Minireview

Modelling gene networks at different organisational levels

Thomas Schlitt*, Alvis Brazma

British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK
European Bioinformatics Institute, EMBL-EBI, Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK

Accepted 24 January 2005

Available online 14 February 2005

Edited by Robert Russell and Giulio Superti-Furga

Abstract **Approaches to modelling gene regulation networks can be categorized, according to increasing detail, as network parts lists, network topology models, network control logic models, or dynamic models. We discuss the current state of the art for each of these approaches. There is a gap between the parts list and topology models on one hand, and control logic and dynamic models on the other hand. The first two classes of models have reached a genome-wide scale, while for the other model classes high throughput technologies are yet to make a major impact.**

Keywords: Gene network; Transcription regulation network; Boolean network; Module; Model

1. Introduction

The term 'gene networks' is used to refer to a wide range of approaches describing complex interrelationships between genes or their products in biological systems. Different types of networks include metabolic networks, signalling networks, transcription regulation networks, and protein–protein interaction networks. Although, in a real biological system all these are closely interrelated, each type has characteristic features and elements, understanding of which is a necessary requirement for understanding the complete 'real world' biological network in a cell or any biological system. Recent advances in high-throughput technologies have opened new possibilities – some aspects of gene networks can now be studied on a genome-wide scale. We will discuss recent advances in gene network modelling, as well as the current limitations and future challenges. We will focus mainly on transcription regulation networks, though to a large extent the same principles are valid for a wide range of biological networks.

We can study gene networks of different sizes and on different level of detail. For instance, we can take a subnetwork consisting of only one gene regulating its own transcription via a feedback loop and describe this system by a set of differential equations capturing every known and hypothetical aspect of the system. Such an approach is often called reductionism and is based on the assumption that it is possible to consider the behaviour of a part of the whole system in isolation. At the opposite extreme, we can take the entire genome and build a network by connecting each transcription factor to the genes it is known to be regulating. Such an approach will not help us in learning much about any particular gene, but it can help in understanding the general properties of the network, such as finding if relatively isolated components (modules) are present, and identifying them. This is a reductionist approach too: complementary to considering one or few genes in isolation, we study specific aspects of the system, but ignore many details. Both approaches can be called modelling or model building. Wide ranges of types of models have been used to study gene regulation. We can distinguish between at least four different levels (layers) of increasing detail in these models:

(i) *Parts lists* (Fig. 1) – the collection, description and systematisation of network elements in a particular organism or a particular biological system (e.g., transcription factors, promoters, and transcription factor binding sites).
(ii) *Topology models* (or *wiring diagrams*) (Fig. 2) – the connection (interaction) diagram between the parts; this can be viewed as a graph where nodes represent genes, while edges or arcs (connections between nodes, which can be directed or undirected) represent different interactions.
(iii) *Control logic models* (Fig. 3) – the description of the combinatorial (synergetic or interfering) effects of regulatory signals – e.g., which transcription factor combinations activate and which repress the transcription of the gene.
(iv) *Dynamic models* (Fig. 4) – the simulation of the real-time behaviour of the network and the prediction of its response to various environmental changes, external, or internal stimuli.

Each next level adds more detail and is more complex for a network with the same number of parts. The current state of the art limits the size of the networks that we are able to model at each particular level. As the result of the genome projects studies, network parts lists have reached the genome scale, though it should be noted that in most genomes the functions of at least a third of the genes are unknown, and it is likely that many of the transcription regulators are also unknown. The network topology is studied on a genome scale for smaller genomes, such as yeast or bacteria (e.g., see [1]). To our knowledge, the largest networks that have been described on the control logics level consist of tens of genes – an example of this is a developmental network for sea urchin [2]. One of the largest dynamic models of biological networks that we are aware of, uses 13 differential equations to describe the role of 5 genes

*Corresponding author.
E-mail address: tsc@bas.ac.uk (T. Schlitt).

