Mathematical Description of Gene Regulatory Units

Reiko J. Tanaka, Hiroyuki Okano, and Hidenori Kimura
Bio-Mimetic Control Research Center, RIKEN, Nagoya, Japan

ABSTRACT Revealing the control mechanisms responsible for the cell's surprisingly well-organized functions should lead directly to a better understanding of how the cell adapts to extraordinarily changing environments. A general framework for describing models that can represent diverse biochemical regulatory functions systematically would help not only systematic interpretation of the various models proposed for certain systems but also further understanding of the general control mechanism and design principles underlying different biological systems. This article presents a unified mathematical framework for describing gene regulatory units. The proposed framework is fairly compatible with the classical control theoretical framework, so it should serve as a connecting bridge between engineering control theory and biological control mechanisms. It should also provide a unified view of different regulatory units and facilitate systematic comparison of different mathematical models proposed in a variety of literature.

INTRODUCTION

Theoretical approaches to molecular biology, mainly based on modeling of biological control mechanisms, have been increasingly appreciated and pursued with the aim of systematically uncovering the design principles of molecular systems (1–3). Although modeling studies of various specific biological systems have successfully provided insights into the structural properties of those systems, they lack systematic ways of constructing models even for the same biological phenomenon. On the one hand, this reflects the fact that models serve as formal and precise extractions of intrinsic mechanisms responsible for extremely complex living cell functions (4) and that the differences in models themselves and in their constructions are a natural consequence of focusing on different characteristics of interest for each study. On the other hand, too much freedom given to modelers in the construction of models leads to difficulty in comparing different models proposed for the same system by different researchers and based on different assumptions, resulting in some confusion and misunderstanding in the interpretation and construction of models. A general framework for describing models that can represent diverse biochemical regulatory functions systematically would help not only systematic interpretation of the various models proposed for certain systems but also further understanding of the general control mechanism and design principles underlying different biological systems. This in turn would promote better cooperation between modelers and experimentalists.

The similarity between biological systems and engineering systems has been pointed out from the viewpoint of well-designed architectures that achieve robustness (5,6). An argument is that both types of systems use tremendously complex regulatory mechanisms that might not be necessary for basic system functions under normal conditions but are necessary for robust functioning against unpredictable and complex external and internal disturbances and for other design specifications, such as noise rejection and efficiency. We further argue here that a unique and intrinsic characteristic of control in biological systems, compared to manmade control, is found in its simultaneous and effective administration of several appropriate regulations to perform a cell task in response to complex and quite often compound environmental changes. We call this characteristic mode of biological control, ''compound control.'' It should be emphasized that the critical points of biological compound control lie in its enhanced flexibility. Different cell actions can be taken by the same control mechanism in response to compound environmental changes, and this characteristic can be considered robustness in the broad sense. The importance of control in molecular biology has been fully recognized by molecular biologists for a long time, ever since the pioneering works of Jacob and Monod (7). Revealing the control mechanisms responsible for the cell's surprisingly well-organized functions, some of which are the same as in engineering systems, should lead directly to a better understanding of how the cell adapts to extraordinarily changing environments (6) and, more generally, of the design principles of cellular systems, a primary interest of recent molecular biology.

Engineering control systems are a good example of advancement in systematic design and synthesis due to using a unified theoretical framework called ''control theory.'' It is thus natural to search for a unified theory for biological control. If biological control mechanisms can be understood in a similar way as engineering control systems, i.e., based on a general theory, application of the principles will improve understanding of biological systems (8). Several studies have already shed light on control theoretical interpretations of certain biological control systems and have provided important insights into their regulatory mechanisms (8–11).

Submitted January 25, 2006, and accepted for publication May 15, 2006.

Address reprint requests to H. Kimura, Bio-Mimetic Control Research Center, RIKEN, Shimo-shidami, Moriyama-ku, Nagoya 463-0003 Japan. Tel.: 81-52-736-5861; Fax: 81-52-736-5862; E-mail: kimura@bmc.riken.jp.

0006-3495/06/08/1235/13 $2.00

doi: 10.1529/biophysj.106.081828

Such mathematical analysis of biological control in terms of control theory is, however, still in its infancy in terms of generality.

As a first step toward building a general theory of biological control systems, we propose a unified mathematical framework for describing gene regulatory units; a rich theory can then be built on this base to help in revealing fundamental design principles of biological control systems. The framework is proposed from the compound control viewpoint and is fairly compatible with the classical control theoretical framework. We thus believe that it will serve as a connecting bridge between engineering control theory and biological control mechanisms. It should also provide a unified view of different regulatory units and facilitate systematic comparison of different mathematical models proposed in a variety of literature.

This article is organized as follows. Compound Control Section describes the concept of compound control in more detail, providing the base for the notion of gene regulatory units we will formulate. Illustrative Example: Lactose Utilization Network Section uses a classic system, a lactose utilization network, as an illustrative example to provide a flavor of the general framework introduced in Mathematical Description of Regulatory Units Section. This general description of regulatory units, the main product of this article, is clarified with various examples in the last section.

COMPOUND CONTROL

Biological control has a characteristic feature that is common in different levels of living organisms (from cellular to neural), compared to manmade control: the simultaneous and effective administration of several appropriate regulations in response to complex and quite often compound environmental changes to perform a task. There is a correspondence, which we call a task, between environmental changes and the actions required to achieve a behavior, which is a unit of activities with a definite purpose evoked by environmental changes. Each action results from the successive administration of appropriate regulations. It is worth noting that basically all regulations in biological control result from successive occurrences of simple control actions, such as firings of neurons at neural level and protein-protein or protein-DNA interactions (PPIs or PDIs) at cellular level. Spatial and temporal combinations of firings in different neurons and of different PPIs or PDIs can result in a tremendous variety of regulations that correspond to compound environmental changes. Homogeneity of regulatory elements, i.e., firings of neurons and PPIs or PDIs, is also a characteristic feature of biological control. The combination of simple homogeneous regulatory mechanisms adaptively creates surprisingly elaborate, heterogeneous, and thus complex regulations corresponding to the compound environmental changes that biological systems experience. We call this characteristic mode of biological control, common to different levels of living organisms (from cellular to neural) and which achieves smooth and versatile adaptation to compound environmental changes, ''compound control.'' It is worth noting that the phrase ''compound control'' is used in brain science in a slightly different but essentially same context (12).

