# Modelling and analysis of gene regulatory networks

*Guy Karlebach and Ron Shamir*

Abstract | Gene regulatory networks have an important role in every process of life, including cell differentiation, metabolism, the cell cycle and signal transduction. By understanding the dynamics of these networks we can shed light on the mechanisms of diseases that occur when these cellular processes are dysregulated. Accurate prediction of the behaviour of regulatory networks will also speed up biotechnological projects, as such predictions are quicker and cheaper than lab experiments. Computational methods, both for supporting the development of network models and for the analysis of their functionality, have already proved to be a valuable research tool.

Stochasticity
The property of a system whose behaviour depends on probabilities. In a model with stochasticity, a single initial state can evolve into several different trajectories, each with an associated probability.

*The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. Correspondence to R.S.*
*e-mail: rshamir@tau.ac.il*
doi:10.1038/nrm2503
Published online
17 September 2008

The genome encodes thousands of genes whose products enable cell survival and numerous cellular functions. The amounts and the temporal pattern in which these products appear in the cell are crucial to the processes of life. Gene regulatory networks govern the levels of these gene products. A gene regulatory network is the collection of molecular species and their interactions, which together control gene-product abundance. Numerous cellular processes are affected by regulatory networks.

Innovations in experimental methods have enabled large-scale studies of gene regulatory networks and can reveal the mechanisms that underlie them. Consequently, biologists must come to grips with extremely complex networks and must analyse and integrate great quantities of experimental data. Essential to this challenge are computational tools, which can answer various questions: what is the full range of behaviours that this system exhibits under different conditions? What changes are expected in the dynamics of the system if certain parts stop functioning? How robust is the system under extreme conditions?

Various computational models have been developed for regulatory network analysis. These models can be roughly divided into three classes. The first class, logical models, describes regulatory networks qualitatively. They allow users to obtain a basic understanding of the different functionalities of a given network under different conditions. Their qualitative nature makes them flexible and easy to fit to biological phenomena, although they can only answer qualitative questions. To understand and manipulate behaviours that depend on finer timing and exact molecular concentrations, a second class of models was developed — continuous models. For example, to simulate the effects of dietary restriction on yeast cells under different nutrient concentrations[1], users must resort to the finer resolution of continuous models. A third class of models was introduced following the observation that the functionality of regulatory networks is often affected by noise. As the majority of these models account for interactions between individual molecules, they are referred to here as single-molecule level models. Single-molecule level models explain the relationship between stochasticity and gene regulation.

Predictive computational models of regulatory networks are expected to benefit several fields. In medicine, mechanisms of diseases that are characterized by dysfunction of regulatory processes can be elucidated. Biotechnological projects can benefit from predictive models that will replace some tedious and costly lab experiments. And, computational analysis may contribute to basic biological research, for example, by explaining developmental mechanisms or new aspects of the evolutionary process.

Here we review the available methodologies for modelling and analysing regulatory networks. These methodologies have already proved to be a valuable research tool, both for the development of network models and for the analysis of their functionality. We discuss their relative advantages and limitations, and outline some open questions regarding regulatory networks, including how structure, dynamics and functionality relate to each other, how organisms use regulatory networks to adapt to their environments, and the interplay between regulatory networks and other cellular processes, such as metabolism.

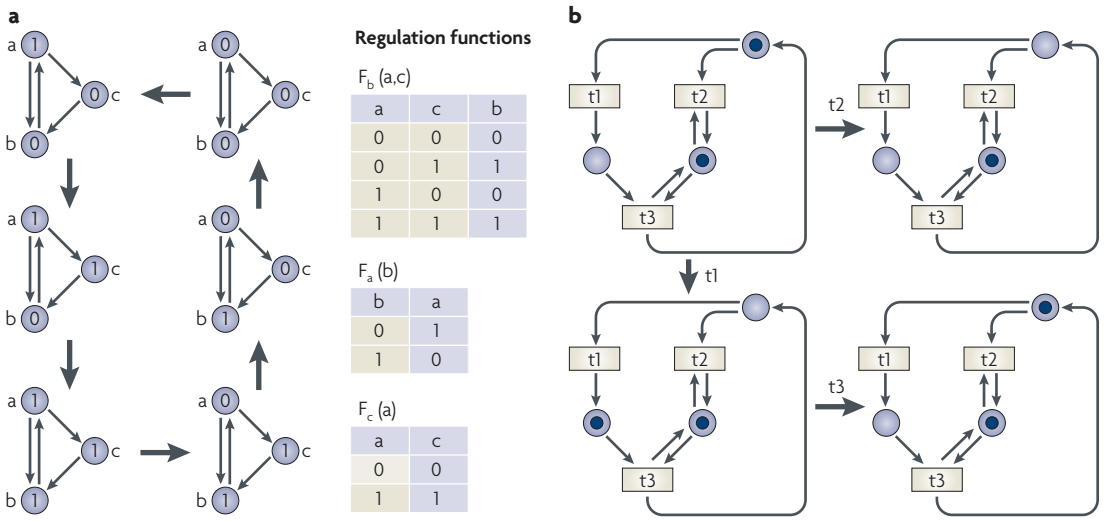



Figure 1 | **Logical models. a** | A Boolean network. Each of the entities a, b and c in the network can be in state 0 or 1. State transitions obey the regulation functions shown on the right, which describe the rules of the model. For example, if a is in state 1 and c is in state 0, at the next time step the state of b will be 0. Thin arrows indicate the regulators of each node. Time steps are represented by thick arrows. The global state of the model is the combination of the three entity states. The system cycles through the six global states. A sequence of consecutive global states is called a trajectory. **b** | A Petri net. The net contains 'places' (light blue circles) that are the model's entities, and 'transitions' (rectangles) that constitute the regulation functions and define the model's dynamics. Arcs connect input places to transitions, and transitions to their output places. Places that receive discrete values are called tokens (dark blue dots). A transition that is activated, or 'fired', reduces the tokens in its input places and increases the number of tokens in each of its output places. At any time step, every transition that has enough tokens in its input places may be fired. In the example, every transition consumes one token from every input place, and produces one token at every output place. Labels at thick arrows indicate which transition fired. Transitions t1 and t3 can be fired in alternation indefinitely, whereas no other transition can be fired after t2 has fired.

Local state
At any time point, the value representing the status of an entity in a model is its (local) state. For example, the state of a protein may indicate whether it is phosphorylated or not (a Boolean value), or the time since its last phosphorylation (a continuous value).

Synchronous model
A model wherein the time steps at which the global state changes are discrete and (usually) equally spaced. On each step, all the states are updated simultaneously, depending on the model's regulation functions and on the global state at the previous step. In asynchronous models, system changes are not confined to specific times and global states do not progress according to 'a common clock'. Time is often continuous, and entities may change their states at different times.

Regulation function
A rule that determines the state of a specific entity in the model as a function of the states of some (other) entities. For example, several transcription factors may together regulate the expression of a gene. The set of entities whose states determine the state of entity X are entity X's regulators.

Global state
The combination of all the local states of a model at one time point.

Steady state
A global state that, once reached, always repeats itself in a trajectory. Another important dynamic behaviour in biological systems is a cycle of global states. For example, the oscillations observed in the cell cycle.

Robustness
A measure of a model's ability to withstand changes without changing its essential properties. For example, in network models, robustness can be quantified as the fraction of edge additions and/or removals that change the trajectory that emanates from some initial state.

## Logical models

The most basic and simplest modelling methodology is discrete and logic-based, and was introduced by Kauffman and Thomas[2,3]. The reconstruction of the regulatory network that controls the development of sea urchin embryos[4,5] is a seminal example of the profound insights that qualitative examination of regulatory network models can provide. This work demonstrates how maternal cues initiate the activity of the regulatory network and how this network orchestrates the developmental process. Logical models represent the local state of each entity in the system (for example, genes, proteins and small molecules) at any time as a discrete level, and the temporal development of the system is often assumed to occur in synchronous, discrete time steps. Entity levels are updated at each time step according to regulation functions (FIG. 1a). Discrete modelling allows researchers to rely on purely qualitative knowledge. Such models can be analysed using a broad range of well established mathematical methods.

***Boolean networks.*** Boolean regulatory networks were first presented by Kauffman[2,6]. In a Boolean network, an entity can attain two alternative levels: active (1) or inactive (0). For example, a gene can be described as expressed or not expressed at any time. The level of each entity is updated according to the levels of several entities, via a specific Boolean function. The 0–1 vector that describes the levels of all entities is called the system's state, or the global state. It is assumed to change synchronously, such that at every time step, the level of each entity is determined according to the levels of its regulators at the previous time step and according to the regulation function (FIG. 1a).

Boolean networks were recently used to analyse the relationship between regulation functions and network stability in the yeast transcriptional network, using only the network's structure[7]. According to this study, the network is stable when random regulation functions are used, and solution stability increases when the regulation functions are biologically meaningful. It also showed that Boolean networks do not correctly model the dynamics of a transcription factor that downregulates its own expression, due to the model's limited level of detail. Another problem is that it is computationally expensive to analyse the dynamics of large networks, as the number of global states is exponential in the number of entities. However, when the number of entities is small and only qualitative knowledge is available, Boolean networks can provide important insights, such as the existence and nature of steady states or network robustness.

