BioSystems 142-143 (2016) 15–24

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Relative stability of network states in Boolean network models of gene regulation in development

Joseph Xu Zhou[a,b,1], Areejit Samal[a,c,d,1], Aymeric Fouquier d'Hérouël[a,e], Nathan D. Price[a], Sui Huang[a,*]

[a] Institute for Systems Biology, Seattle, WA, USA
[b] Kavli Institute for Theoretical Physics, UC Santa Barbara, CA, USA
[c] The Institute of Mathematical Sciences, Chennai, India
[d] The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy
[e] Luxembourg Centre for Systems Biomedicine, Esch-sur-Alzette, Luxembourg

ARTICLE INFO

Article history:
Received 12 October 2015
Received in revised form 27 February 2016
Accepted 2 March 2016
Available online 7 March 2016

Keywords:
Gene regulatory network (GRN)
Cell differentiation
Multistable dynamical system
Boolean network (BN)
Attractor states
Epigenetic landscape

ABSTRACT

Progress in cell type reprogramming has revived the interest in Waddington's concept of the *epigenetic landscape*. Recently researchers developed the quasi-potential theory to represent the Waddington's landscape. The Quasi-potential $U(\mathbf{x})$, derived from interactions in the gene regulatory network (GRN) of a cell, quantifies the relative stability of network states, which determine the *effort* required for state transitions in a multi-stable dynamical system. However, quasi-potential landscapes, originally developed for continuous systems, are not suitable for discrete-valued networks which are important tools to study complex systems. In this paper, we provide a framework to quantify the landscape for discrete Boolean networks (BNs). We apply our framework to study pancreas cell differentiation where an ensemble of BN models is considered based on the structure of a minimal GRN for pancreas development. We impose biologically motivated structural constraints (corresponding to specific type of Boolean functions) and dynamical constraints (corresponding to stable attractor states) to limit the space of BN models for pancreas development. In addition, we enforce a novel functional constraint corresponding to the relative ordering of attractor states in BN models to restrict the space of BN models to the biological relevant class. We find that BNs with *canalyzing/sign-compatible* Boolean functions best capture the dynamics of pancreas cell differentiation. This framework can also determine the genes' influence on cell state transitions, and thus can facilitate the rational design of cell reprogramming protocols.

1. Introduction

A hallmark of multicellular organisms is the co-existence of distinct differentiated cell types with different functions and stable gene expression patterns. A less specialized cell, a stem or progenitor cell, spawns a variety of more specialized progeny cells through cell differentiation. Once differentiated, a specialized cell's gene expression pattern is relatively robust against perturbations emanating from a noisy environment. Where does this stability come from? How do gene expression patterns change as cells differentiate in response to external cues, and thereby, transition from one stable gene expression pattern to another? In principle, such questions can be answered by understanding the interactions between the genes in the underlying gene regulatory network (GRN), which constrain the changes in the gene expression patterns, producing stable and unstable steady states. The dynamical system associated with GRNs can be modeled by a system of ordinary differential equations (ODEs) where continuous variables represent the expression levels of individual genes. However, with ODEs one is quickly limited by the number of configurations of the networks due to the exponential growth of complexity with the number of genes as well as the general lack of information on the parameters that characterize the interactions between genes. A widely used alternative approach to study GRNs is Boolean networks (BNs), a framework that enables modelling of networks with hundreds of genes or analyze large statistical ensembles of networks of random structure (Kauffman, 1969, 1993). Analysis of an ensemble of BNs can yield

Abbreviations: BN, Boolean network; CF, canalyzing function; GRN, gene regulatory network; MFPT, mean first passage time; NCF, nested canalyzing function; ODE, ordinary differential equation; SGN, sign-compatible function.

* Corresponding author.
E-mail address: shuang@systemsbiology.org (S. Huang).
[1] These authors contributed equally to this work

http://dx.doi.org/10.1016/j.biosystems.2016.03.002
0303-2647/

insights on the relationship between structure and dynamics of GRNs (Kauffman, 1969, 1993; Shmulevich and Kauffman, 2004).

In 1969 Kauffman introduced BNs to study the dynamics of GRNs (Kauffman, 1969). Since then BNs have been used to model a wide range of biological phenomena such as cell cycle, cellular differentiation and evolution of GRNs (Huang and Ingber, 2000; Klemm and Bornholdt, 2005; Balleza et al., 2008; Torres-Sosa et al., 2012; Espinosa-Soto et al., 2011; Lau et al., 2007; Samal and Jain, 2008; Li et al., 2004; Albert, 2003; Villani et al., 2011; Krumsiek et al., 2011; Chang et al., 2011; Flöttmann et al., 2012; Villarreal et al., 2012). Specifically, BNs have been extensively used to study developmental processes. Villani et al. (2011) have developed a BN framework for cell differentiation. Krumsiek et al. (2011) have developed a BN model to recapitulate hematopoiesis. Chang et al. (2011) employed a BN model to explain human embryonic stem cell differentiation and the generation of induced pluripotent stem cells (iPSCs). Klipp et al. (Flöttmann et al., 2012) used a BN model to study the influence of gene regulation, methylation and histone modifications on cell differentiation. Alvarez-Buylla et al. (Balleza et al., 2008; Villarreal et al., 2012) used BNs to explain cell differentiation and developmental ordering in the floral organ of Arabidopsis. An important limitation of these reconstructed BN models (Huang and Ingber, 2000; Klemm and Bornholdt, 2005; Balleza et al., 2008; Torres-Sosa et al., 2012; Espinosa-Soto et al., 2011; Lau et al., 2007; Samal and Jain, 2008; Li et al., 2004; Albert, 2003; Villani et al., 2011; Krumsiek et al., 2011; Chang et al., 2011; Flöttmann et al., 2012; Villarreal et al., 2012) for different biological processes is their specification of one defined set of Boolean functions for genes in the network out of a multitude of possible choices (Henry et al., 2013) that can reproduce the biologically relevant cell states as network attractors, and the reason for the chosen set of Boolean functions often remains elusive. Also experimental observations in cell differentiation systems usually are consistent with a large number of possible Boolean functions rather than suggesting a single well-defined set, giving rise to a set of possible BNs that can describe the observed gene expression patterns of the attractors (Henry et al., 2013). Thus, one always wonders whether the reported results would still hold for other choices of functions and how *structurally robust* the predicted dynamics is for the observed attractor states.

