Miha MOŠKON, Nikolaj ZIMIC, Martin STRAŽAR, Miha MRAZ

University of Ljubljana

Comparison of selected performances of biological and electronic information processing structures

Abstract. *We present the information processing perspective on biological systems. Several metrics, similar to the ones used in digital electronic circuits, are introduced. These metrics allow us to compare biological information processing structures with their electronic counterparts, to define the ones with the best dynamical properties, analyse their compatibility and most importantly, automatize their design. Regarding the metric values obtained and used on a simple example, target applications of synthetic information processing biological structures are discussed.*

Streszczenie. *W artykule opisano zagadnienie przepływu informacji w systemach biologicznych. Zastosowano tu odwzorowanie na elementach i obwodach elektronicznych, co pozwoliło na analizę ich własności, w tym dynamicznych oraz zautomatyzowanie projektowania takich modeli. Zawarto także omówienie otrzymanych wyników badań. (**Porównanie wybranych parametrów struktur przepływu informacji w systemie biologicznych i elektronicznym**).*

Keywords: unconventional computing, gene regulatory networks, modelling, computational biology
Słowa kluczowe: obliczenia niekonwencjonalne, GRN, modelowanie, biologia obliczeniowa.

Introduction

Engineered information processing structures were up until recently mainly based on electronic circuits. With the propagation of the computer science to different branches new processing platforms appeared also in the field of biological systems. Numerous synthetic biological information processing structures, especially on the basis of gene regulatory networks, were already constructed in recent years [1]. In this paper we present the basic dynamics of these systems and introduce the metrics similar to the ones used in digital electronic circuits. Establishment of such metrics allows us to evaluate the dynamics of biological structures in the meaning of information processing capabilities and make a comparison of their selected performances with digital electronic circuits. Moreover they allow us to establish criteria for improvements of dynamics of analysed biological systems and for the automation of their design. Metrics are essential to define appropriate fitness functions which are utilized in the metaheuristics for solution space investigation in order to find biological systems which reflect the desired dynamics.

The paper is organized as follows. Firstly, we give a basic description of dynamics in gene regulatory networks, which is followed by an introduction of metrics and demonstration of their evaluation on a sample biological circuit. With the support of evaluated metrics fitness function is established, which is used in a variation of genetic algorithm for optimizing the behaviour of previously described biological system. The behaviour of biological systems as information processing structures is compared to digital electronic circuits. Their target applications are also discussed.

Introduction to dynamics of gene regulatory networks

Networks of interacting genes located within the cell are called *gene regulatory networks*. Their dynamics is based on the gene expression which can be presented by two processes, i.e. *transcription* and *translation*. In the process of transcription, part of the gene called *protein coding sequence* is transcribed to *messenger RNA (mRNA)*, which is in the process of translation translated into a target protein. While we can presume that translation is an unconditional process, transcription requests the presence, absence or specific combination of designated proteins which are called *transcription factors*. We can divide these proteins in two groups regarding their influence on transcription. Transcription factors that activate transcription are called *activators* and transcription factors that repress transcription are called *repressors*. Interaction of genes in gene regulatory networks is achieved if protein coding sequences encode the transcription factors for some other or even for their own genes. With the methods and procedures of *synthetic biology* one can define the transcription factors that will control the expression of a certain gene and the proteins this gene will express and therefore construct *synthetic biological systems*. Gene regulatory networks with specific functionalities, also in the context of information processing, can be constructed accordingly. Three basic examples of gene regulatory networks presenting information processing structures are *inverter*, *NOR gate* and *oscillator*. Inverter can be constructed with the gene expressing an output protein and repressed by an external input (see Fig. 1(a)). If the circuit is extended to two repressors, i.e. two external inputs, where only one needs to be present in order to effectively repress the transcription, the circuit's behaviour reflects the NOR gate (see Fig.1(b)). If the gene represses its own expression we can achieve oscillatory behaviour (see Fig.1(c)). These circuits can be modularly connected into more complex biological circuits, but their characteristics have to be regarded in order to analyse their compatibility. Characteristics of each gene regulatory network do not depend only on its topology, but also on involved chemical species and reactions among them. Detailed analyses must therefore be performed before the construction of such systems. These analyses are mostly based on mathematical models.

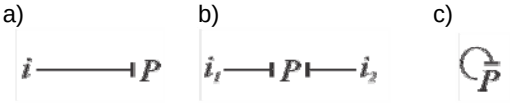

Fig.1. Gene regulatory networks presenting inverter (a), NOR gate (b) and oscillator (c), where i, i_1 and i_2 denote external inputs and P output protein defined with the protein coding sequence.

