

Structure and dynamics of a gene network model incorporating small RNAs

Nicholas Geard

School of ITEE
The University of Queensland
Queensland 4072, Australia
nic@itee.uq.edu.au

Janet Wiles

School of ITEE, School of Psychology
The University of Queensland
Queensland 4072, Australia
j.wiles@itee.uq.edu.au

Abstract-

As advances in molecular biology continue to reveal additional layers of complexity in gene regulation, computational models need to incorporate additional features to explore the implications of new theories and hypotheses. It has recently been suggested that eukaryotic organisms owe their phenotypic complexity and diversity to the exploitation of small RNAs as signalling molecules. Previous models of genetic systems are, for several reasons, inadequate to investigate this theory.

In this study, we present an Artificial Genome model of genetic regulatory networks based upon previous work by Torsten Reil, and demonstrate how this model generates networks with biologically plausible structural and dynamic properties. We also extend the model to explore the implications of incorporating regulation by small RNA molecules in a gene network. We demonstrate how, using these signals, highly connected networks can display dynamics that are more stable than expected given their level of connectivity.

1 Introduction

One of the most surprising results to emerge from the completion of the Human Genome Project was that the number of genes in the genome was far lower than initially expected [14]. The 35,000 or so genes that are estimated to be encoded by the human genome is not significantly greater than the number of genes found in organisms considered to be less “complex”, such as the nematode worm *C. elegans* (around 20,000 genes) and the fruit fly *D. melanogaster* (around 13,500 genes). These figures suggest that the greater complexity of the higher eukaryotes is not due to additional genes, as generally thought, but rather to the complexity of the regulatory interactions controlling their expression.

However, complex genetic regulation has its own costs. It is well known that higher levels of connectivity tend to push network behaviour towards chaos [23]. Furthermore, it has been shown that as the number of genes to be regulated increases, the number of regulatory genes required to control their expression grows quadratically [9]. One method by which higher eukaryotes may have bypassed this complexity limitation is by exploiting small RNA molecules as regulatory signals [28, 29].

Previous models of genetic regulatory networks (GRNs) have generally operated under the assumption that a single mechanism, based on regulatory proteins, is responsible for

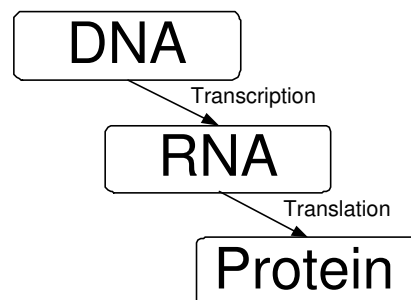


Figure 1: The Central Dogma. Information is transmitted from DNA to protein through RNA.

network connectivity. The possibility of multiple control systems, each with fundamentally different signal costs and speeds, has not been considered. The purpose of this paper is twofold: firstly, to present our GRN model and describe its relevance to the biological systems we aim to simulate; secondly, to demonstrate the ability of this model to increase the range of phenomena that can be simulated by using it to model the role that small RNAs may play in gene regulation.

Initially, some background is provided on the traditional view of gene regulation in biology and the way in which this is reflected in current models. Recent discoveries concerning the role of small RNA molecules that are challenging this view are then described. A new model is proposed that generates networks with multiple levels of regulatory control. Initial results illustrating the structural and dynamic properties of this model are then presented and discussed. Finally the future directions of this model with respect to studies of evolvability are discussed.

2 Gene Regulation

The most simplistic view of gene regulation, and the current dominant paradigm, is that expressed by the “central dogma” of molecular biology, first expressed in 1957, that information flows from DNA through RNA to protein (see figure 1). Implicit in this view is the idea of a unique mapping from gene to protein in which RNA plays only a mediatory role.

The operon model of gene regulation, developed by Jacob and Monod in 1961 [19], refined the central dogma by proposing a distinction between two different types of gene. One type, “structural” genes, encode the proteins that play some functional role in the metabolism of a cell. The other type, “regulatory” genes, encode proteins that act as tran-

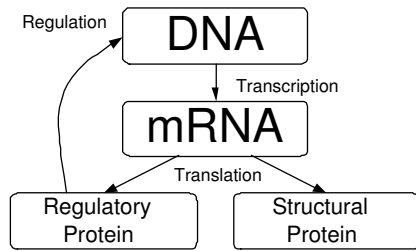


Figure 2: The Operon Model. DNA encodes two classes of proteins. Structural proteins play a functional role in the cell’s metabolism. Regulatory proteins interact with DNA to control the rate of transcription of other genes.

scription factors (TFs), controlling the rate of transcription of other genes without otherwise affecting the proteins they encode (see figure 2).

