Systems biology

Evolutionary design principles of modules that control cellular differentiation: consequences for hysteresis and multistationarity

Junil Kim1, Tae-Geon Kim2, Sung Hoon Jung2, Jeong-Rae Kim1, Taesung Park3, Pat Heslop-Harrison4 and Kwang-Hyun Cho1,∗

1Department of Bio and Brain Engineering and KI for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), 335 Gwahangno, Yuseong-gu, Daejeon, 305-701, 2Department of Information and Communication Engineering, Hansung University, Seoul, 136-792, 3Department of Statistics, Seoul National University, Gwanak-gu, Seoul, 151-747, Korea and 4Department of Biology, University of Leicester, Leicester LE1 7RH, UK

Received on January 12, 2008; revised on April 15, 2008; accepted on May 7, 2008

Advance Access publication May 8, 2008

Associate Editor: Trey Ideker

© The Author 2008. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

ABSTRACT

Motivation: Gene regulatory networks (GRNs) govern cellular differentiation processes and enable construction of multicellular organisms from single cells. Although such networks are complex, there must be evolutionary design principles that shape the network to its present form, gaining complexity from simple modules.

Results: To isolate particular design principles, we have computationally evolved random regulatory networks with a preference to result either in hysteresis (switching threshold depending on current state), or in multistationarity (having multiple steady states), two commonly observed dynamical features of GRNs related to differentiation processes. We have analyzed the resulting evolved networks and compared their structures and characteristics with real GRNs reported from experiments.

Conclusion: We found that the artificially evolved networks have particular topologies and it was notable that these topologies share important features and similarities with the real GRNs, particularly in contrasting properties of positive and negative feedback loops. We conclude that the structures of real GRNs are consistent with selection to favor one or other of the dynamical features of multistationarity or hysteresis.

Contact: ckh@kaist.ac.kr

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

In multicellular organisms, differentiation is a key process in the development of specialized cells from the progeny of unspecialized, pluripotent, stem cells (Ben-Tabou de-Leon and Davidson, 2007) and involves fine control and regulation of gene regulatory networks (GRNs) to express specific sets of genes at a specific body position with the correct timing (Chen et al., 2006; Levine and Davidson, 2005; Oliveri and Davidson, 2007; Shen, 1994). The GRNs for differentiation processes are modularized and the modules are interconnected (Alvarez-Buylla et al., 2007; Ben-Tabou de-Leon and Davidson, 2007; Chickarmane et al., 2006; Jaeger and Reinitz, 2006; Laslo et al., 2006). Let us call the GRNs of differentiation processes ‘differentiation control modules’ (DCMs). Evidently, DCMs are robust against external perturbation (Kitano, 2006; Nakao, 2005) and noise since malfunctioning is rarely seen while, at least in mammals, reprogramming of differentiation is very rare. Although DCMs have diverse consequences, essential structures of many DCMs are similar (Ben-Tabou de-Leon and Davidson, 2007), suggesting evolutionary conservation based on shared evolutionary principles.

Reports on the evolutionary design principles of biological networks suggest some key concepts including the robustness of oscillators (Kim et al., 2005; Wagner, 2005), adaptation to environmental changes (Kashtan and Alon, 2005), reliability of information processing (Klemm and Bornholdt, 2005), redundancy of pathways and multistage regulation that gives stable responses even in cells that are seriously compromised by genetic mutations or extreme environments.

The comparative analysis of contrasting complex biological structures has led to an understanding of their evolution. Here, we aimed to characterize any evolutionary design principles in DCMs, through investigation of the dynamical features of five real DCMs (Alvarez-Buylla et al., 2007; Ben-Tabou de-Leon and Davidson, 2007; Chickarmane et al., 2006; Jaeger and Reinitz, 2006; Jaeger et al., 2004; Laslo et al., 2006; Ma et al., 2006) (Fig. 1). From these investigations, we found that the two representative dynamical features of DCMs are hysteresis (Han et al., 2005) and multistationarity (Thomas and Kaufman, 2001a, b). Hysteresis means a bistable-switching phenomenon with two switching thresholds depending on the state trajectory and multistationarity means the characteristic of having multiple steady states. Why do real DCMs have such dynamical features? Hysteretic switching enables a system to make a reliable decision under noisy environments and a multistationary system can maintain multiple differentiated states robustly. For instance, the DCMs of mammals and sea urchins exhibit hysteresis and the DCMs of fruit flies (Drosophila) and plants show multistationarity. Based
Evolutionary design principles of differentiation