To study the dynamics of cell-cycle regulation in yeast, Li *et al.*[8] constructed a literature-based Boolean network in which all the regulation functions are threshold functions. This model generated trajectories with a high degree of overlap, most of which led into a path that corresponded to the cell-cycle phases of yeast. In addition, most small changes in the model did not significantly change its dynamic behaviour, indicating that it is robust. As the analysis relied on an exhaustive enumeration of all the possible trajectories, this method is only practical for small networks.

In many cases, the regulatory relationships between network components have not been established, and therefore need to be derived from experimental data. For any entity under a Boolean network model, both its regulators and a regulatory function that is consistent with a set of gene-expression profiles can be found efficiently, provided that the number of regulators of each entity does not exceed a set limit[9]. Such an algorithm is faster than a previous one proposed by Akutsu and colleagues[10]. Lahdesmaki *et al.*[9] also presented an algorithm for selecting a set of candidate regulation functions in the presence of contradictory evidence, whereby each expression profile is associated with a certainty level (that is, a numerical value that expresses one's confidence in the profile). This algorithm was tested by deriving regulation functions for 5 yeast cell-cycle regulated genes using expression profiles of 733 candidate regulators[11]; the maximum number of regulators that together regulate a single gene was first set to 1, then 2 and finally 3. The analysis yielded a large number of regulation functions that were equally consistent with experimental data. Some of the suggested functions matched previous findings.

***Probabilistic Boolean networks.*** Often, due to insufficient experimental evidence or incomplete understanding of a system, several candidate regulatory functions may be possible for an entity. This raises the need to express uncertainty in the regulatory logic. Shmulevich *et al.*[12,13] addressed this idea by modifying the Boolean network model such that an entity can have several regulation functions, each of which is given a probability based on its compatibility with prior data. At each time step, every entity is subjected to a regulation function that is randomly selected according to the defined probabilities[12]. Hence the model is stochastic and an initial global state can lead to many trajectories of different probabilities. The new model, the probabilistic Boolean network (PBN), generates a sequence of global states that constitutes a Markov chain[14]. For example, a PBN was used to model a 15 gene sub-network that was inferred from human glioma expression data[13]. This analysis demonstrates that the stationary distributions of entities may indicate possible regulatory relationships among them: entities that have the same states in a significant proportion of the global states are likely to be related. As the number of global states in the gene sub-network was prohibitively large, one study[13] estimated the stationary distribution by sampling the global states[15].

***MetaReg.*** An exponential number of global states makes it difficult to analyse the dynamics of all but tiny models. In some cases, analysis under steady state conditions turns out to be a practical goal. Gat-Viks *et al.*[16] developed the MetaReg model, in which entities can have several levels (typically 3–5) and regulation functions are discrete. Two efficient heuristics were developed: the first detects a network's steady states and the second selects regulation functions that are most consistent with these steady states. The former heuristic can be used to analyse the dynamics of the network, whereas the latter can complete or correct a literature-based network. MetaReg was used to analyse the regulation of lysine biosynthesis in yeast and indicated previously unknown transcriptional controls of several metabolic enzymes.

To express uncertainty in regulation functions, Gat-Viks *et al.*[17] created a probabilistic version of the MetaReg model. In this model, an entity can have one of several possible regulation functions (with the same regulators), and probabilities that each one is correct. Technically, the model is represented as a factor graph (an expansion of Bayesian networks)[18]. Analogously to the model in REF. 16, it can be subjected to steady state identification and optimization of regulation function[18,19]. It can also discover new regulatory relationships. The method has been improved[20] to facilitate changes in the network structure (refinement) and inclusion of additional entities (expansion). Analysis of a network of 4 interconnected osmotic stress-related yeast signalling pathways, which consists of 43 entities, along with 106 expression profiles, identified novel regulatory modules and crosstalks between pathways. Thus, the model can correct and expand a known regulatory network.

***Petri nets.*** The dynamics of a regulatory network can also be analysed using Petri nets[21], non-deterministic models (FIG. 1b). An example of a question that users can ask with a Petri net is: how many transition sequences lead from global state A to global state B? The qualitative description of biochemical reactions using a Petri net is straightforward, and Petri net models are useful analysis tools for large metabolic networks[22–24]. Chaouiya *et al.* showed that Petri nets can also model regulatory networks using Boolean regulatory functions[25], and that the metabolic and regulatory layers can be connected[26]. Steggles *et al.* proved that the synchronous dynamics of a Boolean network can be captured by a Petri net[27] and demonstrated that uncertainty in the regulation functions can also be expressed by the model. Heuristics for analysing the dynamics of Petri nets have been studied extensively in the past 3 decades, and include detection of active pathways, testing if a given system state is reachable and detecting state cycles[28]. Steggles *et al.* modelled the regulatory network of *Bacillus subtilis* sporulation using Petri nets and produced a behaviour that is in good agreement with existing literature[27]. For example, when initializing the system to a global state that corresponds to vegetative growth and activating the sporulation signal, the dynamics of the system lead to a state that corresponds to sporulation. This model also correctly predicted the sporulation capabilities of mutants.

***Inference of particular network properties.*** In certain cases, incomplete information about a regulatory network can be used to infer topological features and regulatory interactions of the network. Due to the noisy nature of biological experiments, inference is usually based on a probabilistic framework that integrates experimental data in a network context. Here we briefly describe some static probabilistic models that infer properties of regulatory networks. These models do not describe in full the regulation of each entity under every possible condition, and do not describe dynamic processes (the concept of

Threshold function
A regulation function is a threshold function if it determines the state of the output entity by summing the states of its inputs and comparing the sum to some fixed value. For example, a gene upregulated if any two out of three transcription factors are active can be modelled by such a function.

Trajectory
In logical models, a trajectory is a sequence of global states that occur consecutively. In continuous models, a trajectory is the change of the level of an entity over time.

Markov chain
A stochastic process in which the next state depends only on the present state, regardless of the trajectory that led to the present state.

Heuristic
An algorithm for solving a problem that does not always provide an optimal solution to it. Heuristics are often used when it is impractical to obtain an exact optimal solution, and in many cases they provide satisfactory solutions.

Bayesian network
A probabilistic model that represents (in)dependencies between variables, taking the form of a directed acyclic graph. Often, both inference and learning can be carried out efficiently in such models. Dynamic Bayesian networks are an extension that describes dynamic behaviour.

Module
A set of genes that have identical regulation functions (and regulators). In other contexts, a module can also be a set of genes with a common function.

Inference
The selection of regulatory functions (or regulators) that best agrees with a dataset.

trajectory is not defined for them), but provide higher level, lower resolution modelling and analysis (REF. 29 is an excellent source on probabilistic inference).

Module networks, introduced by Segal and colleagues, is a model that infers the regulation logic of gene modules given gene-expression data[30]. A regulation logic is represented by a decision tree, in which a path from the root to a leaf is determined by the up- or downregulation of regulatory modules, and a leaf determines the expression level of the corresponding genes. Module networks were tested with experimental data and correctly predicted some regulatory modules. Friedman *et al.* introduced Bayesian networks as a probabilistic tool for the identification of regulatory genes using high-throughput experimental data[29] and showed that they can reproduce certain known regulatory relationships[31,32]. Physical network models combine protein–DNA interactions, protein–protein interactions and knockout experiments for the discovery of regulatory interactions. The network structure of these models predicted knockout effects correctly[33]. Yeang and Vingron integrated perturbation data with knowledge from the literature into a joint model of regulation and metabolism and created a framework for the prediction of regulatory interactions and pathways[34]. They verified the predictive power of their model on the regulatory networks that govern the metabolism of glucose in *Escherichia coli* and found that the use of a joint model explains more perturbations than a regulatory network would explain alone. In all the probabilistic inference models above, predicted properties are assigned a certainty level. A cut-off for deciding which features will be selected for further analysis can then be determined. Examples for using cut-off criteria for network-feature selection can be found in REFS 7,32,35.

## Continuous models

Biological experiments usually produce real, rather than discrete-valued, measurements. Examples include reaction rates, cell mass, cell-cycle length and gene-expression intensities[36–39]. Logical models require discretization of the real-valued data, which reduces the accuracy of the data. Continuous models, using real-valued parameters over a continuous timescale, allow a straightforward comparison of the global state and experimental data and can theoretically be more accurate. In practice, however, quantitative measurements are almost always partial (that is, they cover only a fraction of the system's entities). Therefore, some of the parameters of continuous models are usually based on estimations or inference. Below we describe some types of continuous models and the predictions that they can generate.

***Continuous linear models.*** The defining property of linear models is that each regulator contributes to the input of the regulation function independently of the other regulators, in an additive manner. In other words, the change in the level of each entity depends on a weighted linear sum of the levels of its regulators. This assumption allows a high level of abstraction and efficient inference of network structure and regulation functions.

Discretization
A process that transforms continuous numerical values into discrete ones. For example, real-valued measurements can be discretized to 0, 1 or 2, corresponding to low, medium and high levels.

