, \neg b\}$
$rgx = 10,000$, $rgy = 10,000$, $eg = 10,000$, $x_0 = 0$, $y_0 = 0$, $z_0 = 0$, $a_0 = 0$, $b_0 = 0$



RS Flip-Flop

input: concentration levels of transcription factors $\overline{S}, \overline{R}$
output: concentration level of gene product Q

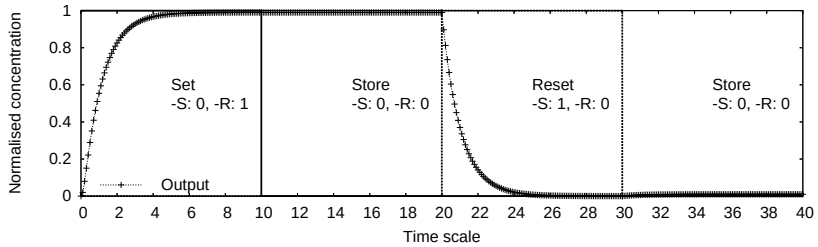


Ordinary Differential Equations

$$\begin{aligned}
\dot{a} &= 1 - \mathrm{h}^{+}(b, \Theta_b, m) \cdot \mathrm{h}^{-}(\overline{S}, \Theta_{\overline{S}}, m) - a \\
\dot{b} &= 1 - \mathrm{h}^{+}(a, \Theta_a, m) \cdot \mathrm{h}^{-}(\overline{R}, \Theta_{\overline{R}}, m) - b \\
\dot{Q} &= \mathrm{h}^{+}(b, \Theta_b, m) \cdot \mathrm{h}^{-}(\overline{S}, \Theta_{\overline{S}}, m) - Q
\end{aligned}$$

Simulation Result (Copasi, ODE solver)

dynamical behaviour depicted for $m = 2$, $\Theta_j = 0.1$, $j \in \{a, b, \overline{R}, \overline{S}\}$



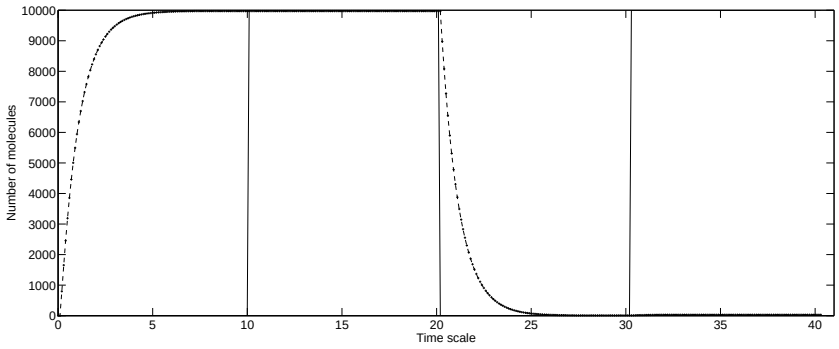
Corresponding P System

$$\begin{aligned}
\Pi_{\mathrm{Hill,GRNrsff}} &= (V_{\mathrm{Genes}}, V_{\mathrm{GeneProducts}}, \Sigma, [\,]_1, L_0, r_1, \ldots, r_6, \mathrm{f}_1, \ldots, \mathrm{f}_6, \Delta\tau, m) \\
V_{\mathrm{Genes}} &= \{\mathit{RegGeneResetState}, \mathit{RegGeneSetState}, \mathit{EffGene}\} \\
V_{\mathrm{GeneProducts}} &= \{Q, \neg\overline{S}, \neg\overline{R}, \neg a, \neg b\} \\
\Sigma &= \{Q\}
\end{aligned}$$