A more stringent requirement on a model capturing the development of multicelluar organisms is the following constraint. In addition to recapitulating the multiple observed attractors of the network, the model of the developmental GRN should also reproduce the experimentally observed *relative stabilities of attractors*, i.e., the model has to relate the different attractor states to each other based on their *relative stabilities*. By that we mean the *relative ease* for transitioning from one attractor state (A) to another state (B) which epitomizes the developmental process. More formally, in a stochastic system, the relative ease of transitioning from state A to state B would be given by the probability $P(A \rightarrow B)$ for transition from A to B (given random flutuations in gene expression). Note that such transition probabilties are typically asymmetric (i.e., $P(A \rightarrow B) \neq P(B \rightarrow A)$)—a property that ultimately accounts for the directionality (irreversibility) of development.

Interactions between genes collectively produce the developmental ordering of different cell types which is robust and repeatable during embryogenesis. Therefore, once the multiple attractors of the dynamical system are determined, it is necessary to evaluate their relative stabilities in order to derive a consistent relative ordering for all attractors in a developmental process (if one exists). Recently, some of us have derived a framework to calculate the relative stabilities of cell attractors in continuous ODE-based GRN models using least action principles (Zhou et al., 2012). However, ODE-based GRN models are not well-suited to model large networks, let alone ensembles of networks, for which BNs are commonly used (Kauffman, 1969, 1993; Shmulevich and Kauffman, 2004).

In this paper, we present a mathematical framework for calculating the relative stabilities of cell attractors and transitions, and hence deriving the notion of a landscape in BN models of development. We use a minimal GRN for pancreas development as an example to demonstrate the utility of our method. Imposing the observed relative ordering of attractors as a novel phenotypic constraint affords evaluation of ensembles of BNs (with a given network structure but different sets of Boolean functions) that are compatible with multiple observed attractors of the GRN. Our method can be used to reconstruct simple BN models for developmental processes from available information on GRN architecture and relative stability of attractor states, and thus, can predict the efforts associated with particular state transitions of interest which in turn can facilitate the rational protocol design for cell reprogramming in regenerative medicine.

2. Modeling framework

2.1. Boolean network (BN) model

BN model for a GRN is specified by its set of nodes, directed edges and Boolean functions. In a BN, the nodes represent genes while the edges represent interactions among genes in the network. Any gene i in a BN at a given time can be in one of two expression states: *on* if its state $x_i = 1$ and *off* if its state $x_i = 0$. For a m-gene BN, the state vector $\boldsymbol{X}^t = (x_1(t), x_2(t), \ldots, x_m(t))$ gives the expression of all genes at discrete time t in the network. For each gene i in a BN, a Boolean function F_i determines the output value x_i at time $t+1$ given the state of its input genes at time t. Thus, the gene expression state of a BN at any time step is governed by the recursive equation:

$$\boldsymbol{X}^{t+1} = \boldsymbol{F}\left(\boldsymbol{X}^t\right) \tag{1}$$

where $\boldsymbol{X}^t$ is a m-dimensional binary vector (0 or 1) that gives the expression of all genes at time step t. F encapsulates both the network topology and Boolean functions at all nodes, and thus, contains the information determining the dynamics of the BN.

For a m-gene BN, there are 2^m possible states. A sequence of states $\boldsymbol{X}^0, \ldots, \boldsymbol{X}^t, \boldsymbol{X}^{t+1}, \ldots$ forms a trajectory in the state space. Trajectories converge in a deterministic (noise-free) system. Since the state space is finite, the trajectories eventually converge either to a single state (point attractor) or a cycle of states (cyclic attractor). In the extreme case, a cyclic attractor encompasses all or almost all possible network states, and given the large number of states 2^m, such behavior will appear *chaotic*. For any given attractor, its associated basin of attraction is the set of initial states that will converge to that attractor. Attractors of a BN are charaterized by the size (and shape) of their associated basin of attraction. The network topology (i.e., the set of nodes and edges) and the Boolean functions at each node fully determine the *attractor structure*—which consists of attractors, trajectories and basins of attraction. The attractor structure can be determined by explicitly evaluating all state transitions $\boldsymbol{X}^1 = \boldsymbol{F}\left(\boldsymbol{X}^0\right)$ for all 2^m possible initial states $\boldsymbol{X}^0$. An example of a 4-gene GRN with Boolean functions and resulting attractor structure is shown in Fig. 1.

2.2. Transition matrix and BN dynamics

Spontaneous transitions between attractors that underlie the *epigenetic landscape* or the quasi-potential landscape requires probabilistic (noise-driven) dynamics. A discrete Markov model can be used to describe the BN dynamics. Let p_i give the probability for the occupation of a state s_i, and T_{ij} give the transition probability from state s_j to state s_i. In a BN with m genes, there are 2^m pos-