2 METHODS

2.1 Genetic algorithm

To simulate the evolutionary process of DCMs, we have employed a genetic algorithm (GA; Beasley et al., 1993). A schematic diagram of GA is described in Figure 2a. First, we have constructed 50 random strings representing the initial chromosomes with the following constraint on connectedness: If a network decoded from a chromosome has isolated parts, then we replace the network with a newly generated connected one. We check this constraint at every step of running the GA. Second, we have evaluated the fitness of each initial chromosome by the criteria given below. Next, we repeated the following procedures 500 times: (i) Make a pool of 25 pairs of chromosomes by selecting two chromosomes out of 50 parent chromosomes with the selection probability proportional to their fitness. (ii) Decide whether to apply crossover to each selected pair of chromosomes (the crossover rate was set to 0.68). If crossover is applied, a pair of child chromosomes is produced by extracting each parent chromosome of the pairs at random positions and swapping each segment of the extracted chromosomes. Otherwise, a pair of parent chromosomes becomes a pair of child chromosomes without any change. (iii) Randomly mutate the selected digits of each child chromosome (the mutation rate was set to 0.03 for each digit except as specified in Fig. S4 of Supplementary Material). (iv) Evaluate the fitness of each child chromosome (Fig. 2a) and replace the child chromosome of the worst fitness with parent chromosome of the best fitness. (v) Move to the next generation.

2.2 Chromosome structure and encoding rule

We have employed the artificial genome of Reil (1999) for the structure of a chromosome defined as a string of values. To test the evolutionary model, we considered networks composed of four nodes related to four genes $x_1$–$x_4$. As seen in Figure 2c, a chromosome consists of 80 base-4 digits. We divided 80-digit string into eight 10-digit strings representing each promoter sequence. Each gene has two kinds of promoters (activation promoter or inhibition promoter) and the ‘binding sequences’ taken to activate (first 10 digits) or inhibit (second 10 digits) the gene were fixed to 01, 21, 23, 03 (Fig. 2d). Here, one chromosome represents one
A stimulus–response curve depicted in Figure 3c is an example simulation. We have employed an ordinary differential equation (ODE) model to measure the network where the stimulus is given to gene 1 (xi) and the response is measured at each gene in the network. The activation and the inhibition matrices, we constructed an ODE system (Fig. 3a) determined by the following encoding rule: Let the number of binding sequences of gene i contained by activation promoter of gene j be n and that by inhibition promoter of gene j be m. If n > m then gene i activates gene j; if n = m then gene i inhibits gene j; if n < m then gene i and j have no relation. By using this rule and the chromosome in Figure 2c, we get the regulation matrix R = (Rij) depicted in Figure 2e where a regulation matrix represents a set of relations between nodes. The element Rij in the regulation matrix R denotes the effect of node j on node i. If Rij is positive (negative) then it implies that node j activates (inhibits) node i.

2.3 ODE simulations and the fitness for hysteresis

We have employed an ordinary differential equation (ODE) model to measure the fitness of hysteresis. For convenience, we decomposed the regulation matrix R into an activation matrix A = (Aij) and an inhibition matrix I = (Iij) where Aij = (Rij + Rj i)/2, Iij = (Rij − Rj i)/2 (Soyet et al., 2006). By using the activation and the inhibition matrices, we constructed an ODE system defined as

\[ \frac{dx_i}{dt} = \sum_{j} \left( V_{ij} \frac{x_j}{1 + S_{ij}^{op}} - x_i + R_i \right) \]

(1)

where \( h \) represents a Hill coefficient, \( x_j \) represents the expression level of each gene, \( S_{ij} \) represents the strength of stimulus ranging from 0 to 5, \( V_{ij} \) represents the maximum velocity parameter, and \( L = 1 \) for \( i = 1 \) and \( L = 0 \) for \( i > 1 \) (\( b = 4 \)).