Time-series data usually contain many more genes than time points. This presents a difficulty in reverse engineering a network's structure and regulation functions. Yeung *et al.*[40] used a linear model and singular value decomposition[41] to generate a family of candidate networks that are consistent with a given dataset, thus compensating for this deficiency in time points. The network that is most consistent with prior knowledge is selected. The authors demonstrated in simulations that this approach is effective in dealing with shortages of data. Weaver *et al.*[42] described a model in which the expression of each gene is regulated by a 'squashing' function that takes as input a weighted linear sum of regulator levels, and presented an algorithm for reverse engineering real networks under these assumptions. One recent study adopted the linear framework to create a model of a regulatory network that is subjected to an arbitrary number of perturbations and studied multiple perturbation scenarios using simulated data and a single-perturbation scenario using experimental data[43]. Another study added time delays to regulatory interactions[44], which can be used to infer the duration of protein synthesis.

Linear models do not require extensive knowledge about regulatory mechanisms and can be used to obtain qualitative insights about regulatory networks, the simplest example being detection of novel regulations. However, when higher sensitivity to detail is desired, more complex models are preferable.

***Models of transcription factor activity.*** The linear model is a crude description of the process of gene expression, and as such it cannot provide answers to subtle questions such as: how does the affinity of a transcription factor to a target promoter affect the network? Nachman *et al.*[45] created a fine-level model of gene regulation. In their model, entities correspond to either genes or transcription factors, and levels represent mRNA abundance or transcription factor activity, respectively. All the regulators are transcription factors. The levels of genes are determined by real-valued, non-linear regulation functions that take the Michaelis–Menten form[46]. The level of a gene is thus determined by that function together with the mRNA-decay rates. The time-dependent transcription factor activities are inferred from microarray time-series data using dynamic Bayesian networks[47,48]. An efficient heuristic aims to discover new regulators and regulatory relationships. Given an established regulatory network of 141 yeast cell-cycle genes, the heuristic successfully predicted the activity levels of the 7 regulators that controlled this network. In addition, it proposed novel regulatory relationships that improved the explanatory power of the model. Moreover, when given the entities, but not the network structure, as input, this method identified the seven regulators.

Shamir and Tanay developed a different model for identifying transcription factor–gene regulations[49]. The method relies on an efficient algorithm that infers transcription factor activity under the assumption that it is a monotone increasing function of both the transcription factor–promoter affinity and the transcription factor dosage. Transcription factor–promoter affinities are

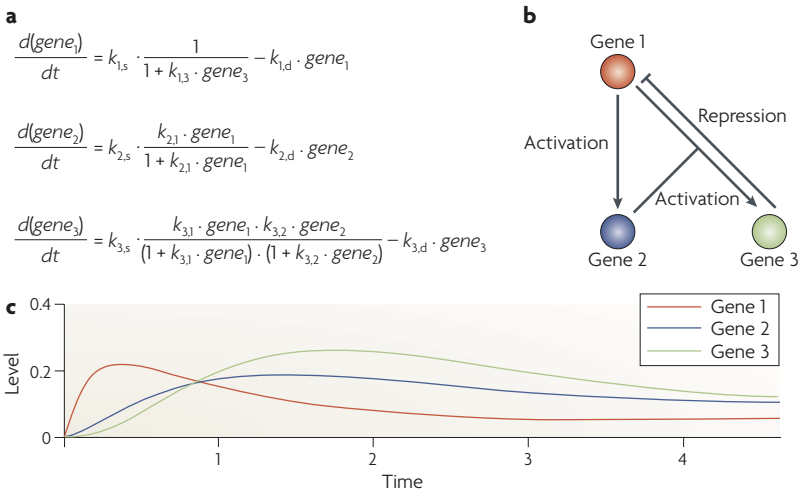



Figure 2 | **Ordinary differential equation model. a** | A network of three genes is modelled using ordinary differential equations (ODEs). Reaction rate constants are denoted by 'k'. **b** | The regulatory relations are depicted graphically. **c** | The trajectories of the model. Each equation shows the change in the level of a gene as a difference of its synthesis and degradation. Gene 1 is constitutively expressed, and is repressed by gene 3. Therefore, its level may reach a maximal rate of increase ($k_{1,s}$; 's' stands for synthesis) when the level of gene 3 is 0, in which case $k_{1,s}$ will be multiplied by 1. When the level of gene 3 is non-zero, the level of gene 1 rises slower than $k_{1,s}$. Transcription of gene 2 is activated by gene 1. This is expressed in the second equation of panel **a**, in which gene 2 level rises as a Michaelis–Menten function of the level of gene 1. Similarly, transcription of gene 3 is activated when both gene 1 and gene 2 levels are non-zero, and this relationship is given in the third equation of panel **a**. Degradation is modelled as a first-order reaction with rate constants $k_{i,d}$ (in which 'i' can be 1, 2 or 3). This formulation assumes that every transcript is immediately translated, and therefore the synthesis constants $k_{i,s}$ refer to both transcription and translation. According to simulation (bottom), the system stabilizes in a steady state at about 4.5 time units. The values of the rates in the simulations were: $k_{1,s}=k_{2,s}=2$; $k_{3,s}=15$; $k_{1,d}=k_{2,d}=k_{3,d}=1$; $k_{2,1}=k_{3,1}=k_{3,2}=1$; and $k_{1,3}=100$. The initial levels were all zero. Equations were solved using DESSolver v1.7 and the fourth order Runge–Kutta method.

inferred based on analysis of the promoters of the regulated genes. The model was applied to 140 genes of the galactose system in yeast, and inferred transcription factor activities that were in accordance with the literature. Two putative novel transcription factors, along with their genomic binding sites, were suggested. This demonstrates that the integration of multiple datasets can yield additional predictions that would be difficult to obtain from either dataset alone. The increased prediction power is obtained by the algorithm's ability to link different, but related, biological phenomena, in this case *cis*-regulatory elements and mRNA abundance.

Recently, Pan *et al.* extended the model developed by Nachman and colleagues by integrating genome sequence data[50]. Although these models offer a detailed description of regulation and provide inference algorithms, they do not directly incorporate interactions between regulatory entities. A similar methodology that uses discrete global states was suggested for inferring transcription factor activities based on the complete regulatory structure[51]. The model of Segal *et al.*[52] reproduced expression patterns that are generated by maternal and zygotic factors in the early *Drosophila melanogaster* embryo and provided interesting insights about the regulatory interactions of this system.

Michaelis–Menten functions
Equations that describe the kinetics of an enzymatic reaction. They can be derived from ordinary differential equations that describe the concentration changes of the involved molecular species under some simplifying assumptions.

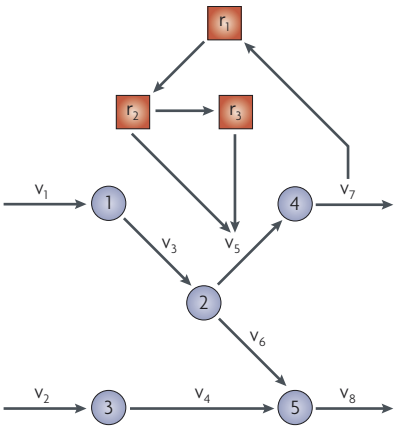
***Ordinary differential equations.*** A more general, detailed model of regulation can be described by ordinary differential equations (ODEs) (FIG. 2). These equations describe the (instantaneous) change in each entity as a function of the levels of some network entities. For simple ODE systems, an analytical solution can be formulated and the resulting set of algebraic equations then describes the change in entity levels over time. (REF. 46 provides a good overview for the use of ODEs in biological context and gives some illustrative examples). The Hill and Michaelis–Menten functions are examples of such analytical solutions of small systems. Larger networks, which often use these functions in addition to linear and bilinear functions, practically always require a numerical solution.

The ODE approach provides detailed information about the network's dynamics, but requires high-quality data on kinetic parameters and it is therefore currently applicable to only a few systems. The idea of using ODEs for modelling regulatory networks was suggested several decades ago[53]. Here we give some recent examples for modelling the dynamics of regulatory networks using ODEs.

Li *et al.* used ODEs to evaluate their model for the cell-cycle regulation in *Caulobacter crescentus*[54]. This bacterium divides asymmetrically into two morphologically distinctive cells, one of which, the stalked cell, is identical in form to the parent[55–58]. Their implementation follows the network dynamics from the parent cell to the stalked daughter cell. Entities correspond to protein concentrations, to the constriction ring at the mid-cell plane, to the process of DNA synthesis and to gene promoters. The system contains 16 equations (one for each variable), and these make use of 44 constants that were initially retrieved from the literature and then adjusted by trial and error. In tests in wild-type and 16 mutant strains, the model's simulations agreed with experimental measurements.

