Definitions, Measures, and Models of Robustness in Gene Regulatory Network

Julio Monte[1], Min Liu[2], Adam Sheya[3], Toshimori Kitami[4]

[1]Department of Medicine, University of California at San Diego, La Jolla, CA, USA
[2]Department of Physics, University of Houston, Houston, TX, USA
[3]Department of Psychology, Indiana University, Bloomington, IN, USA
[4]Department of Genetics, Case Western Reserve University, Cleveland, OH, USA

Abstract

Robustness is a universal feature of biological systems which assures the survival of organisms during evolution. We studied robustness in the context of gene regulatory network models based on prior theoretical studies and its application to experimental studies. We defined robustness by specifying features and perturbations in the biological system, and identified several measures of robustness in Boolean gene regulatory network model. We studied the effect of scale-free distribution on the network dynamics and found that distribution parameter (γ) determines the dynamic phase of the system, as previously observed. Our definitions, measures, and models of robustness will allow us to identify additional design principles that provide robustness to biological systems and offer insight to the origins of cellular complexity.

1. Introduction

One of the defining and universal features of biological system is robustness against genetic mutations and environmental perturbations. In one of the most comprehensive studies to date, 80% of the genes in the yeast genome were found dispensable under favorable laboratory conditions and 65% of the genes were dispensable under non-optimal conditions (Giaever et al. 2002). High levels of robustness are pervasive across all species, assuring their survival during evolution. Although robustness of living organisms has been documented for over a century, design principles responsible for robustness of biological systems are only beginning to emerge (Stelling et al. 2004). These design principles will provide deeper insights into the mechanisms of evolution and origins of cellular complexity and will also provide frameworks for designing diagnosis and treatment of complex human diseases (Kitano 2004). Recent revolutions in genome technologies allow us to obtain part lists of biological systems as well as the interaction of those parts. Both pieces of information are crucial to understanding what give rise to robust biological systems (Lander 1996). Thus far, topology of biological networks, including protein-protein interactions (Jeong et al. 2001, Giot et al. 2003, Li et al. 2004, Han et al. 2004), metabolic pathways (Jeong et al. 2000, Fell and Wagner 2000), and gene regulatory networks (Guelzim et al. 2002), were found to be scale-free. These implied that these networks are robust to random attacks and that scale-free network is a

universal design principle conserved throughout biological systems (Barabasi and Oltavi 2004).

While network topology revealed tolerance against random attacks, analysis of network dynamics offers a new level of insight to the nature of robustness in biological system. Most of the work on the dynamics of biological networks focused on gene regulatory networks using theoretical models (Kauffman 1993). Recently, experimentalists developed tools to monitor dynamics of global gene expression activities (Brown and Botstein 1999) as well as regulatory interactions between genes (Harbison et al. 2004). These tools provide opportunities to entwine theoretical and experimental works on gene regulatory network. However, robustness in biology is often discussed without clear definitions. To provide foundations for applying theoretical network models on experimental data, we set out to define robustness and identify appropriate measures of robustness in gene regulatory networks. We focused mainly on the Boolean network models in which many previous theoretical studies exist. Then, using these definitions and measures, we tested the role of scale-free network on the dynamics of gene regulatory network. Our results agree with previous work by Aldana (2003) and offer new insight to the design principles of robust networks.

2. Definitions of Robustness

Robustness is a feature of complex systems that is often vaguely defined in biology. Erica Jen (2004) in her recent treatise on robust design proposed that one needs to specify both the feature and the perturbation of interest before discussing robustness. An important feature of gene regulatory networks is the ability to achieve functional diversity. In multicellular organisms, functional diversity includes cellular differentiation during development which gives rise to different tissue types and complex body plan. Another important feature of gene regulatory network is homeostasis. Once the cell finds its desired state, it needs to maintain the same cellular state under various stresses from the environment.

Perturbations in gene regulatory network include genetic, environmental, and intrinsic noise. Gene networks must tolerate genetic mutations in order for evolution to take place. Gene networks must also be resilient against changing environmental conditions such as changes in energy source and nutrients. Gene regulatory networks must also tolerate noise within the system itself. The process of gene regulation is a probabilistic process in which gene regulatory proteins within a cell must find their target genes amongst many proteins and sequences that crowd the nucleus. These features and perturbations are important for defining robust biological system in the context of gene regulatory networks.

3. Measures of Robustness

Now that we have defined robustness in gene regulatory networks, we can measure robustness in the network. Measures of robustness involve determining the behavior or feature (output) of a system as a function of perturbation (input) (Giaever et al. 2002). Mathematical and computational models provide mapping function from input to output space. In our project, we focused on Boolean network models and we will discuss input and output functions in the context of these models.

