Simplified Models of Biological Networks

Kim Sneppen, Sandeep Krishna, and Szabolcs Semsey

Niels Bohr Institute, DK-2100, Copenhagen, Denmark; email: sneppen@nbi.dk

Annu. Rev. Biophys. 2010. 39:43–59

First published online as a Review in Advance on January 4, 2010

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi: 10.1146/annurev.biophys.093008.131241

1936-122X/10/0609-0043$20.00

Key Words

feedback loops, bistability, epigenetics, homeostasis, oscillations

Abstract

The function of living cells is controlled by complex regulatory networks that are built of a wide diversity of interacting molecular components. The sheer size and intricacy of molecular networks of even the simplest organisms are obstacles toward understanding network functionality. This review discusses the achievements and promise of a bottom-up approach that uses well-characterized subnetworks as model systems for understanding larger networks. It highlights the interplay between the structure, logic, and function of various types of small regulatory circuits. The bottom-up approach advocates understanding regulatory networks as a collection of entangled motifs. We therefore emphasize the potential of negative and positive feedback, as well as their combinations, to generate robust homeostasis, epigenetics, and oscillations.

Contents

INTRODUCTION

The interior of the cell is crowded. It contains proteins, nucleic acids, and small metabolites and is surrounded by ions and water. Although they have widely disparate properties, most molecules interact specifically with only a few others. These specific interactions define a network of molecular species connected by links that represent interactions and flow of information. The specific interactions may differ in the direction of information flow and in the timescales on which they occur. For example, enzymatic conversion of small metabolites is usually fast and reversible, whereas synthesis of biopolymers is a slower and unidirectional process (25).

Combining such interacting components results in regulatory networks (RNs) that exhibit a rich and varied range of dynamic behavior, even for very small networks, especially when feedback is involved (2, 7, 8, 10, 16, 19–23, 29, 33, 55–57, 59). **Figure 1** shows three examples of RNs. They were chosen to span a range of organisms (phages, prokaryotes, and eukaryotes), molecular mechanisms (transcription and translation regulation, protein-protein interaction, enzymatic reactions), feedback logics (negative, positive, and combinations), and physiological functions (developmental decisions, regulation of metabolism, stress response). We have much to learn before we understand the connection between the structure (i.e., the topology of the network of interactions) and the function (i.e., the dynamical response of the system to external signals and conditions) of such physiologically important RNs.

In this review we present one approach to understanding the structure-function relationship, namely, using simplified models of small subnetworks of RNs. The sign of a feedback loop (FL), and how FLs of different signs are combined, is the primary determinant of the functionality of any RN. Below, we describe what is known about the functional behavior of single FLs and then explore models of two entangled FLs, emphasizing aspects that are more than just a sum of behaviors of single loops. In both cases, we connect the discussion to the cellular contexts in which these structures appear. Before examining feedback behavior, however, we first explore a few different mechanisms by which concentration, stability, and activity of regulatory factors can be controlled in RNs, which also set up the framework for the quantitative models we describe and make explicit the kind of simplifying assumptions used therein.

Regulatory network (RN): a representation of a regulatory system that consists of a collection of nodes, pairs of which are connected by links

Feedback loop (FL): a cyclic chain of links in a regulatory network

TF: transcription factor

ELEMENTARY DYNAMICS OF PROTEIN REGULATION

Binding of TFs to Operator Sites

Regulation of transcription is a dynamic phenomenon involving repeated association and dissociation of transcription factors (TFs) to

operator sites on the DNA. At present, our knowledge of the timescale for a TF to localize the relevant operator sites in vivo is highly uncertain, although some experiments indicate search times on the order of minutes for the lactose repressor (LacI) in *Escherichia coli* (15). An estimate of the association rate, k_{on}, of a TF to a specific operator can be obtained from a calculation of the time required for a single protein molecule, diffusing randomly with a diffusion constant D, to find a target of diameter a in a volume V (53):

$$k_{on} = 4\pi Da/V. \qquad 1.$$

This is an upper bound on the on-rate in particularly because TFs often bind to nonspecific DNA and thereby may lose substantial time searching far distances from specific operators (53). However, Equation 1 captures the general fact that the association rate is higher if the TF diffuses faster, if the target is bigger, or if the search volume is smaller. Assuming D for the TF to be approximately 5 $\mu m^2\ s^{-1}$, as measured for GFP in *E. coli* (17), and assuming a to be the diameter of a typical protein, $\approx$5 nm, we obtain $k_{on} \approx 2\ s^{-1}$ for one freely diffusing TF in a volume $V \approx 1\ \mu m^3$.

The corresponding off-rate for the TF leaving the operator is given by (53)

$$k_{off} = 4\pi Da \cdot (6 \cdot 10^8\ \mu m^{-3}) \cdot (K/1\,M), \qquad 2.$$

where $K = \exp(\Delta G/k_B T) \cdot 1$ M is the equilibrium binding constant in molar units, ΔG is the binding energy, and $k_B T \approx 0.62$ kcal mol^{-1} at room temperature. Thus, even for a strong binding site with, say, $K = 1$ nM, the off-rate

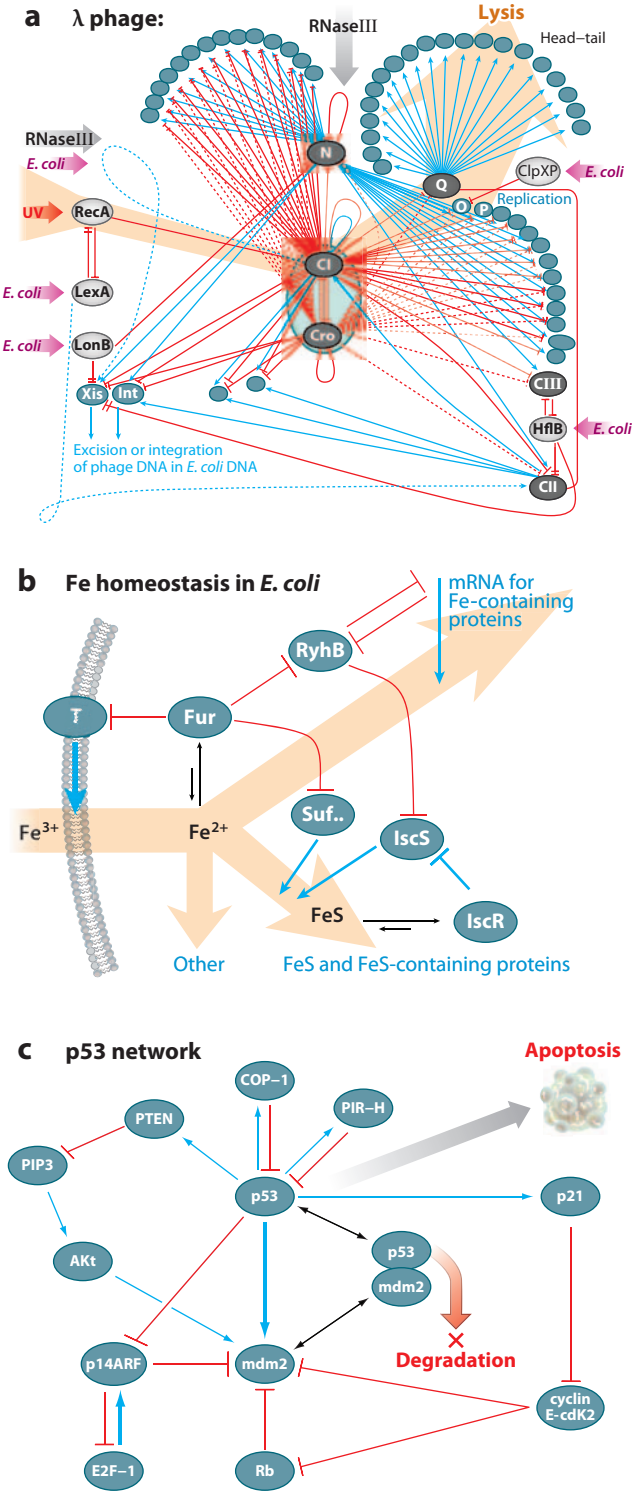

Figure 1

Molecular networks of (*a*) phage λ (58), (*b*) iron homeostasis in *Escherichia coli* (49), and (*c*) the p53-dependent apoptosis decision network in mammals (24). The three networks emphasize transcription regulation, regulation of metabolic fluxes, and posttranslational regulation, respectively. In panel *a* the orange arrow indicates the flow of information across the network, leading from UV induction to lysis. In panel *b*, the orange arrows indicate the flow of iron across the network.

