Vol. 20, e200019 (2023)
doi: 10.2142/biophysico.bppb-v20.0019

Biophysics and Physicobiology
https://www.jstage.jst.go.jp/browse/biophysico/

Review Article (Invited)

Controlling complex dynamical systems based on the structure of the networks

Atsushi Mochizuki

Institute for Life and Medical Sciences, Kyoto University, Kyoto 606-8506, Japan

Received October 31, 2022; Accepted April 20, 2023;
Released online in J-STAGE as advance publication April 21, 2023
Edited by Tamiki Komatsuzaki

Progress of molecular biology resulted in the accumulation of information on biomolecular interactions, which are complex enough to be termed as networks. Dynamical behavior generated by complex network systems is considered to be the origin of the biological functions. One of the largest missions in modern life science is to obtain logical understanding for the dynamics of complex systems based on experimentally identified networks. However, a network does not provide sufficient information to specify dynamics explicitly, i.e. it lacks information of mathematical formulae of functions or parameter values. One has to develop mathematical models under assumptions of functions and parameter values to know the detail of dynamics of network systems. In this review, on the other hand, we introduce our own mathematical theory to understand the behavior of biological systems from the information of regulatory networks alone. Using the theory, important aspects of dynamical properties can be extracted from networks. Namely, key factors for observing/controlling the whole dynamical system are determined from network structure alone. We also show an application of the theory to a real biological system, a gene regulatory network for cell-fate specification in ascidian. We demonstrate that the system was completely controllable by experimental manipulations of the key factors identified by the theory from the information of network alone. This review article is an extended version of the Japanese article, Controlling Cell-Fate Specification System Based on a Mathematical Theory of Network Dynamics, published in SEIBUTSU BUTSURI Vol. 60, p. 349-351 (2020).

Key words: mathematical theory, regulatory network, model-free, linkage logic, ascidian

Significance

While networks of biomolecules governing biological functions have been elucidated, their complexity makes it difficult to understand the behaviors. The mathematical theory introduced in this review, by which key factors are determined from network structure alone, will solve the difficulty of understanding dynamics of complex systems in biology. In a study applied to an actual gene network, the whole dynamical behaviors of the system could be reproduced by manipulating the activity of a small number of factors determined from the theory. By combining prediction by the theory with experimental verification, rational understanding of complex biological systems will be obtained.

Introduction

Throughout the history of biology, it has been shown that numerous species of biomolecules are involved in various biological phenomena and that they have complex regulatory relationships with each other. For example, Figure 1 shows

Corresponding author: Atsushi Mochizuki, Institute for Life and Medical Sciences, Kyoto University, 53 Shogoin Kawahara-cho, Kyoto, 606-8507, Japan. ORCID iD: https://orcid.org/0000-0002-7714-1755, e-mail: mochi@infront.kyoto-u.ac.jp

a gene regulatory network including 85 factors for cell-fate specification in early development of ascidian identified by the research group of Dr. Sato [1]. Seven types of tissues, the epidermis, brain, nerve cords, muscle, notochords, mesenchyme and endoderm are produced during the early development of the ascidian from the 8-cell stage to the tail-bud stage. The regulatory network responsible for the process was created through a series of experiments where each gene was knocked down by morpholino antisense oligonucleotides, and the effects of perturbation were identified by observing changes in gene expression during development.

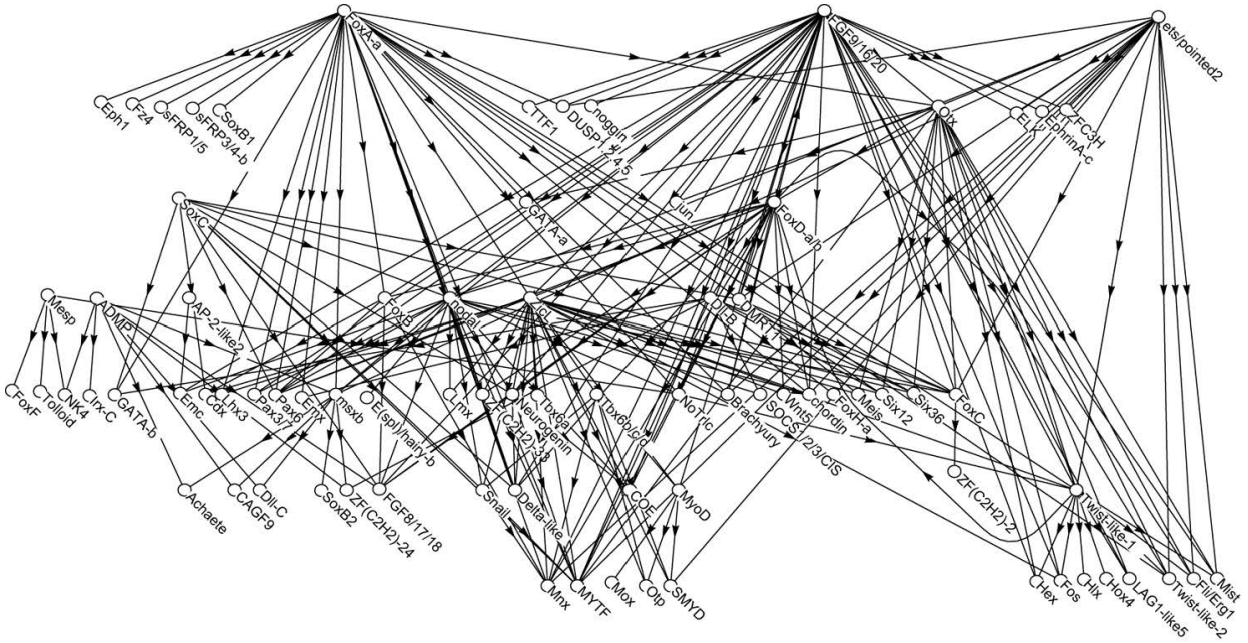

Figure 1 A gene regulatory network for cell-fate specification in early development of ascidian (From Imai et al., 2006; modified by Mochizuki et al., 2013).

In general, during development, the number of cells in the embryo increases through cell division, while the diversity of gene expression patterns within the cell population increases. The term "gene expression pattern" is used to indicate which genes are active/inactive in a cell. As the results of series of inductions of different genes between cells, diversification of gene expression patterns in an embryo is achieved in the developmental process. In the case of ascidians, more than 7 different gene expression patterns are observed at the late gastrula stage, depending on the location of the body. Cells showing different gene expression patterns become different tissues at later developmental stages. In other words, the diversity of cell types observed in the adult ascidian is originated in the diversity of gene expression patterns at a certain developmental stage.

It should be noted that a variety of gene expression patterns are produced from a single gene regulatory system. Essentially all cells have the same set of genes, and the same gene regulatory system. How can we think of a single system producing multiple states? The asymptotic behavior of a dynamical system after a sufficient time is helpful to understand the above question. In general, multiple attractors can appear in dynamical systems depending on the initial states. The emergence of multiple gene expression patterns from a single network system during development is thought to arise from multiple attractors in the corresponding dynamical system. The differences in the expression patterns between cells may arise from different initial condition caused by the cell-cell interactions or the biased cytoplasmic factors during cell division. In the case of ascidian development, gene expression dynamics based on the network shown in Figure 1 is considered to produce seven different expression patterns, which in turn produces seven different tissues. In other words, we can expect that the dynamics based on this gene regulatory network has the property of producing at least seven attractors.

