DESIGN OF GENE CIRCUITS: LESSONS FROM BACTERIA

Michael E. Wall[*‡], *William S. Hlavacek*[§] *and Michael A. Savageau*[‖]

Researchers are now building synthetic circuits for controlling gene expression and considering practical applications for engineered gene circuits. What can we learn from nature about design principles for gene circuits? A large body of experimental data is now available to test some important theoretical predictions about how gene circuits could be organized, but the data also raise some intriguing new questions.

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.

GENETIC REGULATORY CIRCUIT
Also called a gene circuit. The genes and gene products that are involved in the response to a signal.

**Computer and Computational Sciences Division and ‡Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. §Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ‖Department of Biomedical Engineering, One Shields Avenue, University of California, Davis, California 95616, USA. Correspondence to M.E.W. e-mail: mewall@lanl.gov*
doi:10.1038/nrg1244

Early studies of gene regulation began with a small number of bacterial systems and led to the celebrated operon model of Jacob and Monod, which introduced concepts such as OPERON, regulator gene and transcriptional repression[1,2]. This model was elaborated extensively[3,4] as different regulatory mechanisms, such as transcriptional activation[5], were discovered. Today, hundreds to thousands of diverse GENETIC REGULATORY CIRCUITS, mostly in bacteria, have, to some extent, been characterized experimentally. This wealth of knowledge, which is complemented by the availability of many genome sequences, has motivated us to identify clear patterns in the DESIGN of these circuits and to search for DESIGN PRINCIPLES that can explain their natural diversity.

Studies of design will soon become even more compelling, as gene circuits are being characterized experimentally with increasing speed. Technological advances, such as the development of cDNA microarrays, allow gene expression to be monitored crudely on a genome-wide scale[6,7]. *In vivo* fluorescent reporter systems allow gene expression to be accurately monitored on a multigene scale and with fine time resolution[8,9]. This ability to comprehensively and quantitatively monitor dynamic changes in gene expression, together with new genome-scale informatics methods, facilitates high-throughput characterization of genetic regulatory networks[10–12]. This capability is complemented by emerging methods for *in vivo*, high-throughput determination of parameter values in mathematical/computational models of genetic regulatory networks[9]. The impressive body of data already available, and the types of data being generated or contemplated at present, will allow a comprehensive global understanding of gene regulation that cannot be obtained through the study of any individual system[13].

Synthetic regulatory circuits can be readily built, owing to the advanced state of genetic engineering[14–20]. Attention is turning towards manipulating genetic regulatory circuits for therapeutic and technological applications — gene circuits for BIOREMEDIATION[21], metabolic engineering[22] and gene therapy[23] are being constructed. Such applications require a thorough understanding of the functional consequences of genetic manipulations and of the general principles that can guide the design process.

Recent theoretical studies[24–28] of gene regulation have elucidated design principles for transcriptional regulation of bacterial transcription factors (TFs) in ELEMENTARY GENE CIRCUITS. Here, we review these design principles, which provide a framework for understanding and organizing a large body of data, as we illustrate by examining an assembled database that incorporates information about 50 TFs in *Escherichia coli*. The database can be used, for example, to test our understanding of these design principles, to empirically identify patterns of TF regulation and to identify gaps in our knowledge. In the remainder of the review, we provide an overview of the study of gene-circuit design, with an emphasis on the perspective gained from our own studies. We then review specific design principles that we have identified through theoretical studies. These principles have allowed us to make and test predictions about features of elementary circuits of two types: INDUCIBLE-catabolic and REPRESSIBLE-biosynthetic circuits. After a comparison of these predictions with the empirical results collected

DESIGN
The constellation of system components, their specific properties and their pattern of interactions that together determine the integrated behaviour of the system. The term 'structure' might also be used but 'design' is preferred when there is a functional context.

DESIGN PRINCIPLES
General concepts that summarize our understanding of how gene-circuit design relates to gene-circuit function.

BIOREMEDIATION
The use of either naturally occurring or deliberately introduced micro-organisms to consume and break down environmental pollutants.

ELEMENTARY GENE CIRCUIT
A gene circuit in which gene expression is regulated by a single transcription factor in response to a signal under a given set of conditions. When conditions change, however, gene expression might come under the influence of extra regulators.

INDUCIBLE
Describes a gene, the expression of which increases in response to a signal in a given environmental background. An inducible system is one in which the effector transcriptional unit is inducible.

REPRESSIBLE
Describes a gene, the expression of which decreases in response to a signal in a given environmental background. A repressible system is one in which the effector transcriptional unit is repressible.

SIGNAL
A natural molecule that acts directly on the transcription factor to bring about a physiological response.

STABILITY
The ability of a system to return to a steady state after a transient disturbance.

ROBUSTNESS
The ability of a system's steady state to remain unchanged, or not significantly changed, when the structure (that is, the parameter values) of the system significantly changes.

for the database and a survey of these results, we conclude by briefly identifying classes of gene circuits for which, at present, design principles are unavailable and that could be the focus of future research.

Perspective on gene circuit design

Studies of gene circuits (both experimental and theoretical) are similar to many other areas of biological research — the principal aim is to understand the relationship between structure and function. For example, as we discuss below, patterns of regulation in elementary gene circuits can be understood in terms of the functional requirements for biosynthesis and catabolism[25,28]. For natural systems, the important design features are those that can confer a selective advantage in an ecological context[29,30]. This is in contrast to directed evolution and rational improvement of synthetic circuits, in which selection of features is an artefact of engineering.

The biological context determines the functional requirements for the gene circuit. For example, consider the lactose system (*lacI-lacZYA*) in *E. coli*[31], which induces lactose catabolism in response to the SIGNAL allolactose, a catabolic intermediate. The circuit is expected to maintain a stationary basal level of β-galactosidase (the *lacZ* gene product) in the absence of allolactose, to dynamically increase the expression level when the level of allolactose increases, and to maintain a higher, stationary level of expression in the presence of a stationary, inducing level of allolactose. The circuit function requires both the basal and induced states to be stable. If the circuit is to operate in a variety of environments, the function must be robust to environmental changes. If the organism must act quickly to make use of metabolites, catabolism needs to be induced quickly in response to the signal. STABILITY, ROBUSTNESS and RESPONSIVENESS are therefore expected to be important performance criteria for the lactose system and other catabolic gene circuits. We also expect these criteria to be important for biosynthetic gene circuits, such as the tryptophan system (*trpR-trpLEDCBA*) in *E. coli*[32,33], which must meet similar requirements.

Which gene circuit designs, if any, are selected in which contexts, and why? This question is especially intriguing when the same biological functions can be carried out using diverse designs. For example, changes in gene expression in catabolic gene circuits might, in principle, be brought about by a natural signal that is either the substrate, an intermediate or the product in a regulated pathway. To predict which kind of signal would be selected in a given context, we must determine the functional consequences of the alternative designs. A rigorous theoretical approach, known as mathematically controlled comparison, can be used to quantitatively compare the functional characteristics of alternative designs[30,34,35] (BOX 1). Application of this method using performance criteria for catabolism and biosynthesis has resulted in the design principles for the regulation of TF expression that are reviewed here. These principles are in part based on two other design principles: one provides rules for the molecular MODE OF CONTROL of EFFECTOR GENES and the other provides rules for connectivity of the signal molecule in inducible catabolic circuits. Before we summarize the design principles for the molecular mode of gene control, signal connectivity and regulation of TF expression, we discuss some general features of the relevant gene circuits.