Chen *et al.* used the same approach to model the cell-cycle regulatory network in yeast[59]. In their model, entity levels corresponded to protein concentrations, cell mass, DNA mass, the state of the mitotic spindle and the state of the emerging bud from which the daughter cell was formed. The change in cell mass is assumed to depend only on the current cell mass. Therefore, the mass at division time is determined by the duration of the cell cycle. In total, 36 equations and 148 constants were used. After manual fitting, the model generated trajectories that reasonably matched the parent and daughter cell-cycle durations, the lengths of the $G_1$, $G_2$, S and M phases, and some of the experimentally determined ratios between groups of regulatory proteins. Moreover, 120 out of 131 simulated mutant strains had properties that were consistent with experimentally observed properties, including viability, growth rate, size at birth and size at budding.

Thus, ODE models can generate predictions that may subsequently be compared to cellular phenotypes. Additional examples for modelling with ODEs include the *Arabidopsis thaliana* circadian system[60] and osmoregulation in yeast[61]. More restricted types of ODE have also been proposed for modelling regulatory networks[62,63]. These are usually more abstract, require less detail during the modelling process and can be subjected to more powerful analysis.




**Trajectory**

| $r_1$ | $r_2$ | $r_3$ | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 1 | 0.1 | 0.2 | 0.1 | 0.2 | 0 | 0.1 | 0 | 0.3 |
| 0 | 0 | 1 | 0.1 | 0.2 | 0.1 | 0.2 | 0 | 0.1 | 0 | 0.3 |
| 0 | 1 | 1 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0 | 0.2 | 0.2 |
| 1 | 1 | 0 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0 | 0.2 | 0.2 |
| 1 | 0 | 0 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0 | 0.2 | 0.2 |
| 1 | 0 | 1 | 0.1 | 0.2 | 0.1 | 0.2 | 0 | 0.1 | 0 | 0.3 |

**Regulation functions**

$f_{r1}(V_7)$

| $v_7$ | $r_1$ |
|---|---|
| = 0 | 1 |
| ≥ 1 | 0 |

$f_{r2}(r_1)$

| $r_1$ | $r_2$ |
|---|---|
| 0 | 1 |
| 1 | 0 |

$f_{r3}(r_2)$

| $r_2$ | $r_3$ |
|---|---|
| 0 | 1 |
| 1 | 0 |

$f_{v5}(r_2,r_3)$

| $r_2$ | $r_3$ | $v_5$ |
|---|---|---|
| 0 | 0 | ≥ 0 |
| 0 | 1 | = 0 |
| 1 | 0 | ≥ 0 |
| 1 | 1 | ≥ 0 |

**Stoichiometric matix**

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 0 | –1 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 1 | 0 | –1 | –1 | 0 | 0 |
| 0 | 1 | 0 | –1 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 1 | 0 | –1 | 0 |
| 0 | 0 | 0 | 1 | 0 | 1 | 0 | –1 |

**Constraints**

| |
|---|
| $0 \le v_1, v_2 \le 0.2$ |
| $0 \le v_3, v_4, v_5, v_6 \le 0.4$ |
| $0 \le v_7, v_8 \le 0.3$ |

**Objective function**

$v_7 + v_8$

Figure 3 | **Regulated flux balance analysis model.** The model shown contains three regulatory genes (squares) that regulate a metabolic layer. Metabolites are represented by circles, and metabolic fluxes by arrows that connect metabolites. Fluxes are denoted as $v_1$–$v_8$. The objective function that must be maximized is $v_7+v_8$. The metabolic flux $v_7$ regulates $r_1$. If it is non-zero, $r_1$ becomes active. Otherwise, $r_1$ becomes inactive. The regulators $r_2$ and $r_3$ regulate the flux $v_5$. When $r_2$ is not active and $r_3$ is active, $v_5$ is set to zero. Otherwise $v_5$ is not constrained. The regulation functions are shown. When $v_5$ is not constrained, a maximal value of $v_7+v_8$ is obtained by fluxes of magnitude 0.2 in all reactions, except $v_6$, the value of which remains 0. This is one of several possible solutions for the linear programming problem (they are referred to together as the solution space). When $v_5$ is constrained to 0 by the regulatory layer, $v_7$ must also become 0, and, hence, $v_8$ becomes the only outgoing flux. The trajectory cycles through five global states. The stoichiometric matrix describes the metabolites that each reaction consumes and produces. The columns correspond to reactions, and the rows to metabolites. For example, the third column means that the third reaction consumes one molecule of metabolite 1 for each molecule of metabolite 2 that is produced.

***Regulated flux balance analysis.*** The cell-cycle ODE models incorporate cell growth and division by considering the progression of regulatory processes. However, in reality, changes in cell mass depend on metabolic activity. A complete picture of cellular regulation must take into account metabolic reactions and their interplay with the regulatory layer. For example, in the *lac* operon, a regulatory protein, the lac repressor, is regulated by a metabolite, lactose[64]. Regulated flux balance analysis (rFBA)[65,66] is a modelling approach that aims to integrate regulation and metabolism. rFBA is an extension of FBA[67] (see below; for more information on FBA, see REFS 67,68).

A major problem in using ODEs for describing biochemical reactions is the scarcity of experimental data on rate constants. FBA addresses this problem by assuming that the network is in a steady state and therefore that the total concentration of each substance does not change. Under this assumption, a system of ODEs is transformed into a system of linear equations, and its rates can be obtained by solving a linear programming problem that optimizes a certain objective function, for example, cellular growth. Such optimization problems can be solved efficiently. Further constraints are added to narrow the solution space. For example, the rate constants are restricted according to the catalytic capacities of metabolic enzymes[69]. The method has been successfully used to model large metabolic networks covering the near-complete metabolism of several species[70–72].

rFBA extends FBA by adding a layer of Boolean regulatory entities. For example, transcription factors that can be active or inactive and that can regulate enzymes that catalyse metabolic reactions (FIG. 3). Hence, it models interactions of both logical and continuous entities. The reactions of FBA are subjected to Boolean regulation functions that can set the reaction rates to zero if the regulatory logic dictates inactivation. For example, the production rate of a metabolite can drop to zero if the enzyme that produces it is not transcribed. The entities of the regulatory layer may also regulate each other via Boolean functions, and can also depend on discretized levels of metabolic entities. This regulation can be associated with a time delay. For instance, a Boolean entity that corresponds to a transcription factor can switch from 0 to 1 after a delay due to transcription and translation times.

Covert and Palsson[74] used rFBA to model the regulation of the central metabolic network of *E. coli*, which includes 149 genes, 16 regulatory proteins, 73 enzymes, 45 transcriptional regulations and 113 biochemical reactions. Growth predictions agreed well with experimental measurements in 106 out of 116 combinations of mutant strain–growth medium (measurements included viability, metabolite concentrations, cell mass and gene-expression values). A more comprehensive model that accounts for 1,010 genes was later introduced by Covert and colleagues[75].

Solution space
The set of possible solutions to an optimization problem. In the context of flux balance analysis, the solution space corresponds to different combinations of fluxes that optimize the objective function and that satisfy the constraints.