0014-5793/$30.00
doi:10.1016/j.febslet.2005.01.073

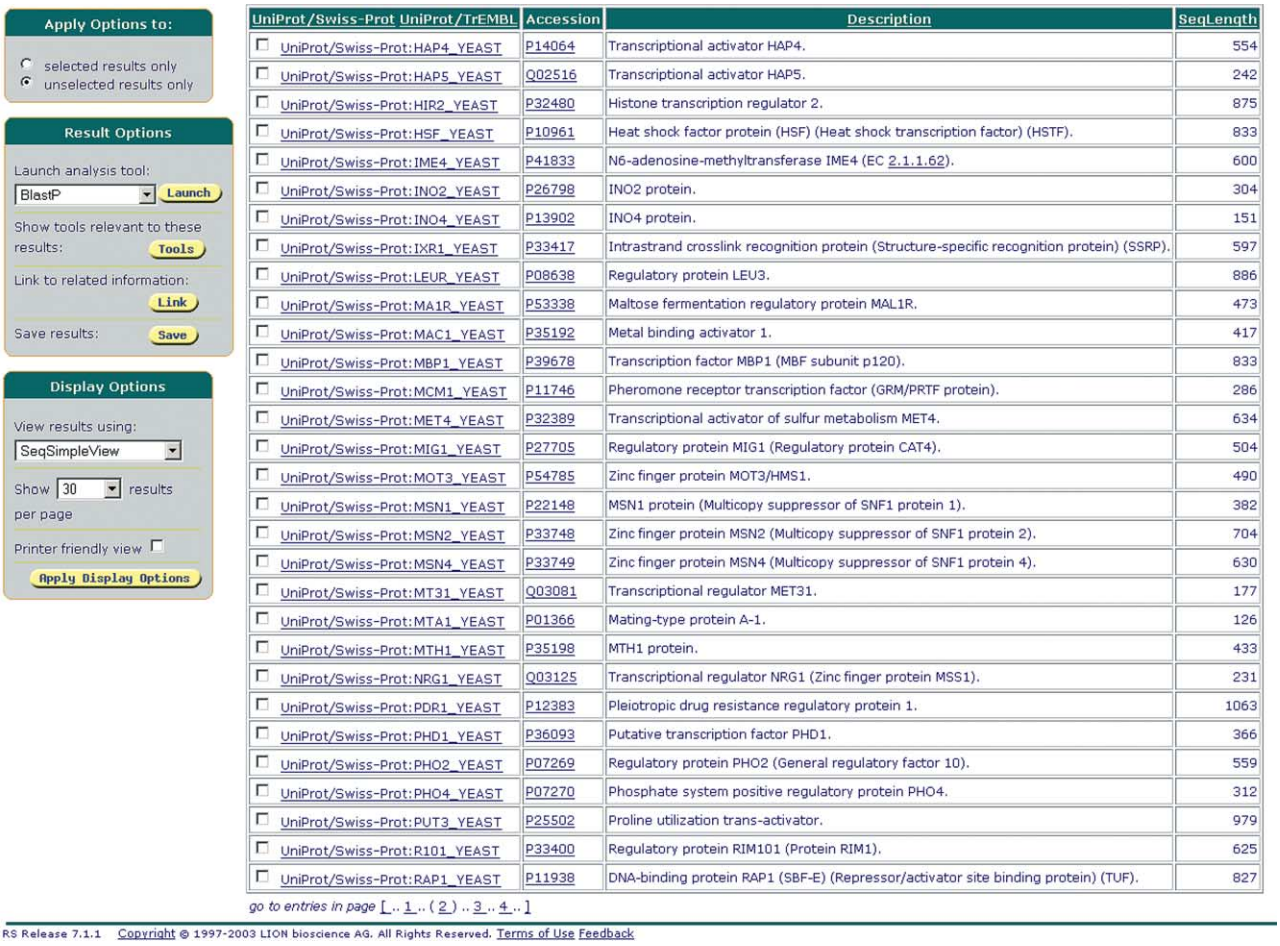


Fig. 1. Parts lists. Existing databases provide information about individual genes and proteins, and whole proteomes. This figure shows the result of a UniProt (www.ebi.ac.uk/uniprot) query for all yeast transcription factors used by Lee et al. [1].

and their products during early embryo development in *Drosophila* [3].

2. Organisational levels of gene network models

2.1. Parts list

Compiling the parts list of the constituent elements is a natural first step in developing any model of some complexity. This can be viewed as building a database of regulatory elements, or as building an ontology of transcription regulation processes (Fig. 1). TRANSFAC is arguably the best-known database of transcription factors [4]. Gene Ontology (GO) contains over 7500 different terms describing biological process 'transcription', including over 6500 terms under process 'regulation of transcription' [5]. Such parts lists can be used to compare different organisms to obtain the indication of the complexity of transcriptional machinery, e.g., [6]. For instance, the number of known and predicted transcriptional regulators in eukaryotic organisms varies from about 300 in yeast, to about 1000 in humans (see Table 1). Babu et al. [7] analysed the domains and protein families of the transcription factors and regulated genes in *Escherichia coli*. They found that many groups of transcription factors have identical domain architectures, and they conclude that roughly threequarters of the transcription factors have arisen as a consequence of gene duplication. In contrast, they found little evidence of duplication of regulatory regions together with regulated genes or of transcription factors together with regulated genes.

Many publications address the identification of transcription factor binding sites, for instance by analysis of promoter sequences of coexpressed genes [8]. However, the identification of regulatory elements in DNA by computational means has turned out to be rather elusive for genomes more complex than that of yeast. Some studies have, therefore, focused on the analysis of higher-level organisation of transcription factor binding sites in promoters, such as frequently occurring combinations of known binding sites [9,10], or restricted the search for regulatory elements to conserved sequence regions, which are identified by genome comparisons [11–13]. Lee et al. [1] identified transcription factor binding sites experimentally in yeast for 106 transcription factors using the ChIP(chromatin immuno-precipitation)-on-chip technology, a chromatin immunoprecipiation technique, which utilizes genomic microarrays (chips) to identify the DNA fragments bound by transcription factors. More recently, Harbison et al. [14] extended this study to over 200 transcription factors by combining information from ChIP-on-chip experiments with phylogenetically conserved sequences, previously published evidence. Once we have identified the transcription factor that