General features of elementary gene circuits

MODULAR (sub-) systems for catabolism and biosynthesis can often be modelled as elementary gene circuits, in which gene expression is regulated specifically by a single TF, the activity of which is modulated by a signal. (Other TFs that are not affected by the local signal might be involved: for example, LacI and cyclic AMP receptor protein (CRP) both regulate expression of the *lac* operon, but only LacI interacts with allolactose.)

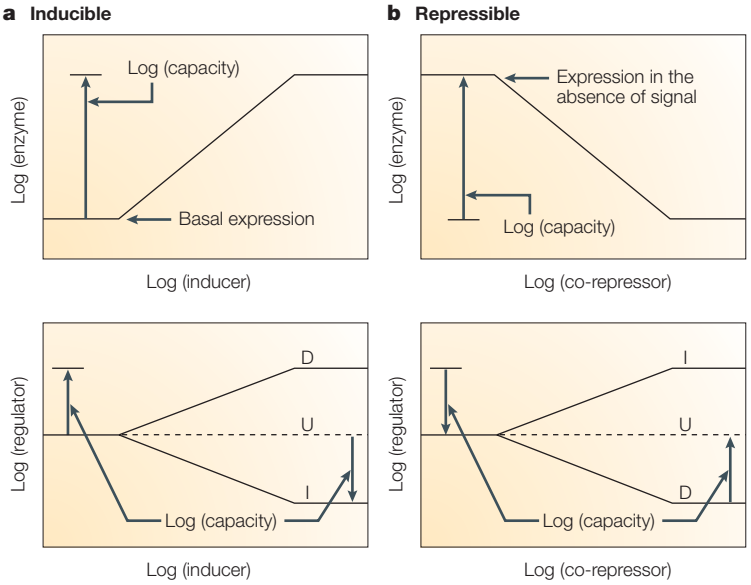
Box 1 | The method of mathematically controlled comparison for studying design principles

This method can be likened to a well-controlled competition experiment: mathematically precise *a priori* conditions for functional effectiveness are defined and a generalized model that mathematically describes the different types of system under consideration is developed[30,34,35]. Each type of system is described by a special case of the model and corresponds to a region in the PARAMETER SPACE that characterizes all systems described by the model. The *a priori* mathematically precise conditions are then used to make comparisons among systems that are selected from these regions; the systems under comparison can differ in ways that are essential for distinguishing system types but are otherwise constrained to be as similar as possible. (For example, systems are allowed to have differences in transcriptional control so that systems of different forms of coupling can be compared, but are constrained to be identical with respect to translation.) If differences in functional effectiveness are observed, the various system types are associated with optimal solutions for distinct regulatory problems. If no differences are observed, variations among system types are considered to be neutral. A mathematically controlled comparison is therefore a mathematically precise way to seek design principles in regulatory networks.

In the theoretical studies we review, when other features of performance are equivalent, the most important differences in functional effectiveness are for three *a priori* conditions: stability, robustness and responsiveness. A measure of stability is a distance in parameter space between a system with a stable steady state and the nearest system with an unstable steady state. Robustness is measured by the sensitivity of the steady-state value of each dependent variable to changes in each model parameter. Responsiveness can be measured by analysing the recovery time of a system after a step change in an independent variable. One measure is the rise time, which is the time it takes for the value of a variable to come close to its new steady-state value. Another measure is the settling time, which is the time after which the value of a variable never strays far from its new steady-state value.

Box 2 | **Examples of qualitative types of steady-state response to a signal in an elementary gene circuit**

The activity of a TF can be affected by a signal molecule, either an inducer or co-repressor, according to whether its presence increases or decreases effector gene expression. The signal molecule need not have the same effect on every promoter. The responses are characterized by the EXPRESSION CHARACTERISTICS **that are illustrated in the figure. For inducible systems, steady-state enzyme levels increase or decrease, respectively, in response to an increase or decrease in the availability of a small-molecule inducer of effector gene expression (a). The** GAIN**, which is positive, is measured as the slope of the inclined region, and the enzyme** EXPRESSION CAPACITY **is measured as the ratio of fully induced expression to basal expression. For repressible systems, steady-state enzyme levels decrease or increase in response to an increase or decrease, respectively, in the availability of a small-molecule co-repressor of effector gene expression (b). The gain is negative, and the enzyme expression capacity is measured as the ratio of expression in the absence of signal to fully repressed expression. The directly coupled (D), uncoupled (U) and inversely coupled (I) patterns of regulator gene expression are illustrated in the bottom panels of (a) and (b).**



RESPONSIVENESS
The ability of a system to settle quickly into a new steady state after an environmental change.

MODE OF CONTROL
Either positive (activator control) or negative (repressor control) according to whether an increase in the level of the transcription factor (other factors being constant) acts to increase or decrease gene expression.

EFFECTOR GENE
A gene that encodes an enzyme and/or another molecule with an effector function (for example, membrane transport).

MODULAR
A (sub-) system of interacting components is modular if it shows behaviour that is independent of the larger system under certain conditions.

PARAMETER SPACE
A list of values for all *N* parameters of a model corresponds to a point in an *N*-dimensional parameter space. A specific system type, as specified by constraints on parameter values, corresponds to a region in parameter space.

EXPRESSION CHARACTERISTIC
A plot of the level of expression versus the level of signal over a range of steady states.

The TF binds near the promoter of one or more TRANSCRIPTIONAL UNITS (TUs). At each of these TUs, the TF might act as either an activator or a repressor (alternatively, repression might be caused by TRANSCRIPTIONAL ATTENUATION, the functional consequences of which have been considered elsewhere[36]).

A specific small-molecule metabolite, such as allolactose, typically acts as a signal to communicate the need for the enzyme. If effector gene expression increases following the addition of a signal, the circuit is termed inducible; if it decreases, the circuit is termed repressible. Catabolic enzymes are needed when substrates are available, and catabolic gene circuits tend to be induced by a metabolic intermediate. By contrast, biosynthetic gene circuits tend to be repressed by a metabolic end product when synthesis of biosynthetic enzymes would be wasteful.

A signal need not have the same effect on the expression of different TUs in an elementary gene circuit — a given signal can bring about one of three qualitatively different patterns of coupling between regulator and effector gene expression: direct coupling, inverse coupling or uncoupling (BOX 2). TFs often regulate their own expression — a phenomenon known as autogenous regulation, or autoregulation. Just as effector gene expression might be under activator or repressor control, regulator gene expression might be under activator control (which yields positive autoregulation), repressor control (which yields negative autoregulation) or it might be unaffected by the TF (FIG. 1). For example, among inducible circuits in *E. coli*, *dsdC-dsdXA*[37–39] shows direct coupling, *cynR-cynTSX*[40–42] shows uncoupling and *metR-metE*[43,44] shows inverse coupling. Among repressible circuits in *E. coli*, *trpR-trpLEDCBA*[32,33] shows direct coupling and *tyrR-(aroF-tyrA)*[45,46] shows uncoupling. Although an inversely coupled circuit remains a formal possibility, there seem to be no examples of this among repressible systems in *E. coli* and other bacteria. All of the above circuits show negative autoregulation; in the inducible circuit *lacI-lacZYA*[31] and the repressible circuit *modEF-modABCD*[47], TF expression is not self-regulating.

