Effects of an RNA control layer on the state space of Boolean models of genetic regulatory networks

Jennifer S. Hallinan, Daniel R. Bradley, John S. Mattick and Janet Wiles

Abstract—The general assumption in biology is that most genes encode proteins. However, it is now evident that much of the genome of humans and other complex organisms is transcribed into non-protein-coding RNAs. Some of these RNAs are processed into small regulatory RNAs, such as microRNAs, that control many aspects of animal and plant development. It has been suggested that regulatory RNAs represent an additional control layer that was critical to the emergence of complex organisms. We examine this possibility using a model of cell differentiation based on attractors in boolean networks. Our simulation studies show that an additional layer of RNA control modeled as fast temporal links can significantly increase the number of attractors in boolean models of genetic regulatory networks (analogous to the number of cell types in a complex organism). However, it also has the power to simplify the state space structure. We explore the conditions under which these different outcomes occur.

I. INTRODUCTION

One measure of organismal complexity is the number of distinct cell types that exist within an organism. Genetic regulation guides the division and differentiation of cells from progenitor cells (stem cells) into mature cells, as well as temporal and spatial movement within the organism. Until recently it has been an accepted tenant of gene regulation that regulatory proteins such as transcription factors controlled gene regulation within the cell and that each gene coded for a single protein product.

However as newer technologies allowed mRNA to be retrieved and sequenced it was found that there is often more than one mRNA transcript per gene. Regulatory mechanisms must be able to control not only when a protein is produced, but also which isoform a specific gene produces in different contexts. Gene regulation has turned out to be more complex than previously thought.

With the advent of genomics and transcriptomics it has become apparent that much more of the genome is transcribed than was previously believed (Mattick, 2003; Frith, Pheasant & Mattick, 2005). At least some of these RNAs are processed to form small regulatory RNAs, termed microRNAs, that are known to control many aspects of animal and plant development, and to be perturbed in cancer (see Bartel, 2004; Mattick & Makunin, 2005).

In a recent paper, Gagen and Mattick (2005) argued that regulation in all integrated complex systems scales more than linearly with function, and that the complexity of prokaryotic organisms has been limited by species reaching the maximum genome size that can effectively be regulated by protein-based regulatory controls. Conversely they suggest that eukaryotes developed an additional RNA-based control layer that enabled their single-celled ancestors to break through this complexity barrier and allowed the evolution of more complex multicellular organisms (Mattick & Gagen, 2001).

How best to study the computational potential of an RNA-based control layer is an open, and clearly critical, question. Such a study requires first translating the question into measurable global properties of genetic regulatory networks. In this project we utilize the decades of experience that has been gained with boolean network models of genetic regulation to explore these issues.

For well over half a century, biologists such as Sewell Wright (1932) and Waddington (1940) have put forward and developed the view of cell development as a trajectory through a landscape, with different cell types taking alternate paths and reaching different end points. The same set of genes controls all the cells in a developing embryo, but each cell type takes its own trajectory through the gene expression space.

This idea was formalized in boolean network models of genetic regulation, in which the long term behavior of a cell (a cell type) is analogous to an attractor in a network's state space. Boolean networks show many interesting properties from a complex systems perspective (Kauffman, 1969; 1993; Gershenson, 2002; Wuensche, 1998). The one that is of particular interest for this paper is that the number of attractors has been equated with the number of cell types (for a comprehensive review of this idea see Kaufman 1993).

The hypothesis that underpins this study is as follows:

Adding an RNA-based control layer to a genetic regulatory network increases the complexity of the state space and hence should increase the number of attractors (cell types).

There are many ways in which RNA-based control could be incorporated into a genetic regulatory network model. The basic assumption for this study is that one major advantage
of RNA-based control over protein-based regulation is speed: In the time that a gene is transcribed and translated into a protein, multiple microRNA signals can be transcribed and sent to targets affecting regulation. Because microRNA does not require translation and the resulting control signals do not need to be reimported into the nucleus to affect transcription, microRNAs are likely to have considerable advantages in their response times for gene regulation (see Figure 1). This leads to the prediction is that one of the key functional attributes of RNA-based control is the rapid time course of the regulatory effects.