$$
\begin{aligned}
L_0 &= \{(\mathit{RegGeneResetState}, rgr), (\mathit{RegGeneSetState}, rgs),\\
&\quad (\mathit{EffGene}, eg), (Q, q_0), (\neg\overline{S}, ss_0), (\neg\overline{R}, rs_0), (\neg a, a_0), (\neg b, b_0)\}\\
r_1 &: s_1\ \mathit{RegGeneResetState} \longrightarrow s_1\ \neg a + s_1\ \mathit{RegGeneResetState}\big|_{\neg\overline{S},\neg b}\\
r_2 &: s_2\ \neg a \longrightarrow \emptyset\\
r_3 &: s_3\ \mathit{RegGeneSetState} \longrightarrow s_3\ \neg b + s_3\ \mathit{RegGeneSetState}\ \big|_{\neg\overline{R},\neg a}\\
r_4 &: s_4\ \neg b \longrightarrow \emptyset\\
r_5 &: s_5\ \mathit{EffGene} \longrightarrow s_5\ Q + s_5\ \mathit{EffGene}\ \big|_{\neg\overline{S},\neg b}\\
r_6 &: s_6\ Q \longrightarrow \emptyset\\
\mathrm{f}_1(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{RegGeneResetState}, \infty)\}| \cdot \right.\\
&\quad \left. \left(1 - \frac{|L_t \cap \{(\neg b,\infty)\}|^m}{|L_t \cap \{(\neg b,\infty)\}|^m + \Theta^m_{\neg b}} \cdot \left(1 - \frac{|L_t \cap \{(\neg\overline{S},\infty)\}|^m}{|L_t \cap \{(\neg\overline{S},\infty)\}|^m + \Theta^m_{\neg\overline{S}}}\right)\right)\right\rfloor\\
\mathrm{f}_2(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(\neg a, \infty)\}| \rfloor\\
\mathrm{f}_3(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{RegGeneSetState}, \infty)\}| \cdot \right.\\
&\quad \left. \left(1 - \frac{|L_t \cap \{(\neg a,\infty)\}|^m}{|L_t \cap \{(\neg a,\infty)\}|^m + \Theta^m_{\neg a}} \cdot \left(1 - \frac{|L_t \cap \{(\neg\overline{R},\infty)\}|^m}{|L_t \cap \{(\neg\overline{R},\infty)\}|^m + \Theta^m_{\neg\overline{R}}}\right)\right)\right\rfloor\\
\mathrm{f}_4(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(\neg b, \infty)\}| \rfloor\\
\mathrm{f}_5(\Delta\tau, L_t, m) &= \left\lfloor \Delta\tau \cdot |L_t \cap \{(\mathit{EffGene}, \infty)\}| \cdot \right.\\
&\quad \left. \frac{|L_t \cap \{(\neg b,\infty)\}|^m}{|L_t \cap \{(\neg b,\infty)\}|^m + \Theta^m_{\neg b}} \cdot \left(1 - \frac{|L_t \cap \{(\neg\overline{S},\infty)\}|^m}{|L_t \cap \{(\neg\overline{S},\infty)\}|^m + \Theta^m_{\neg\overline{S}}}\right)\right\rfloor\\
\mathrm{f}_6(\Delta\tau, L_t, m) &= \lfloor \Delta\tau \cdot |L_t \cap \{(Q, \infty)\}| \rfloor\\
\Delta\tau &\in \mathbb{R}_+\\
m &\in \mathbb{N}_+
\end{aligned}
$$

Simulation Result (MATLAB, P system iteration scheme)

dynamical behaviour depicted for $m = 2$, $\Delta\tau = 0.1$, $\Theta_j = 500$, $j \in \{\neg a, \neg b, \neg\overline{R}, \neg\overline{S}\}$
$rgr = 10,000$, $rgs = 10,000$, $eg = 10,000$, $q_0 = 0$, $ss_0 = 0$, $rs_0 = 0$, $a_0 = 0$, $b_0 = 0$



A homologous analytic model of a bistable toggle switch was introduced in [8]. In case of the forbidden input signalling $\overline{S} = 1$, $\overline{R} = 1$, the normalised concentrations of both inhibitors $\neg a$ and $\neg b$ converge to 0.5. By setting or resetting input signalling, the flip-flop restores.

4 RS Flip-Flop Validation *in vivo*

In addition to prediction and simulation of GRNs acting as logic gates, we demonstrate the practicability of the RS flip-flop by an experimental study *in vivo*. Resulting output protein data measured over the time course can validate the system model. Following the pioneering implementation of a bistable toggle switch [8], we could confirm its function in a previous study [10]. Two extensions were investigated: Firstly, the effects of IPTG and AHL as appropriate intercellular inducers for flip-flop setting were shown. Secondly, flow cytometry was used to quantitatively measure protein concentrations within the flip-flop implementation. In the following, we give a brief overview of the experimental setup and compare obtained experimental results with our simulations based on the models.