A stimulus–response curve depicted in Figure 3c is an example simulation result of (1). Based on this curve, we compute the hysteresis measure, \( H \), defined by

\[ H = \sum_{i} \left( \left| S_{ij}^{op} - S_{ij}^{down} \right| \right) \frac{h}{b} \]

(2)

where \( S_{ij}^{op} \) represents the hysteretic range and \( h/b \) represents irreversibility. In order to formulate fitness that depends not on the parameter \( V_{ij} \) but on the topology of a network, we compute the hysteresis measure for a randomly generated integer \( V_{ij} \) between 1 and 5 and evaluate the fitness defined by the average of the hysteresis measure over 20 repetitions.

2.4 Boolean simulations and the fitness for multistationarity

To measure multistationarity, we have employed a Boolean model as it is difficult to evaluate multistationarity by using a continuous model. The Boolean network model consists of four Boolean variables \( s_1, s_2, s_3 \) and \( x_4 \), having either 0 or 1, representing each gene. The value of each Boolean variable is determined by its value at the previous time step and update function (Fig. 3d). Here, the state is defined as a vector composed of each Boolean variable (00, 01, 10, 11 states in total). Each state changes according to the update function and an attraction field (Fig. 3e) composed of 16 states is constructed. A point attractor is defined as a state whose next state determined by the update function is equal to its current state (e.g., 0000 and 0010 in Fig. 3e). We define the fitness of multistationarity as the number of such point attractors (Fig. 3f).

2.5 Network comparison

To measure the similarity between two networks, we produced regulation matrices of each network and then permuted those matrices by changing their nodes where the number of permutations is 24 (4!). We compared the 24 permuted matrices with the matrix of the other network by counting the number of elements having the same value at the same position. We define the similarity between two networks as the largest value among the 24 counted numbers.

3 RESULTS

3.1 Dynamical features of real DCMs

3.1.1 Hysteresis of real DCMs

Hysteresis means that a system switches its state at different stimulus thresholds depending on the direction of stimulus change (Sedra and Smith, 1997) as illustrated in Figure 3c. If \( \text{on/off} \) is less than zero, the system never switches off once it turned on in response to stimulus \( S \), which is called an irreversible bistable switch (Chickarmane et al., 2006). There are two dynamical characteristics of such a hysteretic switch. First, it is robust to noise and, second, it can memorize its former state (Xiong and Ferrell, 2007). Owing to these characteristics, the hysteretic switch has been widely utilized in various mechanical and electrical systems. Hysteresis is also very important for DCMs in organisms since it makes differentiation processes robust against noise and enables the cellular system to ‘remember’ its former state for sequential developmental processes.

We can find real examples of hysteretic switching systems. For instance, mammals have common hysteretic switching systems to regulate the embryonic state of stem cells, which are composed of three genes, OCT4, SOX2 and NANOG (Chickarmane et al., 2006) (Fig. 1a). Two transcription factors, OCT4-SOX2 complex and NANOG cooperatively activate the gene set associated with the maintenance of stem cells while they cooperatively inhibit the gene set associated with differentiation. A signal such as Wnt activates OCT4 and SOX2, which turns on all of the network nodes and makes the system maintain the state of a stem cell. Another signal such as p53 inhibits NANOG, which turns off all of the network nodes and leads to differentiation. Chickarmane et al. (2006) revealed that the network in Fig. 1a is a bistable switch robust against parameter variations. In particular, the network shows irreversible hysteresis over some parameter ranges and it is this hysteresis that makes a stem cell maintain its own state while prohibiting differentiated cells from returning back to their previous states. Another example is the cell fate-determining switch of hematopoietic cells. A hematopoietic cell can become macrophage or neutrophil as determined by the network in Figure 1b (Laslo et al., 2006). In this network, PU.1 and C/EBPα determine the activities of EgrNab and GfL-1, which repress with each other. EgrNab activates a set of genes that lead to become macrophage and inhibits a set of genes that lead to...
Another example of multistationary DCMs can be found from plants. The resulting evolved networks were then analyzed to examine whether such basic evolutionary design principles of the differentiation of regulatory networks can underlie real networks found in biology. The population size was set to 50 and 500 generations were produced for each simulation. The simulation was repeated 10 times with different initial random networks. We analyzed the resulting evolved networks as follows.