Dissociation constant/equilibrium binding constant: for a complex, this is the ratio of the dissociation rate constant to the association rate constant, and is a measure of the binding strength of the complex

is $\approx 0.05\ s^{-1}$. The times involved in TF binding to operator sites (less than a minute) are often faster than timescales of protein production (transcription and translation, typically a minute or more); therefore, most models assume that association-dissociation occurs so fast that TF-operator complexes are considered to be in quasi-equilibrium. That is, one assumes that the fraction of time the TF spends bound to the operator adjusts instantaneously to changes in concentrations of the involved proteins and DNA.

Under this approximation, the bound fraction is given by the following function of the regulator (i.e., the TF) concentration, R, and the dissociation constant, $K = \exp(\Delta G/k_B T) \cdot 1\ M$, which is related to the binding energy of the TF-operator complex, ΔG (stronger binding implies lower K, and vice versa):

$$\text{bound fraction} = \frac{R/K}{1 + R/K}. \qquad 3.$$

This is similar to the Michaelis-Menten description of enzyme kinetics, with R analogous to substrate concentration, and K analogous to the Michaelis constant describing substrate-enzyme binding. Just as the Michaelis-Menten formulation can be extended to incorporate cooperative effects, so can the above formula be generalized by adding a Hill coefficient:

$$\text{bound fraction} = \frac{(R/K)^h}{1 + (R/K)^h}. \qquad 4.$$

Oligomerization of the TF prior to binding, or cooperative binding of multiple TFs, is represented by Hill coefficients $h > 1$ (dimerization: $h = 2$; tetramerization: $h = 4$). The above equation can be easily extended to account for different TFs competing for multiple operator sites (53). The formalism could also be generalized to the cases in which the numbers of the TF are small by accounting explicitly for all possible combinations of free molecules as described in Reference 53. Thus, small numbers do not exclude the equilibrium approach to gene regulation.

Transcription Regulation

A common type of link in a RN represents a TF activating or repressing a promoter. If the TF activates the promoter, its activity is proportional to the bound fraction from Equation 4:

$$\text{Activity} \sim \frac{(R/K)^h}{1 + (R/K)^h}, \qquad 5.$$

whereas if the TF represses the promoter,

$$\text{Activity} \sim \frac{1}{1 + (R/K)^h}. \qquad 6.$$

A large Hill coefficient makes activation more sensitive to variations in R when R is approximately equal to K, approaching a step function when h becomes very large. Expressing the promoter activities in this manner in fact remains valid even when there is substantial binding of the TF to nonspecific sites, as this simply results in a weaker effective binding of the TF to the specific operator (53), i.e., a larger K. Thus, nonspecific binding can be accounted for by measuring specific binding strength to particular sites relative to genome-wide nonspecific binding. In Equations 5 and 6, we have not explicitly modeled the binding of the RNA polymerase and the subsequent steps that lead to production of mRNA. For a single promoter in steady-state conditions, this is not a serious limitation since nonequilibrium effects can also be absorbed into an effective K, preserving the form of Equations 5 and 6 (44). However, more detailed modeling may be required to reproduce certain dynamical features, such as bunched mRNA production due to slow TF binding (44) and delayed response due to slow initiation of transcription, long-lived mRNAs, slow protein folding (5), or, in eukaryotes, export/import of macromolecules between cellular compartments.

Given the simplifications made above, the dynamics of the concentration, C, of a protein produced from a gene repressed by a TF can then be modeled using the following differential equation:

$$\frac{dC}{dt} = leak + \frac{capacity}{1 + (R/K)^h} - \frac{C}{\tau}. \qquad 7.$$

The positive terms model production, whereas the last term models degradation of the protein. *leak* accounts for cases in which repression cannot reach 100%. *capacity* sets the maximum production rate of the protein over the basal level *leak*. Notice that we do not model transcription and translation separately; i.e., we do not have a separate equation for the mRNA concentration. This is reasonable whenever mRNA turnover is faster than protein turnover. Typically, mRNA half-lives are shorter than the corresponding protein half-lives, so this is often a good assumption. [The half-life of total *E. coli* mRNA is 6.8 min (48).] The lifetime of the protein is given by the parameter τ, which takes into account both passive and active degradation, as well as dilution due to cell growth. Note that using a degradation rate proportional to C involves making the simplifying assumption that none of the active degradation processes is limited by cellular factors.

We especially stress the importance of the parameter τ. A small value of τ on the one hand leads to a fast adjustment to new steady-state levels, but on the other hand increases protein turnover and thus the metabolic cost. **Figure 2*a*** shows how C responds to a sudden 100-fold rise in R, from $K/10$ to 10 K (that is, from a regime in which the operator-bound fraction is close to zero to a regime in which it is close to unity). τ sets the timescale for response in exactly the same way when the TF activates the production of the protein (**Figure 2*c***).

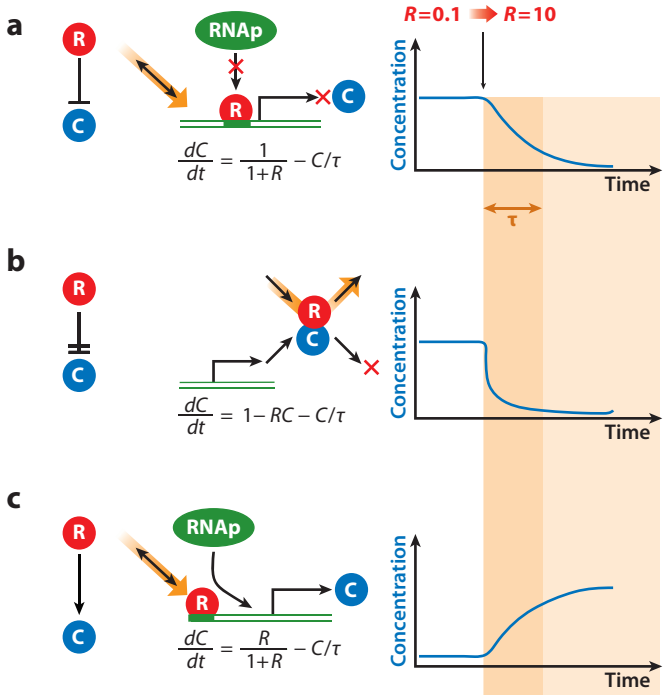

Figure 2

Simple ways to regulate the concentration of a protein, C, through negative and positive control. (*a*) The regulator, R, acts as a transcriptional repressor. (*b*) R catalyzes degradation of C. (*c*) R is a transcriptional activator. In each case, the adjacent plot shows schematically how C responds to a sudden 100-fold change in the concentration of R when each mechanism is modeled by the corresponding differential equations shown in the figure. To focus attention on the mathematical forms used for production and degradation terms, we have set all parameters in these equations (e.g., rate constants, dissociation constants, Hill coefficients) equal to unity, except for the lifetime of the protein, τ. Note that active degradation, panel *b*, can be much faster than transcriptional repression, panel *a*. For example 50% of LexA can be degraded within 1 min after activation of RecA in *Escherichia coli* (38), which degrades LexA actively. Such quick action would be impossible if RecA instead worked by repressing transcription of LexA.

