Emergence of robust regulatory motifs from in silico evolution of sustained oscillation

Yaochu Jin*a, Yan Mengb

a University of Surrey, Department of Computing, FEPS, Guildford, Surrey GU2 7XH, UK
b Stevens Institute of Technology, Hoboken, NJ 07030, USA

Abstract

The relationship between robustness and evolvability (easiness to evolve), and the evolutionary emergence of robust genetic circuits in biology have attracted much attention in systems biology. This paper investigates in silico the influence of the cis-regulation logic and the coupling of feedback loops on the evolvability and robustness of gene regulatory motifs that can generate sustained oscillation. Our simulation results indicate that both evolvability and robustness of the considered regulatory motifs depend on the cis-regulation logic and the way in which positive and negative feedback loops are coupled. Most interestingly, our findings suggest that robust regulatory motifs can emerge from evolution without an explicit selection pressure on robustness and adding noise in the parameters during the evolution is likely to promote the evolution of sustained oscillation.

© 2010 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Complex biochemical networks such as gene regulatory networks can often be analyzed by examining the structure and function of a number of wiring patterns, known as network motifs, such as auto-regulation, feedforward loops and feedback loops (Alon, 2007). Recently, the role of feedback loops, in particular, the coupling of feedback loops and its relation to the robustness of resulting dynamics of the network motifs have received increasing attention. For instance, Crumiere and Sablik (2008) showed that positive feedback loops are essential for generating multi-stability. Kwon and Cho (2008a) demonstrated that coherently coupled feedback loops enhance the robustness to state perturbations in a Boolean network. In another work, the same authors found that networks with a larger number of positive feedback loops and a smaller number of negative feedback loops are likely to be more robust (Kwon and Cho, 2008b). A recent study by Tsai et al. (2008) indicated that interlinked positive and negative feedback loops can achieve oscillatory dynamics with a widely tunable frequency with nearly constant amplitudes.

The cis-regulation logic also plays an important role in the dynamics and functionality of gene regulatory networks. An experimental analysis of cis-regulation logic of the sea urchin genes involved with development was conducted in Yuh et al. (2000). A systematic investigation of control logic in gene regulation was performed in Schilstra and Nehaniv (2008), which concluded that, among others, networks consisting of competitively binding activators and repressors can be controlled more robustly.

In addition to analysis of network motifs, in silico synthesis of typical regulatory dynamics can offer us insight into how nature has shaped the evolution of regulatory motifs. Francois and Hakim (2004) evolved both bi-stable and oscillatory dynamics using a number of predefined basic biochemical reactions, where the difference between the output of the regulatory model and the desired amplitude at the half-integer and integer periods was adopted as the fitness function. To facilitate the evolution of sustained oscillation, Paladugu et al. (2006) employed a fitness function based on the condition for generating Hopf bifurcation. Chu (2007), on the other hand, suggested that the degree of auto-correlation between the outputs of at given time period be used as the fitness function. These research efforts showed that evolution of sustained oscillation for gene regulatory networks is non-trivial due to the fact that the fitness function may be inefficient in guiding the evolutionary search. In Knabe et al. (2008), a gene regulatory network was used for evolving biological clocks in the presence of periodic environmental stimuli, where both the number and activation type (activating or repressive) of regulatory units of each gene were subject to evolution. They reported that the evolved clock tends to be robust to perturbations that evolution has experienced. Jin and Sendhoff (2008) investigated the influence of the genetic encoding
scheme as well as the activation function used in the gene regulatory model of a relaxation oscillation circuit. Their results suggested that evolving sustained oscillation using a step function as the activation function is much easier than using a Hill function. In case the Hill function is used, a larger Hill co-efficient is helpful for improving the evolvability of sustained oscillation. However, the evolutionary system is not able to find a large Hill co-efficient automatically. A step further was reported in Jin et al. (2009), where the parameters as well as the regulatory logic of two regulatory motifs with a predefined structure were evolved.