Molecular mode of gene control

A design principle that accounts for natural selection of either the activator or repressor mode of control is called demand theory[29,30,48]. The central argument of demand theory is that, in an evolutionary context, systems in which demand for effector activity is high (low) are predicted to have activator (repressor) control, because in each case, loss of regulation, owing to mutation, results in a relatively high fitness penalty. By high (low) demand, we mean that expression of effector genes is at the high (low) end of the physiological range most of the time in the natural environment. The qualitative predictions of this theory have been supported by experimental data for many systems and physiological contexts[49,50] and a quantitative version of this theory has recently been applied to the inducible lactose and maltose systems of *E. coli*[51].

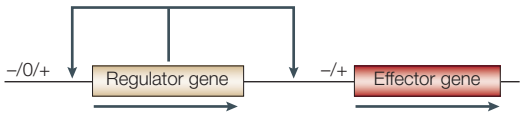


Figure 1 | **Modes of control by a transcription factor in an elementary gene circuit.** A regulator gene encodes a transcription factor (TF) that can act as either repressor or activator at a given promoter. The effector gene encodes an enzyme or some other type of protein with an effector function, such as membrane transport. The mode of control, repressor (–) or activator (+), need not be the same at all promoters, and the regulator gene need not be regulated (0).

As we show below, the molecular mode of gene control must be taken into account when interpreting the rules for coupling of regulator and effector gene expression.

Signal connectivity

A design principle that accounts for the intermediate position of the signalling molecule in an inducible catabolic circuit is based on a comparative analysis of systems with various positions for this molecule[30]. The stability, robustness and responsiveness of the system all depend on whether the signal is the substrate, the product or is an intermediate metabolite in the regulated catabolic pathway. Circuits in which the natural inducer is an intermediate have a steady state, the stability of which is less sensitive to parameter changes than in equivalent circuits with either the substrate or the product as their natural inducers. The most robust and responsive circuits are those in which the natural inducer is a substrate, followed by those in which the natural inducer is an intermediate, and circuits in which the natural inducer is the product are the least robust and responsive of the three types. These results indicate that the best compromise position in terms of robustness and responsiveness is for the natural inducer to be an intermediate in the inducible pathway, as seems to be the case for the circuits that have been best characterized experimentally[52]. For this reason, the studies of inducible systems that are reviewed in the following section focused on models in which the natural inducer is an intermediate. A more recent theoretical study of inducible switches supports the importance of an intermediate position for the natural inducer of a catabolic circuit[53].

Regulation of TF expression

Expanding on earlier analyses of autogenous and constitutive regulation of TF expression[30,52,54,55], a series of theoretical studies has led to the identification of design principles that account for the form of TF autoregulation and the type of coupling between the expression of regulator and effector TUs[24–28]. In these studies, models of inducible catabolic and repressible biosynthetic gene circuits were first developed, and criteria for functionally effective circuits were then postulated on the basis of physiological requirements. Mathematically controlled comparison was used to determine the functional consequences of alternative modes of autoregulation and types of coupling. Systems with activator and repressor control were considered separately, because the mode of control is selected for reasons explained by demand theory. The models considered in the most comprehensive studies[25,27,28] are described in BOX 3. For all models of inducible catabolic circuits, the catabolic intermediate was chosen as the signal, as suggested by the design principle for signal connectivity.

The results of these and earlier studies show that negative autoregulation increases the stability, robustness and responsiveness of elementary gene circuits[24–28,30,54,55]. Experimental studies to address the effect of negative autoregulation on the stability[56] and responsiveness[57] of synthetic gene circuits support this finding. The theoretical results also show that the mode of control for the effector gene and the type of coupling together influence the responsiveness of a gene circuit. For circuits with activator control, inverse coupling is most responsive, whereas for circuits with repressor control, direct coupling is most responsive. The achievement of the most responsive types of coupling is dependent on

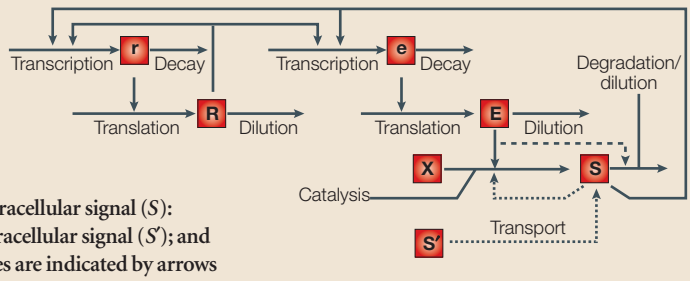
GAIN
If y is the level of enzyme, and x is the level of signal, the gain of the system is defined as $\partial \log y / \partial \log x$ (where ∂ = partial derivative). Gain, according to this definition, is also used interchangeably with the term 'logarithmic gain'. Often, the expression characteristic might be described using the Hill equation, $y = (1 + x^n)^{-1}$, in which case a representative gain of the system in a log-log plot can be $\partial \log y / \partial \log x \approx n$.

EXPRESSION CAPACITY
The ratio of maximal to minimal signal-dependent expression levels.

TRANSCRIPTIONAL UNIT
(TU). A DNA sequence that is transcribed as a single polycistronic mRNA, and might encode one or more individual genes.

TRANSCRIPTIONAL ATTENUATION
A decrease in transcription that results from a disengagement of mRNA polymerase from the DNA before reading through a leader sequence. Attenuation is enhanced by an increase in the level of an amino acid that corresponds to codons that are transcribed from the leader sequence.

Box 3 | Model of elementary gene circuits



Inducible and repressible gene circuits have been modelled[25,27,28] by considering the following processes: transcription and decay of regulator (r) and effector (e) mRNA; translation and dilution of regulator (R), which is a transcription factor (TF), and enzyme (E); and processes that influence the level of intracellular signal (S): reaction of substrates (X); transport of extracellular signal (S'); and degradation/dilution. Regulatory influences are indicated by arrows that terminate in the middle of another arrow. For inducible circuits, the effect of enzyme on degradation of signal is represented by a dashed arrow. For repressible circuits, transport of extracellular signal and feedback inhibition of catalysis are represented as dotted arrows. System features can be defined with reference to the model:

• In systems with an activator (repressor) mode of control, the regulator exerts a positive (negative) influence on transcription of effector mRNA.

• In systems with positive (negative) autoregulation, the regulator exerts a positive (negative) influence on transcription of regulator mRNA. The regulator might also have no influence on transcription of regulator mRNA.

• In inducible (repressible) systems, an increase in the steady-state level of intracellular signal leads to an increase (decrease) in the steady-state level of enzyme.

• In directly (inversely) coupled systems, an increase in the steady-state level of intracellular signal leads to an increase (decrease) in the steady-state level of regulator. In uncoupled systems, changes in the steady-state level of intracellular signal have no effect on the steady-state level of regulator.