Active Degradation and Irreversible Inactivation

The crucial role during the lifetime of the protein in setting its response time suggests that a particularly efficient way of achieving a fast response is to use a regulator that, instead of repressing transcription, actively degrades the protein. Such a regulatory link is displayed in **Figure 2*b***. The corresponding equation shows the simplest way of modeling such active degradation, where the rate of degradation is taken to be γRC. This assumes standard second-order kinetics of association of R and C, followed by rapid degradation (or irreversible inactivation) of C. Therefore, one does not need to consider dissociation of the R-C complex. Thus, $1/\gamma R$ becomes the effective average lifetime of the protein in the presence of the regulator. The advantage of such a mechanism is that one can achieve a fast response to changes in R while maintaining a low degradation rate, and therefore low metabolic cost, when R is low. Note also that in the mathematical formalism used here it does not matter whether C is a protein or

SMALL RNA REGULATION

In the text we discussed active degradation or irreversible inactivation of a protein (see Active Degradation and Irreversible Inactivation). We mentioned also that inactivation of mRNA by a sRNA could be modeled similarly. However, there is an additional interesting feature of regulation by sRNA, namely, that it is often associated with degradation of both the sRNA (s) and its target mRNA (m). This can be modeled by the equations $dm/dt = 1-\gamma ms - m/\tau$ and $ds/dt = \alpha - \gamma ms - s$, where α quantifies the regulatory input and γ the mutual interaction strength (35, 36, 42, 43). High γ or α allows a high (regulated) degradation rate of m, a rate that increases with $\gamma\alpha$. The speed of response, after a change in α, increases with the value of γ. Furthermore, one expects switch-like behavior of steady-state values of m, as a function of α, provided that γ is high (36). For RyhB-SodB in *E. coli*, one finds $(\alpha;\gamma) \approx (4;400)$, reflecting a large mutual interaction. In contrast, for the Spot42-galK system $(\alpha;\gamma) \approx (18;2)$, where a large α indicates a large potential for overproducing sRNA relative to its target (43). Notice that the above equations apply to any mechanisms where two molecules mutually inactivate each other by forming an irreversible complex.

an mRNA. Thus, the same sort of model could be used to model inactivation of a target mRNA (C) by a small- or mRNA (R), for instance, by antisense pairing and subsequent degradation (35, 36, 40, 42, 43) (see sidebar, Small RNA Regulation). Active degradation of proteins or mRNA is a major part of many metabolic and stress response systems.

DYNAMICS OF SINGLE FEEDBACK LOOPS

FLs are formed when links like those in **Figure 2** are combined into a closed cyclic chain; thereby, the influence of each regulator in this cycle eventually loops back to reach itself. Depending on the types of links in the cycle, this effective self-interaction can be positive or negative, and this strongly determines the behavior of the FL. RNs typically have multiple FLs of various signs. However, we begin by describing, in this section, the behavior of single, isolated negative and positive FLs.

Negative Feedback: Homeostasis and Oscillations

In the RN shown in **Figure 1*b***, an increase in intracellular iron increases the amount of active Fur, a TF that represses production of iron transporters. Thus, an increase in active Fur reduces the intake of iron, thereby closing the cycle and forming a FL. Following the links, one sees that the overall logic of the loop is to counteract any perturbation in the intracellular iron level. This is what we mean by negative feedback. **Figure 1*c*** contains another example: p53 activates the production of Mdm2, which binds to p53 and inactivates it. In general, a FL is negative if it has an odd number of repression links. The logic of negative feedback makes it ideal for stabilizing systems and minimizing fluctuations (12). Thus, negative feedback is associated mostly with maintenance of homeostasis. The most common form is where a small molecule binds to and inhibits an enzyme that catalyzes one of the earlier steps in its metabolic pathway. A similar effect is produced in riboswitches where the metabolite binds to and inactivates the mRNA of the necessary enzyme (39). These negative FLs use only posttranscriptional regulation.

Negative FLs can also involve transcriptional regulation. In particular, many TFs repress the transcription of their own gene, thereby buffering their regulation of other genes against global variations in protein-DNA binding, for example, due to modulations in intercellular salt concentrations (60). Negative autoregulation of a TF can be readily modeled using a straightforward modification of Equation 7 (50):

$$\frac{dR}{dt} = leak + \frac{capacity}{1+(R/K)^h} - \frac{R}{\tau}. \qquad 8.$$

The product $\tau \cdot capacity$ is a characteristic concentration that is important: R cannot exceed this value and, therefore, if $\tau \cdot capacity < K$, negative feedback is effectively always inactive. Conversely, $\tau \cdot capacity$ should be larger than K for the feedback to be active and relevant. In steady state, Equation 8 results in R growing slower than linearly as *capacity* is increased.

That is, the steady-state level of R is affected relatively little by changes in cellular factors that affect *capacity*, which is what makes a negative FL good for homeostasis. In addition, the steady state is reached faster than in the absence of feedback (47, 54), which is also useful for homeostasis.

When negative feedback is delayed, in time it can give rise to oscillations (**Figure 3*b***). We can add an explicit time delay to Equation 8, giving

$$\frac{dR}{dt} = leak + \frac{capacity}{1 + (R(t - \tau_d)/K)^b} - \frac{R}{\tau}. \quad 9.$$

In the parameter regime where negative feedback is active, the system favors homeostasis and shows no oscillations when the time delay, τ_d, is small (typically, of the order of the regulator lifetime, τ, or less), and shows oscillations when the delay is larger than a critical amount. Such time-delayed differential equations have been used to model oscillations in the developmental regulator Hes1, which inhibits its own transcription (26, 27), as well as more complex negative FLs in p53 response (55, 56) and zebrafish somitogenesis (34). Time delay can arise in many ways. Transcription, and translation in particular, can be relatively slow in eukaryotes. For example, Hirata et al. (26) suggest that delays caused by transcription and translation in production of Hes1 are of the order of 25 min.

Other mechanisms can also cause an effective time delay, for instance, when a protein binds to a TF catalyzing its degradation. Such saturated degradation (13, 29, 56) has been suggested as the source of delay underlying oscillations observed in negative feedback systems of p53, NF-κB, and Wnt. In such a case, it may be more appropriate to model the complex formation explicitly instead of hiding it within a parameter like τ_d in Equation 9. Overall, when there are oscillations, the period is set by the slowest rate in the system, which is usually the decay time of a protein. Therefore, the pattern of oscillation can be changed by genetic engineering of system components (32).