Early complex systems models of gene regulation were based on this idea of a simple, protein-based regulatory logic. In particular, the Random Boolean Network (RBN) model developed by Kauffman was designed to study the global patterns of behaviour exhibited by networks of interacting genes [21, 22]. RBNs exhibit many properties common to genetic systems, including complex periodic behaviour, self-organisation and robustness [2, 23].

3 A Role for RNA

Since the 1960s, advances in molecular biology have revealed a more complicated picture of gene regulation, particularly in eukaryotic organisms. Initially, it was found that there was considerably more genetic information in the genome than was required to specify the protein-coding genes. Separating these genes were long sequences of DNA whose purpose was unknown. Towards the end of the 1970s, it was also discovered that the DNA coding for an individual protein was not necessarily arranged in a continuous sequence. Most eukaryotic genes consist of coding segments (exons) broken up by long noncoding segments (introns) that are removed prior to translation (see figure 3). Exons are frequently joined together in a variety of different combinations, permitting several different proteins to be produced from a single DNA sequence. This alternative splicing mechanism is likely to contribute significantly to the greater complexity of eukaryotes [37].

A widespread early view of these intronic and intergenic sequences, which constitute over 90% of the human genome, was that they were “junk” DNA. Explanations for the origin of these DNA sequences included that they are the non-functional remnants of gene duplication events or the result of parasitic DNA sequences, conferring no advantage to the genome as a whole [40]. Another suggestion was that these noncoding regions act as separating regions to allow “exon-shuffling” – rapid exploration of evolutionary space via the recombination of functional building blocks [16].

Subsequent investigations revealed that damage to these intronic and intergenic regions frequently results in developmental defects [31], suggesting that the specific content

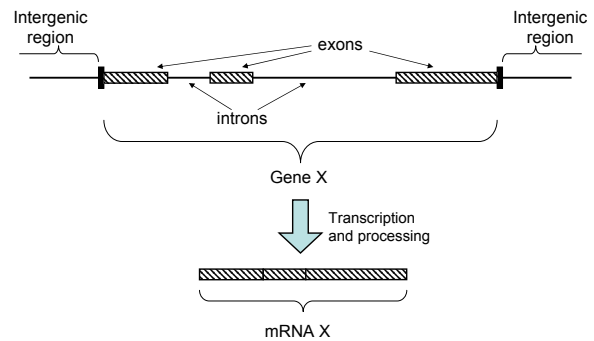


Figure 3: The Intron/Exon Structure of Genes. Genes are generally separated on the chromosome by long, non-coding intergenic regions. A single gene consists of both protein-coding exons and non-coding introns, which are removed before an mRNA is translated into a protein.

of these “junk”, or noncoding, sequences is significant. It is now known that many of these noncoding regions are transcribed into RNA, and that these RNA molecules, although not translated into proteins, play an important functional role [3].

In the last few years, many more studies have provided evidence for the important role that RNA plays in the regulation and control of genetic events (see [8] for a recent review). A large quantity of RNA is transcribed from the genome, of which only a small fraction is messenger RNA molecules (mRNAs) that will be translated into proteins. Many other noncoding RNAs (ncRNAs) are also transcribed, ranging in size over several orders of magnitude from several kilobases down to around 20 nucleotides. These ncRNAs appear to play a role in a number of genetic processes, including regulation of transcription and translation, the modification of chromatin structure, and the processing, modification and stability of RNA and protein molecules [11] (see figure 4).

As noncoding regions and introns are characteristic features distinguishing eukaryotic genomes from those of the simple, more compact genomes of prokaryotes, it has been proposed that regulatory RNA may form the basis of a parallel control system [27]. This more sophisticated level of regulatory control may have been one of the key features that enabled the evolution of multicellularity and the subsequent increase in phenotypic complexity and diversity found in eukaryotic organisms. This idea has been further explored and developed into the idea that information processing by small RNA molecules (sRNAs) may provide a “meta”-level of regulation allowing for the evolution of new and complex functions by modulating the control architecture of an otherwise stable core proteome [28, 29]. To date, no computational models of GRNs exist that explicitly incorporate multiple levels of regulatory control.