Let us reconsider what is the information on the regulatory network from dynamical point of views. A regulatory network represents the information about “which gene influence which gene”, that is, dependencies between variables in the system. When we consider an ordinary differential equation (ODE) system corresponding to the focal system, the network provides information of the argument set of right-hand-side functions of the corresponding ODE system (Figure 2). On the other hand, since the network does not contain information of mathematical formulation of the function or

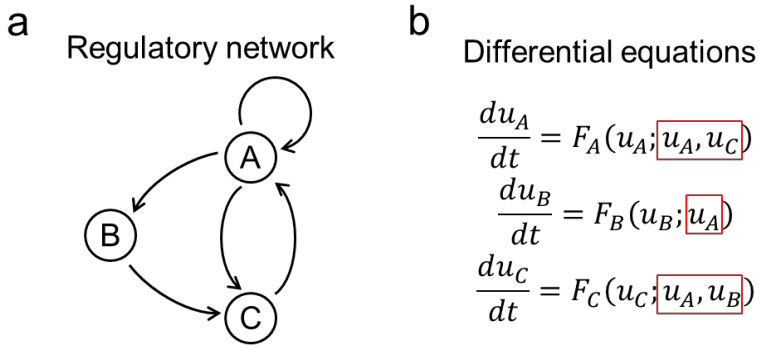

Figure 2 The relationship between a regulatory network (a), and differential equations (b) (Modified from Kobayashi et al., 2018). See Appendix for the meaning of the red boxes.

parameter values in the functions, it is impossible to specify the details of the dynamics explicitly from the information of network, alone. For studying dynamics of network systems, many mathematical studies have taken a modeling approach where mathematical models are constructed by assuming functional forms and parameter values, and tried to reproduce and explain observed biological phenomena.

However, the system size should be small for the modeling approach working effective. To construct a mathematical model for a complex system such as the one shown in Figure 1, one must assume a large number of mathematical formulations of functions and determine an enormous number of parameter values. Finding appropriate values of parameters alone requires a great deal of effort, and the results obtained will include countless assumptions. It is obvious to anyone that such a mathematical model is not useful. The difficulty in understanding the behavior of biological networks is not so much their complexity, but rather the lack of information to determine their dynamics.

In the following, we take a different approach from the standard modeling approaches, namely model-free approach. We study an aspect (but an important aspect) of dynamical behaviors determined from network information alone. When we carefully consider what is important to know in life sciences, we may not necessarily need to determine all the details on dynamics. If we obtain a logical conclusion from network information alone without introducing other assumptions, it will be important and useful not only for the mathematical sciences, but also for the life sciences.

Structural Theory of Regulatory Networks - Linkage Logic

Consider dynamics on a regulation network $\Gamma = (V, E)$, where V is a set of vertices and E is a set of directed edges [2-5]. A time-evolving quantity u_n $(n \in V)$ is assigned to each vertex, and the quantity varies according to the dependencies indicated by the regulatory network. Since the activity of biomolecules is the concentration of substances such as mRNA or proteins, these dynamics should include degradation or decay of activity. For example, let us consider a system of ordinary differential equations

$$\frac{du_n}{dt} = f_n(\boldsymbol{u}_{I_n}) - d_n(u_n), \ n = 1, \cdots, |V| \tag{1}$$

where u_n is a positive value indicating activity or concentration of a biomolecule. f_n is a regulatory function for the activity of molecule n. Here, I_n denotes the argument set of f_n, indicating the factor that influence the activity of molecule n, i.e. $I_n := \{i | (i \to n) \in E\}$. $\boldsymbol{u}_{I_n}$ is a vector whose elements are u_i $(i \in I_n)$. d_n is a positive increasing function that represent decay or degradation of biomolecule n. In many biological systems, the specific forms of f_n and d_n are unknown. I_n contain n itself $(n \in I_n)$, if and only if there is self-"activation" regulation. That is, the self-inhibition effect of u_n can be included in the decay term d_n, and need not be shown separately. Among self-regulations, only the self-activation, that cannot be represented by d_n, is included in f_n by setting $n \in I_n$. In the following, we consider regulatory networks redefined by I_n constructed in this way. In other words, any self-regulatory arrows in the network below implies self-activation.

The form of equation (1) with the separate decay term was introduced for easier understanding. This form is not actually necessary, and the theory described here can be applied to more general formula of ODEs. Details are shown in Appendix.

Here we introduce an important concept in graph theory called a feedback vertex set (FVS) [6,7]. An FVS is “a subset of the vertices in a directed graph, removal of which leaves a graph without any directed cycles”. A directed cycle is a path in a graph that starts from a vertex, follows edges in the direction of the arrow, and returns to the original vertex.

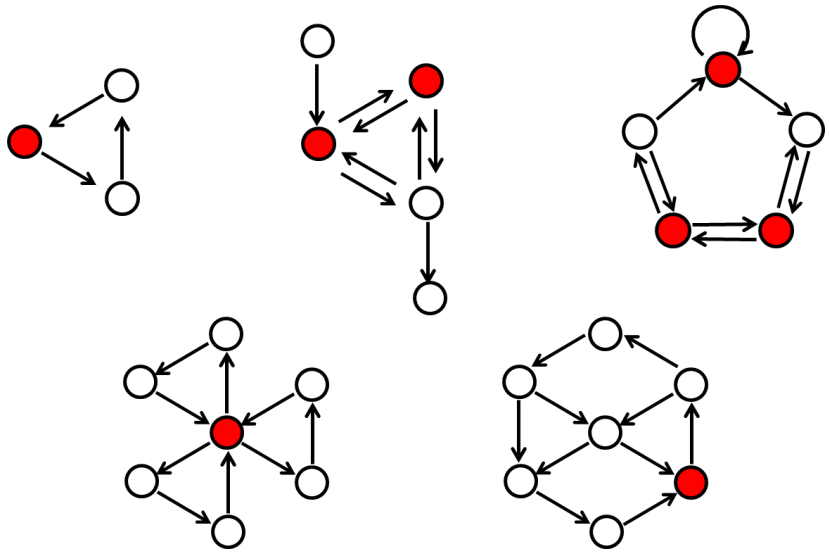

Figure 3 Examples of hypothetical regulatory networks and minimum FVS. FVS is not always unique. An example is colored here.

Figure 3 shows examples of simple networks with their FVS. The top-left graph is a simple directed cycle containing three genes, and any one of these three constitutes the minimum FVS, since removing any one would eliminate the cycle. In the top-center graph, two vertices at the top and bottom can be excluded since the cycles only pass through the three central vertices. Removing only one of the central three vertices still leaves a cycle. Removing any two of the three vertices eliminates all cycles, so any two of the three vertices are the minimum FVS. In the upper-right graph, a vertex $n \in I_n$ with self-regulation is always contained in the FVS. In the bottom-left graphs, three loops pass through a central vertex. The bottom-right graph looks more complicated, but since all loops pass through the bottom right (or top right) vertex, the minimum FVS contains only one vertex in the bottom right (or top right). Choice of FVS and minimum FVS may not be unique.

FVS is a concept in graph theory. Mochizuki and Fiedler et al. found that this definition also has important implications for dynamics on a graph [4,5]. The theorem claims that for a given dynamics of FVS, the dynamics of all variables in the system is determined uniquely in long term regardless of the regulatory functions (See Appendix for the detail). Note that the "unique" determination does not imply determining the dynamics of all variables "constructively": exact values of non-FVS variables cannot be determined without knowing the functions f_n (or F_n in (A1)). However, the unique determination of dynamics of non-FVS for a given dynamics of FVS is sufficient to induce the following two practical significances. First, if a system has multiple asymptotic behaviors, the diversity of behaviors of the whole system should be detectable by observing dynamics of FVS only. That is (1) dynamics of the whole system is observable by observing FVS. Second, by controlling behaviors of FVS to converge to one of any asymptotic trajectories, the whole system spontaneously come to converge to the target trajectory. That is (2) dynamics of the whole system is controllable by controlling FVS. These claims hold no matter what the nonlinear function of each vertex is. Conversely, the set of vertices that realize (1) observation and (2) control of the whole system without depending on model (functions and parameters) is only FVS. In other words, given information of network, the set of vertices to observe or to manipulate can be determined without knowing the functions.

The proof is given in the Appendix. Here, we give an intuitive explanation. First, consider a network with only one layer consisting of regulators (upper-side) and regulated (lower-side) nodes (Figure 4a). Let us assume that the behavior of the upper side are observable, i.e. the diversity of the behaviors of the upper-side can be captured.

We focus on one of the observed upper-side behaviors. Given a behavior of the upper side uniquely, the behavior of the lower side should asymptotically approach a unique behavior after sufficient time (unless the lower side has self-loop, i.e. self-activation). We do not care the details of the asymptotic behavior of the lower side; the fact that the lower side is uniquely determined is rather important. For the purpose of capturing the diversity of behaviors, it is not needed to observe the lower behaviors.