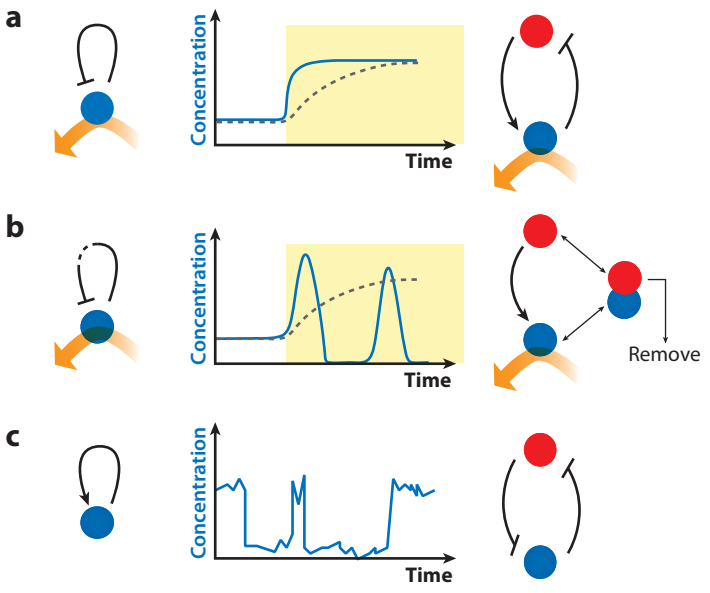

Figure 3

Schematic illustration of negative and positive feedback loops along with their respective dynamical behavior. (*a*) Negative feedback stabilizes the output to a near-constant level and allows for fast transient increase in production in response to stress or perturbations. (*b*) If negative feedback is delayed, the protein concentrations may oscillate in time. (*c*) Positive feedback can result in bistability; i.e., the system can exist stably in either of two distinct steady states. A second transcription factor that modulates the shift from one state to another then typically causes an ultrasensitive response (18, 52).

Negative Feedback in Stress Response

Cells often need to respond to environmental stresses, which may be biotic (e.g., viral infection) or abiotic (e.g., osmotic stress, heat shock, DNA damage, presence of toxic substances, and mechanical damage). In any of these situations, the cell has a repertoire of appropriate proteins that can mitigate the stress. Accordingly, when stress is experienced, the cell communicates the need for stress response proteins to the transcription/translation factors that can activate production of the required proteins—this is a situation suitable for negative feedback regulation. However, stress response usually needs to be initiated as quickly as possible to minimize the damage caused by the stress. Therefore, one usually finds that negative FLs in stress response systems involve protein-protein interactions in which a TF regulates the production of a

Saturated degradation: when degradation of a protein is catalyzed by limiting cellular factors, the rate saturates at a value independent of the protein concentration

protein that in turn catalyzes the proteolytic inactivation of the TF.

Protein-protein binding/unbinding is a fast process that can be used to reduce the half-life of an active protein, thereby making it respond much faster than if the protein was controlled by transcription regulation only. A classic example of a negative feedback motif that uses protein-protein binding is shown in **Figure 1*c***: The p53-dependent response to DNA damage in mammals involves inactivation of p53 by binding to Mdm2, among other things. Similar feedback motifs are found in the heat shock (6) and in the SOS response systems in *E. coli* (1, 30). The main role of negative feedback in such systems is to bring the system back to its quiescent state, in which stress response protein production is halted once the stress is brought down to a harmless level. From a dynamic point of view, this kind of behavior is similar to that in metabolic homeostasis systems—the overall goal in either case is to stabilize the system and counteract perturbations.

Motif: small, well-defined, and common subnetwork of regulatory systems

Bistability: refers to a system that has two distinct and stable steady states

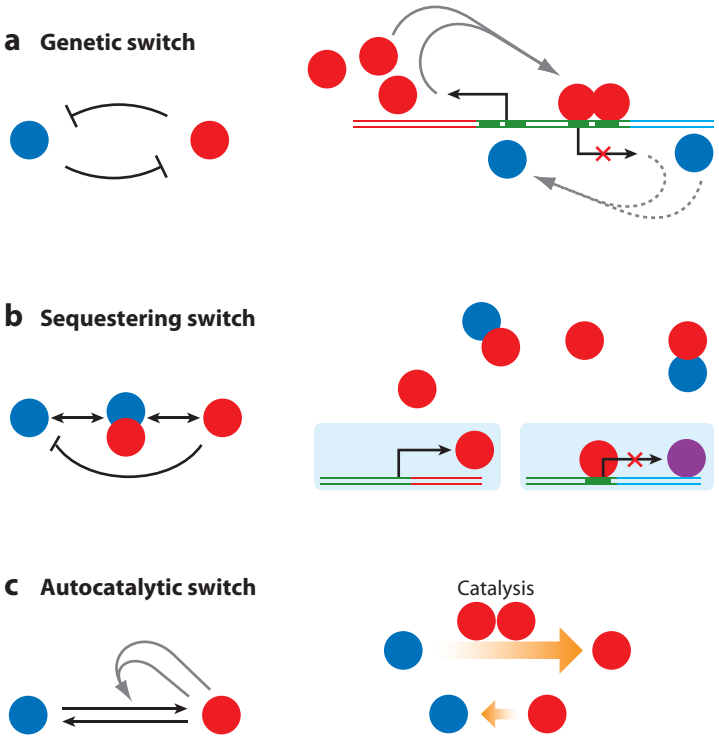

Figure 4

Bistability mechanisms. (*a*) Mutual transcriptional repression, also known as a genetic switch. The double-tailed arrows in the right panel indicate that the transcription factors need to cooperatively repress the operators of the opposing protein. This motif is found in the regulatory network of phage λ (45). (*b*) A positive feedback loop that uses strong protein-protein binding so that only the abundant protein species is available in free form. Here, the blue promoter is stronger and the heterodimer does not act as a transcriptional repressor. This motif is found in phage TP901-1 (46). (*c*) A switch consisting of catalyzed modifications in which, for example, proteins in the red state can direct proteins in the blue state to change to red. The directed processes compete with random background conversions. A version of this mechanism is found in nucleosome modification processes associated with mating-type genes in *Schizosaccharomyces pombe* (14).

Positive Feedback: Switch-Like Responses and Bistability

In contrast to negative feedback, positive feedback is often associated with instability and tends to amplify perturbations. Positive feedback can result from self-activation or from double negative feedback, as illustrated in **Figure 3*c***. In the extreme, such a regulatory motif can lead to bistability. That is, the system can stably exist in one of two distinct states: one state in which the concentration is low, below the threshold at which self-activation is triggered, and the other state in which the concentration is high, where self-activation results in a much higher production rate. Even in less extreme situations, where there is no bistability, a positive FL of this kind will show a switch-like response in which the steady state of the protein concentration shifts sharply and nonlinearly from a low level to a high level, or vice versa, as cellular factors affecting protein production are varied. With noise, the long-term dynamic behavior of a switch may look like that shown in **Figure 3*c***, exhibiting periods of time when the system is in a high or low state, with relatively rapid transitions between these states (8).

Figure 4*a* shows a common transcriptional positive FL consisting of two mutually repressing TFs. Extending the formalism of Equation 7, we can model this as follows:

$$\frac{dR}{dt} = \frac{c_R}{1 + (L/K_L)^{b_L}} - \frac{R}{\tau_R} \qquad 10.$$

and

$$\frac{dL}{dt} = \frac{c_L}{1+(R/K_R)^{b_R}} - \frac{L}{\tau_L}, \qquad 11.$$

where for simplicity we have ignored the *leak*, assuming it to be negligible for both proteins. As before, for the feedback to have a chance of working, the parameters should satisfy $\tau_L \cdot c_L > K_L$ and $\tau_R \cdot c_R > K_R$. This system is bistable when (*a*) $c_L \cdot \tau_L / K_L$ and $c_R \cdot \tau_R / K_R$ are not too different from each other, and (*b*) Hill coefficients are greater than 1. When both of these conditions are met, the system can be in one of two states: one in which R is large and L is low, and vice versa. In other words, the first criterion implies that the strength of each repression must be similar; symmetry aids bistability. The second criterion is important: Cooperativity in the interactions is necessary for bistability. The bistability is more stable and easier to achieve, the higher the Hill coefficients (i.e., the more the cooperativity).