Biological Principles and Prerequisites

Quorum Sensing and Autoinduction via AHL

In quorum sensing, bacterial species regulate gene expression based on cell-population density [17]. An alteration in gene expression occurs when an intercellular signalling molecule termed autoinducer, produced and released by the bacterial cells reaches a critical concentration. Termed as quorum sensing or autoinduction, this fluctuation in autoinducer concentration is a function of bacterial cell-population density. *Vibrio fischeri*, a well studied bacterium, colonises the light organs of a variety of marine fishes and squids, where it occurs at very high densities ($10^{10}\,\frac{\text{cells}}{\text{ml}}$) and produces light. The two genes essential for cell density regulation of luminescence are: luxI, which codes for an autoinducer synthase [22]; and luxR, which codes for an autoinducer-dependent activator of the luminescence genes. The luxR and luxI genes are adjacent and divergently transcribed, and luxI is the first of seven genes in the luminescence or lux operon. LuxI-type proteins direct AHL synthesis while LuxR-type proteins function as transcriptional regulators that are capable of binding AHL signal molecules. Once formed, LuxR-AHL complexes bind to target promoters of quorum-regulated genes. Quorum sensing is now known to be widespread among both Gram-positive and Gram-negative bacteria.

Bioluminescence in *Vibrio fischeri*

Bioluminescence in general is defined as an enzyme catalysed chemical reaction in which the energy released is used to produce an intermediate or product in an electronically excited state, which then emits a photon. It differs from fluorescence or phosphorescence as it is not depended on light absorbed. The mechanism for gene expression and the structure of the polycistronic message of the lux structural genes in *Vibrio fischeri* have been thoroughly characterised [9]. Briefly, there are two substrates, luciferin, which is a reduced flavin mononucleotide ($FMNH_2$), and a long chain ($7-16$ carbons) fatty aldehyde (RCHO). An external reductant acts via flavin mono-oxygenase oxidoreductase to catalyse the

reduction of FMN to $FMNH_2$, which binds to the enzyme and reacts with O_2 to form a 4a-peroxy-flavin intermediate. This complex oxidises the aldehyde to form the corresponding acid (RCOOH) and a highly stable luciferase-hydroxyflavin intermediate in its excited state, which decays slowly to its ground state emitting blue-green light $h\nu$ with a maximum intensity at about 490nm:

$$FMNH_2 + RCHO + O_2 \xrightarrow{\text{lucif.}} FMN + H_2O + RCOOH + h\nu$$

Transcription Control by LacR and λCI Repressor Proteins

Escherichia coli cells repress the expression of the lac operon when glucose is abundant in the growth medium. Only when the glucose level is low and the lactose level is high, the operon is fully expressed. The Lac repressor LacR is a 360 residue long protein that associates into a homotetramer. It contains a helix-turn-helix (HTH) motif through which it interacts with DNA. This interaction represses transcription by hindering association with RNA polymerase and represents an example of "combinatorial control" widely seen in prokaryotes and eukaryotes [4]. The CI repressor of bacteriophage lambda is the key regulator in lambda's genetic switch, a bistable switch that underlies the phage's ability to efficiently use its two modes of development [21].

Flow Cytometry

Flow cytometry refers to the technique where microscopic particles are counted and examined as they pass in a hydro-dynamically focused fluid stream through a measuring point surrounded by an array of detectors. Previously, flow cytometry analyses were performed by us using a BD LSRII flow cytometer equipped with 405nm, 488nm and 633nm lasers. 488nm laser was used for *gfp* and yellow fluorescent protein (*yfp*) quantification.

Experimental Setup and Implementation

We have shown that an *in vivo* system [10] can potentially be used to mimic a RS flip-flop [13] and have quantified its performance using flow cytometry. The presence or absence of the inducers IPTG or AHL in combination with temperature shift acts as an input signal. The toggle switch comprising of structural genes for reporter/output proteins fused to promoter regions that are regulated by input signals is visualised as a RS flip-flop. The functional modularity of input and output circuits is maintained so that the artificial GRN used can be easily extended for future studies.

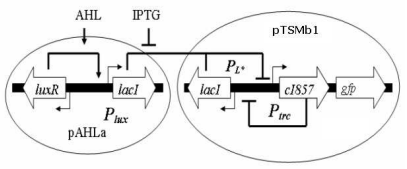


Fig.3. A schematic diagram of an AHL biosensor module interfaced with the genetic toggle switch adapted from [10]. The transgenic artificial GRN consists of a bistable genetic toggle switch [8] which is interfaced with genes from the lux operon [6] of the quorum sensing signalling pathway of *Vibrio fischeri* [22].