4 Modeling Framework

Current complex systems approaches to modeling GRNs are generally based on the outdated “central dogma” view

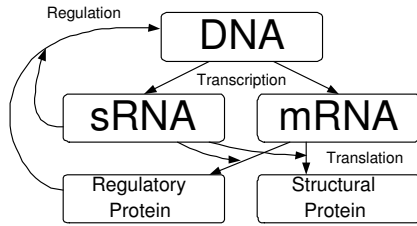


Figure 4: Small RNA Regulation. Produced in parallel with mRNA, sRNA molecules appear to play a number of roles involved in the regulation of transcription and translation events.

of gene regulation described above, and rarely incorporate such features as alternative splicing, post-transcriptional processing and RNA-mediated regulation. More significantly, most models make no distinction between regulatory and structural genes and proteins. In Kauffman’s networks, for example, all nodes are regulatory and no environmental input/output or system functionality is modeled. While many insights have been gained from such environmentally isolated models, crucial issues relating to the functionality and control of complex behaviour call for the ability to embed a system in a dynamic environment.

One of the most important contributions that systems level models can make is to provide mappings from biological processes to computational analogues that reveal new insights into the way systems are structured and controlled. The mapping from a DNA sequence to a computational string is a widely accepted analogy, as is the mapping from genetic and metabolic systems to networks. What is less frequently considered, however, is the process by which information encoded in a string is transformed into a network. One field in which this mapping has been investigated is Artificial Life, where developmental models have been used to increase the potential phenotypic complexity that can be encoded in a genotype (e.g., [10, 12, 34]). The model that we have chosen as our starting point, the Artificial Genome (AG) [33], uses a DNA-inspired representation for its genotype (see figure 5).

The AG model provides a simple yet elegant means of generating regulatory networks that addresses several criticisms levelled at previous models, including their abstraction from biology and limited extensibility [13]. In contrast to the random generation of networks used by many existing models, the structure and function of AG networks are extracted from an underlying sequence, analogous to a biological genome. Utilising a sequence level of representation provides an intuitive link to the biological emergence of regulatory networks from DNA sequences. It also permits an increased level of control over individual facets of gene expression, such as transcription, translation and regulatory binding. We have previously demonstrated how a sequence-based model closely related to that described here can be used to explore the effects of more biologically plausible mutation operators [39].

The AG model as initially presented by Reil [33] possesses many interesting features, such as point and cyclic

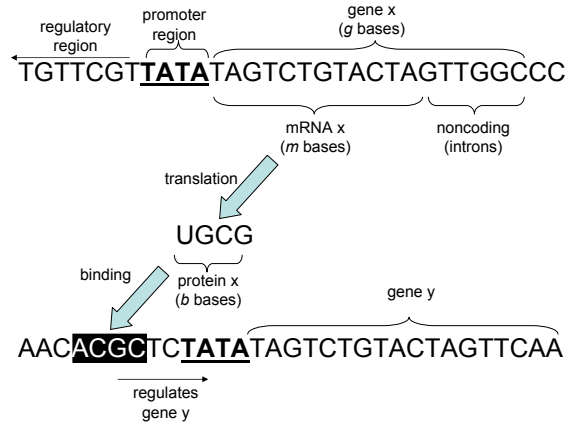


Figure 5: Artificial Genome Model - Gene Structure: A gene is identified by a promoter sequence (TATA) and consists of the following g bases. The first m bases specify the mRNA sequence, the remainder is non-coding. An mRNA sequence is translated into a protein binding sequence of length b using an artificial genetic code (see figure 6). A protein sequence regulates a gene by binding with a matching sequence in its regulatory region.

attractors that exhibit robustness to perturbation and sensitivity to initial conditions. However, some of the rules used to define the network structure and behaviour in the model are arbitrary and amenable to a greater level of biological justification. A more biologically grounded version of the AG model is presented here. In particular, the process of gene expression, modeled as a single step in the AG model, is expanded to explicitly include DNA transcription and RNA processing and translation. The modified model is used to generate genetic networks with both protein and RNA-mediated regulatory interactions.