This paper investigates in silico the role of cis-regulation logic and the coupling of feedback loops in evolving oscillatory dynamics. The investigation has been divided into two stages. In the first stage, we analyzed the robustness of two motifs with a predefined structure to perturbations in parameters and initial states. In the second stage, part of the structure of the regulatory motifs and their regulatory logic were also subject to evolution, for which a hybrid evolutionary algorithm was used. Our results demonstrated that the robust motif emerged from in silico evolution without an explicit selection pressure on robustness, which suggests that robustness to perturbations in parameters and system states is an emergent property of evolution, confirming the findings in a related but different study reported recently by Krishnan et al. (2008).

2. Gene Regulatory Network Motifs

The gene regulatory motifs considered in this work consist of three genes, namely, two regulatory genes and one target gene. During the in silico evolution, the target gene is expected to generate sustained oscillatory dynamics. There are regulatory interactions, either activating or repressive, between the two regulatory genes and the target gene, though no direct interactions between the two regulatory genes were considered for the sake of simplicity, thus forming two feedback loops. With this simplification, we can have ten different network topologies in total, taking into account the structural symmetry between the two regulatory genes. Each motif has two feedback loops, which can be both positive or negative feedback loops, or one positive feedback and one negative feedback loop. In evolving the structure of the motifs, we made a further simplification by fixing the structure between one of the regulatory genes and the target gene. Consequently, we can have four different motifs, as shown in Fig. 1. In the motifs shown in Fig. 1(b) and (d), the target gene is consistently regulated by the two regulatory genes (both are activating), which are termed as consistently regulated motifs (CRMs). In contrast, the target gene is inconsistently regulated in the motifs in Fig. 1(a) and (c), which are then termed as inconsistently regulated motifs (IRMs). Moreover, in motifs in Fig. 1(a) and (d), two positive feedback loops are coupled. In contrast to that, a positive feedback loop is interconnected with a negative feedback loop in the motifs in Fig. 1(b) and (c).

In this work, the dynamics of the regulatory motifs is described using a set of differential equations. For instance, the regulatory dynamics of the CRM in Fig. 1(b) can be described using the following ordinary differential equations (ODEs):

\[
\dot{x}_1 = a_{12} H_{12}(x_2) - a_{11} x_1, \tag{1}
\]

\[
\dot{x}_2 = a_{23} H_{23}(x_3) - a_{22} x_2, \tag{2}
\]

\[
\dot{x}_3 = a_3 I (H_{31}(x_1), H_{32}(x_2)) - a_{33} x_3, \tag{3}
\]

where \(x_1\) and \(x_2\) are the concentration of the protein product of the regulatory genes, \(x_3\) is the concentration of the protein of the target gene, \(a_{11}, a_{22}\) and \(a_{33}\) are the degradation rate of the proteins, \(a_{12}, a_{23}\) and \(a_{31}\) are the parameters representing the strength of gene–protein interactions. In the CRM, there are three activating regulations and one repressive regulation distinguished by different Hill functions. An activating Hill function, e.g., \(H_{13}\) can be described as follows:

\[
H_{13}(x_3) = \frac{\beta x_3^n}{\beta x_3^n + x_3^3}, \tag{4}
\]

where \(n\) is called the Hill coefficient. By contrast, a repressive Hill function, e.g., \(H_{23}\), can be represented by

\[
H_{23}(x_3) = \frac{\beta}{1 + (x_3/\theta_2)^3}. \tag{5}
\]

In Eq. (3), both \(H_{12}\) and \(H_{32}\) are activating Hill functions similar to \(H_{13}\), and \(I (H_{31}(x_1), H_{32}(x_2))\) is a logic function combining the regulatory activity of the two regulatory genes. The dynamics of other regulatory motifs can be described in a similar fashion by simply changing the type of the Hill functions according to the coupling structure in the motif. For example for the IRM in Fig. 1(c), \(H_{23}\) should be activating and \(H_{32}\) should be repressive.