Information processing perspective on gene regulatory networks

Several metrics that are similar to characteristics which describe the properties of digital electronic circuits are used in order to objectively analyse the information processing capabilities of target biological systems. We can use the

evaluated metrics in order to define the weaknesses of analysed biological circuits, determine the ones with the best dynamical properties and study compatibility among them. Moreover, they can be used for the automation of design of synthetic biological systems. This section describes an introduction of such metrics from information processing perspective.

In order to use biological systems as information processing structures information has to be presented with designated chemical species in the system. This role is assigned to certain proteins, i.e. *output species*, which are expressed by certain segment of the biological system and usually also have a role of transcription factors, i.e. *inputs*, to other segments of the biological system. Information carried by these proteins can be encoded with their concentrations. While the behaviour of biological systems is more or less deterministic, logic levels of input and output species can be defined. For the reasons of its convenience binary logic is used in the majority of modern computer structures and therefore also in biological systems. Information is presented by the presence (logical 1), respectively absence (logical 0) of observed chemical species. In order to use these structures in more complex biological systems logic levels have to be defined more accurately, similarly as in digital electronic circuits. These levels are defined based on the concentration values the system achieves in optimal scenario and based on the maximal expected noise system will be exposed to (*external noise*) and system will express (*intrinsic and extrinsic noise*) in its expected environment [2–4]. Absolute noise values can differ significantly for concentrations that present different logic states and must be evaluated for the concentrations presenting logic 0 and concentrations presenting logic 1 separately. Various techniques for noise evaluation using the experimental results [5] or using mathematical approaches [6–8] exist and can therefore be used for the estimation of noise values in both logic states. Let's presume that C_{OL} and C_{OH} present ideal concentrations of output specie presenting logic 0, logic 1 respectively. They can be determined as extreme, i.e. minimal and maximal concentration levels without the noise consideration whether on simulation whether on experimental results. Furthermore, let's presume that N_L and N_H present maximal expected noise in concentrations around C_{OL}, C_{OH} respectively. Output specie concentrations that have to be accepted as a valid signal by other segments of the system can thus be calculated as

$$C_{OL(\max)} = C_{OL} + N_L \tag{1}$$

$$C_{OH(\min)} = C_{OH} - N_H \tag{2}$$

where $C_{OL(\max)}$ and $C_{OH(\min)}$ present maximal low level output concentration and minimal high level output concentration respectively. The region of concentrations in the interval $]C_{OL(\max)}, C_{OH(\min)}[$ has no valid interpretation and is called *invalid range* in digital electronic circuits. The size of invalid range on the one hand defines the sensitivity of the system to the noise and on the other hand the productivity of the system, i.e. larger invalid ranges reflect in larger robustness, but larger energy consumption for the transition from one logic state to another. Trade-off between sensitivity and productivity must therefore be sought [3]. Characteristics describing logic levels in biological systems slightly differ from the ones in digital electronic circuits. We need to define input signal levels separately, i.e. concentration levels that bring the system to valid output concentrations, i.e. concentrations in the interval $[0, C_{OL(\max)}]$ or $[C_{OH(\min)}, \infty]$. Two more characteristics are therefore introduced, i.e. maximal low level input concentration ($C_{IL(\max)}$) and minimal high level input concentration ($C_{IH(\min)}$). We are able to analyse the compatibility of various biological systems in this context: two biological systems are compatible in the meaning of logic levels if $C_{OL(\max)}(out) \leq C_{IL(\max)}(in)$ and if $C_{OH(\min)}(out) \geq C_{IH(\min)}(in)$, where *out* denotes output and *in* input biological system. Furthermore, switching times, i.e. rise time and fall time, can be measured. Switching time is defined by the time the concentration of output chemical specie is located within the invalid range after the switch is initiated with the modification of input signal. We can evaluate this times with the initiation of a switch with a boundary input values, i.e. $C_{IL(\max)}$ and $C_{IH(\min)}$, and measuring the transitions of the output specie from the value $C_{OL(\max)}$ to $C_{OH(\min)}$ when evaluating the rise time and from the value $C_{OH(\min)}$ to $C_{OL(\max)}$ when evaluating the fall time.