Boolean networks are generalization of Boolean cellular automata (CA) where the state of each node is affected by distant nodes in the network rather than by its neighbors. The original Boolean gene regulatory network was proposed by Stuart Kauffman (1969, 1993). He found that these networks exhibit dynamic behaviors which are classified into three phases: ordered, chaotic, and critical (edge of chaos). He found that in ordered and critical phases, gene regulatory network reached limited number of attractor states despite changes in the initial condition of the network. He compared these attractor states to different cell types in organisms and argued that these networks capture dynamics of cellular processes.

If we define attractors as cell types, we can measure robustness by creating specific perturbations (input) to the network model. We previously defined perturbations as genetic, environmental, and intrinsic noise. Genetic mutations could be represented by removal or addition of edges that connect genes in the network. Molecularly, edge removal is analogous to loss of transcription factor binding site on the target gene and edge addition represent creation of new regulatory binding site. Environmental perturbation and stochastic noise in the system can be represented by changing the state of a node. Molecularly, these are analogous to temporary inhibition or activation of gene activities by chemical inhibitors and activators in the environment or delay in the changes of gene activity state due to intrinsic noise.

After we induced perturbations, we can compare the dynamics (outputs) of the original network and the perturbed network and examine whether the networks reach the same number of attractors in the same time steps (functional diversity) or determine whether two networks remain in the same attractor that it started in (homeostasis). A more sophisticated analysis of network dynamics is Derrida's coefficient (Derrida and Pomeau 1986) which is equivalent to Lyapunov exponent in continuous dynamic system. These coefficients classify networks into one of three dynamic phases. However, this analysis requires a computationally intensive search of all possible network state configurations and was not feasible for our project.

4. Building Gene Regulatory Network Model

In this section, we describe the construction of scale-free Boolean network which we used to examine the effect of scale-free distribution on network dynamics. We first selected a set number of nodes in the network (N=20), defined as $\{\sigma_1, \sigma_2, \ldots, \sigma_{20}\}$. Each element takes on a value of 0 or 1 which corresponds to inactive and active gene states, respectively. Each element σ_i is controlled by k_i elements in the network where k_i is chosen randomly from the scale-free distribution $P(k) = Ck^{-\gamma}$ in which we selected γ = 1.5 and γ = 2.6. We could only define in-degree distribution of the network such that we assumed random (Poisson) distribution for out-degree edges.

Once we defined the network topology, we assigned rules that govern the dynamics of the system. To each σ_i, we assigned Boolean function, $f_i(\sigma_{i1}, \ldots, \sigma_{iki})$, such that each configuration of the controlling element has equal probability (p=0.5) of f_i=1 and f_i=0. In other words, in the entire rule table of network, half of the rules yield gene activation (f_i=1) and the other half yield gene inactivation (f_i=0). After rules were assigned, we updated the network synchronously with the dynamical equation: $\sigma_i(t+1) = f_i(\sigma_{i1}(t), \ldots, \sigma_{iki}(t))$. The configuration of the entire system at time t was defined as, $\Sigma_t = \{\sigma_1(t), \sigma_2(t), \ldots, \sigma_N(t)\}$ and at t+1 as, $\Sigma_{t+1} = F[\Sigma_t]$ where F was the set of Boolean function $f_i(\sigma_{i1}, \ldots, \sigma_{iki})$.

5. Network Dynamics

To examine the dynamics of our scale-free Boolean networks, we selected two different initial starting conditions for each network and performed synchronous updating of network as described in section 4. In our model, two initial starting conditions resemble transient perturbation to the system such as environmental stress or stochastic noise. As shown in Figure 1(a,b), we plotted the network state (upper panels) for two initial starting conditions and the differences in the network state using Hamming distance (bottom panels). For a Boolean network with scale-free distribution of γ = 2.6 (Figure 1a), networks with two initial starting conditions quickly converged after three time points and settled into periodic attractor of size two. The Hamming distance plot (bottom panel), which indicates the difference in the state of two networks, showed rapid decline during the first three time steps. These results indicated that at γ = 2.6, the scale-free network was in the ordered phase of the dynamics.

In contrast, in the Boolean network with scale-free distribution of γ = 1.5 (Figure 1b), we did not observe any overlap in the system state (top panel) over 100 time steps. We observed periodicity in the network state for each of the initial starting conditions but these attractor cycles did not converge. In other words, at γ = 1.5, same network with two initial conditions found separate periodic attractors that do not overlap with each other. The Hamming distance plot indicated that the difference in the system state oscillates unpredictably and ranges in the order of 0.2 to 0.6 but never converged to zero. These results indicated that at γ = 1.5, the network dynamics were in chaotic phase.