This article is concerned with compound control at the cellular level, whose good example can be found in the classic diauxie, manifested when cultures of *Escherichia coli* are grown in mixtures of two carbon sources, such as glucose and lactose (13). The diauxie is not apparent in the presence of only a single carbon source. In the presence of both glucose and lactose, the cells use glucose first as a carbon source and then use lactose when the glucose runs out, in principle. Transcription of *lac* operons coding for the enzymes necessary for lactose catabolism is repressed during the first phase when only glucose is used. The repression of *lac* transcription is thus precisely regulated based on the availability of both glucose and lactose, resulting in the efficient use of available carbon sources. In other words, the cellular control mechanism for *lac* transcription is so well-designed that it can complete the cell task at hand: namely, efficient catabolism of lactose, despite compound environmental changes, and the availability of both glucose and lactose, by simultaneous and effective administration of multiple PPI- and PDI-based regulations for transcriptions.

This article aims to describe a general framework for systematic description of a compound control scheme at the cellular level that will hopefully help in revealing the fundamental design principles of biological systems. The arguments above suggest that the key to revealing the complexity and uniqueness of biological control at the cellular level lies in understanding the PPI- and PDI-based regulatory actions in response to compound environmental changes. We will consider operons in particular as central physical entities responsible for compound control, since compound environmental changes are, in most cases, processed at the transcription level where the activity of most bacterial promoters is determined depending on multiple environmental cues, except for some rare cases where regulatory proteins can integrate multiple signals (14). Transcriptional regulations are frequently the main control mechanisms for realizing cell tasks by initiating intracellular biochemical processes in response to compound environmental changes. Given a cell task, defined as the correspondence between environmental changes and the required actions, we can specify an elementary building block, which we call a ''regulatory unit'' and for which we provide a mathematical description. We naturally define it as a task-oriented module from the compound-control viewpoint, although we do not claim its originality, since similar notions have been already proposed. Its inputs are the relevant environmental changes, and the operons situated at its core are responsible for the realization of the actions (see Mathematical Description of Regulatory Units Section for a general treatment of regulatory units).

ILLUSTRATIVE EXAMPLE: LACTOSE UTILIZATION NETWORK

Before proposing a general treatment of regulatory units, we provide in this section its flavor with an illustrative example using a classic system, the lactose utilization network (15).

Lac regulatory unit

A regulatory unit is first defined from the compound-control viewpoint for the system of interest, and then the general descriptions are proposed. The central task of the lactose utilization network is efficient catabolism of lactose, a nutritious sugar that a cell finds in its environment. The task is accomplished through appropriate actions, that is, efficient transcription of the *lacZYA* operon precisely regulated by several PPI- or PDI-based regulatory actions, in response to changes in the availability of carbon sources, especially lactose and glucose in the environment. Catabolite repression in the presence of glucose in addition to lactose is a complex regulation apparent at the transcription level of the *lacZYA* operon, encoding β-galactosidase for conversion of lactose into allolactose, lactose permease for uptake of external lactose, and acetyltransferase for sugar metabolism.

Fig. 1 shows our proposed lac regulatory unit for handling compound environmental changes in the availability of external lactose and external glucose. The *lacZYA* operon is placed in the core of the unit (the *hatched module* in Fig. 1). The operon can be considered a functional operator that determines the concentration of transcribed mRNA from the concentrations of the transcription factors, cAMP-CRP and Lac*I*, in this case. The final output of the regulatory units is the concentration of translated proteins from the mRNA (*three green arrows* in Fig. 1). Although we are interested in the dynamic change in the mRNA concentration of the core operon (*lacZYA*, in this example) or its resulting proteins with the environmental changes as inputs (*two blue arrows* in Fig. 1), there is an additional step between the external changes (inputs to regulatory units) and concentrations of the transcription factors (inputs to the operon). The external environmental changes are transmitted through the cell by the concentrations of second messengers, which in turn affect the concentrations of transcriptional factors. In other words, transcription factors couple the expression of the target gene to environmental signals, and they are regulated by second messengers. In this example, the second messengers are

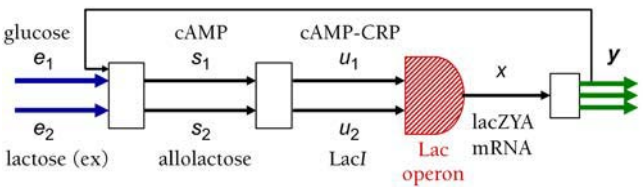

FIGURE 1 Lac regulatory unit of lactose utilization network.

cAMP for cAMP-CRP and allolactose for Lac*I*. Repressor Lac*I* is inactivated when it is bound by four allolactose molecules. The concentrations of second messengers are determined by the availabilities of external glucose and lactose and also by the lactose permease produced from LacY (the *long feedback arrow* in Fig. 1). To be precise, the concentration of cAMP rises sharply as the glucose is depleted and then drops rapidly to almost the initial level when the cell begins to use lactose (16).

General description of Lac regulatory unit

The dynamics of the Lac regulatory unit defined above are described as

$$\dot{x} = F(\boldsymbol{u}) - \alpha x, \tag{1}$$

$$\dot{\boldsymbol{y}} = G(x) - \beta \boldsymbol{y}, \tag{2}$$

$$\dot{\boldsymbol{u}} = H(\boldsymbol{s}) - \gamma \boldsymbol{u}, \tag{3}$$

$$\dot{\boldsymbol{s}} = K(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{y}) - \delta \boldsymbol{s}, \tag{4}$$

where x, $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$, $\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$, $\boldsymbol{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$, and $\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ are the concentrations of *lacZYA* mRNA, produced proteins (y_1 for β-galactosidase, y_2 for lactose permease, and y_3 for acetyltransferase), external carbon sources (e_1 for glucose and e_2 for lactose), second messengers (s_1 for cAMP and s_2 for allolactose), and regulator molecules (u_1 for cAMP-CRP and u_2 for Lac*I*), respectively, and α, $\beta(= \mathrm{diag}\{\beta_1, \beta_2, \beta_3\})$, $\gamma(= \mathrm{diag}\{\gamma_1, \gamma_2\})$, and $\delta(= \mathrm{diag}\{\delta_1, \delta_2\})$ represent the degradation rates (together with the growth rate) for $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{s}$, respectively. Functions F, G, H, and K describe the production rates of $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{s}$, respectively, and are rational functions of the arguments. Note that they explicitly depend on $\boldsymbol{u}$, $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{e}$, respectively, as is obvious from the signaling flows in Fig. 1 and K depends also on $\boldsymbol{s}$ itself and on $\boldsymbol{y}$ corresponding to the feedback regulation.