An example of tunable biological oscillator

Establishment and evaluation of metrics described in preceding section gives us objective criteria that can be used in the design of synthetic biological systems with desired dynamics. Evaluated metrics can help us with the investigation of compatibility among different systems, with their comparison and most importantly they can be a basis for the establishment of fitness functions for various metaheuristic approaches, which can be used for automatic design of synthetic biological systems with desired dynamics. In this section we will demonstrate an application of *genetic algorithm* to the design of a simple biological oscillator (see Fig. 1(c)), which can be described with the following ordinary differential equation (ODE) based model [9]:

$$\frac{dM}{dt} = \frac{r_M}{1+\left(\frac{P(t)}{k}\right)^n} - q_M \cdot M(t) \tag{3}$$

$$\frac{dP}{dt} = r_P M(t-\tau)^m - q_P \cdot P(t) \tag{4}$$

where $M(t)$ and $P(t)$ present the concentrations of mRNA and observed protein respectively, r_M and r_P production rates, q_M and q_P degradation rates, τ delay in translation, n Hill coefficient, m nonlinearity in protein synthesis cascade and k scaling constant. Let's presume that the parameters that define the ODE models are as follows: $r_M = r_P = 1h^{-1}$, $q_M = q_P = 0.21h^{-1}$, $n=2$, $m=3$, $\tau=4h$ and $k=1$ [9] and that the role of output specie is designated to protein P (its time evolution is presented in Fig. 2).

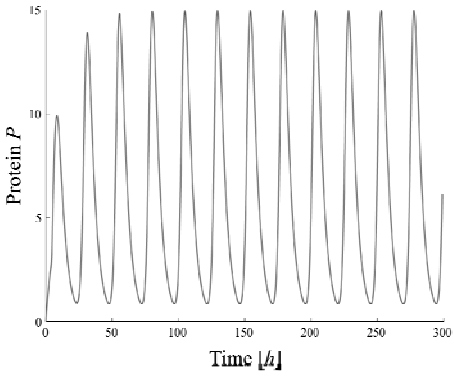

Fig.2. Time evolution of protein P according to the model presented in Equations (3) and (4) and the parameter values derived from [9].

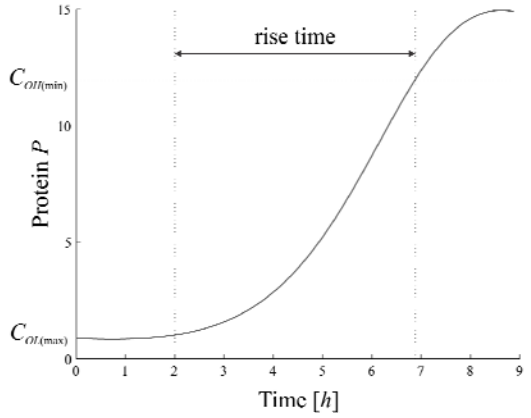

Fig.3. Evaluating the rise time of protein P according to the model presented in Equations (3) and (4) and the parameter values derived from [9]

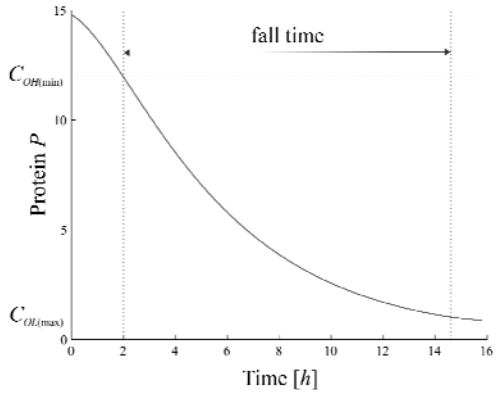

Fig.4. Evaluating the fall time of protein P according to the model presented in Equations (3) and (4) and the parameter values derived from [9]

Introduced model will be used to demonstrate the evaluation of metrics established in preceding section and to determine the parameter values according to which the model reflects the desired behaviour. Presented model is relatively simple, but its dynamics is on the other hand complex enough to demonstrate the introduced concept. While its behaviour is independent of external inputs, only the metrics describing the properties of output specie will be evaluated.