Positive Feedback in Epigenetic Behavior

Bistability is, in fact, a prerequisite for epigenetic memory—the ability of genetically identical cells to differ in phenotypes and to maintain their distinct phenotypes from one generation to the next.

Epigenetic behavior based on bistability (or, in general, multistability) is ubiquitous in development. Perhaps the simplest, and best-studied, example of a developmental decision is the lysis-lysogeny decision of bacteriophage λ (45). The core of the λ network (**Figure 1*a***) contains positive feedback implemented by two TFs (CI and Cro) that mutually repress each other (the scheme of **Figure 4*a***). Thus, the case of phage λ can be understood largely on the basis of the analysis of Equations 10 and 11, even though this motif ignores many details of the phage RN. With R corresponding to CI and L corresponding to Cro, the two states would correspond to the lysogenic (R high) and lytic (L high) states exhibited by λ (57). In λ, this basic motif is augmented by a positive FL due to CI self-activation and additional regulatory links that are important for other functions λ needs to perform (45, 58).

Bacteriophages that have a functional repertoire similar to that of phage λ have structurally similar RNs (58), indicating that mutual repression is a robust way to achieve bistability. However, alternative designs also exist. For example, phage TP901-1 achieves bistability using the regulatory motif shown in **Figure 4*b*** (46). As opposed to the motif shown in **Figure 4*a***, this implementation of a positive FL uses protein-protein binding. For bistability the binding must be strong enough so that only the protein having a larger concentration exists in its free form, having mopped up most of the minority protein into the complex. This results in a large effective Hill coefficient and therefore helps to achieve bistability.

The scheme in **Figure 4*c*** shows a third possible implementation of a positive FL that can exhibit bistability. This implementation can be applied to epigenetics based on nucleosome modification in eukaryotes. Consider two types of opposing histone modifications (**Figure 4*c***). If the interactions between modified nucleosomes and the associated modifying enzymes are such that two nucleosomes of one type can convert an opposing nucleosome to their type, as in **Figure 4*c***, then the system can exhibit bistability. The system will then be found in a state in which most nucleosomes are of the same type. Once again, therefore, bistability requires some sort of effective cooperativity. This scheme has been used to explain silencing of mating-type genes in yeast (14).

DYNAMICS OF TWO ENTANGLED FEEDBACK LOOPS

Metabolism and Uptake of Small Molecules

Phage RNs miss one component, namely, metabolism. Phages use the energy and resources generated by host metabolism, but substantial fractions of free-living organisms are devoted to conversion of food molecules into raw material for anabolic processes. Metabolic

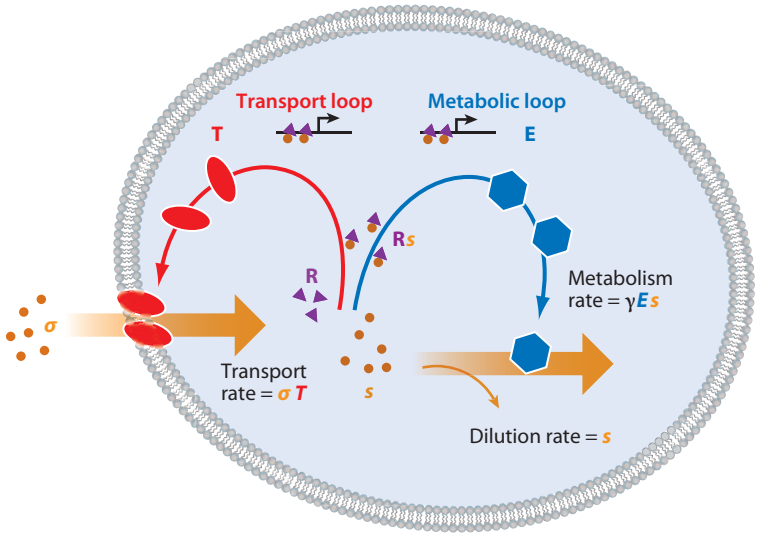

Figure 5

Entangled feedback loop (FL) motifs for regulation of uptake and metabolism of small molecules. The small molecule has an extracellular concentration σ and an intracellular concentration s. It is imported by transport proteins, T, into the cell, where it is consumed in reactions catalyzed by enzymes, E. In addition, it may be diluted by cell growth. A regulator, R, binds to the small molecule, thereby sensing its concentration. R in turn affects production of T and E, thereby closing two FLs. Several logical possibilities exist for all these interactions: The active form of R may be either the free, unbound form or when it is complexed to the small molecule; R may repress or activate production of T or E. These alternatives result in different combinations of the two FLs: both positive, both negative, or one negative and one positive.

networks may be viewed as a mixture of intertwined production lines. The flow of matter along these routes can be quantified computationally through flux balance and other analyses (28). RNs controlling metabolism contain a vast number of FLs. The feedback typically involves the formation of a complex between a small molecule and a TF that regulates enzymes that, in turn, affect metabolism of the small molecule. Not only are such FLs much more common than purely transcription FLs, they also function on faster timescales. In *E. coli*, about 50% of all TFs bind small molecules, thus forming a zoo of feedback circuits on the interface between the RN and the metabolic network (3, 4).

A typical feedback motif associated with small-molecule metabolism is shown schematically in **Figure 5**. A transcriptional regulator (R) senses the intracellular concentration of a particular small molecule (s) and regulates transcription of the transport proteins (T) using one FL, facilitating the influx of the small molecule. With the second FL, R controls transcription of enzymes (E) responsible for the metabolism of the small molecule. The dynamics of the intracellular concentration, s, of the small molecule will be given by an equation of the form

$$\frac{ds}{dt} = \sigma T - \gamma E s - s. \qquad 12.$$

Here, again we make a simplification, namely, that the enzymes E and T are functioning in the linear, unsaturated regime. It would be more correct to use Michaelis-Menten terms, but this does not change the qualitative behavior we describe below. We choose units of concentration so that T and E lie between zero and unity, and the rate of dilution due to cell growth is unity (the last term above). σ is the extracellular concentration of the small molecule, and γ specifies

the rate at which the molecule is metabolized in units of the dilution rate due to cell growth.

γ is an important characteristic parameter because it controls the ratio between the flux of small molecules and the size of the intracellular pool: To begin, let us ignore the feedback, i.e., set T and E fixed to unity. Then, in steady state, the intracellular pool will be $s = \sigma/(\gamma+1)$, while the flux will be σ per cell generation. Typically, small metabolites are processed quickly, and at any time only a minute fraction of the small molecules are freely available in the cell, compared with the flux. In other words, γ is typically large. This implies that any change in the uptake or usage of a small molecule is quickly (within a timescale of the order *generation time*/γ) converted into a change in the intracellular concentration of the small molecule.

For iron in *E. coli*, when growing in rich medium, a cell will have 10,000 free or loosely bound Fe^{2+} ions intracellularly, whereas the flux is of the order of 10^6 per cell generation (61). Therefore $\gamma_{Fe} \sim 100$. The biological implication is that most of the iron in an *E. coli* cell is in fact already integrated into various proteins, which, in our framework, is the same as irreversible consumption of iron. For nutritional molecules like glucose, galactose and lactose, the estimate for γ is much higher (61).