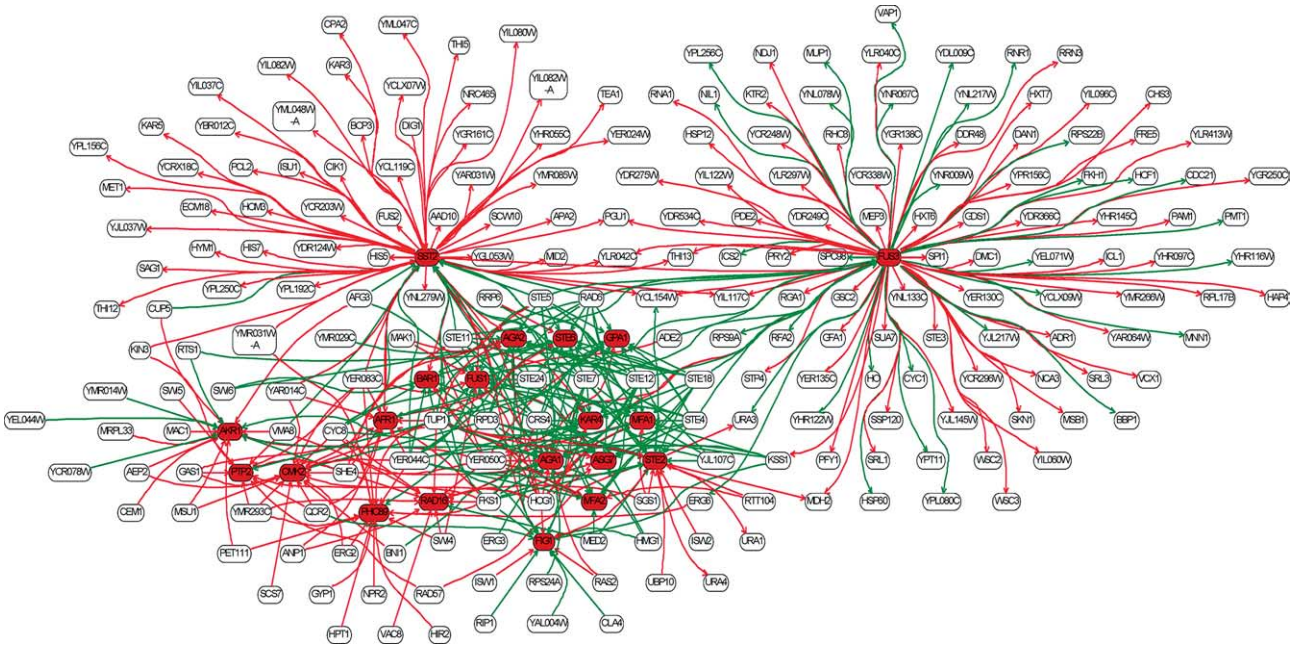


Fig. 2. Topology models. Mutant network according to Rung et al. [18]. This network is based on a microarray dataset of 270 yeast strains, each carrying a single gene deletion [16]. The deleted genes are connected to the genes, which show significant upregulation (green), or downregulation (red) of the wild type yeast strain compared to the respective mutant strain. This figure shows only part of the full network; only the genes highlighted in red (pheromone response genes) and their direct neighbours were selected. For more details see [18].

binds to a particular binding site, we effectively have information about the network topology.

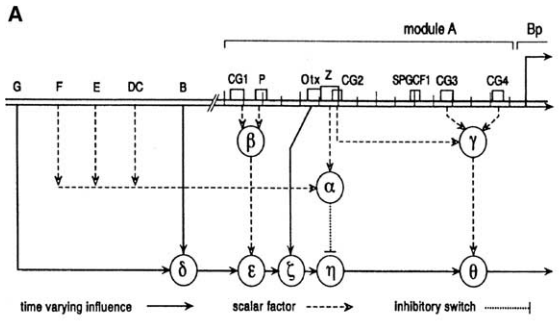
2.2. Topology models

The gene network topology (or wiring diagram) can be represented as a graph; where nodes represent genes, while edges or arcs denote the interactions between the respective genes (Fig. 2) [15]. To define a specific model, we need to define the meaning that we assign to the connections. For instance, an arc from a gene *A* to *B* may mean that gene *A* is a transcription factor, which is known to bind to the promoter of gene B. A rather different network will be obtained, if an arc from *A* to *B* denotes the observation that the disruption (e.g., mutation) of gene *A* changes the expression of *B*. In the first model we describe physical interactions, but it does not tell us anything about the effects of these interactions. The second model is similar to the one used in gene networks built by classical genetics means – we know that a mutation (perturbation) of the first gene has an effect on the second one, but it does not necessarily mean a direct physical interaction – there may be a long transcriptional or signalling cascade leading from the first gene to the second. An important question is how these two relate.

Recent advances in microarray technologies have been used to generate both types of networks. In their ChIP-on-chip experiments Lee et al. [1] observed nearly 4000 interactions for 106 transcription factors (*P*-value equal to 0.001). The promoter regions of over 2300 of the about 6000 yeast genes were bound by at least one of the studied transcription factors. In 30 promoter regions, 10 or more transcription factors were found. Relative to randomised networks, a disproportional high number of promoter regions were bound by four or more regulators. The number of different promoter regions bound by each transcription factor ranges from 0 to 181, with an average of 38. (Note that the particular numbers depend on the chosen *P* value.)

Hughes et al. [16] used whole genome gene expression microarrays to study the effect of mutations of about 300 genes. Comparisons between the networks of Lee et al. [1] and Hughes et al. [16] show that not all physical interactions (reported by the particular experimental techniques) result in significant functional effects, and that the relationships between these two models are quite complex (Schlitt and Brazma, unpublished data). Both networks seem to have roughly so-called scale-free topology [17]; there are *hubs* (genes with many connections) in these networks, while most genes have low connectivity ([18] and unpublished data).