In this project we formalize the idea of integrating an RNA control layer into a Boolean network by incorporating delay steps in a Boolean network model. In a series of simulation studies we measured the dynamic properties of the networks for a range of parameters. It should be noted that RNA-based control may also have other advantages over the use of regulatory proteins beyond those we consider here, including the efficient transmission of regulatory information as short sequence-specific strings (like microRNAs) that may be recognized and decoded by a common protein infrastructure. The scope of this project is to focus on understanding the relationship between the time course of regulation and state space structure.

II. MATERIALS AND METHODS

In our CBNs, each node of a network represents a transcription factor gene, while each edge represents the relationship between the product of that gene and the gene(s) that it activates or inhibits.

A variety of simulation strategies have been proposed to model synchronous and asynchronous node updates. In the studies reported here synchronous models were used. One update cycle is analogous to the time for a gene to be transcribed, translated and its protein product(s) to be transported back into the nucleus to regulate other genes.

Each node of the network maintains a boolean state that reflects whether it is active or inactive – in biological terms, whether it is being transcribed or not. Some nodes are defined to be constitutive and are activated unless specifically inhibited. The state of the network is updated at each time step. The next state is produced from the current state using the following rules:

1. A node in the next state will be inactive if a currently active node inhibits it.
2. A node that is not inhibited will be activated either if it is constitutive, or if one or more currently active nodes activates it.
3. If a node is not constitutive nor activated it will become inactive.
4. During RNA-only update steps, if a node is not
inhibited, constitutive nor activated it remains in its previous state. Each update step represents transcription factors being expressed and affecting their targets.

A. The constitutive attractor

Networks such as CBNs which have one or more constitutively-active nodes have a property that distinguishes them from most networks used in artificial GRN studies to date. In standard RBNs, no state of the system is in principle different from any other state. However, with low connectivity, if few or no nodes are active, a basin of attraction can form around the null point attractor (i.e. no active nodes). By contrast, in networks with constitutively active nodes, if all nodes are simultaneously inhibited, the subsequent state will be one in which all the constitutive nodes (and only those nodes) are active and it will lead to a non-null attractor. To distinguish the state in which only the constitutive nodes are active, we use the term constitutive state, and the subsequent dynamics follow the constitutive transient to the constitutive attractor. Constitutive nodes ensure that the null state cannot be a point attractor in a network regardless of its connectivity.

The constitutive attractor is likely to be important in (real) genetic regulatory systems.

B. Adding an RNA-control layer

To model a simple RNA control layer we added to the network model a new type of edge, which represented either an inhibitory or excitatory RNA link. Each update step in the original CBNs was replaced with an update cycle and different update rules were associated with the different types of links. On the first step of each cycle, every edge is used to update the next step (universal update step). Then for the following n steps only RNA links are used to update the nodes (RNA update steps). For n=0 the CBN is identical to our previous synchronous deterministic simulations. For non-zero n, in the biological analogy, the network changes from protein based regulation to protein+RNA regulation. The benefit of this design is that the addition of RNA links maintains the same network size, in terms of total number of nodes and number of links, while allowing for varying numbers of RNA update steps per protein update.

C. Transients and attractors in variable-time networks

In CBNs with RNA update steps, a state may lead to different states depending on whether the next update is a universal or an RNA update step. Thus a state may be encountered twice and yet lead to different attractors.

Since the aim of the study is to test whether the addition of non-zero RNA steps changes the complexity of the state space, we only test the states at the universal update step. This corresponds to the states of the protein network of a system. To detect these types of attractor it is necessary to test for reencountered states only during universal update steps, and then only from the set of states produced during previous universal update steps.

D. Network architecture

To enable characterization of the entire state spaces of each network, the size of the networks was restricted. The fixed parameters were the number of nodes, n = 10, with five constitutively active; and average out-degree of 2.0 (comprised of an average protein out-degree of 1.0 and average RNA out-degree also of 1.0). The variable parameters were the proportions of inhibitory protein and RNA links, which ranged from 0.0 to 1.0 with increments of 0.1 (i.e. 0.0, 0.1, 0.2 ... 1.0). Each network was generated probabilistically.