Different Ways of Combining Two FLs

In order to model the feedback from the intracellular small-molecule concentration to the transport and metabolic systems, we must introduce equations for T and E to supplement Equation 12. As explained above, feedback typically uses a transcriptional regulator that senses the small-molecule concentration by forming a complex with it.

Assuming that s is always much larger than the regulator concentration, R, we obtain the concentrations of free and bound R:

$$R_{free} = R^{tot}\frac{1}{1+s}; \quad \{Rs\} = R^{tot}\frac{s}{1+s}, \quad 13.$$

where R^{tot} is the total amount of regulator and s is counted in units of the dissociation constant of the $\{Rs\}$ complex. The derivation of these formulae is similar to the derivation of the fraction of TFs bound to an operator site—after all, the two situations are analogous because both of them involve complex formations considered to be in quasi-equilibrium.

The concentration of the active form of R, denoted R^*, is either R_{free} or $\{Rs\}$ depending on details of the regulation. In case R^* is a repressor, the dynamics of E is given by

$$\frac{dE}{dt} = leak + \frac{capacity}{1+(R^*/K_E)^{b_E}} - E, \quad 14.$$

where again *leak* represents a small basal activity, and *capacity* is chosen so that E lies between 0 and 1. A similar equation governs the transport system T. $K_{E,T}$ and $b_{E,T}$ will then parameterize the corresponding dissociation constants and Hill coefficients.

Each FL in **Figure 5** can be characterized by a sign that denotes whether the loop implements positive (+) or negative (−) feedback on s. Note that a positive metabolism FL does not mean an increase in the metabolic rate with increase of s. Rather, it means exactly the opposite: An increase of s leads to a decrease in the metabolic rate, hence there is a positive feedback of the s level onto itself. In describing the logic of the entangled loop motifs, we use the notation of two signs; e.g., (+ −) means that the transport loop is positive and the metabolism loop negative.

There are four possible logical structures, each of which can be implemented in two distinct but logically equivalent ways depending on whether s inhibits or activates the regulator. Overall, we find that there is no qualitative difference between the two implementations (33). Each of the four structures constrains its function (**Figure 6**):

- (− −) Maintains homeostasis of intracellular small-molecule concentration, i.e., exhibits relatively small variation of s with σ; useful for essential molecules that are harmful at high intracellular concentrations, e.g., iron.
- (+ −) Differentiates between low- and high-consumption states in a step-like manner that strongly favors consumption

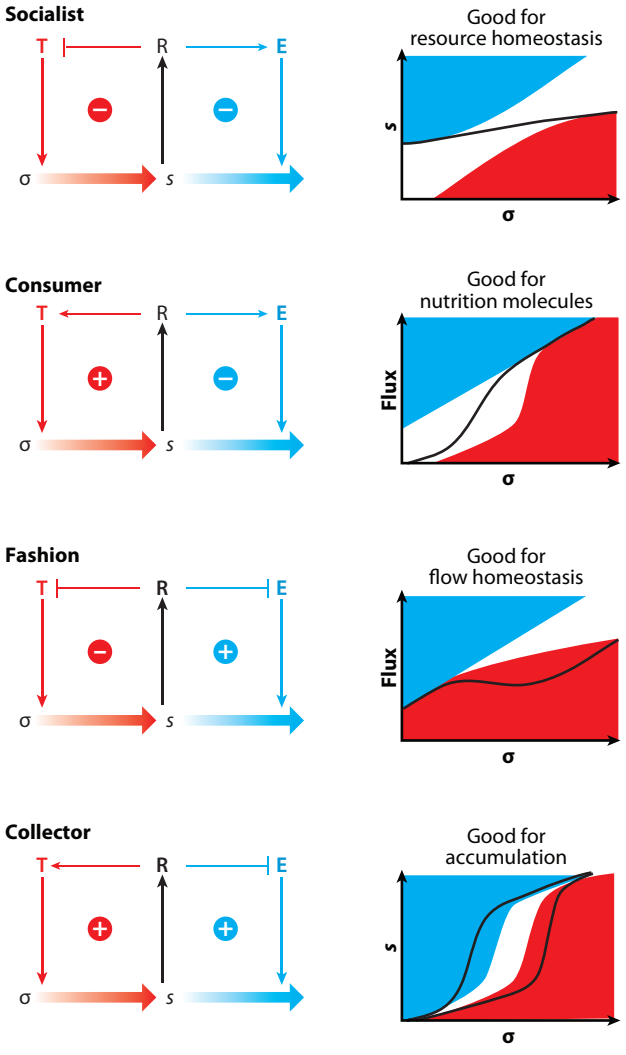

Figure 6

Steady-state values of s or flux ($\sigma T = \gamma E s + s$) as a function of σ, illustrated schematically, for each of the four feedback systems. The black curve in each graph corresponds to standard parameter values where the two loops are equally strong. The boundary of the red (or blue) region is the behavior when E (T) is fixed at unity, i.e., when only the transport (metabolism) feedback loop is active.

of highly abundant molecules; found in many sugar-uptake systems.

- (− +) Homeostasis of small-molecule flux. This motif is a subpart of the Fe-uptake system in *E. coli*.
- (+ +) Generates robust bistability in both the flux and the intracellular small-molecule level.

sRNA: small regulatory RNA

The functional behavior of each two-loop motif has tempted us to name them the socialist (− −), the consumer (+ −), the fashion (− +), and the collector motif (+ +), respectively (33).

In addition to steady-state behavior, each motif exhibits a dynamic response to sudden changes in resources. The response time is limited primarily by the relatively long time it takes to stabilize to new levels of E and T. In general, the timescale for complete recovery is set by the decay times of E and T. As E and T are usually abundant workhorse proteins, they are not actively degraded and decay times are therefore typically set by dilution during one cell generation. For some particularly valuable metabolites, the relaxation time may be modulated by regulating decay times of E. An example of this is found in the Fe-uptake system in *E. coli*, where the role E is taken by mRNA and its degradation time is modulated by the small regulatory RNA (sRNA) RyhB (49).

BEYOND MOTIFS

We can now return to some of the more complex networks shown in **Figure 1** to summarize the lessons we have learned from simple models of regulation and feedback.

The λ network essentially involves only transcription regulation. Transcription is relatively slow and thus not prone to fast feedback. The λ circuit is a decision-making and decision-locking network, where the initial choice is made quite fast using the rapidly degraded CII protein. The central memory maintaining the initial choice is centered around the mutual repression of CI and Cro (**Figure 1*a***). The role of this switch is to maintain bistability—one state is lytic growth, the other is lysogeny. The rest of the network involves unidirectional signal propagation with little feedback. One case in which such a signal has been quantified is induction of phage λ, where UV radiation activates RecA, which in turn destabilizes the CI-Cro switch. Therefore, it facilitates excision of phage DNA and production of sufficient holins to induce lysis (**Figure 1*a***) (2). Thus, even pure transcriptional RNs are able to distribute

information on distances comparable to the famous six degrees of separation suggested by studies of human social networks (41). The central switch in the core of the network works as a way to focus signal propagation. A possible role of the many positive FLs observed in extended RNs in living systems could be to direct and focus signals to relevant parts of the network. Adding switches seems to correlate with robustness in a number of systems (37).

In contrast to the transcription-dominated λ network is the p53 system of **Figure 1*c***, in which protein-protein binding plays a central role. Posttranslational regulation generally helps FLs to work on faster timescales than transcription FLs do and are therefore useful in stress response systems. Indeed, similar core motifs are found in the heat shock and SOS systems of *E. coli* (1, 30). The unfolded protein response system also uses mRNA manipulation and translation attenuation to overcome the stress (9). Strangely enough, not all stress response systems establish homeostasis quickly. In the p53 system, for example, oscillations with a period of 2–4 h have been observed in response to DNA damage (11). The cause of these oscillations is likely to be an effective time delay due to saturated degradation (56), but their physiological function is unknown.

