Branko ŠTER, Rok GABER, Monika AVBELJ, Roman JERALA, Andrej DOBNIKAR

University of Ljubljana

Design of information processing in cells using artificial gene repressors

Abstract. *The progress of synthetic biology allows one to design artificial repressors that inhibit selected genes. Combination of repressors enables construction of NOR logical gates that could form the foundation for information processing within cells. The theoretical potentials and limitations of constructing NOR gates were analyzed. They could be experimentally realized in bacterial cells. The number of required artificial repressors was analysed and temporal simulations of an example function were performed.*

Streszczenie*. Postęp biologii syntetycznej pozwala zaprojektować sztuczne represory, które hamują rozwój wybranych genów. Połączenie represorów umożliwia budowę bramek logicznych NOR, które mogą stanowić podstawę do przetwarzania informacji w komórkach. Teoretyczny potencjał i ograniczenia budowy bramy NOR były analizowane. Mogą one być realizowane w komórkach bakteryjnych. Dokonano analizy liczby wymaganych represorów i wykonano symulacje funkcji przykładowych. (****Projektowanie przetwarzania informacji w komórkach przy użyciu sztucznych represorów****)*

Keywords: information processing, synthetic biology, artificial repressors, logical gates
Słowa kluczowe: przetwarzanie informacji, syntetyczna biologia, sztuczne represory, bramki logiczne

Introduction

Synthetic biology and computer science are beginning to collaborate in designing logical functions within the living cells. The combination of several repressors allows the construction of NOR gates, which is the universal operator for building any logical function. The paper shows the original procedures that are relevant to design any possible logical function of three input variables with NOR gates implemented with repressors. Such implementation could be experimentally realized in bacterial cells. For the purpose of this paper, however, only simulation results of the proposed circuits are shown.

The paper is organized as follows. The biological system for constructing a NOR gate with artificial repressors and for building any circuit of NOR gates is shown first, based on the scheme of biological processes and related equations. They are transformed into the differential equations, which are used for the simulation purposes. It is followed by the description of the genetic approach, for designing of all possible logical functions of three variables with NOR operators, together with some imposed limitations that are important for the experimental implementation, such as the maximal number of inputs to the NOR operators, maximal number of layers and operators and the minimal number of the zinc finger based repressors. Next, the results of genetic design are given for two runs, one with limitation of up to two inputs per NOR gates and the other with limitation of up to three inputs per NOR gates. In both cases, however, the minimal number of zinc finger repressors and the minimal number of inputs are searched for all possible functions of three variables. The simulation results of some optional function realized with genetically designed NOR circuits based on zinc finger repressors are finally given, followed by conclusion, where the recapitulation and the plans for future work are outlined.

Construction of NOR based circuits with zinc finger repressors

Regulation of the cellular response to the combination of input signals can be achieved at several levels, including protein and DNA level. While the response based on proteins is significantly faster then DNA-based response, the latter, which involves regulation of gene expression, allows using the same type of modules at several levels. The basic type of regulation involves transcriptional repression, where the constitutive gene expression is prevented by binding of a repressor to the appropriate region in front of the selected gene [1]. Cells contain many different repressors that bind to specific sequences and can be used to engineer modified response, however the number of well-characterised natural repressors limits the achievable complexity of synthetic information processing networks. Zinc fingers are protein domains that bind to a specific nucleotide sequence. Their main advantage is that they are modular, which allows us to design thousands of different variants that bind to selected sequence [2-4]. Therefore zinc fingers can be used as artificial repressors that limit the expression of selected genes, which are engineered to contain the zinc finger binding site [5]. This removes the limitation to naturally available repressors what is the main weakness of previously reported artificial bio-logic circuits [6-8]. Using zinc finger based transcriptional repressors is the foundation of complex artificial gene regulatory networks as information processing circuit. Introduction of two different zinc finger binding sites in the gene regulatory region creates a logical NOR gate with two inputs, which is open only in the absence of both zinc finger repressors, while any or both of them shut down gene transcription. If the gene in question is coding for another zinc finger repressor it can be used in the next level of information processing, allowing constructing, in principle, any level of logical information processing complexity.

