COMPUTATIONAL STUDIES OF GENE REGULATORY NETWORKS: *IN NUMERO* MOLECULAR BIOLOGY

Jeff Hasty, David McMillen, Farren Isaacs and James J. Collins

Remarkable progress in genomic research is leading to a complete map of the building blocks of biology. Knowledge of this map is, in turn, setting the stage for a fundamental description of cellular function at the DNA level. Such a description will entail an understanding of gene regulation, in which proteins often regulate their own production or that of other proteins in a complex web of interactions. The implications of the underlying logic of genetic networks are difficult to deduce through experimental techniques alone, and successful approaches will probably involve the union of new experiments and computational modelling techniques.

NONLINEAR DYNAMICS
In a system governed by nonlinear dynamics, the rate of change of any variable cannot be written as a linear function of the other variables. Most real systems are nonlinear and show interesting behaviours not seen in linear systems (for example, only nonlinear systems can be multistable).

STOCHASTIC
Probabilistic; governed by chance.

Centre for BioDynamics and Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts 02215, USA.
Correspondence to J.H.
e-mail: hasty@bu.edu

An important theme in post-genomic research will probably be the dissection and analysis of the complex dynamical interactions involved in gene regulation. Although the concepts of protein–DNA feedback loops and network complexity are not new, experimental advances are inducing a resurgence of interest in the quantitative description of gene regulation. These advances are beginning to allow a 'modular' description of the regulatory processes that underlie basic cellular function[1–5]. In the light of nearly three decades of parallel progress in the study of complex NONLINEAR and STOCHASTIC processes, the project of quantitatively describing gene regulatory networks is timely.

Pioneering theoretical work on gene regulatory networks has anticipated the emergence of post-genomic research, and has provided a mathematical framework for the current description and analysis of complex regulatory mechanisms[6–18]. Although these studies have identified the need for a quantitative description of gene regulation, their true significance has only recently emerged with experimental techniques that can determine their validity. The diagram in FIG. 1 depicts some of the known components of the regulatory network that involve the tumour-suppressor protein p53 (REFS 19–20). These types of schematic resemble circuit diagrams[2,21] and in many regards this analogy highlights the motivation for a quantitative description of gene regulation. If this were a complex electrical circuit, there would be an accompanying set of equations that would faithfully describe its functionality. This description would be built from a knowledge of the properties of the individual components (resistors, capacitors, inductors and so on) and provide a framework for predicting behaviour that results from modification of the circuit. An acceptable model that describes the p53 activation network (FIG. 1) would thus be built from knowledge of the basic regulatory themes and could predict the effects of genetic perturbations to the system.

In this article, we review recent advances in the mathematical modelling of genetic regulation. Most of this work has focused on networks that involve transcription factors and we restrict ourselves to work in this class. We begin with the modelling of specific genetic networks and discuss representative models that have been used for several relatively simple networks. We then turn to the recent progress in designing and testing synthetic gene networks. Although these networks have important biotechnological implications in their own right[22,23], we highlight their use in determining the primary themes of gene regulatory networks. In this regard, the accurate mathematical description of synthetic networks provides the foundation for describing complex, naturally

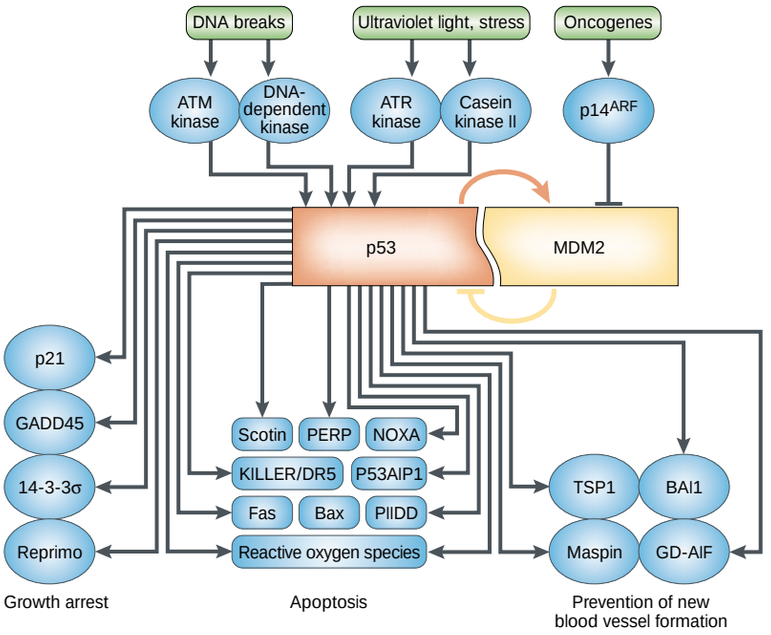

Figure 1 | **Regulatory diagram for the activation of the tumour-suppressor protein p53.** The complexity of the p53 network highlights the need for a quantitative description of genetic circuitry. (ATM, ataxia telangiectasia mutated; ATR, AT and Rad3-related; MDM2, mouse double minute 2; GADD45, growth arrest and DNA-damage inducible; PERP, p53 apoptosis-associated target; KILLER/DR5, death receptor 5; P53AIP1, p53-regulated apoptosis-inducing protein 1; Bax, Bcl2-associated X protein; PIDD, p53 protein induced, with death domain; TSP1, tumour-suppressor region 1; BAI1, brain-specific angiogenesis inhibitor 1; GD-AIF, glioma-derived angiogenesis inhibitory factor.)(Redrawn with permission from REF. 19 © (2000) Macmillan Magazines Ltd.)

occurring networks. Throughout this review, we ground our discussion through the use of illustrative examples and primarily focus on modelling efforts that are directly connected to experiments.

Modelling genetic networks

Several traditional approaches to analysing gene regulation have been based on modelling specific natural systems (BOX 1). By contrast, some researchers have concentrated on analysing abstract models to obtain general results. The power of the latter approach is that it can offer insight into the behaviour of entire classes of biological system. This approach has recently been reviewed in some detail[24] and here we discuss several representative examples.

In a typical experimental situation, regulatory networks are quantified by the concentrations of the constitutive gene products. When a gene product is in an equilibrium state (that is, its rate of production is balanced by its rate of degradation), small perturbations to this steady state will be accompanied by an exponential decay or rise back to equilibrium. Several studies have focused on the conditions that are required for the existence of multiple steady states in generic gene regulatory systems[13,25–29]. The existence of more than one stable FIXED POINT implies the possibility of switch-like behaviour: the system will remain near one of its stable states until a sufficiently large perturbation drives it into the vicinity of another stable state, in which it will remain even after the perturbation is removed. Understanding how MULTISTABILITY arises is thus relevant to understanding the operation of natural biological switches (such as the lysogeny/lysis switches that occur in viruses, for example λ-bacteriophage), as well as to the design of synthetic switching networks. A recent analysis[29] considered a generic rate-equation model to determine the precise conditions required for the existence of multiple stable fixed points in a two-gene system. Another study[28] used a model (derived from REF. 17) to determine the relationship between the number of OPERATOR sites that constitute a given promoter and the number of stable steady states that the system could support (see below).

In addition to considering the existence of stable fixed points, their degree of stability can be considered. States vary in how quickly they recover from perturbations, the size of perturbation they can withstand before being forced into another state and, in noisy systems, the expected length of time it will take for noise to induce a transition to another state. NEGATIVE FEEDBACK increases stability in generic gene regulatory systems, whereas POSITIVE FEEDBACK decreases stability[7]. It has often been assumed that switches based on the action of small numbers of individual molecules could not be very stable, because small numbers of molecules generally imply that the system will be subjected to larger fluctuations than systems with greater numbers of molecules. However, a recent analysis, using a general formulation applicable to any biological switch[30], indicates that switches based on only tens of molecules could flip states in milliseconds and remain stable for years. This degree of stability is an upper bound and thus does not necessarily indicate how stable any particular biochemical switch will be; however, it does point out the possibility of achieving the required stability. With larger numbers of molecules (hundreds rather than tens), even greater degrees of stability should be achievable.

Many cellular processes are characterized by oscillations that are generated at the genetic level and several investigations have focused on the general conditions under which oscillations are to be expected from a given gene network[27,31–33]. One of these[33] concludes that oscillatory behaviour cannot exist in systems with only positive-feedback interactions; this result applies to systems with and without time delays. Additionally, systems with only negative feedback can generate oscillations in the presence of time delays[34], and mixed positive and negative feedback can also generate oscillatory behaviour[32,33].

Naturally occurring gene networks

Although abstract models can offer an insight into basic mechanisms, modelling must ultimately be connected to specific systems so that verifiable predictions can be made. So far, few computational modelling studies have involved tight coupling between modelling and experiment. Part of the difficulty is the high degree of complexity inherent in natural systems and the difficulty of carrying out experiments on them.

FIXED POINT
A point at which the rates of change of all variables in a system are exactly zero. A system precisely at its fixed point (or steady state) will remain there permanently. Small perturbations to a system that is initially poised at a 'stable' fixed point will be accompanied by a return to the stable fixed point.

MULTISTABILITY
The property of having more than one stable fixed point.

OPERATOR
A prokaryotic DNA regulatory element that interacts with a repressor to control the transcription of adjacent genes.