In this system, F satisfies $(\partial F/\partial u_1) > 0$ and $(\partial F/\partial u_2) < 0$, since cAMP-CRP ($u_1$) acts as an activator and Lac*I* (u_2) acts as a repressor. As a whole system, function F is implicitly affected by both e_1 and e_2, suggesting that changes in the two carbon sources' concentrations (e_1 and e_2) are compound environmental changes. The choice of functions F, G, H, and K together with the choice of variables to be included in the model leads to different models for the same system, as will be shown in the next subsection. The general description of the *lac* regulatory unit in Eqs. 1–4 provides a unified view for different models, enabling them to be compared.

Comparison of different models

Several mathematical models for the lactose utilization network have been proposed. Here we focus on four of them and clarify their focus in terms of the regulatory unit shown

in Fig. 1 by specifying their variables, loops, and detailed expressions for F, G, H, and K. This enables us to illustrate the generality of the regulatory unit and its description, Eqs. 1–4, proposed above.

Setty et al. model

Setty et al. (17) proposed a dynamic model of the *lac* operon as a functional unit with inputs of cAMP (s_1) and allolactose (or IPTG, s_2). It does not consider the external environmental changes or the feedback regulation via lactose permease (y_2). Fig. 2 shows the regulatory unit in their model, with the parts not considered shown as shaded dotted lines.

The dynamics of their model are described as

$$\dot{x} = F(\boldsymbol{u}) - \alpha x, \tag{5}$$

$$\dot{\boldsymbol{u}} = 0, \tag{6}$$

corresponding to Eq. 1 and Eq. 3, respectively, while Eq. 2 and Eq. 4 are not considered, as shown in Fig. 2. Function F in Eq. 5 is given as a rational function of u_1 and u_2 (details omitted here), and the static relationship corresponding to Eq. 6 is

$$\boldsymbol{u} = H_s(\boldsymbol{u}) = \begin{bmatrix} \dfrac{s_1^2}{1+s_1^2} \\ \dfrac{1}{1+s_2^4} \end{bmatrix}, \tag{7}$$

where the subscript s in H_s stands for ''steady state.''

Yildirim and Mackey model

Yildirim and Mackey (18) proposed a detailed model for a lactose utilization network with lactose as the only carbon source. It takes into account the time delay due to the transcription and translation processes and feedback regulation via lactose permease. Fig. 3 shows the regulatory unit in their model, with the parts not considered shown as shaded dotted lines. In addition to the variables introduced in our proposed regulatory unit, it considers the concentration of intracellular lactose (L), although it does not appear explicitly in the general description of the regulatory unit. External glucose (e_1), cAMP (s_1), and cAMP-CRP (u_1) are not considered.

The dynamics of their model are described as

$$\dot{x} = F(u_2) - \alpha x, \tag{8}$$

$$\dot{\boldsymbol{y}} = G(x) - \beta \boldsymbol{y}, \tag{9}$$

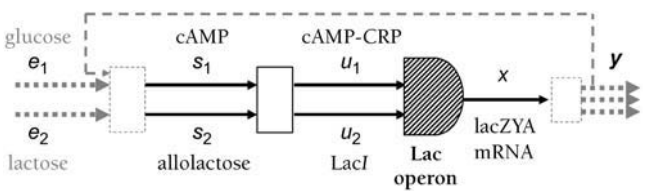

FIGURE 2 Setty et al.'s model of lactose utilization network (17).

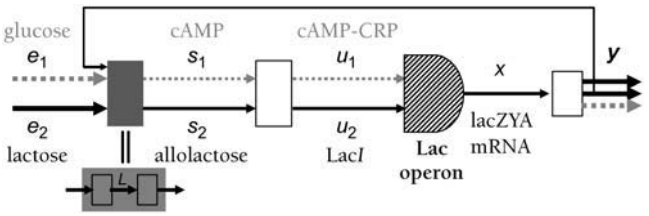

FIGURE 3 Yildirim and Mackey's model of lactose utilization network (18).

$$\dot{u}_2 = 0, \tag{10}$$

$$\dot{\boldsymbol{z}} = K(e_2, z, \boldsymbol{y}) - \delta \boldsymbol{z}, \quad s_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \boldsymbol{z}, \tag{11}$$

where $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Note that Eq. 10 leads to a static relationship, $u_2 = H_s(s_2)$, similar to the previous example, and that Eq. 11 is an extended representation of Eq. 4 for an augmented variable, $\boldsymbol{z} = \begin{bmatrix} s_2 \\ L \end{bmatrix}$.

The detailed expressions for F, G, H_s, and K are

$$F(u_2) = \Gamma + a \frac{1}{1 + \Phi_1 u_2 (t - \tau_u)}, \tag{12}$$

$$G(x) = \begin{bmatrix} b_1 \exp^{-\mu \tau_{y_1}} x(t - \tau_{y_1}) \\ b_2 \exp^{-\mu(\tau_{y_1} + \tau_{y_2})} x(t - \tau_{y_1} - \tau_{y_2}) \end{bmatrix}, \tag{13}$$

$$H_s(s_2) = \frac{c}{1 + \Phi_2 s_2^4}, \tag{14}$$

$$K(e_2, z, \boldsymbol{y}) = \begin{bmatrix} \left(d \dfrac{L}{\Phi_3 + L} - e \dfrac{s_2}{\Phi_4 + s_2} \right) y_1 \\ \left(f \dfrac{e_2}{\Phi_5 + e_2} - g \dfrac{L}{\Phi_6 + L} \right) y_2 - h \dfrac{L}{\Phi_7 + L} y_1 \end{bmatrix}, \tag{15}$$

