Modelling Gene Regulatory Networks: Systems Biology to Complex Systems *ACCS Draft Technical Report*

Nicholas Geard

May 6, 2004

Note to the reader...

Please note that this is a draft version, but one on which I would particularly appreciate feedback (`nic@itee.uq.edu.au`). Almost all sections are as complete as time will allow, however the document is still subject to revision. The one section remaining to be written is the conclusion, which will summarize major themes, general principles and current research directions.

A PDF version of this document may be downloaded from:

http://www.itee.uq.edu.au/~nic/_accs-grn/modelling-grns.pdf

Overview

This document provides an overview of approaches to the modelling of genetic regulatory networks, with an emphasis on techniques from complex systems.

Section 2 provides a basic introduction to the biological processes that are involved in gene regulation. When a gene is expressed, information stored in an organism's genome is transcribed and translated into proteins. Some of these proteins are transcription factors that regulate the expression of other genes. These proteins are themselves under regulatory control, resulting in complex networks of interacting genes. These gene regulatory networks control a number of important cellular processes including responding to the environment, regulating the cell cycle and guiding the development of an organism.

Regulatory systems are generally too complex to allow abstract reasoning about their dynamics. Mathematical and computational formalisms therefore allow the creation of models in which all assumptions about a system are made explicit. Section 3 introduces some modelling concepts and motivations. Systems biology entails a cooperative cycle between model construction and experimental validation to study the emergent properties of biological systems. The various approaches to modelling may be broken down on their representation

of system state, their use of spatial and temporal dimensions and the questions that the model is being used to investigate.

The next sections of the document describe some of the major approaches to modelling regulatory networks. Section 4 reviews logical activation models, in which state variables take one of a number of discrete values. The most common approach is to allow two possible values (on and off) and represent system transitions using Boolean functions. There is a long history of using Boolean networks to model both the dynamics of abstract classes of regulatory networks as well as the behaviour of specific systems. A number of models have also been proposed that allow multivariate logic and more detailed updating functions. While these models are frequently restricted to systems of a limited size, they do allow a higher level of biological fidelity.

Section 5 describes continuous activation models, in which state variables take the form of continuous concentrations and systems are modelled using ordinary differential equations. While this theoretically allows a greater level of biological accuracy, the size and non-linear nature of biological systems renders many models analytically intractable and computationally expensive. One advantage to these formalisms however is the large body of dynamical systems theory that may be applied to such models. Hybrid approaches that incorporate elements of both logical and continuous formalisms have been proposed in an effort to allow the implementation of larger networks.

Many models of regulatory systems make the simplifying assumption that genes are expressed at a continuous rate. However, the biological processes involved are inherently noisy, and a number of formalisms have been developed to allow this aspect of regulation to be incorporated into models. Section 6 outlines some of the implications of stochasticity and noise and outlines some of the approaches to dealing with these issues. Again, while allowing a greater level of biological fidelity, stochastic models are frequently difficult to solve analytically and expensive to compute numerically.

A complementary body of work derived from the theory of random graphs has been produced analysing the statistical properties of the structure of regulatory networks. One of the key findings from the field of network theory is that real networks in many different domains, including biology, have certain structural properties that may have implications for their behavioural characteristics, such as system robustness. Results from this field of modelling are reviewed in Section 7.

1 Introduction

One of the most exciting challenges in biology today is the task of deciphering how the genome controls the development of complex organisms. This endeavour is utilising the skills and techniques of a wide range of academic disciplines. Researchers in molecular biology have access to sophisticated experimental technologies capable of gathering large amounts of data on genetic processes. The quantity of information obtained is too vast to be manipulated and processed

manually, leading to an increased usage of pattern detection, machine learning and data mining techniques from computer science. In addition, theory and formalisms from mathematics are being used to build models of systems. These models can help to clarify intuitions, manage data and assist in the development of a theoretical understanding of biological organisms.

For the past 50 years, the research program in molecular biology has been directed towards understanding biological systems at the level of their most fundamental components, such as genes, proteins and cells. In the last decade, the new field of systems biology has established a program aiming to reverse this reductionist trend. One of the primary aims of systems biology is to use a computational models to integrate diverse sources of experimental data back into a systems level description of biological organisms.

Another development of the last few decades is the field of complex systems, which is interested in the description and analysis of the systems consisting of large numbers of interacting parts. Such systems exist in many domains from ecology and biology to communication networks and engineering, leading to a strong emphasis on interdisciplinary studies.

Both complex biology and systems biology have much to gain from the other: systems biology can benefit from the tools and theoretical insights generated by studies of complex systems in other domains, while complex systems in turn has much to learn from the progress made in undestanding biological systems.

This document reviews some of the motivations for modelling biological systems and provides an overview of some of the the major formalisms that have been used to model genetic regulatory networks. In each section, sources for further reading are recommended, including pointers to further theoretical results and technical details, reviews of specific areas, as well as studies that are of particular historical interest.

2 Biological background

Biological systems are incredibly complex. One of the major challenges in modelling is deciding on an appropriate level of detail to include in a model. Too much detail results in a complicated model with reduced explanatory power. Too little detail risks omitting critical processes and mechanisms. What constitutes the "right" level of detail will vary depending on what question the model is being used to address. This section aims to present a sufficient level of detail about the biological processes involved in gene regulation to allow an appreciation of what is included and omitted by different formalisms. By necessity, it presents a simplified view of current biological knowledge; pointers to more detailed reviews are therefore provided.

2.1 The basics of gene expression

Information in a biological organism is stored in its genome. The genome of all complex organisms consists of long molecules of DNA made up of chains of

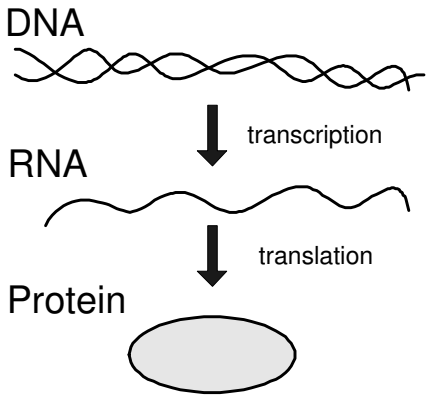

Figure 1: The Central Dogma. The creation of a protein molecule from a DNA double helix occurs in two stages, transcription and translation. In the transcription stage, the two strands of DNA are separated at the site of the gene, and the RNA polymerase enzyme copies the noncoding strand of DNA into a complementary mRNA strand. The mRNA is then transported from the nucleus to the cytoplasm where it is translated by ribosomes into an amino acid chain.

nucleotides in a double-helix structure. The basic functional unit of the genome is a gene. The central dogma of molecular biology states that information stored in the DNA of a given gene is transcribed into RNA, which is then translated into proteins (see Figure 1).

Proteins are the fundamental structural and functional units in cells. Each one is specialised to carry fill one of a variety of important roles, such as a structural element, enzyme catalyst or antibody. A large subset of proteins known as transcription factors (TFs) also play a regulatory role, determining when, where and how much a particular gene is expressed into proteins. Because regulatory proteins are themselves the products of expressed genes, they too are under regulatory control, giving rise to complex networks of interacting genes.

This section describes the processes of transcription and translation that mediate the path from DNA to protein in prokaryotic and eukaryotic cells. While the gene expression mechanism in both types of cells is generally very similar, there are several significant differences [121]. In eukaryotic cells, DNA is stored in the nucleus, whereas prokaryotic cells have no nucleus. All complex, multicellular organisms are eukaryotic, and their cells tend to have a considerably higher level of regulatory complexity than single-celled prokaryotic organisms such as bacteria.

2.1.1 Transcription

A gene consists of a regulatory region, which controls when the gene will be activated, and a coding region, which specifies the shape of the protein that will be produced when the gene is activated (see Figure 2). In prokaryotes, the regulatory region is generally located directly upstream of the coding region, whereas in eukaryotes elements of the regulatory region may be located at a considerable distance both upstream and downstream from the coding region. A regulatory region contains binding sites for a number of transcription factors (TFs). Individual TFs may exert either positive or negative control on the activation of a gene, increasing or decreasing its rate of transcription. When the activation conditions for a given gene are fulfilled, a large molecule called RNA Polymerase binds to the TF complex and the DNA in the gene's coding region is unwound. The sequence of nucleotides on the coding strand of the DNA is then used as a template to create a single-stranded messenger RNA (mRNA) molecule [92].

In prokaryotes, the coding region is contiguous. In eukaryotes however, the coding region is broken up into a series of coding exons and non-coding introns, which must be spliced out of the initial RNA transcript. A number of other processing mechanisms are also possible at this stage. In many cases, a single eukaryotic gene can be spliced and edited in multiple ways to produce a variety of different protein products [110] (see Figure 3). As the next step of gene expression, translation, occurs in the cytoplasm of the cell, mRNA molecules in eukaryotes must also be transported outside of the cell nucleus.

2.1.2 Translation

Once in the cytoplasm, mRNA molecules bind to another large molecule called a ribosome. A ribosome reads an mRNA molecule in triplet known as codons. Each codon maps to one of twenty possible amino acids, that are chained together in the order specified by the mRNA. The newly created amino acid chain then folds into a complex three-dimensional protein structure.

Whereas DNA is a stable molecule, mRNA and proteins have only limited lifespan before they are broken down and their constituent nucleotides and amino acids are reused. Both mRNA and proteins may be degraded at different rates depending on their conformation and the presence or absence of other chemicals in the cell. While the most well understood form of regulation occurs at the transcriptional level, control of gene expression may be exercised at the at almost any stage of protein synthesis. Regulation is also known to occur at the level of RNA processing, mRNA transport and translation, protein modification and mRNA and protein degradation.

2.2 The control tasks of the genome

The genome is responsible for controlling cellular tasks such as response to environmental conditions, the cell division cycle and cell differentiation. Each of these require the regulation of gene expression in both space and time.

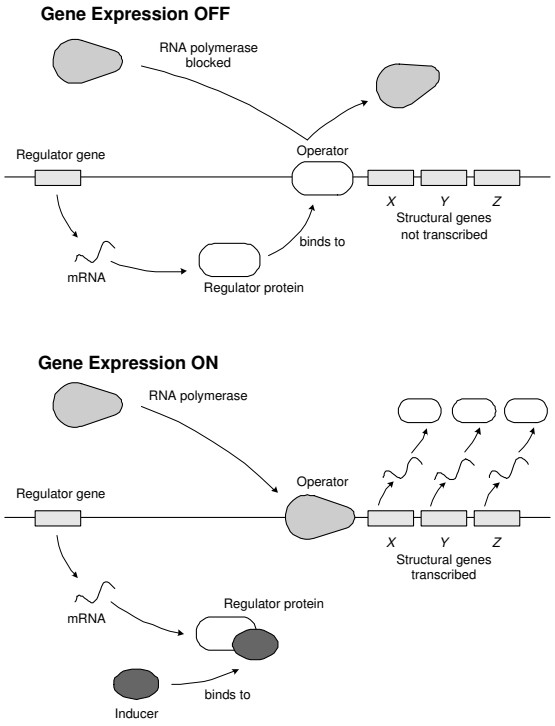

Figure 2: Regulation of transcription initiation (the operon model). In the operon model, the operator region of a gene may be bound by a regulator protein, preventing the transcription of structural genes by RNA Polymerase (top). When an inducer molecule is present, it binds to the regulator protein, releasing the operator and allowing RNA polymerase access to transcribe the structural genes (bottom).

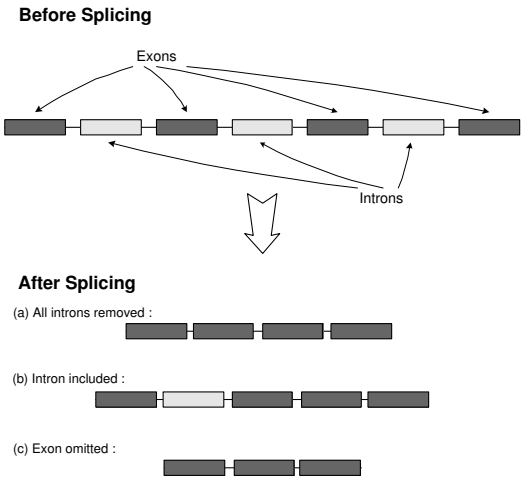

Figure 3: Alternative splicing. In eukaryotes, the DNA coding for a protein is not stored continuously, but as multiple coding sequences (exons) interspersed with noncoding introns. A variety of proteins may be generated from the same base sequence by selectively including or omitting exons or introns.