NEGATIVE FEEDBACK
A component of a system is subject to negative feedback when it inhibits its own level of activity. For example, a gene product that acts as a repressor for its own gene is applying negative feedback.

Box 1 | Modelling methods

Several levels of detail have traditionally been used in modelling gene regulation. In the 'logical' or 'binary' approach[6–10], each gene is treated as having two discrete states, ON or OFF, and the dynamics describe how groups of genes act to change one another's states over time. Such models are relatively easy to implement, simplifying the examination of large sets of genes. A disadvantage of the logical approach is that the abstraction of genes to ON/OFF switches makes it difficult or impossible to include many of the details of cellular biology.

A more detailed level of description is used in the 'chemical kinetics' or 'rate-equation' approach, in which the variables of interest are the concentrations of individual proteins within the cell, and the dynamics describe the rates of production and decay of these proteins. The models consist of a system of ordinary differential equations, permitting the modeller to apply the analytical techniques of nonlinear dynamics. These techniques have been developed considerably in recent decades, making the rate-equation approach a promising avenue for combining mathematical analysis and computational simulation.

Although the basic rate-equation approach is completely deterministic (no random component exists in the dynamical equations), the equations can be augmented with noise terms to account for fluctuations in concentration within the cell[30,94]. The 'stochastic kinetics' modelling approach[3,21,58] provides the most detailed level of description; techniques for simulating the behaviour of chemical reactions, which involve small numbers of molecules[105], are applied to the reactions involved in protein–DNA binding, transcription and translation. This approach is impressively complete and yields a detailed picture of the behaviour of the system modelled. However, such completeness comes at a high computational cost and sacrifices any immediate prospect of analytical treatment. Alternatively, the effects of internal noise can be incorporated into stochastic terms, which have magnitudes that are concentration dependent. This approach has recently been used to explore the reliability of genetic switches in the presence of internal noise[30]. The advantage of this formulation is that stochastic effects can be viewed as a perturbation to the deterministic picture, so that analytical tools can be used.

Each of the modelling techniques has its own merits and drawbacks, and the appropriate level of description for a given system depends on the goals of the modelling effort and the nature of the experiments proposed.

This complexity indicates that modelling efforts should centre on describing relatively simple systems and be closely linked with experiments. We now turn to several representative examples in this class. As we focus on systems in which the principal means of control is through transcription factors, we will not discuss closely related topics, such as bacterial chemotaxis[35–41], intracellular signalling[42–45], and cell-cycle control[1,46–50]. The modelling techniques used in these important studies are similar in spirit to the work presented here and the reader is encouraged to consult the primary sources for further investigation.

Bacteriophage-λ. The λ-bacteriophage has been a fertile topic for modelling studies. The biochemical reactions that constitute the control of λ-phage are well characterized and the fundamental biochemical reactions are understood[51–55]. When the virus infects a bacterium such as *Escherichia coli*, one of two developmental pathways is followed. In the lysogenic pathway, the viral DNA is incorporated into the genome of the host and the virus lies dormant, replicating with the bacterium. In the lytic pathway, the virus expresses the proteins necessary to replicate new phages, then lyses the host cell and releases its progeny into the environment. When the host DNA is damaged, which is signalled by activation of the bacterial protein RecA (Recombinase A), the virus can switch from lysogeny to lysis, allowing it to 'abandon ship' when the host is threatened. The key section of the phage DNA lies in the right operator region (O_R), in which three DNA-binding sites are recognized by two phage-encoded regulatory proteins, λ-repressor protein (also called cI) and cro. The three operator sites (O_R1 to O_R3) overlap the promoter regions of the genes that encode these same proteins: the P_{RM} (where RM is repressor maintenance) promoter controls expression of *cI* and the P_R (where R is repressor) promoter controls expression of *cro* (FIG. 2a). The pattern of cI/cro binding to the three operator sites determines whether the lysogenic or lytic pathway will be followed (see REF. 56 for an excellent review).

It is often the case that multiple operator sites constitute a promoter region. In such cases, one would like to know the probability that a particular protein–DNA binding configuration will occur. One generally applicable modelling technique uses thermodynamic quantities derived from experimental data to formulate a model and this technique has been applied to the λ-switch[16–17]. The λ-repressor protein concentrations at which the operator sites in the O_R region are half-occupied[57] can be converted into a set of thermodynamic free energies[16], which can be used to calculate the probability that the system will occupy a given state. For example, in the binding of λ-repressor to sites in O_R, the relevant states are eight possible patterns of binding of λ-repressor to the three operator sites.

By examining the states in which each promoter is repressed and calculating the probabilities of these states, it is possible to obtain a graph of the degree of repression of each promoter as a function of cI concentration (FIG. 2b). A curve might also be calculated for the repression of *cro* in the hypothetical situation in which adjacent cI proteins do not interact. This 'computational experiment', which would have been difficult or impossible to do in the laboratory, supported the view[16] that cooperative interactions stabilize the lysogenic state by tightening the repression of *cro* at lysogenic concentrations of λ-repressor.

The model was extended to include the binding of cro and RNA polymerase in addition to the λ-repressor[17]; in the resulting formulation, there were 40 possible binding states rather than 8. In addition, the previously static model was extended to include the dynamics, using a rate-equation approach. The behaviour of the resulting model was in qualitative agreement with the known biological behaviour of the infected cell. However, no quantitative evaluation of the validity of the model was possible.

Next, a rate-equation-based model was constructed in which the parameters were tuned using experimental results[18]. To be a reasonable model for the λ-switch, it was argued that the rate equations must show BISTABILITY. One stable state (high repressor concentration, low cro concentration) corresponds to a lysogenic cell; the other stable state (low repressor, high cro) corresponds to the situation prevailing in an anti-immune bacterium in which the lytic machinery has been decoupled from the repressor/cro switch, so that a high level of cro does not lyse the cell. The ranges of activity of the two promoters for which this bistability exists can be obtained from the model equations and the validity of the model can be

POSITIVE FEEDBACK
A component of a system is subject to positive feedback when it increases its own level of activity. For example, a gene product that activates the expression of its own gene is subject to positive feedback.

COOPERATIVITY
Interaction between binding sites in which the binding of one molecule modifies the ability of a subsequent molecule of the same type to bind to its binding site.

BISTABILITY
The property of having two stable fixed points. See also the definition for multistability.

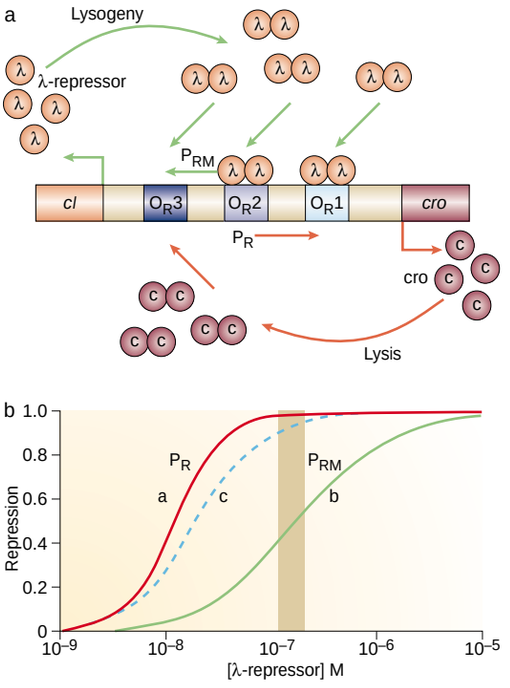

Figure 2 | **Modelling the λ-bacteriophage circuitry.** **a** | The right operator region (O_R) of bacteriophage-λ. The three operator sites O_R1, O_R2 and O_R3 bind two proteins: λ-repressor (λ, encoded by *cI*) and cro (c). Transcription from the P_{RM} promoter (which drives expression of *cI*) is blocked when either protein is bound to O_R3 and is enhanced when the λ-repressor binds to O_R2. High levels of λ-transcription lead to the lysogenic state (green arrows); high levels of cro transcription induce phage replication and lysis (red arrows). Transcription from the P_R promoter (which drives the expression of *cro*) has no stimulated state and proceeds only when O_R3 is either unoccupied or has cro bound to it; binding of either protein to O_R1 or O_R2 halts transcription of *cro*. **b** | Graph depicting the degree of repression of the P_R and P_{RM} promoters as a function of λ-repressor concentration. The dashed curve indicates the repression curve that would exist if there were no cooperative interactions between adjacent λ-repressor proteins. The graph shows that cooperative interactions between adjacent λ-repressor proteins tighten the repression of *cro* (from the P_R promoter) at lysogenic concentrations of repressor (shaded vertical box). (Redrawn from REF. 16.)

assessed by comparing these with experimentally determined values of the promoter activities. If the experimental parameters do not correspond to a bistable state, a flaw must exist in the model. This study provides an excellent example of tight coupling between model and experiment by considering a carefully chosen piece of a larger system (here, separating the operation of the switch from the rest of the bacterial metabolism). This approach has become even more attractive with the advent of tailor-made synthetic networks.

The central conclusion[18] was that the experimentally estimated parameter values did not place the model in a regime in which bistability existed. Clearly, this implies a difficulty with the assumptions underlying the model. The authors suggested that the model omitted some important aspect of the regulation of the switch and that such an aspect was also omitted from the standard 'word model' that describes the behaviour of the λ-switch. This speculation has not, so far, been either confirmed or refuted.