Biological processes that are relevant to function of the artificial gene regulatory networks include:

1. Binding of zinc fingers to the gene regulatory (operator) region, which is a reversible reaction. In case of NOR gate, different combinations of zinc fingers can be bound or released.
2. Transcription of the gene in the absence of bound zinc finger repressor, which involves binding of RNA polymerase and its processing and release of the mRNA
3. Translation of the mRNA into the protein (another zinc finger)
4. Proteolytic degradation (turnover) of the zinc finger

The following figures present the bio-schemes of the 2-input NOR gate (Fig. 1) and the function f_{19} (Fig. 2), shown also in standard form in Fig. 5.

Each of the species, either being a single molecule or molecular complex can be formed or degraded through one or several pathways, which can be represented by differential equations. Rate constants and concentrations of each species can be estimated from the literature data or empirically determined. Experimental readout of the

performance of the system is to put at the end of the information processing chain a gene that releases a coloured or fluorescent protein reporter that can be assayed as the output.

The additional advantage of zinc finger repressors is that they have comparable properties in terms of stability and affinity for DNA, which allows us to use the same rate constants for all stages.

Transformation of biological processes to the system of differential equations

The dynamics of a two-input NOR gate which is connected with the (global) inputs can be described with a system of 21 (7 for each input zinc finger and 7 for output) differential equations. NOR gates not connected with the inputs require a system of 7 differential equations. Due to the lack of space it is not possible to show the whole system here. Therefore we show only the differential equations for the first zinc finger, which are replicated for all other input zinc fingers taking into account the adjustment for their respective inputs.

$$\begin{aligned} d[pBAD_{ZNF1}]/dt &= kr_{AraC}[pBAD_{ZNF1}'] \\ &+ kr_{RNAP\text{-}pBAD}[pBAD_{ZNF1}^*] \\ &+ k_{trsca}[pBAD_{ZNF1}^*] \\ &- kf_{AraC}[pBAD_{ZNF1}][AraC^*] \\ &- kf_{RNAP\text{-}pBAD}[pBAD_{ZNF1}][RNAP] \end{aligned}$$

$$\begin{aligned} d[pBAD_{ZNF1}^*]/dt &= kf_{RNAP\text{-}pBAD}[pBAD_{ZNF1}][RNAP] \\ &- kr_{RNAP\text{-}pBAD}[pBAD_{ZNF1}^*] \\ &- k_{trsca}[pBAD_{ZNF1}^*] \end{aligned}$$

$$\begin{aligned} d[pBAD_{ZNF1}']/dt &= kf_{AraC}[pBAD_{ZNF1}][AraC^*] \\ &- kr_{AraC}[pBAD_{ZNF1}'] \end{aligned}$$

$$\begin{aligned} d[AraC^*]/dt &= kr_{AraC}[pBAD_{ZNF1}'] \\ &+ kr_{ara}[AraC'] \\ &- kf_{ara}[AraC^*][arabinose] \\ &- kf_{AraC}[pBAD_{ZNF1}][AraC^*] \end{aligned}$$

$$\begin{aligned} d[AraC']/dt &= kf_{ara}[AraC^*][arabinose] \\ &- kr_{ara}[AraC'] \end{aligned}$$

$$\begin{aligned} d[mRNA_{ZNF1}]/dt &= k_{trscc}[pBAD_{ZNF1}^*] \\ &- k_{degm}[mRNA_{ZNF1}] \end{aligned}$$

$$\begin{aligned} d[ZNF1]/dt &= k_{trsl}[mRNA_{ZNF1}] \\ &+ kr_{ZNF1}[pA_{Rep\,{}_{iZNF1}}] \\ &+ kr_{ZNF1}[pA_{Rep\,{}_{iZNF1\text{-}2}}] \\ &- kf_{ZNF1}[pA_{Rep\,{}_{iZNF2}}][ZNF1] \\ &- kf_{ZNF1}[pA_{Rep}][ZNF1] \\ &- k_{degZNF}[ZNF1] \end{aligned}$$