Genome: At the heart of the model is the artificial genome itself, a sequence of bases of length ℓ . We have chosen to use four bases (A, C, T and G) to match those present in real DNA sequences.

Transcription: The beginning of a transcription unit is defined by the presence of a standard promoter region (an alternating sequence of T’s and A’s of length p), analogous to the “TATA box” that indicates an RNA polymerase binding site in eukaryotic genomes. The primary transcript (gene) is defined as the g bases following the promoter region (see figure 5).

Processing: In biological systems, the primary transcript is subject to several processing steps that result in the production of a protein-coding mRNA. In our model, the mRNA sequence is defined to be the first m bases of the primary transcript.

Translation: In a biological system, the mRNA sequence is translated into an amino acid chain that is then folded into a protein. A subset of these amino acids define the binding domain of the protein that interacts with DNA to control gene activity. In our model, the artificial genetic code is used to translate an mRNA into a sequence of bases that specify the regulatory motif recognised by the protein’s

		2 nd			
		A	C	T	G
1 st	A	A	C	U	G
	C	U	G	A	C
	T	G	A	C	U
	G	C	U	G	A

Figure 6: The Artificial Genetic Code: The above code is used to translate mRNA sequences into protein binding sequences. The code has been designed such that mutations to the third base are always neutral, while mutations to the first and second bases will always alter the binding sequence.

DNA-binding domain. The mRNA sequence is translated into a binding domain of length b according to an artificial genetic code. The mRNA sequence is read in triplets, in which the first and second bases specify the target and the third base is redundant (see figure 6).

GRN Structure: A subset of the protein products encode the functional outputs of the network. The remainder are TFs that regulate other genes. Targets for each TF are found by searching for subsequences matching the TF binding sequence in the regulatory region of each gene, which extends from the end of the preceding gene to the beginning of the current gene. As any one TF can act as either a repressor or an activator depending on context, the binding affinity of a particular TF is determined by the three bases preceding its binding site. The strength of the binding affinity is calculated by converting the base-4 sequence to a real-valued weight in the range $[-w, w]$. A positive weight indicates activation and a negative weight, inhibition.

GRN Dynamics: In this simple model, time is measured in discrete units and the activation state of each gene is Boolean (i.e., either on or off). The input to a node at each time step is determined by summing the inputs (weighted by the strength of the binding affinity) from each TF that regulates it:

$$net_i(t) = \sum_j w_{ij} a_j(t-1) \quad (1)$$

where $a_j(t-1)$ is the activation state of gene j at the previous time step and w_{ij} is the strength of the regulatory interaction between genes i and j . The current activation state of a node is determined by thresholding the net input:

$$a_i(t) = \begin{cases} 1 & \text{if } net_i(t) \geq 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

5 Structure

5.1 Differences between Artificial Genome networks and random networks

The first issue we address is how networks generated using the AG framework differ from randomly generated networks (e.g., the random Boolean networks developed by Kauffman [23]).

The number of genes, N , in AG networks scales linearly with the length of the genome, ℓ :

$$N = \ell/4^p \quad (3)$$

where p is the length of the promoter sequence. Average network connectivity in AG networks also scales linearly:

$$K = \ell/4^b \quad (4)$$

where b is the length of the protein binding sequence. For large systems, this relation results in an average connectivity considerably higher than has previously been argued to occur in biological systems (e.g., in [23, 38], where it is suggested that an average connectivity of around two is likely). However, there is a growing opinion that actual connectivity in gene networks may be denser rather than sparser. The combinatorial nature of gene regulation in eukaryotes is proving to be more complex than initially suspected [26, 36]. In particular, factors such as competition for metabolic resources and transcription factors [6], the diverse roles of co-regulators and other “accessory” proteins in chromatin remodeling and DNA binding modulation [26] and the structure of network architecture [7] may greatly increase the number of input signals that affect transcription of a given gene.

The most significant difference between AG networks and randomly constructed networks is in the distribution of input and output connections (K_{in} and K_{out} respectively). Randomly constructed networks are usually generated with a fixed number of inputs per node and randomly assigned outputs, resulting in a Poisson output distribution. For networks generated using the AG framework, K_{in} follows an exponential distribution and K_{out} follows a Poisson distribution.