Next, let us consider a network containing a large number of vertices (Figure 4b). Even if the structure of the network is complicated, by choosing a set of vertices to observe "appropriately", the behavior of the remaining vertices is uniquely determined after a long time, and there is no need to observe them. Of course, the validity of this argument depends on how we choose the set of vertices to observe. Then, let us consider how can we minimize the set of observations while

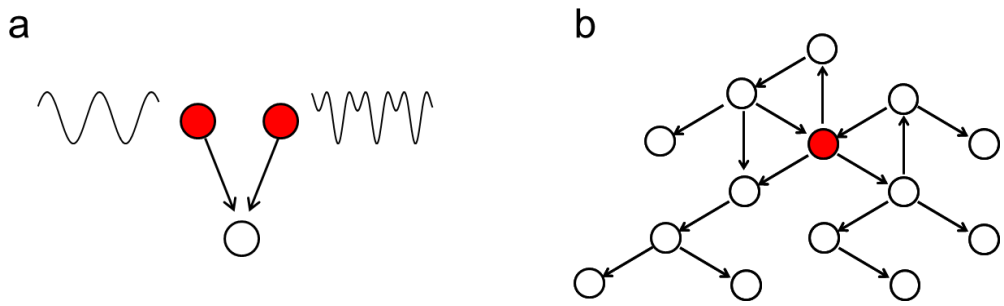

Figure 4 An intuitive explanation of linkage logic theory. (a) Given the behavior of the upper vertices, the behavior of the lower vertex is determined uniquely. (b) Similarly, by giving (observing/manipulating) dynamics of an appropriate subset of vertices, the dynamics of the remaining vertices are determined uniquely. To minimize the set of vertices, where dynamics are given, under the constraint that dynamics of the remaining are determined uniquely, we should choose the minimal FVS.

satisfying the condition that "the behavior of the remaining vertices is uniquely determined"? The answer is to choose the minimal FVS. From the definition of FVS, the rest of the network excluding FVS has no directed cycles. Therefore, it should be possible to arrange all remaining vertices in one direction from top to bottom, and their behavior should be uniquely determined sequentially from top to bottom. Therefore, it is sufficient to observe FVS only in order to observe the behaviors of the entire network.

By simply replacing "observation" in the above explanation with "manipulation", the controllability of the entire system by FVS is shown in the same way. The underlying logic here is simply that, given the upper-side behavior, the uniqueness of the asymptotic behavior of the lower-side is guaranteed independent of the choice of functions. This logic is universal beyond the formalism of dynamics. In fact, the importance of FVS in network dynamics was first pointed out in Boolean networks, which are discrete-time and discrete-state systems, by Akutsu et al. [8].

A few additional explanations are in order. (1) For a given graph structure, FVS is the minimum target of observation/manipulation that guarantees observation/control of the whole system independent of the choice of functions. In other words, depending on the choice of function, it may be possible to observe/control the whole system by a smaller set of variables. In other words, given only network information, observation/manipulation of the FVS is a "necessary and sufficient" condition for observing/controlling the whole dynamics, and for network systems with functions known, observation/manipulation of the FVS is a "sufficient" condition for observing/controlling the dynamics. (2) A vertex with self-activation is always included in the FVS. In the logic shown in Figure 4a, if the lower vertex has self-activation, its behavior is not uniquely determined from the upper vertex. However, this is not a problem, since this vertex has a self-activation, it is always included in the set to observe/manipulate i.e. FVS. (3) If there are vertices that do not receive any regulatory edge, they are not included in the minimal FVS. Considering the meaning of Eq. (1) (or (A1), (A2)), it is obvious that the dynamics of such vertices converge uniquely to an equilibrium point, and do not contribute to the diversity of behaviors.

Control of Network Systems Based on Structural Theory

We show a numerical demonstration of this theory in this section [5]. Figure 5a shows the regulatory network for the human circadian rhythm. From the network, oscillation of gene expression with 24-hour cycle appears, which is the origin of the physiological oscillations in humans. Figure 5a shows a network consist of 21 biomolecules determined from experiments, and it's minimal FVS contains 7 molecules. As well as experimental studies, there have also been many mathematical studies on human circadian rhythms. In one of such mathematical studies, a 24-hour oscillation is reproduced in numerical simulation of a mathematical model, where the network in Figure 5a is used, Hill functions or mass action kinetics are used for regulatory functions, and parameter values are chosen appropriately [9].

We changed the parameter values slightly from the original study [9] to obtain behaviors different from that of the real circadian clock. Under a new parameter, the system has four invariant sets (Figure 5b,c). Let us call the two stable periodic solutions P1 and P2, the unstable periodic solution UP, and the unstable equilibrium point USS. In the following, we consider the control of switching dynamics between these four solutions. The procedure is as follows (1) First, numerical simulations are performed and the behaviors of FVS on the four solutions is recorded as a time series. (2) The dynamics is calculated numerically according to the ordinary differential equations as the unmanipulated state. Then, the manipulation of FVS is started at a certain time point, where the behaviors of vertices in FVS are brought closer to the

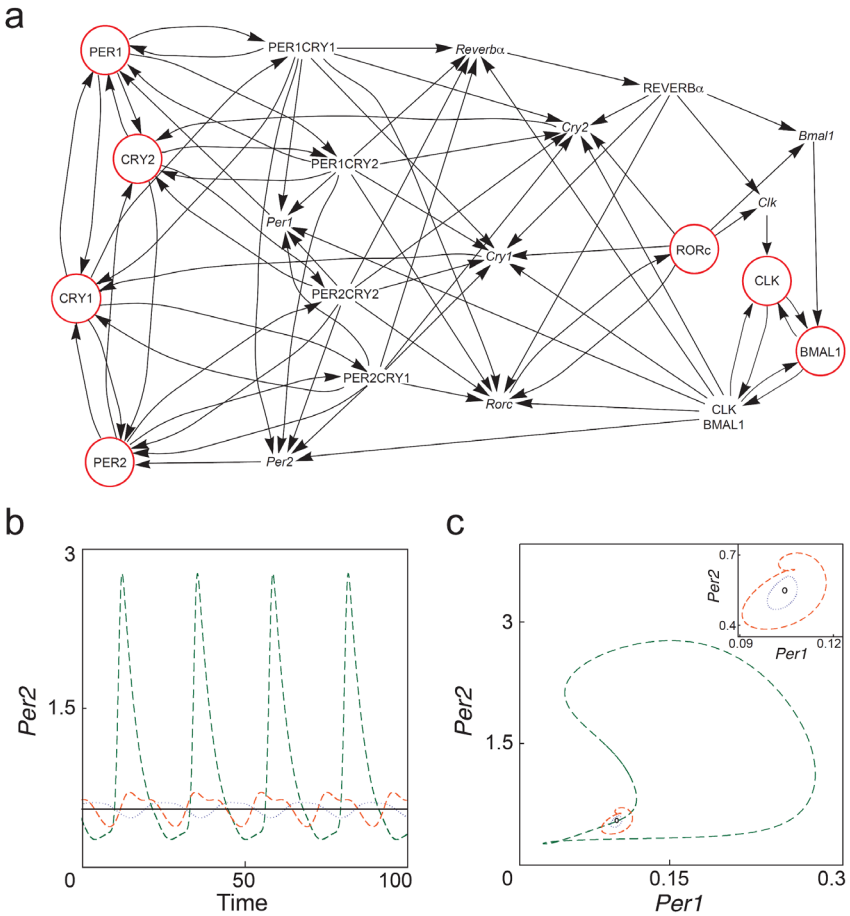

Figure 5 The gene network that governs the human circadian rhythm (a). The red circle indicates one of the minimal FVS. (b, c) Four invariant sets of the dynamics: Green: P1, Blue: P2 (stable periodic oscillations), Red: UP (unstable periodic oscillations), Black: USS (unstable equilibrium point).

target behavior (one of the time series prepared in (1)). (3) On the other hand, dynamics of variables other than the FVS factors are not manipulated, and calculated following the ordinary differential equations as is. The behavior of these variables not included in the FVS is observed.

Figure 6 shows the behavior of two variables (Per1, Per2) not included in the FVS, obtained by the above manipulations. In Figure 6a, the dynamics was on P1 before the manipulation. After starting manipulation of FVS to match the behavior on P2, the dynamics of the whole system quickly leaves P1, approaches P2, and then oscillates periodically on P2. Figure 6b shows the opposite, where the dynamics was oscillating on P2 before the manipulation. After manipulating the FVS to match the behavior on P1, the dynamics of whole system converges to P1. In conclusion, controlling the whole system to converge to the target trajectory was achieved by manipulating FVS only.