Why does the gene network topology matter? First of all it tells us which gene products may interact with each other and which are mutually independent, which is important information when building models of increasing levels of complexity. Probably, the most important question is if we can find *modules*, i.e., subnetworks that are relatively isolated from the rest of the network. If such modules are found, they can help us to use the reductionist approach later on by allowing modelling the parts of the network independently on a more detailed level (e.g., by a dynamic model). The existence of modules in biological systems has often been taken as an axiom [19]. However, a precise definition for what constitutes a module is elusive, and, therefore, this term has been used in various contexts. In a graph representation, it is natural to define a module as a 'relatively' isolated component, and indeed such components were found in protein–protein interaction networks. In contrast, isolated components have hitherto not been found in the wiring diagrams of eukaryotic transcription regulation networks [18]. Several methods have been proposed to identify modules as groups of genes coexpressed under specific conditions [20,21], still there remain controversial



B

if (F = 1 or E = 1 or CD = 1) and (Z = 1)	Repression functions of modules F, E, and DC mediated by Z site
α = 1	
else α = 0	
if (P = 1 and CG_1 = 1)	Both P and CG_1 needed for synergistic link with module B
β = 2	
else β = 0	
if (CG_2 = 1 and CG_3 = 1 and CG_4 = 1)	Final step up of system output
γ = 2	
else γ = 1	
δ(t) = B(t) + G(t)	Positive input from modules B and G
ε(t) = β*δ(t)	Synergistic amplification of module B output by CG_1-P subsystem
if (ε(t) = 0)	Switch determining whether Otx site in module A, or upstream modules (i.e., mainly module B), will control level of activity
ξ(t) = Otx(t)	
else ξ(t) = ε(t)	
if (α = 1)	Repression function inoperative in endoderm but blocks activity elsewhere
η(t) = 0	
else η(t) = ξ(t)	
Θ(t) = γ*η(t)	Final output communicated to BTA

Fig. 3. Control logics models. Computational model for module A of the promoter for the *Endo*16 gene in sea urchin according to Yuh et al. [39]. *Endo*16 is a gene that encodes a polyfunctional secreted protein of the midgut in the late embryo and larva. (A) Schematic diagram of interrelations and functions. Interrelations between upstream modules (G to B) and specific module A target sites demonstrated experimentally, and among the module A target sites, are indicated beneath the line representing the DNA. Each circle or node represents the locus in the system of a specific quantitative operation, conditional on the state of the system; operations are specified for all relevant states in (B). Operations at each node are carried out on inputs designated by the arrows incident on each circle, and produce outputs designated by arrows emergent from each circle. Open arrowheads indicate inputs to the indicated node that are constant through time, the values of which are specified according to the logic sequence in (B); closed arrowheads indicate time-varying inputs. The terminated bar indicates a Boolean repression function that under given conditions extinguishes activity at node h. (B) Logic sequence for operation of model shown in (A). The value 0 denotes that a given site or module site has been mutationally destroyed or is inactive because its factor (or factors) is missing or inactive; the value 1 indicates that the site or module is present and productively occupied by its cognate transcription factor. For the case of modules F, E, and DC, a Boolean representation is chosen because ectopic expression is essentially zero (beyond technical background) in ectoderm and mesenchyme when these modules (together with module A) are present in the construct (5); otherwise, ectopic expression occurs. Sites within module A are designated as above. The logic sequence specifies the values attained at each operation locus [circles in (A)], either as constants determined experimentally and conditional on the state of the relevant portions of the system, or in terms of time-varying, continuous inputs designated by the symbol (t). The constants are based on experimental evidence; for details of the model see [39]. Reprinted with permission from Yuh et al. Science 279, 1896–1902 [39]. Copyright 1998 AAAS.

opinions regarding the existence and nature of modules in gene networks [22,23].

The topology and dynamics of gene regulation networks are not entirely independent. ChIP-on-chip experiments demonstrate that the localisation of a number of transcription factors – and thus the network topology – depend on the experimental conditions [14,24–26]. Therefore, one should be cautious in treating ChIP-on-chip based network topology as a static feature (e.g., [27]). For protein–protein interaction networks a dynamic component may help in identifying functional modules: Han et al. [28] showed that hub proteins can be divided into two groups, based on the level of coexpression with their neighbours in the network. Hubs, which are less tightly coexpressed, seem to link functionally separate modules and removing these hubs leads to more rapid disintegration of the network [28]. However, so far this has not been observed to happen in eukaryotic transcription networks.

There are many questions that can be addressed at the topology level. For instance, the number of connections can indicate to which functional class a gene belongs [18], and it is possible to identify functionally related genes by comparing neighbourhoods of genes in network graphs [29]. Manke et al. [30] found directly interacting transcription factors and those, which are members of a protein complex, to occur more likely together as putative DNA-binding modules. It has been proposed that the existence of hubs in a network might make these networks more tolerant to random failure of network elements [31,32]. Lee et al. [1] and Milo et al. [33] identified reoccurring structural elements (motifs) in the networks, such as feed-forward and feedback loops. These motifs may partly be the result of gene duplications during genome evolution [34].