Another approach to modelling the λ-phage involved stochastic simulation techniques[58]. Rather than considering switching behaviour, this study addressed the initial decision between the two available developmental pathways (lysis and lysogeny) made shortly after a bacterium is first infected with the phage. Initially, the two key regulatory proteins are completely absent from the cell; in a deterministic system, fixing the initial condition [cI] = 0, [cro] = 0 would lead to identical results in every run of the simulation. In the stochastic simulation, the protein concentrations (including several proteins other than cI and cro) vary with a strong random component and the fluctuations are large enough to send some cells down the lysogenic pathway (those in which cI rises to a high enough concentration that it dominates and shuts off production of cro), whereas other cells proceed down the lytic pathway (those in which cro comes to dominate, shutting off cI production). The predictions of the model agree quite well with the results of experiments that measured the fraction of lysogens produced in a cell population[59] for various ratios of phage particles to cells. The power of the stochastic kinetics approach lies in its completeness and attention to detail. Although the simulations are computationally expensive compared with other methods, the *in numero* experiments are still rapid compared with *in vivo* work, allowing researchers to examine many hypotheses and concentrate effort on the most promising of them. This might lead, for example, to more rapid hypothesis testing, by indicating which experiments would be expected to distinguish most sharply among the competing hypotheses.

Other systems. Work on a similar system, the bacteriophage T7 (another lytic phage that preys on *E. coli*)[60] has also yielded computational results that have been compared directly with experimental predictions. An elaborate rate-equation model[61] tracks the behaviour of the T7 phage from initial injection of its genome through to the production of progeny phages; the concentrations of each of 52 mRNA transcripts and their corresponding protein products are simulated, along with rates of viral genome injection and progeny formation. Experimental comparisons were obtained by constructing mutant phages with genomes in which the physical position of one gene was altered relative to the wild type. Although the model achieved some success at predicting early mRNA concentrations and RNA polymerase activity[61], the attempt to predict the complete phage behaviour was largely unsuccessful. This underscores the difficulty of large-scale modelling of natural systems and the comparative immaturity of our understanding. Admirably, REF. 60 frankly discusses the mismatches between the model and the experiment, correctly viewing such discrepancies as opportunities for gaining insight into where the problems with the model might lie.

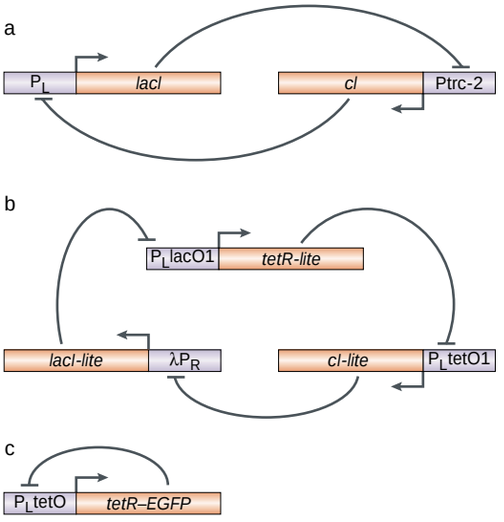

Figure 3 | **Schematic diagrams of three negatively regulated synthetic gene networks. a** | The toggle switch is composed of a two-gene co-repressive network. The constitutive P_L promoter drives the expression of the *lacI* gene, which produces the lac repressor tetramer. The lac repressor tetramer binds the lac operator sites adjacent to the Ptrc-2 promoter, thereby blocking transcription of *cI*. The constitutive Ptrc-2 promoter drives the expression of the *cI* gene, which produces the λ-repressor dimer. The λ-repressor dimer cooperatively binds to the operator sites native to the P_L promoter, which prevents transcription of *lacI*. **b** | The repressilator is composed of a three-gene repressive network driven by three strong constitutive promoters. Expression of *tetR-lite* is driven by the constitutive $P_L$lacO1 promoter. The tet repressor binds to the tetO1 operator sites on the $P_L$tetO1 promoter, turning off constitutive transcription of *cI-lite*. Transcription of *cI-lite* produces a λ-repressor protein, which binds to native operator sites on the P_R promoter. The constitutive P_R promoter drives the expression of *lacI-lite*, which produces lac repressor. The lac repressor binds to lacO1 operator sites on the $P_L$lacO1 promoter, thus completing the repressilator circuit. The '*-lite*' extensions on each gene describe the production of proteins that are encoded with a sequence that targets them for expedient degradation by native bacterial proteases. **c** | An autorepressive network is composed of a single-gene negative-feedback circuit. The strong constitutive P_L promoter with tetO operator sites drives the expression of the open reading frame consisting of the tetR and enhanced green fluorescent protein (EGFP) genes. Production of the tetR–EGFP fusion protein negatively regulates its own production by binding to the tetO operator sites on the P_L promoter.

The concept of OPERON regulation was introduced over 40 years ago[62–64] and a general descriptive theory arose shortly thereafter[65–66]. Several modelling efforts have focused on the dynamical behaviour of the *lac* (lactose) and *trp* (tryptophan) operons[67–77]. Tryptophan is an amino acid that is incorporated into proteins that are essential to bacterial growth. When tryptophan is present in the growth media, it forms a complex with the tryptophan repressor and the complex binds to the promoter of the *trp* operon, effectively switching off production of tryptophan biosynthetic enzymes. In the absence of tryptophan, the repressor cannot bind to the promoter and the essential tryptophan biosynthetic enzymes are produced. One of the first mathematical models of the tryptophan operon[69] used dynamical equations to describe operon repression by the *trp* repressor. The parameter estimates were based on experimental data and the model reproduced derepression experiments. However, the model omitted interactions between the *trp* operon and the repressor molecules. A more recent model of the *trp* operon[72,73] accounted for repression, enzyme feedback inhibition, inherent time delays and transcriptional attenuation (premature termination of transcription — a feature of the *trp* regulatory operon). Model parameters were closely estimated from experimental data, and numerical results from a system of differential equations were compared with experimental results. The model successfully predicted changes in the concentration of biosynthetic enzymes in bacteria grown in minimal media with and without a tryptophan supplement. In addition, simulations qualitatively reproduced identical growth experiments involving mutant *E. coli* strains.

A recent modelling study[78] used rate equations to examine the regulation of segmentation in the fruitfly *Drosophila melanogaster*. The large PARAMETER SPACE of the model was searched to find solutions that qualitatively matched the experimental data. The initial assumed pattern of gene network connectivity made it very difficult to find parameter sets that yielded the desired behaviour, whereas the addition of several key connections made such parameter sets relatively common. In earlier modelling work on *Drosophila* segmentation[79], the rate-equation model made no assumptions about the nature of the connections among genes. Rather, experimental data were used to determine the connections by searching the parameter space for those solutions best fitting the observed behaviour of the system. Both studies illustrate the use of model construction in the determination of underlying network connectivity. Such 'reverse-engineering' approaches are particularly attractive in the light of recent advances in gene chip technology[80–86].

Synthetic networks: the road to reductionism

The ability to design synthetic networks offers the exciting prospect of extracting carefully chosen subsystems from natural organisms, and focusing both modelling and experimental effort on determination of the behaviour of the subsystems in isolation. The long-range goal of such work would be to assemble increasingly complete models of the behaviour of natural systems, while maintaining at each stage the ability to test models in a tractable experimental system. Further, simple networks represent a first step towards logical cellular control, whereby biological processes can be manipulated or monitored at the DNA level[87]. Such control could have a significant effect on post-genomic biotechnology. Current examples of potential applicability range from the use of genetically engineered microorganisms for environmental clean-up purposes[22], to the flipping of genetic switches in mammalian neuronal cells[23]. From the construction of simple switches or oscillators, the

OPERON
A genetic unit or cluster that consists of one or more genes that are transcribed as a unit and are expressed in a coordinated manner.

PARAMETER SPACE
The set of all possible values of all parameters.

Box 2 | A synthetic λ-switch

The *in numero* exploration of the λ-switch consisted of a model based on two plasmids. On one was a synthetic network comprising the P_{RM} promoter of λ-phage and the *cI* gene, which encodes the protein λ-repressor. To this, a second plasmid was added, which contained the P_R promoter that was used to control the expression of cro. Each plasmid contains the right operator regions O_R1, O_R2 and O_R3, and both proteins are capable of binding to these regions on either plasmid. On the P_{RM}-promoter plasmid, transcription of *cI* takes place whenever there is no protein (of either type) bound to O_R3; when cI is bound to O_R2, the rate of *cI* transcription is enhanced. On the P_R-promoter plasmid, *cro* is transcribed only when operator site O_R3 is either clear or has a cro dimer bound to it; either protein being bound to O_R1 or O_R2 has the effect of halting the transcription of *cro*.

The genetic network of λ-phage switches from the dormant lysogenic state to the lytic growth state in its host bacterium in roughly 20 minutes[56]. Under the conditions studied, the λ-switch model showed a significantly faster transition between its stable states than those seen in the model of the toggle switch. This indicates that the properties of the λ-switch might offer an advantage in terms of the speed of transitions, indicating that it might be fruitful to study synthetic models based on this natural system. The future construction of a synthetic λ-switch, based on modelling results such as these, might permit more precise statements regarding the source of this advantage.

design of genetic code, or software, capable of performing increasingly elaborate functions[43,88] can be imagined. In this section, we review the recent advances in the modelling and construction of synthetic networks.