Throughout its lifetime, a cell must respond to many different types of environmental signals. Single-celled bacteria are able to detect and move towards nutrient sources, they also react to changes in temperature and acidity. Multicellular cells must also respond to chemical signals emitted by neighbouring cells in the organism. These external signals are transmitted to the genome via a series of chemical reactions known as signal transduction pathways [50].

As well as responding to external signals, the genome is also subject to internal control. The cell cycle plays the role of a cell's internal clock [88]. In order for an organism to develop, each embryogenic cell goes through a process of growth, replication and division. During cell growth, a cell increases in size. Its entire genome is then replicated to produce two identical copies. When the cell divides, each of its daughter cells contains one complete copy of the genome. The signals that tell a cell when to switch from growth to replication and from replication to division are controlled by a subset of genes that regulate timing.

Each cell of a multicellular organism contains identical genetic information (with some rare exceptions). The feature that distinguishes cells of different types is the set of genes that are active in a particular cell. This pattern activation determines which proteins are produced, and hence the functional properties of the cell. When an egg cell is initially fertilised, it is fully undifferentiated and has the potential to become any type of cell. As an organisms developmental program unfolds, its cells divide and undergo physical and chemical changes that result in their final fates (for example, as blood or skin cells) becoming more specified [142]. The role of the gene regulatory network in this process

is to integrate the internal dynamics of the cell and external signals from the environment and other cells to control the differentiation process.

2.3 Further reading

Molecular Biology of the Cell [6] and *Genes* [77] are two well known textbooks in the area of molecular biology and genetics. Both are comprehensive, clearly presented and regularly revised. For a lighter, more general introduction to genes and gene regulation, Enrico Coen's *The Art of Genes* [27] and Evelyn Fox Keller's *The Century of the Gene* [72] are recommended. In *A Genetic Switch* [95], Mark Ptashne provides a short, highly readable overview of the simple, yet remarkably powerful, genetic circuit that controls the developmental pathway of phage-λ. *Genomic Regulatory Systems* [28] by Eric Davidson and *From DNA to Diversity* [26] by Sean Carroll are thorough overviews of gene regulation that assume a little more background knowledge.

3 A diversity of models

A wide variety of formalisms for modelling genetic regulatory networks (GRNs) have been proposed. Before reviewing several important models in detail, this section provides a high level introduction to the field. It is important to note that the choice of an appropriate modelling formalism is very dependant on the aim of a study. This section begins describes the main ways in which modelling formalisms differ, the dynamic system concepts used to frame models, and some of the different goals that motivate GRN modelling.

3.1 Why build models?

Modelling a system involves building a formal description of the system on the basis of current knowledge and understanding. Traditionally, models are contructed to allow a system to be conceptualised and communicated and to assist in determining the course of further research. Over the last fifty years, there has been an increasing trend towards the use of mathematical and computational formalisms to frame models of regulatory systems in biology. The structure of such systems is frequently complex, consisting of multiple intertwined feedback loops and non-linear interactions. This structural complexity, combined with the varying timescales on which different biological processes act, makes it particularly difficult to develop intuitions about how regulatory systems operate. Building a formal model of such a system requires all assumptions about the timing and connectivity of regulatory elements to be made explicit. Modelling can therefore provide a valuable check on intuitions during the development of hypotheses [82].

In addition, formal models are frequently complemented with computer simulation, in which a model is built and then used to make some form of prediction

about system behaviour. Running simulations using models based on known systems can provide validation of a particular modelling approach. Furthermore, such simulations can also provide valuable guidance to target future studies by enabling experiments to be carried out *in silico* that would be expensive, time consuming or otherwise infeasible to perform *in vitro*.

3.2 Complex systems and systems biology

The field of complex systems is interested in the complicated systems consisting of many interacting components that occur in many different fields. Economic markets, ant colonies, the Internet and metabolic networks are all examples of complex systems. A fundamental characteristic shared by all these systems is that they can be described as a network in which nodes are components and edges between nodes are interactions between components. Each individual compononent in the system may be relatively system, however complex behaviour frequently emerges as a result of the interactions between large numbers of such simple components.

Within a context of a genetic regulatory network, the system parts are genes and proteins while the emergent properties of interest include oscillatory behaviour, pattern formation, robustness and a number of other complex control phenomena. The field of complex systems is highly interdisciplinary and much of the literature is focussed on systems in a particular domain, such as biology, ecology or economics, and the extrapolation of insights between domains. There is also a growing emphasis on general techniques, theories and insights that may be applied across domains [108].

Recently, the cooperative efforts of theoreticians and experimentalists have been embodied in the new field of systems biology [74]. The tools of systems biology are the large quantities of data generated by high-throughput experimental techniques and the increasingly sophisticated range of mathematical modelling techniques. The aim of systems biology is to integrate models at multiple biological scales and investigate systems-level properties of biological organisms. This aim includes understanding at four levels: (a) the structure of biological interaction networks; (b) their dynamics, how states change over time in different conditions; (c) the methods biological systems use to control the state of a cell; and (d) the design of systems, including both how they have evolved and how they may potentially be artificially constructed [73].

A key feature of systems biology is the integration of both theoretical modelling and empirical investigation, in which current biological knowledge informs the development of models and the analysis of these models produces a set of predictions that may then be tested in the laboratory (see Figure 4).

3.3 Key features of dynamic systems

As a large numbers of different formalisms have been used to model GRNs, it is useful to have an underlying conceptual framework that can be used to categorize and compare particular models. A common view of a regulatory network is

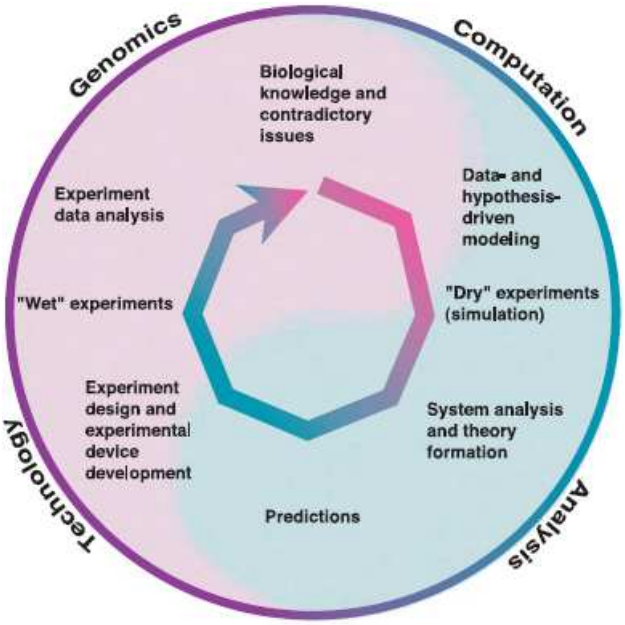

Figure 4: The modelling–experimentation cycle in systems biology. From [74].

as a dynamic system, consisting of a set of components (genes, gene products) whose properties change in response to internal interactions and external signals. The two fundamental concepts in a dynamic systems description are state and transition.

A state of a system is a description of the properties of each component at a given point in time. In a GRN model, this may include levels of gene activation, concentration of chemical species or even the number and location of individual molecules, depending on the level of resolution of the model. A related concept is a state space, the total set of possible states a system can be in. The state space of a system will have a dimensionality equal to the number of components in the system (see Figure 5).

States in a state space are linked together by transitions, which describe how the state of a system is updated. The set of transitions that can be applied to any given state will determine the possible state or states into which a system can move. The path of a system through state space over time is often referred to as its trajectory.

3.4 Logical, continuous and stochastic models

The four main model categories that will be considered in this review are: logical models, in which the a state variable takes one of a number of discrete

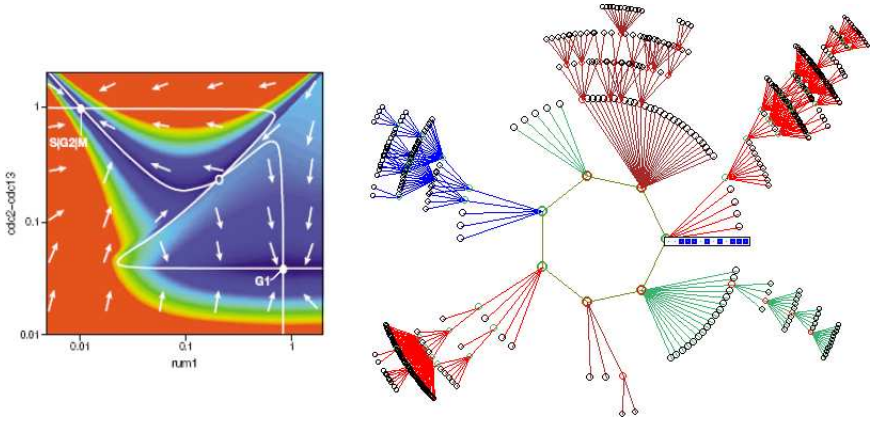

Figure 5: State spaces. Example representations of both a continuous (left) and a discrete (right) state space.

values; continuous models, in which each state variable is a continuous value; and stochastic models, in which each state is a probability distribution of possible configurations. The final category consists of network models, which are primarily concerned with the structure of networks and their evolution, and only secondarily with dynamics. Within these primary categories, models may also be distinguished by their treatment of space and time.

Of the models that are concerned purely with intracellular dynamics (i.e., regulation within the cell), many omit any reference to regulation taking place in a spatial domain, Other models acknowledge the fact that cells have complex spatial structures by including time delays due to diffusion and molecular transport. Some models, such as investigating pattern formation, model the spatial arrangement of groups of cells. This spatial arrangement may be in either two or three dimensions and may be represented either on a grid or in a continuous space.

Similarly, the evolution of a system behaviour in the temporal dimension may be modelled in a discrete or continuous fashion. When time is measured in a discrete fashion, as it often must be if a model is being simulated, or solved numerically, a further choice arises of whether state variables are updated all at once (synchronously), or independently (asynchronously). Some models are concerned purely with the static structural properties of interaction networks and include no temporal dimension. A number of models have been designed to investigate how regulatory networks have evolved. In these models, *evolutionary* time may also be a factor.

3.5 Different motivations for building models

Finally, models may be differentiated by the question that motivates their development. This motivation may range from a desire to obtain quantifiable values for some aspect of a system that can then be experimentally validated through to exploring high-level principles of cellular control. While it can be an oversimplification to categorise a model according to its purpose, as many models will overlap across different categories, several broad approaches can be discerned.

Crafted models of specific systems Many attempts at modelling regulatory interactions focus on small, well understood systems that can be modelled by hand from available empirical knowledge. These models generally have a high level of fidelity to the underlying biological system, with each component in the model system corresponding to a particular element of the biological system. Numerical and computer simulations are used to make predictions about systems that are too complex to allow for analytical solution. This category includes both continuous models, such as the various models of phage-λ [109, 102, 9], as well as logical models, such as Bodnar's Boolean characterisation of *Drosophila* embryogenesis [18].

Phenomenological models of biological mechanisms Another approach, at a slightly higher level of abstraction, is to use systems of generalised components to reproduce observed biological behaviour, such as morphogenesis and pattern formation. In these models, there is no longer a direct mapping between components in the model and copmonents in the biological system, however the high level behaviour of the system is preserved. An example of this category is the gene circuit models developed by Mjolsness, Reinitz and Sharp for modelling segmentation in *Drosophila* [87].

General models of classes of networks Other researchers, rather than investigating individual systems, have taken the approach of characterising the behaviour of classes of networks with particular structural and dynamic properties. These approaches frequently work with simplified descriptions of gene activation that allow much larger and more complex networks to be simulated than would otherwise be possible. A common technique is to generate a large number of random networks (an ensemble) governed by a specified set of local rules and observe the statistical properties of the global behaviour [69]. Another type of modelling that falls into this category is the exploration of networks whose structures share particular statistical properties, such as scale-free connectivity distribution [12], or hierarchical patterns of modularity [98].