In this work, we assume that various transcription factors bind to the binding site of the target gene independently (Schilstra and Nehaniv, 2008). We considered two situations in this work. First, both transcription factors are needed to activate the expression of the target gene. Second, either of the two regulatory genes will be adequate to activate the expression of the target gene. In the first case, we used the fuzzy logic ‘AND’, which can be described using either the Zadeh ‘AND’ or the probabilistic ‘AND’ (Jin, 2003):

Zadeh ‘AND’ : \(L(x, y) = \min(x, y)\), \(\text{Prob. ‘AND’} : L(x, y) = xy\) \(\text{Prob. ‘AND’} : L(x, y) = \min(x, y)\), \(\text{Prob. ‘OR’} : L(x, y) = \max(x, y)\), \(\text{Prob. ‘OR’} : L(x, y) = x + y - xy\).

For the second case, the following Zadeh ‘OR’ or probabilistic ‘OR’ logic operators were adopted:

Zadeh ‘OR’ : \(L(x, y) = \max(x, y)\), \(\text{Prob. ‘OR’} : L(x, y) = x + y - xy\).

In addition, we also considered the case where a summation operator is used for the logic function, which can be seen as a combination of the probabilistic ‘AND’ and probabilistic ‘OR’:

\[
\text{summation} : L(x, y) = \frac{1}{2} (x + y). \tag{10}
\]
The motifs, though with only three nodes, are coupled nonlinear dynamic systems. It is very hard, if not impossible, to determine the parameters analytically to design oscillatory dynamics with a desired frequency. Meanwhile, we also aim to verify, in a computational environment, if oscillatory dynamics can be evolved for such small motifs. Therefore, in the next section, we are going to synthesize in silico oscillatory dynamics using simulated evolution.

3. In Silico Evolution of Regulatory Motifs

Since the late 1960s, a larger number of evolutionary algorithms (EAs) have been developed in the field of artificial intelligence (Fogel, 1995; Bäck, 1996). EAs were inspired by natural evolution, and have been applied to a wide range of research areas such as optimization (Osyiczka, 2001; Jin and Sendhoff, 2009), robotics (Nolfi and Floreano, 2000), and bioinformatics (Fogel and Corne, 2003; Fogel et al., 2008), to name a few. Recently, evolutionary algorithms have also been employed to synthesize gene regulatory dynamics (Francois and Hakim, 2004; Paladugu et al., 2006; Chu, 2007; Knabe et al., 2008; Jin and Sendhoff, 2008; Jin et al., 2009).

In this work, we used a canonical evolution strategy (Bäck, 1996, Chapter 2) for evolving the parameters in a model of regulatory motifs, and an evolutionary algorithm with a hybrid representation for simultaneously evolving the structure, regulation logic and the parameters of the motifs. The details of the hybrid evolutionary algorithms will be given in Sections 3.1 and 3.2.

3.1. Genetic Representations

In the first stage of the work, the parameters of two motifs with predefined regulation logic and connecting structure are evolved using an ES with one global step-size only as the strategy parameter for self-adaptation (Bäck, 1996). In the ES, the genome of each individual consists of two chromosomes. The first chromosome contains 12 real-valued loci, including three parameters representing the strength of protein–protein interactions, three degradation parameters, four regulation thresholds for the four Hill functions, one parameter for the maximum gene expression level, and one for the Hill coefficient. The second chromosome has one locus only, the strategy parameter.

For evolving the structure and regulatory logic of the motifs in the second stage, each genome has a binary chromosome in addition to the two real-valued ones used in first stage. The binary chromosome is composed of two loci representing the interactions between regulatory gene 2 and the target gene in the motifs, and two loci encoding the regulatory logic that combines the activity of the two regulatory genes. The representation schemes used in the two stages are illustrated in Fig. 2(a) and (b), respectively. In Fig. 2(b) and (c), \( c_{23} \) represents either an activating (\( c_{23} = 1 \)) or repressive (\( c_{23} = 0 \)) regulatory interaction from the target gene (\( g_3 \)) to regulatory gene 1 (\( g_1 \)). Similarly, \( c_{32} \) denotes an activating or repressive regulation from the regulatory gene (\( g_2 \)) to the target gene (\( g_3 \)). If both \( c_{23} \) and \( c_{32} \) are activating, or both are repressive, then \( g_2 \) and \( g_3 \) form a positive feedback loop. Otherwise, \( g_2 \) and \( g_3 \) form a negative feedback loop.