Analyses of regulatory networks in biology have revealed an exponential distribution of both inputs and outputs [20]. Computational studies have suggested that this “scale free” connectivity confers several advantages, including increased stability at higher levels of connectivity [1, 15, 30]. Clearly, both random networks and the AG model are failing to capture an important aspect of GRN structure.

The biologically unrealistic K_{out} distribution generated by the AG model results from the assumption of fixed-length binding sequences. It is possible to rectify this anomaly by modifying the model to allow the length of genes, and their resulting regulatory products, to vary by defining a “gene end” sequence (three consecutive ‘A’s) similar to the way in which a promoter region is used to identify gene start sites. Due to the relatively short genome

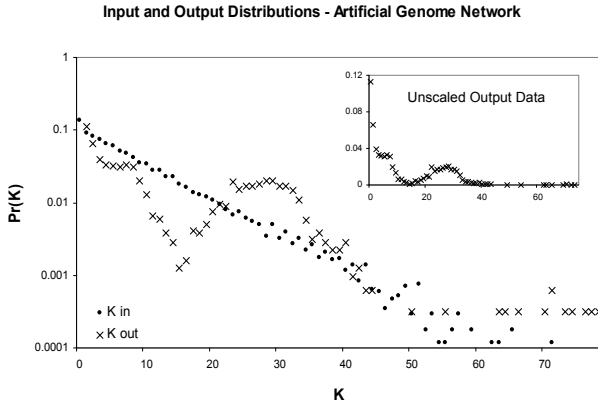


Figure 7: Input and Output Distributions for the Artificial Genome Network (Genome length = 500,000; promoter length = 4; gene-end length = 3; data points were averaged over 20 genomes). Both input and output distributions show an exponential trend. Inset: Unscaled data for the output distribution (see text for details).

lengths used, regulatory products fell into a small number of discrete classes according to size. As before, the output distribution of each of these classes follows a Poisson distribution. However, when all of these distributions are superimposed, the overall trend is exponential (see figure 7). It is also likely that evolution plays a role in determining the distribution of K_{out} , and studies are planned to investigate the plausibility of modelling these processes.

5.2 The effect of adding sRNAs

The next issue we address with regard to network structure is the effect of adding a second class of signalling molecules (i.e., sRNA molecules) to the model. The additional regulatory links were defined by modifying the processing stage described in Section 4 to include the production of both a mRNA sequence of length m as well as an sRNA sequence of length s . Unlike mRNA sequences, which are translated into proteins prior to binding, sRNA sequences bind in their untranslated state. As would be expected, the primary effect on network structure is to increase the level of connectivity that may be obtained from a genome of given length:

$$K = \ell/4^b + \ell/4^s \quad (5)$$

One of the key features of small RNA molecules is that they require much less DNA to encode than proteins. If an AG network regulated only by protein signals were required to achieve the same level of connectivity as a network regulated by both protein and sRNA signals, it would require G to be increased by a factor of $1 + 4^{(b-s)}$. In situations where high connectivity is advantageous but there is some cost associated with genome size, the potential for a single gene to provide multiple different outputs already provides some reduction in these costs for a regulatory system using sRNAs.

5.3 The evolvability of sRNA-regulated networks

From an evolvability perspective, we hypothesize that systems regulated by sRNA molecules as well as proteins are likely to have an advantage over those regulated by proteins alone.

The shape of the search landscape is likely to differ considerably between protein-mediated regulation and sRNA-mediated regulation. In biological systems, several levels of indirection exist between a nucleotide sequence and the functional product it encodes. Not only is the primary sequence substantially edited and translated into amino acids, but the function of a protein molecule depends largely on the three-dimensional shape into which these amino acids fold. Protein folding is an extremely complex problem, containing a high degree of neutrality, and is not yet fully understood. A complex, nonlinear mapping between sequence-level mutations and changes to protein function will result in rugged and difficult to search landscapes. On the other hand, sRNA molecules are untranslated and frequently interact with DNA and RNA targets via template matching [17]. In these cases, there will be a simpler, one-to-one mapping between mutations to the coding sequence and changes to the signalling function. The search landscape will be therefore be smoother and more easily searched. We are currently investigating the evolvability of various classes of AG networks.