2.3. Control logics models

The network topology tells us which genes in the network depend on which other genes, but it does not tell us anything about the regulatory effects of these dependencies, i.e., about the control logics. For instance, some transcription factors act as inhibitors and some as activators; moreover, promoters may consist of tens of binding sites for many different transcription factors. Combinatorial effects between the transcription factors bound to a promoter are non-trivial and sometimes have to be described by quite complex algorithms. Linear functions, Boolean functions (AND, OR, NOT and combinations of these), decision trees, and Bayesian probability distributions have all been used to describe the network logic. We can distinguish between discrete control functions,

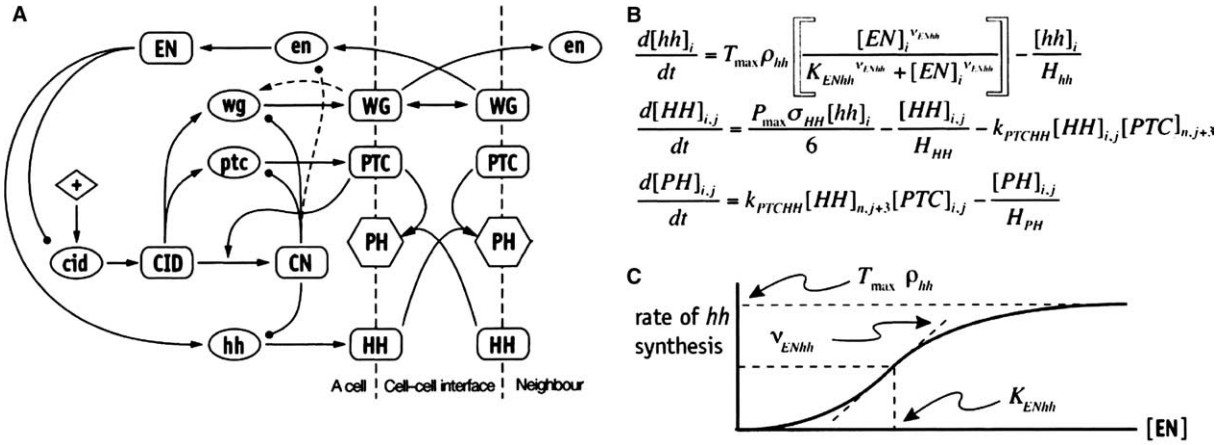


Fig. 4. Dynamic models. A continuous dynamical model of the segment polarity gene network in *Drosophila melanogaster* according to von Dassow et al. [3]. Interactions among products of the five genes in the model: WG, wingless; EN, engrailed; HH, hedgehog; CID, cubitus interruptus (whole protein); CN, repressor fragment of cubitus interruptus; PTC, patched; PH, patched hedgehog complex. Dashed lines were added according to the insufficiencies of the bold lines alone. Ellipses, mRNAs; rectangles, proteins; arrows, positive interactions; circles, negative interactions. cid is basally expressed (+ in rhombus). (B) Examples of differential equations constituting the model. These simplified dimensional-form equations govern dynamics of hedgehog mRNA. (C) Simple doseresponse curve governing transcriptional activation (brackets in B), illustrating parameterization of the model. Transcription rate saturates because of inherent limits on how fast RNA polymerase can move (T_{max}) multiplied by a gene-specific efficiency parameter (ρ_{hh}). For every monotonic regulator there is some concentration at which it has a halfmaximal effect on its target (κ_{ENhh}). Each such interaction may exhibit nonlinearity (ν_{ENhh}). In the case of cooperative binding, n is equivalent to a Hill coefficient. For details of the model see [3]. Figure reproduced from [3], with permission from Nature Publishing Group.

Table 1
Number of transcription regulators in different organisms

Organism	Number of genes	Number of transcription regulators
Yeast	6682	312
Fly	13525	492
Human	22287	1034

The number of genes and transcriptional regulators (genes annotated with GO term GO:0030528 "transcription regulator activity" for yeast (*Saccharomyces cerevisiae*) was taken from SGD (www.db.yeastgenome.org/cgibin/SGD/search/featureSearch) and for fly (*Drosophila melanogaster, DROM3*) and human (*Homo sapiens, NCBI 34 dbSNP120*) was taken from ENSEMBL (www.ensembl.org/Multi/martview) (on 13.07.2004).

which are based on the assumption that a gene can be in a finite number of states (e.g., *expressed*, or *not expressed*), and continuous control functions, where the expression level can be characterized by a real value in a certain range. Kauffman introduced the notion of *canalizing function* – a Boolean function that has at least one input variable and one value (0 or 1), which determines the value of the output of the function, regardless of other variables [35]. He hypothesized that genes are predominantly controlled by such functions. Others (e.g., [36]) have used *additive functions*, where output can be expressed as a sum of inputs each taken with a particular weight (which can be positive or negative). There are other examples of control logic based network analysis. Soinov et al. [37] used a supervised learning approach to build decision-tree-related classifiers, which predict gene expression from the expression data of other genes. Segal et al. [20,38] applied a learning procedure based on probabilistic graphical models to networks consisting of groups of coregulated genes.

Although few promoters have been studied in great detail, there are excellent examples, such as the description of the promoter action logics of sea urchin developmental gene *Endo16* [39]. The *Endo16* promoter consists of almost 30 regulatory elements stretched over a region of 2.3 kb. Based on experimental data collected using modified promoter constructs, Davidson and co-workers constructed a model expressed as an algorithm combining Boolean and linear functions (Fig. 3). This algorithm takes as an input the occupancy information from 12 binding sites and outputs a value, that 'can be thought of as the factor by which, at any point of time, the endogenous transcription activity (...) is multiplied as a result of the interactions mediated by the *cis*-regulatory control system' [39]. Predictions of promoter manipulations based on this model have largely been confirmed in subsequent experiments. Extending their earlier work the group of Davidson compiled a regulatory network containing over 40 genes by the construction of a model that integrates extensive experimental evidence on early development of sea urchin embryos [40].