A genetic toggle switch. Co-repressive switches have long been proposed as a common regulatory theme[89], and the synthetic toggle switch[90] serves as a model system in which to study such networks. This study involved a network in which each of two proteins negatively regulates the synthesis of the other. A specific example of one of the toggle designs is depicted in FIG. 3a. Intuitively, one might anticipate that there could be two possible steady states for this system. Because lacI production is repressed by the cI protein, an initial high concentration of cI would be self-sustaining and lead to a state with high cI and low lac repressor concentrations. Conversely, because cI production is repressed by the lac repressor, if the lac repressor is initially present in high concentrations, a second stable state would entail high lac and low cI concentrations.

One counter-intuitive finding to emerge from the toggle work was that not all co-repressive systems will show bistability. In fact, a central feature of this work was the use of mathematical and computational tools in deducing *a priori* the criteria for a robust toggle switch. The feasibility of a toggle switch is manifest in the existence of two stable fixed points; any initial state above the dividing line in FIG. 4a will evolve to the fixed point that is characterized by a high cI (low lac repressor) concentration, whereas initial states below the dividing line will evolve to a high lac repressor (low cI) concentration. The design of an operating toggle thus depended on parameter choices that led to bistability. These criteria included the use of strong and balanced constitutive promoters, effective transcriptional repression, the formation of protein multimers and similar protein degradation rates.

The reliable toggling between states was induced experimentally through the transient introduction of either a chemical or a thermal stimulus, and shown to be significantly sharper than for that of a network designed without co-repression. Specifically, isopropyl-β-D-thiogalactopyranoside (IPTG), which binds to lac repressor tetramers, was used to render the lac repressor unable to repress its promoter. Likewise, a temperature-sensitive cI protein was used, so that its degradation rate was an increasing function of temperature. For detection, the green fluorescent protein (GFP GENE) was transcribed POLYCISTRONICALLY with the *cI* gene, so that GFP concentrations were proportional to the concentration of the λ-repressor (encoded by *cI*). Results for one of the toggles are presented in FIG. 4b. The system, beginning in the high-lac repressor/low-cI state, was toggled to the high-cI/low-lac repressor state with IPTG. As expected, the system remained in the high-cI state after removal of the IPTG stimulus. Toggling to the low-cI state was then accomplished by tuning the temperature to 42 °C, and this state was subsequently stable upon the return of the system to 32 °C. These results show that synthetic toggle switches can be designed and used in a cellular environment.

Recently, a synthetic switch based on the switching mechanism of the λ-phage was compared with the toggle switch[91] (BOX 2). It was numerically shown that the λ-like system offers a faster switching time from the dormant lysogenic state to the lytic growth state than seen in the model of the toggle switch and thus might represent a more fruitful synthetic model.

The repressilator. In tandem with the toggle work, a synthetic network that can generate oscillations in the concentrations of cellular proteins was presented[92]. Although the ideas in this study were developed independently, the problem was approached in a fashion similar to the toggle: model equations were used to determine design criteria, and a functioning network was built and studied in the context of a mathematical model. The repressilator consisted of three repressible promoters designed with cyclic repressibility (FIG. 3b). Although this network can produce oscillatory behaviour, the construct, in itself, is not sufficient for oscillations. Specific properties, such as the protein synthesis and degradation rates and cooperativity of protein–DNA binding, must be properly chosen. Instead of determining these properties through the lengthy process of trial-and-error experimentation, model equations were used as an efficient alternative.

In determining the design criteria for the repressilator, equations similar to those used for the toggle switch were derived and analysed, and the results were summarized in a parameter-space plot (FIG. 5). Taken as a whole, FIG. 5 implies that oscillations are favoured by high synthesis and degradation rates, large cooperative binding effects and efficient repression. These conclusions led directly to specific design choices — strong and tightly repressible hybrid promoters were selected, and the effective protein degradation rates were increased by using SMALL STABLE RNA TAGS (SsrA tags).

In the experiment, a second plasmid was designed with a *tet*-repressible (tetracyclin) promoter that directed the transcription of a *GFP* reporter gene. So, GFP

GFP GENE
A gene encoding the green fluorescent protein (GFP). GFP can be transcribed in tandem with another gene of interest, so that one GFP molecule is produced for each molecule of the target protein. The fluorescence level in a cell then provides an indication of the concentration of the protein of interest.

POLYCISTRONIC
A form of gene organization that results in transcription of an mRNA that codes for multiple gene products, each of which is independently translated from the mRNA.

SMALL STABLE RNA TAG (SsrA)
A short peptide tag that is added to the carboxy-terminal end of the incomplete protein product of a stalled ribosome. This *trans*-translation process, which is catalysed by a small stable RNA (also known as tmRNA (tRNA-like and mRNA-like)), targets the abnormal proteins for proteolysis.

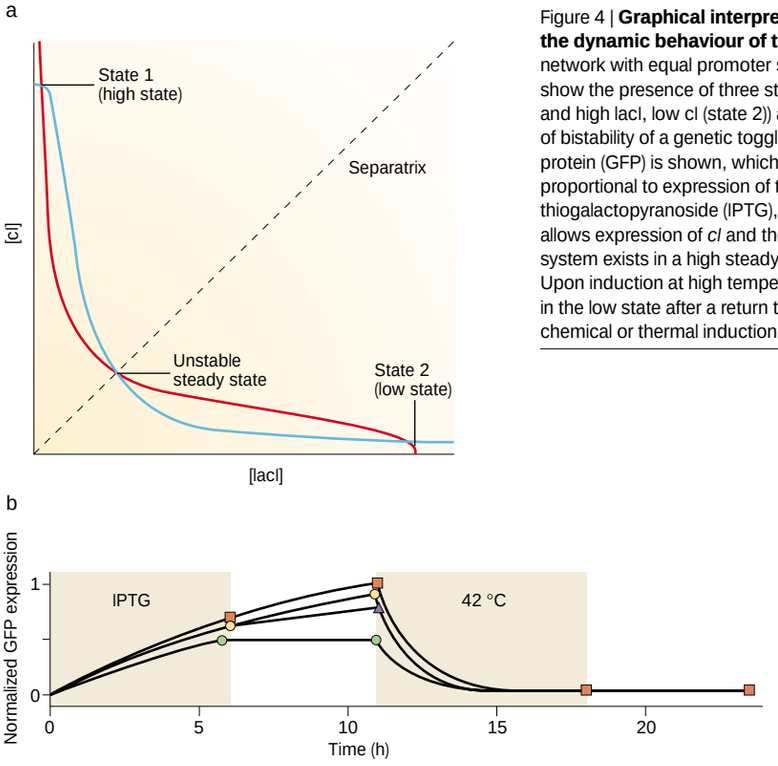

Figure 4 | **Graphical interpretation of the toggle switch equations and illustration of the dynamic behaviour of the co-repressive network. a** | Analysis of a bistable toggle network with equal promoter strengths driving the expression of lacI and cI proteins. The plots show the presence of three steady states: two stable steady states (low lacI, high cI (state 1) and high lacI, low cI (state 2)) and one unstable steady state. **b** | Experimental demonstration of bistability of a genetic toggle switch in *Escherichia coli*. The response of green fluorescent protein (GFP) is shown, which corresponds to expression of the *cI* gene and is inversely proportional to expression of the *lacI* gene. Upon induction with isopropyl-β-D-thiogalactopyranoside (IPTG), lac repressor protein is removed from its operator site and allows expression of *cI* and the reporter protein, GFP. As observed, after six hours, the system exists in a high steady state. The system remains in a high state after removal of IPTG. Upon induction at high temperature (42 °C), the system returns to the low state and remains in the low state after a return to low temperature. The shaded areas indicate periods of chemical or thermal induction. The symbols represent four toggle plasmids.

would be inversely proportional to the prevalence of tet, so that tet oscillations would emerge as GFP oscillations. The repressilator showed self-sustained oscillations over the entire growth phase of the host *E. coli* cells. Interestingly, the period of the oscillations was longer than the bacterial SEPTATION PERIOD, indicating that cellular conditions that are important to the oscillator network might be transmitted to the progeny cells.

Although oscillations were observed as anticipated, the repressilator results raised some additional questions regarding fluctuations. For example, only 40% of the cells contained oscillating GFP, and significant variations in the oscillatory phases and amplitudes were observed between daughter cells. Further modelling indicated that, to circumvent the effects of noise, naturally occurring oscillators might need some additional form of control. Indeed, an important aspect of this study was its focus on the use of synthetic networks as tools for biological inference. In this regard, the repressilator work provides potentially valuable information about the design principles of other oscillatory systems, such as circadian clocks.

Effect of regulation on stability. One of the earlier models explored the effect of AUTOREGULATION on equilibrium stability[7]. When a gene product is initially in an equilibrium state, small perturbations to the steady state will be accompanied by an exponential decay or rise back to equilibrium. The stability of the fixed point is given by the time constant in the exponential response; high stability corresponds to a fast return to equilibrium, whereas low stability correlates with a slow return. In this model, two side-by-side comparisons were made between single-gene networks. The first compared networks with varying degrees of repression, whereas the second compared networks with varying activation. With regard to stability, the central result was that stability is increased with repression and decreased with activation (FIG. 6a, b).