Network models inferred from experimental data The rapid increase in available experimental data in recent years has shifted some of the focus towards techniques that are able to automatically construct models of larger, less well-understood regulatory systems. Advances here are divided between both the

formalism that is used to model the system, and the learning algorithms that are used to derive the model from the available data [134].

3.6 Further reading

There are a number of reviews of gene network models in the literature, many of which either focus on a particular model system or modelling formalism, or are targeted at an audience with knowledge in a particular background area. Two recent overviews that provide a good coverage of the field have been published by Hidde de Jong [31] and Paul Smolen [112]. In addition, *Computational Modeling of Genetic and Biochemical Networks* [23] by James Bower and Hamid Bolouri provides a good overview of modelling formalisms (chapter 2), as well as a more in depth look at several particular techniques. A general description of the complex systems approach to modelling (not specific to biological systems) can be found in *Emergence* by John Holland [62].

4 Logical models

The defining characteristic of the models classified as "logical" is that their state variables are measured discretely. While this frequently represents a high level of abstraction from actual biology, the models described below have nevertheless been responsible for providing a number of theoretical insights and have had a significant influence on thinking about regulatory networks. In particular, the Boolean network model was critical in defining the complex systems view of biology.

4.1 Boolean networks

One of the earliest approaches to modelling large networks of interacting genes was to view a genetic regulatory system as a network of logical elements [67, 68, 70].

4.1.1 Assumptions

The Boolean network approach makes a number of assumptions to simplify analysis [119]. First, the activation of a single gene is represented as a Boolean switch that can be either on or off. In effect, a gene can be either expressed or not expressed and there is no possibility of intermediate levels of activation. This assumption is reasonable when a gene spends most of its time either at a floor value of zero or at some positive saturation level and the time required for a gene to switch is negligible with respect to the time scale of the model. The second assumption is that the regulatory control of a gene is described by a combination of Boolean logic rules, such as AND, OR and NOT. The final assumption is that timing is synchronous, that is, the states of all genes are updated simultaneously at each time step.

One of the immediate advantages of these assumptions was that the computational requirements of simulating regulatory systems were massively reduced, allowing the exploration of much larger systems. On the other hand, the validity of the above assumptions, and the value of the Boolean approach in general, has been questioned by a number of people, particularly in the biological community, where there is a perceived lack of connection between simulation results and empirically testable hypotheses [42].

As there was little knowledge of the connectivity patterns in real biological networks, Kauffman used an *ensemble* approach, generating large numbers of randomly connected networks with randomly chosen Boolean updating functions [69]. His goal was to measure the generic properties of certain classes of networks and observe how their global dynamics resulted from local interactions.

4.1.2 Model description

Kauffman's model of Boolean networks have two primary parameters: network size, N, the number of elements in the network and network connectivity, K, the number of inputs regulating the activity of each element. Each of the N elements is associated with a rule table specifying outputs for each of the 2^K possible input combinations. As each element in the network is updated simultaneously, the system is deterministic and the state at time $t + 1$ can be determined on the basis of the state at time t (see Figure 6). The rule tables for each element can be defined in a number of different ways [7]: They may be fixed over time (the *quenched* model), as is usually the case when a single network is being simulated. Alternatively, a new set of rule tables may be generated at each step (the *annealled* model), which simplifies theoretical analysis of network behaviour.

4.1.3 Theoretical results and hypotheses

Classes of behaviour The dynamics of a system will fall into three different phases depending on the value of K. There are a number of different metrics for distinguishing between these phases, one of which is information transfer. If two identical systems are initialized with similar, but not identical, starting states, the distance between their subsequent states (measured by a Hamming metric) will change over time. This property reflects the localisation of information transfer. If the Hamming distance stays small, information is communicated across only a local portion of the network. If the Hamming distance increases, it indicates that information is being transferred to a much larger portion of the network.

When $K > 2$, the Hamming distance grows exponentially with time and the system is in the chaotic or disordered phase. When $K < 2$, the Hamming distance decays exponentially with time and the system is in the frozen or ordered phase. For $K = 2$, the Hamming distance remains stable, subject to fluctuations. This phase has been referred to as the critical or complex phase. It is also colloquially known in some contexts as "the edge of chaos".

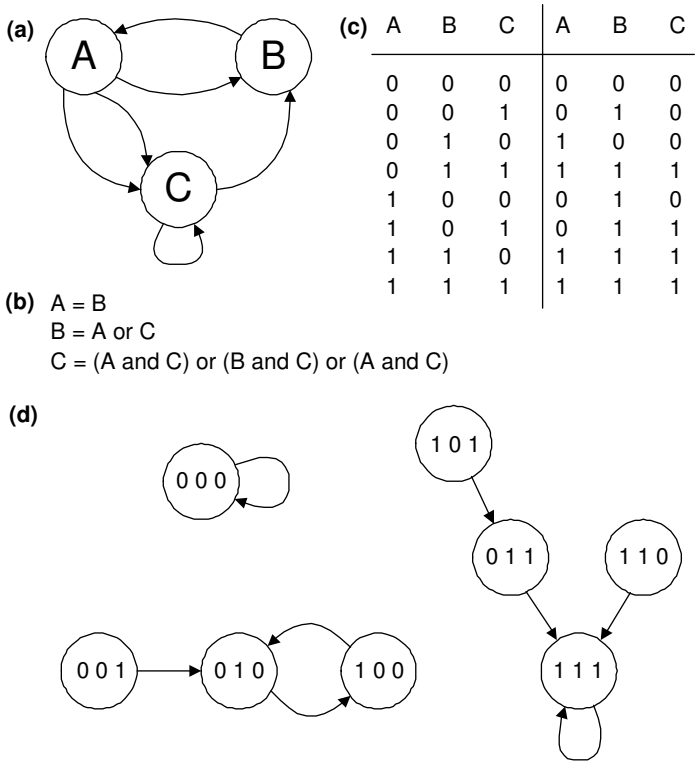

Figure 6: An example of a Boolean network: (a) the wiring diagram; (b) the updating rules; (c) a state transition table, showing how network activation at time $t + 1$ depends on network activation at time t; and (d) the state space of the network, with two point attractors and a limit cycle with a period of two.

For quenched networks (i.e., those with updating functions fixed over time), a system will eventually return to a previously visited point and the dynamics will form a cycle with a given period. All possible states in a system will either be a part of one of these cycles, or a transient point in a path leading to one of these cycles. Taking language from the field of dynamic systems, a cycle can be referred to as an *attractor* and the set of all points that lead to a particular cycle as its *basin of attraction*. A *garden of eden state* is a particular type of state that has no predecessors. An attractor can either be a fixed point (period equal to 1), or a limit cycle (period greater than 1) (see Figure 6 (d)).

Attractors as cell types Chaotic systems tend to contain cycles with long periods and long transients. Frozen systems tend to have much shorter cycles and transients. The behaviour of critical systems is intermediate between these. As mentioned in Section 2, different cell types are distinguished primarily on the basis of which of their genes are expressed. Kauffman draws an analogy between an attractor in a Boolean network and a particular cell type or fate. The transient period then corresponds to the process of cell differentiation. In the chaotic regime, these transients would appear to be unrealistically long. Furthermore, systems in the chaotic regime tend to be highly sensitive to perturbations, which does not correspond to the robust behaviour displayed by biological systems. On the other hand, systems in the frozen regime, while displaying acceptably short transient lengths, have virtually zero sensitivity to perturbations, which would appear to preclude any differentiation whatsoever.

Kauffman therefore proposed that life occurs in the vicintiy of the critical regime [70], and argued that the relationship between attractor number and system size in Boolean networks mirrored the observed relationship between cell types and number of genes in various biological organisms [67]. The exact properties of the scaling law between system size and attractor number has been the subject of continued debate [15, 17, 115]. Regardless, at a qualitative level, systems in the critical regime tend to display both short transient lengths and a small, but significant, level of sensitivity to perturbations. These features are consistent with a biological system in which cell types are relatively stable but have a small possibility of mutating to one of a few "neighbouring" cell types. The properties of Boolean network state spaces and the analogy between basins of attraction and cell types have been extensively explored by Wuensche [143, 144].

4.1.4 Extensions and applications

Updating rules A major problem with Kauffman's argument is that the level of connectivity of networks displaying such complex behaviour ($K = 2$) is much lower than has been observed in real systems (where some genes may be controlled by as many as 20 regulatory factors). Several modifications to the Boolean model have been proposed that address this issue. By default, a random Boolean function has an equal probability of switching a given gene on or off. However, the model can be extended by the addition of a bias term,

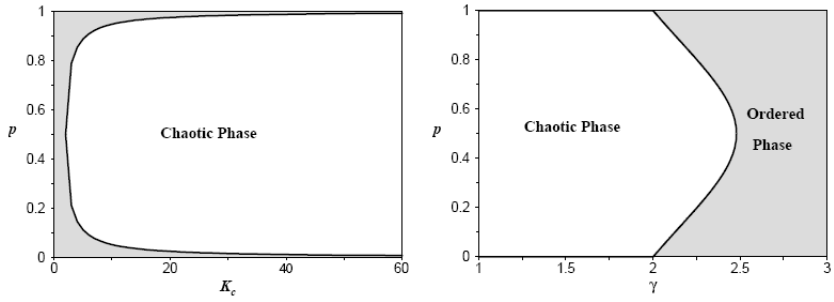

Figure 7: The phase space of Boolean networks with random (left) and scale-free (right) connectivity.

p, specifying the probability that a certain combination of inputs result in an element being switched off.

The behaviour of a network now depends on both K and p For a fixed K, these three phases can also be reached by altering p, the bias in the rule tables (see Figure 7). As p is changed, the level of connectivity corresponding to the critical phase, K_c is given by:

$$K_c = \frac{1}{2p(1-p)} \tag{1}$$

Kauffman has also proposed that biology may use only a subset of the total possible range of Boolean functions, termed canalizing functions, in which the state of a single input is sufficient to determine at least one of the possible output states [70]. A network using canalizing functions displays more stable behaviour than a non-canalizing network, however, as K increases, the proportion of functions that are canalizing decreases rapidly. It has been argued that canalizing functions are likely to be extremely rare at realistic levels of connectivity [7]. However, reviews of the biological literature have suggested a strong bias towards canalizing functions in regulatory interactions with 3, 4 and 5 inputs [53]. A number of other definitions of Boolean updating functions that produce more stable behaviour have also been proposed [105, 96].

Network structure While the approaches to increasing network stability mentioned so far have focused on modifying the updating rules, or the network dynamics, it has also been suggested that changing network structure may stabilize network behaviour. In particular, if nodes are connected with a scale-free distribution (see Section 7 below), rather than a random distribution, the position of the order/disorder boundary in the state space will be modified, increasing the size of the ordered region [7, 43, 91] (see Figure 7).

Another approach that has been taken to generating Boolean networks in a non-random fashion is to extract the network structure and updating rules

from a lower level of description. In the Artificial Genome model [99], network structure and functions are generated by parsing a string of bases (the artificial genome). This method results in networks with a significantly different degree distribution and a restricted set of updating rules [46].

Timing Most of the results mentioned so far have relied on the assumption that network updating is carried out synchronously, that is, the activation of every node is updated simultaneously. It has been pointed out that relaxing this assumption and allowing asynchronous updating introduces a level of indeterminism that interferes with many of the interesting phenomena displayed by traditional Boolean networks [55]. Using two different definitions of asynchrony (essentially with and without replacement), Harvey and Bossomaier found that cyclic attractors disappeared, point attractors remained, and a new category of "loose attractors" appeared, in which the network passes indefinitely through some subset of its possible states [55]. The nature of basins of attraction also changes, with some being definite basins, from which all paths lead to the attractor, and others being possible basins, form which at least one path leads to the attractor.