3.2.1 Feedback loops

To find out the structural properties of the evolved networks, we counted the average numbers of PFLs and NFLs at each generation. Figure 4a and c show the profiles of the average numbers of PFLs and NFLs in the case of evolution towards hysteresis and multistationarity, respectively. Figure 4b and d show the corresponding fitness profiles. The fitness profiles are increased with each evolving step. Although hysteresis and multistationarity have similar dynamical features, they are different fundamentally. To examine the different evolutionary driving forces of hysteresis and multistationarity, we have evaluated the multistationarity (hysteresis, respectively) fitness in the case of evolution towards hysteresis (multistationarity, respectively) (see Fig. S1 and S2 in Supplementary Material) and found that the fitness was not changed significantly. This implies that the two fitness functions are rather uncorrelated. From these figures, we found that hysteresis does not affect the number of PFLs but it decreases the number of NFLs (Fig. 4a). On the other hand, multistationarity increases the number of PFLs and decreases the number of NFLs (Fig. 4c). These results suggest that DCMs might lose NFLs to attain multistationarity, and loose NFLs while preserving PFLs to attain hysteresis. It was notable that such topological characteristics were also observed in the real DCMs shown in Figure 1. The DCMs of mammals and sea urchins (Fig. 1a–c) showing hysteretic switches actually contain many PFLs and no NFL. We also found that the DCMs of fruit flies and plants (Fig. 1d and f) give rise to multistationarity. Moreover, the DCMs shown in Figure 1d and f contain many PFLs but only a few NFLs. To examine whether these results are mainly dominated by the fitness, we have simulated the evolution of networks with a constant fitness and found that the numbers of PFLs and NFLs are not significantly changed (see Fig. S3 in Supplementary Material). We have also evolved the networks with the preference of hysteresis and multistationarity for 10 times repetition (Fig. 4b and d) and then the network is switched off. Due to the positive feedback loop (PFL) between Otx1 and Gain1 that Gain1 maintain their own expression levels without activation of Blimpl1, regulating the endoderm specification genes such as Bmp and Foxa. Hysteresis keeps the status of this network system unless a sufficient switching-off signal is applied.
3.2.2 Network comparisons with the two real DCMs

What is the relevance of artificially evolved networks to understanding the behavior of the real DCMs? To answer this question, we measured the similarity between the evolved networks (the best individuals at each generation of each simulation, so 2000 individuals in total (200 generations * 10 simulations)) and the two real DCMs (Fig. 1a and d), and obtained the similarity distributions shown in Figure 5a and b (these two networks were chosen as they consist of four genes). For statistical analysis, we also took measures of the similarity between the real DCMs and 2000 networks generated by evolving the random networks with a constant fitness. Then, we performed two-sample two-sided t-tests between the two similarity distributions. From the analysis results, we found that the network of mammalian embryonic stem cell switch (Fig. 1a) might have been evolved to prefer hysteresis (Fig. 5a) and the fruit fly gap gene network (Fig. 1d) might have been evolved to prefer multistationarity (Fig. 5b).

3.3 Network motifs

To get a further insight into the simulation results, we investigated the network motifs present in the DCMs. In particular, we considered the human GRN composed of 537 nodes and 1645 links from KEGG (Kanehisa et al., 2008). In this case, the number of two-node PFLs is seven and that of two-node NFLs is 17. From the gene ontology (GO; Consortium, 2006), it turns out that, among the 7 two-node PFLs, 5 PFLs are composed of the nodes related to differentiation while, out of the 17 two-node NFLs, none is related to differentiation (see Table S1 and Fig. S7 in Supplementary Material). This implies that the two-node PFL might be an important motif of the DCM. We also obtained the same results for GRNs of other two species (mouse and rat) (see Table S2, S3 and Fig. S7 in Supplementary Material) and for another human GRN extracted from TRANSFAC (Matys et al., 2006) (see Table S4 in Supplementary Material). There were some experimental studies reporting that the two-node PFL is an important motif of the gene network regulating the differentiation process (Ben-Tabou de-Leon and Davidson, 2006; Chickarmane et al., 2006; Ingolia, 2004; Ma et al., 2006; Oliveri and Davidson, 2007). These reports support our analysis on the feedback loops.