More recently, experimental work with simple synthetic networks has supported these stability predictions[93]. In this work, both a negatively controlled promoter and an unregulated promoter were used to study the effect of repression on variations in cellular protein concentration (FIG. 3c). To test the role of negative feedback, the system was first modelled with equations similar to those used in the earlier work. To compare directly with the experiment, a random term was added to the governing equations and protein concentration distributions were generated. In this framework, increased stability corresponds to tighter distributions (FIG. 6c, d).

The general experimental design used a *tet*-repressible promoter that directed the production of a *tet* repressor–EGFP (enhanced green fluorescent protein) fusion protein. The system was first studied with repression, then the negative-feedback loop was eliminated in two separate designs. The first involved the mutation of the tet protein and the second entailed operator replacement at the promoter. The results confirmed the prediction that repression decreases the cell-to-cell fluctuations in protein concentration measurements. This was shown empirically through the measurement of protein fluorescence distributions over a population of cells. The findings showed that, for a repressive network, the fluorescence distribution is significantly tightened. Taken together with the earlier modelling study, these results indicate that negative feedback might be used in cellular design as a means for mitigating variations in cellular protein concentrations. Because the number of proteins per cell is typically small, internal fluctuations are thought to be important and this study is relevant to issues regarding the reliability of cellular processes in the presence of internal noise.

Activation. The construction of an activating circuit seems to be the next logical step. Results from modelling have provided an insight into two important issues pertaining to a positive-feedback network[7,94,95]. First, it has been shown that activation should

SEPTATION PERIOD
The time that it takes a bacterium to divide.

AUTOREGULATION
The property of a system whereby a component of the system controls its own activity. See also the definitions for positive feedback and negative feedback.

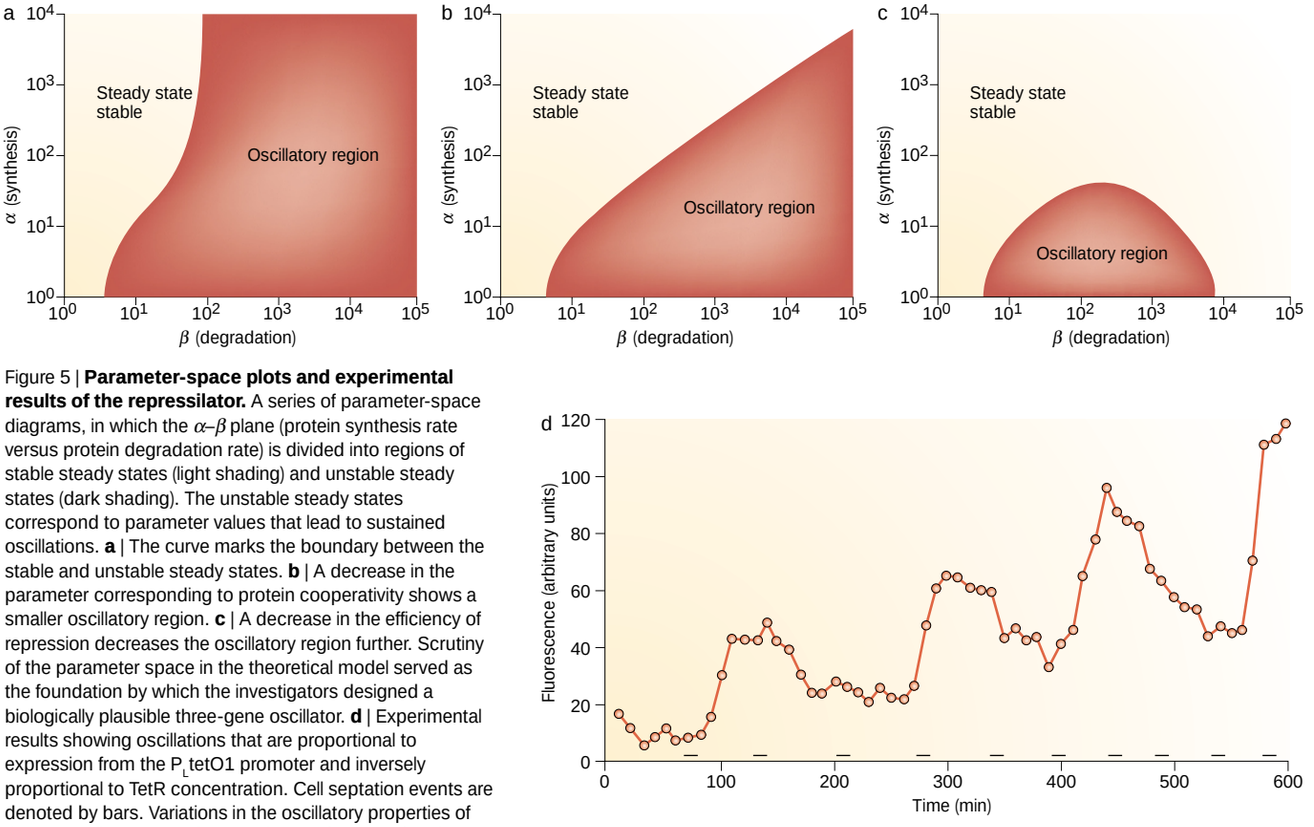

Figure 5 | **Parameter-space plots and experimental results of the repressilator.** A series of parameter-space diagrams, in which the α–β plane (protein synthesis rate versus protein degradation rate) is divided into regions of stable steady states (light shading) and unstable steady states (dark shading). The unstable steady states correspond to parameter values that lead to sustained oscillations. **a** | The curve marks the boundary between the stable and unstable steady states. **b** | A decrease in the parameter corresponding to protein cooperativity shows a smaller oscillatory region. **c** | A decrease in the efficiency of repression decreases the oscillatory region further. Scrutiny of the parameter space in the theoretical model served as the foundation by which the investigators designed a biologically plausible three-gene oscillator. **d** | Experimental results showing oscillations that are proportional to expression from the $P_L$tetO1 promoter and inversely proportional to TetR concentration. Cell septation events are denoted by bars. Variations in the oscillatory properties of the repressilator network were observed across daughter cells in the form of phase delays. K_M, the Michaelis constant (defined as the substrate concentration at half the maximum velocity of the reaction). (The dimensionless parameter α is defined as the ratio of the rates of mRNA synthesis and degradation, whereas dimensionless β is the ratio of the rates of protein degradation and mRNA degradation.) (Redrawn with permission from REF. 92 © (2000) Macmillan Magazines Ltd.)

decrease the stability of the equilibrium state (FIG. 6b), thus leading to variations that are characterized by wider distributions as compared with networks without feedback. Second, a single-gene network with positive regulation is capable of bistability. This implies that a single-gene switch can be constructed as an alternative to the co-repressive toggle discussed above. Additionally, such a switch can be used as the basic element for an oscillator[96].

As a concrete example of an activating system, consider a synthetic network that consists of the P_{RM} promoter of λ-phage and the *cI* gene, which encodes the protein λ-repressor (FIG. 2). The nonlinearity of the governing equation for this network leads to a bistable regime in the steady-state concentration of the repressor (FIG. 7a). The bistability arises as a consequence of the competition between the production and dimerization of the repressor and its degradation. For certain parameter values, the initial concentration is irrelevant, but for those that more closely balance production and loss, the final concentration is determined by the initial value. Because activation is the primary mode of regulation in eukaryotes[97], the experimental demonstration of bistability in a single-gene network should provide valuable information regarding gene regulation in higher organisms.

HYSTERESIS
As a parameter that represents some property of a system is increased, the behaviour makes a sudden jump at a particular value of the parameter. But, as the parameter is then decreased, the jump back to the original behaviour does not occur until a much lower value. In the region between the two jumps, the system is bistable.

Noise resistance and circadian clocks. The repressilator study has led to proposals regarding the genetic architecture that underlies circadian rhythms[96]. As discussed above, an interesting property of the repressilator oscillations was the existence of significant cell-to-cell variations in amplitude and period, apparently arising from small fluctuations in molecule number[92,96]. With regard to circadian clocks, it is thought that the ability to maintain constant periodicity is paramount[98], perhaps implying that these systems generate oscillations through mechanisms that are qualitatively different from those of the repressilator[96]. In other words, an important design principle for circadian networks might be that the resulting oscillatory period is resilient to noise.

To circumvent variability, the use of HYSTERESIS-based oscillations was recently proposed[96]. In this work, it was shown how a model circadian network can oscillate reliably in the presence of internal noise. Although the underlying genetic architecture for any of the various known circadian systems has not been deduced, it was pointed out that these networks seem to involve both positive and negative control elements[99]. This information was used to construct a generic model capable of oscillations that are resistant to fluctuations.