The differential equations for simulating the NOR function are derived from the limited set of biological equations, representing a simplified description of actual processes that are happening in biological cells. We consider several processes: transition of inducible promoter that controls expression of first level zinc finger from active $pBAD_{ZNF1}^*$ to undefined $pBAD_{ZNF1}$ and later to inactive $pBAD_{ZNF1}'$ state or vice versa; transition of active repressor that inhibits first level zinc finger expression AraC* to inactive repressor AraC' or vice versa; changing of mRNA concentration coding for individual zinc finger protein $mRNA_{ZNF1}$; changing of individual zinc finger protein concentration ZNF1. Dynamics of described processes can be defined by observing the concentration of selected parameters. Quantity of individual parameter in system is given as a relative intracellular concentration of biological counterpart and is labeled with square brackets. Dynamics of system is defined with concentration of individual parameters and constants that reflect dynamics previously observed in natural systems.

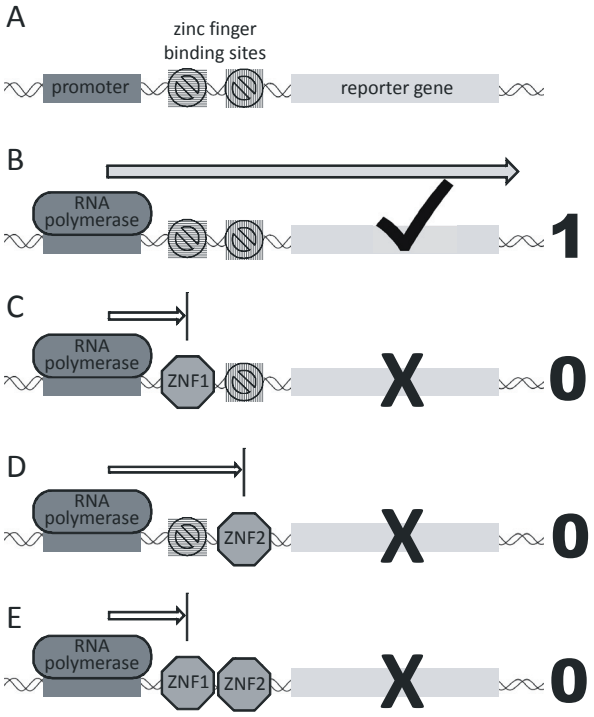


Fig.1. Schematic presentation of biologically realized NOR gate

DNA cassette coding for NOR gate consists of constitutive promoter, binding sites for two different DNA binding proteins (zinc fingers ZNF) and reporter gene (A). When neither of zinc fingers is present in cell, RNA polymerase binds to the promoter region, slides through both zinc finger binding sites and transcribes the reporter gene. The result of the reporter gene transcription can be experimentaly observed as a fluorescence signal in cell (B). Presence of one (C or D) or both zinc fingers in the cell blocks progression of the RNA polymerase and prevents transcription of reporter gene. In that case, no fluorescence can be observed in biological cell (E). This arrangement results in the logical NOR gate.

Because transitions in biological systems can occur in both directions (for instance AraC' to AraC* and vice versa) with different dynamics, all constants have their forward and reverse version. kf_{ara} and kr_{ara} define association dynamics between inducer arabinose and AraC* and therefore determine dynamics of transition between AraC* and AraC'; kf_{AraC} and kr_{AraC} define association dynamics of AraC* and $pBAD_{ZNF1}$ and therefore determine dynamics of transition between $pBAD_{ZNF1}$ and $pBAD_{ZNF1}'$; $kf_{RNAP\text{-}pBAD}$ and $kr_{RNAP\text{-}pBAD}$ define association dynamics between RNA polymerase (RNAP) and $pBAD_{ZNF1}$ and therefore determine dynamics of transition between $pBAD_{ZNF1}$ and $pBAD_{ZNF1}^*$; k_{trsca} defines dynamics of total transcription and therefore determines transition from $pBAD_{ZNF1}^*$ to $pBAD_{ZNF}$; k_{trscc} defines dynamics of successful transcription and therefore determines $mRNA_{ZNF1}$ concentration; k_{degm} defines dynamics of $mRNA_{ZNF1}$ degradation and therefore also determines $mRNA_{ZNF1}$ concentration; k_{trsl} defines dynamics of $mRNA_{ZNF1}$ translation to ZNF1 protein. All association events temporarily reduce concentrations of free associating partners. Two parameters have constant values and are not changed with association events. These are: arabinose and RNAP. Initial concentrations of individual parameters should be given at simulation start point. $pBAD_{ZNF1}$ and AraC* have positive initial values all other parameters have initial value 0.