2.4. Dynamic models

Compared to the approaches above, the dynamic models can be described as 'classical' approaches to gene network modelling, as many of them have been developed and studied long before the current *genome era*. They aim at describing and often simulating the dynamic changes in the state of the system and try to predict the network's response to various environmental changes and stimuli.

The simplest dynamic models – *synchronous Boolean network models* – are based on the assumption that (a) the genes in the network can be in one of two states: *expressed*–1, or not expressed 0; (b) the network control logic can be described by Boolean functions and (c) the genes in the network switches from a state to state in a synchronous manner at regular intervals (discrete time-points) depending on the values of the Boolean functions at the previous time-point (e.g., [41,42]). We can introduce the concept of the *state of the network* – an *n*-tuple of 0 s and 1 s describing which genes in the network are expressed and which are not at a particular moment (for instance, for a three gene network the possible states are (0,0,0), (0,0,1),...,

(1,1,1), the state (0,0,0) means that no genes are expressed - it can be called 'death'). As time progresses, the network navigates through the 'state space', switching from one state to another. For a network of n genes, in total there are 2^n different states. However, this does not mean that all 2^n states are possible for a particular network. One can ask the question – how many of the states are possible biologically, and which ones?

In 1969, Kauffman [43] studied the properties of random Boolean networks. He found that under certain assumptions about the network topology – the number of incoming connections at each node are bound by a certain constant – and logics – promoters are predominantly controlled by canalizing functions – there are only a small number of states, in which the network will be for most of the time. These states are called *attractors*; any other state, if possible at all, will lead to an attractor state in a relatively small number of steps. Moreover, the system either reaches a steady state or fluctuates between the attractor states in a regular fashion. Kauffman hypothesized that attractors correspond to different cell types of an organism. The number of cell types predicted by this model corresponds well with our current knowledge [35].

Although Boolean networks can reveal important network properties, generally they are considered to be too crude to capture many important aspects of network dynamics. Difference and differential equations allow more detailed descriptions of network dynamics [44,45], by explicitly modelling the concentration changes of molecules over time. For instance, von Dassow et al. [3] constructed a system of differential equations describing the activity of *Drosophila* segment polarity genes and pattern formation (Fig. 4). Their system included 48 parameters, such as the half-lives of messenger RNAs and proteins, binding ranges and cooperativity coefficients. The initial model described all known interactions, but it also revealed that the additions of at least two new hypothetical interactions were needed to ensure that the behaviour of the model was consistent with the observations.

Although predictions of small models have been successfully tested experimentally using specifically engineered control circuits, such as feedforward loops [46] and feedback loops [47–50], models usually depend on a large number of numerical parameters, which are difficult to estimate experimentally. Therefore, an important question for these models is the *stability* does the behaviour of the system depend on the exact values of these parameters, or is it similar for different variations. It seems unlikely that an unstable system represents a biologically realistic model, while on the other hand, if the system is stable, the exact values of the parameters are not so crucial (this does not exclude the possibility of phase transitions, where some threshold values determine the future development path of the system). For instance, the above-mentioned *Drosophila* developmental model [3] is stable it tolerates tenfold or more variation in the values of most individual parameters.

In the real world systems both continuous aspects and discrete aspects are present. In general, concentrations are expressed as continuous values, whereas the binding of a transcription factor to DNA is expressed as a discrete event (bound or unbound). However, on single cell level the concentrations may have to be expressed by molecule counts and become discrete, whereas if we use thermodynamic equilibrium to model the protein–DNA binding, the variable describing the binding state becomes continuous. Hybrid models have been developed in an attempt to describe both, discrete and continuous aspects, in one model. One example is the phage λ model by McAdams and Shapiro [51], where elements similar to ones used to describe electronic circuits have been exploited. A very simple model combining the discrete and continuous aspects of gene networks called the Finite State Linear Model, was introduced in [52].

The network models mentioned so far are all deterministic – they assume that the next state of the system is determined by the current state and the inputs. However, in real world systems stochastic effects may play an important role. For instance, for some genes in yeast the number of mRNA molecules is close to one copy per cell [53]. This means that it is likely that there is a considerable intrinsic noise element present some cells apparently have more mRNA molecules of the given species present than others. Thus modelling a cell by using continuous concentrations effectively means modelling an ensemble of cells by mean values of stochastic variables. It is not obvious to what extent this is possible. It has been demonstrated that the stochastic effects are important for the phage λ switch decision between lysis and lysogeny [54]. Lately, experimental studies have tried to measure the level of intrinsic noise in eukaryotic cells (e.g., [55,56]).

3. Future challenges

In essence, network modelling is the realisation and acceptance that a model describes only some properties of the 'real world' system, and ignores others. A rigorously defined model can be studied independently from the 'real world' network, but the ultimate test of its usefulness is in the prediction of system properties that can be tested in experiments performed on the respective 'real world' biological system.

One cannot fail to notice the gap between the models describing the network topology on one hand, and network logics or dynamics on the other hand – the first are approaching the whole genome scale, while the second are typically modelling a handful of genes. High-throughput experiments, most notably microarrays, provide us with temporal information about transcriptional processes in time series experiments. These have been used to study control logics as well as some dynamics aspects of transcription regulation in processes such as the cell cycle [57–59], stress response [60,61], or galactose utilization [62]. Although these studies have produced valuable observations and hypotheses, they have not yet yielded large-scale models predicting the behaviour of systems that can be rigorously tested in experiments. As far as rigorous model building is concerned, high-throughput technologies have yet to have a direct impact beyond the network topology-level studies.