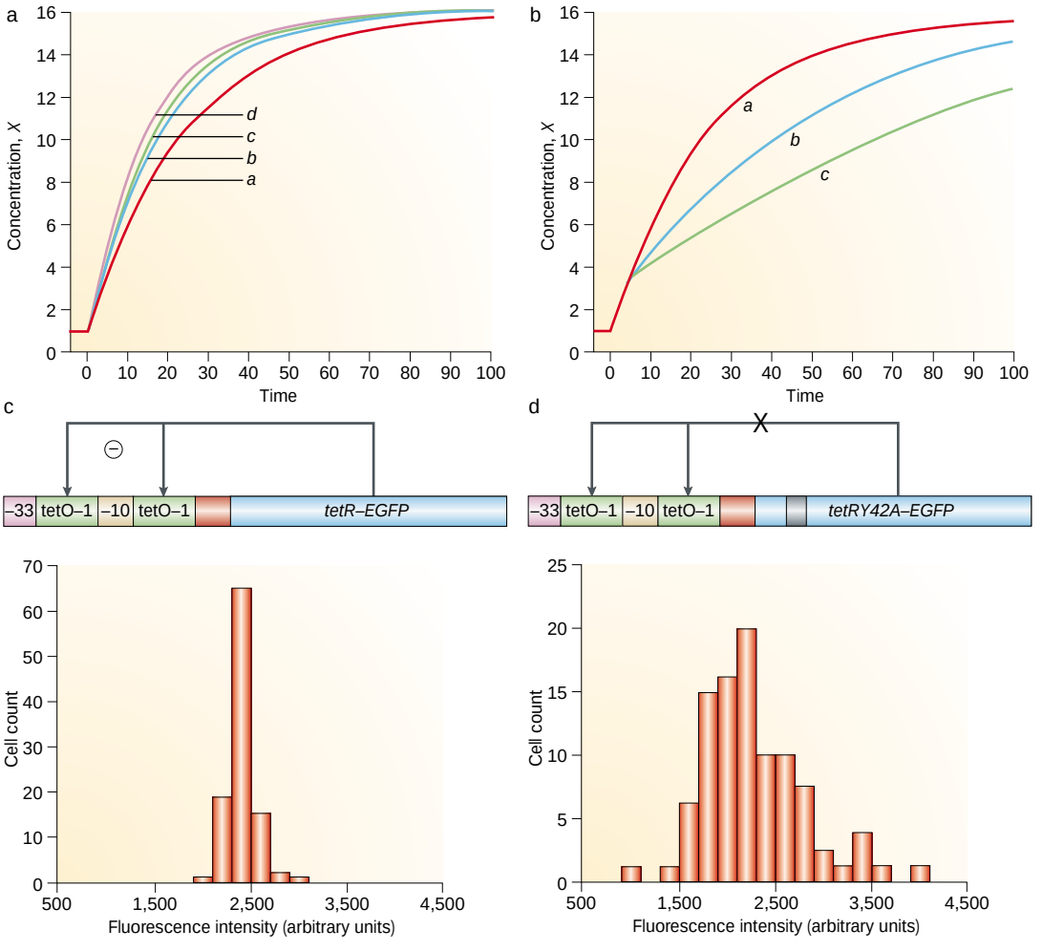

Figure 6 | **Theoretical predictions and experimental results of an autorepressive gene network. a** | The response of autorepressive gene systems predicts decreases in repressor protein fluctuations with increases in the degree of autorepression (curves representing increases in the degree of repression: $a < b < c < d$). The system with the lowest degree of repression (*a*) has the slowest temporal response to perturbation; whereas the system with the highest degree of repression (*d*) has the quickest temporal response to fluctuations. **b** | The response of autoactivated gene systems is opposite to that of autorepressive systems: increases in the degree of autoactivation (curves representing increases in the degree of activation: $a < b < c$) correspond with increases in protein fluctuations. The system with the highest degree of activation (*c*) has the slowest temporal response to perturbation and the system with the lowest degree of activation (*a*) has the quickest temporal response to perturbations. (Panels **a** and **b** redrawn from REF. 7 © (1974) Macmillan Magazines Ltd.) **c, d** | Experimental confirmation of the theoretical predictions. **c** | The autoregulated negative-feedback network in which a tetR–EGFP (enhanced green fluorescent protein) fusion protein binds to tetO operator sites in the promoter that drives its production (top panel). Green fluorescent protein (GFP) measurements show a tightly distributed population (bottom panel). **d** | An unregulated network is constructed by mutating the DNA-binding domain of the tetR protein (grey box, top panel), thereby preventing binding at the tetO operator sites. In the unregulated case, GFP measurements depict a wider distribution of fluorescence (bottom panels). Each circuit is composed of an operator at position V (–33 hexamer) and an operator at position IV (–10 hexamer); untranslated region (red) and fusion protein (blue). (Panels **c** and **d** redrawn with permission from REF. 93 © (2000) Macmillan Magazines Ltd.)

The circuitry of their proposed oscillator can be understood by considering the following hypothetical synthetic network. The hysteretic effect in FIG. 7a can be used to induce oscillations, provided that the network can be coupled to a slow subsystem that effectively drives the degradation parameter. This can be done by inserting a protease under the control of a separate promoter region. The activating network is on one plasmid: here, the P_{RM} promoter controls the expression of repressor protein cI, which stimulates its own production at low concentrations and represses the promoter at high concentrations. The P_{RM} promoter region is also on a second plasmid, but here it drives the gene that encodes the protein RcsA. The crucial interaction is between RcsA and cI; RcsA is a protease for repressor, effectively inactivating the ability of cI to control the P_{RM} promoter region[100] (FIG. 7b).

The central point of REF. 96 was that hysteresis-based oscillations are comparatively more resistant to noise. This was shown through a side-by-side comparison of numerical results obtained from several models. This conclusion, although intriguing, must nevertheless be evaluated with caution because the various alternative models[101–104] have many parameters and an extensive

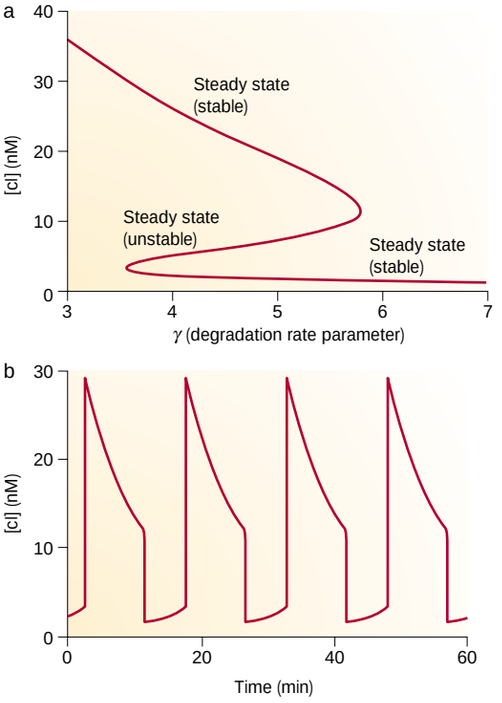

Figure 7 | **A synthetic system based on the λ-switch.** **a** | Bifurcation plot for the steady-state concentration of repressor versus the model parameter γ, representing the degradation rate. Note that, for values of γ between 3.6 and 5.8, there are three possible steady-state concentrations (arrows). The top and bottom branch values are stable (concentrations near these values will remain nearby despite small fluctuations), whereas the middle branch value is unstable (any tiny fluctuation will be amplified, driving the protein concentration towards one of the two stable states on the upper and lower branches). **b** | Results from a relaxation oscillator network in which two identical promoters each drive the expression of different genes. Promoter P_{RM} drives the expression of λ-repressor protein, which activates its own production as well as the production of a second protein, RcsA (regulator of capsule synthesis A), driven by a second copy of the P_{RM} promoter. Protein RcsA cleaves the λ-repressor protein, which prevents λ-repressor from activating the P_{RM} promoters that transcribe *cI* and *rcsA*. Oscillations arise as the RcsA-induced degradation of repressor causes a traversal of the hysteresis diagram. Suppose one begins with a parameter value of $\gamma = 4$ and on the upper branch of panel **a**. The large value of repressor will then serve to activate the promoter for the RcsA. An increase in RcsA production acts as an additive degradation term for repressor and thus effectively induces slow motion to the right on the upper branch of panel **a**. This motion will continue until the repressor concentration falls off the upper branch at $\gamma \approx 5.8$. At this point, with the repressor concentration at a very low value, the promoters are essentially turned off. Then, as RcsA begins to degrade, the repressor concentration slowly moves to the left along the lower branch of panel **a**, until it encounters the bifurcation point at $\gamma \approx 3.6$. It then jumps to its original high value, with the entire process repeating and producing the oscillations.

side-by-side investigation has not been pursued. In the future, the construction of a synthetic hysteresis-based oscillator, such as the one described here, might provide additional information about the generation of circadian rhythms.

Concluding remarks

The modelling of gene regulatory networks relies on characterization of the behaviour of small subsystems, formation of hypotheses about how these subsystems interconnect, translation of these hypotheses into a mathematical model and experimentation to yield results that indicate necessary changes to the original hypotheses. Of course, the same general procedure might be carried out without reduction of the hypotheses to mathematical form and much of what we now know about gene regulation has been garnered in this fashion. However, working with equations has the advantage of making it clear what assumptions have been made and where contradictions arise when comparisons are made with experiment. Furthermore, the complexity of these systems is such that it is nearly impossible to predict all of the consequences of a given hypothesis simply by abstract reasoning.

The synthetic network studies represent important advances in the engineering-based methodology of network design. In the studies presented so far, the experimental behaviour is consistent with predictions that arise from *in numero* modelling. Furthermore, theoretical models were used to determine design criteria, which lent support to the idea of an engineering-based approach to genetic network design. These criteria included the use of strong constitutive promoters, effective transcriptional repression, cooperative protein interactions and similar protein degradation rates.