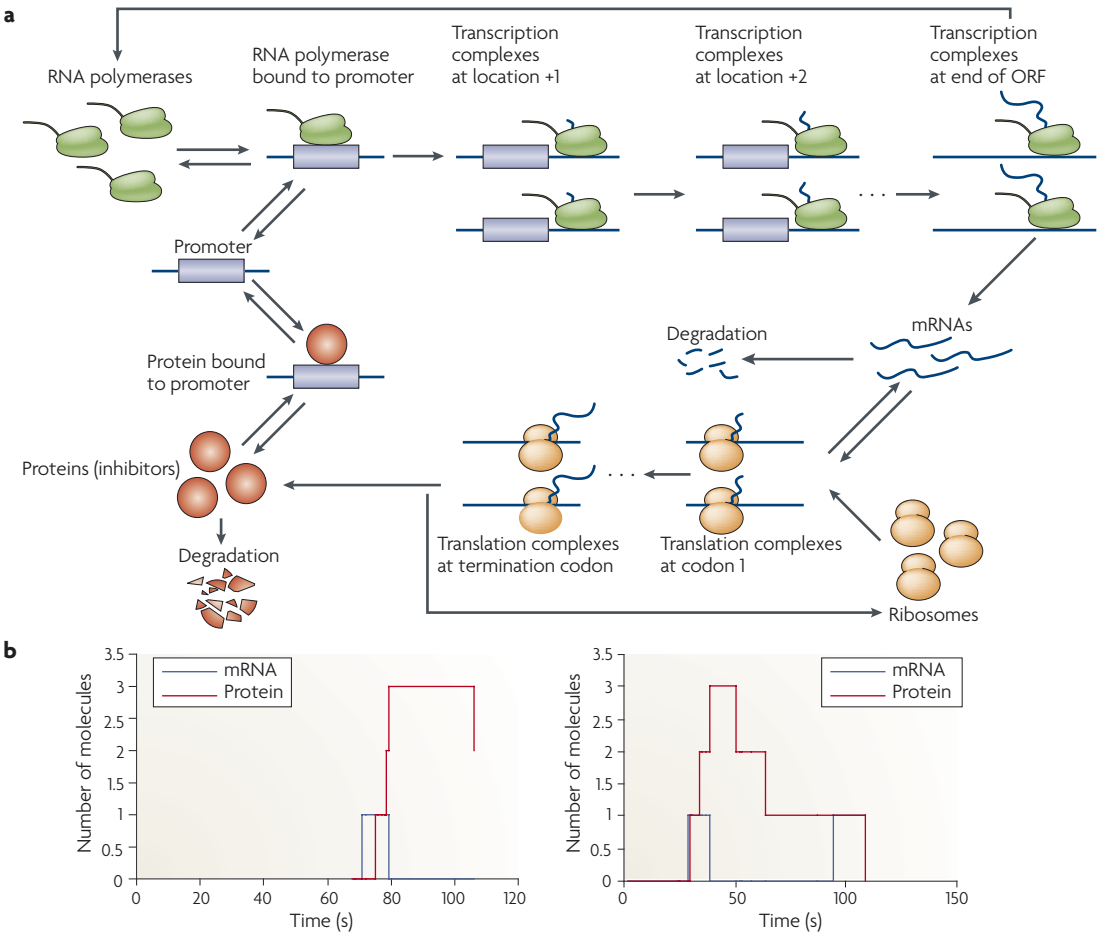



Figure 4 | **Single-molecule level model. a** | Stochastic model for a negative-feedback loop. The system contains a single gene, the product of which represses its own promoter. The diagram shows the different interactions between molecules, each represented by a different entity. For example, the transcription complex is represented by a distinct entity for every location of the transcription complex on the open reading frame (ORF). Arrows represent transformations of molecular species that occur during a reaction. The tails of the arrows point to the substrates and the arrowheads point to the products. For example, the dissociation of the complex RNA polymerase + promoter is represented by the two arrows pointing from the complex to RNA polymerase and to the naked promoter (top left). **b** | Two possible trajectories for the mRNA and protein entities of the model. In the first trajectory, a transcription event occurs, followed by a translation event. Next, several ribosomes initiate translation consecutively and produce two additional proteins (the model allows this as initiations of translation do not consume an mRNA molecule, as is depicted in panel **a**). At the same time, the only transcript degrades. The last event is protein degradation. In the second trajectory, a transcript is produced at an earlier time, and also degrades earlier. Three proteins are generated and then gradually degrade. At about 90 seconds, RNA polymerase manages to bind the promoter and produces a second transcript. Simulations performed using STOCKS 2.0 (REF. 138). The values of the rates in seconds$^{-1}$ were: 100 for elongation of transcript; 30 for elongation of the polypeptide chain; 1 for termination of transcription and/or translation; 0.04 for transcript degradation; 0.025 for protein degradation; and 0.1 for all other reactions. Transcript size was 100, and polypeptide chain size was 30. Initial levels were 1 promoter. The initial number of RNA polymerase molecules is selected from the normal distribution N(35,3.5), and the initial number of ribosome molecules is selected from the normal distribution N(15,3.5), and 0 for all other entities.

Barrett and Palsson[76] created an algorithm that uses rFBA to design a series of experiments for reverse engineering a regulatory network. Before every lab experiment, the algorithm chooses a set of transcription factors that will be knocked out and two growth environments between which the cells will be shifted. The goal is to minimize the total number of experiments. Given probabilistic knowledge about regulatory interactions, the algorithm simulates cell growth for every possible combination of environments and knockout sets, and selects one under which the largest number of novel regulatory interactions are most likely to occur. The lab experiment that follows applies the suggested perturbations and environmental shift, generates an expression profile and verifies all the indicated regulatory interactions using chromatin immunoprecipitation (ChIP). Experimentally verified interactions are added to the model, and the process can be repeated. The algorithm's selections showed good agreement with the decisions of scientists in the reconstruction of an *E. coli* network. A similar methodology was proposed[77] and tested experimentally[33,78] for selecting experiments in Boolean network reconstruction.

Box 1 | **Stochastic simulation of phage λ development**

Phage λ is a bacteriophage that infects *Escherichia coli* cells. A network of regulatory interactions between phage molecules determines if the phage selects the lysogenic pathway or the lytic one[137]. When a phage chooses the lytic pathway, the concentration of the Cro protein in the host is relatively high and the concentration of the CI protein is relatively low. If the lysogenic pathway is chosen, the opposite is true. McAdams and Arkin simulated the pathway-decision process by using an stochastic simulation algorithm (SSA) under several simplifying assumptions (for example, that the host's housekeeping molecules are present in constant concentrations)[104]. Their model defined 26 reaction types, 40 parameters and 18 molecular species (not including complexes). For example, elongation of a polypeptide chain is a single reaction with the same rate for all amino acids. They view the DNA as one species, although the position of RNA polymerase affects transcription rate, and consider the translation of any mRNA transcript by the ribosome as a single reaction type. The simulations showed that the trajectories of CI and Cro concentrations may vary substantially as a result of the intrinsic stochasticity of the system. Furthermore, the fraction of lysogens as a function of the average number of phages per host was in good accordance with experimental data.

This work demonstrated, for the first time, that a real regulatory network can generate profoundly different trajectories due to stochasticity. Subsequently, Weinberger *et al.*[103], on the basis of experiments and simulations, proposed that a positive-feedback loop created by the Tat protein and affected by stochasticity generates fluctuations in latency time. Schultz *et al.*[102] used SSA to explain the transition between vegetation and competence in *Bacillus subtilis*. Gonze and Golbeter[100] investigated the effects of noise on circadian clocks and the conditions that promote their robustness. More efficient methods are needed to carry out simulations of larger networks.

The rFBA approach offers a detailed description of the metabolic layer and also accounts for the interplay between regulation and metabolism. Although the modelling of the regulatory layer is qualitative and less detailed than in other continuous approaches, this is compensated for by the model's capability to infer metabolic fluxes. (For another example of rFBA, see REF. 79 for an analysis of the regulation of metabolism in yeast.) Shlomi *et al.* extended rFBA to study the regulation of metabolism in the steady state[80]. In their model a steady state is obtained by solving a mixed integer linear programming problem rather than by following a trajectory. A different constraint-based approach that allows analysis of the regulatory network in various environments was introduced by Gianchandani and colleagues[81].

## Single-molecule level models

Every biological network is composed of stochastic components, and therefore it may manifest different behaviours, even starting from the same initial conditions[82,83]. When the number of involved molecules of each species is large, the law of mass action[46] can be used to accurately calculate the change in concentrations, and little or no stochastic effect is observable. However, when the number of molecules is small, significant stochastic effects may be seen (FIG. 4). This is particularly true for regulatory networks, in which the number of regulatory molecules is often low[84–87]. Recently, single-cell experimental assays demonstrated the stochastic behaviour of the processes of transcription[88–90] and translation[89,91,92]. Here we present models that incorporate the stochastic nature of regulation by accounting for the fluctuations that occur on the molecular level (reviewed in REF. 93).

***Gillespie's stochastic simulation algorithm.*** McAdams and Arkin[94] showed that fluctuations in time intervals between biochemical reactions, and consequently in the occurrence times of regulatory events, can be expressed by a model that follows biochemical reactions at single-molecule resolution. The model is based on Gillespie's stochastic simulation algorithm (SSA)[95,96]. SSA takes as input the initial number of molecules of several species (for example, mRNAs and proteins) and reaction-probability constants, and simulates the dynamics of the system, reaction by reaction. A reaction probability is the probability that the necessary combination of specific molecules will participate in that reaction in an infinitesimal time interval. For example, consider the phosphorylation reaction:

$$\text{Kinase–phosphate} + \text{target} \xrightarrow{c_1} \text{kinase} + \text{target–phosphate}$$

The reaction probability $c_1 \cdot dt$ is the probability that a specific kinase molecule will phosphorylate a specific protein molecule in the infinitesimal time interval dt. Gillespie has shown how reaction probability constants can be derived from deterministic reaction rates.

The basic assumption of the algorithm is that the system is 'well stirred' — that is, that each molecule always has an equal chance of being anywhere in the system's volume. This assumption applies, for example, if most of the collisions between molecules are non-reactive. Although it overlooks some biological processes that affect regulation, such as diffusion[97] and transportation[98,99], the algorithm proved useful in describing the time evolution of several small regulatory networks and mechanisms[100–104]. BOX 1 provides an example of how SSA can be used to analyse a biological system.

***Approximations to SSA.*** Although Gisbon and Bruck introduced a way to speed up SSA[105], SSA still requires extensive computational resources because it simulates every individual reaction. Consequently, SSA is not ideal for modelling large-scale networks. Therefore, researchers further modified SSA, sacrificing a certain level of detail for the sake of faster simulation.

τ-leaping is a variation of SSA that trades accuracy for efficiency[106]. Instead of generating every single reaction, τ-leaping 'leaps' over a time interval of size τ and randomly selects the number of reactions of each type that occurred in this interval. Gillespie suggested[106] a procedure for selecting τ that was later improved and implemented as part of a stochastic simulation toolkit[107] (REF. 97 describes in detail different SSA approximation methods). When some of the reactions can be described using ODEs, a more efficient strategy is to separate reactions into two regimes: discrete and continuous (see, for example, REFS 108, 109). The integration algorithm of E-Cell version 3 (see Supplementary information S1 (table)) combines multiple stand-alone algorithms (for example SSA and a numerical ODE solver[110]). The use of effective reactions, which amalgamate several simple reaction steps into a single complex step, is a method for abstraction and increasing simulation speed[111,112]. Reaction steps can also be eliminated by applying a steady state assumption[113]. Additional approximation methods are described in REF. 93.