The two binary bits for encoding the regulation logic can be decoded into an integer number between 0 and 3, which represents the Zadeh ‘AND’ logic (0), the probabilistic ‘AND’ (1), the probabilistic ‘OR’ (2), and the summation logic (3), respectively. The Zadeh ‘OR’ logic is excluded from this representation since we never succeeded in evolving any sustained oscillation in the first stage with this logic operator.

3.2. Genetic Variations

For evolving the real-valued parameters in the gene regulatory network model in the first stage, Gaussian mutations as well as discrete recombination were employed for genetic variations. Both crossover and bit-flip mutations were applied to the binary chromosome encoding the regulatory logic and the interactions between regulatory gene 2 and the target gene in the second stage. Refer to Section 4.1 for the setup of the crossover and mutation probabilities used in this work.

3.3. Fitness Function and Selection Strategy

To evolve limit cycle (sustained oscillation), the sinus function was adopted to define the desired state for the target gene:

\[
x_3(t) = \sin \left( \frac{2\pi t}{\tau} \right) + 1.0,
\]

where \( x_3(t) \) is the desired state denoting the protein concentration of the target gene, \( \tau \) is the desired period of the limit cycle, and the amplitude was set arbitrarily to 1.0. The following fitness function was adopted to minimize the difference between the real state of the target gene (\( x_3(t) \)) and the desired state:

\[
f = \sum_{t=0}^{N} (x_3(t) - x_3^d(t))^2,
\]

Fig. 2. Genetic representations for in silico evolution of sustained oscillation. (a) Evolution of the parameters only; (b) evolution of the regulatory logic, interactions between the genes, and the parameters; (c) a generic regulatory motif whose interactions between regulatory gene 2 and the target gene are to be determined in evolution.
where \( N \) is the number of time steps simulated in solving the dynamics of the regulatory motifs.

For evolution strategies, elitism is often not recommended for solving continuous optimization problems (Beyer and Schwefel, 2002). In our simulations, however, we found that elitism helped a lot in improving the convergence of the evolutionary search and therefore was adopted in all simulations.

4. Simulation Results

4.1. Experimental Setups

The differential equations are solved using the modified Euler’s method with a step-size of 0.1. In the simulations, all regulatory parameters were randomly initialized between 0 and 4, and the initial step-size of the ES was set to 0.5. The parent and offspring population sizes were 60 and 400, respectively. For parameter optimization in the first stage, the Gaussian mutation and discrete recombination were applied with a probability 1, and each evolution was run for 500 generations. For both structural and parameter optimization in the second stage, the parameter setup for changing real-valued chromosomes was the same as in the first stage, and a probability of 0.9 and 0.02 was applied for the crossover and bit-flip-mutation, respectively. Each evolution was also run for 500 generations.

4.2. Evolvability of Predefined Motifs Through Parameter Optimization

In this stage, we chose two regulatory motifs with a predefined structure, one is the CRM in Fig. 1(b), and the other is IRM in Fig. 1(c), both of which having a coupled positive and negative feedback loops. The main reason for choosing these two motifs was based on the recent evidence suggesting that motifs having a negative feedback loop coupled with a positive feedback loop are able to generate more robust and flexible oscillation (Tsai et al., 2008). For the two motifs, we also considered the five different logic operations described in Eqs. (6)–(10), thus resulting in ten different motifs in total.

In this paper, we investigate evolvability by examining the easiness of evolving sustained oscillation. Two remarks should be made before we present the empirical results. First, it should be noted that this evolvability measure is fitness-dependent. More general, fitness-independent evolvability measures can also be defined, see, e.g., Jin and Trommer (2010). Second, it is non-trivial to numerically judge if the system dynamics is a sustained oscillation for continuous dynamic systems (Leier et al., 2006; Kuo et al., 2006). To evaluate the evolvability of the two different motifs, we attempted to evolve sustained oscillation for different desired periods, namely, \( T = 1, \ldots, 11, 12 \). For each desired period, we performed 10 independent runs. Then, we counted the number of runs in which sustained oscillation was evolved successfully. The evolution was said to be successful if the change in the height of the peaks is less than 1% within 400 time steps. The average success percentages for the CRM and IRM over the 12 desired periods are presented in Figs. 3 and 4, respectively.