The third example in **Figure 1*b***, the iron utilization network, is a model system for studying regulation of small molecules and sRNA. Fast responses are possible because of the large γ for Fe consumption and because sRNA regulation involves fast decay of key mRNA in the Fe utilization pathway (49). If feedback were not fast, small perturbations in gene expression of transporters or enzymes would cause the internal pool of Fe to fluctuate wildly. In order to maintain Fe homeostasis, the system uses multiple negative FLs. In fact, the core of the iron regulation is the (– –) motif. The Fe system involves an additional positive feedback on the Fe utilization side; compare the FL involving RhyB and mRNA with the FL involving Suf and FeS in **Figure 1*b***. The purpose of this mixing of the socialist and the fashion motifs, schematized in **Figure 7*a***, appears to be prioritization

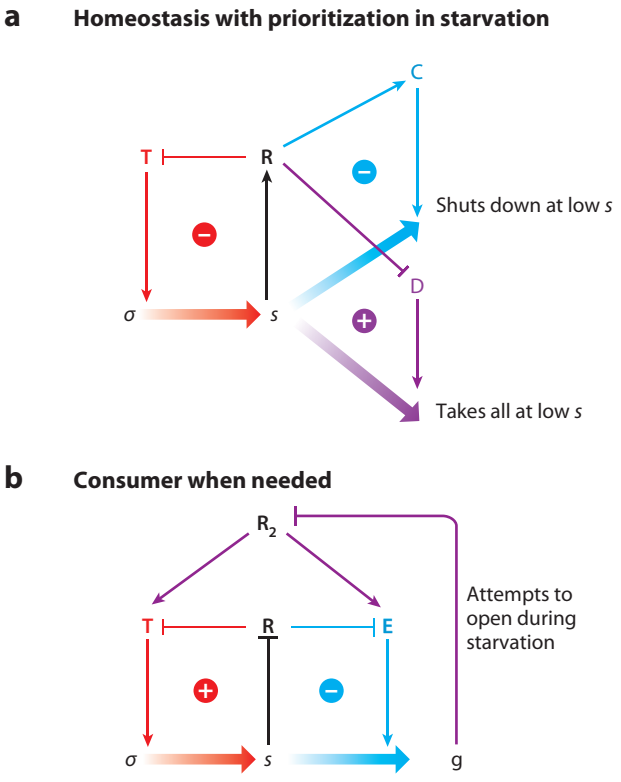

Figure 7

Different ways to combine multiple feedback loops (FLs). (*a*) A subset of the *Escherichia coli* Fe system of Figure 1*b*. The red FL controls transport. Blue and purple FLs function to redirect the flux of small-molecule consumption to the purple arrow when the extracellular concentration of the small molecule falls to starvation levels. Here, the sign of the FLs is the primary determinant of the behavior of this motif. (*b*) A subset of a typical sugar-uptake system, e.g., galactose regulation in *E. coli*. Again, the red FL controls transport. The blue FL controls metabolism of the sugar. The purple FLs use a second transcription factor to sense carbon starvation, and respond by increasing import of galactose and its subsequent conversion to glucose.

during Fe depletion: When Fe gets scarce, negative feedback (socialist) decreases its utilization, whereas positive feedback (fashion) tends to strengthen the depletion by using the little Fe that was left. As a result, Fe is channeled into the FeS channel (by enzymatic conversion, E) and away from the Fe-using proteins (which are regulated by the sRNA RyhB) (40).

One must keep in mind, however, that with an increasing number of FLs combined in one system it will be harder to pinpoint clear principles connecting structure and function. One can

see this already with more than two entangled FLs. Consider sugar-uptake systems, which at the core are consumer motifs but also have an overarching negative FL that can increase *E* and *T* production during carbon starvation (see **Figure 7*b***). In contrast to the iron three-loop motif of **Figure 7*a***, this motif uses a second TF to sense the level of another small molecule (cAMP, a measure of carbon starvation). Both TFs regulate the same set of genes; therefore, the logic of signal integration at the production of E and T (51) is also a major determinant of network behavior (as opposed to the iron three-loop motif for which the main determinant remains the signs of the FLs). The logic of signal integration is determined by molecular details of network components and interactions, e.g., the structure of promoters and regulatory regions, the mechanism of transcription regulation, and the interplay between TFs (31). The challenge for the bottom-up approach is to nevertheless extract principles governing the behavior of multiple entangled FLs. Understanding such principles would emphasize core parameters to be measured to predict the function and dynamics of larger RNs.

SUMMARY POINTS

1. Feedback is an essential part of molecular networks. It allows the cell to adjust the repertoire of functional proteins to current needs.
2. A FL is primarily characterized by its sign: negative feedback for maintaining homeostasis, positive feedback for obtaining ultrasensitivity or multiple stable states of the cellular composition.
3. Negative feedback can cause oscillations if signal propagation around the FL is sufficiently slow. High Hill coefficients, additional positive FLs, or saturated degradation facilitates oscillations in a negative FL.
4. Positive feedback can come from strong self-activation of a gene, from mutual repression between proteins, or by autocatalytic processes. In all cases one can obtain bistability if reactions involve some sort of cooperativity.
5. Metabolism of small molecules is characterized by a separation of scales. Typically, the intracellular pool of available small molecules is much smaller than the total amount of small molecules consumed during one cell generation.
6. Combinations of FLs in small-molecule uptake and metabolism can result in new behavioral features that are significantly different from a simple sum of the behaviors of single loops.

FUTURE ISSUES

1. How often are simplifying assumptions (e.g., quasi-equilibrium of TF-operator binding) good approximations, in practice, in living cells?
2. How should one characterize RNs with multiple entangled FLs?
3. To what extent is positive feedback used to focus signals in RNs and thereby facilitate modularity of the network?
4. How important is the role of physical space in modulating the behavior of a RN?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