where a, b_1, b_2, c, d, e, f, g, and h are constants. The Γ, Φ_1, and τ_u in Eq. 12 denote the spontaneous rate of mRNA production, the equilibrium constant for the operator-repressor reactions, and the time delay required to produce the mRNA, respectively. The τ_{y_1} and τ_{y_2} values in Eq. 13 denote the time delay for the mRNA translation of β-galactosidase (y_1) and of lactose permease (y_2), respectively. The power 4 in Eq. 14 comes from the fact that the repressor Lac$I$ is deactivated by the tetramer of allolactose with Φ_2 being the equilibrium constant for the repressor-allolactose reaction. The terms in Eq. 15 denote, respectively, the β-galactosidase-mediated conversion of lactose into allolactose, the β-galactosidase-mediated allolactose loss due to conversion into glucose and galactose, the permease-facilitated transport of external lactose, the intracellular lactose loss due to extracellular fluid because of the reversible nature of the permease-mediated transport, and the β-galactosidase-mediated conversion of lactose to allolactose. The Φ_i-values ($i = 3, 4, 5, 6, 7$) are the corresponding equilibrium constants. The nonlinear dynamics of this model depend on the system parameters and are further analyzed elsewhere (19).

Ozbudak et al. model

Ozbudak et al. (20) investigated the mechanism creating bistability in the lactose utilization network using a simple mathematical model. Feedback regulation via produced lactose permease for lactose uptake and inducer exclusion by glucose is explicitly taken into account, although the translation, thus $\boldsymbol{y}$, is not (see Fig. 4).

The dynamics of their model are described as

$$\dot{x} = \frac{a}{1 + u_2/R_0} - \alpha x, \tag{16}$$

$$\dot{\boldsymbol{u}}_2 = 0, \tag{17}$$

$$\dot{\boldsymbol{s}}_2 = K_1(e_1)K_2(e_2)x - \delta s_2, \tag{18}$$

corresponding to Eqs. 1, 3, and 4, respectively, where a and R_0 are constants, and K_1 and K_2 are functions corresponding to catabolite repression and lactose uptake, respectively (details omitted here). Equation 17 leads to the static relationship $u_2 = (R_T/1+s_2^2)$ with constant R_T.

Santillán and Mackey model

Santillán and Mackey (21) proposed yet another detailed model of the lactose utilization network with all of the variables in the proposed regulatory unit (see Fig. 5).

The dynamics of their model are described as

$$\dot{x} = F(\boldsymbol{u}) - \alpha x, \tag{19}$$

$$\dot{\boldsymbol{y}} = G(x) - \beta \boldsymbol{y}, \tag{20}$$

$$\dot{\boldsymbol{u}} = 0, \tag{21}$$

$$\dot{\boldsymbol{s}} = K(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{y}) - \delta \boldsymbol{s}, \tag{22}$$

corresponding to Eqs. 1–4, respectively. Equation 21 leads to $\boldsymbol{u} = H(\boldsymbol{s}_f)$, where $\boldsymbol{s}_f (= H_f(\boldsymbol{s}))$ denotes the vector for the concentrations of free second messengers (subscript f stands for ''free''), i.e., free cAMP (not bound to CRP) and free allolactose (not bound to the repressor Lac*I*). The **s** denotes the total (free and bound) concentration of cAMP and allolactose. Detailed expressions for F, G, H, H_f, and K are omitted here. Function F is determined by the configuration of the binding sites on the DNA and by the binding energies of possible binding states of the *lac* operon.

MATHEMATICAL DESCRIPTION OF REGULATORY UNITS

Using the illustrative examples in the previous section, here we present our general framework for describing models of general regulatory units.

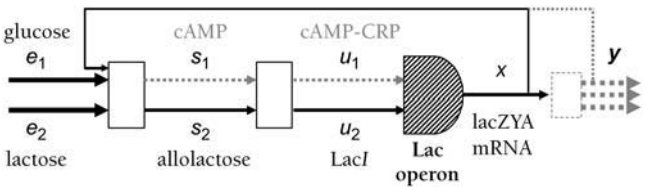

FIGURE 4 Ozbudak et al.'s model of lactose utilization network (20).

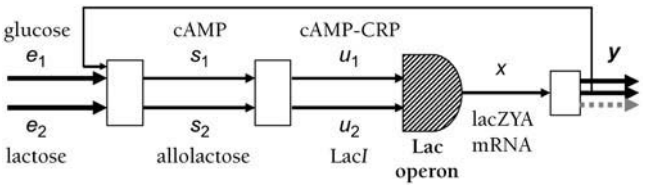

FIGURE 5 Santillán and Mackey's model of lactose utilization network (21).

Regulatory units

As we did for the lactose utilization network in the previous section, we first define a general regulatory unit for which mathematical descriptions are then proposed. A regulatory unit is defined as a task-oriented module from the compound control viewpoint. As described in Compound Control Section, a task in the compound control scheme is the correspondence between environmental changes and the actions required to realize a behavior, which is a unit of activities with a definite purpose evoked by environmental changes. For a given task, the relevant environmental changes and the actions required to perform the task are specified; a set of the main operons responsible for the actions is then naturally chosen (*lacZYA* for the lactose utilization network in Illustrative Example: Lactose Utilization Network Section). A regulatory unit is thus defined with the main operons in the core (Fig. 6 for the case with only one operon). The inputs and outputs of an operon are the concentrations of regulatory molecules (transcription factors) and of transcribed mRNA, respectively. The ultimate output of the regulatory unit is the concentration of translated proteins. Since transcription factors couple the expression of the target genes to environmental signals and they must be regulated, we consider an additional variable—the concentration of second messengers that regulate transcription factors. Second messengers include small ligands (such as allolactose in lac system) that regulate DNA-binding affinity of transcription factors, sensor kinases (such as NarX and NarQ) that regulate the activity of response regulators (such as NarL) by covalent modification as in two-component systems, and regulatory proteins that regulate effective concentration of a transcription factor by sequestration (14). The main flow of signals in the regulatory unit along these variables is shown by the thick arrows in Fig. 6. Additional regulatory relationships among the variables can be systematically defined to complete the regulatory unit. Fig. 6 shows the general scheme of a regulatory unit from the

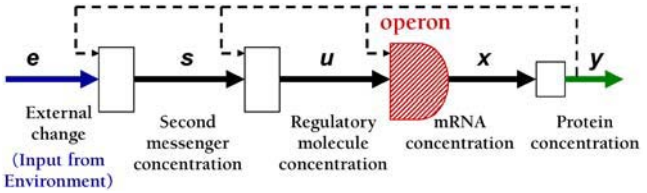

FIGURE 6 Regulatory unit from compound-control viewpoint.

compound control viewpoint with feedback loops (shown by *dotted arrows*) from the output as representative regulatory relationships.