From Figs. 3 and 4, we can see that the success percentage profiles for the CRM and IRM are quite different. The CRMs using the probabilistic ‘AND’ logic and summation logic exhibited the highest and second highest success percentages on average among all CRMs. By contrast, the IRMs with the summation and probabilistic ‘OR’ logic have the highest success percentage among all IRMs. As a whole, the success percentage for evolving sustained oscillation is quite low. Nevertheless, different to the relaxation motif studied in Jin and Sendhoff (2008), a high Hill coefficient is required neither for the CRM nor for the IRM.

From these results, we observed that at least one run was successful in 120 independent runs in evolving sustained oscillation for the CRM with probabilistic ‘AND’, probabilistic ‘OR’ and summation, and the IRM with probabilistic ‘OR’ and summation, respectively. Among them, the IRM with summation showed the highest success percentage.

4.3. Robustness Analysis of the Evolved Motifs

We examined now the robustness of the evolved oscillatory dynamics by varying the initial concentrations or by perturbing the parameters of the regulatory motifs. The robustness of regulatory dynamics to perturbations in initial concentrations of the regulatory genes is of interest since the perturbations in initial concentration reflect to a certain degree the influences of genes from outside the motif. During the evolution, the initial concentrations of the three proteins were set to \((1, 1, 0)\). To study the robustness of the evolved oscillation, the initial concentrations were set between 0 and 4 randomly. We found that when the initial states of the evolved CRM with the probabilistic ‘AND’ were randomly initialized, the state of the target gene converged to a point attractor instead of producing a limit cycle for all 16 evolved motifs showing sustained oscillation if their initial state is \((1, 1, 0)\). For the CRM with the summation logic, sustained oscillation was evolved in four runs. In the robustness test, all the four motifs were able to maintain their limit cycle when their states were initialized randomly between 0 and 4. From these results, we can conclude that the sustained oscillation of the CRM with the summation logic is less sensitive to perturbations in the initial concentrations than the CRM with the probabilistic ‘AND’ logic.

![Fig. 3. Average success percentage for the CRM with probabilistic ‘AND’ (pAND), Zadeh ‘AND’ (zAND), probabilistic ‘OR’ (pOR), Zadeh ‘OR’ (zOR), and summation (SUM), respectively, in evolving limit cycle with a desired period of \( T = 1, 2, \ldots, 12 \).](image)

![Fig. 4. Average success percentage for the IRM with probabilistic ‘AND’ (pAND), Zadeh ‘AND’ (zAND), probabilistic ‘OR’ (pOR), Zadeh ‘OR’ (zOR), and summation (SUM), respectively, in evolving limit cycle with a desired period of \( T = 1, 2, \ldots, 12 \).](image)
For the IRM, the results were quite different. In all the 29 successful runs of the IRM with the summation logic and in 15 successful runs for the IRM with the probabilistic ‘OR’ logic, all randomly initialized states converge to the evolved limit cycle. This indicates that the evolved oscillatory dynamics of IRM is robust to changes in the initial states.

The different motifs showed also very different behaviors regarding their sensitivity to perturbations in parameters. It was found that the CRMs, with either the probabilistic logic or the summation logic, are vulnerable to noise in parameters. Fig. 5 shows the 50 state-space trajectories of a CRM with the summation logic under 1% (left panel) and 5% (right panel) noise level. The 50 initial states were generated randomly in the interval of [0, 4]. The motif was still able to generate sustained oscillation for all the initial states when the noise level in the parameters was set to 1%. However, when the noise level rises to 5%, the oscillation became damped, and systems states finally converged to a point attractor.