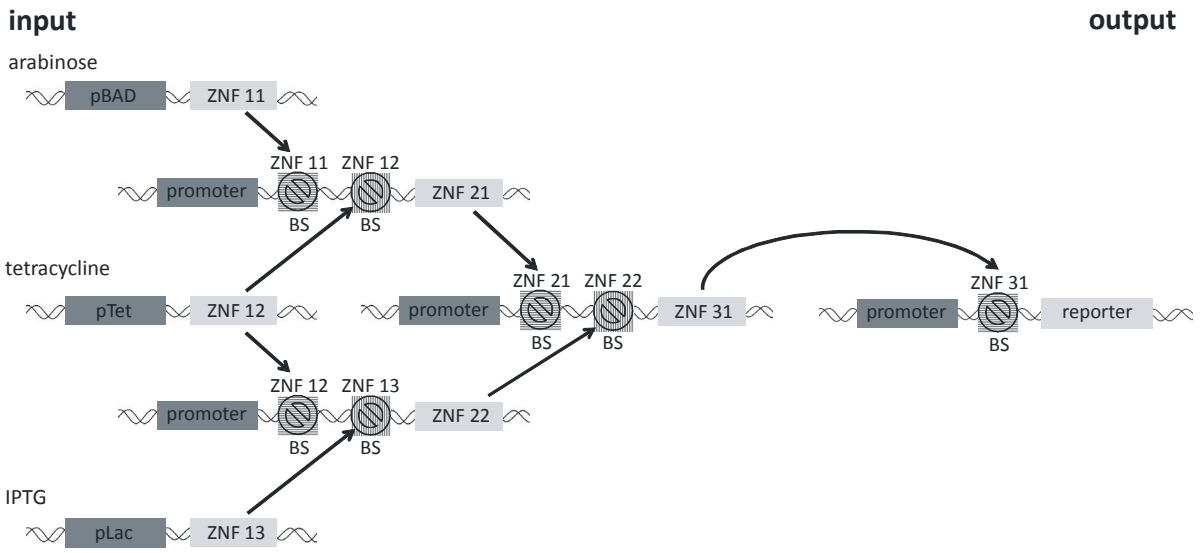


Fig.2. Schematic presentation of biologically realized logical function f_{19} (see Fig.5). pBAD, pTet and pLac are inducible promoters activated by chemical inputs arabinose, tetracycline and IPTG, respectively. Induction results in production of the first level zinc fingers. Those affect the first level of NOR logic gates. Second layer zinc fingers are outputs of the first layer logical gates and enter as inputs of the second level NOR gates. Those control expression of third level zinc finger, which controls expression of the reporter gene as the output of the synthetic genetic circuit.

However, the most important quantities for the simulation are zinc fingers and reporter gene, which serves as the output of the circuit.

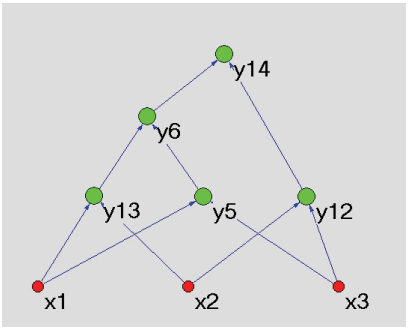
Genetic design of logical functions with NOR elements

In order to prepare the theoretical background for the design of synthetic information processing networks in cells, the natural approach of designing any logical function of three variables based on exclusively NOR operators (either up to two and up to three inputs) is taken. The solution is based on genetic algorithm (GA), which searches for an optimal implementation of each possible logical function. The goal is to identify the required complexity in terms of the number or artificial repressors and number of levels, necessary to implement any of 256 possible logical functions. The identification function number is represented with the decimal number between 0 and 255, which is obtained from the binary function values $f_7f_6f_5f_4f_3f_2f_1f_0$, where f_i is the binary function value of the i-th input combination and index i means the decimal value of the binary input combination. For example, the function f_6 is defined with $6_{10} = 00000110_2$.