A view of regulatory units from the control theoretical viewpoint suggests that the operon is a plant to be controlled and the rest of the regulatory unit is a controller, producing control inputs for the plant (Fig. 7). Fig. 8 shows a typical control system treated in control theory, with a feedforward controller and a feedback controller. This view provides a guide to carry out system identification of the total regulatory unit. A standard way of designing or identifying control systems begins with identification of the plant, the operon in this case, and then proceeds to identification or design of the controller, which is responsible for achieving the appropriate functioning of the regulatory unit in response to compound environmental changes. It should be noted that this classification of a plant and a controller is provided within the regulatory unit, which is simply a part of a more complex and larger biological system. Another way to classify the modules of a biological system into a plant and a controller is to consider the regulatory unit defined in this article as a controller and the physical phenomenogical part outside the regulatory unit as a plant (9,22).

Note that the notion of regulatory units is similar to that of gene circuits (23), which sense their environmental context and orchestrate the expression of a set of genes to produce appropriate patterns of cellular response. Similarly to regulatory units, a gene circuit is organized around transcription units, the simplest of which is an operon. Different ways of coordinating the expression of functions in a gene circuit are provided depending on the connectivity of transcription units, in a similar manner for regulatory units to achieve different tasks. Regulatory units put more emphasis than gene circuits do on the compound environmental changes related to a given task and on the general mathematical description that are compatible with classical control theory. Regulatory units are also similar but different from the classic notions of regulons, modulons, and stimulons (24) in the sense that the former are determined based on the operons responsible for the given task despite compound and thus (basically) several external changes, whereas the latter designate groups of operons controlled by a common regulator or responding to a given environmental change. Still, the notions of regulons, modulons, and stimulons can help define a regulatory unit for certain systems and, in some cases, they might help elucidate new actions, tasks, and connectivity that have not been considered.

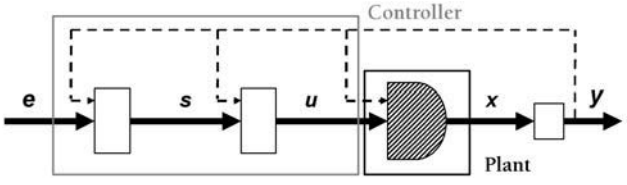

FIGURE 7 Regulatory unit can be decomposed into a plant and a controller, corresponding to the operon and the rest, respectively.

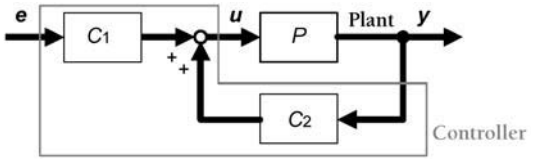

FIGURE 8 Typical control system used in control theory, with a plant (P), a feedforward controller (C_1), and a feedback controller (C_2). Inputs to and outputs from the plant are denoted as $\boldsymbol{u}$ and $\boldsymbol{y}$, respectively, and the external inputs from the environment are denoted as $\boldsymbol{e}$.

Mathematical description

The mathematical description of the general regulatory unit shown in Fig. 6 is

$$\dot{\boldsymbol{x}} = F(\boldsymbol{u},\boldsymbol{y}) - \alpha\boldsymbol{x}, \tag{23}$$

$$\dot{\boldsymbol{y}} = G(\boldsymbol{x},\boldsymbol{y}) - \beta\boldsymbol{y}, \tag{24}$$

$$\dot{\boldsymbol{u}} = H(\boldsymbol{s},\boldsymbol{y},\boldsymbol{u}) - \gamma\boldsymbol{u}, \tag{25}$$

$$\dot{\boldsymbol{s}} = K(\boldsymbol{e},\boldsymbol{y},\boldsymbol{s}) - \delta\boldsymbol{s}, \tag{26}$$

where $\boldsymbol{x} = [x_1 x_2 \ldots x_{n_x}]^T$, $\boldsymbol{y} = [y_1 y_2 \ldots y_{n_y}]^T$, $\boldsymbol{e} = [e_1 e_2 \ldots e_{n_e}]^T$, $\boldsymbol{s} = [s_1 s_2 \ldots s_{n_s}]^T$, and $\boldsymbol{u} = [u_1 u_2 \ldots u_{n_u}]^T$ are the concentrations of mRNA for the operons, produced proteins, external changes, second messengers, and regulator molecules, respectively, and $\alpha(= \mathrm{diag}\{\alpha_1, \ldots, \alpha_{n_x}\})$, $\beta(= \mathrm{diag}\{\beta_1, \ldots, \beta_{n_y}\})$, $\gamma(= \mathrm{diag}\{\gamma_1, \ldots, \gamma_{n_u}\})$, and $\delta(= \mathrm{diag}\{\delta_1, \ldots, \delta_{n_s}\})$ represent the degradation rates (together with the growth rate) for $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{s}$, respectively. Functions F, G, H, and K describe the production rates of $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{s}$, respectively, and are rational functions of the arguments. Note that F, G, H, and K explicitly depend on $\boldsymbol{u}$, $\boldsymbol{x}$, $\boldsymbol{s}$, and $\boldsymbol{e}$, respectively, as is obvious from the main signaling flow, and on the output $\boldsymbol{y}$ via feedback regulation. Functions H and K generally depend additionally on $\boldsymbol{u}$ and $\boldsymbol{s}$ themselves, respectively, reflecting the fact that the production rates of $\boldsymbol{u}$ and $\boldsymbol{s}$ are affected by their own concentrations (see, for example, the model of the tryptophan metabolic pathway in Tryptophan Metabolic System Section).

