

METHODOLOGY

OPEN ACCESS

Full open access to this and thousands of other papers at <http://www.la-press.com>.

Design of Synthetic Genetic Oscillators Using Evolutionary Optimization

Yen-Chang Chang, Chun-Liang Lin and Tanagorn Jennawasin

Department of Electrical Engineering, National Chung Hsing University, Taichung, Taiwan, ROC.
Corresponding author email: chunlin@dragon.nchu.edu.tw

Abstract: Efforts have been made to establish computer models of genetic oscillation. We have developed a real structured genetic algorithm (RSGA) which combines advantages of the traditional real genetic algorithm (RGA) with those of the structured genetic algorithm (SGA) and applies it as an optimization strategy for genetic oscillator design. For the generalized design, our proposed approach fulfils all types of genes by minimizing the order of oscillator while searching for the optimal network parameters. The design approach is shown to be capable of yielding genetic oscillators with a simpler structure while possessing satisfactory oscillating behavior. In silico experiments show effectiveness of the proposed algorithm to genetic oscillator design. In particular, it is shown that the proposed approach performs better than the traditional GAs in the sense that a cheaper structure of genetic oscillators can be obtained.

Keywords: synthetic biology, biological oscillator, real structure genetic algorithm

Evolutionary Bioinformatics 2013:9 137–150

doi: [10.4137/EBO.S11225](https://doi.org/10.4137/EBO.S11225)

This article is available from <http://www.la-press.com>.

© the author(s), publisher and licensee Libertas Academica Ltd.

This is an open access article. Unrestricted non-commercial use is permitted provided the original work is properly cited.



Introduction

Building a biological circuit by using artificial biologic components has been regarded as a significant tendency in synthetic biology, which is an interdisciplinary application in molecular biology and engineering. Information of the bio-computing process is based on the concentrations of biochemical molecules, like voltage signal in an electrical circuit. The ultimate goal is to construct systems which integrate very large scale bio-circuits and bio-chips, similar to those in the field of electrical engineering. Inspired by electronic logic elements, several recent studies have stated the possibility of realizing a genetic circuit, which is foreseen to have significant contributions to cancer medicine as well as to a variety of bio-energy sources.^{1,2}

For the design of bio-systems, several optimization algorithms were taken into consideration. Among such algorithms, genetic algorithms (GAs) (see, eg, Holland, 1975)³ have been widely applied to engineering optimization problems. The methods were firstly introduced by Holland³ in 1970s. Conventional genetic algorithms emphasize binary coding in chromosomes. However, binary genetic algorithms (BGAs) require excessive computing time when dealing with high-dimensional problems, and the premature convergence of solutions often occurs. To compensate for the disadvantages of BGAs, real genetic algorithms (RGAs) have made changes in floating point coding of chromosomes, and are proven to have advancement on computing speed and precision.⁴ Furthermore, structured genetic algorithms (SGAs)⁵ and hierarchical genetic algorithm (HGAs)⁶ have been proposed to address the premature convergence issue in solving optimization problems. Combining the advantages of RGA and SGA, a novel real structured genetic algorithm (RSGA)^{7,8} was proposed to solve complex multi-objective optimization problems. This approach exhibits advantages to simultaneously deal with the parameter and structure optimization problems based on a specific structural mapping operator used to determine appropriate numeric values of effective parameter genes in individuals.

Among a variety of problems in the bio-systems design, a problem concerning design of a genetic network with desired dynamic behaviors has attracted much attention from many researchers.

In the pioneering work of Elowitz and Leibler,⁹ a genetic ring oscillator, known as a repressilator, consisting of three genes: *tetR*, *lacI*, *λcI* was proposed. The three genes repress each other like a ring oscillation circuit comprising three BioLogic “NOT” gates, and induce oscillating light in the green fluorescent protein. The transcriptional regulation that generates the oscillatory behavior can be modeled based on several factors, including the transcription rates on repressor concentration, the decay rates of the protein, and the translation rates of mRNA.⁹ Chen and Chen¹⁰ proposed a GA-based approach to achieve a robust design of oscillatory behaviors in genetic networks.

Design of a genetic oscillator with desired amplitude, phase, and frequency could be thought as a tracking design problem in automatic control engineering. However, the stochastic intrinsic dynamic noise and extrinsic disturbances can perturb the oscillatory behaviors, and preservation of robustness to these noise perturbations has become an important issue.² The existing GA-based design searches for an optimal pair of parameters including the decay rates of protein concentrations and the transcription rates of mRNA. However, this approach can only solve the parameter optimization problem. In the presently described work, we combine the advantages of RGA and SGA to simultaneously determine the optimal structure and the optimal parameters for synthetic biological oscillators. In particular, designing the optimal transcription rates of mRNA and the decay rates of protein concentrations to track the desired sinusoidal signals is considered. Moreover, this paper proposes two methods to improve the tolerance of the intrinsic and extrinsic noises. The first method is based on the increase of searching undefined parameters to extend the dimension of the design parameters. The second method is to adjust hill coefficients to bring changes in protein concentration output. These approaches not only extend the design freedom but also simplify the structure of the oscillator module.

In summary, this paper attempts to develop a more efficient method than the traditional evolutionary algorithm-based approaches for solving the structure and parameter optimization problem while synthesizing the biological genetic oscillators and biological logic gates. In silico experiments show that the RSGA approach is effective in obtaining a genetic oscillator with the cheapest structure.

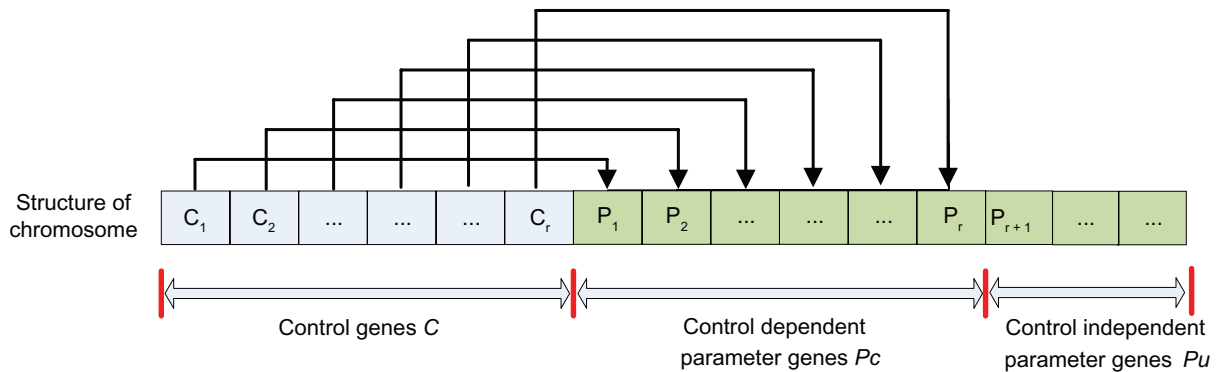


Figure 1. Chromosome structure of RSGA.

Real Structured Genetic Algorithm

RSGA is a combination of RGA with SGA, and is a method developed for both the optimal parameter and structure searching. The genetic operations of RSGA include reproduction, crossover, and mutation which were pioneered by Tsai, Huang, and Lin.^{7,8} The major difference between RSGA and SGA is that both the control genes and parameter genes in the former are real numbers (control genes in SGA are binary numbers whereas parameter genes are real numbers), which improves the mathematical mechanism due to the consistent operators of crossover and mutation, and requires less computing time.