6 Dynamics

6.1 The effect of sRNAs on system stability

As previously described by Reil [33], the AG model, as it stands, can produce a wide range of dynamics, from single-point and cyclic attractors to “chaotic” behaviour, depending upon the parameters used and the resulting level of connectivity. More generally, it has been established that the phase space of dynamic network models can be divided into disordered and ordered regions depending on the level of connectivity and the bias determining whether a given node is active or not [24]. It has been hypothesised that the conditions necessary for interesting behaviour are likely to occur at the phase transition between these two regimes (the so-called “edge of chaos”) [25].

One effect of incorporating sRNA interactions into a network with low connectivity is to shift its dynamics from the ordered region of phase space toward the ordered/disordered boundary, simply due to the increase in connectivity. When both sRNA and protein signals operate with the same time constant, network behaviour was found to be more chaotic.

6.2 The behaviour of networks with multiple rates of interaction

A more interesting possibility is to consider the role that sRNAs may play in stabilizing the behaviour of highly connected, chaotic networks when they operate at an increased rate with respect to proteins. Several different approaches, such as artificial evolution [5], have previously been used

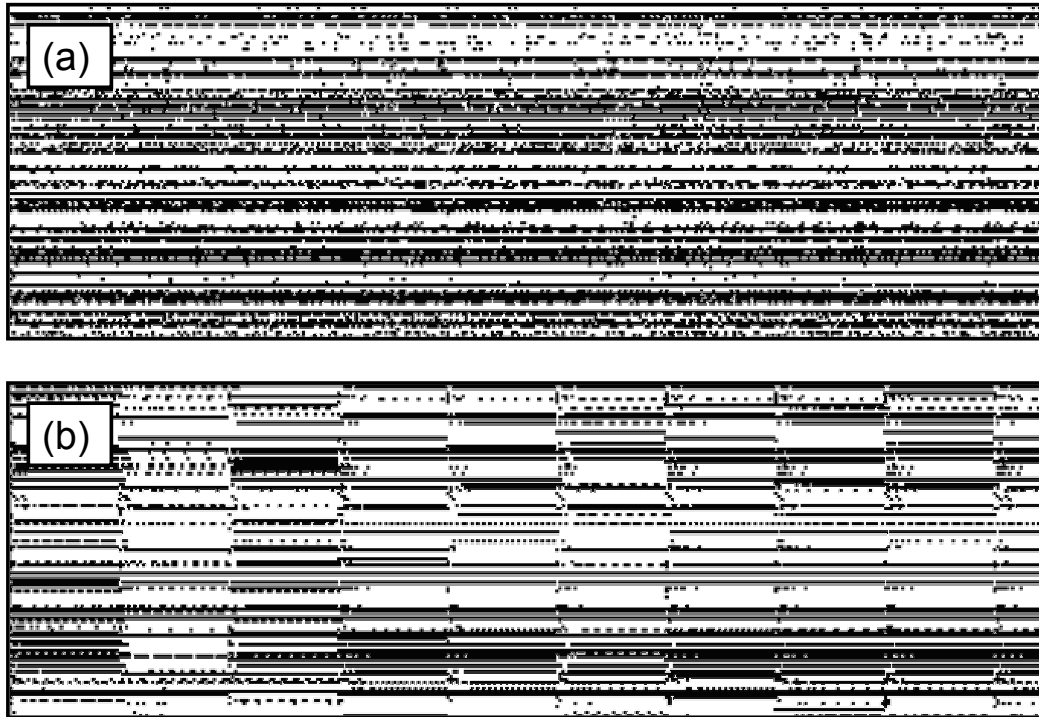


Figure 8: Expression Patterns for AG Network: (a) with proteins only; (b) with proteins and sRNA. In figure (b), the sRNA signals operate faster than the protein signals by a factor of 35. In these graphs, each row represents the state of a single gene, black when expressed, white when not expressed. A vertical column therefore describes the state of the entire network at a given time step and the entire graph shows how the state of the network activity changes over time. Network parameters: $N = 188$ (only the first 100 genes are shown), $K = 12.475$. No claim is made for the biological plausibility of the exact parameter values, which have been chosen purely to illustrate the characteristic behaviour of the model.

to generate highly connected networks that display ordered behaviour. Structural properties such as scale-free topology [15, 30, 1] and modularity [4, 32] also appear to provide an intrinsic level of stability to highly connected networks.