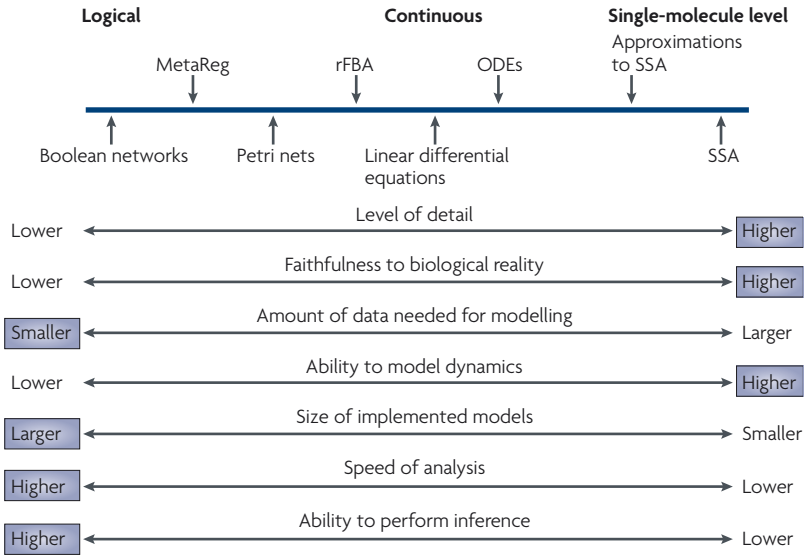



Figure 5 | **A schematic comparison of regulatory network models.** Models are listed along an imaginary scale, in which the level of detail of the models decreases, and the amount of detail increases, from left to right. Several pertinent criteria are indicated below the scale. Boolean networks are the purest form of logical models. They are highly abstract and hence require the least amount of data, but at the same time can display only qualitative dynamic behaviour. MetaReg is closer to biological reality because it can express intermediate regulator concentrations and accommodate probabilities, but requires more knowledge about the network and is limited to analysis of steady states. Petri nets can reveal finer detail to metabolic and signalling networks, and can therefore be used to describe integrated regulatory and metabolic/signalling networks and handle some dynamics. The analysis Petri nets offer is still qualitative. Regulated flux balance analysis (rFBA) produces metabolic predictions that can be compared to experimental measurements, but requires biochemical knowledge and is more challenging to analyse. Linear differential equations can model and predict experimentally observed concentrations of regulatory entities, and possess more detailed dynamics than the former models. General ordinary differential equations (ODEs) are more consistent with biochemical mechanisms than linear ODEs, but are harder to analyse. Single-molecule level models, the most detailed, can capture stochasticity, but are computationally expensive. To deal with this computational burden, approximations to stochastic simulation algorithms (SSAs) were developed, which sacrifice some detail for better performance. Methods that infer particular properties (not shown) can fall anywhere on the left half of the scale, depending on the properties of the chosen model.

## Summary

The introduction of novel and powerful experimental methods for studying gene regulation has created an upsurge of interest in modelling regulatory networks. In this article, three approaches to modelling were highlighted and some representative examples were discussed. We also discussed key differences among these approaches and rules of thumb for selecting an appropriate model (FIG. 5). Available modelling tools from each approach, as well as relevant databases, are listed in Supplementary information S1 (table).

A model's quality can be assessed by how similar its predictions are to experimental data. If two models generate predictions that match the same data equally well, then the simpler model is preferable, because it can be better understood and is less prone to over-fitting. When available observations are qualitative in nature, logical models can be accurate and have the advantage of having a modest number of global states. This enables more intuitive and efficient analysis methods. When data include real-valued measurements, such as time[74] or space[55,114], real-valued predictions can be more accurate. In addition, the simplified dynamics of logical models are often less appropriate for the complex behaviours that generate such measurements, and this motivates the use of continuous models or models that combine logical and continuous approaches.

The stochastic nature of gene expression influences the dynamics of regulatory networks, and this aspect is usually not modelled by continuous approaches[115]. Single-molecule level models are the most detailed and can explain stochastic behaviour in several scenarios. While accounting for the full complexity of gene regulation, single-molecule level models are also the hardest to study analytically, and stochastic experimental data are currently very scarce.

Limited availability of reaction rate constants and incomplete understanding of gene regulation are major impediments for building accurate models. In this respect, lower model resolution is an advantage, as it requires fewer parameters and less detailed understanding of the regulatory mechanisms[116–123]. Analytical methods that cope with these problems were developed for logical and continuous models, and some of these were presented above. As a brute force alternative, the space of potential parameters can be scanned for certain dynamic behaviours, provided that the model is both computationally simple and has a sufficiently small number of global states. For example, one study[103] searches the parameter space of a continuous model and derives molecular level parameters from results. Another problem associated with building accurate models is that experimental data are usually derived from a population of cells that needs to be synchronized[124]. The mean behaviour of a population (for example, as measured by gene expression) does not always exhibit fluctuations that can be observed at a single cell level. In such cases, the accuracy of deterministic and stochastic approaches is equally limited.

Despite substantial progress in modelling regulatory networks over the past decade, nature's design of regulatory networks confronts us with many open questions. Although it is clear that structure alone does not determine network dynamics[125,126], the role of different network architectures in generating dynamic behaviours[127,128], and the evolutionary processes that produced them, are far from understood. And, what is the effect of noise on regulatory networks? General strategies for overcoming stochastic effects are known[82], but a large-scale quantitative study has not yet been performed. Stochastic effects can also give rise to evolutionary advantages in a population by creating diversity[129,130]. Characterization of the beneficial role of stochasticity remains a future challenge. Notably, stochastic effects have been extensively studied in other types of dynamic biological systems, including population genetics and theoretical ecology[131–134].

Our current picture of how regulation is carried out is probably still missing several significant pieces. More experimental work is needed, and we must incorporate results into improved network models. Experimental design approaches[76–78] will help us to select the most efficient

set of experiments. In addition to understanding regulation as a stand-alone process, models for the interplay of regulation with other processes, for example metabolism and cell–cell signalling, need to be created[135,136]. The benefits of accurate, large-scale regulatory network models for medicine and biotechnology provide a strong incentive for cooperation between experimentalists and computational scientists.