For the current purpose in biogenetic oscillator design, a structured genetic mapping in the RSGA that varies parameters in three ways was applied. The structured mapping generates different structures and searches out the optimal solution structure according to the pre-specified fitness function. During evolution of offspring, it improves the individuals' fitness by crossover and mutation unceasingly. Based on this feature, the RSGA may achieve the optimal structure and parameters simultaneously for application in design of synthetic biological devices if the index reflects the oscillator structure and the desired amplitude, frequency, and phase. A demonstrative structure of its representation of the individual (chromosome) is shown in Figure 2 which shows a case consisting of 6 control genes, 6 control dependent parameter genes, and some control independent parameter genes.

Structured genetic mapping

The chromosomes $Y = \langle c, p \rangle$ of each individual in RSGA is an order set consisting of the set of control genes c and the set of parameter gene $p = \langle p_c, p_u \rangle$

with p_c and p_u representing the control dependent and control independent parameter genes respectively. The order set of control genes c and the order set of parameter genes p are both real numbers within (R_{\min}, R_{\max}) . The parameter genes of p_c are regulated by the control genes. The RSGA structured mapping is equipped with three gene operations: activate, inactivate, and linear ratio variation. The operating function of the control genes is determined by the correlation between the boundary values of the control genes, denoted B_{\max} and B_{\min} . When the value of control gene is greater than B_{\max} , the corresponding parameter genes are activated (ON). If its value is less than B_{\min} , the corresponding parameter genes are inactivated (OFF). When the value is within (B_{\min}, B_{\max}) , the parameter gene is regulated by a linear scaling factor. The functions of mapping for various statuses are explained as those shown in Figure 2.

For the current purpose, the order set of parameter genes contains the key parameters for transcription of the biogenetic oscillator which are to be determined, such as transcription rate and sensitivity of mRNA, and decay rates of protein and mRNA, etc. Variations of these parameters are alternated by the control genes during the computational evolutionary process.

The structured genetic mapping from Y to \tilde{Y} is defined as

$$\tilde{Y} = \langle c, \tilde{p} \rangle \quad (1)$$

where the chromosome \tilde{Y} represents an ordered set consisting of c and \tilde{p} , and

$$\tilde{p} = \langle \tilde{p}_c, p_u \rangle \quad (2)$$

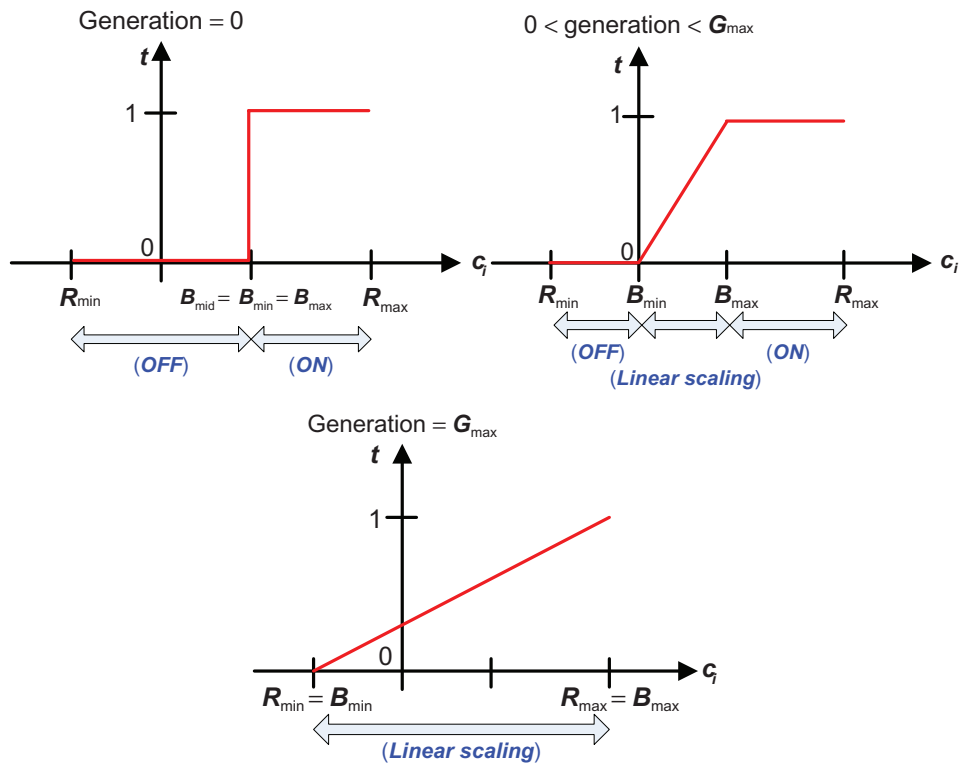


Figure 2. Activation function under boundary sizing.

An operator \otimes denoting the genetic switch is defined by

$$\tilde{P}_c = [c_i] \otimes [p_{ij}]_{n_j} \equiv \begin{cases} p_{ij}, & \text{if } B_{\max} \leq c_i \\ p_{ij}t, & \text{if } B_{\min} \leq c_i \leq B_{\max} \\ \phi, & \text{if } c_i \leq B_{\min} \end{cases} \quad (3)$$

where $j = 1, \dots, n_j$ and

$$t = \frac{c_i - B_{\min}}{B_{\max} - B_{\min}}$$

with ϕ denoting an empty element. Variations of B_{\max} and B_{\min} are generationally dependent and are defined by the following rules:

$$\begin{cases} B_{\max} = B_{\min} = B_{mid}, & \text{if } g_i = g_{init} \\ B_{\max} = R_{mid} + 0.5\Delta B, \\ \quad B_{\min} = R_{mid} - 0.5\Delta B, & \text{if } 0 < g_i < g_{fin} \\ B_{\max} = R_{\max}, B_{\min} = R_{\min}, & \text{if } g_i = g_{fin} \end{cases} \quad (4)$$

with

$$R_{mid} = \frac{1}{2}(R_{\max} + R_{\min}) \quad (5)$$

where R_{\max} and R_{\min} are the maximum and minimum boundaries, g_i denotes the current generation, g_{init} and g_{fin} are the initial and final generations, and $\Delta B = kg_i$ with k being a positive constant. The current boundaries B_{\max} and B_{\min} would be shifted when g_i increases such that the parameter searching range extends. Through the boundary sizing technique, the RSGA is able to provide a switch function in the early stage of evolution, emphasizing structural optimization. While the boundary increases gradually with generation number, it is a linear scaling function. When gradually reaching the end of the generation, (B_{\min}, B_{\max}) replaces (R_{\min}, R_{\max}) and the working algorithm gradually focuses onto the parameter optimization.

Reproduction

In the reproduction process, the algorithm follows the usual manner utilized in evolutionary computational algorithms to create a new population for the next generation from a population in the current generation. The selection operation imitates the

mechanism that describes the survival of the fittest in natural selection. The chromosomes are selected for mating, which depends on their relative fitness values, i.e., roulette wheel selection. The chromosome selection probability is given by

$$P_r = \frac{f_i}{\sum_{j=1}^m f_j} \quad (6)$$

where f_i is the fitness value of the i -th member, and m is the population size, as with usual GAs. The chromosomes of high probability associate with a relative high-fitness value among the population.