A significant feature of employing sRNAs as transcriptional regulators is that they have considerably shorter coding sequences and hence may be transcribed in less time. Furthermore, sRNAs are able to operate without being translated. Therefore, functional signals can be produced far more rapidly than would be the case if proteins alone were used [35]. We investigated the effect of reducing the rate of protein interactions with respect to sRNA interactions. This time delay was implemented by updating the input from sRNA-regulatory links every time step, but only updating the input from protein-regulatory links every τ time steps. It was found that, in almost all cases, chaotic dynamic trajectories collapsed into comparatively short periodic cycles. Interestingly, in some cases, network behaviour exhibited different classes of behaviour at two different time-scales. In the short term, at time-scales $< \tau$, the network moved towards an attractor with short period length. The change in input that occurred each time the protein-regulatory links were updated was sufficient to shift the network between attractors, resulting in the network exhibiting more complex dynamics over longer time-scales (see figure 8)

It has been shown that much of the interesting behaviour of Boolean networks disappears when the assumption of synchronous updating is removed [18]. To test what occurred when synchrony was not assumed, we altered the updating rule such that, rather than all proteins being updated together every τ time steps, each protein was updated with probability $1/\tau$ at each time step. Under this updating scheme, the periodic behaviour seen above disappears, however, the increased rate of the sRNA interactions does still appear to stabilize network dynamics, based upon measurements of the Hamming distance between successive states. We are investigating alternate statistical measures to characterize network stability in asynchronous models.

7 Conclusions and Further Work

This study has presented a model of genetic regulation incorporating a number of biological correspondences. We have demonstrated that when networks are extracted from a sequence representation, rather than being randomly generated, they possess structural properties more closely approximating those found in biological organisms. The model has been used to simulate the networks that result when two qualitatively different regulatory mechanisms, protein-mediated and sRNA-mediated, are used. Current work is

focussing on quantifying the effect of multiple interaction rates on network dynamics and using artificial evolution to further investigate the properties of gene networks with sRNA regulation.

Acknowledgements

For useful feedback on an early version of this paper, the authors would like to thank Jennifer Hallinan. This work was supported by an APA to NG and an ARC grant to JW.

Bibliography

- [1] M. Aldana. Dynamics of Boolean networks with scale-free topology. Available as arXiv pre-print cond-mat/0209571, 2003.
- [2] M. Aldana, S. Coppersmith, and L. P. Kadanoff. Boolean dynamics with random couplings. In E. Kaplan, J. E. Marsden, and K. R. Sreenivasan, editors, *Perspectives and Problems in Nonlinear Science. A celebratory volume in honor of Lawrence Sirovich*, Springer Applied Mathematical Sciences Series. Springer, 2003.
- [3] V. Ambros. Control of developmental timing in *Caenorhabditis elegans*. *Current Opinion in Genetics & Development*, 10:428–433, 2000.
- [4] U. Bastolla and G. Parisi. The modular structure of Kauffman networks. *Physica D*, 115:219–233, 1998.
- [5] S. Bornholdt and K. Sneppen. Robustness as an evolutionary principle. *Proceedings of the Royal Society of London, Series B*, 267:2281–2286, 2000.
- [6] P. Brazhnik, A. de la Fuente, and P. Mendes. Gene networks: how to put the function in genomics. *Trends in Biotechnology*, 20(11):467–472, 2002.
- [7] N. E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial transcription logic. *Proceedings of the National Academy of Science, USA*, 100(9):5136–5141, 2003.
- [8] J. C. Carrington and V. Ambros. Role of micrnas in plant and animal development. *Science*, 301:336–338, 2003.
- [9] L. J. Croft, M. J. Lercher, M. J. Gagen, and J. S. Mattick. Prokaryotic complexity is limited by a quadratic increase in regulatory overhead. Submitted to the Proceedings of the National Academy of Science, USA, 2003.
- [10] F. Dellaert and R. D. Beer. A developmental model for the evolution of complete autonomous agents. In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack, and S. W. Wilson, editors, *From animals to animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, pages 393–401, Cambridge, MA, 1996. The MIT Press/Bradford Books.
- [11] S. R. Eddy. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, 2:919–929, 2001.
- [12] P. Eggenberger. Evolving morphologies of simulated 3D organisms based on differential gene expression. In P. Husbands and I. Harvey, editors, *Fourth European Conference on Artificial Life*, pages 205–213, Cambridge, MA, 1997. The MIT Press/Bradford Books.
- [13] D. Endy and R. Brent. Modelling cellular behaviour. *Nature*, 409:391–395, 2001.
- [14] B. Ewing and P. Green. Analysis of expressed sequence tags indicates 35,000 human genes. *Nature Genetics*, 25:232–234, 2000.
- [15] J. J. Fox and C. C. Hill. From topology to dynamics in biochemical networks. *Chaos*, 11(4):809–815, 2001.
- [16] W. Gibson. The exon theory of genes. *Cold Spring Harbor Symp. Quant. Biol.*, 52:901–905, 1987.
- [17] S. Gottesman. Stealth regulation: biological circuits with small RNA switches. *Genes & Development*, 16:2829–2842, 2002.
- [18] I. Harvey and T. Bossamaier. Time out of joint: attractors in asynchronous random boolean networks. In P. Husbands and I. Harvey, editors, *Fourth European Conference on Artificial Life*, Cambridge, MA, 1997. The MIT Press/Bradford Books.
- [19] F. Jacob and J. Monod. On the regulation of gene activity. *Cold Spring Harbor Symp. Quant. Biol.*, 26:193–211, 1961.
- [20] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, 2000.
- [21] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.
- [22] S. A. Kauffman. Gene regulation networks: a theory for their global structure and behaviours. *Current Topics in Dev. Biol.*, 6:145–182, 1971.
- [23] S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford, UK, 1993.
- [24] K. E. Kürten. Critical phenomena in model neural networks. *Physics Letters A*, 129:157–160, 1988.
- [25] C. Langton. Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D*, 42:12–37, 1990.
- [26] B. Lemon and R. Tjian. Orchestrated response: a symphony of transcription factors for gene control. *Genes and Development*, 14:2551–2569, 2000.