Box 4 | **Limitations on the type of coupling when the gain is sufficiently high**

Specific combinations of coupling and mode of control are prevented for systems in which the steady-state level of enzyme is sufficiently sensitive to signal, as measured by the gain L[25,27,28] (see BOX 2). L can be calculated as the sum of two terms: $L=L_1+L_2$. L_1 is the gain in the absence of autoregulation and is derived from the direct influence of signal on effector mRNA transcription. L_2 is the contribution to the gain from autoregulation and is derived from the influence of signal on regulator mRNA transcription and the subsequent influence of regulator on effector mRNA transcription. The magnitude of L_1 is limited by a factor, the value of which must be smaller than the total number of molecules of the signal that bind to control TF interactions near the promoter of the effector TU[25,28]. For a given sign of L, there is therefore a critical magnitude of the gain $|L|=|L^*|$ above which the sign of L_2 must be the same as that of L. The signs of L_2 and L are the same for direct (inverse) coupling with activator (repressor) control and opposite for inverse (direct) coupling with activator (repressor) control. The consequences of the constraint on $|L_1|$ are therefore as follows: for systems with $|L|>|L^*|$, only direct (inverse) coupling is possible for activator (repressor) control; for systems with $|L|=|L^*|$, uncoupling is also possible; and for systems with $|L|<|L^*|$, all types of coupling are possible.

the effector expression characteristic (BOX 2). In particular, the type of coupling is limited when the system is required to have a sufficiently large steady-state gain of enzyme with signal (BOX 4). Natural selection for a particular type of coupling might therefore be influenced by the advantages of both maximizing responsiveness and providing an appropriate steady-state gain. Design principles based on these results predict relationships between the gain and the type of coupling for elementary gene circuits (TABLE 1).

The design principles for the regulation of TF expression have been tested using experimental data for a small number of well-studied inducible[25] and repressible[28] systems in bacteria. The predictions can be tested rigorously through the measurement of six system features; two of these features are commonly measured and are readily available for many well-studied *E. coli* systems[13,58] (data for the other features must, for now, be obtained from the primary literature). The two commonly measured features are the mode of control both for effector gene expression (activator control or repressor control) and for regulator gene expression (positive autoregulation, negative autoregulation or no self-regulation). Data that characterize the former might be used to test the predictions of demand theory; they also influence the predictions about the type of coupling (see TABLE 1). Data that characterize the latter might be used to test predictions about the mode of control in the self-regulation of TFs. Two other features that are required to test predictions of coupling type are commonly characterized experimentally: response of effector gene expression to the signal (inducible or repressible), and response of regulator gene expression to the signal (inducible, repressible or constitutive). Together, these features determine the type of coupling in the system (direct, inverse or uncoupled). In special cases, however, direct coupling can be inferred when the TF gene is found in an effector TU, without the need to measure the response to the signal.

The final two features that are needed to test predictions for the type of coupling, the steady-state gain and the CRITICAL GAIN, are not commonly characterized experimentally. These two properties together determine whether the gain is high, intermediate or low, which has important implications for the predictions (TABLE 1). The steady-state gain can be readily determined experimentally by measuring the expression characteristic (BOX 2). The value of the critical gain is model-dependent and cannot be measured directly; its magnitude can, however, be estimated as the total number of molecules of the signal that bind to control TF interactions near the promoter of the effector TU[25,28].

As measurements of the gain are not readily available, it is not possible at present to test rigorously the predictions about the coupling type in many systems. Instead, attempts to understand whether the predictions are in reasonable agreement with the available data have so far relied on examination of the enzyme expression capacity (BOX 2) — a commonly measured quantity that is expected to be correlated with the steady-state gain[25]. Owing to the varying experimental methods and measurement conditions, as well as high scatter in cases for which several measurements are available, the precise value of any one-expression capacity is not expected to be very meaningful. Nevertheless, expression capacities obtained from the literature do show an overall trend that supports the predictions[25,28].

CRITICAL GAIN
A model-dependent quantity that is used as a reference to determine whether the system gain is high, intermediate or low. The value can be estimated as the total number of molecules of the signal that bind to control transcription-factor interactions near the promoter of the effector transcriptional unit[25,28].

Table 1 | **Predictions of coupling type for elementary gene circuits***

Effector TU type	Low gain gene circuit	Intermediate gain gene circuit	High gain gene circuit
Inducible (+)	Inverse coupling	Uncoupling	Direct coupling
Inducible (–)	Direct coupling	Uncoupling	Inverse coupling
Repressible (+)	Inverse coupling	Direct coupling	Direct coupling
Repressible (–)	Direct coupling	Uncoupling	Inverse coupling

*Predictions depend on both the mode of control of effector expression and the magnitude of the steady-state gain (BOX 4). The predictions for inducible and repressible systems are identical except for the case of activator control with intermediate gain; (–) indicates a repressor mode of control, and (+) indicates an activator mode of control.

Database to characterize TF expression

A large body of experimental data on regulation of gene expression is now available and is being assembled into genome-scale databases, such as RegulonDB (see online links box)[58]. Although RegulonDB does not document the influence of signals on gene expression, it can be used to show the prevalence of negative autoregulation in *E. coli*[13,59] and to raise new questions about patterns in the connectivity of transcriptional regulatory interactions[13]. In contrast to RegulonDB, a database for genome-wide testing of design principles for the regulation of TF

Table 2 | **Distribution of system types among 49 *E. coli* TFs***

Effector TU type	Repressor mode of control at regulator TU			Activator mode of control at regulator TU			No TF self-regulation
	I	U	D	I	U	D	U(0)
Inducible (+)	4‡	3§	4‖	0	0	5¶	4#
Inducible (–)	0	0	9**	0	0	0	4‡‡
Repressible (+)	0	3§§	0	0	0	0	2‖‖
Repressible (–)	0	1¶¶	9##	0	0	0	1***

*The following footnotes indicate the sets of transcription factors that correspond to the Table entries (see online table 1): ‡(AraC, IlvY, MetR, SoxS); §(CynR, SoxR, TorR); ‖(CsyB, DsdC, MelR, RhaS); ¶(CpxR, IdnR, MarA, RhaR, XylR); #(MalT, MhpR, Rob, XapR); **(BetI, CytR, EmrR, GalS, MarR, NagC, PdhR, PutA, UxuR); ‡‡(GalR, GlpR, LacI, RbsR); §§(AsnC, GcvA, PspF); ‖‖(FadR, FruR); ¶¶(TyrR); ##(ArgR, DnaA, Fur, H-NS, IscR, MazEF, MetJ, PurR, TrpR); ***(ModE). Due to a lack of published data on self-regulation, TreR was not used to compile the numbers in this table. D, direct coupling; I, inverse coupling; TF, transcription factor; TU, transcriptional unit; U, uncoupling with TF self-regulation; U(0), uncoupling with no TF self-regulation, a special type of uncoupling; (–) indicates a repressor mode of control; (+) indicates an activator mode of control.

expression would require assembly of the features mentioned in the previous section for a comprehensive set of TFs in an organism. As it would include detailed information on the effects of signals on gene expression, the database would also serve broader interests by providing easy access to important information that is, at present, not available in other databases.