Crossover

As with usual GAs, the probability of the chromosome being selected to crossover is P_c , in general $0.5 \leq P_c \leq 0.9$. The crossover mechanism utilizes extrapolation or interpolation to generate new individuals. Initially, it operates in extrapolation. When the parameters of the offspring exceed the allowable ranges R_{\max} and R_{\min} , it then switches to interpolation. Therefore, it can avoid parameters varying over the range. The crossover operation is determined by

$$\begin{cases} \tilde{x}_{di} = x_{di} - \lambda(x_{di} - x_{mi}) \\ \tilde{x}_{mi} = x_{mi} + \lambda(x_{di} - x_{mi}) \end{cases}, \quad \text{if } x_{di} > R_{\max} \text{ or } x_{mi} < R_{\min} \quad (7)$$

$$\begin{cases} \tilde{x}_{di} = x_{di} + \lambda(x_{di} - x_{mi}) \\ \tilde{x}_{mi} = x_{mi} - \lambda(x_{di} - x_{mi}) \end{cases}, \quad \text{if } R_{\min} \leq x_{mi} \leq x_{di} \leq R_{\max} \quad (8)$$

$$\lambda = \lambda_0 \left(1 - \frac{g_i}{g_{fin}} \right) \quad (9)$$

where x_{di} and x_{mi} refer to the parent individuals, \tilde{x}_{di} and \tilde{x}_{mi} are the offspring of x_{di} and x_{mi} respectively, and $\lambda_0 \in [0, 1]$ is a uniform random number. Both the control and parameter genes are real numbers and share the same operation of mutation. It simplifies

the mechanism of two crossover operators applied in traditional SGAs.⁵

Mutation

The mutation operator applies randomly chosen individuals to gain fine tuning in chromosomes. The probability of the chromosome being selected to mutate is P_{mut} , in general $0.01 \leq P_{mut} \leq 0.1$. For the mutation operator, we adopt the non-uniform mutation method to change genes in a chromosome which can be realized as

$$\tilde{x}_{ij} = \begin{cases} x_{ij} + \Delta(g_i, x_{ij\max} - x_{ij}), & \text{if } h = 0 \\ x_{ij} - \Delta(g_i, x_{ij} - x_{ij\min}), & \text{if } h = 1 \end{cases} \quad (10)$$

and

$$\Delta(g_i, y) = y \lambda_0 \left(1 - \frac{g_i}{g_{fin}} \right) \quad (11)$$

where $x_{ij} \notin \{\phi\}$ is the j -th element of the i -th individuals in the current generation, the value $h \in \{0, 1\}$ depends on the random production, $x_{ij\max}$ ($x_{ij\min}$) is the maximal (minimal) j -th element of the i -th individual in the current generation, and y is a scaling factor.

Dynamic probability

Inspired by the characteristic advantages that the gain of Butterworth filter in Bode plot is flat in the passband and approaches zero near the stopband, a dynamic probability is proposed to ensure that the emphasis is placed on the object's structure first and then on its parameters:

$$P_{new} = P_{cur} \frac{1}{1 + \left(\frac{g_i}{g_c} \right)^{2q}} \quad (12)$$

$$\Delta P = \frac{1}{1 + \left(\frac{g_i}{g_c} \right)^{2q}} \quad (13)$$

where P_{new} is the new probability, P_{cur} is the current probability, ΔP is the dynamic probability factor,



g_c is the cut-off generation, and q is the order of the dynamic probability. Applying the new dynamic probability, the crossover and mutation probabilities become constant after reaching g_c . Related probability rates are defined as:

$$P_{crossnew} = P_{crosscur} \Delta P \quad (14)$$

$$P_{mutnew} = P_{mutcur} \Delta P \quad (15)$$

where $P_{crossnew}$ is new crossover probability, $P_{crosscur}$ is current crossover probability, P_{mutnew} is new mutation probability, and P_{mutcur} is current mutation probability. In general, g_c is selected to be the median of the total generation number. This ensures that the emphasis of the optimization will switch towards parameter optimization after reaching g_c . The design also ensures that the effect of crossover and mutation for parameter genes will not decay to zero.

We can sieve out good genes by evaluating the fitness values and generating elite chromosomes to the next generation. Figure 3 displays the schematic diagram for illustrating the operational process. All chromosomes in the operational process of RSGA are real numbers, thus it does not require any encoding step.

Biologically Genetic Oscillator Design

This section provides a general framework for the design of genetic oscillators based on the RSGA. Several methods to improve tolerance of the intrinsic and extrinsic noises are also proposed in the current framework.

Genetic oscillator

Elowitz and Leibler⁹ used three transcriptional repressor systems (*tetR*, *lacI*, *λcI*) to build an oscillating network. The first repressor, *lacI*, inhibits the transcription of the second repressor, *tetR*, which in turn inhibits the expression of the third gene, *λcI*. Finally, *tetR* inhibits *lacI*'s expression. Because the green fluorescence protein is coupled to a protein in the oscillating network, the oscillations are observed by the fluorescence in the cell. That transcriptional regulation oscillation behavior can be modeled in details based on several factors including mRNA, the dependence of transcription rates on repressor concentration, the decay rates of the protein, and the

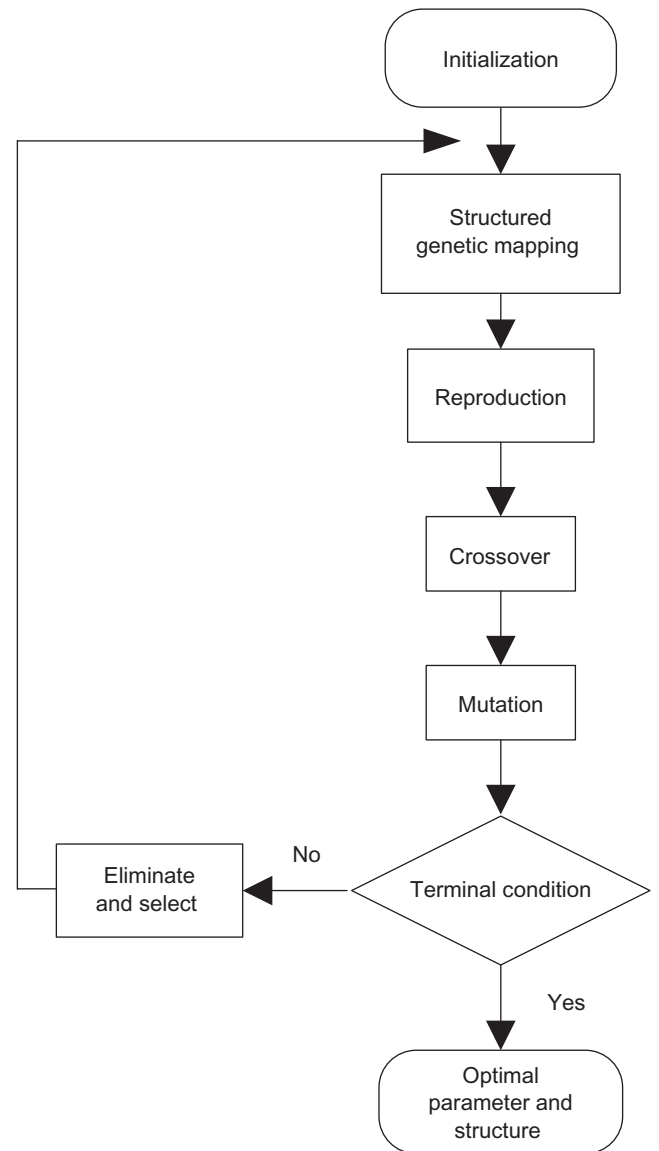


Figure 3. Schematic flowchart of real structured genetic algorithm.