Can the gap between the number of genes in dynamic network models and in genomes be bridged? Is rigorous modelling of gene regulation network dynamics possible at all on the whole genome scale? We think that the answer largely depends on the robustness and modularity properties of the real world biological networks. To what extent can the transcription regulation networks be decoupled from other networks, such as signal transduction networks? And to what extent can specific processes be decoupled from each other? How much do the exact quantitative values, such as substance concentrations, matter in determining the more general patterns of system behaviour, such as cell differentiation?

If we can find modules – units behaving independently of each other – it would be possible to build the complete model as a set of modules. If the exact values of parameters and substance concentration are not crucial, we can hope to describe and predict the states of the system in a simpler and more robust way.

The belief that real world biological networks 'must be' robust and 'must be' modular is quite popular. However precise definitions of biological robustness and modularity and, moreover, the proofs of their presence remain elusive. The principles of modularity and robustness used in engineering are sometimes given as a reason that the same must be true in biological systems, but there are many examples when the 'designs' in nature, which are obtained by natural selection are different from the designs one would use in engineering. There are other arguments why biological networks could be modular, such as reuse of the components after genome duplications, but they are no proofs.

Nevertheless there are numerous indications that, on the dynamic level, network modules exist. For instance, cell growth can be decoupled from cell cycle in yeast (e.g., [63]), indicating that to some extent independent modules control these two processes. Similarly, the above-mentioned example of the *Drosophila* developmental network indicates that the exact values of the model parameters may not be crucial in large-scale systems behaviour. If we are not interested in predicting the exact concentrations of different substances, but only in the patterns of the systems behaviour such as steady states, we can often use simplified Boolean-type networks instead of differential equations [42].

Whether rigorous modelling of gene network dynamics is possible on a genomic scale remains to be seen. Obtaining high quality systematic quantitative data characterizing systems parameters such as mRNA, protein and metabolite levels, interactions and spatial and temporal localization of different molecules will be important in such model building. Nevertheless, the data will not provide new insights automatically. We believe that hypotheses expressed as rigorously defined models, the properties of which can be studied independently and tested on experimental data, will play an important role in understanding the living systems on genome-wide level.

Acknowledgements: We thank Jurg Bahler for helpful discussion of the manuscript. The project is funded by the European Commission as the TEMBLOR, contract-no. QLRI-CT-2001-00015 under the RTD programme "Quality of Life and Management of Living Resources".

References

[1] Lee, T.I., et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298, 799–804.
[2] Davidson, E.H., et al. (2002) A genomic regulatory network for development. Science 295, 1669–1678.
[3] von Dassow, G., Meir, E., Munro, E.M. and Odell, G.M. (2000) The segment polarity network is a robust developmental module. Nature 406, 188–192.
[4] Matys, V., et al. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res. 31, 374–378.
[5] Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. Gene Ontol. Consortium. Nat. Genet. 25, 25–29.
[6] Pruess, M., et al. (2003) The Proteome Analysis database: a tool for the in silico analysis of whole proteomes. Nucleic Acids Res. 31, 414–417.
[7] Madan Babu, M. and Teichmann, S.A. (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. Nucleic Acids Res. 31, 1234–1244.
[8] Brazma, A., Jonassen, I., Vilo, J. and Ukkonen, E. (1998) Predicting gene regulatory elements in silico on a genomic scale. Genome Res. 8, 1202–1215.
[9] Brazma, A., Vilo, J., Ukkonen, E. and Valtonen, K. (1997) Data mining for regulatory elements in yeast genome. Proc. Int. Conf. Intell. Syst. Mol. Biol. 5, 65–74.
[10] Werner, T., Fessele, S., Maier, H. and Nelson, P.J. (2003) Computer modeling of promoter organization as a tool to study transcriptional coregulation. FASEB J. 17, 1228–1237.
[11] Dieterich, C., Rahmann, S. and Vingron, M. (2004) Functional inference from nonrandom distributions of conserved predicted transcription factor binding sites. Bioinformatics 20 (Suppl. 1), I109–I115.
[12] Wasserman, W.W. and Fickett, J.W. (1998) Identification of regulatory regions which confer muscle-specific gene expression. J. Mol. Biol. 278, 167–181.
[13] Kellis, M., Patterson, N., Endrizzi, M., Birren, B. and Lander, E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423, 241–254.
[14] Harbison, C.T., et al. (2004) Transcriptional regulatory code of a eukaryotic genome. Nature 431, 99–104.
[15] Schlitt, T. and Brazma, A. (2002) Learning about gene regulatory networks from gene deletion experiments. Comp. Funct. Genom. 3, 499–503.
[16] Hughes, T.R., et al. (2000) Functional discovery via a compendium of expression profiles. Cell 102, 109–126.
[17] Albert, R. and Barabási, A.-L. (2002) Statistical mechanics of complex networks. Rev. Mod. Phys., 74.
[18] Rung, J., Schlitt, T., Brazma, A., Freivalds, K. and Vilo, J. (2002) Building and analysing genome-wide gene disruption networks. Bioinformatics 18, S202–S210.
[19] Hartwell, L.H., Hopfield, J.J., Leibler, S. and Murray, A.W. (1999) From molecular to modular cell biology. Nature 402, C47–C52.
[20] Segal, E., Shapira, M., Regev, A., Pe'er, D., Botstein, D., Koller, D. and Friedman, N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. Nat. Genet. 34, 166–176.
[21] Ihmels, J., Friedlander, G., Bergmann, S., Sarig, O., Ziv, Y. and Barkai, N. (2002) Revealing modular organization in the yeast transcriptional network. Nat. Genet. 31, 370–377.
[22] Wolf, D.M. and Arkin, A.P. (2003) Motifs, modules and games in bacteria. Curr. Opin. Microbiol. 6, 125–134.
[23] Snel, B. and Huynen, M.A. (2004) Quantifying modularity in the evolution of biomolecular systems. Genome Res. 14, 391–397.
[24] Ren, B., et al. (2000) Genome-wide location and function of DNA binding proteins. Science 290, 2306–2309.
[25] Simon, I., et al. (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. Cell 106, 697–708.
[26] Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R. and Young, R.A. (2003) Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. Cell 113, 395–404.
[27] Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A. and Gerstein, M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312.
[28] Han, J.D., et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature 430, 88–93.
[29] Schlitt, T., Palin, K., Rung, J., Dietmann, S., Lappe, M., Ukkonen, E. and Brazma, A. (2003) From gene networks to gene function. Genome Res. 13, 2568–2576.
[30] Manke, T., Bringas, R. and Vingron, M. (2003) Correlating protein–DNA and protein–protein interaction networks. J. Mol. Biol. 333, 75–85.
[31] Albert, R., Jeong, H. and Barabasi, A.L. (2000) Error and attack tolerance of complex networks. Nature 406, 378–382.
[32] Albert, R., Jeong, H. and Barabasi, A.L. (2001) Correction: error and attack tolerance of complex networks. Nature 409, 542.