A chromosome consists of integer numbers in groups of four. The first number represents the level where the logical operator resides, the other three numbers (after the comma) correspond to inputs to the operator. If a level is zero, it means the absence of the operator. If an input is zero, it means the absence of the input. Each chromosome has $N = 14$ groups because we set this as a reasonable maximal number of operators that would be allowed for the circuit solution of the logical function. Each group describes one operator; its number is shown in parenthesis and is not actually coded in the chromosome. Figure 3 shows an example of the circuit, obtained with GA and its corresponding chromosome for the function $f_6 = x_1'x_2'x_3$ V $x_1'x_2x_3'$, realized with the circuit of 5 NOR operators with 2 inputs in 3 layers. In this case only five operators are actually existent (y5, y6, y12, y13 and y14, bolded in Figure 3). The global inputs x_1, x_2 and x_3 are coded as N+1, N+2 and N+3 (15, 16 and 17).

The parameters of the GA are the following:

- $N = 14$ (maximal number of operators or groups)
- $N_L = 6$ (maximal number of logic levels)
- $n = 2$ or 3 (maximal number of inputs to the NOR operator)
- $gn = 3$ (number of global inputs or variables of logical function)
- $nC = 1000$ (number of chromosomes or size of population)
- $T = 2000$ (number of generations or steps of evolution/design)
- $p_{mut} = 0.01$ (mutation probability)
- crossover not applied



(1) 0, 17 0 15	(2) 0, 15 0 0	(3) 0, 17 0 15
(4) 0, 15 17 16	**(5) 1, 17 0 15**	**(6) 4, 0 5 13**
(7) 0, 16 0 17	(8) 0, 15 0 17	(9) 0, 16 0 15
(10) 0, 16 15 0	(11) 0, 17 0 15	**(12) 1, 17 0 16**
(13) 1, 16 0 15	**(14) 6, 0 6 12**	

Fig.3. An example of the chromosome for a logical function of three variables f_6 and its circuit consisting of 5 NOR operators with two inputs

The evaluation or fitness function of the GA can be described with the expression:

(1) $$\frac{u^4}{\sqrt[5]{0.8N_{zn}+0.2N_{in}}}$$

(where u is the proportion of the correct among all 2^{gn} logical values, or 8 in case $gn = 3$. N_{zn} is the number of zinc fingers and N_{in} the total number of inputs to all operators.

The GA is searching for the circuit with the maximal proportion of correct values, minimal number of zinc fingers and with the minimal number of inputs to all operators. The GA is therefore looking for the chromosome or its phenotype (the circuit of NOR elements) with maximal value of the fitness function. With p_{mut} the numbers within the chromosome are randomized, the level number within the $N_L = 6$ and the input number within the $N = 14$.

Results of the genetic algorithm

The GA found optimal or near-optimal (due to limited number of generations) representations of all 256 logical functions over three input variables, from f_0 to f_{255}. The most complex function is f_{105}, which corresponds to XNOR (negation of XOR) of three variables, and requires 8 up to 3-input NOR operators. The circuit of f_{105} is shown in Figure 4.

The distribution of the function representations regarding to the number of NOR operators, the number of inputs and the number of zinc fingers is outlined in Figure 6. In this case the three-input, two-input and one-input NOR gates (the latter is simply the invertor) were at our disposal. Figure 7 shows the same distributions when only two-input and one-input gates are applied.