translation rates of mRNA. The dynamic system can be described by the following coupled first-order differential equations:

$$\begin{aligned} \frac{dm_i}{dt} &= -\gamma_{m_i} m_i + \alpha_i b_i(p_j^{n_i}) + \alpha_0 + w_{i1}(t), \\ \frac{dp_i}{dt} &= \beta m_i - \gamma_{p_i} p_i + w_{i2}(t) \end{aligned} \quad (16)$$

$$b_i(p_j^{n_i}) = \frac{1}{1 + p_j^{n_i}}$$

where $(i,j) \equiv (lacI, \lambda cI)$, $(tetR, lacI)$, or $(\lambda cI, tetR)$, $m_i \in \mathbb{R}_+$ is the concentration of mRNA, $p_i, p_j \in \mathbb{R}_+$ are

concentrations of proteins in three genes, $\alpha_i \in \mathbb{R}$ is the transcription rate of mRNA, α_0 is the leakiness of the promoter, which is usually zero for stable state, $\beta \in \mathbb{R}_+$ denotes the ratio of the protein decay rate to the mRNA decay rate, $\gamma_{p_i} \in \mathbb{R}_+$ and $\gamma_{m_i} \in \mathbb{R}_+$ are decay rates of proteins and mRNA, respectively, $n_i \in \mathbb{R}$ is the Hill coefficient, and $w_{ij}; j = 1, 2$ denotes the effect of environmental perturbations. Fundamental oscillation behavior can exhibit in the concentration of the three repressor proteins.

The three-genes oscillation model was extended to an N coupled genetic model by Strelkowa and Barahona,¹¹ Hori, Hara and Kim,¹² Zeiser, Muller and Liebscher,¹³ and Hori and Hara.¹⁴ That is, one can use different number of genes to synthesize an oscillator. Figure 4 shows the configuration of a generalized N-stage gene oscillator. For compactness, we represent a class of stochastic models for the N-gene oscillator with intrinsic fluctuations and extrinsic disturbances as follows:

$$\dot{X} = f(X) + \sum_{i=1}^4 g_i(X)v_i + w \quad (17)$$

where $X = [m_1 p_1 m_2 p_2 \dots m_N p_N]^T$ is the state vector, $f(X)$ represents the nonlinear interactions of gene oscillation, $g_i(X)v_i$ are intrinsic parameter fluctuations due to random intrinsic noise sources, and $w = [w_{11} w_{12} \dots w_{N1} w_{N2}]^T$ is the extrinsic disturbance vector. The nonlinear terms $f(X)$ and $g_i(X)v_i$ are defined, respectively, by

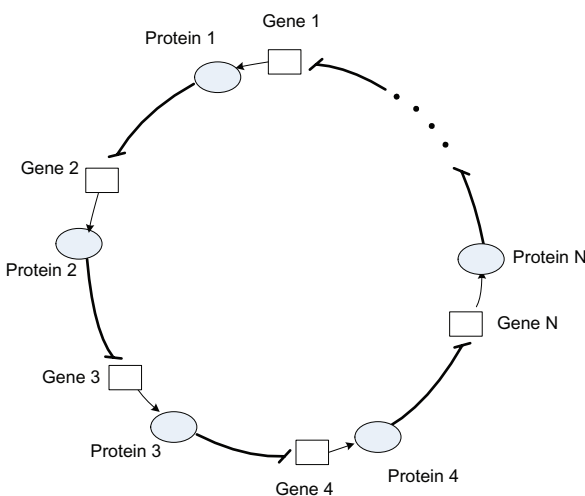


Figure 4. N-gene oscillator.

$$f(X) = \begin{bmatrix} -\gamma_{m_1} m_1 + b_1 (p_N^{n_1}) \\ \beta_1 m_1 - \gamma_{p_1} p_1 \\ -\gamma_{m_2} m_2 + b_2 (p_1^{n_2}) \\ \beta_2 m_2 - \gamma_{p_2} p_2 \\ -\gamma_{m_3} m_3 + b_3 (p_2^{n_3}) \\ \beta_3 m_3 - \gamma_{p_3} p_3 \\ \vdots \\ -\gamma_{m_N} m_N + b_N (p_{N-1}^{n_N}) \\ \beta_N m_N - \gamma_{p_{N+1}} p_N \end{bmatrix}, \quad g_1(X)v_1 = \begin{bmatrix} \delta\gamma_{m_1} \\ 0 \\ \delta\gamma_{m_2} \\ 0 \\ \delta\gamma_{m_3} \\ 0 \\ \vdots \\ \delta\gamma_{m_N} \\ 0 \end{bmatrix} \quad (18)$$

$$g_2(X)v_2 = \begin{bmatrix} \delta b_1 (p_N^{n_1}) \\ 0 \\ \delta b_2 (p_1^{n_2}) \\ 0 \\ \delta b_3 (p_2^{n_3}) \\ 0 \\ \vdots \\ \delta b_N (p_{N-1}^{n_N}) \\ 0 \end{bmatrix}, \quad g_3(X)v_3 = \begin{bmatrix} 0 \\ \delta\beta_1 \\ 0 \\ \delta\beta_2 \\ 0 \\ \delta\beta_3 \\ \vdots \\ 0 \\ \delta\beta_N \end{bmatrix}, \quad g_4(X)v_4 = \begin{bmatrix} 0 \\ \delta\gamma_{p_1} \\ 0 \\ \delta\gamma_{p_2} \\ 0 \\ \delta\gamma_{p_3} \\ \vdots \\ 0 \\ \delta\gamma_{p_N} \end{bmatrix} \quad (19)$$

and the Hill function

$$b_i (p_j^{n_i}) = \frac{1}{1 + p_j^{n_i}} \quad (20)$$



where

$$n_i \triangleq \begin{cases} \geq 0, & \text{for repression} \\ < 0, & \text{for activation} \end{cases}$$

Stable behavior is exhibited when oscillators are constructed with an odd number of repressor genes.¹⁵ In other words, protein concentration of that kind of oscillator attracts globally to a stable limit cycle. However, oscillators with even number of genes tend to be a quasi-stable periodic cycle that would diverge after a long period of time.¹⁶ When the number of genes is even, the number of repressive loops is even as well. Thus, the traditional way generally adjusts the set of nonlinear equations to ensure the number of repressive loops to be odd. This will guarantee that the bio-system insufficiently robust and attracted to the stable limit cycle.^{14,16} For example, a four-gene oscillator has four genes and four loops. One should change one nonlinear term of the four to be in activation form, thereby leaving the number of repressive loops at three. While applying the RSGA for oscillator design, one does not have to be concerned about the issue. Rather, the algorithm is able to determine the optimal combination of the repressors and activators.