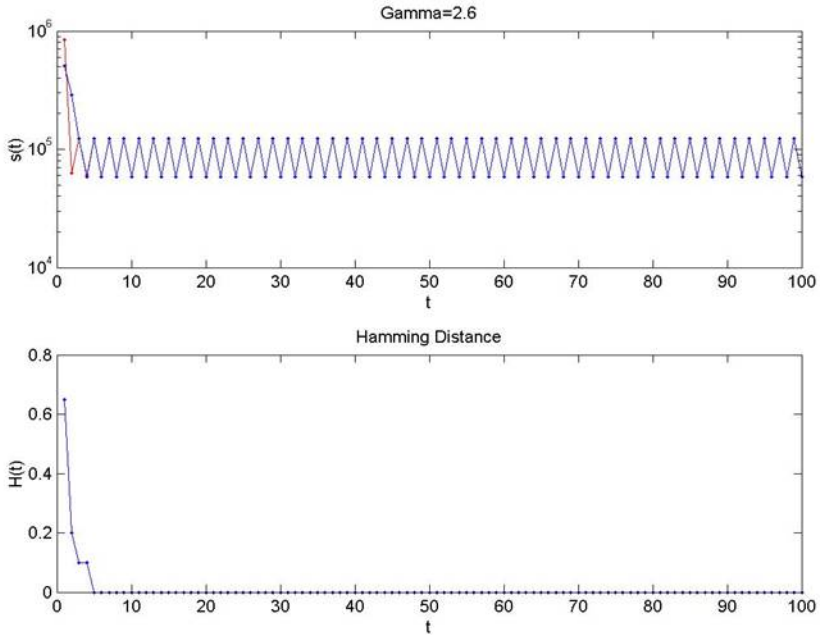
The above results agree with analytical estimate of the phase diagram across different ranges of γ in the scale-free Boolean network. Aldana (2003) showed that transition between ordered and chaotic phase occurs in the γ range of 2.0 to 2.5 with ordered phase at higher γ value. This range reflects many of the γ values found in various biological networks. Our two γ parameters were 1.5 and 2.6, corresponding to chaotic and ordered phase in the analytical work of Aldana (2003). His phase diagram of scale-free Boolean networks shows surprising contrast to phase diagram of random Boolean networks. In random Boolean networks, average connectivity of genes (k_i) must be constrained to less than three in order to reach ordered phase (Gershenson 2002). If the network contains higher number of average connectivity (k_i), the bias parameter for the rule table (p) must be small in order to achieve ordered phase. However, scale-free network eliminates such fine-tuning of parameters given that the exponent of scale-free network (γ) is beyond the threshold of 2.0 to 2.5. We were able to achieve ordered phase with p=0.5 given γ value above 2.5. Therefore, scale-free network contributes to robust network dynamics given appropriate distribution parameter (γ).

6. Conclusion & Future Direction

Robustness is a universal feature of biological systems important for understanding evolutionary processes, origin of cellular complexity, and the mechanisms of health and disease. We studied robustness in Boolean gene regulatory network based on its prior work in the field as well as its potential application to modeling experimental data. We described both features and perturbations of gene regulatory network in defining robustness and identified several approaches in measuring robustness. Based on previous observations that scale-free networks provide robustness on topological level, we studied whether scale-free networks provide robustness on the level of dynamics. We found that similar to Aldana (2003), scale-free networks give rise to robust network (ordered phase) given appropriate γ value. Scale-free networks eliminated requirement for fine-tuning of bias parameter (p) in order to achieve ordered phase.

Our work involved perturbations to the scale-free Boolean network by changing the initial starting condition of the network, analogous to environmental perturbation or stochastic noise in gene regulatory network. However, we could also create alternative perturbation by deleting the edges of the network, analogous to genetic mutations, and study the changes in dynamic phases of the system. Also, there are several applications of Boolean networks in constructing models of experimental data such as in gene regulation of Arabidopsis plant development (Espinosa-Soto et al. 2004) or comprehensive gene regulatory network of yeast (Kauffman et al. 2003). We could begin to apply some of our definitions, measures, and models of robustness to not only identify additional design principles that generate robust biological system but also apply these models to predict the outcome of specific genetic mutations and environmental perturbations in bioengineering and medicine.