1. Weindruch, R. & Walford, R. L. *The Retardation Of Aging And Disease By Dietary Restriction* (Thomas, Springfield, 1988).
2. Glass, L. & Kauffman, S. A. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* **39**, 103–129 (1973).
3. Thomas, R. Boolean formalization of genetic control circuits. *J. Theor. Biol.* **42**, 563–585 (1973).
4. Davidson, E. H. *et al.* A genomic regulatory network for development. *Science* **295**, 1669–1678 (2002).
**A highly detailed reconstruction of the regulatory network that controls the first 24 hours of sea urchin embryo development.**
5. Smith, J., Theodoris, C. & Davidson, E. H. A gene regulatory network subcircuit drives a dynamic pattern of gene expression. *Science* **318**, 794–797 (2007).
6. Kauffman, S. A. *The Origins Of Order: Self-Organization And Selection In Evolution* (Oxford University Press, Oxford, 1993).
**An accessible description of Kauffman's work, including the Boolean network model and its analysis.**
7. Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. Random Boolean network models and the yeast transcriptional network. *Proc. Natl Acad. Sci. USA* **100**, 14796–14799 (2003).
8. Li, F., Long, T., Lu, Y., Ouyang, Q. & Tang, C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA* **101**, 4781–4786 (2004).
9. Lähdesmäki, H., Shmulevich, I. & Yli-Harja, O. On learning gene regulatory networks under the Boolean network model. *Machine Learning* **52**, 147–167 (2003).
10. Akutsu, T., Miyano, S. & Kuhara, S. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac Symp. Biocomput.*, 17–28 (1999).
11. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
12. Shmulevich, I., Dougherty, E. R., Kim, S. & Zhang, W. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18**, 261–274 (2002).
13. Shmulevich, I., Gluhovsky, I., Hashimoto, R. F., Dougherty, E. R. & Zhan, W. Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. *Comp. Funct. Genomics* **4**, 601–608 (2003).
14. Bhattacharya, R. N. & Majumdar, M. *Random Dynamical Systems: Theory And Applications* (Cambridge University Press, Cambridge, 2007).
15. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. *Markov Chain Monte Carlo In Practice* (Chapman & Hall, Boca Raton, 1998).
16. Gat-Viks, I., Tanay, A. & Shamir, R. Modeling and analysis of heterogeneous regulation in biological networks. *J. Comput. Biol.* **11**, 1034–1049 (2004).
17. Gat-Viks, I., Tanay, A., Raijman, D. & Shamir, R. A probabilistic methodology for integrating knowledge and experiments on biological networks. *J. Comput. Biol.* **13**, 165–181 (2006).
**An exposition of the MetaReg methodology, including the probabilistic layer.**
18. Kschischang, F. R., Frey, B. J. & Loeliger, H. A. Factor graphs and the sum-product algorithm. *IEEE Trans. Info. Theory* **47**, 498–519 (2001).
19. MacKay, D. J. C. *Introduction To Monte Carlo Methods In Learning In Graphical Models* (ed. Jordan, M. I.) (Kluwer Academic Press, New York, 1998).
20. Gat-Viks, I. & Shamir, R. Refinement and expansion of signaling pathways: the osmotic response network in yeast. *Genome Res.* **17**, 358–367 (2007).
21. Petri, C. A. Kommunikation mit Automaten. *Schriften des Instituts für Instrumentelle Mathematik* (1962).
22. Koch, I., Schueler, M. & Heiner, M. STEPP — search tool for exploration of Petri net paths: a new tool for Petri net-based path analysis in biochemical networks. *In Silico Biol.* **5**, 129–137 (2005).
23. Reddy, V. N., Liebman, M. N. & Mavrovouniotis, M. L. Qualitative analysis of biochemical reaction systems. *Comput. Biol. Med.* **26**, 9–24 (1996).
24. Kuffner, R., Zimmer, R. & Lengauer, T. Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics* **16**, 825–836 (2000).
25. Chaouiya, C., Remy, E., Ruet, P. & Thieffry, D. in *Proceedings of the 25th International Conference on Applications and Theory of Petri Nets* (eds Cortadella, J. & Reisig, W.) (Springer, Berlin, 2004).
26. Simao, E., Remy, E., Thieffry, D. & Chaouiya, C. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in *E. coli*. *Bioinformatics* **21**, 190–196 (2005).
27. Steggles, L. J., Banks, R., Shaw, O. & Wipat, A. Qualitatively modelling and analysing genetic regulatory networks: a Petri net approach. *Bioinformatics* **23**, 336–343 (2007).
28. Peterson, J. *Petri Net Theory and the Modeling of Systems* (Prentice Hall PTR, New Jersey, 1981).
29. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, San Francisco, 1988).
**An excellent source on the foundation of probabilistic inference.**
30. Segal, E. *et al.* Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genet.* **34**, 166–176 (2003).
31. Friedman, N. Inferring cellular networks using probabilistic graphical models. *Science* **303**, 799–805 (2004).
32. Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* **7**, 601–620 (2000).
**A pioneering work that describes the use of Bayesian networks for recovering regulatory interactions from experimental data.**
33. Yeang, C. H., Ideker, T. & Jaakkola, T. Physical network models. *J. Comput. Biol.* **11**, 243–262 (2004).
34. Yeang, C. H. & Vingron, M. A joint model of regulatory and metabolic networks. *BMC Bioinformatics* **7**, 332 (2006).
35. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
36. Sauer, U. *et al.* Physiology and metabolic fluxes of wild-type and riboflavin-producing *Bacillus subtilis*. *Appl. Environ. Microbiol.* **62**, 3687–3696 (1996).
37. Ness, S. A. Basic microarray analysis: strategies for successful experiments. *Methods Mol. Biol.* **316**, 13–33 (2006).
38. Kingsmore, S. F. Multiplexed protein measurement: technologies and applications of protein and antibody arrays. *Nature Rev. Drug Discov.* **5**, 310–320 (2006).
39. Hellerstein, M. K. *In vivo* measurement of fluxes through metabolic pathways: the missing link in functional genomics and pharmaceutical research. *Annu. Rev. Nutr.* **23**, 379–402 (2003).
40. Yeung, M. K., Tegner, J. & Collins, J. J. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl Acad. Sci. USA* **99**, 6163–6168 (2002).
41. Golub, G. H. & Van Loan, C. F. *Matrix computations* (Johns Hopkins University Press, Maryland, 1996).
42. Weaver, D. C., Workman, C. T. & Stormo, G. D. Modeling regulatory networks with weight matrices. *Pac Symp. Biocomput.*, 112–123 (1999).
43. Bansal, M., Gatta, G. D. & di Bernardo, D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles. *Bioinformatics* **22**, 815–822 (2006).
44. Dasika, M. S., Gupta, A. & Maranas, C. D. A mixed integer linear programming (MILP) framework for inferring time delay in gene regulatory networks. *Pac Symp. Biocomput*, 474–485 (2004).
45. Nachman, I., Regev, A. & Friedman, N. Inferring quantitative models of regulatory networks from expression data. *Bioinformatics* **20**, 248–256 (2004).
46. Klipp, E. *Systems Biology In Practice: Concepts, Implementation And Application* (Wiley-VCH, Weinheim, 2005).
**Explains basic concepts in modelling biological networks and provides an excellent introduction for the use of ODEs.**
47. Friedman, N., Murphy, K. & Russell, S. in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence* (eds Cooper, G. F. & Moral, S.) 129–138 (Morgan Kaufmann, San Francisco, 1998).
48. Kim, S. Y., Imoto, S. & Miyano, S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief Bioinform.* **4**, 228–235 (2003).
49. Shamir, R. & Tanay, A. in *RECOMB'03* (ACM, Berlin, 2003).
50. Pan, Y., Durfee, T., Bockhorst, J. & Craven, M. Connecting quantitative regulatory-network models to the genome. *Bioinformatics* **23**, 367–376 (2007).
51. Li, Z., Shaw, S. M., Yedwabnick, M. J. & Chan, C. Using a state-space model with hidden variables to infer transcription factor activities. *Bioinformatics* **22**, 747–754 (2006).
52. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. & Gaul, U. Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. *Nature* **451**, 535–540 (2008).
53. Goodwin, B. C. *Temporal Organization In Cells; A Dynamic Theory Of Cellular Control Processes* (Academic Press, New York, 1963).
54. Li, S., Brazhnik, P., Sobral, B. & Tyson, J. J. A quantitative study of the division cycle of *Caulobacter crescentus* stalked cells. *PLoS Comput. Biol.* **4**, e9 (2008).
55. McAdams, H. H. & Shapiro, L. A bacterial cell-cycle regulatory network operating in time and space. *Science* **301**, 1874–1877 (2003).
56. Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M. & Shapiro, L. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**, 2144–2148 (2000).
57. Holtzendorff, J. *et al.* Oscillating global regulators control the genetic circuit driving a bacterial cell cycle. *Science* **304**, 983–987 (2004).
58. Biondi, E. G. *et al.* Regulation of the bacterial cell cycle by an integrated genetic circuit. *Nature* **444**, 899–904 (2006).
59. Chen, K. C. *et al.* Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell* **15**, 3841–3862 (2004).
**An ODE model for the yeast cell-cycle regulatory network that is supported by phenotypes of more than 100 mutant strains.**
60. Locke, J. C. *et al.* Extension of a genetic network model by iterative experimentation and mathematical analysis. *Mol. Syst. Biol.* **1**, 0013 (2005).
61. Klipp, E., Nordlander, B., Kruger, R., Gennemark, P. & Hohmann, S. Integrative model of the response of yeast to osmotic shock. *Nature Biotechnol.* **23**, 975–982 (2005).
62. Gebert, J., Radde, N. & Weber, G. Modeling gene regulatory networks with piecewise linear differential equations. *Eur. J. Oper. Res.* **181**, 1148 (2007).
63. Ropers, D., de Jong, H., Page, M., Schneider, D. & Geiselmann, J. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *Biosystems* **84**, 124–152 (2006).
64. Mèuller-Hill, B. *The lac Operon: A Short History Of A Genetic Paradigm* (Walter de Gruyter, Berlin, 1996).
65. Covert, M. W., Schilling, C. H. & Palsson, B. Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73–88 (2001).
66. Kauffman, K. J., Prakash, P. & Edwards, J. S. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* **14**, 491–496 (2003).
67. Palsson, B. Ø. *Systems Biology: Properties Of Reconstructed Networks* (Cambridge University Press, New York, 2006).
**An excellent book about the study of biological networks by one of the leading researchers in this field.**
68. Edwards, J. S., Covert, M. & Palsson, B. Metabolic modelling of microbes: the flux-balance approach. *Environ. Microbiol* **4**, 133–140 (2002).