Ten networks were generated from each set of parameters using different random seeds. Hence 11 * 11 * 10 = 1210 unique networks were generated. These were used in pilot studies to determine the effects of parameter sizes. From this set of networks, we selected a subset of networks with both protein and RNA inhibitory links equal to 0.6.

E. RNA update steps and measures

The study used a 2x3x5 design, varying the number of protein links (P1, P2), number of RNA links (E0, E1, E2), and RNA update steps (0-4). For each set of parameters, 100 networks were generated from random seeds (3000 networks in total). For each network we recorded the length of the constitutive attractor, the length of the transient to the constitutive attractor, the number of attractors, the period of the longest attractor and the longest transient to an attractor.

III. RESULTS AND DISCUSSION

Adding RNA links (E) to networks with a fixed number of protein links (P) showed a significant increase in the number of attractors (see Figure 2 first three columns). The control condition (P1E0, step 0) showed the fewest number of attractors (1.27). This low value is consistent with expectations for networks with low total connectivity (average out-degree of 1.0). Even before considering the effects of non-zero RNA steps, it is clear that adding RNA links increases the complexity of the space (P1E1 1.71 attractors, P1E2 2.07, step 0). This result is consistent with current understanding of Boolean network models, since the effective total out-degree is increasing with the addition of RNA links.

The interesting cases occur as the number of RNA steps is increased, since the number of links is effectively held constant. Here, the networks appear to show two different types of behaviors, depending on the number of RNA links. For an RNA step-size of 1, P1E1 had a small increase in the number of attractors (from 1.71 to 1.84), whereas P1E2 had a much larger increase (from 2.07 to 2.66). As the step-size increased further, the P1E1 network showed a general decrease in average number of attractors, whereas the P1E2 network maintained the number of attractors at the same or higher level than for step-size zero. The length of the constitutive attractor also decreased as the RNA step-size increased (results measured but not shown).

A plausible explanation for the different trends in these results is that in networks with fewer RNA links, the RNA
steps cause the network to converge between universal update cycles, whereas the networks with higher numbers of links maintain the complexity of the space. The direction of convergence is towards the constitutive state.

How general are these results likely to be? We repeated the simulation with networks with higher protein out-degree (average 2.0, 3.0 and 4.0) and recorded the same measures (see Figure 3 for P2E1 and P2E2). The same trends were observed, in that low RNA out-degree reduced the complexity of the space and higher RNA out-degree maintained or increased the complexity as measured by the number of attractors.

These findings support the hypothesis that adding an RNA-based control layer to a genetic regulatory network can increase the size and complexity of the state space and the number of attractors. But the situation is more subtle than just more attractors.

The first part of the hypothesis is true by virtue of the network update function: The RNA-based control layer (implemented as described in this study) increases the complexity of the state space by increasing the number of states that can succeed a given state, depending on whether the update is a universal update or sequence of RNA updates.

The second part of the hypothesis, that the increase in complexity of the state space results in increased numbers of attractors, is supported by the simulation results for RNA-out-degree of 2.0 and step-size of 1. But for low RNA out-degrees of 1.0, the state space is actually simplified (indicated by fewer attractors).

The findings show that RNA has a range of effects on a network, depending on the out-degree of the RNA links. These results suggest that the addition of microRNA links may have a range of control powers, which can be tuned depending on the complexity of the underlying protein networks. The networks in this study all had relatively low connectivity, which was designed to maintain stable behavior. If the out-degree of the protein links is increased, it is well-understood from the Boolean networks literature that the networks enter the chaotic regime. An interesting conjecture from the current study is the roles that RNA can play in such networks. In current simulations we are testing the power of RNA links to stabilize behavior in otherwise chaotic networks.

Figure 2. Attractors for varying numbers of RNA links and RNA step sizes. The key shows the average number of protein and RNA links: P1E0 means 1.0 protein and no RNA links; P1E1 – 1.0 protein and 1.0 RNA links; P1E2 – 1.0 protein and 2.0 RNA links on average. See text for discussion.

IV. ACKNOWLEDGEMENTS

This research was funded by the ARC Centre in Bioinformatics (CE0348221) and an ARC grant to JH, JM and JW.

REFERENCES