Simulation results

The purpose of the simulation is to verify the logical correctness and to estimate the propagation delays and concentration values of the relevant signals. It is of interest to simulate the time course of the response of NOR gates made of zinc fingers as a function of selected values of the inputs. For the purpose of simulation the function f_{19} is chosen, shown also in Figure 5 with NOR gates:

$$f_{19} = \overline{(x_1 \downarrow x_2) \downarrow (x_2 \downarrow x_3)}$$

Inputs x_1, x_2 and x_3 consist of different concentrations of inductors, arabinose, tetracycline and IPTG, respectively. In order to capture the qualitative behaviour of the system the input concentration of 1000 units is used as the logical 1, while the absence of an reactant at the input corresponds to the logical 0.

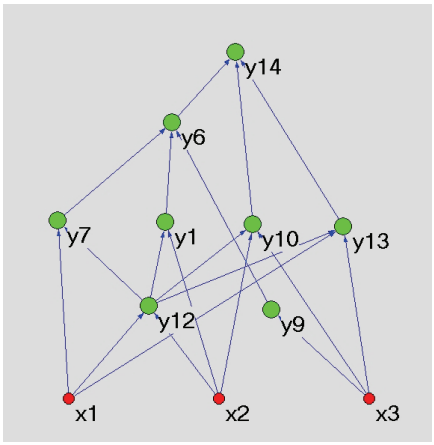


Fig.4. Realization of function f_{105} with up to 3-input NOR operators

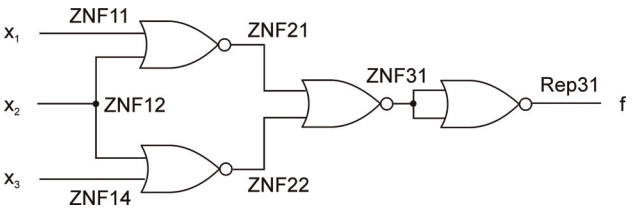


Fig.5. Logical function f_{19} for the simulation

The correct logical values for $f_{19} = 00010011_2$ emerge for each of the 8 input combinations. Figure 8 shows three typical cases. The propagation delay in case of function value 0 is approximately 10, while in case of value 1 it is about 100 time units.

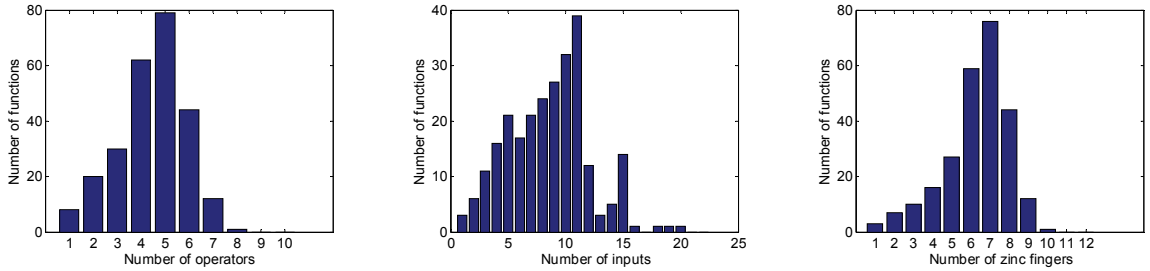


Fig.6. The distribution of the function representations regarding to the number of NOR operators, the number of inputs and the number of zinc fingers for up to three-input NOR gates

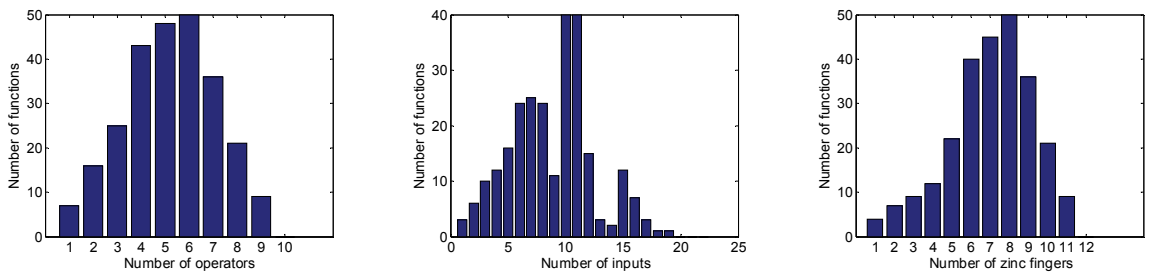


Fig.7. The distribution of the function representations regarding to the number of NOR operators, the number of inputs and the number of zinc fingers for up to two-input NOR gates