Firstly, we evaluate the optimal concentration levels that present logic 0 and logic 1 as minimal and maximal concentration values, i.e. $C_{OL} \approx 0.85$ and $C_{OH} \approx 15.3$. Secondly noise has to be evaluated. According to [10] the coefficient of variation in a system similar as ours stays below 10% due to its autoregulatory structure. However, coefficient of variation can increase due to the system's low copy number which is dependent on number of plasmid copies within the cell and is a consequence of so called finite number effect [11]. We presume that the maximal noise values equal 20% of concentration levels. Maximal expected noise levels can be estimated as $NL = 0.17$ and $NH = 3.06$. According to Equations (1) and (2) $C_{OL(\max)}$ and $C_{OH(\min)}$ can be calculated as 1.02 and 12.24 respectively. Switching times can be measured as the times the concentrations are located within the interval $]C_{OL(\max)}, C_{OH(\min)}[=]1.02, 12.24[$. Rise time and fall time thus equal $4.88h$ and $12.6h$ respectively for the proposed parameter values (see Fig. 3 and Fig. 4).

We evaluated several metrics on the example of simple biological oscillator. Let's presume that our goal is to find the parameter values that would minimize the switching times of proposed model without significant changes in concentration levels. In order to achieve our goal the following variation of genetic algorithm is proposed:

1. Initialize the parameter set population.
2. Evaluate the population.
3. While the desired behaviour is not reached:
   i) Select parents and produce offspring according to the values of their fitness functions.
   ii) Mutate the resulting offspring.
   iii) Evaluate new members.
   iv) Select individual members for the next generation.

The fitness functions are defined according to the switching times of each member of population and according to the concentration levels members achieve for states presenting logic 0 and logic 1. In order to gain applicable results parameter set was limited to the values which are comparable to the ones found in nature, i.e. $r_M \in [0.36h^{-1}, 3600h^{-1}]$, $r_P \in [1h^{-1}, 1000h^{-1}]$ q_M, $q_P \in [0.01h^{-1}, 100h^{-1}]$, $n,m \in [1,10]$, $\tau \in [0h, 10h]$ and $k^n \in [0.1, 100]$ [12].

Given the constraints and the desired behaviour an approximation of optimal solution was determined with the genetic algorithm in approximately 300 iterations. The following parameter values were obtained: $r_M \approx 6.61h^{-1}$, $r_P \approx 2.11h^{-1}$, $q_M \approx 1.25h^{-1}$, $q_P \approx 10.66h^{-1}$, $n \approx 5.12$, $m \approx 7.37$, $\tau \approx 0.1h$ and $k \approx 1.3$. Simulation results obtained on the basis of given parameter values are presented in Fig. 5. Rise time and fall time equal approximately $0.1545h$ and $0.3991h$ respectively for the proposed parameter values. Even though the switching times drastically improved when compared to the original model, they are still much higher than the ones of digital electronic circuits. These times could be improved to some degree with fewer constraints (i.e. without predefined concentration levels) or with the employment of some different gene regulatory network topology (for example the one presented in [13]).

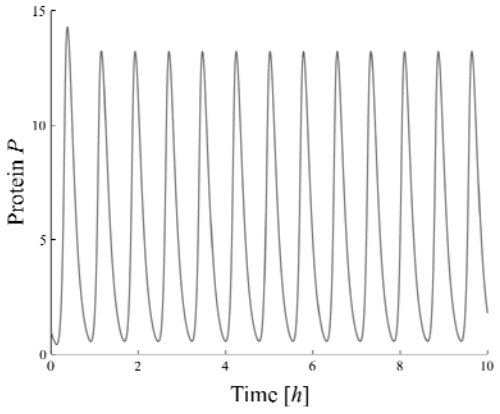

Fig.5. Time evolution of protein P according to the model presented in Equations (3) and (4) and the parameter values obtained with genetic algorithm.

Discussion

Our examples show that the time scales of biological circuits can be much higher than the ones of digital electronic circuits. Switching times can be improved to some degree, but because of the speed limits of elementary gene expression reactions such as transcription and translation the absolute boundary of these times is located much higher than the one in electronic circuits. The other

disadvantage of processing with gene regulatory networks is high presence of noise. On the other hand noise effects can be mostly eliminated and sometimes even exploited [14]. Similar as in digital electronic circuits, crosstalk is also inherent to biological systems. It derives from the unpredicted interactions among different chemical species and can be eliminated to some level with the use of so called orthogonal transcription-translation networks [15]. The problem that arises here is in the scalability of such circuits because the number of orthogonal proteins, which can be found in nature is too small to allow the construction of more complex synthetic biological networks [16]. One of the approaches to solve this problem is in the construction of artificial gene repressors [17]. In spite of all, there are numerous advantages of processing with biological systems, which can be exploited in their target applications. Noticeable accelerations of response times can be achieved with the accomplishment of massive parallelism which is inherent to these systems. Moreover their main advantage is in the possibility to combine the synthetic biological systems with the ones already present in nature in the fields such as biomedicine, pharmacy, agriculture and energy.