Structured genetic mapping for genetic oscillator

A chromosome Y consists of two parts: the control gene set c and the parameter gene set p . For designing genetic oscillators, we rearranged the structured mapping operator. The first control gene controls the first and second parameter genes, which represent the decay rate of protein concentration and the transcription rate of mRNA in (16) respectively. The second control gene controls only the third parameter gene, which represents the hill coefficient. The third control gene controls the fourth and fifth parameter genes, which represent the decay rate of protein concentration and the transcription rate of mRNA. The fourth control gene controls only the sixth parameter gene, and so on.

Let the original chromosome be

$$Y = \langle c, p \rangle = \langle \overbrace{c_1(\alpha_1, \gamma_{p1}) \ c_2(n_1) \ c_3(\alpha_2, \gamma_{p2}) \ c_4(n_2) \ \cdots \ c_{2N-1}(\alpha_N, \gamma_{pN}) \ c_{2N}(n_N)}^c, p \rangle \tag{21}$$

where

$$p = (\alpha_1, \gamma_{p1}, n_1, \alpha_2, \gamma_{p2}, n_2, \dots, \alpha_N, \gamma_{pN}, n_N)$$

No control independent genes are considered in (21) for brevity of demonstration. After performing the structural mapping, the new chromosome becomes

$$\tilde{Y} = \langle c, \tilde{p} \rangle$$

where

$$\tilde{p} = [c_i] \otimes [p_i] \equiv \begin{cases} p_i, & \text{if } B_{\max} \leq c_i(\alpha_i, \gamma_{pi}) \\ p_i t, & \text{if } B_{\min} \leq c_i(\alpha_i, \gamma_{pi}) \leq B_{\max}, \\ \phi, & \text{if } c_i(\alpha_i, \gamma_{pi}) \leq B_{\min} \end{cases}$$

$$t = \frac{c_i(\alpha_i, \gamma_{pi}) - B_{\min}}{B_{\max} - B_{\min}} \tag{22}$$

Consider, for example, a randomly generated chromosome given and will be transformed as follows:

$$Y = \langle c, p \rangle = \langle \overbrace{0.52 \ 1.5 \ 1.35 \ 2 \ 4.6 \ 3.8 \ 0.02 \ 3.6 \ 2.33 \ 5.1}^c, \overbrace{1.58 \ 3.45 \ 4.97 \ 0.98 \ 0.11 \ 2.63 \ 3.01 \ 1.44 \ 0.7 \ 1.33 \ 3.52 \ 1.02 \ 4.4 \ 0.1 \ -2}^p \rangle$$

Given $R_{\max} = 5$, $R_{\min} = 0$, $g_{fin} = 10$, $g_i = 5$, $k = 0.1$ and the variations of B_{\max} and B_{\min} defined, respectively, by

$$B_{\max} = R_{mid} + \frac{1}{2} \Delta B = 2.75$$

$$B_{\min} = R_{mid} - \frac{1}{2} \Delta B = 2.25$$

Then, the structured genetic mapping yields



$$\tilde{p} = [c_i] \otimes [p_i] \equiv \begin{cases} p_i, & \text{if } 2.75 \leq c_i \\ p_i t, & \text{if } 2.25 \leq c_i \leq 2.75, \\ \phi, & \text{if } c_i \leq 2.25 \end{cases}$$

$$t = 2.25 \frac{c_i(\alpha_i, \gamma_{pi})}{2.75 - 2.25}$$

After mapping, the chromosome Y is denoted by \tilde{Y} :

$$\tilde{Y} = \langle c, c \otimes p \rangle$$

$$= \langle \overbrace{0.52 \ 1.5 \ 1.35 \ 2 \ 4.6 \ 3.8 \ 0.02 \ 3.6 \ 2.33 \ 5.1}^c, \overbrace{\underbrace{\phi \ \phi \ \phi \ \phi \ \phi}_{\text{off}} \ \underbrace{3.01 \ 1.44}_{\text{on}} \ \underbrace{0.7}_{\text{on}} \ \underbrace{\phi \ \phi}_{\text{off}} \ \underbrace{1.02}_{\text{on}} \ \underbrace{0.704 \ 0.016}_{\text{linear scaling}} \ \underbrace{-2}_{\text{on}}}^{\tilde{p}} \rangle$$

The corresponding linear ordinary differential equations with six key variables ($\alpha_1, \gamma_{p1}, n_1, \alpha_2, \gamma_{p2}, n_2$), after rearrangement of gene order, are obtained accordingly as follows

$$\begin{cases} \dot{m}_1 = -\gamma_{m1} m_1 + \frac{3.01}{1 + p_2^{0.7}}, \\ \dot{p}_1 = \beta m_1 - 1.44 p_1, \\ \dot{m}_2 = -\gamma_{m1} m_2 + \frac{0.704 p_1^2}{1 + p_1^2}, \\ \dot{p}_2 = \beta m_2 - 0.016 p_2 \end{cases}$$

Fitness function

The objective function consisting of two parts, related to parameter and solution structure, is defined by

$$J_{tot}(\alpha_i, \gamma_{pi}, N, n) = \min_{\substack{\alpha_i \in [\alpha_{ia}, \alpha_{ib}], \gamma_{pi} \in [\gamma_{pia}, \gamma_{pib}], \\ n \in [n_{ia}, n_{ib}], i=1,2,\dots,N}} [\rho J_p + (1-\rho) J_s] \quad (23)$$

where $\rho \in [0,1]$ is the weighting factor, J_p is the normalized performance index, and J_s is the normalized structure index.

The design objective is to search for the optimal decay rates of proteins (α_i), the transcription rates of mRNA (γ_{pi}), and the number of genes in order for the

objective function J_{tot} to be minimized. Furthermore, we would like to discover a solution candidate with the simplest structure for the requirement of a compact structure.

To specify the oscillating signal we consider the reference signal denoted by

$$r_i(t) = A_i \sin(\omega_i t + \phi_i), \quad i = 1, 2, \dots, N \quad (24)$$

where A_i , ω_i , and ϕ_i are amplitude, frequency, and phase, respectively. The goal is to design α_i and γ_{pi} that make concentrations of each protein emerging oscillating behavior. The normalized J_p and J_s are thus defined as

$$J_p = \frac{\sum_{i=1}^N \int_0^T (p_i - A_i \sin(\omega_i t + \phi_i))^2 dt}{T \sum_{i=1}^N A_i^2}, \quad J_s = \frac{N}{n_{max}} \quad (25)$$

where T is the integration time period, J_p can be explained as the objective value of tracking error, J_s represents the objective function for gene number, and the constants N and n_{max} are the order of genetic oscillator and the maximum order of genetic oscillator, respectively. The smaller J_p , the more accurate of tracking will be. A small J_s implies a simpler object structure. The corresponding fitness function can be expressed as follows

$$F(\alpha_i, \gamma_{pi}, N, n) = 1 / J_{tot}(\alpha_i, \gamma_{pi}, N, n) \quad (26)$$

Appropriately selecting α_i , γ_{pi} and N to maximize $F(\alpha_i, \gamma_{pi}, N, n)$, or equivalently, minimize $J(\alpha_i, \gamma_{pi}, N, n)$ to obtain the small tracking error, is the objective of the succeeding approach. See Figure 5 for illustration.