In this general description, a priori knowledge of whether the regulators are acting as activators or repressors is not necessary; however, it can be specified by the sign of the partial derivative of F: u_i is a repressor if $(\partial F/\partial u_i)<0$ and is an activator if $(\partial F/\partial u_i)>0$. This formulation of the regulators without a priori knowledge of their activity is advantageous for a unified treatment of regulatory units. It is consistent with the fact that regulation by transcriptional factors is initiated by their interactions with the binding sites on DNA irrespective of their activities and that the transcription rate is determined by the configuration of binding sites on DNA and by the binding energies of the possible binding states (H. Kimura, H. Okano, and R. J. Tanaka,

unpublished data), although the resulting actions that affect the functioning of RNAP may completely differ between activators and repressors. The fact that $\{e_i\}_{i=1}^{n_e}$ represents compound environmental changes can be verified by determining whether F depends on all of $\{e_i\}_{i=1}^{n_e}$.

In the example in Yildirim and Mackey Model (Fig. 3), the internal state (L) between the environmental change (e_2) and the second messenger (s_2) is considered with its dynamics, and an augmented variable, z, including the internal state is introduced in the model, as in Eq. 11. Although it is possible to augment the variables to include as much of the internal dynamics of the controller as we want, the variables shown in Eqs. 23–26 (Fig. 6) are minimally necessary for describing the regulatory unit from the compound-control viewpoint. The only exception occurs when other operons should be included in the controller part, as shown in Fig. 9. Assume the additional operon in the controller part (the *dotted black module* in Fig. 9) has input $\boldsymbol{u}_c$ and output $\boldsymbol{x}_c$. The dynamics of this part of the controller can then be given as

$$\dot{\boldsymbol{u}}_c = \eta(\boldsymbol{s}, \boldsymbol{u}, \boldsymbol{u}_c), \tag{27}$$

$$\dot{\boldsymbol{x}}_c = \xi(\boldsymbol{u}_c, \boldsymbol{u}, \boldsymbol{x}_c), \tag{28}$$

$$\dot{\boldsymbol{u}} = \phi(\boldsymbol{x}_c, \boldsymbol{u}). \tag{29}$$

Introducing augmented state $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{u}_c \\ \boldsymbol{x}_c \\ \boldsymbol{u} \end{bmatrix}$, we can replace Eq. 25 in the general description with

$$\dot{\boldsymbol{z}} = H(\boldsymbol{s}, \boldsymbol{y}, \boldsymbol{z}) - \gamma \boldsymbol{z}, \tag{30}$$

$$\boldsymbol{u} = L\boldsymbol{z}, \quad L = \mathrm{diag}\{O, O, I\}. \tag{31}$$

EXAMPLES

To illustrate the generality of the mathematical description we proposed in the previous section, here we provide more examples of regulatory units, in addition to that for the lactose utilization network described above.

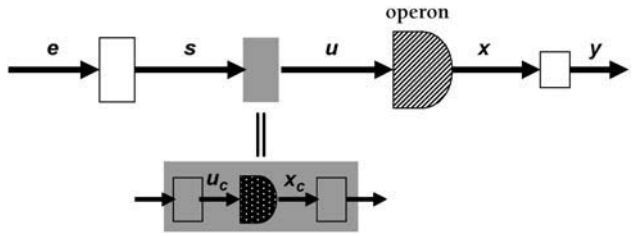

FIGURE 9 Regulatory unit with the dynamic controller.

Arabinose utilization network

The task of an arabinose utilization network is arabinose catabolism, and is achieved by efficient transcription of the *araBAD* operon, the required actions in response to changes in the availability of carbon sources, especially glucose and arabinose in the environment. As in the classic lactose-glucose systems, an arabinose utilization network shows diauxie in the presence of both arabinose and glucose because of catabolite repression by glucose (13).

The regulatory unit of the arabinose utilization network is shown in Fig. 10. The external availabilities of glucose and arabinose are the environmental changes and they affect the concentrations of second messengers, cAMP and internal arabinose. The regulators for the main operon, *araBAD*, are cAMP-CRP, AraC, and AraC-arabinose. AraC by itself acts as a repressor, whereas AraC-arabinose acts as an activator (25). The *araBAD* operon is responsible for the catabolism of arabinose, and arabinose uptake requires transcription of two additional operons, *araE* and *araFGH*. We include only *araFGH* in the scheme for simplicity and omit *araE*. Regulator AraC is produced by an *araC* operon, which is regulated by the same set of regulators as for *araBAD* and *araFGH*, as shown in the shaded box in Fig. 10. The lac operon considered in Illustrative Example: Lactose Utilization Network Section is a special case in which an operon provides the proteins necessary for the catabolism as well as the uptake of lactose, and its repressor Lac*I* is produced constitutively so that there is no need to include the *lacI* operon in the regulatory unit.

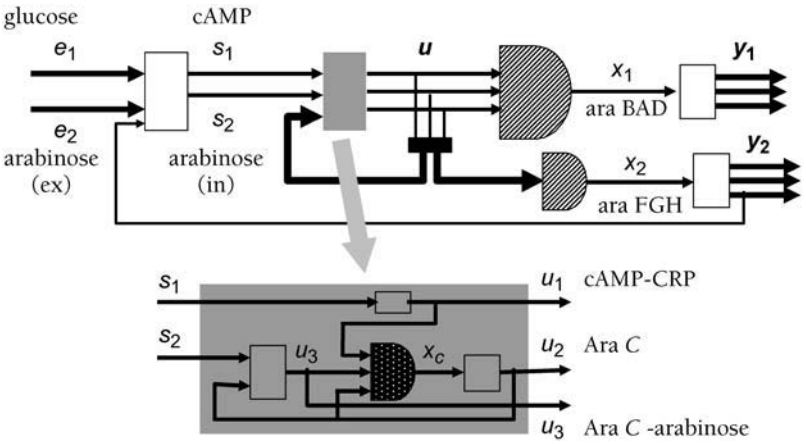

FIGURE 10 Regulatory unit of the arabinose utilization network; x_c is internal state of dynamic controller.