This work was followed up by Di Paolo, who defined a measure of "pseudo-periodicity" in which an autocorrelation function is used to measure the probability of a given state approximately recurring with a particular regularity. He demonstrated that it was possible to evolve systems that were able to display rhythmic behaviour [94]. Analysis of these evolved systems has been carried out to determine what properties of networks allow the emergence of robust rhythmic behaviour from inherently noisy components [103]. A pruning algorithm is presented that allows evolved rhythmic networks to be reduced to their functional core and reveal that a common feature of these networks is a ring of elements that produces travelling waves of activation. This architectural component acts as a cellular clock for the entire system, other nodes in the network being either stationary or entrained by the central clock. One limitation of this analysis is that it favours the evolution of rhythmic behaviour in networks with relatively low values of K. An advantage of these systems is their intrinsic robustness to external perturbation. The evolutionary search mechanism used to evolve these networks also biases the discovery of networks operating with a single timescale, whereas biological systems can accommodate more complex temporal designs [103].

Applications Boolean networks of genetic regulation have also been applied in a number of other domains, including:

- to build models of specific systems, such as, *Drosophila* embryogenesis [18] and the endothelial cell cycle [63];
- as the basis for phenomenological models of a morphogenetic processes [59, 60, 61];
- to study the evolutionary dynamics of regulatory networks [21, 22, 44];

- as a framework for inferring regulatory networks from gene expression data [78, 1, 2, 3]; and
- as a biologically-inspired control mechanism for autonomous agents [35].

4.1.5 Strengths and limitations

The main strengths of the Boolean network model are its analytical tractability and the ease and efficiency with which it can be simulated. The primary limitations of the model are its perceived lack of applicability to biological systems. Some of these issues, such as connectivity and synchrony, have been raised in the section above. A more fundamental objection concerns the starting point for these models, the validity of the Boolean assumption. Some genes are known to have different regulatory effects depending on their level of expression and in some situations the transient period between as a gene switches may be significant. While a Boolean representation may be sufficient for a product that tends to be present either in excess, or in insignificant quantities, products whose concentration varies in a more smoothly continuous fashion may require a continuous function to accurately capture their dynamics [112, 20]. A number of researchers have also demonstrated that there is not a direct correlation between the dynamic behaviour of Boolean systems and that of corresponding continuous systems [49, 11], suggesting a qualitative loss of behavioural information.

4.2 Generalised logic

The generalised logic formalism for modelling GRNs has been developed by René Thomas and a number of colleagues over the past three decades [127]. While its origins lie in similar areas to the Boolean models described above, it is distinguished by several features: it is inherently asynchronous, it allows variables to take multiple logical values and it allows for a more sophisticated definition of logical interactions, involving multiple thresholds and parameters. Generalised logic is also motivated by a different set of questions. While Kauffman's networks were developed to investigate the theoretical properties of an entire class of networks, generalised logic tends to focus on models of actual systems. It provides a set of tools with which to characterise and analyse networks derived either from known interactions or from measured patterns of gene expression in terms of their dynamic steady states.

4.2.1 Assumptions

Although the initial version of the generalized logic formalism described the state of a gene in a Boolean fashion [123], later iterations introduce the possibility of state variables assuming more than two levels [133, 124]. The argument for multivariate logic is that when a particular element acts in more than one context, it cannot necessarily be assumed that the thresholds required for each of these actions to occur is going to be equal. For example product X may have

an effect on gene Y when it reaches concentration c_1 and also have a further effect on gene Z at concentration c_2 (see Figure - multiple thresholds).

The generalized logical formalism also allows for a considerably more sophisticated form of logical updating than the Boolean rules used in RBNs. The first refinement is the introduction of logical parameters, which allow for weighted gene interactions [113]. The argument for allowing this complication is that genes may be expressed to different extents in different circumstances and therefore may affect the expression of another gene to varying degrees. The second refinement concerns the possibility that some steady states of a system, particularly unstable ones, may be located at the threshold values [113]. This issue is dealt with by introducing logical values for the thresholds, as well as for expression levels below and above thresholds.

Unlike RBNs, in which time is measured discretely, the generalized logic formalism uses continuous time, allowing for asynchronous updating of elements [124]. It is important to note that the form of asynchrony used here is deterministic in its ordering of element updating. Instead of a set of deterministic state transition rules, a generalized logic model defines a set of functions mapping current states to their image, or the state towards which a system would tend to move if all variable updates were carried out. This transition is enhanced by the inclusion of two time delays, one describing the period between a gene switching on and its product reaching functional levels and the other describing the period between a gene swtiching off and its product dropping below functional levels (see figure - time delays). The use of asynchrony produces to systems containing more complex sets of periodic attractors than standard synchronous networks and the dynamics of such systems tend to be closer to equivalent differential models.

4.2.2 Model Description

The first stage in building a logical description of a system is to specify the graph of positive and negative interactions between logical elements. From this diagram, logical equations, and a corresponding image table may be inferred. It is important to note that, unlike the Boolean network approach described above, the image table does not show deterministic transitions. Whereas the standard Boolean network assumed synchronous updating of all elements, the generalised logic formalism is inherently asynchronous. Therefore, in a transition involving the change of state of two genes, the probability of both genes being updated simulataneously is infinitesimally small. Therefore, one of the two possible transitions will occur first, dependant on the time delay for that element, and determine the next state. Carrying out this process for all states results in a transition graph, from which steady states and cycles can be identified (see Figure 8). The path that will actually be taken from this graph can be determined by considering the time delays of each transition. A more thorough description of the model, including more advanced elements such as logical parameters and multi-valued logical variables is given in [127].

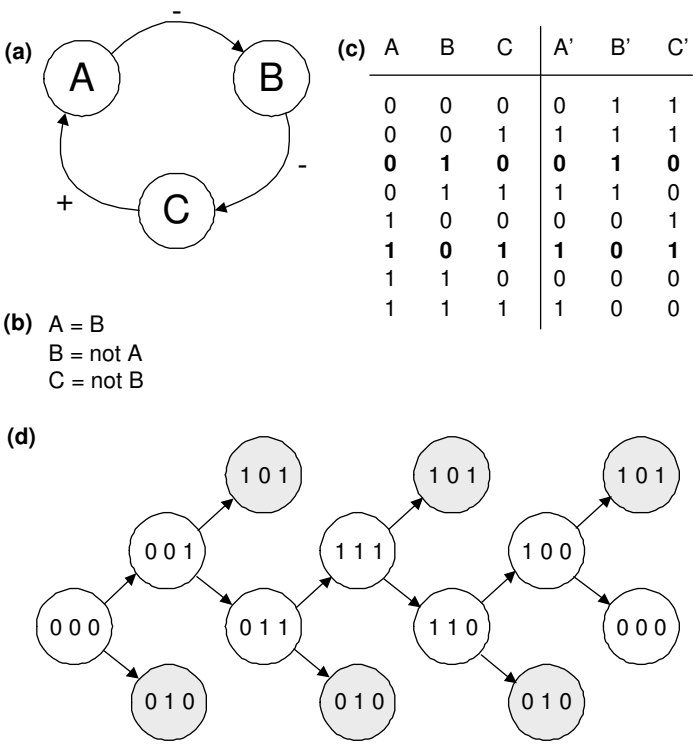

Figure 8: An example of a generalised logic network: (a) a simple network with two negative interactions and one positive interaction; (b) the logical updating rules; (c) the image table. Note that two of the states, 100 and 101 have themselves as image, these represent the steady states towards which a system will ultimately move; and (d) the transition graph.

4.2.3 Theoretical results and Applications

Once a model has been built, the set of logical equations can be analysed to determine the logical steady states of the system, analogous to attractors in Boolean systems. A state space can then be constructed in which each state corresponds to a qualitative behaviour of the system.

The ease with which qualitative analyses can be carried out has been increased by theorems which allow the identification of steady states of a system by considering the *characteristic state* of the circuits that make up a system [114, 126]. A circuit is a complete loop in the system's interaction graph. The characteristic state of a circuit is the intersection of the thresholds beyond which each variable in the circuit is active. The properties of these circuits will determine which of them are functional in given conditions, and from this knowledge, the steady states of the system can be determined. While the number of logical states grows rapidly with the size of the system, the number of circuits increases much more slowly, therefore the ability to derive steady states from characteristic states greatly improves the scalability of this type of analysis.

A feedback circuit can also be described as positive or negative, depending on whether it contains an even or odd number of inhibitory interactions respectively. Negative circuits generate homeostasis, while positive circuits are involved in multistationarity, and hence differentiation. Circuits may interact to produce multistationarity in a number of different conditions [128].

The generalised logic formalism has been applied to the analysis of a number of real genetic systems, including phage-λ [122], dorso-ventral patterning in *Drosophila* [106] and flower morphogenesis in *Arabidopsis thaliana* [84].

4.2.4 Strengths and limitations

The generalized logic formalism is a powerful method for analysing networks whose interactions are well known. It enables the possible qualitative behaviours of a system to be determined in a rigorous and scalable fashion. The use of logical values corresponding to functional thresholds removes the necessity of having to set the values of large numbers of real parameters. The process is amenable to being automated by a computer and it has been demonstrated to be effective for the induction of gene networks from expression data. The explicit inclusion of time delays leads to a considerably more accurate picture of biological systems than synchronous Boolean networks and a number of theoretical insights into necessary conditions for multistationarity have been shown.

One of the primary limitations of this approach is that, because it has been designed for the detailed analysis of relatively small systems consisting of well characterized interactions, its scalability is limited. It is less suited to the exploration of different classes of behaviour and of large, less well-known systems. Furthermore, phenomena such as cyclic behaviour in generalized logic models are quite sensitive to those parameters which do require specification, such as time delays.

4.3 Continuous logic

Continuous logic is used to refer to models of regulatory systems in which the activation of a given gene is again considered to be Boolean, but the analytical treatment of the models is more similar to that used in continuous models than previous approaches. Furthermore the system states that are measured are generally qualitative in nature (in comparison to true continuous approaches, where system states are frequently quantitative). Such continuous logic models have been used both for ensemble approaches to determining classes of behaviour as well as qualitative simulation approaches for modelling the behaviour of particular systems.

Glass networks use piecewise linear differential equations (PLDEs) to describe the switching of gene states in continuous time [49]. This methodology has the advantage of rendering systems amenable to analysis, while still allowing complex periodic and chaotic dynamic patterns. The motivating question for this formalism is: "Given a network with a certain logical structure, what are the possible dynamics that can be found in this network?" [40].

4.3.1 Assumptions

In order to simplify mathematical analysis, nonlinearities in the updating function are eliminated by replacing continuous sigmoidal functions with discontinuous step functions. The rate equations that result from this approximation are in the form of piecewise linear differential equations. The n-dimensional phase space of a model may therefore be pictured as being divided by threshold hyperplanes into volumes corresponding to qualitative states of the system (spaces in which the system behaves in a qualitatively distinct way) (see Figure 9). Transitions between neighbouring qualitative states occurs whenever a solution starting in one region ends in another region. These systems have two types of steady states: regular steady states, lying within a volume and singular steady states, lying on one or more threshold planes between volumes.

One of the primary advantages of using differential equations to model updating functions is that time may incorporated in a continuous fashion. A disadvantage of this approach is that analytical methods frequently scale poorly, limiting analysis to small systems, the use of unrealistic simplifications, or the use of numerical simulation, which typically requires the introduction of some form of temporal discretization.

A common feature of these formalisms is their motivations is frequently to render the dynamics of complex regulatory systems tractable to mathematical analysis. As a result, they omit many complicating features of real biological systems, including time delays, spatial structure, sigmoidal activation and regulatory control of decay rates. A framework that allows regulatory mechanisms to be described more comprehensively has been proposed by Mestl and colleagues [85], however, this additional complexity limits the application of these techniques to relatively small systems.

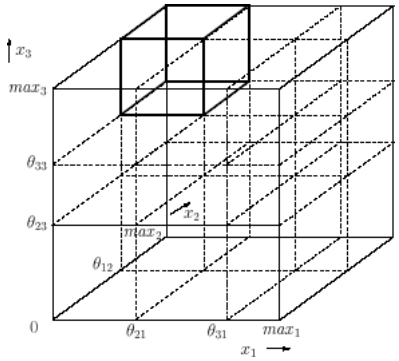

Figure 9: The phase space of a continuous logic model showing the volumes defined by the various activation thresholds.

4.3.2 Applications

Early usage of this formalism was restricted to analytical techniques [39] and ensemble approaches [49]. The mathematical complexity of the approach frequently limited application to only very small systems (two or three interacting genes). Recently, a streamlined qualitative simulation technique based on PLDE models has been proposed [34, 33] that is extendable to large systems and has been used to model sporulation in *B. subtilis* [32].