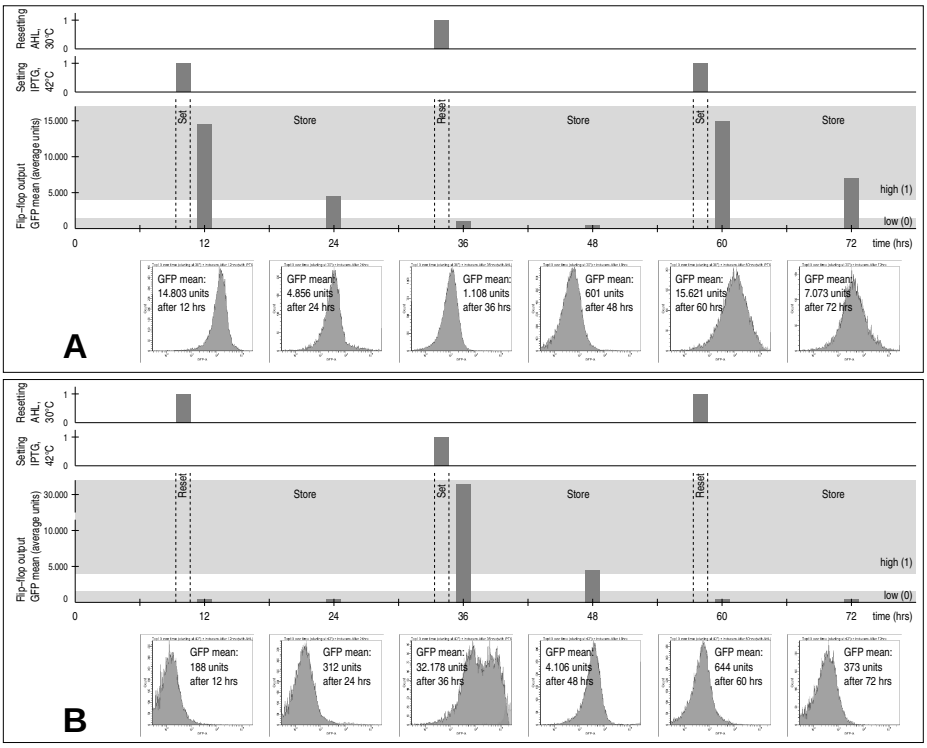


Fig. 4. Inducer-dependent switching. Repeated activation and deactivation of the toggle switch based on inducers and temperature. Temperature was switched every 24 hours. Cells were incubated with inducers for 12 hours, followed by growth for 12 hours without inducers, initially kept at 30°C (**A**) and 42°C (**B**). The cells successfully switched states thrice.

This design endows cells with two distinct phenotypic states: where the λCI activity is high and the expression of lacI is low (referred to as high or 1 state), or where the activity of LacR is high and the expression of λCI is low (referred to as low or 0 state). *gfp* is expressed only in the high λCI/low LacR state.

Results and Discussion

For co-relational purposes, all experiments were conducted with both BL21 and Top10 strains of *Escherichia coli*. The concentration of IPTG used in all the experiments was 2mM and that of AHL was 1μM. Experiments conducted without the use of inducers, lead to an unreliable shifting of the states, signifying the use if inducers in a tightly, mutually regulated circuit. Further experiments conducted to understand the switching dynamics of the circuit revealed that in the current scenario, it was easier to switch from a high to a low state than vice versa. This discrepancy in switching behaviour is attributed to the differing

modes of elimination of LacR and λCI repressor proteins. While switching from low to high state, the repression due to IPTG-bound Lac repressor needs to be overcome by cell growth. Switching from high to low state is effected by immediate thermal degradation of the temperature-sensitive λCI. Experiments were also conducted to test the sustainability of states. The plug and play property of the circuit was examined by employing *yfp* as the reporter gene instead of *gfp*. As shown in Figure 4, the circuit could reliably mimic a RS flip-flop. The massive parallelism permissible by the use of large quantities of cells can compensate for the slow speed of switching. Further tests are to be performed to confirm this hypothesis.

5 Conclusions

The dynamical behaviour of GRNs is able to emulate information processing in terms of performing computations. In order to formalise this capability, we have introduced P systems of the form Π_{Hill} incorporating cooperativity and competitivity between transcription factors based on Hill kinetics. Its transformation to a dedicated iteration scheme for a discrete term rewriting mechanism with variable stoichiometric factors in Π_{Hill} provides a homogeneous approach that allows to compose GRNs towards functional units like computing agents. Examples address computational units (inverter, NAND gate, RS flip-flop), each defined by GRN, its ODE model, and the corresponding P system. Simulations of the dynamical behaviour quantitatively show the switching characteristics as well as the expected quality of binary output signals. Along with the prediction of GRNs acting as computational units, an experimental study *in vivo* demonstrates their practicability. Although the measurement of the dynamic switching behaviour was condensed to 12 points in time, they approximate the expected course. At the crossroad of modelling, simulation, and verification of biological reaction networks, the potential of amalgamating analytic, stochastic, and algebraic approaches into the P systems framework seems promising for applications in systems biology to explore network functions.