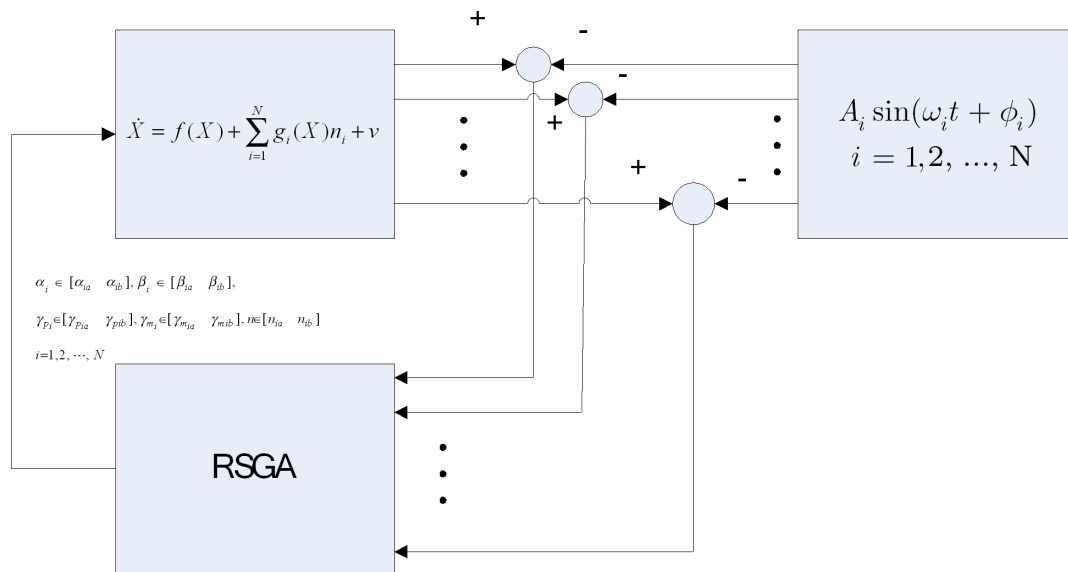


Figure 5. RSGA based synthetic genetic oscillator design.

It is notable that oscillators with different gene orders have limited capabilities of amplitude and frequency. However, the structured genetic mapping technique makes it possible to work in different gene-number oscillator module. Therefore, the optimal order can be discovered to fit the desired oscillated amplitude, frequency, and phase. In regard to the accurate performance, we improve the selection, crossover, and mutation probability to a dynamic probability. This idea will be illustrated in the next section.

In Silico Experiments

Consider the cyclic gene regulatory network composed of 1–5 genes to track the sinusoidal wave $A_i \sin(\omega_i t + \phi_i)$ where $A_i = 0.362$, $\omega_i = 0.075\pi, \forall i$, $\phi_i = i2\pi/N$. The parameters $\beta_i = 0.3465$ and $\gamma_{m_i} = 0.167$ were set as those by Chen and Chen.¹⁰

By considering only maximization of the fitness value F with three genes, the method of conventional RGA generated an oscillator with related responses (Fig. 6) with the corresponding fitness value of F converging to 4.2 (Fig. 7).

We applied a variable number of genes into experiments by the proposed RSGA. The results obtained are summarized in Table 1. While considering maximization of F , the RSGA yielded an oscillator with only two genes and the corresponding fitness value is 4.4. The related responses are illustrated in Figure 8. Convergence of the fitness value is displayed

in Figure 9. The result demonstrates superiority of the proposed approach over the conventional GA-based approach,¹⁰ in the sense that ours yields comparative tracking performance while requiring less number of genes to realizing the oscillator.

Robust design of genetic oscillator to tolerate intrinsic and extrinsic noises is more meaningful in vivo. To bring uncertain fluctuations into the module, we consider (17) and show responses of the repressor protein concentration in Figure 10. As is seen in the figure, the amplitude and phase are seriously affected by the intrinsic parameter fluctuations and extrinsic disturbances. This shows that existence of the intrinsic and extrinsic noises would deteriorate accuracy of the oscillation.

However, the RSGA approach described above and the traditional GA-based approach only determine the decay rate of protein concentration γ_{p_i} and the transcription rate of mRNA α_i . Their capabilities are restricted because the searching spaces are not robust enough to cover extra performance requirements. The resulting design would fail to track the oscillating commands with different amplitudes and phases under noise interference.

There are two possible ways to improve the deficiency. The first is to add extra parameters for performance tuning: the decay rates of mRNA γ_{m_i} and the ratio of the protein decay rate to the mRNA decay rate β in (16). This extends the parameter search from two to four parameters. It is also possible to

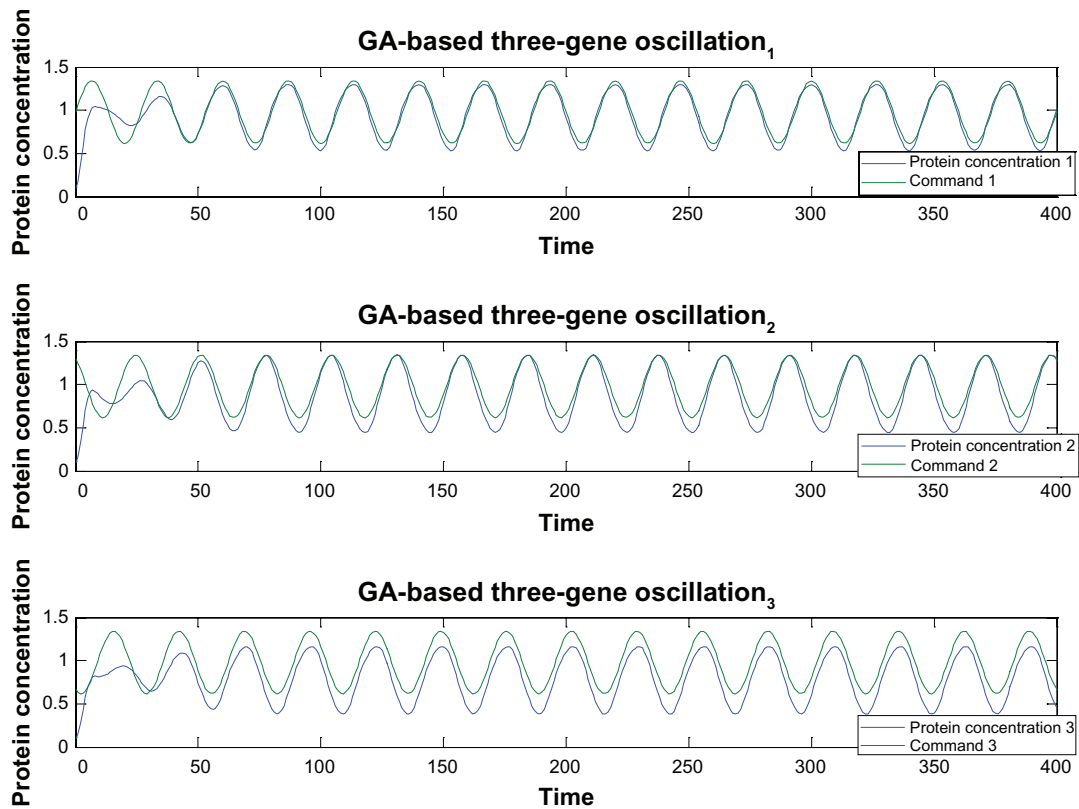


Figure 6. Three-gene oscillator obtained by RGA.