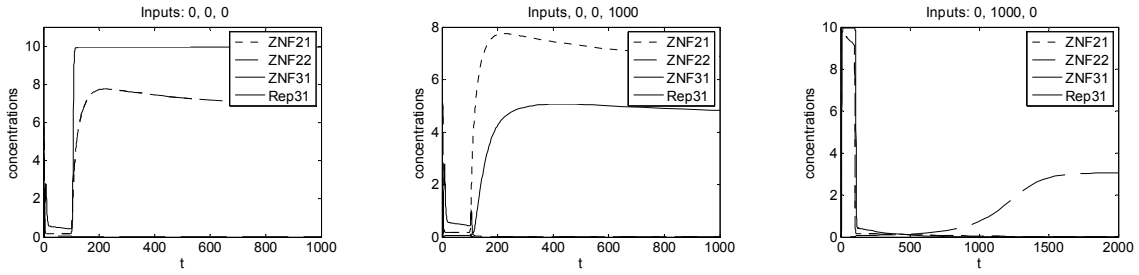


Fig.8. Zinc fingers and the reporter gene in case of three logical combinations at the input

Conclusion

Synthetic biological information processing devices have great potentials as we can use them to regulate biological processes, either for industrial or medical applications. Additionally, the ability of biological systems to self replicate and operate at high energetic efficiency represents the opportunity for other information processing applications. It is demonstrated in the paper that artificial zinc finger-based repressors can represent the core module of complex information processing networks. Computational simulation showed that all possible functions of three inputs can be realized by experimentally accessible number of zinc fingers. The genetic algorithm can be used to optimize construction of synthetic information processing network, where one can optimize either the number of levels, number of used zinc fingers, type of NOR gates or their combination, which is important for the experimental implementation of the information processing systems. The plan for future work is to confirm the results with the experimental work in the lab.

REFERENCES

[1] Browning D.F., Busby S.J. The regulation of bacterial transcription initiation, *Nature reviews, Microbiology,* 2 (2004), 57-65

[2] Hurt J. et al. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection, *Proceedings of the National Academy of Sciences of the United States of America,* 100 (2003), 12271-6

[3] Gonzalez B. et al. Modular system for the construction of zinc-finger libraries and proteins, *Nature protocols,* 5 (2010), 791-810

[4] Fu F. et al. Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays, *Nucleic acids research,* 37 (2009), D279-283

[5] Sera T. Zinc-finger-based artificial transcription factors and their applications, *Advanced drug delivery reviews,* 61 (2009), 513-526

[6] Tamsir A., Tabor J.J., Voigt C. Robust multicellular computing using genetically encoded NOR gates and chemical “wires”, *Nature,* 1-4 (2010) .doi:10.1038/ nature09565

[7] Regot S. et al. Distributed biological computation with multicellular engineered networks, *Nature,* 2-6 (2010), doi:10.1038/nature09679

[8] Anderson J.C., Voigt C., Arkin, A.P. Environmental signal integration by a modular AND gate, *Molecular systems biology,* 3 (2007), 133

***Authors**: Branko Šter, Faculty of Computer and Information Science, University of Ljubljana, Tržaška 25, 1000 Ljubljana, E-mail: branko.ster@fri.uni-lj.si; Rok Gaber, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana and Centre of Excellence EN-FIST E-mail: rok.gaber@ki.si; Monika Avbelj, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana and Centre of Excellence EN-FIST, E-mail: monika.avbelj@ki.si; Roman Jerala, National Institute of Chemistry, Hajdrihova 19, 1000 Ljubljana, Faculty of chemistry and chemical technology, University of Ljubljana and Centre of Excellence EN-FIST E-mail: roman.jerala@ki.si; Andrej Dobnikar, Faculty of Computer and Information Science, University of Ljubljana, Tržaška25,1000 Ljubljana, E-mail: andrej.dobnikar@fri.uni-lj.si.*