Figures 6c and 6d are particularly interesting. In Figure 6c, the dynamics was on P1 before the manipulation. Then, FVS was manipulated to match the behavior of the unstable periodic solution UP. As a result, the whole system approaches and converges on this trajectory as if UP was a stable periodic solution. In Figure 6d, for the dynamics on P1, FVS was manipulated and fixed to the value of the unstable equilibrium point USS. As a result, the whole system approaches and converges to this point, as if USS was the stable equilibrium point. In the original 21-variable system, UP and USS were of course unstable solutions. However, under the FVS manipulation, the dynamics show behaviors as if they were stable. The proof of the theorem shown in Appendix implies that under FVS manipulation, the dynamics always asymptotically approaches a unique behavior from any initial state as if the system has a global solution.

Control of the Ascidian Gene Regulatory Network

In this section, we will demonstrate the utility of the theory by applying it to an actual biological network. As introduced at the beginning (Figure 1), by repeated perturbation experiments, Sato's group determined a gene regulatory network

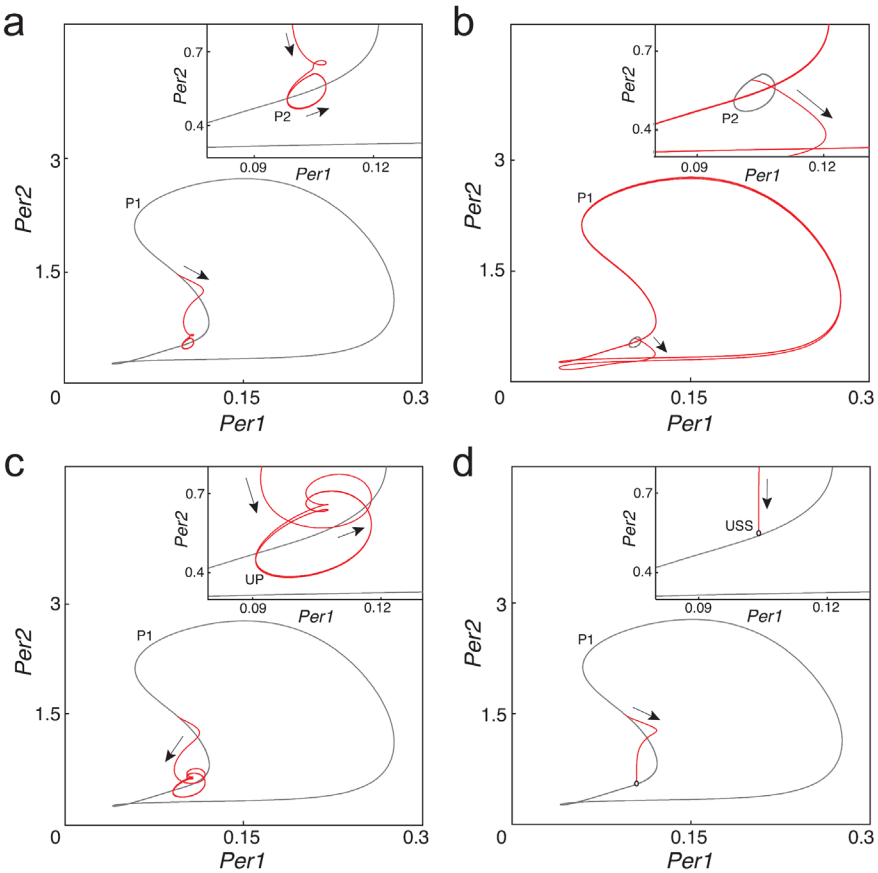

Figure 6 Controlling dynamics of circadian rhythm by seven FVS factors. Dynamical behaviors of (Per1, Per2), that are not included in the FVS, are plotted on the plane. (a) From P1 to P2. (b) From P2 to P1. (c) From P1 to UP. (d) From P1 to USS.

including 92 factors, from which 7 different cell types are generated during early development of ascidian [1]. Here, we attempt to understand the structure of this network in connection with the dynamics of cell fate specification.

Analysis of this network revealed that, despite its apparent complexity, the minimal FVS contains only five vertices [10]. The 5 factors are Zic-r.b, Erk signalilng, (Foxd or Twist-like1), (Neurogenin or Delta.b), and (Foxa.a, Nodal, or Snail) (Figure 7). There are 12 possible ways to choose the set. In other words, if the network information is complete, and cell fates are indeed specified based on the dynamics of this network, then manipulation of these five factors should allow us to reproduce the seven tissues arbitrarily. To verify this prediction, Kobayashi et al. conducted experiments in which they manipulating five factors [10].

From 12 minimum FVSs, we selected the one that was easy to manipulate experimentally (Foxa.a, Foxd, Neurogenin, Zic-r.b, Erk). Note that these genes (except for Erk signaling) code transcription factors. Each of these factors is known active in multiple different lineages in ascidian development, and not specific to unique cell type. Next, in order to exclude the effects of cell-cell interactions, we created a single-cell embryo system by cytochalasin B treatment. Under the treatment, cytokinesis was halted while expression dynamics proceeded (Figure 8). The activities of Foxa.a, Foxd, Neurogenin, and Zic-r.b were manipulated by introducing mRNA (activator) or morpholino antisense oligonucleotides (inhibitor) into the fertilized eggs, and Erk signal was manipulated by adding FGF or U0126 to the culture medium. After the manipulated embryos were cultured and grown for a while, the expression of marker genes corresponding to the seven tissues (epidermis, brain, nerve cord, muscle, chorda tympani, mesenchyme, and endoderm) were measured, and the results of manipulation were analyzed.

Figures 8(b) and (c) show the expression of marker genes without any manipulation of the FVS factor as a control. It turns out that only a limited cell fates i.e. epidermis, brain, or nerve cord were induced in the single-cell embryo. The fate of embryo is not uniquely determined, as marker gene expressions were varied between embryo.

In Figure 9, shows results of an example manipulation, where three genes were suppressed and two factors were activated. Foxa.a, Foxd, and Neurogenin were suppressed using morpholino antisense oligos, Zic-r.b was activated by introducing the mRNA, and the Erk signaling pathway was activated by adding FGF to the culture medium. The