1. Aksenov SV. 1999. Dynamics of the inducing signal for the SOS regulatory system in *Escherichia coli* after ultraviolet irradiation. *Math. Biosci.* 157:269–86
2. Amir A, Kobiler O, Rokney A, Oppenheim AB, Stavans J. 2007. Noise in timing and precision of gene activities in a genetic cascade. *Mol. Syst. Biol.* 3:71
3. Babu M, Teichmann SA. 2003. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* 31:1234–44
4. Anantharaman V, Koonin EV, Aravind L. 2001. Regulatory potential, phyletic distribution and evolution of ancient, intracellular small-molecule-binding domains. *J. Mol. Biol.* 307:1271–92
5. Andrews BT, Gosavi S, Finke JM, Onuchic JN, Jennings PA. 2008. The dual-basin landscape in GFP folding. *Proc. Natl. Acad. Sci. USA* 105:12283–88
6. Arnvig KB, Pedersen S, Sneppen K. 2000. Thermodynamics of heat shock response. *Phys. Rev. Lett.* 84:3005–8
7. Aurell E, Brown S, Johansen J, Sneppen S. 2002. Stability puzzles in lambda phage. *Phys. Rev. E* 65:051914
8. Aurell E, Sneppen K. 2002. Epigenetics as a first exit problem. *Phys. Rev. Lett.* 88:048101
9. Axelsen JB, Sneppen K. 2004. Quantifying the benefits of translation regulation in the unfolded protein response. *Phys. Biol.* 1:159–65
10. Barkai N, Leibler S. 1997. Robustness in simple biochemical networks. *Nature* 387:913–17
11. Bar-Or RL, Maya R, Segel LA, Alon U, Levine AJ, Oren M. 2000. Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proc. Natl. Acad. Sci. USA* 97:11250–55
12. Becskei A, Serrano L. 2000. Engineering stability in gene networks by autoregulation. *Nature* 405:590–93
13. Bliss RD, Painter PR, Marr AG. 1982. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* 97:177–93
14. Dodd IB, Micheelsen MA, Sneppen K, Thon G. 2007. Theoretical analysis of epigenetic cell memory by nucleosome modification. *Cell* 129:813–22
15. Elf J, Li GW, Xie XS. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* 316:1191–94
16. Elowitz MB, Leibler S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–38
17. Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 181:197–203
18. Ferrell JE Jr. 1996. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* 21:460–66
19. Ferrell JE Jr. 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* 14:140–48
20. Francois P, Hakim V. 2004. Design of genetic networks with specified functions by evolution in silico. *Proc. Natl. Acad. Sci. USA* 101:580–85
21. Gardner TS, Cantor CR, Collins JJ. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–42
22. Goldbeter A, Koshland DE Jr. 1984. Ultrasensitivity in biochemical systems controlled by covalent modification. Interplay between zero-order and multistep effects. *J. Biol. Chem.* 259:14441–47
23. Gonze D, Halloy J, Goldbeter A. 2002. Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. USA* 99:673–78
24. Harris SL, Levine AJ. 2005. The p53 pathway: positive and negative feedback loops. *Oncogene* 24:2899–908
25. Hasty J, Pradines J, Dolnik M, Collins JJ. 2000. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. USA* 97:2075–80

26. Hirata H, Yoshiura S, Ohtsuka T, Bessho Y, Harada T, et al. 2002. Oscillatory expression of the bHLH factor Hes1 regulated by a negative feedback loop. *Science* 298:840–43
27. Jensen MH, Tiana G, Sneppen K. 2003. Sustained oscillations and time delays in gene expression of protein Hes1. *FEBS Lett.* 541:176–77
28. Kauffman KJ, Prakash P, Edwards JS. 2003. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* 14:491–96
29. Krishna S, Jensen MH, Sneppen K. 2006. Minimal model of spiky oscillations in NF-kB signaling. *Proc. Natl. Acad. Sci. USA* 103:10840–45
30. Krishna S, Maslov S, Sneppen K. 2007. UV-induced mutagenesis in the *Escherichia coli* SOS response: a quantitative model. *PLoS Comput. Biol.* 3:e41
31. Krishna S, Orosz L, Sneppen K, Adhya S, Semsey S. 2009. Relation of intracellular signal levels and promoter activities in the *gal* regulon of *Escherichia coli*. *J. Mol. Biol.* 391:671–78
32. Krishna S, Semsey S, Jensen MH. 2009. Frustrated bistability as a means to engineer oscillations in biological systems. *Phys. Biol.* 6:036009
33. Krishna S, Semsey S, Sneppen K. 2007. Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proc. Natl. Acad. Sci. USA* 104:20815–19
34. Lewis J. 2003. Autoinhibition with transcriptional delay: a simple mechanism for the zebrafish somitogenesis oscillator. *Curr. Biol.* 13:1398–408
35. Lenz DH, Mok KC, Lilley BN, Kulkarni RV, Wingreen NS, Bassler BL. 2004. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* 118:69–82
36. Levine E, Zhang Z, Kuhlman T, Hwa T. 2007. Quantitative characteristics of gene regulation mediated by small RNA. *PLoS Biol.* 5:e229
37. Li F, Lu Y, Long T, Ouyang Q, Tang C. 2004. The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. USA* 101:4781–86
38. Little JW. 1983. The SOS regulatory system: control of its state by the level of RecA protease. *J. Mol. Biol.* 167:791–808
39. Mandal M, Breaker RR. 2004. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* 5:451–63
40. Masse E, Gottesman S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 99:4620–25
41. Milgram S. 1967. The small world problem. *Psychol. Today* 1:60–67
42. Mitarai N, Andersson AMC, Krishna S, Semsey S, Sneppen K. 2007. Efficient degradation and expression prioritization with small RNAs. *Phys. Biol.* 4:164–71
43. Mitarai N, Benjamin J-AM, Krishna S, Semsey S, Csiszovszki Z, et al. 2009. Dynamic features of gene expression control by small regulatory RNAs. *Proc. Natl. Acad. Sci. USA* 106:10655–59
44. Nakanishi H, Mitarai N, Sneppen K. 2008. Dynamical analysis on gene activity in the presence of repressors and an interfering promoter. *Biophys. J.* 95:4228–40
45. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. 2005. Switches in bacteriophage lambda development. *Annu. Rev. Genet.* 39:409–29
46. Pedersen M, Hammer K. 2008. The role of MOR and the CI operator sites on the genetic switch of the temperate bacteriophage TP901-1. *J. Mol. Biol.* 384:577–89
47. Rosenfeld N, Elowitz M, Alon U. 2002. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* 323:785–93
48. Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* 13:216–23
49. Semsey S, Andersson AMC, Krishna S, Jensen MH, Masse E, Sneppen K. 2006. Genetic regulation of fluxes: iron homeostasis of *Escherichia coli*. *Nucleic Acids Res.* 34:4960–67
50. Semsey S, Krishna S, Erdossy J, Horvath P, Orosz L, et al. 2009. Dominant negative autoregulation limits steady-state repression levels in gene networks. *J. Bacteriol.* 191:4487–91
51. Semsey S, Krishna S, Sneppen K, Adhya S. 2007. Signal integration in the galactose network of *Escherichia coli*. *Mol. Microbiol.* 65:465–76
52. Sneppen K, Micheelsen MA, Dodd IB. 2008. Ultrasensitive gene regulation by positive feedback loops in nucleosome modification. *Mol. Syst. Biol.* 4:182

53. Sneppen K, Zocchi G. 2005. *Physics in Molecular Biology*. Cambridge, UK: Cambridge Univ. Press
54. Svenningsen SL, Waters CM, Bassler BL. 2008. A negative feedback loop involving small RNAs accelerates *Vibrio cholerae*'s transition out of quorum-sensing mode. *Genes Dev.* 22:226–38
55. Tiana G, Jensen MH, Sneppen K. 2002. Time delay as a key to apoptosis induction in the p53 network. *Eur. Phys. J. B* 29:135–40
56. Tiana G, Krishna S, Pigolotti S, Jensen MH, Sneppen K. 2007. Oscillations and temporal signalling in cells. *Phys. Biol.* 4:R1–17
57. Toman Z, Dambly-Chaudiere C, Tenenbaum L, Radman M. 1985. A system for detection of genetic and epigenetic alterations in *Escherichia coli* induced by DNA-damaging agents. *J. Mol. Biol.* 186:97–105
58. Trusina A, Sneppen S, Dodd IB, Shearwin KE, Egan JB. 2005. Functional alignment of regulatory networks: a study of temperate phages. *PLoS Comput. Biol.* 1:e74
59. Tsai TY, Choi YS, Ma W, Pomerening JR, Tang C, Ferrell JE Jr. 2008. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* 321:126–29
60. Wall ME, Hlavacek WS, Savageau MA. 2004. Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.* 5:34–42
61. Wong P, Gladney S, Keasling JD. 1997. Mathematical model of the *lac* operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol. Prog.* 13:132–43

Annual Review of Biophysics
Volume 39, 2010

Contents