4.4 Further reading

The most comprehensive, if not always the most accessible, source of Kauffman's work on random Boolean networks as models for gene regulation is contained in *The Origins of Order* [70]. Some of his ideas are developed in a less formal, and more speculative, way in *At Home in the Universe* [71]. A readable overview of the analogy between basins of attraction and cell types is provided by Wuensche in [144]. The best starting point for a description of the generalised logic formalism is given in a pair of recent articles in *Chaos* [128, 127], which review and summarize many previous results. A good non-technical description of the theoretical findings from this research program can be found in [125]. Similarly, a comprehensive overview of the De Jong's approach to qualitative simulation, describing both the formalism and its relation to other similar approaches, is [33].

5 Continuous models

The modelling formalisms described in Section 4 above all share the assumption that state variables can be represented in a discrete fashion. In reality, while it is true that, at a given point in time, a gene is either being transcribed or

not being transcribed, levels of activation, rates of transcription and product concentrations can all vary in a continuous fashion. This section describes a number of approaches to modelling GRNs in which a continuous representation is used for state variables.

5.1 Ordinary Differential Equations

There is a long history of using systems of ordinary differential equations (ODEs) to model the reaction kinetics of regulatory systems. These approaches have several advantages. In principle, their more detailed representation of regulatory interactions provides a more accurate representation of the physical system under investigation. Additionally, there is a large body of dynamical systems theory that can be used to analyse such models. The primary disadvantage of ODE approaches is that they can be much more computationally intensive to analyse and solve than discrete models, especially for realistically sized systems.

5.1.1 Assumptions

Biological processes are almost inevitably highly complicated, and most mathematical models of gene regulation make two simplifying assumptions. The first of these is that the control of gene expression resides in the regulation of gene transcription. This assumption is known to be incorrect, as control may also be exercised at a number of other levels, including the post-transcriptional processing and translation of RNA and the control of RNA and protein degradation. While models have been developed that do investigate some of these processes, they are rarely integrated into a comprehensive framework. The second assumption is that genes are expressed and proteins produced at a continuous rate. Again, this assumption does not always hold. In some systems where the number of molecules involved is very small, the production and movement of individual molecules may be important, and there may be a degree of randomness. Stochastic approaches to modelling have been developed that reduce reliance upon this assumption, these are described in Section 6 below.

5.1.2 Model description

The basis for many ODE descriptions of regulatory systems is chemical rate equations, which describe the relationship between the rate of a reaction and the concentrations of the reactants. For example, consider a simple regulatory system in which a transcription factor X associates with an empty binding site Y_0 to give a bound site Y_1 at some rate k_1 and dissasociates at some rate k_{-1}. A bound site results in transcription and the production of a product P and an empty binding site (Y_0) at rate k_2. This system may be represented by the following rate equations:

$$X + Y_0 \quad \rightleftharpoons_{k_{-1}}^{k_1} \quad Y_1 \tag{2}$$

$$Y_1 \quad \rightarrow^{k_2} \quad P + Y_0 \tag{3}$$

which can then be translated into the following set of differential equations:

$$\frac{dx}{dt} = -k_1 x y_0 + k_{-1} y_1 \quad (4)$$

$$\frac{dy_0}{dt} = -k_1 x y_0 + k_{-1} y_1 + k_2 y_1 \quad (5)$$

$$\frac{dy_1}{dt} = k_1 x y_0 - k_{-1} y_1 - k_2 y_1 \quad (6)$$

$$\frac{dp}{dt} = k_2 y_1 \quad (7)$$

Introducing a number of assumptions: that the total number of bound and unbound sites is constant, $y_0 + y_1 = b$; and that the number of transcription factors is significantly higher than the number of binding sites, $x \gg b$, such that all of the binding sites will generally be occupied, this set of equations can be simplified to:

$$\frac{dx}{dt} = \frac{-K_{max} x}{k_n + x} \quad (8)$$

$$K_{max} = k_2 b \quad (9)$$

$$k_n = \frac{k_{-1} + k_2}{k_1} \quad (10)$$

These equations correspond to the Michaelis-Menten kinetic scheme and describe a situation where the rate of expression increases with transcription factor availability up to some limiting value [38] (see Figure - graph).

Early work investigating the existence and properties of various steady, periodic and chaotic solutions to these sets of equations has been summarized in [132]. The equations above can be generalised to a set of reaction-rate equations in which the concentration of a gene product is described in terms of the concentrations of the other elements of the system:

$$\frac{dx_i}{dt} = f_i(\mathbf{x}) \quad (11)$$

where $\mathbf{x}$ is the vector of gene product concentrations and f_i is an update function. This form of equation can be extended to include the influence of external input signals, product degradation and time delays:

$$\frac{dx_i}{dt} = f_i(\mathbf{x}, \mathbf{u}) \quad (12)$$

$$\frac{dx_i}{dt} = f_i(\mathbf{x}(t - \tau)) \quad (13)$$

$$\frac{dx_i}{dt} = f_i(\mathbf{x}) - \gamma_i x_i \quad (14)$$

where $\mathbf{u}$ is a vector of input signals, τ is a time delay and γ_i is the degradation rate of product i. Another possible extension is to model transcription and translation as indpendent processes, in which the production of messenger RNA depends upon the concentrations of protein transcription factors and the production of proteins depends on the concentrations of messenger RNAs:

$$\frac{dr_i}{dt} = f_i(\mathbf{p}) \tag{15}$$

$$\frac{dp_i}{dt} = g_i(\mathbf{r}) \tag{16}$$

where $\mathbf{p}$ and $\mathbf{r}$ are vectors of protein and mRNA concentrations respectively. In eukaryotic organisms, protein and mRNA are each produced in different cellular compartments and must be transported between them. An advantage of this approach is that it allows time delays due to mRNA and protein transport to be explicitly incorporated into a model [112].

A number of different functions have been used for f_i, the updating function. A common feature is their sigmoidal shape, which experimental evidence has suggested is plausible. Possibile updating functions include the hill curve and the logistic function, respectively:

$$f(x_j, \theta_{ij}, m) = \frac{{x_j}^m}{{x_j}^m + {\theta_{ij}}^m} \tag{17}$$

$$f(x_j, m) = \frac{1}{1 + e^{-mx_j}} \tag{18}$$

where $m > 0$ is a steepness parameter and $\theta_{ij} > 0$ is a threshold for the influence of x_j on x_i.

Due to nonlinearity of the updating functions, analytical solutions are not normally possible. In some cases, qualitative properties can be established, such as existance of steady styates, limit cycles and critical points [131]. The analysis of feedback dynamics carried out by Thomas [125] (described in Section 4.2) can be extended to continuous systems.

Another approach is to simplify the equations by replacing non-linear sigmoidal functions with step functions, or some other form of piecewise-linear function as described in Section 4.3 above.

Finally, it is sometimes possible to use numerical techniques to solve sets of equations. In numerical simulation, the exact solution of an equation is approximated by calculating values for each of the state variables at a series of discretised time steps. A number of systems have been characterised and solved in this manner, some of which are described below. A significant problem with the numerical approach is the lack of measurement of the various kinetic parameters in a system. The number of systems for which detailed parameter values are known is very small, and the size of most systems makes it unfeasible to obtain *in vitro* or *in vivo* measurements of many parameter values. Some

researchers have dealt with this problem by searching the parameter space of a system for combinations that allow the qualitative behaviour to be reproduced [138]. Another possible solution is to use the rapidly increasing amounts of available gene expression data to estimate parameter values, as described in the section on 'reverse engineering' below.

5.1.3 Applications

Model systems There exist only a small number of systems for which sufficient experimental data has been obtained to enable accurate models to be built. One of the best characterised systems is phage-λ [95]. This system has been the subject of a number of mathematical models [109, 102], including a hybrid model [82] and a stochastic model [9].

Other areas of modelling include the circadian clock [76], and the cell cycle [130]. Further models are reviewed in [57].

Reverse engineering of network structure Many of the approaches to modelling and simulation described above focused on either characterising a small, well-known regulatory system, or exploring the possible behaviour of a particular class of model networks. The relatively recent development of high-throughput experimental techniques in molecular biology has opened a new avenue of investigation. For the first time, there exists sufficient data to potentially enable network structure and dynamics to be inferred automatically with little or no *a priori* knowledge. DNA microrarrays can be used to generate thousands of measurements of gene expression levels during the course of a single experiment. The reverse engineering approach begins with the assumption that the interactions between these genes can be modelled as a network and aims to infer these interactions from the expression data.

Two of the main problems hampering reverse engineering efforts are the highly complex, combinatorial nature of the problem, and the relatively poor information content of the available data. Whereas microarrays are capable of collecting data on a large number of genes, the number of data points for each gene is typically very small. Furthermore the data is typically very noisy.

Both discrete and continuous modelling formalisms have been used for the task of network induction, as well as a number of different approaches to parameter learning. A recent overview of the different models and learning strategies used is provided by van Someren and colleagues [134].

Forward engineering of novel networks As some of the basic control modules in regulatory networks become more well understood, the construction of synthetic networks *in vitro* has become possible [56]. These novel networks not only have many potential therapeutic uses, they also allow understanding of regulatory processes to be refined. Systems constructed so far include a toggle switch [45] and an oscillator [41]. In addition, suites of networks have been created by randomly combining low level modules, allowing the combinatorial possibilities of synthetic networks to be explored [52].

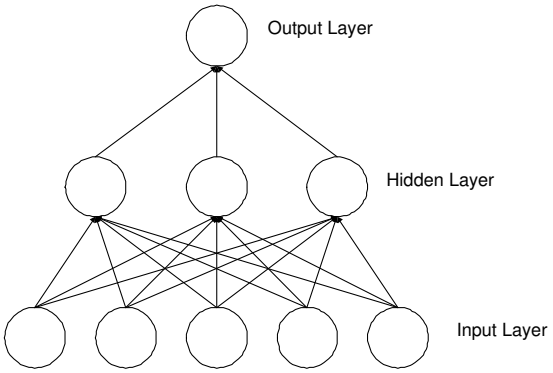

Figure 10: A standard feed-forward neural network. Nodes are divided into input nodes, output nodes and hidden nodes, which significantly increase the computational abilities of the network. This basic structure may be modified by changing the number of nodes in a layer, the number of layers and the arrangement of the links. In particular adding feedback links from the output layer back to the input layer allows the network to process temporal information, such as grammatical structure and patterns of gene expression.

The role of modelling in this process is to enable the behaviour of complex networks to be predicted via simulation before the circuit is implemented *in vitro*. Modelling formalisms with a high level of biological fidelity are therefore preferred, and several quantitative and semiquantitative approaches incorporating both deterministic and stochastic dynamics [65].

5.2 Neural network models

Artificial neural networks are mathematical models of information processing originally inspired by networks of neurons in the brain [58]. A neural network typically consists of a collection of nodes, some of which may be designated as input or output nodes, connected by weighted links (see Figure 10). Each node contains a transfer function that transforms a set of weighted input signals into an output signal. These networks can be trained to match particular patterns of activation via a number of learning processes.

Mathematically, it is possible to create a mapping between a neural network and a system of ODEs. Conceptually, a relatively straightforward analogy may be drawn between an information processing system in which the constituent elements are neurons and the links are synaptic interactions and a system in which the elements are genes and the links are regulatory interactions. Consequently, a number of researchers have used network architectures and concepts taken directly from neural networks and connectionist models [87, 136, 137].

The regulatory input to gene i is described as a the sum of the weighted inputs modified by the gene's activation threshold θ:

$$g_i = \sum_j w_{ij} y_j + \theta_i \tag{19}$$

where $w_i j$ is the strength of the regulatory interaction between genes i and j. A gene's level of activation is determined on the basis of this regulatory input and degradation:

$$\frac{dx_i}{dt} = \alpha_i f_i(g_i) - \gamma_i x_i \tag{20}$$

where α and γ are activation and degradation rates and f_i is a sigmoid transfer function as described above.