- [27] J. S. Mattick. Introns: evolution and function. *Current Opinion in Genetics & Development*, 4:823–831, 1994.
- [28] J. S. Mattick. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports*, 2(11):986–991, 2001.
- [29] J. S. Mattick and M. J. Gagen. The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol. Biol. Evol.*, 18(9):1611–1630, 2001.
- [30] C. Oosawa and M. A. Savageau. Effects of alternative connectivity on behavior of randomly constructed Boolean networks. *Physica D*, 170:143–161, 2002.
- [31] A. E. Pasquinelli and G. Ruvkun. Control of developmental timing by microRNAs and their targets. *Annu. Rev. Cell Dev. Biol.*, 18:495–513, 2002.
- [32] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltavi, and A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297:1551–1555, 2002.
- [33] T. Reil. Dynamics of gene expression in an artificial genome - implications for biological and artificial ontogeny. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life: 5th European Conference, ECAL '99*, volume 1674 of *Lecture Notes in Artificial Intelligence*, pages 457–466, Berlin, 1999. Springer-Verlag.
- [34] L. M. Rocha. Contextual genetic algorithms: Evolving developmental rules. In F. Moran, A. Moreno, J. J. Merelo, and P. Chacon, editors, *Third European Conference on Artificial Life*, volume 929 of *Lecture Notes in Computer Science*, pages 368–382, Berlin, 1995. Springer-Verlag.
- [35] G. Ruvkun. Glimpses of a tiny RNA world. *Science*, 294:797–799, 2001.
- [36] Y. Setty, A. E. Mayo, M. G. Surette, and U. Alon. Detailed map of a cis-regulatory input function. *Proceedings of the National Academy of Science, USA*, 100(13):7702–7707, 2003.
- [37] C. W. Smith and J. Valcarel. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in Biochemical Science*, 25:381–388, 2000.
- [38] D. Thieffry, A. M. Huerta, E. Perez-Rueda, and J. Collado-Vives. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays*, 20:433–440, 1998.
- [39] J. Watson, N. Geard, and J. Wiles. Towards more biological mutation operators in gene regulation studies. To be published in the Proceedings of the 5th International Workshop on Information Processing in Cells and Tissues, Lausanne, Switzerland, 2003.
- [40] A. S. Wu and R. K. Lindsay. A survey of intron research in genetics. In H.-M. Voight, W. Ebeling, I. Rechenberg, and H.-P. Schwefel, editors, *Proceedings of the Fourth International Conference on Parallel Problem Solving from Nature*, pages 101–110, Berlin, 1996. Springer-Verlag.