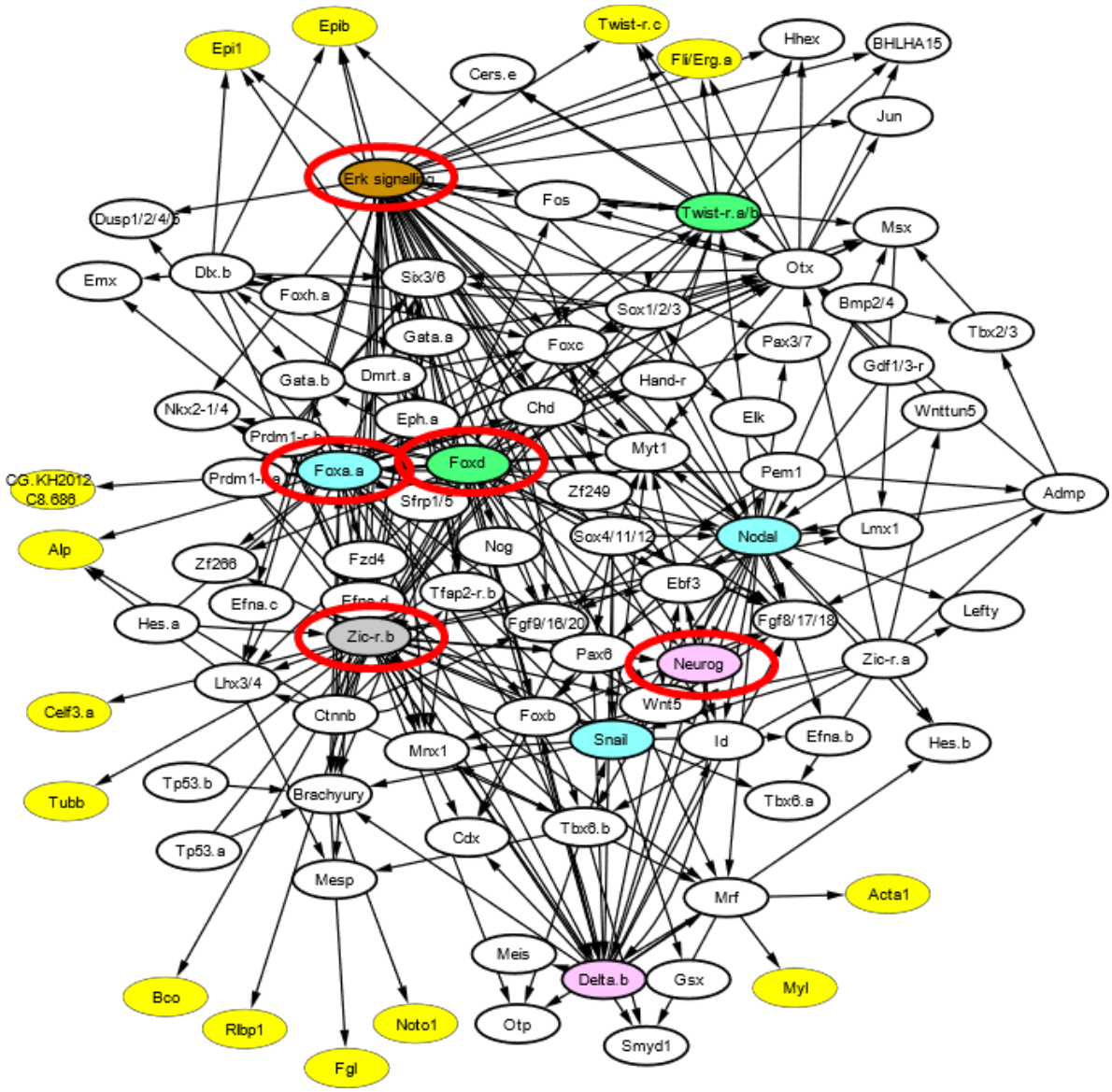

Figure 7 Gene regulatory network for cell-fate specification in ascidian early development. This network contains 92 factors. The vertices circled in red indicate the selection of FVS factors by Kobayashi et al. [10]. Colored vertices mean that the FVS factors can be alternated between the same color. Vertices in yellow indicate marker genes for cell fates.

experiment was repeated 25 times using different embryos, and in all results (except for one case in which no marker genes was expressed), it was observed that only the mesenchyme markers were activated. In other words, under this manipulation, the fate of the embryo is uniquely and deterministically directed to the mesenchyme.

In FVS manipulation, each of the five FVS factors is either activated or suppressed, i.e. there are $32(=2^5)$ possible manipulations. A total of 32 exhaustive manipulation experiments were performed to capture all possible dynamics behaviors that this system can produce. The results are summarized in Figure 10. In many cases, the manipulations resulted in the unique and deterministic marker gene induction similarly to the example shown in Figure 9. For example, when all five genes were inactivated, only epidermal markers were always expressed, and when only Foxa.a was activated and the remaining four were inactivated, only endodermal markers were always expressed. It is important to note that in FVS manipulation, not only the activated factors are important for induction, but the manipulation of each FVS factor, including those are inactivated, is equally important.

Throughout the exhaustive manipulations, there are methods by which marker expression for each of epidermis, endoderm, notochord, nerves, brain, and mesenchyme is induced uniquely and deterministically. We were indeed able to

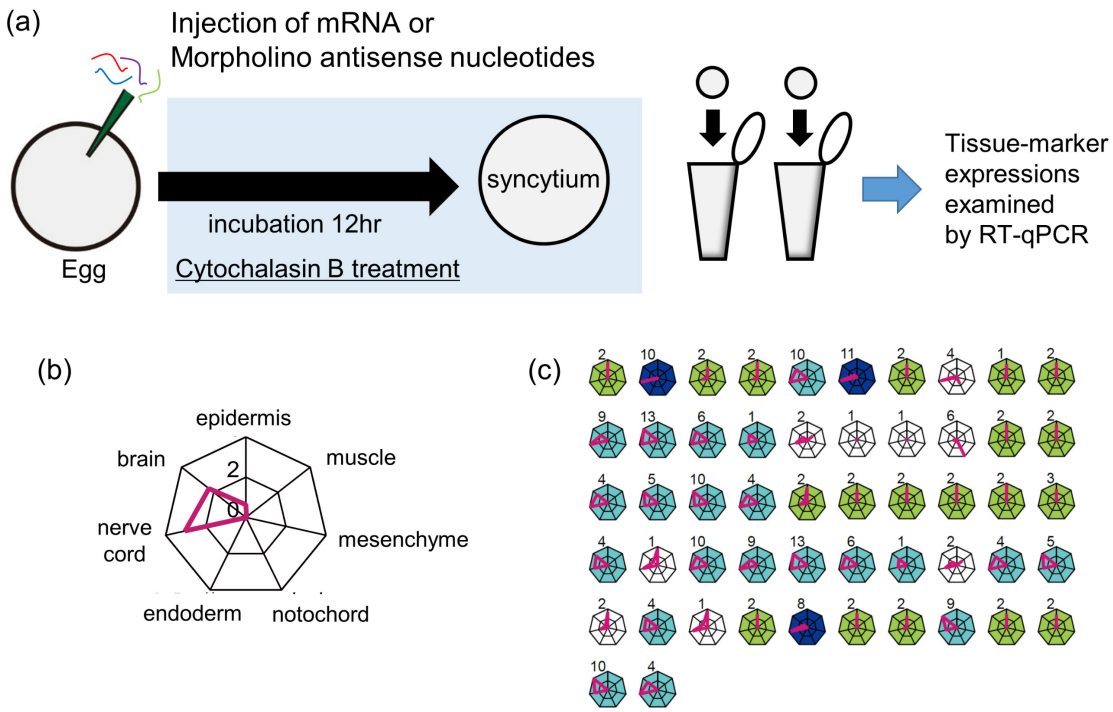

Figure 8 Experiments to control the gene regulatory network of ascidians. (a) Manipulate molecules in the FVS by introducing nucleotides into fertilized eggs or by adding chemicals to the culture medium. After culture, the cell fate is identified by the expression of marker genes. (b) The population average of marker gene expression under no manipulation of FVS genes. (c) Marker gene expression of each individual experiment under no manipulation. The numbers on the radar chart indicate how many times larger the maximum of the chart is compared to the expression level of normal embryos. In (c), the background color of the radar chart indicates that tissue-specific expression was observed. Green: epithelium, light blue: brain, dark blue: nerve cord.

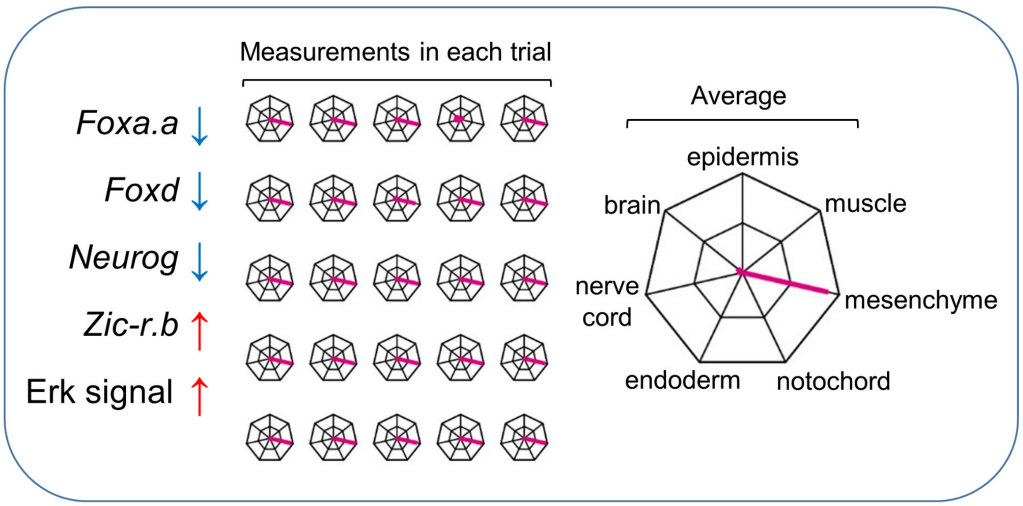

Figure 9 An example of FVS factor manipulation. Under this manipulation, only a single marker corresponding to the mesenchyme tissue was always active without variety between individuals.