This type of formalism has been used in several different types of models. Mjolsness et. al. developed a phenomonological model of segmentation in the *Drosophila* blastoderm that used a neural network model to describe the internal dynamics of a cell as well as a generative grammar that described higher-level developmental processes such as cell division and differentiation [87]. This model has also been applied to other aspects of pattern formation and neurogenesis in *Drosophila* [100, 101, 79]. In these models, network parameters were trained such that the dynamics matched observed experimental behaviour.

Vohradský used a similar approach to model the lysis/lysogeney decision in phage λ [136, 137]. Here, the network structure is determined *a priori* from known interactions and the interaction weights are learned from experimental data. Several variations on the basic network are investigated, including connected networks and multi-compartment models, in which protein and RNA products are represented by separate network layers [137] (see Figure 11).

Neural network models have also been widely used in network inference (see the appropriate sections of [134] for a comprehensive review). In this domain, D'haeseleer has performed a comparison between the performance of network models using both linear and non-linear updating functions and obtained several analytical results [36].

5.3 Hybrid models

In the last decade, a number of models have been developed that take a hybrid approach to modelling gene regulatory networks. In these models, biochemical processes that are characterised by sharp thresholds are represented by Boolean elements, while genes whose activations vary more continuously with time, or for which intermediate levels of activation are significant, are modelled continuously. Early work in this direction was carried out by McAdams and Shapiro, who characterised the phage-λ circuit in terms of an electrical circuit, incorporating discrete and continuous elements, time delays and feedback dynamics [82].

A similar approach has been developed by Eric Davidson and colleagues, who have taken a strongly integrative approach to modelling the regulatory networks responsible for development [28, 29]. This work has ranged from detailed

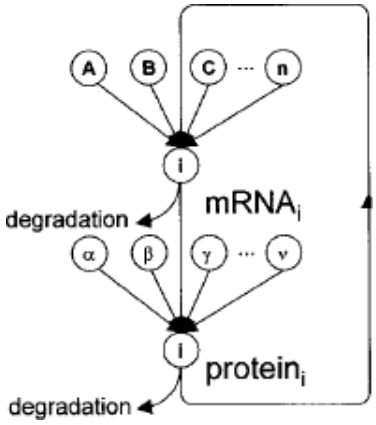

Figure 11: A two-compartment model of gene expression. Regulatory proteins $A, B, C, \ldots, n$ control the level of expression of gene i within the nucleus. The resulting mRNA, along with additional factors $\alpha, \beta, \gamma, \ldots, v$ controls the production of the corresponding protein i. Note also the presence of mRNA and protein degradation and an autoregulatory feedback loop. (From [137]).

characterisation of the logic underlying individual regulatory interactions [145] through to a network level view of regulatory dynamics [30]. One of the novel conceptual distinctions drawn in this approach is between the "view from the genome" and the "view from the nucleus" [10]. The former represents all possible regulatory interactions that genome encodes, while the latter restricts itself to those that are active in a particular cell at a particular time.

A key feature of this approach is the use of both continuous and Boolean functions, which result in a model lying somewhere between a continuous kinetic model and a Boolean model. The primary advantages of this level of abstraction is the clarity with which complex circuits may be represented, computationally simulated and empirically validated. The main cost associated with this approach is the loss of many of the analytical techniques that can be applied to more "pure" continuous or logical models. As the motivation for this work lies more in the direction of integrating and guiding experimental data, with less focus on abstract theoretical results, this trade-off is considered to be acceptable. An emphasis has been placed on the role of models as "the developmental biologists essential organizer for getting causal relationships between genes straight" [19].

5.4 Spatial models

Many of the models described above do not include any consideration of the physical space in which gene regulation is occurring. However, there are at least two possible situations in which spatial information may be important.

All regulatory events involve a physical interaction between molecules, some of which are present only in very small numbers and all of which are several orders of magnitude smaller than the size of a cell. Therefore a molecule may take some time, and require the assistance of some additional mechanism, before it is located in a position to act. This importance of localisation is compounded in eukaryotic cells, which have a complicated internal structure. As an example, while mRNA molecules are transcribed from DNA in the cell nucleus, they must be transported through the nucleur membrane and into the cytoplasm before they can be transcribed. While including the location and momentum of every single molecule would quickly become computationally infeasible, some models incorporate time delays to allow for the diffusion and transportation of molecules [111].

The second situation when spatial information may be required arises in models that incorporate interactions between cells. One of the most apparent distinctions between prokaryotic and eukaryotic organisms is that, while prokaryotes all consist of a single cell, a large majority of eukaryotes are multicellular. A human, for example, consists of around a trillion cells. Specifying the morphogenetic processes that transform a single cell into a complete organism requires a substantial increase in regulatory complexity. It also introduces several new issues related both to intercellular communication and to the mechanical processes of development, such as migration and cell adhesion.

Intercellular communication One of the simplest ways of implementing intercellular communication is to simply allow network connections to exist not only between elements within a cell, but also between elements in adjacent cells [138, 83].

Mechanisms of development Controlling the formation of spatial patterns during development presents a significant computational challenge. In addition to the the internal dynamics of the cell, external factors such as protein gradients and physical interactions between cells also play a role. One of the earliest mathematical attempts at modelling pattern formation was by Turing. His approach used a pair of coupled reaction-diffusion equations to describe a system consisting of two chemicals, known as morphogens (see [16] for a review). As the two morphogens diffuse across a spatial field and react with one another, a variety of patterns emerge, depeding on parameter values. One problem with this approach is the lack of any evidence for morphogens actually existing in a biological system. A gap therefore exists between the phenomenological description of the pattern formation process and the regulatory process that controls it at a genetic level.

The gene circuit approach of Mjolsness et. al. [87] mentioned above goes some way towards addressing this issue. The geometric aspect of the model uses a diffusion mechanism to describe communication between cells. Solé and Salazar-Ciudad also use a reaction-diffusion mechanism linked directly to a regulatory network to investigate developmental dynamics [104, 118]. Their model

is based on that of Mjolsness:

$$\frac{dx_{ij}}{dt} = f_j(\mathbf{x_i}) - \gamma_j x_{ij} + D_j \bigtriangledown^2 x_{ij} \quad (21)$$

where x_ij represents the concentration of gene product j in cell i, the first term specifies the production of x_{ij}, the second term its degradation, and the final term specifies the diffusion component, at rate D_j. The networks are connected together in a random fashion, inspired by the ensemble approach of random Boolean networks [69], and the behaviour of the networks under different parameter settings is explored. One finding of this study was that networks capable of producing spatial patterns such as gradients, stripes, spots and noise (chaos) are relatively common once a connectivity threshold is crossed [118].

5.5 Further reading

A good introduction to some of the mathematical modelling techniques used to studt biological systems is Leah Edelstein-Keshet's *Mathematical Models in Biology* [38]. Early results on the mathematical modelling and analysis of regulatory circuits are reviewed and extended in [132]. Some more recent results are reviewed by Smolen [112]. An overview of the gene network approach developed by Mjolsness and Marnellos, among others, can be found in Chapter 2 of *Modeling Neural Development* [80]. The hybrid approach to modelling developmental regulatory networks has been the subject of a number of recent reviews [30, 20, 19].

6 Stochastic models

6.1 Noise from within and without

As described in Section 2, gene activation is controlled by molecular signals, some of which are proteins producted by other transcription events. In general, genes are activated when the concentration of signal molecules crosses a threshold. Although many mathematical models make the simplifying assumption that proteins are produced at a continuous rate, evidence suggests they are actually produced stochastically in short "bursts" [81]. Therefore, the time taken for a concentration to reach it's critical threshold will vary stochastically. This variability time delay length can result in significant differences in the timing of similar events across an otherwise homogeneous population of cells. Due to the complex nature of interactions between regulatory elements, it is possible that individual cells may take different branches of regulatory pathway. Another potential source of stochasticity in the timing of events arises from the fact that, if a particular signal is represented by only a very small number of molecules, random molecular fluctuations may affect the time taken for a signal to be transferred [93].

Stochastic events in gene expression have several implications [97]. First, identical systems provided with similar inputs may produce different outputs as a result of stochastic elements of their regulatory mechanism. Second, it is likely that the evolution of gene regulatory networks has been driven in part by the requirement to produce deterministic outputs from a system constructed from noisy components operating in a noisy environment. While it is the case that in many situations regulatory systems are able to produce ordered results from chaotic starting points, in other instances, noise is exploited to the benefit of the system. The final implication is that deterministic modelling techniques may be insufficient to capture some of the dynamics of inherently noisy systems [75, 97].

Mechanisms by which the effects of noise may be diminished include negative and integral feedback (intensifying intermediate frequencies and dampening high and low frequencies), redundancy mechanisms and regulatory "checkpoints". In some systems, noise is amplified and used to generate heterogeneity in a population and hence increase diversity. Simulations have also found that complex systems involving many interacting feedback loops may be stabilised by noise [97]. In several studies [14, 138, 66], systems have been found whose robustness to noise appears to be a systemic product of network structure, rather than any explicit combinations of parameter settings or attenuating mechanisms.

6.2 Stochastic modelling approaches

Two main approaches have been developed to modelling stochastic events in gene expression, stochastic differential equations and the stochastic simulation algorithm. Stochastic differential equations extend the standard differential equation description of the reaction dynamics to include a noise term

$$\frac{dx_i}{dt} = f_i(x_i) + \nu_i(t) \tag{22}$$

where $\nu_i(t)$ is an additive noise term. This equation, known as the Langevin equation, can be developed into an alternative formulation that describes the evolution of the probability density function. These equations are generally too complex to be solved using analytic or numerical techniques, therefore a Monte-Carlo approach is generally used.

A characteristic of stochastic differential equation approaches is that they treat molecular concentrations as continuous variables. As mentioned above, in many situations signal molecules may exist in very small numbers, therefore it may be more appropriate to model them as discrete entities. An alternative approach formulates an equation in terms of the probability that a molecule undergoes a transition in a particular small time slice. This approach, known as the master equation, produces equations that are mathematically simple, but for realistic systems, are too numerous and too large to be feasibly solved. Again, the approach typically taken is to simulate the system a number of times and estimate a probability density function. A number of approaches to the stochastic simulation of such equations have been developed [47, 48, 89].

A major problem with both of these approaches is efficiency. Running multiple simulations of systems involving large numbers of reactions is computationally expensive. An important area of further research is the development of multiscale approaches. These models would be able to use continuous representations where individual events are not important, but still allow for the possible occurrence of rare, but significant, events [25].

In addition, a number of other methods for modelling stochasticity have been developed including stochastic Petri Nets [51] and stochastic neural networks [129]. Some of the logical modelling approaches described in Section 4, such as the asynchronous Boolean model and the generalised logic formalism, also include an degree of noises arising from the non-deterministic timing of regulatory events.

6.3 Further reading

Stochastic modelling formalisms tend to be of a higher level of mathematical sophistication than other approaches. A non-technical review of the important issues and techniques can be found in [97].

7 Network models

Most of the methods for the modelling and simulation of gene regulatory networks described above have taken a "bottom up" approach. They take the interactions between the indvidual elements of a system as their starting point and then observe the global behaviour that results when the system is solved or simulated. Recently, there has been an increasing amount of interest in a "top down" approach, focusing first and foremost on the genetic system as a network [90, 24]. The most basic feature of any network is its structure, the way in which individual elements are connected together. The structure of a network is constrained by the growth process that produced it and, in turn, constrains the possible dynamics of the system.

7.1 Small world and scale-free networks

Starting with a large set of elements, it is possible to connect them up in a number of different fashions. At one extreme, each element may be connected to its nearest spatial neighbours, leading to a network known as a regular lattice. At the other extreme, pairs of elements may be connected together at random, leading to the type of random networks investigated by Kauffman [67] (see Section 4). In the absence of any more detailed data on the architecture of biological networks, the random approach seemed reasonable. In addition, it allowed the use of a number of results from graph theory concerning the properties of random graphs. Real networks however, whether they be social networks, telecommunication networks or genetic networks, are not connected at random, and two recent models seem to offer a more representative model [64, 139].