To illustrate the potential utility of such a database, we have documented and compiled information about some key design features for 50 TFs in *E. coli*. The compilation is publicly available at the EcoTFs web site (see online links box), and in online TABLE S1. The activity of each of these TFs is modulated by a signal, the identity of which is recorded. TUs that are known to be regulated by the TF are also recorded, as is the mode of control at the effector TUs for each TF. The mode of control at the regulator TU and the type of coupling are recorded for each TF, except for TreR. The distribution of the system types in this database is given in TABLE 2.

The EcoTFs database, like RegulonDB[58], can be used to test the theoretical prediction that negative autoregulation should be preferred on the basis of stability, robustness and responsiveness. A total of 33 of the 49 TFs for which autoregulation has been characterized are negatively autoregulated, which is consistent with the results of previous global analyses of TF interactions in *E. coli*[13,59]. The number of negatively autoregulated TFs is therefore consistent with an expectation that stability, robustness and responsiveness are important for many of the surveyed systems. Although we have not compiled the biological functions of each of these systems, there is a bias towards catabolism and biosynthesis among known gene circuits. This bias can, in part, explain the relatively high percentage of TFs that are negatively autoregulated. However, because stability, robustness and responsiveness are important properties for a wide variety of man-made automatic controllers, we also expect these criteria to be important for cellular functions other than catabolism and biosynthesis.

In cases for which information is partial, the theory can be used to make predictions about systems in the database. For example, self-regulation has not been documented for TreR. Assuming that stability, robustness and responsiveness are important performance criteria for the *treR-treBC* system, TreR is predicted to be negatively autoregulated. In addition, this database does not document the steady-state gain or critical gain for any of the systems. Assuming that stability, robustness and responsiveness are important performance criteria for all of the systems, TABLE 1 might be used to predict a classification for the gain using the documented coupling type and mode of control for the circuits in TABLE 2. The following TFs are predicted to have a smaller than critical gain: AraC, IlvY, MetR, SoxS, BetI, CytR, EmrR, GalS, MarR, NagC, PdhR, PutA, UxuR, ArgR, DnaA, Fur, H-NS, IscR, MazEF, MetJ, PurR and TrpR. Conversely, the following TFs are predicted to have a higher than critical gain: CysB, DsdC, MelR, RhaS, CpxR, IdnR, MarA, RhaR and XylR. The remaining TFs are predicted to have a gain that is equal to the critical gain: CynR, SoxR, TorR, MalT, MhpR, Rob, XapR, GalR, GlpR, LacI, RbsR, AsnC, GcvA, PspF, FadR, FruR, TyrR and ModE.

The database can be used to identify systems that do not agree with the predictions. This allows us to pose new questions, such as, why are CpxR, IdnR, MarA, RhaR and XylR all positively autoregulated? Positive autoregulation is not expected on the basis of stability, robustness and responsiveness, and therefore, the theory indicates that other performance criteria are important for the functions that are regulated by these TFs (CpxR and MarA are both regulators of genes that encode drug-resistance determinants[60,61], and IdnR, RhaR and XylR are involved in regulating expression of genes that encode L-idonate[62], L-rhamnose[63,64] and xylose[65] catabolic enzymes). The theory provides an explanation for positive autoregulation in inducible catabolic systems: a requirement for very high gain (a criterion that we have not emphasized here) can only be realized by the use of positive autoregulation[24,25]. Some of the systems that have positively autoregulated TFs (such as rhamnose[63,64] and some non-*E. coli* systems[66–70]) do indeed show relatively large expression capacities. Other explanations might involve functional requirements that are outside the scope of the theoretical studies. Systems that operate as discontinuous switches[14,17,53] and oscillators[16,17,71], for example, are expected to have very different requirements.

The absence of certain system types in the database also raises questions. For example, we did not find examples of inverse coupling among repressor-controlled systems in any of the systems that we surveyed. Assuming that this observation is indicative of an important pattern, what is implied by the absence of such examples? According to the theoretical predictions, systems with high gain are expected to have inverse coupling for repressor-controlled effector TUs (TABLE 1). The absence of inverse coupling might therefore indicate that the gain need not be high (relative to the critical gain) for systems that show repressor control. Alternatively, the degree of inverse coupling that is required might be small, and the experiments might not be sensitive enough to detect it. The same reason might account for the apparent enhancement in the number of systems with uncoupling.

HYSTERESIS
A possible attribute of a switch. A switch with hysteresis has a different threshold for the transition from the OFF state to the ON state compared with the transition from the ON state to the OFF state.

The database also shows a pattern that indicates a potential difference in the preferred mechanism of direct coupling between systems that have activator or repressor control. Direct coupling either might be indicated when the TF gene lies within an effector TU, or might only result from a similarity between the effects of a signal on expression of different TUs for regulator and effector. Among negatively autoregulated TFs, the only examples of the former type of direct coupling are for repressor control. The lack of this type of direct coupling for negatively autoregulated, activator-controlled systems means that they would be especially good targets of experiments to identify more inversely coupled systems, examples of which are scarce at present.

Finally, theoretical studies indicate specific ways in which the database can be expanded in order to increase its utility — for example, by recording the data that are required for testing theoretical predictions about the type of coupling. An example of such data is provided by a recent study of gene expression as a function of cyclic AMP and isopropyl-β-D-thiogalactoside (IPTG) levels in the *lacI-lacZYA* system[72]. In this study, the IPTG-dependent expression of a *lacZ-gfp* fusion was measured by fluorescence using a plasmid reporter system. The Hill equation, $y = (1 + x^{-n})^{-1}$, with a Hill coefficient n of approximately four, which is consistent with a steady-state gain of four, was found to be a good fit for the IPTG dependence of the *gfp* expression. This gain is the same as the estimated critical gain, which is four because there are four IPTG binding sites on the LacI tetramer and one tetramer that binds near the *lacZ* promoter. The steady-state gain is therefore intermediate, leading to a prediction of uncoupling because *lacZ* expression is inducible (TABLE 1). As *lacI* expression is indeed uncoupled from *lacZ* expression (see online TABLE S1), the coupling type in the *lacI-lacZYA* system, as determined from this experiment, is consistent with the predictions. However, earlier measurements using a β-galactosidase assay of chromosomal gene expression[73] are consistent with a Hill coefficient of two, which corresponds to a low gain and a prediction of direct coupling. As the plasmid *gfp* expression has been reported to be similar to chromosomal *lacZ* expression[72], the theoretical predictions lend support to the results of more recent fluorescent reporter assays. Measurements of steady-state gains that are unavailable at present, ideally obtained through systematic genome-wide assays using consistent methods, will be required in order to test coupling-type predictions for all systems in the database.

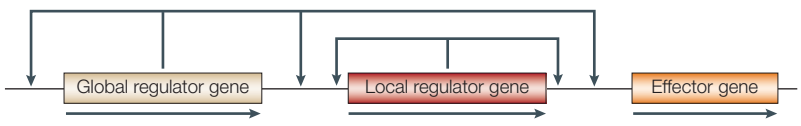


Figure 2 | **Transcriptional regulatory interactions in a gene circuit with a global and a local transcription factor.** Each arrow begins at a transcription factor (TF) gene and terminates at the promoter of a gene, transcription of which might be regulated by the TF (only the TF interactions that regulate effector gene expression are necessarily present). The illustration of the gene circuit shown here is similar to that of the elementary gene circuit shown in FIG. 1. Analysis of the functional consequences of alternative designs requires a model at the level of detail shown in BOX 3, including both the local and global signals.