Foxa.a	*Foxd*	*Neurog*	*Zic-r.b*	Erk signal	marker expression	*Foxa.a*	*Foxd*	*Neurog*	*Zic-r.b*	Erk signal	marker expression
Down	Down	Down	Down	Down	epidermis	Up	Up	Down	Down	Up	none
Down	Down	Down	Down	Up	none	Up	Down	Up	Down	Up	**nerve cord**
Up	Down	Down	Down	Down	endoderm	Up	Down	Down	Up	Up	**nerve cord**
Down	Up	Down	Down	Down	notochord	Down	Up	Up	Down	Up	**nerve cord**
Down	Down	Up	Down	Down	**nerve cord**	Down	Up	Down	Up	Up	none
Down	Down	Down	Up	Down	**brain**	Down	Down	Up	Up	Up	**nerve cord**
Up	Down	Down	Down	Up	none	Up	Up	Up	Down	Down	notochord
Down	Up	Down	Down	Up	none	Up	Up	Down	Up	Down	endoderm
Down	Down	Up	Down	Up	**verve cord**	Up	Down	Up	Up	Down	**nerve cord**
Down	Down	Down	Up	Up	mesenchyme	Down	Up	Up	Up	Down	endoderm
Up	Up	Down	Down	Down	notochord	Up	Up	Up	Down	Up	notochord
Up	Down	Up	Down	Down	**nerve cord**	Up	Up	Down	Up	Up	endoderm
Up	Down	Down	Up	Down	**nerve cord**	Up	Down	Up	Up	Up	**brain**
Down	Up	Up	Down	Down	none	Down	Up	Up	Up	Up	none
Down	Up	Down	Up	Down	none	Up	Up	Up	Up	Down	none
Down	Down	Up	Up	Down	**brain**	Up	Up	Up	Up	Up	none

Figure 10 Marker gene expression in the exhaustive 2^5=32 manipulations of the gene regulatory network in ascidians. Marker genes specifically expressed in each FVS-manipulated embryo are shown. “None” indicates that there is no predominantly expressed gene.

induce six out of the seven tissues observed in normal development, except for the muscle tissue, using methods within the same framework [10]. Various experimental methods for tissue induction have been developed in developmental biology, but they are all empirical and specific to the focal tissue types. There has never been an achievement that manipulation criteria were determined rationally from mathematical theory and various cell fates were induced systematically within a single framework.

On the other hand, muscle cells could not be induced despite the exhaustive manipulation of FVS (albeit in a binary range). Mathematically, the FVS is the minimal and sufficient target for manipulation to reproduce all the behaviors of the system. This result may imply that there is still some missing network information. Another possibility is that factors outside of the regulatory network, such as cell-cell interactions, play important roles in determining muscle fate. However, considering the fact that ascidian muscle tissue is a well-known example of autonomous development and has been known for over 100 years, cell-cell interactions are unlikely to play essential roles in this regard.

There are cases where addition of a single regulatory edge to the regulatory network changes the FVS [11]. These may be potential candidates for unknown regulations. Therefore, a list of candidate unknown regulations was created by adding a hypothetical regulation one by one that could change the FVS. The list was examined by the Sato lab, and it was confirmed that one of them, the regulation from Mrf to Tbx6-r.b, actually exists and works. Adding this new regulation to the network, an FVS of size 6 was obtained including Tbx6-r.b, {Foxa.a, Foxd, Neurogenin, Zic-r.b, Erk, Tbx6-r.b}. In the previous experiment, only five factors without including Tbx6-r.b were manipulated. That may be the reason why the experiment could not reproduce all the tissue types.

However, an exhaustive manipulation of the six new FVS factors would require $64(=2^6)$ manipulation experiments, making this project even more difficult than the 32 manipulation experiments. Therefore, an experiment was conducted to activate the newly added factor, Tbx6-r.b, and inactivate the five factors originally included in FVS. Although this was only a pilot experiment, surprisingly, muscle markers were induced. Furthermore, it was confirmed that the other 6 tissues, which had been known to be induced by manipulating the 5 factors, were also induced by manipulating the 6 FVS factors. Thus, it was confirmed that the insufficiency of manipulation with 5 factors and the sufficiency of manipulation with 6 factors [11].

Let us summarize the results of this section. By applying linkage logic theory to the ascidian gene regulatory network, it was demonstrated that the cell-fate specification system is indeed controllable by manipulating the activity of the FVS factors. Identifying the important molecules from the network structure alone, without using information on the biological properties of the molecules. Controlling the cell fate system completely is a remarkable achievement from both biological and mathematical viewpoints. In addition, it is interesting that the network, which serves as the basic information, was updated through the exchange of mathematical prediction and experimental test. We would say that this was possible since linkage logic is a model-free theory that does not assume functions or parameter values. It is a fairly strong proposition from theoretical side that the expected results were not obtained because the data were incomplete. This argument was possible because the theory does not introduce model assumptions. Experimental verification of the prediction from linkage logic is not just confirmation, but has the meaning of testing the completeness of the network information.

Concluding Remark and Outlook

The linkage logic presented in this review is a mathematical theory that directly connects the network structure and dynamics of regulatory systems. The conclusion from the theory is quite simple: To understand the dynamics of complex regulatory networks, we should focus on the FVS. This is because the FVS is the minimal sufficient subset of variables to capture the dynamics of the whole system, and at the same time, the minimal sufficient subset of variables to control the whole system to converge to a desired behavior.

We have so far introduced a method to determine driver nodes based on network structure in order to control a nonlinear dynamical system to converge to an arbitrary behavior. On the other hand, in the field of control theory, another criterion for selecting the driver nodes has been discussed for linear dynamical systems. Kalman [12] defined the controllability of a linear dynamical system based on the full rankness of the controllability matrix. Lin [13] gave a criteria to choose driver nodes from a network structure based on Kalman's definition. Later, Liu et al. [14] adapted the method to many examples of regulatory networks by an algorithm using concept of 'maximum matching'. Liu et al. calculated the size of the smallest set of driver nodes determined by this algorithm for various network systems, including those describing biological and social systems. Zanudo et al. [15] similarly compared the size of the driver node set between FVS control and maximum matching method for various network systems. Which method is more advantageous (i.e. yields a smaller set of driver nodes) depends on the structure of the network. Although proof of controllability is given only for linear systems, numerical control of nonlinear systems has been studied using the same guidelines. There have been both positive and negative reports on the possibility of controlling nonlinear systems with this method.

The linkage logic theory introduced here is not only valid for specific system of ascidian cell-specification, but can be applied to any regulatory networks. Once we have enough information on regulatory relationships between biomolecules, we can discuss the behavior of the system of any organisms. Moreover, the mathematical formulation of our definition of a regulatory system implies that this theory is applicable not only to biological regulatory networks, but to any regulatory relationships. We may expect that the theory, which originally developed for living systems, will be used for understanding behaviors of high-dimensional dynamical systems, like economic systems or human relationships networks.

Appendix

(1) Formulation

Consider a directed graph $\Gamma = (V, E)$ where V is a set of vertices and E is a set of directed edges, where nodes represent biomolecules and edges represent regulatory linkages. Suppose that the dynamics of molecular activity on this graph is described by a system of ordinary differential equations in the form [4,5]:

$$\begin{aligned} \frac{du_n}{dt} &= F_n(\boldsymbol{u}) \\ &= F_n\left(u_n, \boldsymbol{u}_{I_n}\right) \end{aligned}, \quad n = 1, \cdots, |V| \tag{A1}$$

Here, $\boldsymbol{u}_{I_n}$ is a vector notation with u_i $(i \in I_n)$ as elements.