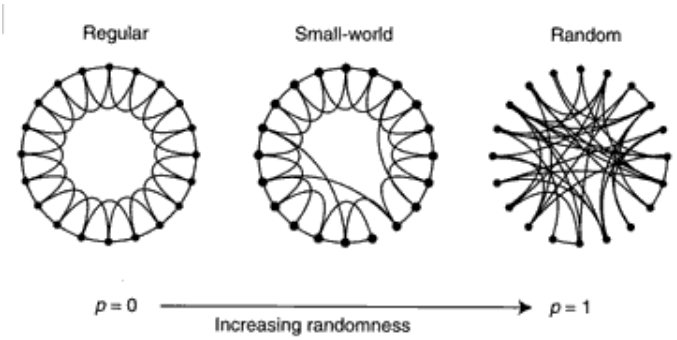

Figure 12: Regular, Small World and Random Networks. The networks are generated by starting with a regular lattice (left) and randomly rewiring a proportion, ρ of the links. When all of the links are rewired, the result is a random network (right). For intermediate values of ρ, the result is a small world network (centre), which is highly clustered, like a regular lattic, yet has a short distance between nodes, as in a random network. (From [141]).

Two related models, small world networks and scale-free networks have recently been proposed to address this gap. Small world networks [141, 140] are obtained when a small portion of the links in a random lattice are randomly rewired (see Figure 12). The resulting networks have two important properties. Firstly, the average distance between any two nodes in a network is very short, as it is in a randomly connected network. Secondly, the clustering coefficient, the number of nodes whose neighbours are also neighbours of each other, is high. It turns out that this model is a reasonable description of a number of naturally occurring networks [120, 116].

The closely related scale-free network model [13] is characterised by a power law decay in the probability of a node interacting with k other nodes, according to $P(k) \sim k^{-\gamma}$. This structural property arises from a growth dynamic termed preferential attachment. Under this dynamic, a network is grown from an initially small number of nodes by successively adding new nodes. A new node has a probability of being connected to each existing node dependant on the number of connections that node already has. Thus networks with a large number of connections are likely to attract more, whereas minimally connected nodes are more likely to stay that way.

An interesting property of scale free networks is their robustness to the failure of individual nodes. Because many of the nodes have very low connectivity, the random removal of any individual node is unlikely to fundamentally affect the structure of the network, providing a degree of robustness to error. On the other hand, the fact that some nodes act as 'hubs' and are connected to many other nodes, such networks may be particularly vulnerable to attacks that target highly connected nodes [5].

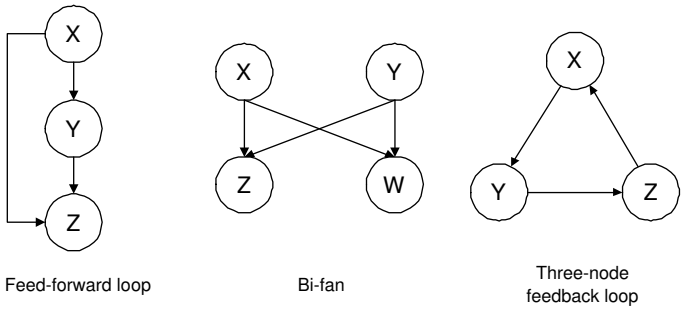

Figure 13: Examples of network motifs found in gene regulatory networks, electronic circuits, food webs and neural circuits. (Based on [86]).

One of the mechanisms by which the genome is hypothesized to have evolved is by gene duplication. Certain copying errors can result in a segment of the genome being duplicated and, because the duplicated segment encodes redundant information, it can subsequently diversify and possibly increase the functionality of the genome. The dynamics of this process have been modelled and demonstrated to show similar properties, such as response to failure and attack, to networks based on real data [117, 135].

7.2 Modularity, motifs and other structural features

The notion of clustering in a network model corresponds intuitively with the idea of functional modules in regulatory networks [90]. It has been suggested that functional modules are an important level at which to consider biological organisation for a number of reasons [54]. Modules involve a small fraction of network components working together in a relatively autonomous fashion and, as such, they represent a possible route to reducing the complexity of regulatory networks. Furthermore, empirical evidence suggests that such independent control substructures may actually exist [138]. It has also been suggested that functional modules may be one of the units on which evolution operates [54].

Several statistical properties of networks have been identified that provide a potential means if identifying and measuring modularity in systems. Ravasz et. al. describe a hierarchically structured network, in which small functional modules combine in a hierarchical fashion into progressively larger units [98]. They acknowledge however, that more work is still required to be able to accurately and usefully characterise modularity in network structure.

A second approach to the investigation of modularity in networks is the identification of "regulatory motifs", small, repeated patterns of interaction that occur with greater regularity than would be expected in a random network [86] (see Figure 13).

7.3 Further reading

Readable overviews of the small world and scale free network models have been produced by both Strogatz [120] and Solè [116]. Alternatively, more comprehensive, and significantly more technical, reviews of network structure [4] and network evolution [37] also exist.

References

[1] T. Akutsu, S. Miyano, and S. Kuhara. Identification of genetic networks from a small number of gene expression patterns under the boolean network model. In R. B. Altman, K. Lauderdale, A. K. Dunker, L. Hunter, and T. E. Klein, editors, *Pacific Symposium on Biocomputing '99*, pages 17–28, Singapore, 1999. World Scientific. [PDF].

[2] T. Akutsu, S. Miyano, and S. Kuhara. Algorithms for inferring qualitative models of biological networks. In R. B. Altman, A. K. Dunker, L. Hunter, K. Lauderdale, and T. E. Klein, editors, *Pacific Symposium on Biocomputing 2000*, pages 290–301, Singapore, 2000. World Scientific. [PDF].

[3] T. Akutsu, S. Miyano, and S. Kuhara. Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics*, 16:727–734, 2000. [Link].

[4] R. Albert and A.-L Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, 2002. [Link].

[5] R. Albert, H. Jeong, and A.-L Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000. [Link].

[6] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. *Molecular Biology of the Cell*. Garland Publishing Inc., New York, NY, 3rd edition, 1994.

[7] 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. [Link].

[8] R. B. Altman, A. K. Dunker, L. Hunter, and T. E. Klein, editors. *Pacific Symposium on Biocomputing '98*, Singapore, 1998. World Scientific.

[9] A. Arkin, J. Ross, and H. A. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics*, 149:1633–1648, 1998.

[10] M. I. Arnone and E. H. Davidson. The hardwiring of development: organization and function of genomic regulatory systems. *Development*, 124:1851–1864, 1997.

[11] R. J. Bagley and L. Glass. Counting and classifying attractors in high dimensional dynamical systems. *Journal of Theoretical Biology*, 183:269–284, 1996. [Link].

[12] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999. [Link].

[13] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[14] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387:913–917, 1997. [Link].

[15] U. Bastolla and G. Parisi. The modular structure of kauffman networks. *Physica D*, 115:203–218, 1998. [Link].

[16] R. Behrevan, P. J. Bentley, and R. Carlisle. Exploring reaction-diffusion and pattern formation. In *Proceedings of the First Australian Conference on Artificial Life (ACAL2003)*, pages 13–27. University of New South Wales, Australia, 2003.

[17] S. Bilke and F. Sjunnesson. Stability of the kauffman model. *Physical Review E*, 65:016129, 2002. [Link].

[18] J. W. Bodnar. Programming the *Drosophila* embryo. *Journal of Theoretical Biology*, 188:391–445, 1997. [Link].

[19] H. Bolouri and E. H. Davidson. Modeling dna sequence-based *cis*-regulatory gene networks. *Developmental Biology*, 246(1):2–13, 2002. [Link].

[20] H. Bolouri and E. H. Davidson. Modeling transcriptional regulatory networks. *BioEssays*, 24:1118–1129, 2002. [Link].

[21] S. Bornholdt and K. Sneppen. Neutral mutations and punctuated equilibrium in evolving genetic networks. *Physical Review Letters*, 81(1):236–239, 1998. [Link].

[22] S. Bornholdt and K. Sneppen. Robustness as an evolutionary principle. *Proceedings of the Royal Society of London, Series B*, 267:2281–2286, 2000.

[23] J. M. Bower and H. Bolouri. *Computational Modeling of Genetic and Biochemical Networks*. MITB, Cambridge, MA, 2001.

[24] D. Bray. Molecular networks: The top-down view. *Science*, 301:1864–1865, 2003. [Link].

[25] K. Burrage, T. Tian, and P. Burrage. A multi-scaled approach for simulating chemical reaction systems. Submitted to special issue of Progress in Biophysics and Molecular Biology. Preprint available [PS], 2004.

[26] S. B. Carroll, J. K. Grenier, and S. D. Weatherbee. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science, Oxford, 2001.

[27] E. Coen. *The Art of Genes: How Organisms Make Themselves*. Oxford University Press, Oxford, 1999.

[28] E. H. Davidson. *Genomic Regulatory Systems*. Academic Press, San Diego, CA, 2001.

[29] E. H. Davidson, D. R. McClay, and L. Hood. Regulatory gene networks and the properties of the developmental process. *Proceedings of the National Academy of Science, USA*, 100:1475–1480, 2003.

[30] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C.-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. Jun Pan, M. J. Schilstra, P. J. C. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri. A genomic regulatory network for development. *Science*, 295:1669–1678, 2002. [Link].

[31] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103, 2002. [Link].

[32] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *Bacillus subtilis*. *BMB*, 66:261–200, 2004. [Link].

[33] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualtiative simulation of genetic regulatory networks using piecewise-linear models. *Bulletin of Mathematical Biology*, 66:301–340, 2004. [Link].

[34] H. de Jong, M. Page, C. Hernandez, and J. Geiselmann. Qualitative simulation of genetic regulatory networks: Method and application. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI-01)*, pages 67–73, San Francisco, CA, 2000. Morgan Kauffman. [Link].

[35] 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. [Link].

[36] P. D'haeseleer. *Reconstructing Gene Networks from Large Scale Gene Expression Data*. PhD thesis, The University of New Mexico, Albuquerque, New Mexico, 2000.

[37] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51(4):1079–1187, 2002. [Link].

[38] L. Edelstein-Keshet. *Mathematical Models in Biology*. Random House, New York, NY, 1988.

[39] R. Edwards. Analysis of continuous-time switching networks. *Physica D*, 146:165–199, 2000.

[40] R. Edwards and L. Glass. Combinatorial explosion in model gene networks. *Chaos*, 10(3):691–704, 2000. [Link].

[41] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000. [Link].

[42] D. Endy and R. Brent. Modelling cellular behaviour. *Nature*, 409:391–395, 2001. [Link].

[43] J. J. Fox and C. C. Hill. From topology to dynamics in biochemical networks. *Chaos*, 11(4):809–815, 2001. [Link].

[44] S. A. Frank. Population and quantitative genetics of regulatory networks. *Journal of Theoretical Biology*, 197:281–294, 1999. [Link].

[45] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403:339–342, 2000. [Link].

[46] N. Geard and J. Wiles. Structure and dynamics of a gene network model incorporating small rnas. In Sarker et al. [107], pages 199–206.

[47] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25), 1977.

[48] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115:1716–1733, 2001.

[49] L. Glass and S. A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *Journal of Theoretical Biology*, 39:103–129, 1973.

[50] B. D. Gomperts, P. E. R. Tatham, and I. M. Kramer. *Signal Transduction*. Academic Press, San Diego, CA, 2002.

[51] P. J. E. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets. *Proceedings of the National Academy of Science, USA*, 95:6750–6755, 1998. [Link].

[52] C. C. Guet, M. B. Elowitz, W. Hsing, and S. Leibler. Combinatorial synthesis of genetic networks. *Science*, 296:1466–1470, 2002. [Link].

[53] S. E. Harris, B. K. Sawhill, A. Wuensche, and S. A. Kauffman. A model of transcriptional regulatory networks based on biases in the observed regulation rules. *Complexity*, 7(4):23–40, 2002. [Link].

[54] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402:C47–C52, 1999.

[55] 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. [Link].

[56] J. Hasty, D. McMillen, and J. J. Collins. Engineered gene circuits. *Nature*, 420:224–230, 2002. [Link].

[57] J. Hasty, D. McMillen, F. Isaacs, and J. J. Collins. Computational studies of gene regulatory networks: *In numero* molecular biology. *Nature Reviews Genetics*, 2:268–279, 2001.

[58] J. Hertz, A. Krogh, and R. G. Palmer. *Introduction to the Theory of Neural Computation*. Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Redwood City, CA, 1990.