To figure out the dynamical characteristics of a two-node PFL, we have investigated the attraction fields of two kinds of two-node PFLs: double-PFL (Fig. 6a) and double-NFL (Fig. 6c). For comparison, we have also carried out the simulations of two-node NFLs (Fig. 6e). We found that both two-node PFLs have two point attractors (Fig. 6b and d) while the two-node NFL has no point attractor (Fig. 6f). From these, we infer that the two-node PFL helps to induce bistability but it is not the case for the two-node NFL. We have obtained the same results as we further extended the model to three-node cases (Fig. 6h and j). In this case, we found that the network becomes multistationary (Fig. 6j where three point attractors exist) if there are coupled double-NFLs. As the coupling of double-NFLs increases the number of steady states, the DCMs of fruit flies and plants (Fig. 1d and e) which exhibit multistationarity might have many double-NFLs. Although NFLs weaken the multistationarity, a few NFLs are still found in the DCMs of Fig. 1d and e. This might be due to the coupling of double-NFLs. For instance, if three double-NFLs are coupled as in Figure 6i, two NFLs are found at the three nodes. Moreover, the effect of such a few NFLs is suppressed by that of PFLs since there are already a few NFLs.
lot of NFLs in the DCMs shown in Figure 1d and e. In summary, the DCMs shown in Figure 1d and e can exhibit multistationarity due to the coupled double-NFLs.

Through ODE simulations, we took the measure of the hysteretic effect of the four network motifs (Fig. 7a) for different values of $V_f$. These simulations showed that both a single PFL and coupled PFLs exhibit hysteresis over a broad parameter range. In contrast, the networks having at least one NFL show no hysteresis even for a broad parameter range (Fig. 7b). Although a PFL is very important to hysteresis, hysteresis does not affect the number of PFLs in the simulation result of the previous section (Fig. 4a). This result indicates that a certain number of PFLs is enough to ensure hysteresis. On the other hand, since an NFL can suppress hysteresis, the number of NFLs is decreased during the evolutionary process (Fig. 4a). The real DCMs (Fig. 1a–c) support our results. Hence, we conclude that the DCMs having hysteresis might have been evolved by formulating a certain number of PFLs while not favoring of NFLs.

4 CONCLUSION

Hysteresis and multistationarity are key dynamic features of differentiation control modules, and a consequence of the relevant GRNs that are amenable to simulation and analysis. To unravel the evolutionary design principle of gene networks governing the differentiation processes, we have carried out artificial evolution of random networks such that they favor the two dynamical features. Remarkably, the outcome of the selection lead to contrasting patterns of the number of feedback loops during selection for the two features. We have further analyzed the topological characteristics of resulting networks through motif analysis and comparison with real networks. Our results suggested that the preference of enforcing hysteresis and the preference of enforcing multistationarity might be mutually exclusive design principles in gene network architecture, and examples of each have been well characterized in several model systems. Moreover, we found that the numbers of PFLs and NFLs are key parameters in determining the dynamical features of DCMs. Hysteresis is well known to increase the robustness of systems to noise, whether arising from external sources or within a controlled system, and is critical to avoid damaging levels of oscillation while maintaining appropriate control of cell parameters. In other cellular control systems, the property of multistationarity is important to allow differentiation of cells to proceed with an appropriate cascade of regulation following the initial signal and early responses. Evolutionary simulations presented with different criteria of fitness have proved that selection towards hysteresis or towards multistationarity is possible in relatively simple, four-gene systems, and has contrasting outcomes in the nature of the feedback loops that are revealed. To enforce hysteresis, DCMs might have been evolved to decrease the number of NFLs while maintaining PFLs. To enforce multistationarity, DCMs might have been evolved to decrease the number of NFLs and to increase the number of PFLs (i.e. coupled double-NFLs). Both hysteresis and multistationarity are amenable to experimental measurement and manipulation, and it is possible that this might be a powerful approach to the discovery of networks and the genes involved, leading to the modeling of cellular regulation.

ACKNOWLEDGEMENTS

Funding: This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea Ministry of Education, Science & Technology through the Systems Biology grant (M10503010001-07N030100112), the Nuclear Research Grant (M20708000001-07B0800-00110), and the 21C Frontier Microbial Genomics and Application Center Program. Conflict of Interest: none declared.

REFERENCES