For IRMs with either the probabilistic ‘OR’ logic or the summation logic, the regulatory system was able to generate sustained oscillation in the presence of 1% or 5% noise level in the parameters. Fifty trajectories of an IRM using the summation logic are presented in Fig. 6, where the left panel shows the results with 1% noise and the right panel the results with 5% noise in the parameters. It can be seen that in both cases, the regulatory motif was able to maintain a sustained oscillation.

To summarize, our analysis suggested that the IRM using either the summation logic or the probabilistic ‘OR’ is robust to changes in both initial concentrations and regulatory parameters.

4.4. Evolutionary Emergence of Robust Motifs

After evolving and analyzing the regulatory motifs with the given regulation logic and a predefined structure, we raised the question whether the robust regulatory motif can emerge when part of the motif structure and the regulatory logic are subject to evolution without an explicit selection pressure on robustness. Two sets of simulations were performed to answer this question. In the first set of simulations, the structure (interactions between
regulatory gene 2 and the target gene) and regulatory logic were evolved, together with the parameters of the regulatory motif for generating limit cycle of a desired period $T = 4, 5, \ldots, 18$. For each desired period, 20 independent runs were conducted. The resulting motif structures and logic operators evolved in the successful runs (in which limit cycle was evolved successfully) are presented in Fig. 7(a) and (b), respectively. From 300 runs, limit cycle was evolved successfully in a total of 65 runs, resulting in a success rate of 21.7%. From the results, we can see clearly that the IRM (structure ‘10’) using summation as the regulation logic, which was shown to be the most robust motif in the previous section, emerged in 63 among the 65 successful runs.

In the second set of simulations, all experimental setups were the same as in the first set, except that 5% noise was added to the parameters during evolution. The purpose was to investigate the influence of the noise on the evolvability of the system. The resulting motif structures and regulation logic operators are plotted in Fig. 8(a) and (b), respectively. Among a total of 300 runs, 79 runs were successful in evolving limit cycles, resulting in a success rate of 26.3%. This indicates that adding a certain level of noise to the parameters during the evolution enhanced the evolvability. The IRM with summation logic emerged in 78 of the 79 successful runs. Comparing the success rates in the two different setups, we surmise that adding noise in the parameters may be helpful in evolving limit cycles, although a statistical significance cannot be shown.

One question arose regarding the above results is that why the IRM using probabilistic ‘OR’ evolved only once in all the 600 evolutionary runs, although it had also been shown to be insensitive to changes in initial states and regulatory parameters. Our hypothesis is that IRM with the probabilistic ‘OR’ may be less robust compared to the IRM with the summation logic, which can in part be confirmed by the fact that in evolving the parameters with a fixed structure, the IRM using summation achieved sustained oscillation 29 times in 120 independent runs, while the IRM with probabilistic ‘OR’ was successful 15 times. Nevertheless, a more rigorous conclusion can be made only if we compare the robustness of the two motifs using a quantitative measure, which will be our future work.

5. Conclusions

This work studied the influence of coupling pattern of the feedback loops and regulation logic on the evolvability and robustness of small regulatory motifs having three genes. Our results suggested that both regulation logic and the way in which the feedback loops are coupled play a central role in evolving sustained oscillatory dynamics. We also showed that the IRM is more robust to perturbations in the parameters of the regulatory model than the CRM. Finally, we found that the IRM with summation logic, which is the most robust construction among the studied motifs, emerged from the evolution and adding noise in the parameters during evolution further increased the success rate of evolving sustained oscillation.

The reported work could be further extended in several aspects. First, the network motifs studied in this work are of a very small
size and only part of their structure was evolved. Second, the analysis with respect to the robustness and evolvability of the motifs was rather qualitative and less quantitative. In our future research, we will at first evolve all interactions in the motifs to verify if the robust motif can still emerge. In addition, larger motifs with multiple feedback loops and various feedback loop lengths should be investigated to see if robust networks such as those analyzed in Kwon and Cho (2008a,b) can also emerge from in silico evolution. Lastly, quantitative measures for evolvability and robustness will be designed for a more rigorous analysis of the robustness and evolvability.

References