The dynamics of the regulatory unit are given as

$$\dot{\boldsymbol{x}} = F(\boldsymbol{u}) - \alpha \boldsymbol{x}, \tag{32}$$

$$\dot{\boldsymbol{y}} = G(\boldsymbol{x}) - \beta \boldsymbol{y}, \tag{33}$$

$$\dot{\boldsymbol{z}} = H(\boldsymbol{s}, z) - \gamma \boldsymbol{z}, \quad \boldsymbol{u} = L\boldsymbol{z}, \tag{34}$$

$$\dot{\boldsymbol{s}} = K(\boldsymbol{e}, \boldsymbol{s}, \boldsymbol{y}) - \delta \boldsymbol{s}, \tag{35}$$

where $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$, $\boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$, $\boldsymbol{s} = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$, and $\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ are the concentrations of mRNA (x_1 for *araBAD* and x_2 for *araFGH*), produced proteins ($\boldsymbol{y}_1$ for AraB, A, and D and $\boldsymbol{y}_2$ for AraF, G, and H), external carbon sources (e_1 for glucose and e_2 for external arabinose), second messengers (s_1 for cAMP and s_2 for internal arabinose), and regulator molecules (u_1 for cAMP-CRP, u_2 for AraC, and u_3 for AraC-arabinose), respectively; and $\boldsymbol{z} = \begin{bmatrix} x_c \\ \boldsymbol{u} \end{bmatrix}$ is the augmented variable with x_c being the concentration of *araC* mRNA, $L = \mathrm{diag}\{0, 1, 1, 1\}$, and $\alpha(= \mathrm{diag}\{\alpha_1, \alpha_2\})$, $\beta(= \mathrm{diag}\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2\})$, $\gamma(= \mathrm{diag}\{\gamma_1, \boldsymbol{\gamma}_2\})$, and $\delta(= \mathrm{diag}\{\delta_1, \delta_2\})$ are the degradation rates (together with the growth rate) for $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, and $\boldsymbol{s}$, respectively.

Tryptophan metabolic system

The tryptophan metabolic system is another classic example of gene regulation systems and has been extensively studied both experimentally and theoretically (26). The task is the efficient supply of tryptophan, one of the amino acids that the cell requires; it is achieved by efficient transcription of the *trpEDCBA* operon, which encodes the genes for the enzymes necessary in the tryptophan biosynthesis pathway in response to the changes in availability of tryptophan in the cell environment.

The regulatory unit for this system can be described as shown in Fig. 11 for the *trpEDCBA* operon. The regulators for the operon are charged tRNA and active repressor TrpR-tryptophan, which are responsible for transcription attenuation and repression, respectively. In addition to these two mechanisms, this system can use enzyme inhibition for feedback regulation. A produced enzyme, anthranilate synthase, which is the first enzyme to catalyze a reaction in the tryptophan biosynthesis pathway beginning with chorismate, is a heterotetramer consisting of two TrpE and two TrpD polypeptides. The enzymatic activity of anthranilate synthase is inhibited when the TrpE subunits of an anthranilate are bound by two tryptophan molecules. A change in the external concentration of tryptophan in the environment is the only environmental change that affects the actions we consider, and the second messenger is internal tryptophan. We assume here that tRNA is constantly available for charged tRNA and omit its production mechanism, whereas the production of TrpR requires a *trpR* operon, which is also regulated by TrpR-tryptophan and is shown in the shaded box in Fig. 11.

The dynamics of the regulatory unit in Fig. 11 are described as

$$\dot{x} = F(\boldsymbol{u}) - \alpha x, \tag{36}$$

$$\dot{\boldsymbol{y}} = G(x) - \beta \boldsymbol{y}, \tag{37}$$

$$\dot{\boldsymbol{z}} = H(s, \boldsymbol{z}) - \gamma \boldsymbol{z}, \quad \boldsymbol{u} = L\boldsymbol{z}, \tag{38}$$

$$\dot{s} = K(e, \boldsymbol{y}, s) - \delta s, \tag{39}$$

where x, $\boldsymbol{y}(= [y_1 y_2 y_3 y_4 y_5]^{\mathrm{T}})$, e, s, and $\boldsymbol{u}(= [u_1 u_2]^{\mathrm{T}})$ are the concentrations of mRNA, produced proteins (y_1, y_2, y_3, y_4, y_5 are for TrpA,B,C,D,E, respectively), external tryptophan, internal tryptophan, and regulatory molecules (u_1 for charged tRNA and u_2 for TrpR-tryptophan), respectively, $\boldsymbol{z} = \begin{bmatrix} x_c \\ \boldsymbol{u} \end{bmatrix}$ is the augmented variable with x_c being the concentration of mRNA for inactive transcription factor TrpR, $L = \mathrm{diag}\{0, 1, 1\}$, and α, $\beta(= \mathrm{diag}\{\beta_1, \beta_2, \beta_3\})$, $\gamma(= \mathrm{diag}\{\gamma_1, \boldsymbol{\gamma}_2\})$, and δ are the degradation rates (together with the growth rate) for x, $\boldsymbol{y}$, $\boldsymbol{z}$, and s, respectively.

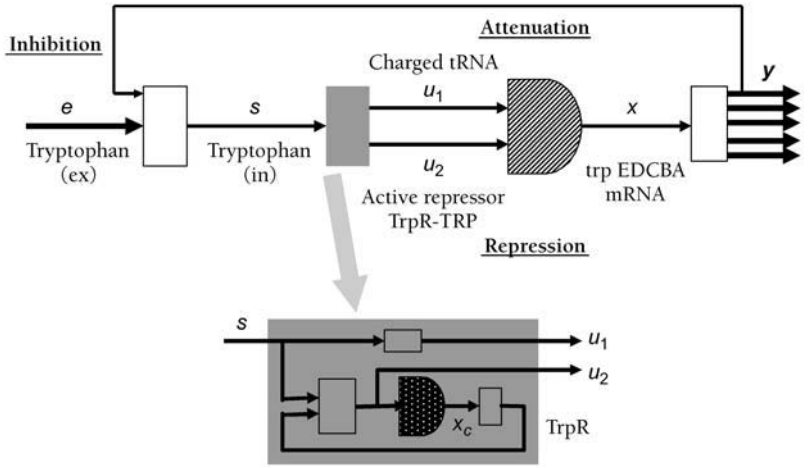

FIGURE 11 Regulatory unit of the tryptophan metabolic system.

A number of theoretical models for the tryptophan metabolic system have been used to study different system characteristics. However, they have considered neither the gene regulation for TrpR nor the external input e of tryptophan outside the cell. As a result, the general description used for the tryptophan system does not need augmented variable z. For example, Sinha (27) proposed a simple mathematical model with transcription repression as the only regulation mechanism (ignoring transcription attenuation and enzyme inhibition) (Fig. 12). Bliss et al. (28) proposed a model including both transcription repression and enzyme inhibition but not transcription attenuation (Fig. 13). They also considered time delay for the translation. It is easy to verify that both models can be interpreted in the general framework given by Eqs. 23–25.