Acknowledgements

This work is part of the ESIGNET project (Evolving Cell Signalling Networks *in silico*), which has received research funding from the European Community's Sixth Framework Programme (project no. 12789). Further funding from the Federal Ministry of Education and Research (BMBF, grant 0312704A) and from German Research Foundation (DFG, grant DI852/4-1) is acknowledged. We are very grateful to J.J. Collins for providing us with the plasmids and their sequences; to W. Pompe, G. Rödel, K. Ostermann, and L. Brusch from Dresden University of Technology for their scientific support and V. Helms from Saarland University for administrative support.

References

1. U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall, 2006
2. N. Barbacari, A. Profir, C. Zelinschi. Gene Regulatory Network Modelling by Membrane Systems. In R. Freund et al., eds., *PreProc. WMC6*, p. 162-178, 2005
3. F. Bernardini, M. Gheorghe, N. Krasnogor. Quorum Sensing P Systems. *Theor. Comput. Sci.* **371(1-2)**:20-33, 2007
4. N.E. Buchler, U. Gerland, T. Hwa. On Schemes of Combinatorial Transcription Logic. *Proc. Natl. Acad. Sci. USA* **100(9)**:5136-5141, 2003
5. N. Busi, C. Zandron. Computing with Genetic Gates, Proteins, and Membranes. In H.J. Hoogeboom et al., eds., *Membrane Computing. LNCS* **4361**:233-249, 2006
6. J. Engebrecht, M. Silverman. Identification of Genes and Gene Products Necessary for Bacterial Bioluminescence. *Proc. Natl. Acad. Sci. USA* **81**:4154-4158, 1984
7. F. Fontana, V. Manca. Discrete Solutions to Differential Equations by Metabolic P Systems. *Theor. Comput. Sci.* **372(1)**:165-182, 2007
8. T.S. Gardner, C.R. Cantor, J.J. Collins. Construction of a Genetic Toggle Switch in *Escherichia coli. Nature* **403**:339-342, 2000
9. J.W. Hastings, K.H. Nealson. Bacterial Bioluminescence. *Annu. Rev. Microbiol.* **31**:549-595, 1977
10. S. Hayat, K. Ostermann, L. Brusch, W. Pompe, G. Rödel. Towards *in vivo* Computing: Quantitative Analysis of an Artificial Gene Regulatory Network Behaving as a RS Flip-Flop and Simulating the System *in silico. Proc. Bionetics*, 2006
11. T. Hinze, T. Lenser, P. Dittrich. A Protein Substructure Based P System for Description and Analysis of Cell Signalling Networks. In H.J. Hoogeboom et al., eds., *Membrane Computing. LNCS* **4361**:409-423, 2006
12. S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, U. Kummer. Copasi – a COmplex PAthway SImulator. *Bioinformatics* **22**:3067-74, 2006
13. D.A. Huffman. The Synthesis of Sequential Switching Circuits. *Journal of the Franklin Institute* **257(3)**:161-190, **257(4)**:275-303, 1954
14. M.O. Magnasco. Chemical Kinetics is Turing Universal. *Physical Review Letters* **78(6)**:1190-1193, 1997
15. V. Manca. Metabolic P Systems for Biomolecular Dynamics. *Progress in Natural Sciences* **17(4)**:384-391, 2006
16. T. Mestl, E. Plahte, S.W. Omholt. A Mathematical Framework for Describing and Analysing Gene Regulatory Networks. *J. Theor. Biol.* **176**:291-300, 1995
17. M.B. Miller, B.L. Bassler. Quorum Sensing in Bacteria. *Annu. Rev. Microbiol.* **55**:165-199, 2001
18. A. Păun, M.J. Perez-Jimenez, F.J. Romero-Campero. Modeling Signal Transduction Using P Systems. In H.J. Hoogeboom et al., eds., *Membrane Computing. LNCS* **4361**:100-122, 2006
19. G. Păun. Computing with Membranes. *Journal of Computer and System Sciences* **61(1)**:108-143, 2000
20. G. Păun. *Membrane Computing: An Introduction.* Springer-Verlag Berlin 2002
21. M. Ptashne. *Genetic Switch: Phage Lambda and Higher Organisms.* Blackwell, Cambridge, MA, 1992
22. A.L. Schaefer, D.L. Val, B.L. Hanzelka, J.E. Cronan jr., E.P. Greenberg. Generation of Cell-to-Cell Signals in Quorum Sensing: Acyl Homoserine Lactone Synthase Activity of a Purified *Vibrio fischeri* LuxI Protein. *Proc. Natl. Acad. Sci. USA* **93**:9505-9509, 1996