Common themes

A common theme running through the three design principles reviewed here is the importance of system robustness, or insensitivity to parameter variation, which has long been used as a performance criterion in comparative analyses[74]. Recent experimental evidence provides direct support for the importance of this criterion in natural biochemical systems[75–77]. Other common themes include the importance of stability and responsiveness, criteria that have been used in the studies reviewed here. Theoretical results on negative autoregulation that are based on these criteria[24–28,54,55] have been supported by direct experimental studies of the effect of autoregulation on the stability[56] and responsiveness[57] of synthetic gene circuits. The latter comes closest to an experimental counterpart to a mathematically controlled comparison. Taken together, these theoretical and experimental results provide a rationale for the observation that most *E. coli* transcription factors regulate expression of their own genes, with many being negatively autoregulated[13,59].

Other gene circuits

Gene circuits that include discontinuous switching with HYSTERESIS represent a class of system that functions according to criteria that are different from those that we have considered so far. The most prominent difference is the stability of the steady state that lies on the inclined portion of the steady-state expression characteristic (BOX 2). This state must be stable for the class of circuit that is characterized by continuously variable expression[24,25,53], whereas it must be unstable for the class of discontinuously switching circuits[14,17,53]. Moreover, rigorous comparative analysis has shown that the discontinuous switch is less responsive and less robust, with regard to switching thresholds and switching time, than is the comparable continuously variable switch[53].

Discontinuous switching with hysteresis serves other functions and is likely to follow different rules. For example, these switches are more likely to be prominent in cells that undergo commitment to alternative developmental fates for which bi-stability is crucial. In these cases, the commitment to a particular pattern of gene expression is less susceptible to noise in the signal and more likely to result in an irreversible commitment. This class of circuit provides a functional role for positive autoregulation[17,53], a design that does not fit the predictions, except in cases that are expected to be unusual for the inducible catabolic or repressible biosynthesis systems that are the primary focus in this discussion.

Genetic oscillators comprise another class of circuit that functions according to separate criteria. Again, the most prominent difference is the stability of the steady state that lies on the inclined portion of the steady-state expression characteristic. This state must be stable for the circuits with continuously variable expression, whereas it must be unstable for the class of oscillatory circuits[16,17,71]. This class of circuit provides another example of the functional effectiveness of positive autoregulation[17,71], a design that does not fit the predictions for the circuits that have been the primary focus here.

NETWORK MOTIF
A common pattern of connections in a network.

Finally, it should be noted that molecules that are involved in gene regulation are often present in small numbers per cell, leading to notable stochastic effects that can influence the functional requirements of genetic circuits[78–85]. For example, fluctuations can cause unwanted noise that is then smoothed by downstream circuits that act as filters and produce the required mean behaviour. Stochasticity can also increase the robustness of oscillations to parameter changes[86], and might cause phenotypic diversity that is essential for long-term survival, as has been modelled for virulence factors that must evade immune surveillance[87]. These cases, which we have not considered, might well be at odds with the predictions reviewed here.

Future directions

If we look beyond the results observed for elementary gene circuits, it is clear that genome-wide approaches will allow the detection of higher-order patterns, such as those already seen in the genetic regulatory network of *E. coli*. For example, statistical analysis of the transcriptional regulatory networks of *E. coli* and *Saccharomyces cerevisiae* have shown NETWORK MOTIFS that involve two or more transcription factors and that occur more often than would be expected by chance[13,88,89]. *In vivo* fluorescent plasmid reporters have already been used to determine the combined effect of two input signals on gene expression[72] (a two-dimensional expression characteristic). This reporter system and other technologies[12] will allow the assembly of genome-wide transcriptional regulatory networks that include detailed information about signals, and that will motivate studies of increasingly complicated gene circuits, which might show even richer patterns of gene circuit design.

Analysis of more complex models that include local and global TFs will be needed to discover the design principles that underlie these more complicated gene circuits (FIG. 2). For example, despite having similar biological functions, the arabinose (*araC-araBAD*[31,90-92]) and lactose (*lacI-lacZYA*[31]) systems in *E. coli* are regulated in different ways by the same global TF — CRP. In the arabinose system, CRP regulates expression of the regulator TU, *araC*, whereas in the lactose system, CRP does not regulate expression of the regulator TU, *lacI*. These and other inducible systems have been analysed, using an elementary gene circuit model, for the study of regulation of gene expression by the local TF[25] (BOX 3), which requires the assumption that the effect of the global TF is constant. However, studying regulation of gene expression by the global TF requires a model that further includes both the global TF and the global signal (in this case, CRP and cAMP). The expanded model not only needs to take into account an increased number of transcriptional regulatory interactions, but also needs to account for the dynamic relationship between the local signal (such as arabinose or allolactose) and the global signal (such as cAMP) through cellular metabolism. In such cases, determination of the functional consequences of alternative designs is too complex for an intuitive analysis, and therefore, more systematic quantitative approaches are required for this task.

Although the studies reviewed here specifically involve bacteria, they also provide a framework for understanding the design of gene circuits in other cell types. For example, they should prove useful for studies of TF expression in eukaryotes, for which other effects (such as compartmentalization by the nuclear membrane) must be considered. Researchers are already beginning to address questions of gene circuit design in eukaryotes by, for example, asking why the bone morphogenic protein activation gradient in *Drosophila* embryos is robust in response to changes in gene dosage[77,93], and why transcriptional regulatory cascades tend to be longer in *Drosophila* developmental systems than in *E. coli* or *S. cerevisiae* sensory systems[94]. Answers to these and other questions about the design of gene circuits in all cell types will require not only theoretical and experimental studies of natural systems, but also rigorous analysis of the functional consequences of alternative designs.

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
2. Jacob, F. & Monod, J. On the regulation of gene activity. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 193–211 (1961).
3. Reznikoff, W. S. The operon revisited. *Annu. Rev. Genet.* **6**, 133–156 (1972).
4. Neidhardt, F. C. & Savageau, M. A. in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 1310–1324 (American Society for Microbiology, Washington DC, 1996).
5. Englesberg, E., Irr, J., Power, J. & Lee, N. Positive control of enzyme synthesis by gene C in the *L-arabinose* system. *J. Bacteriol.* **90**, 946–957 (1965).
6. Khodursky, A. B. *et al.* DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **97**, 12170–12175 (2000).
7. Martin, R. G. & Rosner, J. L. Genomics of the *marA/soxS/rob* regulon of *Escherichia coli*: identification of directly activated promoters by application of molecular genetics and informatics to microarray data. *Mol. Microbiol.* **44**, 1611–1624 (2002).
8. Kalir, S. *et al.* Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science* **292**, 2080–1083 (2001).
9. Ronen, M., Rosenberg, R., Shraiman, B. I. & Alon, U. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc. Natl Acad. Sci. USA* **99**, 10555–10560 (2002).
10. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
11. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
12. Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* **301**, 102–105 (2003).
13. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64–68 (2002).
Reports the discovery that certain kinds of patterns of mutual connections among genes, termed network motifs, occur more often than would be expected at random in the transcriptional regulatory network of *E. coli*. Also reports that regulatory connections are relatively shallow, emphasizing the importance of elementary gene circuits in *E. coli*.
14. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
15. Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
16. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
17. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607 (2003).
18. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100**, 7714–7719 (2003).
19. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. *Science* **296**, 1466–1470 (2002).
20. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. *Proc. Natl Acad. Sci. USA* **99**, 16587–16591 (2002).
21. Ronchel, M. C. & Ramos, J. L. Dual system to reinforce biological containment of recombinant bacteria designed for rhizoremediation. *Appl. Environ. Microbiol.* **67**, 2649–2656 (2001).
22. Farmer, W. R. & Liao, J. C. Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nature Biotechnol.* **18**, 533–537 (2000).
23. Ramachandra, M. *et al.* Re-engineering adenovirus regulatory pathways to enhance oncolytic specificity and efficacy. *Nature Biotechnol.* **19**, 1035–1041 (2001).
24. Hlavacek, W. S. & Savageau, M. A. Subunit structure of regulator proteins influences the design of gene circuitry: analysis of perfectly coupled and completely uncoupled circuits. *J. Mol. Biol.* **248**, 739–755 (1995).
25. Hlavacek, W. S. & Savageau, M. A. Rules for coupled expression of regulator and effector genes in inducible circuits. *J. Mol. Biol.* **255**, 121–139 (1996).
Reports the definitive relevant theoretical study of elementary inducible gene circuits in bacteria.