Although the experimental techniques used in studies of this nature are certainly impressive, it is clear that reliable theoretical tools would be of enormous value. On a strictly practical level, such techniques could potentially reduce the degree of trial-and-error experimentation. More importantly, computational and theoretical approaches will lead to testable predictions regarding the current understanding of complex biological networks.

Although it has been nearly 30 years since the pioneering theoretical work on interacting genetic networks[6–10], the true significance of these studies had to await technological advances. Current progress in the study of both naturally occurring and synthetic genetic networks indicates that computational modelling should have an important role in the description and manipulation of the dynamics that underlie cellular control.

Links

DATABASE LINKS p53 | RecA | SsrA | RcsA

FURTHER INFORMATION Modelling work on eukaryotic cell cycle control | Adam Arkin's home page | *Chaos* special Focus Issue "Molecular, Metabolic, and Genetic Control" in March | Jeff Hasty's home page | Farren Isaacs' lab | James Collins' home page | James Collins' lab

ENCYCLOPEDIA OF LIFE SCIENCES Genetic networks | Bacteriophage λ and its relatives

1. Novak, B. & Tyson, J. J. Modeling the control of DNA replication in fission yeast. *Proc. Natl Acad. Sci. USA* **94**, 9147–9152 (1997).
2. McAdams, H. H. & Shapiro, L. Circuit simulation of genetic networks. *Science* **269**, 650–656 (1995).
3. McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl Acad. Sci. USA* **94**, 814–819 (1997).
4. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47–C51 (1999).
5. Lauffenburger, D. A. Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl Acad. Sci. USA* **97**, 5031–5033 (2000).
6. Glass, L. & Kauffman, S. A. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* **39**, 103–129 (1973).
7. Savageau, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **252**, 546–549 (1974).
   References 6 and 7 are pioneering papers in the modelling of gene regulation. They correctly anticipated the complex nonlinearity of gene regulation and the potential for computational approaches.
8. Kauffman, S. A. The large-scale structure and dynamics of gene control circuits: an ensemble approach. *J. Theor. Biol.* **44**, 167–174 (1974).
9. Glass, L. Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.* **54**, 85–107 (1975).
10. Glass, L. Combinatorial and topological methods in nonlinear chemical kinetics. *J. Chem. Phys.* **63**, 1325 (1975).
11. Savageau, M. A. *Biochemical System Analysis* (Addison Wesley, Reading, 1976).
12. Goodwin, B. C. *Analytical Physiology of Cells and Developing Organisms* (Academic, London, 1976).
13. Tyson, J. J. & Othmer, H. G. The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.* **5**, 1 (1978).
14. Palsson, B. O. & Lightfoot, E. N. Mathematical modelling of dynamics and control in metabolic networks. V. Static bifurcations in single biochemical control loops. *J. Theor. Biol.* **113**, 279–298 (1985).
15. Moran, F. & Goldbeter, A. Onset of birhythmicity in a regulated biochemical system. *Biophys. Chem.* **20**, 149–156 (1984).
16. Ackers, G. K., Johnson, A. D. & Shea, M. A. Quantitative model for gene regulation by λ phage repressor. *Proc. Natl Acad. Sci. USA* **79**, 1129–1133 (1982).
17. Shea, M. A. & Ackers, G. K. The O_R control system of bacteriophage λ: a physical–chemical model for gene regulation. *J. Mol. Biol.* **181**, 211–230 (1985).
    Development of a modelling technique in which thermodynamic quantities derived from experimental data are used to formulate a quantitative model. When the technique was applied to the λ-switch, it strongly supported the view that cooperative interactions stabilize the lysogenic state by tightening the repression of cro at lysogenic concentrations of λ-repressor.
18. Reinitz, J. & Vaisnys, J. R. Theoretical and experimental analysis of the phage λ genetic switch implies missing levels of co-operativity. *J. Theor. Biol.* **145**, 295–318 (1990).
    One of the earliest papers to apply ideas from the burgeoning field of nonlinear dynamics to a molecular biology experiment. It describes the mathematical framework for the study of the λ-switch with an intriguing (and, so far, untested) conjecture about missing levels of regulation.
19. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
20. Agarwal, M. L., Taylor, W. R., Chernov, M. V., Chernova, O. B. & Stark, G. R. The p53 network. *J. Biol. Chem.* **273**, 1–4 (1998).
21. McAdams, H. H. & Arkin, A. Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199–224 (1998).
22. Szafranski, P. *et al.* A new approach for containment of microorganisms: dual control of streptavidin expression by antisense RNA and the T7 transcription system. *Proc. Natl Acad. Sci. USA* **94**, 1059–1063 (1997).
23. Harding, T. C., Geddes, B. J., Murphy, D., Knight, D. & Uney, J. B. Switching transgene expression in the brain using an adenoviral tetracycline-regulatable system. *Nature Biotechnol.* **16**, 553–555 (1998).
24. Smolen, P., Baxter, D. A. & Byrne, J. H. Mathematical modeling of gene networks. *Neuron* **26**, 567 (2000).
    Review of the use of general abstract models to obtain insights into the behaviour of entire classes of biological systems.
25. Keller, A. D. Model genetic circuits encoding autoregulatory transcription factors. *J. Theor. Biol.* **172**, 169–185 (1995).
26. Thomas, R. & Thieffry, D. Feedback loops, the wheels of biological regulatory networks. *Med. Sci.* **11**, 189–197 (1995).
27. Thomas, R., Thieffry, D. & Kaufman, M. Dynamical behaviour of biological regulatory networks. I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.* **57**, 247–276 (1995).
28. Wolf, D. M. & Eeckman, F. H. On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J. Theor. Biol.* **195**, 167–186 (1998).
29. Cherry, J. L. & Adler, F. R. How to make a biological switch. *J. Theor. Biol.* **203**, 117–133 (2000).
30. Bialek, W. Stability and noise in biochemical switches. *Adv. Neural Info. Process.* **13** (MIT Press, Cambridge, Massachusetts) (in the press).
31. Smith, H. Monotone semiflows generated by functional differential equations. *J. Differ. Equat.* **66**, 420–442 (1987).
32. Mestl, T., Lemay, C. & Glass, L. Chaos in high-dimensional neural and gene networks. *Physica* D **98**, 33–52 (1996).
33. Smolen, P., Baxter, D. & Byrne, J. Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems. *Am. J. Physiol.* **277** (*Cell Physiol.* **46**), C777–C790 (1999).
34. An der Heiden, U. Delays in physiological systems. *J. Math. Biol.* **8**, 345–364 (1979).
35. Falke, J., Bass, R., Butler, S., Chervitz, S. & Danielson, M. The two-component signaling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu. Rev. Cell Dev. Biol.* **13**, 457–512 (1997).
36. Bray, D., Bourret, R. & Simon, M. Computer-simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **4**, 469–482 (1993).
37. Spiro, P., Parkinson, J. & Othmer, H. A model of excitation and adaptation in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **94**, 7263–7268 (1996).
38. Eisenbach, M. Control of bacterial chemotaxis. *Mol. Microbiol.* **4**, 161–167 (1996).
39. Abouhamad, W. *et al.* Computer-aided resolution of an experimental paradox in bacterial chemotaxis. *J. Bacteriol.* **180**, 3757–3764 (1998).
40. Levin, M., Morton-Firth, C., Abouhamad, W., Bourret, R. & Bray, D. Origins of individual swimming behavior in bacteria. *Biophys. J.* **74**, 175–181 (1998).
41. Yi, M. *et al.* Perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl Acad. Sci. USA* **97**, 4649–4653 (1999).
42. Bray, D. & Lay, S. Computer-simulated evolution of a network of cell-signaling molecules. *Biophys. J.* **66**, 972–977 (1994).
43. Bray, D. Protein molecules as computational elements in living cells. *Nature* **376**, 307–312 (1995).
    A clear exposition of the idea that gene networks might be viewed as fulfilling computational operations, if input–output relationships are established between protein concentrations inside a cell. Includes biologically relevant examples.
44. Fisher, M., Paton, R. & Matsuno, K. Intracellular signalling proteins as 'smart' agents in parallel distributed processes. *Biosystems* **50**, 159–171 (1999).
45. Fisher, M., Malcolm, G. & Paton, R. Spatio-logical processes in intracellular signalling. *Biosystems* **55**, 83–92 (2000).
46. Dutta, A. & Bell, S. Initiation of DNA replication in eukaryotic cells. *Annu. Rev. Cell Dev. Biol.* **13**, 293–332 (1997).
47. Mahaffy, J. & Zyskind, J. A model for the initiation of replication in *Escherichia coli*. *J. Theor. Biol.* **140**, 453–477 (1989).
48. Novak, B. & Tyson, J. J. Quantitative analysis of a molecular model of mitotic control in fission yeast. *J. Theor. Biol.* **173**, 283–286 (1995).
49. Sveiczer, A., Csikasz–Nagy, A., Gyorffy, B., Tyson, J. J. & Novak, B. Modeling the fission yeast cell cycle: quantized cycle times in *wee1⁻cdc25Δ* mutant cells. *Proc. Natl Acad. Sci. USA* **97**, 7865–7870 (2000).
50. Chen, K. C., Csikasz-Nagy, A., Gyorffy, B., Novak, B. & Tyson, J. J. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* **11**, 369–391 (2000).
51. Ptashne, M. *et al.* How the λ repressor and cro work. *Cell* **19**, 1–11 (1980).
52. Meyer, B. J., Maurer, R. & Ptashne, M. Gene regulation at the right operator (O_R) of bacteriophage λ. II. O_R1, O_R2, and O_R3: their roles in mediating the effects of repressor and cro. *J. Mol. Biol.* **139**, 163–194 (1980).
53. Johnson, A. D. *et al.* λ repressor and cro — components of an efficient molecular switch. *Nature* **294**, 217–223 (1981).
54. Johnson, A. D., Pabo, C. O. & Sauer, R. T. Bacteriophage λ repressor and cro protein: interactions with operator DNA. *Meth. Enzymol.* **65**, 839–856 (1980).
55. Ohlendorf, D. H. & Matthews, B. W. Structural studies of protein–nucleic acid interactions. *Annu. Rev. Biophys. Bioeng.* **12**, 259–284 (1983).
56. Ptashne, M. *A Genetic Switch: Phage λ and Higher Organisms* (Cell Press, Cambridge, Massachusetts, 1992).
    Excellent reference for anyone interested in beginning an exploration of gene regulation. Especially good for engineers, physicists or applied mathematicians who would like to understand the primary issues.
57. Johnson, A., Meyer, B. & Ptashne, M. Interactions between DNA-bound repressors govern regulation by the λ phage repressor. *Proc. Natl Acad. Sci. USA* **76**, 5061–5065 (1979).
58. 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).
    Pioneering work on the role of fluctuations in gene regulation. The central conclusion is that fluctuations cannot always be viewed as simply small perturbations as they can, in fact, induce different developmental pathways.
59. Kourilsky, P. Lysogenization by bacteriophage λ. I. Multiple infection and the lysogenic response. *Mol. Gen. Genet.* **173**, 183–195 (1973).
60. Endy, D., You, L., Yin, J. & Molineux, I. J. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl Acad. Sci. USA* **97**, 5375–5380 (2000).
61. Endy, D., Kong, D. & Yin, J. Intracellular kinetics of a growing virus: a genetically structured simulation for bacteriophage T7. *Biotech. Bioeng.* **55**, 37–44 (2000).
62. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
63. Jacob, F., Perrin, D., Sanchez, C. & Monod, J. L'operon: groupe de gene a expression par un operatour. *C.R. Seances Acad. Sci.* **250**, 1727–1729 (1960).
64. Monod, J., Wyman, J. & Changeux, J. P. On the nature of allosteric transitioins: a plausible model. *J. Mol. Biol.* **12**, 88–118 (1965).
65. Goodwin, B. Oscillatory behaviour in enzymatic control process. *Adv. Enzyme Regul.* **3**, 425–438 (1965).
66. Griffith, J. S. Mathematics of cellular control processes. I. Negative feedback to one gene. *J. Theor. Biol.* **2**, 202–208 (1968).
67. Dickson, R., Abelson, J., Barnes, W. & Reznikoff, W. S. Genetic regulation: the Lac control region. *Science* **187**, 27–35 (1975).
68. Betenbaugh, M. J. & Dhurjati, P. A comparison of mathematical model predictions to experimental measurements for growth and recombinant protein production in induced cultures of *Escherichia coli*. *Biotech. Bioeng.* **36**, 124–134 (1990).
69. Bliss, R. D., Painter, R. P. & Marr, A. G. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* **97**, 177–193 (1982).
70. Koh, B. T. & Yap, M. G. S. A simple genetically structured model of *trp* repressor–operator interactions. *Biotechnol. Bioeng.* **41**, 707–714 (1993).
71. Qaddour, J. S. Mathematical model of negative control of *E. coli* bacteria. *Proc. 36th Conf. Decision Control* 838–839 (IEEE, San Diego,1997).
72. Santillan, M. & Mackey, M. C. Dynamic regulation of the tryptophan operon: a modeling study and comparison with experimental data. *Proc. Natl Acad. Sci. USA* **98**, 1364–1369 (2001).
73. Santillan, M. & Mackey, M. C. Dynamic behaviour in mathematical models of the tryptophan operon. *Chaos* **11**, 261–268 (2001).
74. Sen, A. K. & Liu, W. Dynamic analysis of genetic control and regulation of amino acid synthesis: the tryptophan operon in *Escherichia coli*. *Biotechnol. Bioeng.* **35**, 185–194 (1989).
75. Sinha, S. Theoretical study of tryptophan operon: application in microbial technology. *Biotechnol. Bioeng.* **31**, 117–124 (1988).
76. Wong, P., Gladney, S. & Keasling, J. D. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Prog.* **13**, 132–143 (1997).
77. Xiu, Z. L., Zeng, A. P. & Deckwer, W. D. Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. *J. Biotechnol.* **58**,