(a)



(b)

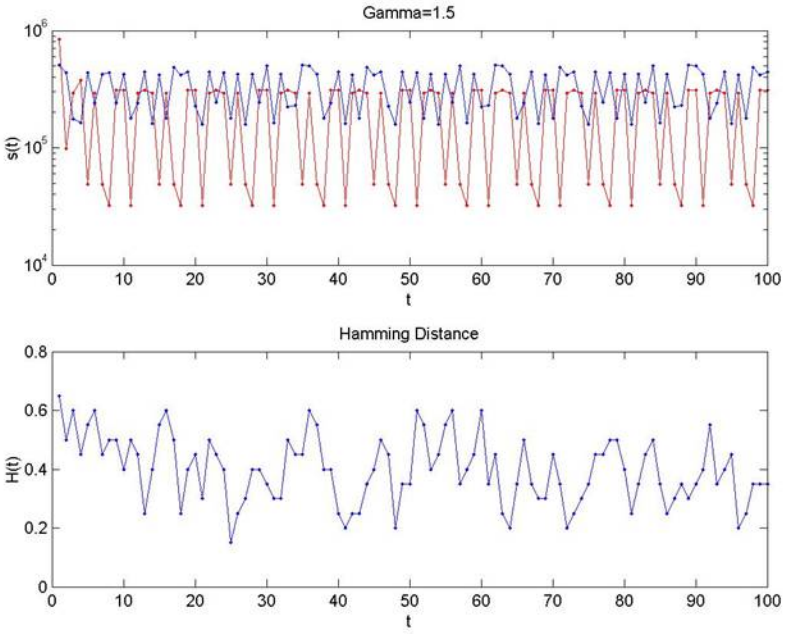


Figure 1. Plot of Boolean network states (upper panels) across time with two different initial starting conditions for different scale-free distributions, and the Hamming distance (bottom panels) between the networks with two different starting conditions. (a) $\gamma = 2.6$. (b) $\gamma = 1.5$.

7. Acknowledgments

We thank Drs. Melanie Mitchell, Tom Carter, Owen Densmore, Stephen Guerin, and Maximino Aldana for helpful discussions and comments on our project.

8. References

Barabasi, A.L., Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet*. **5**: 101-13.

Brown, P.O. and Botstein, D. (1999). Exploring the new world of the genome with DNA microarrays. *Nat Genet.* **21:** 33-37.

Derrida, B. and Pomeau, Y. (1986). Random networks of automata: A simple annealed approximation. *Europhys Lett.* **1**: 45-49.

Espinosa-Soto, C., Padilla-Longoria, P., and Alvarez-Buylla, E.R. (2004). A gene regulatory network model for cell-fate determination during *Arabidopsis thaliana* flower development that is robust and recovers experimental gene expression profiles. *Plant Cell.* **16**: 2923–2939.

Fell, D.A. and Wagner, A. (2000). The small world of metabolism. *Nat Biotech*. **18**: 1121–1122.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*. **418**: 387-391.

Giot, L., J.S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y.L. Hao, C.E. Ooi, B. Godwin, E. Vitols et al. (2003). A protein interaction map of *Drosophila melanogaster*. *Science.* **302:** 1727-1736.

Gershenson, C. (2002). Classification of random Boolean networks. In Standish, R.K., Bedau, M.A., and Abass, H.A., editors, *Artificial Life VIII: Proceedings of the Eight International Conference on Artificial Life*, 1-8. MIT Press.

Guelzim, N., Bottani, S., Bourgine, P., and Kepes, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet.* **31:** 60-63.

Han, J.D., Bertin, N., Hao, T., Goldberg, D.S., Berriz, G.F., et al. (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature.* **430:** 88-93.

Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., et al. (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature.* **431:** 99-104.

Jen, E. (2004). *Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies*. Oxford University Press.

Jeong, H., Tombor, B., Albert, R. , Oltvai, Z.N., and A.L. Barabasi. (2000). The large-scale organization of metabolic networks. *Nature.* **407:** 651-654.

Jeong, H., Mason, S.P., Barabasi, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. *Nature.* **411:** 41-42.

Kauffman, S. A. (1969). Metabolic stability and epigenesist in randomly constructed genetic nets. *J Theor Biol.* **22**: 437–467.

Kauffman, S.A. (1993). *The Origins of Order*. Oxford University Press.

Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. (2003). Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci, USA*. **100**: 14796–14799

Kitano H. (2004). Cancer as a robust system: implications for anticancer therapy. *Nat Rev Cancer*. **4**: 227-35.

Lander, E.S. (1996). The new genomics: global views of biology. *Science* **274:** 536-539.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., et al. (2004). A map of the interactome network of the metazoan *C. elegans*. *Science* **303:** 540-543.

Stelling, J., Sauer, U., Szallasi, Z., Doyle, F., and Doyle, J. (2004). Robustness of cellular functions. *Cell*. **118**: 675-685.