[33] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. and Alon, U. (2002) Network motifs: simple building blocks of complex networks. Science 298, 824–827.
[34] Teichmann, S.A. and Babu, M.M. (2004) Gene regulatory network growth by duplication. Nat. Genet. 36, 492–496.
[35] Kauffman, S.A. (2002) Investigations, Oxford University Press Inc, USA.
[36] Thieffry, D., Colet, M. and Thomas, R. (1993) Formalization of regulatory networks: a logical method and its automation. Math. Model. Sci. Comput. 55, 144–151.
[37] Soinov, L.A., Krestyaninova, M.A. and Brazma, A. (2003) Towards reconstruction of gene networks from expression data by supervised learning. Genome Biol. 4, R6.
[38] Friedman, N. (2004) Inferring cellular networks using probabilistic graphical models. Science 303, 799–805.
[39] Yuh, C.H., Bolouri, H. and Davidson, E.H. (1998) Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. Science 279, 1896–1902.
[40] Davidson, E.H., et al. (2002) A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. Dev Biol 246, 162–190.
[41] Akutsu, T., Miyano, S. and Kuhara, S. (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. Pac. Symp. Biocomput., 17–28.
[42] Thomas, R. (1973) Boolean formalization of genetic control circuits. J. Theor. Biol. 42, 563–585.
[43] Kauffman, S. (1969) Homeostasis and differentiation in random genetic control networks. Nature 224, 177–178.
[44] D'Haeseleer, P., Wen, X., Fuhrman, S. and Somogyi, R. (1999) Linear modeling of mRNA expression levels during CNS development and injury. Pac. Symp. Biocomput., 41–52.
[45] Chen, T., He, H.L. and Church, G.M. (1999) Modeling gene expression with differential equations. Pac. Symp. Biocomput., 29–40.
[46] Basu, S., Mehreja, R., Thiberge, S., Chen, M.T. and Weiss, R. (2004) Spatiotemporal control of gene expression with pulse-generating networks. Proc. Natl. Acad. Sci. USA 101, 6355–6360.
[47] Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. Nature 405, 590–593.
[48] Elowitz, M.B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338.
[49] Gardner, T.S., Cantor, C.R. and Collins, J.J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403, 339–342.
[50] Becskei, A., Seraphin, B. and Serrano, L. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J. 20, 2528–2535.
[51] McAdams, H.H. and Shapiro, L. (1995) Circuit simulation of genetic networks. Science 269, 650–656.
[52] Brazma, A. and Schlitt, T. (2003) Reverse engineering of gene regulatory networks: a finite state linear model. Genome Biol. 4, P5.
[53] Holstege, F.C., et al. (1998) Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95, 717–728.
[54] McAdams, H.H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. Proc. Natl. Acad. Sci. USA 94, 814–819.
[55] Raser, J.M. and O'Shea, E.K. (2004) Control of stochasticity in eukaryotic gene expression. Science 304, 1811–1814.
[56] Paulsson, J. (2004) Summing up the noise in gene networks. Nature 427, 415–418.
[57] Cho, R.J., et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell. 2, 65–73.
[58] Spellman, P.T., et al. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol. Biol. Cell 9, 3273–3297.
[59] Rustici, G., et al. (2004) Periodic gene expression program of the fission yeast cell cycle. Nat. Genet. 36, 809–817.
[60] Chen, D., et al. (2003) Global transcriptional responses of fission yeast to environmental stress. Mol. Biol. Cell 14, 214–229.
[61] Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D. and Brown, P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell 11, 4241–4257.
[62] Ideker, T., et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292, 929–934.
[63] Jorgensen, P., Nishikawa, J.L., Breitkreutz, B.J. and Tyers, M. (2002) Systematic identification of pathways that couple cell growth and division in yeast. Science 297, 395–400.