125–140 (1997).
78. von Dassow, G., Meir, E., Munro, E. M. & Odell, G. The segment polarity network is a robust developmental module. *Nature* **406**, 188–192 (2000).
79. Reinitz, J. & Sharp, D. Mechanism of eve stripe formation. *Mech. Dev.* **49**, 133–158 (1995).
References 78 and 79 illustrate the use of model construction in the determination of underlying gene network connectivity in a eukaryotic network.
80. Selinger, D. W. *et al.* RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. *Nature Biotechnol.* **18**, 1262–1268 (2000).
81. Brown, P. O. & Botstein, D. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* **21**, S33–S37 (1999).
82. Southern, E., Mir, K. & Shchepinov, M. Molecular interactions on microarrays. *Nature Genet.* **21**, S5–S9 (1999).
83. Lipshutz, R. J., Fodor, S. P., Gingeras, T. R. & Lockhart, D. J. High density synthetic oligonucleotide arrays. *Nature Genet.* **21**, S20–S24 (1999).
84. Cole, K. A., Krizman, D. B. & Emmert-Buck, M. R. The genetics of cancer — a 3D model. *Nature Genet.* **21**, S38–S41 (1999).
85. Cheung, V. G. *et al.* Making and reading microarrays. *Nature Genet.* **21**, S15–S19 (1999).
86. Duggan, D. J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. M. Expression profiling using cDNA microarrays. *Nature Genet.* **21**, S10–S14 (1999).
87. Chen, W., Kallio, P. & Bailey, J. E. Construction and characterization of a novel cross-regulation system for regulating cloned gene expression in *Escherichia coli*. *Gene* **130**, 15 (1993).
88. Weiss, R. & Knight, T. F. *Engineered Communications for Microbial Robotics. DNA6: Sixth International Meeting on DNA-Based Computers* (Leiden, The Netherlands, 2000).
89. Monod, J. & Jacob, F. General conclusions: teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harbor Symp. Quant. Biol.* **26**, 389 (1961).
90. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
91. Hasty, J., Isaacs, F., Dolnik, M., McMillen, D. & Collins, J. J. Designer gene networks: towards fundamental cellular control. *Chaos* **11**, 207–220 (2001).
92. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
93. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
References 90, 92 and 93 discuss important advances in the engineering-based methodology of synthetic network design. In all three, the experimental behaviour is consistent with predictions that arise from continuum dynamical modelling. Furthermore, theoretical models lend support to the idea of an engineering-based approach to genetic network design.
94. Hasty, J., Pradines, J., Dolnik, M. & Collins, J. J. Noise-based switches and amplifiers for gene expression. *Proc. Natl Acad. Sci. USA* **97**, 2075–2080 (2000).
95. Smolen, P., Baxter, D. A. & Byrne, J. H. Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. *Am. J. Physiol.* **43**, C531–C542 (1998).
96. Barkai, N. & Leibler, S. Biological rhythms: circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
97. Struhl, K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**, 1–4 (1998).
98. Glossop, N. R., Lyons, L. C. & Hardin, P. E. Interlocked feedback loops within the *Drosophila* circadian oscillator. *Science* **286**, 766–768 (1999).
99. Dunlap, J. Molecular bases for circadian clocks. *Cell* **96**, 271–290 (1999).
100. Rozanov, D. V., D'Ari, R. & Sineoky, S. P. RecA-independent pathways of lambdoid prophage induction in *Escherichia coli*. *J. Bacteriol.* **180**, 6306–6315 (1998).
101. Leloup, J. C. & Goldbeter, A. A model for circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J. Biol. Rhythms* **13**, 70–87 (1998).
102. Tyson, J. J., Hong, C. I., Thron, C. D. & Novak, B. A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM. *Biophys. J.* **77**, 2411–2417 (1999).
103. Leloup, J. C. & Goldbeter, A. Modeling the molecular regulatory mechanism of circadian rhythms in *Drosophila*. *BioEssays* **22**, 84–93 (2000).
104. Roussel, M. R., Gonze, D. & Goldbeter, A. Modeling the differential fitness of cyanobacterial strains whose circadian oscillators have different free-running periods: comparing the mutual inhibition and substrate depletion hypotheses. *J. Theor. Biol.* **205**, 321–340 (2000).
105. Gillespie, D. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340 (1977).

Acknowledgements
This work is supported by The Fetzer Institute (J.H.) and the Office of Naval Research.