26. Hlavacek, W. S. & Savageau, M. A. Completely uncoupled and perfectly coupled gene expression in repressible systems. *J. Mol. Biol.* **266**, 538–558 (1997).
27. Hlavacek, W. S. & Savageau, M. A. Method for determining natural design principles of biological control circuits. *J. Intell. Fuzzy Syst.* **6**, 147–160 (1998).
28. Wall, M. E., Hlavacek, W. S. & Savageau, M. A. Design principles for regulator gene expression in a repressible gene circuit. *J. Mol. Biol.* **332**, 861–876 (2003).
Reports the definitive relevant theoretical study of elementary repressible gene circuits in bacteria.
29. Savageau, M. A. Genetic regulatory mechanisms and the ecological niche of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **71**, 2453–2455 (1974).
30. Savageau, M. A. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, Massachusetts, 1976).
The definitive source of information about mathematically controlled comparisons. Addresses numerous questions of biochemical system design, including the mode of self-regulation in inducible catabolic and repressible biosynthetic gene circuits, and the position of the natural inducer in an inducible catabolic gene circuit.
31. Beckwith, J. in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 1444–1452 (American Society for Microbiology, Washington DC, 1987).
32. Klig, L. S., Carey, J. & Yanofsky, C. *trp* repressor interactions with the *trp*, *aroH* and *trpR* operators. Comparison of repressor binding *in vitro* and repression *in vivo*. *J. Mol. Biol.* **202**, 769–777 (1988).
33. Somerville, R. The Trp repressor, a ligand-activated regulatory protein. *Prog. Nucleic Acid Res. Mol. Biol.* **42**, 1–38 (1992).
34. Savageau, M. A. A theory of alternative designs for biochemical control systems. *Biomed. Biochim. Acta* **44**, 875–880 (1985).
35. Alves, R. & Savageau, M. A. Extending the method of mathematically controlled comparison to include numerical comparisons. *Bioinformatics* **16**, 786–798 (2000).
36. Elf, J., Berg, O. G. & Ehrenberg, M. Comparison of repressor and transcriptional attenuator systems for control of amino acid biosynthetic operons. *J. Mol. Biol.* **313**, 941–954 (2001).
37. Heincz, M. C. & McFall, E. Role of small molecules in regulation of D-serine deaminase synthesis. *J. Bacteriol.* **136**, 104–110 (1978).
38. McFall, E. & Heincz, M. C. Identification and control of synthesis of the *dsdC* activator protein. *J. Bacteriol.* **153**, 872–877 (1983).
39. Nørregaard-Madsen, M., McFall, E. & Valentin-Hansen, P. Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus. *J. Bacteriol.* **177**, 6456–6461 (1995).
40. Sung, Y. C. & Fuchs, J. A. The *Escherichia coli* K-12 *cyn* operon is positively regulated by a member of the *lysR* family. *J. Bacteriol.* **174**, 3645–3650 (1992).
41. Guilloton, M. B. *et al.* A physiological role for cyanate-induced carbonic anhydrase in *Escherichia coli*. *J. Bacteriol.* **175**, 1443–1451 (1993).
42. Lamblin, A. F. & Fuchs, J. A. Functional analysis of the *Escherichia coli* K-12 *cyn* operon transcriptional regulation. *J. Bacteriol.* **176**, 6613–6622 (1994).
43. Urbanowski, M. L. & Stauffer, G. V. Regulation of the *metR* gene of *Salmonella typhimurium*. *J. Bacteriol.* **169**, 5841–5844 (1987).
44. Urbanowski, M. L. & Stauffer, G. V. Role of homocysteine in *metR*-mediated activation of the *metE* and *metH* genes in *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.* **171**, 3277–3281 (1989).
45. Camakaris, H. & Pittard, J. Autoregulation of the *tyrR* gene. *J. Bacteriol.* **150**, 70–75 (1982).
46. Pittard, A. J. in Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 458–484 (American Society for Microbiology, Washington DC, 1996).
47. Grunden, A. M., Ray, R. M., Rosentel, J. K., Healy, F. G. & Shanmugam, K. T. Repression of the *Escherichia coli modABCD* (molybdate transport) operon by ModE. *J. Bacteriol.* **178**, 735–744 (1996).
48. Savageau, M. A. Demand theory of gene regulation. I. Quantitative development of the theory. *Genetics* **149**, 1665–1676 (1998).
49. Savageau, M. A. Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl Acad. Sci. USA* **74**, 5647–5651 (1977).
50. Savageau, M. A. in *Theoretical Biology — Epigenetic and Evolutionary Order* (ed. Saunders, P. T.) 42–66 (Edinburgh Univ. Press, Edinburgh, 1989).
51. Savageau, M. A. Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of *Escherichia coli*. *Genetics* **149**, 1677–1691 (1998).
52. Savageau, M. A. in *Biological Regulation and Development* Vol. 1 (ed. Hood, L. E.) 57–108 (Plenum, New York, 1979).
53. Savageau, M. A. Alternative designs for a genetic switch: analysis of switching times using the piecewise power-law representation. *Math. Biosci.* **180**, 237–253 (2002).
54. Savageau, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **252**, 546–549 (1974).
55. Savageau, M. A. Significance of autogenously regulated and constitutive synthesis of regulatory proteins in repressible biosynthetic systems. *Nature* **258**, 208–214 (1975).
56. Becskei, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
Reports an experiment that showed that fluctuations in concentrations of mRNA transcripts are decreased by introducing negative autoregulation into a synthetic elementary gene circuit in *E. coli*. This result is consistent with the theoretical finding that stability is increased for negative autoregulation.
57. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**, 785–793 (2002).
Reports an experiment that used synthetic elementary gene circuits in *E. coli*, showing that the response time of a negatively autoregulated circuit is smaller than the response time of a similar circuit without TF self-regulation. This result is consistent with the theoretical finding that responsiveness is increased for negative autoregulation.
58. Salgado, H. *et al.* RegulonDB (version 3.2): transcriptional regulation and operon organization in *Escherichia coli* K-12. *Nucleic Acids Res.* **29**, 72–74 (2001).
Describes a well-known public database of transcriptional regulatory interactions in *E. coli*. The database does not include information on the influence of signals.
59. Thieffry, D., Huerta, A. M., Pérez-Rueda, E. & Collado-Vides, J. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* **20**, 433–440 (1998).