adjust the hill coefficient n to bring changes in transcription sensitivity. However, it is more difficult to realize in lab implementation. It was observed from the experimental results that the amplitude and frequency of oscillation increase as the hill coefficient increases. In contrast, a smaller hill coefficient results in smaller amplitude and frequency. As a result, if the

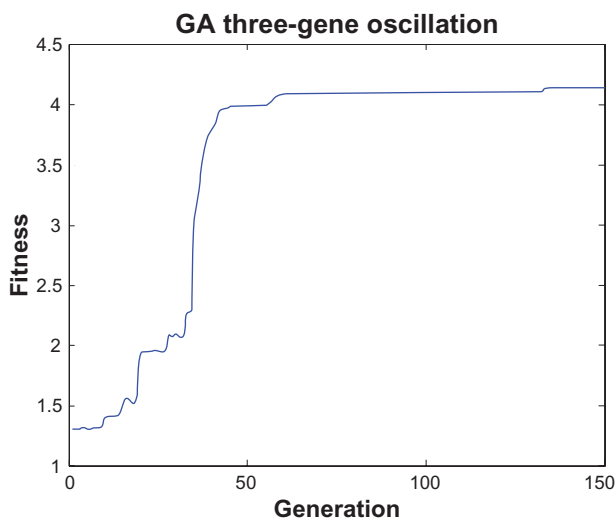


Figure 7. Fitness convergence of the oscillator obtained by RGA.

hill coefficient is relatively small, the protein concentration will converge to a steady level. See Figure 11 for the previous design.

Combining the two approaches, searching for four parameters γ_{m_i} , α_p , β_p , γ_{p_i} and n in each gene, as shown in Figure 12, improves the tracking performance and the corresponding fitness value f rises to 5.25. Figure 13 illustrates convergence of the fitness function. The resulting parameters are summarized in Table 2. However, the downside is that it loses the benefit of cheap network structure as the order of the oscillator is increased to three. The resulting oscillator is still of a simple structure, though.

From the numerical results presented above, it can be seen that the RSGA design approach is easy and efficient to apply while dealing with the situation of trade-off in simultaneous parameter and solution structure optimization problem.

Table 1. Parameters of two-gene oscillator design using RSGA.

α_1	γ_{p1}	α_2	γ_{p2}	N	n
0.34	1.285	0.079	0.277	2	5

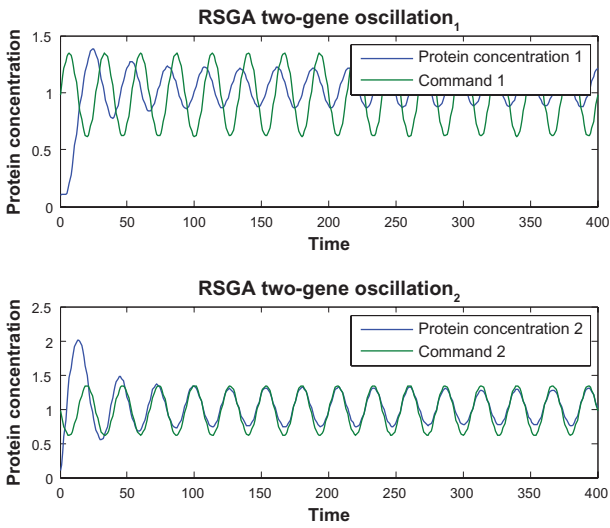


Figure 8. Two-gene oscillator obtained by RSGA.

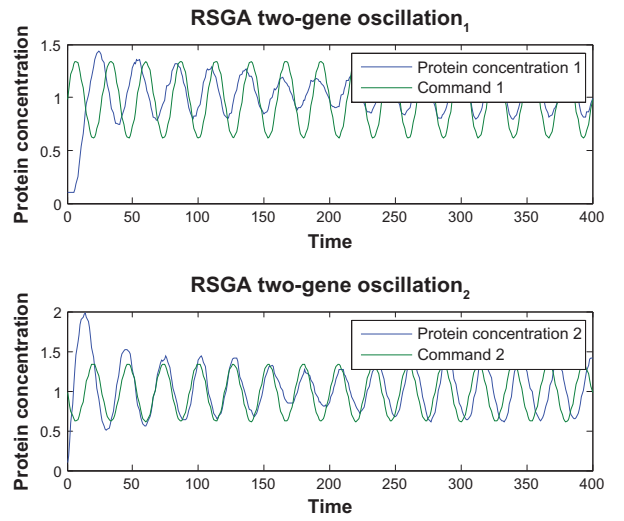


Figure 10. Oscillator with noise corruption.

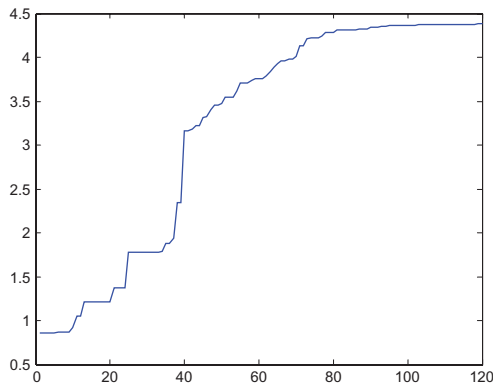


Figure 9. Fitness convergence of the oscillator obtained by RSGA.

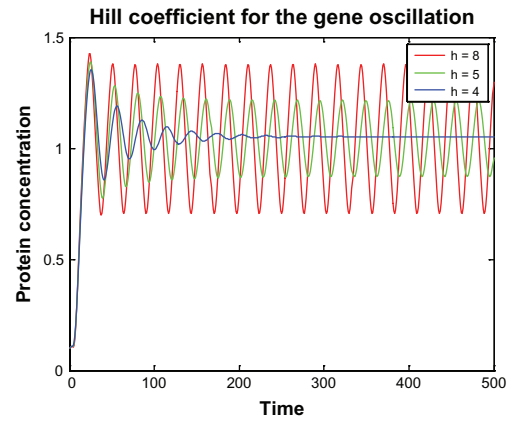


Figure 11. Oscillators with different hill coefficients.

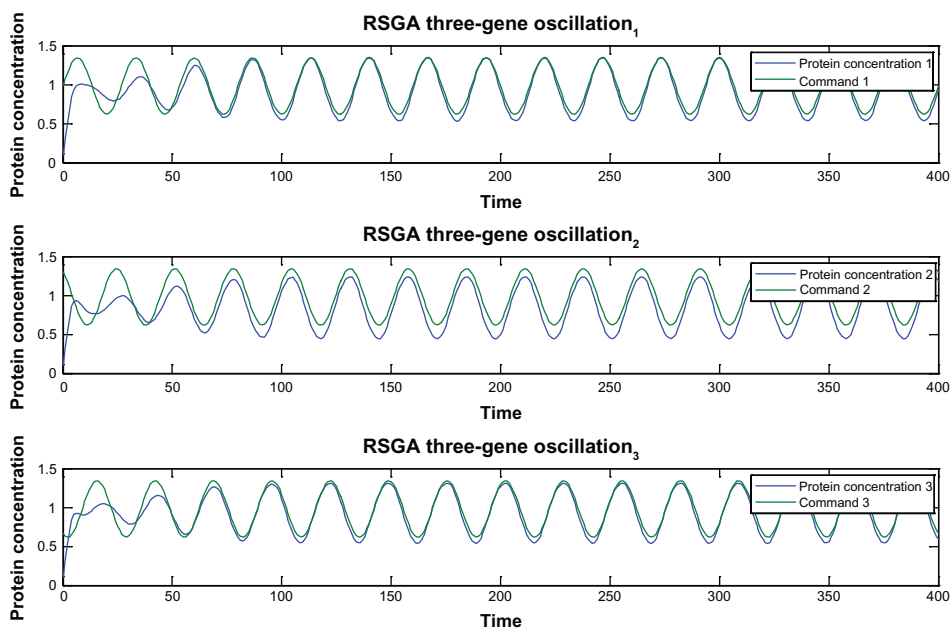


Figure 12. Oscillator obtained by RSGA while extending the searching parameters to $\alpha_r, \gamma_{pr}, \gamma_{mr}, \beta_r, n$.