Conclusion

Numerous weaknesses of biological systems show that it is utopian to think that they will substitute modern information processing platforms in the near future. On the other hand there are many possibilities to use their advantages in other kind of applications, even in the meaning of information processing. Here we presented an introduction of metrics which allow us to define the biological systems with the best information processing capabilities, investigate their compatibility and most importantly establish fitness functions that can be used in various metaheuristics for the automation of the design of synthetic biological systems with desired functionalities.

The research was supported by the scientific-research programme Pervasive Computing (P2-0359) financed by Slovenian Research Agency in years from 2009 to 2012.

REFERENCES

[1] A. Cheng and T. K. Lu, Synthetic Biology: An Emerging Engineering Discipline, *Annual Review of Biomedical Engineering*, 14 (2012), 155–178

[2] M. Thattai and A. v. Oudenaarden, Intrinsic noise in gene regulatory networks, *Proceedings of the National Academy of Sciences*, 98 (2001), 8614–8619

[3] P. R. Patnaik, External, extrinsic and intrinsic noise in cellular systems: analogies and implications for protein synthesis, *Biotechnology and Molecular Biology Review*, 1 (2006), 121–127

[4] I. Lestas, J. Paulsson, N. E. Ross, and G. Vinnicombe, Noise in gene regulatory networks, *IEEE Transactions on Automatic Control*, 53 (2008), 189–200

[5] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, *Science*, 297 (2002), 1183–1186

[6] P. S. Swain, M. B. Elowitz, and E. D. Siggia, Intrinsic and extrinsic contributions to stochasticity in gene expression, *Proceedings of the National Academy of Sciences*, 99 (2002), 12795–12800

[7] D. Orrell, S. Ramsey, P. de Atauri, and H. Bolouri, A method for estimating stochastic noise in large genetic regulatory networks, *Bioinformatics*, 21 (2005), 208–217

[8] A. Hilfinger and J. Paulsson, Separating intrinsic from extrinsic fluctuations in dynamic biological systems, *Proceedings of the National Academy of Sciences*, 108 (2011), 12167–12172

[9] T. o. Scheper, D. Klinkenberg, C. Pennartz, and J. v. Pelt, A mathematical model for the intracellular circadian rhythm generator, *The Journal of Neuroscience*, 19 (1999), 40–47

[10] A. Becskei and L. Serrano, Engineering stability in gene networks by autoregulation, *Nature*, 405 (2000), 590–593

[11] W. K. Smits, O. P. Kuipers and J. W. Veening, Phenotypic variation in bacteria: the role of feedback regulation, *Nature Reviews Microbiology*, 4 (2006), 259–271

[12] U. Alon, *An Introduction to Systems Biology,* Chapman & Hall, (2007)

[13] M. B. Elowitz and S. Leibler, A synthetic oscillatory network of transcriptional regulators, *Nature*, 403 (2000), 335–338

[14] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler, Mechanisms of noise-resistance in genetic oscillators, *Proceedings of the National Academy of Sciences*, 99 (2002), 5988–5992

[15] W. An and J. W. Chin, Synthesis of orthogonal transcription-translation networks, *Proceedings of the National Academy of Sciences*, 106 (2009), 8477–8482

[16] T. K. Lu, A. S. Khalil and J. J. Collins, Next-generation synthetic gene networks, *Nature Biotechnology*, 27 (2009), 1139–1150

[17] B. Šter, R. Gaber, M. Avbelj, R. Jerala and A. Dobnikar, Design of information processing in cells using artificial gene repressors, *Przegląd Elektrotechniczny*, 88 (2012), 105–109

***Authors**: Miha Moškon, Faculty of Computer and Information Science, University of Ljubljana, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia, E-mail: miha.moskon@fri.uni-lj.si; Nikolaj Zimic, Faculty of Computer and Information Science, University of Ljubljana, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia, E-mail: niko.zimic@fri.uni-lj.si; Martin Stražar, Faculty of Computer and Information Science, University of Ljubljana, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia, E-mail: martin.strazar@ksok.si; Miha Mraz, Faculty of Computer and Information Science, University of Ljubljana, Tržaška cesta 25, SI-1000 Ljubljana, Slovenia, E-mail: miha.mraz@fri.uni-lj.si*.