60. Hirakawa, H., Nishino, K., Hirata, T. & Yamaguchi, A. Comprehensive studies of drug resistance mediated by overexpression of response regulators of two-component signal transduction systems in *Escherichia coli*. *J. Bacteriol.* **185**, 1851–1856 (2003).
61. Martin, R. G., Jair, K. W., Wolf, R. E. Jr & Rosner, J. L. Autoactivation of the *marRAB* multiple antibiotic resistance operon by the MarA transcriptional activator in *Escherichia coli*. *J. Bacteriol.* **178**, 2216–2223 (1996).
62. Bausch, C. *et al.* Sequence analysis of the GntII (subsidiary) system for gluconate metabolism reveals a novel pathway for L-idonic acid catabolism in *Escherichia coli*. *J. Bacteriol.* **180**, 3704–3710 (1998).
63. Egan, S. M. & Schleif, R. F. A regulatory cascade in the induction of *rhaBAD*. *J. Mol. Biol.* **234**, 87–98 (1993).
64. Via, P., Badia, J., Baldoma, L., Obradors, N. & Aguilar, J. Transcriptional regulation of the *Escherichia coli rhaT* gene. *Microbiology* **142**, 1833–1840 (1996).
65. Song, S. & Park, C. Organization and regulation of the D-xylose operons in *Escherichia coli K-12*: XylR acts as a transcriptional activator. *J. Bacteriol.* **179**, 7025–7032 (1997).
66. Roof, D. M. & Roth, J. R. Autogenous regulation of ethanolamine utilization by a transcriptional activator of the *eut* operon in *Salmonella typhimurium*. *J. Bacteriol.* **174**, 6634–6643 (1992).
67. Azakami, H., Sugino, H., Yokoro, N., Iwata, N. & Murooka, Y. *moaR*, a gene that encodes a positive regulator of the monoamine regulon in *Klebsiella aerogenes*. *J. Bacteriol.* **175**, 6287–6292 (1993).
68. Jeter, R. M. Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium*. *J. Gen. Microbiol.* **136**, 887–896 (1990).
69. Bobik, T. A., Ailion, M. & Roth, J. R. A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation. *J. Bacteriol.* **174**, 2253–2266 (1992).
70. Ailion, M., Bobik, T. A. & Roth, J. R. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*. *J. Bacteriol.* **175**, 7200–7208 (1993).
71. Barkai, N. & Leibler, S. Circadian clocks limited by noise. *Nature* **403**, 267–268 (2000).
72. Setty, Y., Mayo, A. E., Surette, M. G. & Alon, U. Detailed map of a cis-regulatory input function. *Proc. Natl Acad. Sci. USA* **100**, 7702–7707 (2003).
73. Sadler, J. R. & Novick, A. The properties of repressor and the kinetics of its action. *J. Mol. Biol.* **12**, 305–327 (1965).
74. Savageau, M. A. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* **229**, 542–544 (1971).
75. Alon, U., Surette, M. G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).
Reports an experiment that showed that the tumbling frequency in *E. coli* chemotaxis is robust in response to changes in the level of the signalling protein CheR, supporting the importance of robustness as a performance criterion for naturally occurring biochemical systems.
76. Little, J. W., Shepley, D. P. & Wert, D. W. Robustness of a gene regulatory circuit. *EMBO J.* **18**, 4299–4307 (1999).
Reports an experiment that showed that phage λ decision circuitry is robust in response to changes in transcriptional control, supporting the importance of robustness as a performance criterion for naturally occurring gene circuits.
77. Eldar, A. *et al.* Robustness of the BMP morphogen gradient in *Drosophila* embryonic patterning. *Nature* **419**, 304–308 (2002).
Describes an experiment that supports the importance of robustness in eukaryotic development. The bone morphogenic protein gradient in *Drosophila* was found to be robust in response to both temperature changes and to heterozygous mutations in important genes.
78. Løbner-Olesen, A. Distribution of minichromosomes in individual *Escherichia coli cells*: implications for replication control. *EMBO J.* **18**, 1712–1721 (1999).
79. Paulsson, J. & Ehrenberg, M. Noise in a minimal regulatory network: plasmid copy number control. *Q. Rev. Biophys.* **34**, 1–59 (2001).
80. Thattai, M. & van Oudenaarden, A. Intrinsic noise in gene regulatory networks. *Proc. Natl Acad. Sci. USA* **98**, 8614–8619 (2001).
81. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nature Genet.* **31**, 69–73 (2002).
82. Swain, P. S., Elowitz, M. B. & Siggia, E. D. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl Acad. Sci. USA* **99**, 12795–12800 (2002).
83. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
84. Blake, W. J., KAErn, M., Cantor, C. R. & Collins, J. J. Noise in eukaryotic gene expression. *Nature* **422**, 633–637 (2003).
85. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237 (2002).
86. Vilar, J. M., Kueh, H. Y., Barkai, N. & Leibler, S. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl Acad. Sci. USA* **99**, 5988–5992 (2002).
87. Wolf, D. M. & Arkin, A. P. Fifteen minutes of fim: control of type 1 pili expression in *E. coli*. *Omics* **6**, 91–114 (2002).
88. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
89. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
90. Doyle, M. E., Brown, C., Hogg, R. W. & Helling, R. B. Induction of the *ara* operon of *Escherichia coli* B-r. *J. Bacteriol.* **110**, 56–65 (1972).
91. Casadaban, M. J. Regulation of the regulatory gene for the arabinose pathway, *araC*. *J. Mol. Biol.* **104**, 557–566 (1976).
92. Hahn, S. & Schleif, R. *In vivo* regulation of the *Escherichia coli araC* promoter. *J. Bacteriol.* **155**, 593–600 (1983).
93. Eldar, A., Rosin, D., Shilo, B. Z. & Barkai, N. Self-enhanced ligand degradation underlies robustness of morphogen gradients. *Dev. Cell* **5**, 635–646 (2003).
94. Rosenfeld, N. & Alon, U. Response delays and the structure of transcription networks. *J. Mol. Biol.* **329**, 645–654 (2003).

Acknowledgements

This work was supported by the National Institutes of Health and by the Department of Energy. Many thanks to the experimentalists who responded to various questions of ours about the gene circuits they have studied.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

EcoCyc: http://ecocyc.org

araC-araBAD | *cynR-cynTSX* | *dsdC-dsdXA* | *lacI-lacZYA* | *modEF-modABCD* | *treR-treBC* | *trpR-trpLEDCBA* | *tyrR-(aroF-tyrA)*

FURTHER INFORMATION

EcoGene: http://bmb.med.miami.edu/EcoGene/EcoWeb

EcoTFs: http://EcoTFs.lanl.gov

RegulonDB:
http://www.cifn.unam.mx/Computational_Genomics/regulondb

Uri Alon's web site: http://www.weizmann.ac.il/mcb/UriAlon

Access to this interactive links box is free online.