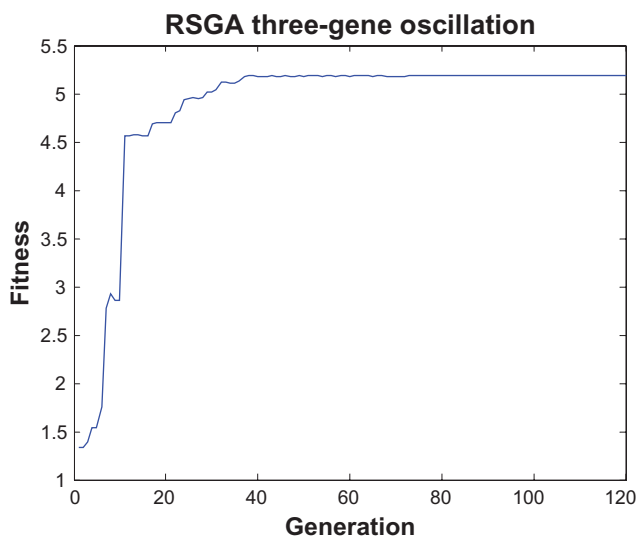


Figure 13. Fitness convergence of the oscillator obtained by RSGA while extending the searching parameters to $\alpha_r, \gamma_{pr}, \gamma_{mr}, \beta_r, n$.

Conclusions

RSGA approach has been applied to deal with the optimal design of synthetic genetic oscillator. The proposed approach is able to achieve the synthesized oscillator with a simplified structure and a lower number of parameters than that used by the existing evolutionary computational approach. For the noisy situation, it was observed that more tuning parameters, such as the decay rates of mRNA and the ratio of the protein decay rate to the mRNA decay rate, could be considered to improve the performance of the oscillator while clarifying the benefit of the compact network structure. Finally, a possible extension of our approach is to establish a conceptualized designing framework of steady-state combinational and sequence logic circuits, which may become a foundation while constructing Boolean computing devices.

Acknowledgement

This research was sponsored by National Science Council, Taiwan, ROC under the Grant NSC-98-2221-E-005-087-MY3.

Author Contributions

Conceived and designed the experiments: Y-CC, C-LL. Analyzed the data: Y-CC, C-LL. Wrote the first draft of the manuscript: Y-CC, C-LL. Contributed to the writing of the manuscript: Y-CC, C-LL, T.J. Agree with manuscript results and conclusions:

Table 2. Improved oscillator design using RSGA.

γ_{m1}	α_1	β_1	γ_{p1}	γ_{m2}	α_2	β_2
1.328	0.395	1.621	0.541	1.621	0.541	1.295
γ_{p2}	γ_{m3}	α_3	β_3	γ_{p3}	N	n
0.624	1.295	0.264	0.908	0.806	3	4.05

Y-CC, C-LL, T.J. Jointly developed the structure and arguments for the paper: Y-CC, C-LL, T.J. Made critical revisions and approved final version: Y-CC, C-LL, T.J. All authors reviewed and approved of the final manuscript.

Competing Interests

Author(s) disclose no potential conflicts of interest.

Disclosures and Ethics

As a requirement of publication author(s) have provided to the publisher signed confirmation of compliance with legal and ethical obligations including but not limited to the following: authorship and contributorship, conflicts of interest, privacy and confidentiality and (where applicable) protection of human and animal research subjects. The authors have read and confirmed their agreement with the ICMJE authorship and conflict of interest criteria. The authors have also confirmed that this article is unique and not under consideration or published in any other publication, and that they have permission from rights holders to reproduce any copyrighted material. Any disclosures are made in this section. The external blind peer reviewers report no conflicts of interest.

References

- Hasty J, McMillen D, Collins JJ. Engineered gene circuits. *Nature*. 2002; 420(6912):224–30.
- Dasika MS, Maranas CD. OptCircuit: an optimization based method for computational design of genetic circuits. *BMC Syst Biol*. 2008;2:24.
- Holland JH. *Adaptation in Natural and Artificial Systems*. Cambridge, MIT Press, MA, 1975.
- Goldberg DE. Real-coded genetic algorithm, virtual alphabets and blocking. *Complex System*. 1991;5:139–67.
- Dasgupta D, McGregor DA. *A Structured Genetic Algorithm: The Model and the First Result*. Res Rep IKBS-2-91, University of Strathclyde, 1991.
- Lai CC, Chang CY. A hierarchical genetic algorithm based approach for image segmentation. *Proceedings of IEEE International Conference on Networking, Sensing and Control*. Taipei, 2004:1284–8.
- Tsai CW, Huang CH, Lin CL. Structure-specified IIR filter and control design using real structured genetic algorithm. *Applied Soft Computing*. 2009;9(4):1285–95.
- Tsai CW, Huang CH, Lin CL. Microbrushless DC motor control design based on real coded structural genetic algorithm. *IEEE/ASME Transactions on Mechatronics*. 2011;16(1):151–9.



9. Elowitz MB, Leibler S. A synthetic oscillatory network of transcriptional regulators. *Nature*. 2000;403(6767):335–8.
10. Chen BS, Chen PW. GA-based design algorithms for the robust synthetic genetic oscillators with prescribed amplitude, period and phase. *Gene Regul Syst Bio*. 2010;4:35–52.
11. Strelkova N, Barahona M. Switchable genetic oscillator operating in quasi-stable mode. *J R Soc Interface*. 2010;7(48):1071–82.
12. Hori Y, Hara S, Kim TH. Existence criteria of periodic oscillations incyclic gene regulatory networks. Mathematical Engineering Technical Reports, METR 2010–11, The University of Tokyo, 2010.
13. Zeiser S, Muller J, Liebscher V. Modelling the Hes1 oscillator during somitogenesis. *J Comput Biol*. 2007;14(7):984–1000.
14. Hori Y, Hara S. Oscillation pattern analysis for gene regulatory networks with negative cyclic feedback. *Mathematical Engineering Technical Reports*. The University of Tokyo, 2010.
15. O'Brien EL, Itallie EV, Bennett MR. Modeling synthetic gene oscillators. *Mathematical Biosciences*. 2012;236(1):1–15.
16. Wang R, Li C, Chen LN, Aihara K. Modeling and analyzing biological oscillations in molecular networks. *Proceedings of the IEEE*. 2008;96(8):1361–85.