We also assume that (i) each F_n is a continuous and differentiable function, (ii) there are no divergent solutions (F_n is bounded), and (iii) F_n satisfies the "decay condition":

$$\partial_1 F_n\left(u_n, \boldsymbol{u}_{I_n}\right) < 0. \tag{A2}$$

Here $\partial_1 F$ has a special meaning: "the partial derivative of F with respect to the first argument". This equation (A2) corresponds to the decay term or self-inhibition in equation (1). The set $I_n \subseteq V$ is the input set of n, a subset of molecules that regulate molecule n, i.e. $I_n := \{i \mid (i \to n) \in E\}$. The set I_n includes n $(n \in I_n)$ i.e. a self-regulation, if and only if $\partial F_n / \partial u_n$ is "not always negative". In other words, I_n should not include the effects of decay or self-inhibition because they are already represented by the decay condition (A2). In the discussion below, we consider that the sets of I_n $(\forall n \in V)$ directly represent the graphical structure of the regulatory network, i.e. the regulatory network is defined by the set of I_n. That is, if there exists a self-regulatory loop in a graph, it indicates an autoactivation regulation.

Note that the decay condition $\partial_1 F_n < 0$ does not always imply $\partial F_n / \partial u_n < 0$. For example, suppose that we have a differential equation where $\partial F_n / \partial u_n < 0$ does not hold. In that case, using $\tilde{I}_n := I_n \cup \{n\}$ we can reconstruct the function as

$$\tilde{F}_n\left(\eta_n, \boldsymbol{u}_{\tilde{I}_n}\right) := F_n\left(\boldsymbol{u}_{I_n}\right) + u_n - \eta_n \tag{A3}$$

where the decay condition (A2) is satisfied. In other words, when there is no decay term in the mathematical formula, we can interpret that there is a self-activating term that cancels out the decay term, which is assumed to exist. Thus, the requirement of decay conditions does not constrain the mathematical formula of the differential equations.

(2) Two definitions

Definition 1 (Feedback Vertex Set: FVS)

In a directed graph $\Gamma = (V, E)$, a subset $I \subseteq V$ of vertices is called a feedback vertex set (FVS), if the graph $\Gamma \setminus I$ excluding the subset I has no directed cycles.

FVS is one of the major concepts in graph theory, and it has been shown that finding an FVS for any graph is NP-hard problem [7,8].

Definition 2 (Determining nodes)

In a dynamical system, a subset $J \subseteq V$ of variables is called a set of determining nodes, if the following condition is satisfied:

For any two solutions $\boldsymbol{u}(t)$ and $\tilde{\boldsymbol{u}}(t)$ of the dynamics, convergence $\tilde{\boldsymbol{u}}_J(t) - \boldsymbol{u}_J(t) \to 0$ $(t \to +\infty)$ of two solutions $\boldsymbol{u}_J(t)$, $\tilde{\boldsymbol{u}}_J(t)$ on the subset J always implies the convergence $\tilde{\boldsymbol{u}}(t) - \boldsymbol{u}(t) \to 0$ $(t \to +\infty)$ of two solutions $\boldsymbol{u}(t)$, $\tilde{\boldsymbol{u}}(t)$ of the whole system V.

A concept of determining nodes has been discussed in the field of dynamical systems theory, especially in the context of fluid dynamics. A set of determining nodes is a subset of variables such that one can identify any asymptotic behaviors of the whole system by observing them without observing all the variables in the system [16]. By considering the contrapositive of Definition 2, it is easy to understand that observations of determining nodes capture the diversity of the asymptotic behaviors of a dynamical system. The contrapositive ($\tilde{\boldsymbol{u}}(t) - \boldsymbol{u}(t) \not\to 0 \Rightarrow \tilde{\boldsymbol{u}}_J(t) - \boldsymbol{u}_J(t) \not\to 0$) means that "if two solutions do not asymptote to the same behavior, then (at least one of) the variables in the subset J do not asymptote to the same behavior." A set of determining nodes is originally a concept related to a finite number of points in a space that should be measured to identify the solution of the Navier-Stokes equation, and its relationship with graphs has not been discussed at all.

(3) The theorem and its proof

Theorem (FVS = set of determining nodes)

In the dynamics (A.1) on a directed graph, an FVS of the graph is a set of determining nodes regardless of the choice of F_n. Conversely, if a subset of variables is a set of determining nodes regardless of the choice of F_n , then it is an FVS of the graph.

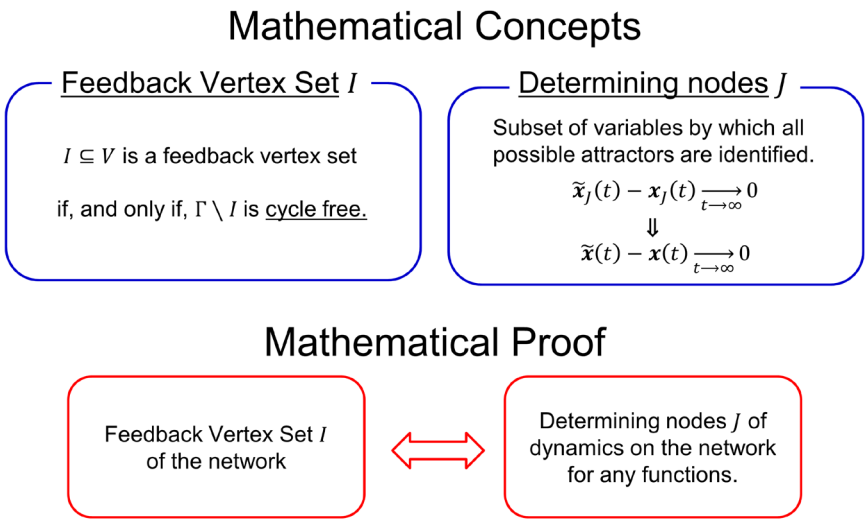

Figure A1 The structure of the theorem.

Here is an outline of the proof. First, we show that "FVS ⇒Set of determining nodes".

(1) For a given FVS I, a set $K := V \setminus I$ of vertices not included in I is called the nonFVS. From the definition of FVS, a directed cycle cannot be composed only of the vertices in nonFVS K. Therefore, for each vertex $k \in K$ in the nonFVS, the number k can be reassigned so that $I_k \subseteq I \cup \{1, \dots, k-1\}$ $(\forall k \in K)$. That is, $k \in K$ is reordered so that a regulator vertex has a younger number than the regulated vertex for each regulatory edge.

(2) Let $w_n(t) := \tilde{u}_n(t) - u_n(t)$ be the difference in variable u_n between two dynamical solutions $\boldsymbol{u}(t)$ and $\tilde{\boldsymbol{u}}(t)$ at time t. Suppose that $w_i(t) \to 0$ is given for all vertices $\forall i \in I$ in the FVS I, and show by mathematical induction that $w_k(t) \to 0$ for each vertex $k \in K$ in the nonFVS K. For the dynamics of the difference of solutions $\boldsymbol{w}(t) := \tilde{\boldsymbol{u}}(t) - \boldsymbol{u}(t)$, we can derive

$$\begin{aligned}\frac{d}{dt}\boldsymbol{w}(t) &= \frac{d}{dt}\tilde{\boldsymbol{u}}(t) - \frac{d}{dt}\boldsymbol{u}(t)\\ &= \boldsymbol{F}\big(\boldsymbol{u}(t) + \boldsymbol{w}(t)\big) - \boldsymbol{F}\big(\boldsymbol{u}(t)\big)\\ &= \big[\boldsymbol{F}\big(\boldsymbol{u}(t) + \theta\boldsymbol{w}(t)\big)\big]_{\theta=0}^{1}\\ &= \int_0^1 \frac{d}{d\theta}\boldsymbol{F}\big(\boldsymbol{u}(t) + \theta\boldsymbol{w}(t)\big)d\theta\\ &= \int_0^1 \frac{\partial}{\partial \boldsymbol{u}}\boldsymbol{F}\big(\boldsymbol{u}(t) + \theta\boldsymbol{w}(t)\big)\cdot\boldsymbol{w}(t)d\theta\\ &= A(t)\boldsymbol{w}(t)\end{aligned} \tag{A4}$$

$$A(t) := \left(\int_0^1 \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{u}(t)+\theta\boldsymbol{w}(t)} d\theta\right) \tag{A5}$$