[59] P. Hogeweg. Evolving mechanisms of morphogenesis: on the interplay between differential adhesion and cell differentiation. *Journal of Theoretical Biology*, 203(4):317–333, 2000. [Link].

[60] P. Hogeweg. Shapes in the shadow: Evolutionary dynamics of morphogenesis. *Artificial Life*, 6:85–101, 2000. [Link].

[61] P. Hogeweg. Computing an organism: on the interface between informatic and dynamic processes. *BioSystems*, 64:97–109, 2002. [Link].

[62] J. Holland. *Emergence*. Addison-Wesley, Redwood City, CA, 1998.

[63] S. Huang and D. E. Ingber. Shape-dependent control of cell growth, differentiation, and apoptosis: Swtiching between attractors in cell regulatory networks. *Experimental Cell Research*, 261:91–103, 2000. [Link].

[64] 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. [Link].

[65] M. Kærn, W. J. Blake, and J. J. Collins. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.*, 5:179–206, 2003. [Link].

[66] J. Kastner, J. Solomon, and S. Fraser. Modeling a *Hox* gene network *in silico* using a stochastic simulation algorithm. *Developmental Biology*, 246:122–131, 2002. [Link].

[67] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.

[68] S. A. Kauffman. Gene regulation networks: a theory for their global structure and behaviours. *Current Topics in Developmental Biology*, 6:145–182, 1971.

[69] S. A. Kauffman. The large-scale structure and dynamics of gene control circuits: An ensemble approach. *Journal of Theoretical Biology*, 44:167–190, 1974.

[70] S. A. Kauffman. *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, Oxford, UK, 1993.

[71] S. A. Kauffman. *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*. Penguin, London, UK, 1996.

[72] E. F. Keller. *The Century of the Gene*. Harvard University Press, Cambridge, MA, 2000.

[73] H. Kitano. *Foundations of Systems Biology*. The MIT Press/Bradford Books, Cambridge, MA, 2001.

[74] H. Kitano. Systems biology: a brief overview. *Science*, 295:1662–1664, 2002. [Link].

[75] M. Ko. A stochastic model for gene induction. *Journal of Theoretical Biology*, 153:181–194, 1991.

[76] J.-C. Leloup and A. Goldbeter. Toward a detailed computational model for the mammalian circadian clock. *Proceedings of the National Academy of Science, USA*, 100:7051–7056, 2003.

[77] B. Lewin. *Genes VII*. Oxford University Press, Oxford, UK, 2000.

[78] S. Liang, S. Fuhrman, and R. Somogyi. REVEAL, a general reverse engineering algorithm for inference of genetic network architectures. In Altman et al. [8], pages 18–29. [PDF].

[79] G. Marnellos and E. Mjolsness. A gene network approach to modeling early neurogenesis in *Drosophila*. In Altman et al. [8], pages 30–41. [PDF].

[80] G. Marnellos and E. D. Mjolsness. Gene network models and neural development. In A. van Ooyen, editor, *Modeling Neural Development*, pages 27–48. The MIT Press/Bradford Books, Cambridge, MA, 2003.

[81] H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Science, USA*, 94:814–819, 1997. [Link].

[82] H. H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269:650–656, 1995.

[83] E. Meir, G. von Dassow, and E. Munro. Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. *Current Biology*, 12:778–786, 2002. [Link].

[84] L. Mendoza, D. Thieffry, and E. R. Alvarez-Buylla. Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics*, 15(7):593–606, 1999. [Link].

[85] T. Mestl, E. Plahte, and S. W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176:291–300, 1995. [Link].

[86] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298:824–827, 2002. [Link].

[87] E. Mjolsness, D. H. Sharp, and J. Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152:429–453, 1991.

[88] A. Murray and T. Hunt. *The Cell Cycle: An Introduction*. Oxford University Press, New York, NY, 1993.

[89] N. Le Novere and T. S. Shimizu. STOCHSIM: modelling of stochastic biomolecular processes. *Bioinformatics*, 17:575–576, 2001.

[90] Z. N. Oltvai and A.-L. Barabási. Life's complexity pyramid. *Science*, 298:763–764, 2002. [Link].

[91] C. Oosawa and M. A. Savageau. Effects of alternative connectivity on behavior of randomly constructed Boolean networks. *Physica D*, 170:143–161, 2002. [Link].

[92] G. Orphanides and D. Reinberg. A unified theory of gene expression. *Cell*, 108:439–451, 2002. [Link].

[93] E. M. Ozbudak, M. Thattai, I. Kurster, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature Genetics*, 31:69–73, 2002.

[94] E. A. Di Paolo. Rhythmic and non-rhythmic attractors in asynchronous random boolean networks. *BioSystems*, 59(3):185–195, 2001. [Link].

[95] M. Ptashne. *A Genetic Switch: Phage Lambda and Higher Organisms*. Blackwell Science, Oxford, UK, 1992.

[96] L. Raeymaekers. Dynamics of boolean networks controlled by biologically meaningful functions. *Journal of Theoretical Biology*, 218:331–341, 2002. [Link].

[97] C. V. Rao, D. M. Wolf, and A. P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002. [Link].

[98] 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. [Link].

[99] 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. [PS].

[100] J. Reinitz, E. Mjolsness, and D. H. Sharp. Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. *Journal of Experimental Zoology (Mol. Dev. Evol.)*, 271:47–56, 1995.

[101] J. Reinitz and D. H. Sharp. Mechanism of *eve* stripe formation. *Mechanisms of Development*, 49:133–158, 1995.

[102] J. Reinitz and J. R. Vaisnys. Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity. *Journal of Theoretical Biology*, 145:295–318, 1990.

[103] P. Rohlfshagen and E. A. Di Paolo. The circular topology of rhythm in asynchronous random boolean networks. *BioSystems*, 73:141–152, 2004. [Link].

[104] I. Salazar-Ciudad, J. Garcia-Fernández, and R. V. Solé. Gene networks capable of pattern formation: From induction to reaction-diffusion. *Journal of Theoretical Biology*, 205:587–603, 2000.

[105] J. A. De Sales, M. L. Martins, and D. A. Stariolo. Cellular automata model for gene networks. *Physical Review E*, 55:3262–3270, 1997. [Link].

[106] L. Sánchez, J. van Helden, and D. Thieffry. Establishment of the dorso-ventral pattern during embryonic development of *Drosophila melanogaster*: A logical analysis. *Journal of Theoretical Biology*, 189:377–389, 1997.

[107] R. Sarker, R. Reynolds, H. Abbass, K.-C. Tan, R. McKay, D. Essam, and T. Gedeon, editors. *Proceedings of the IEEE 2003 Congress on Evolutionary Computation*, Piscataway, NJ, 2003. IEEE Press.

[108] C. R. Shalizi. Methods and techniques of complex systems science: An overview. In T. S. Deisboeck, J. Y. Kresh, and T. B. Kepler, editors, *Complex Systems Science in Biomedicine*. Kluwer, 2003. [Link].

[109] M. A. Shea and G. K. Ackers. The o_r control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of Molecular Biology*, 181:211–230, 1985.

[110] C. W. Smith and J. Valcarel. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends in Biochemical Science*, 25:381–388, 2000. [Link].

[111] P. Smolen, D. A. Baxter, and J. H. Byrne. Effects of macromolecular transport and stochastic fluctuations on the dynamics of genetic regulatory systems. *American Journal of Physiology*, 274:C777–C790, 1999.

[112] P. Smolen, D. A. Baxter, and J. H. Byrne. Modeling transcriptional control in gene networks - methods, recent results, and future directions. *Bulletin of Mathematical Biology*, 62:247–292, 2000. [Link].

[113] E. H. Snoussi. Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach. *Dynamics and Stability of Systems*, 4:189–207, 1989.

[114] E. H. Snoussi and R. Thomas. Logical identification of all steady states: The concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 55:973–991, 1993.

[115] J. E. S. Socolar and S. A. Kauffman. Scaling in ordered and critical random boolean networks. *Physical Review Letters*, 90(6):068702, 2003. [Link].

[116] R. Solé, R. F. Cancho, J. M. Montoya, and S. Valverde. Selection, tinkering and emergence in complex networks. *Complexity*, 8(1):20–33, 2002. [Link].

[117] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler. A model of large-scale proteome evolution. *Advances in Complex Systems*, 5(1):43–54, 2002.

[118] R. V. Solé, I. Salazar-Ciudad, and J. Garcia-Fernández. Common pattern formation, modularity and phase transitions in a gene network model of morphogenesis. *Physica A*, 305:640–654, 2002.

[119] R. Somogyi and C. A. Sniegoski. Modelling the complexity of of genetic networks: Understanding multigenic and pleiotropic regulation. *Complexity*, 1:45–63, 1996.

[120] S. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001. [Link].

[121] K. Struhl. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*, 98:1–4, 1999. [Link].

[122] D. Thieffry and R. Thomas. Dynamical behaviour of biological regulatory networks - II: Immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology*, 57:277–297, 1995.

[123] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42:563–585, 1973.

[124] R. Thomas. Regulatory networks seen as asynchronous automata: A logical description. *Journal of Theoretical Biology*, 153:1–23, 1991.

[125] R. Thomas. Laws for the dynamics of regulatory networks. *Int. J. Dev. Biol.*, 42:479–485, 1998. [Link].

[126] R. Thomas and M. Kaufman. Dynamical behaviour of biological regulatory networks - I: Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bulletin of Mathematical Biology*, 57:247–276, 1995.

[127] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. I. logical analysis of regulatory networks in terms of feedback circuits. *Chaos*, 11(1):180–195, 2001. [Link].

[128] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. II. structural conditions of multistationarity and other nontrivial behaviour. *Chaos*, 11(1):170–179, 2001. [Link].

[129] T. Tian and K. Burrage. Stochastic neural network models for gene regulatory networks. In Sarker et al. [107], pages 162–169. [PDF].

[130] J. J. Tyson. Models of cell cycle control in eukaryotes. *Journal of Biotechnology*, 71:239–244, 1999.

[131] J. J. Tyson, K. Chen, and B. Novak. Network dynamics and cell physiology. *Nature Reviews Molecular Cell Biology*, 2:908–916, 2001. [Link].

[132] J. J. Tyson and H. Othmer. The dynamics of feedback control circuits in biochemical pathways. *Progress in Theoretical Biology*, 5:1–62, 1978. [PDF].

[133] P. van Ham. How to deal with more than two levels. In R. Thomas, editor, *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, volume 29 of *Lecture Notes in Biomathematics*, pages 326–343. Springer-Verlage, Berlin, 1979.

[134] E. P. van Someren, L. F. A. Wessels, E. Backer, and M. J. T. Reinders. Genetic network modeling. *Pharmacogenomics*, 3(4):507–525, 2002.

[135] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani. Modelling of protein interaction networks. cond-mat/0108043, 2001.

[136] J. Vohradský. Neural model of the genetic network. *The Journal of Biological Chemistry*, 276(39):36168–36173, 2001. [Link].

[137] J. Vohradský. Neural network model of gene expression. *The FASEB Journal*, 15:846–854, 2001. [Link].

[138] G. von Dassow, E. Meir, E. M. Munro, and G. M. Odell. The segment polarity network is a robust developmental module. *Nature*, 406:188–192, 2000. [Link].

[139] A. Wagner and D. A. Fell. The small world inside large metabolic networks. *Proceedings of the Royal Society of London, Series B*, 268:1803–1810, 2001. [Link].

[140] D. J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton Univ. Press, Princeton, NJ, 1999.

[141] D. J. Watts and S. Strogatz. Collective dynamics of 'small world' networks. *Nature*, 393:440–442, 1998. [Link].

[142] L. Wolpert. *The Principles of Development*. Oxford University Press, Oxford, UK, 1998.

[143] A. Wuensche. The ghost in the machine: Basin of attraction fields of random boolean networks. In C. G. Langton, editor, *Artificial Life III*, Sante Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA, 1994.

[144] A. Wuensche. Genomic regulation modeled as a network with basins of attraction. In Altman et al. [8], pages 89–102. [Link].

[145] C.-H. Yuh, H. Bolouri, and E. H. Davidson. Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science*, 279:1896–1902, 1998. [Link].