69. Beg, Q. K. *et al.* Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc. Natl Acad. Sci. USA* **104**, 12663–12668 (2007).
70. Duarte, N. C., Herrgard, M. J. & Palsson, B. O. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* **14**, 1298–1309 (2004).
71. Resendis-Antonio, O., Reed, J. L., Encarnacion, S., Collado-Vides, J. & Palsson, B. O. Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*. *PLoS Comput. Biol.* **3**, 1887–1895 (2007).
72. Feist, A. M. *et al.* A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121 (2007).
73. Papin, J. A., Price, N. D., Edwards, J. S. & Palsson, B. B. The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy. *J. Theor. Biol.* **215**, 67–82 (2002).
74. Covert, M. W. & Palsson, B. O. Transcriptional regulation in constraints-based metabolic models of *Escherichia coli*. *J. Biol. Chem.* **277**, 28058–28064 (2002).
75. Covert, M. W., Knight, E. M., Reed, J. L., Herrgard, M. J. & Palsson, B. O. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–96 (2004).
**A large-scale model of *E. coli* metabolic regulation that uses quantitative metabolic flux values.**
76. Barrett, C. L. & Palsson, B. O. Iterative reconstruction of transcriptional regulatory networks: an algorithmic approach. *PLoS Comput. Biol.* **2**, e52 (2006).
77. Ideker, T. E., Thorsson, V. & Karp, R. M. Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac Symp. Biocomput*, 305–316 (2000).
78. Yeang, C. H. *et al.* Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol.* **6**, R62 (2005).
79. Herrgard, M. J., Lee, B. S., Portnoy, V. & Palsson, B. O. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* **16**, 627–635 (2006).
80. Shlomi, T., Eisenberg, Y., Sharan, R. & Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.* **3**, 101 (2007).
81. Gianchandani, E. P., Papin, J. A., Price, N. D., Joyce, A. R. & Palsson, B. O. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput. Biol.* **2**, e101 (2006).
82. McAdams, H. H. & Arkin, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* **15**, 65–69 (1999).
83. Ross, I. L., Browne, C. M. & Hume, D. A. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* **72**, 177–185 (1994).
84. Bae, K., Lee, C., Hardin, P. E. & Edery, I. dCLOCK is present in limiting amounts and likely mediates daily interactions between the dCLOCK–CYC transcription factor and the PER–TIM complex. *J. Neurosci.* **20**, 1746–1753 (2000).
85. Guptasarma, P. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays* **17**, 987–997 (1995).
86. Bailone, A., Levine, A. & Devoret, R. Inactivation of prophage λ repressor *in vivo*. *J. Mol. Biol.* **131**, 553–572 (1979).
87. Shea, M. A. & Ackers, G. K. The OR control system of bacteriophage λ. A physical-chemical model for gene regulation. *J. Mol. Biol.* **181**, 211–230 (1985).
88. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036 (2005).
89. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
90. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
91. Cai, L., Friedman, N. & Xie, X. S. Stochastic protein expression in individual cells at the single molecule level. *Nature* **440**, 358–362 (2006).
92. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. Probing gene expression in live cells, one protein molecule at a time. *Science* **311**, 1600–1603 (2006).
93. Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55 (2007).
94. McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
**McAdams and Arkin describe their modelling methodology, which uses Gillespie's SSA, and present simulation results of a simple system that exhibits protein bursts.**
95. Gillespie, D. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.* **22**, 403–434 (1976).
**Gillespie presents his SSA, the understanding of which is of major importance for computational modelling.**
96. Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
97. Halford, S. E. & Marko, J. F. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
98. Terry, L. J., Shows, E. B. & Wente, S. R. Crossing the nuclear envelope: hierarchical regulation of nucleocytoplasmic transport. *Science* **318**, 1412–1416 (2007).
99. Rapoport, T. A., Jungnickel, B. & Kutay, U. Protein transport across the eukaryotic endoplasmic reticulum and bacterial inner membranes. *Annu. Rev. Biochem.* **65**, 271–303 (1996).
100. Gonze, D. & Goldbeter, A. Circadian rhythms and molecular noise. *Chaos* **16**, 026110 (2006).
101. Niemitalo, O. *et al.* Modelling of translation of human protein disulfide isomerase in *Escherichia coli* — a case study of gene optimisation. *J. Biotechnol.* **120**, 11–24 (2005).
102. Schultz, D., Ben Jacob, E., Onuchic, J. N. & Wolynes, P. G. Molecular level stochastic model for competence cycles in *Bacillus subtilis*. *Proc. Natl Acad. Sci. USA* **104**, 17582–17587 (2007).
103. Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. & Schaffer, D. V. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell* **122**, 169–182 (2005).
104. Arkin, A., Ross, J. & McAdams, H. H. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics* **149**, 1633–1648 (1998).
**A seminal work that quantitatively explains how phage λ makes a stochastic decision between lysis and lysogeny using regulatory network simulation.**
105. Gibson, M. & Bruck, J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.* **104**, 1876–1889 (1999).
106. Gillespie, D. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**, 1716–1733 (2001).
107. Li, H., Cao, Y., Petzold, L. R. & Gillespie, D. T. Algorithms and software for stochastic simulation of biochemical reacting systems. *Biotechnol. Prog.* **24**, 56–61 (2008).
108. Kiehl, T. R., Mattheyses, R. M. & Simmons, M. K. Hybrid simulation of cellular behavior. *Bioinformatics* **20**, 316–322 (2004).
109. Griffith, M., Courtney, T., Peccoud, J. & Sanders, W. H. Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network. *Bioinformatics* **22**, 2782–2789 (2006).
110. Takahashi, K., Kaizu, K., Hu, B. & Tomita, M. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics* **20**, 538–546 (2004).
111. Ribeiro, A., Zhu, R. & Kauffman, S. A. A general modeling strategy for gene regulatory networks with stochastic dynamics. *J. Comput. Biol.* **13**, 1630–1639 (2006).
112. Zhu, R., Ribeiro, A. S., Salahub, D. & Kauffman, S. A. Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. *J. Theor. Biol.* **246**, 725–745 (2007).
113. Rao, C. & Arkin, A. Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm. *J. Chem. Phys.* **118**, 4999–5010 (2003).
114. Furlong, E. E. A topographical map of spatiotemporal patterns of gene expression. *Dev. Cell* **14**, 639–640 (2008).
115. Gillespie, D. The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306 (2000).
116. Bernstein, J. A., Khodursky, A. B., Lin, P. H., Lin-Chao, S. & Cohen, S. N. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA* **99**, 9697–9702 (2002).
117. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
118. van Driel, R. & Fransz, P. Nuclear architecture and genome functioning in plants and animals: what can we learn from both? *Exp. Cell Res.* **296**, 86–90 (2004).
119. Ogden, S. K. *et al.* p53 targets chromatin structure alteration to repress α-fetoprotein gene expression. *J. Biol. Chem.* **276**, 42057–42062 (2001).
120. Wong, W. W., Tsai, T. Y. & Liao, J. C. Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator. *Mol. Syst. Biol.* **3**, 130 (2007).
121. Tetko, I. V. *et al.* Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in *Arabidopsis thaliana*. *PLoS Comput. Biol.* **2**, e21 (2006).
122. Chapman, R. D. *et al.* Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science* **318**, 1780–1782 (2007).
123. Brodersen, P. *et al.* Widespread translational inhibition by plant miRNAs and siRNAs. *Science* **320**, 1185–1190 (2008).
124. Zhou, T. *et al.* Identification of primary transcriptional regulation of cell cycle-regulated genes upon DNA damage. *Cell Cycle* **6**, 972–981 (2007).
125. Mangan, S. & Alon, U. Structure and function of the feed-forward loop network motif. *Proc. Natl Acad. Sci. USA* **100**, 11980–11985 (2003).
126. Ingram, P. J., Stumpf, M. P. & Stark, J. Network motifs: structure does not determine function. *BMC Genomics* **7**, 108 (2006).
127. Remy, É., Ruet, P. & Thieffry, D. Graphic requirements for multistability and attractive cycles in a Boolean dynamical framework. *Adv. Appl. Math.* (2008).
128. Richard, A. & Comet, J. Necessary conditions for multistationarity in discrete dynamical systems. *Dis. Appl. Math.* **155**, 2403–2413 (2007).
129. Stern, S., Dror, T., Stolovicki, E., Brenner, N. & Braun, E. Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Mol. Syst. Biol.* **3**, 106 (2007).
130. Bar-Even, A. *et al.* Noise in protein expression scales with natural protein abundance. *Nature Genet.* **38**, 636–643 (2006).
131. Fisher, R. A. On the dominance ratio. Reprint in *Bull. Math. Biol.* **52**, 297–318 (1990).
132. May, R. M. *Theoretical Ecology: Principles And Applications* (Blackwell Scientific Publications, Oxford, 1981).
133. Nei, M. *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
134. Wright, S. Evolution in mendelian populations. *Genetics* **16**, 97–159 (1931).
135. *Min.* Lee, J., Gianchandani, E. P., Eddy, J. A. & Papin, J. A. Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput. Biol.* **4**, e1000086 (2008).
136. Tomita, M. *et al.* E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**, 72–84 (1999).
137. McAdams, H. H. & Shapiro, L. Circuit simulation of genetic networks. *Science* **269**, 650–656 (1995).
138. Kierzek, A. M. STOCKS: Stochastic kinetic simulations of biochemical systems with Gillespie algorithm. *Bioinformatics* **18**, 470–481 (2002).



**Acknowledgements**
We thank I. Ulitsky and I. Gat-Viks for their comments on the manuscript. This study was supported in part by a grant to the APO-SYS consortium from the European Community's Seventh Framework Programme, and by a grant from the Ministry of Science, Culture and Sport, Israel, and the Ministry of Research, France. G.K. was supported in part by a fellowship from the Edmond J. Safra Bioinformatics program at Tel Aviv University, Israel.


## FURTHER INFORMATION

**Ron Shamir's homepage:**
www.cs.tau.ac.il/~rshamir

## SUPPLEMENTARY INFORMATION

**See online article:** S1 (table)

**ALL LINKS ARE ACTIVE IN THE ONLINE PDF**