where A is the matrix obtained by integrating each component of the Jacobian matrix of $\boldsymbol{F}$ at $\boldsymbol{u}(t) + \theta\boldsymbol{w}(t)$ with respect to θ. Note that any vertex in K does not have self-regulation i.e. self-activation. Then dynamics of each $k \in K$ can be represented by a linear nonautonomous dynamic equation $\dot{w}_k(t) = -a_k(t)w_k(t) + \boldsymbol{b}_k^T(t)\boldsymbol{w}_{\boldsymbol{I_k}}(t)$, where the first term correspond to the decay condition (A.2) and the second term is from the second argument set in (A.1). The nonautonomous coefficients $a_k(t) \in \mathbb{R}$ and $\boldsymbol{b}_k(t) \in \mathbb{R}^{|I_k|}$ are bounded by positive constants a_0 and b_0 as $0 < a_0 \le a_k(t)$ and $|\boldsymbol{b}_k(t)| \le b_0$, respectively. Assuming that $w_n(t) \to 0$ has already been shown for $\forall n \in \{1, \cdots, k-1\}$, let us show for $n = k$. Solving the equations yields:

$$\begin{aligned}w_k(t) &= \exp\left(-\int_0^t a_k(s)ds\right)w_k(0) + \int_0^t \exp\left(-\int_s^t a_k(\sigma)d\sigma\right)\boldsymbol{b}_k^T(s)\boldsymbol{w}(s)ds\\ &\le \exp\left(-\int_0^t a_k(s)ds\right)|w_k(0)| + \textstyle\sum_{j\in I_k}\int_0^t \exp\left(-\int_s^t a_k(\sigma)d\sigma\right)|\boldsymbol{b}_k(s)|\big|w_j(s)\big|ds\\ &\le \exp(-a_0 t)\,|w_k(0)| + \textstyle\sum_{j\in I_k}\int_0^t \exp\big(-a_0(t-s)\big)\,b_0\big|w_j(s)\big|ds\\ &\to 0 \quad (t \to \infty)\end{aligned} \tag{A6}$$

Note that the convergence of the first term is shown by $\exp(-a_0 t) \to 0$ $(t \to \infty)$, and the convergence of the second term is shown by the fact that $w_j(s) \to 0$ $(s \to \infty)$ for $j \in I_k \subseteq I \cup \{1, \dots, k-1\}$. When $k = 1$, then $I_1 \subseteq I$, it is obvious that $w_1(t) \to 0$ from the similar argument. From the above the first half of the theorem is proved.

"Set of determining nodes ⇒ FVS"' is shown by taking the contrapositive, that is, a subset of vertices which is not an FVS is not a set of determining nodes. Suppose I' is not an FVS, i.e. $\Gamma \setminus I'$ contains directed cycles. By choosing the function of the vertices appropriately, dynamics can be constructed so that I' is not a set of determining nodes. For example, all functions included in I' are taken to be simple decay $F_n\big(x_n, \boldsymbol{x}_{I_n}\big) := -x_n$. From this, the behaviour of $\Gamma \setminus I'$ cannot be captured by I'. However, there is a cycle within $\Gamma \setminus I'$ and, by choosing these functions, it is possible to create diversity in the solutions such as multiple stationary points. In other words, I', which is not an FVS, is not always a set of determining nodes for arbitrary functions. (End of proof)

Here we add a supplementary note on the control of dynamic by manipulating the FVS. The theorem claims that when the behaviors $\boldsymbol{u}_{FVS}(t)$ of vertices in the FVS are made closer to the target behaviors $\tilde{\boldsymbol{u}}_{FVS}(t)$ $(\tilde{\boldsymbol{u}}_{FVS}(t) - \boldsymbol{u}_{FVS}(t) \to 0$ $(t \to +\infty))$, the behavior $\boldsymbol{u}(t)$ of the whole system spontaneously approaches the target behavior $\tilde{\boldsymbol{u}}(t)$ $(\tilde{\boldsymbol{u}}(t) - \boldsymbol{u}(t) \to 0$ $(t \to +\infty))$. The target $\tilde{\boldsymbol{u}}_{FVS}(t)$ does not have to be an ω-limit set such as stable equilibria or stable periodic oscillations, but can be an α-limit set such as unstable equilibrium points or unstable periodic solutions. Moreover, if we pay careful attention to the theorem, the target need not be the original asymptotic behaviors of the dynamical systems. Given any behaviors in the FVS, the entire system will asymptotically approach a uniquely determined behavior (whatever that may be). In other words, under any manipulation of FVS, the system loses the diversity of asymptotic behaviors. When FVS approaches the system's original asymptotic behavior, the entire system uniquely asymptotes to that behavior.

Conflict of Interest

None

Author Contributions

AM wrote the review.

Data Availability

The evidence data generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

The achievement in this review were realized by the great efforts of the following collaborators: K. Kobayashi, K. Maeda, Y. Satou and M. Tokuoka. I express my sincere thanks to them for their supports and critical discussions.

References

[1] Imai, K. S., Levine, M., Satoh, N., Satou, Y. Regulatory blueprint for a chordate embryo. Science 312, 1183-1187 (2006). https://doi.org/10.1126/science.1123404
[2] Mochizuki, A. Structure of regulatory networks and diversity of gene expression patterns. J. Theor. Biol. 250, 307-321 (2008). https://doi.org/10.1016/j.jtbi.2007.09.019
[3] Mochizuki, A., Saito, D. Analyzing steady states of dynamics of bio-molecules from the structure of regulatory networks. J. Theor. Biol. 266, 323-335 (2010). https://doi.org/10.1016/j.jtbi.2010.06.007
[4] Fiedler, B., Mochizuki, A., Kurosawa, G., Saito, D. Dynamics and control at feedback vertex sets. I: Informative and determining nodes in regulatory networks. J. Dyn. Differ. Eqns. 25, 563-604 (2013). https://doi.org/10.1007/s10884-013-9312-7
[5] Mochizuki, A., Fiedler, B., Kurosawa, G., Saito, D. Dynamics and control at feedback vertex sets. II: A faithful monitor to determine the diversity of molecular activities in regulatory networks. J. Theor. Biol. 335, 130-146 (2013). https://doi.org/10.1016/j.jtbi.2013.06.009
[6] Bondy, J. A., Murty, U. S. R. Graph Theory (Graduate Texts in Mathematics). (Springer-Verlag, London, 2008).
[7] Festa, P., Pardalos, P. M., Resende, M. G. C. Feedback set problems. in Handbook of Combinatorial Optimization. (Du, D.-Z., Pardalos, P. M. eds.) Supplement vol. A, 209-259 (Springer, Boston, 2000).
[8] Akutsu, T., Kuhara, S., Maruyama, O., Miyano, S. A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. Genome Inform. Ser. Workshop Genome Inform. 9, 151-160 (1998).
[9] Mirsky, H. P., Liu, A. C., Welsh, D. K., Kay, S. A., Doyle III, F. J. A model of the cell-autonomous mammalian circadian clock. Proc. Natl. Acad. Sci. U.S.A. 106, 11107-11112 (2009). https://doi.org/10.1073/pnas.0904837106
[10] Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A., Satou, Y. Controlling cell fate specification system by key genes determined from network structure. iScience 4, 281-293 (2018). https://doi.org/10.1016/j.isci.2018.05.004
[11] Kobayashi, K., Maeda, K., Tokuoka, M., Mochizuki, A., Satou, Y. Controlling dynamics of a gene regulatory network of a chordate embryo using a structural theory. Sci. Rep. 11, 4001 (2021).
[12] Kalman, R. E. Mathematical description of linear dynamical systems. J. Soc. Indus. Appl. Math. Ser. A 1, 152-192 (1963). https://doi.org/10.1137/0301010
[13] Lin, C. T. Structural controllability. IEEE Transacitons on Automatic Control 19, 201-208 (1974).
[14] Liu, Y. Y., Slotine, J. J., Barabasi, A. L. Controllability of complex networks. Nature 473, 167-173 (2011). https://doi.org/10.1038/nature10011
[15] Zanudo, J. G. T., Gang, Y., Albert, R. Structure-based control of complex networks with nonlinear dynamics. Proc. Natl. Acad. Sci. U.S.A. 114, 7234-7239 (2017). https://doi.org/10.1073/pnas.1617387114
[16] Foias, C., Temam, R. Determination of the solutions of the Navier-Stokes equations by a set of nodal values. Math. Comput. 43, 117-133 (1984). https://doi.org/10.2307/2007402