Santillán and Zeron (29) recently proposed a detailed model of the tryptophan metabolic system with all three control mechanisms and the time delay due to translation. The related regulatory unit is shown in Fig. 14, and the dynamics are described as

$$\dot{x} = F(\boldsymbol{u}) - \alpha x, \tag{40}$$

$$\dot{y}_5 = G(x) - \beta y_5, \tag{41}$$

$$\dot{\boldsymbol{u}} = H(s_\mathrm{f}) - \gamma \boldsymbol{u}, \quad s_\mathrm{f} = H_\mathrm{f}(s, y_5), \tag{42}$$

$$\dot{\boldsymbol{s}} = K(s_\mathrm{f}, y_5) - \delta s, \tag{43}$$

where s_f is the free tryptophan (not bound to anthranilate). The detailed expressions for F, G, H, K, and H_f are

$$F(\boldsymbol{u}) = a g(u_1) h(u_2), \tag{44}$$

$$G(x) = b x(t - \tau), \tag{45}$$

$$H_1(s_\mathrm{f}) = \frac{s_\mathrm{f}}{K_G + s_\mathrm{f}} G_\mathrm{T}, \tag{46}$$

$$H_2(s_\mathrm{f}) = c\left(\frac{s_\mathrm{f}}{K_T + s_\mathrm{f}}\right)^2, \tag{47}$$

$$K(s_\mathrm{f}, y_5) = k y_5 \left(\frac{K_I}{s_\mathrm{f} + K_I}\right)^2 - \rho \frac{s_\mathrm{f}}{K + s_\mathrm{f}}, \tag{48}$$

$$H_\mathrm{f}(s, y_5) = \frac{1}{2}\sqrt{(K_I + 2y_5 - s)^2 + 4K_I s} - \frac{1}{2}(K_I + 2y_5 - s), \tag{49}$$

with $g(u_2) = (1+2\mathrm{u}_1/\mathrm{K_C})/((1+\mathrm{u}_1/\mathrm{K_c})^2)$ and $h(u_2) = (\mathrm{P}/\mathrm{K_P})/(1+\mathrm{P}/\mathrm{K_P}+\mathrm{u}_2/\mathrm{K_R})$, where a, b, c, K_G, K_T, K_I, K, K_C, K_P, P, and G_T are constants.

Although we did not include the transporter protein in the regulatory unit shown in Fig. 11, nor is it included in the three models shown above, it is used to take up tryptophan from outside the cell to inside. Tryptophan outside the cell is important from the compound-control viewpoint, so the transporter, the box from e to s in Fig. 11, should be taken into consideration in the regulatory unit. However, *mtr* gene encoding of the tryptophan-specific transporter is regulated not only by TrpR but also by TyrR-tyrosine, a compound of TyrR and another aromatic amino acid, tyrosine. We thus have to include the tyrosine biosynthesis pathway and the operon for TyrR with their regulators as well if we want to include the transporter proteins. Actually, there are several genes that are co-regulated by transcription factors related to three aromatic amino acids (tryptophan, tyrosine, and phenylalanine). They usually code for the proteins that are used in the common pathway for three aromatic amino acids. Consequently, it is reasonable to consider the whole aromatic amino-acid biosynthesis pathway, and not only the tryptophan biosynthesis pathway, when we discuss the regulatory unit from the compound-control viewpoint with external changes as inputs.

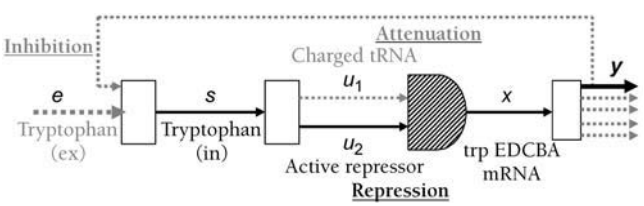

FIGURE 12 Sinha's model of the tryptophan metabolic system (27).

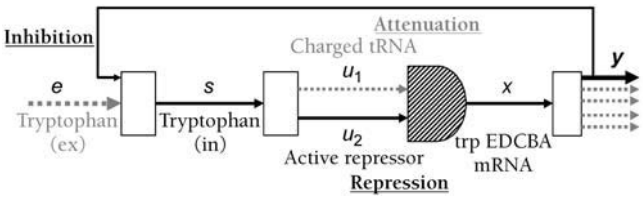

FIGURE 13 Bliss et al.'s model of the tryptophan metabolic system (28).

Heat shock response system

As an example of a more complex system, we consider systems for heat shock response (30). The task is to counter the effects of heat such as protein unfoldings and malfunctions. It is done by making heat shock proteins (molecular chaperones) in response to environmental changes in temperature. Based on a recent mathematical analysis of the heat shock response system (9), we consider a simple description of the system with DnaK/J as representative molecular chaperones and FtsH as a representative protease.

The regulatory unit for this simple heat shock response system is shown in Fig. 15. The main operons are *dnaK* for DnaK/J and *hflB* for FtsH. Transcription of both operons is

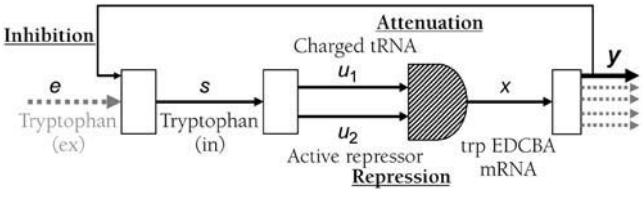

FIGURE 14 Santillán and Zeron's model of the tryptophan metabolic system (29).

promoted when σ_{32} factors are bound to RNAP, so that free σ_{32} factors (which are ready to bind to RNAP and promote transcription and are not bound to chaperones or proteases) act as regulatory molecules, whereas the total σ_{32} factors are considered as second messengers. The environmental changes are temperature shifts outside the cell. The system uses three different regulatory mechanisms: feedforward by which the temperature upshift promotes translation of the σ_{32} factors, sequestration feedback by which the σ_{32} factors are sequestered by chaperones and thus cannot bind to RNAP and activate the transcription of the operons, and degradation feedback by which the σ_{32} factors are degraded when bound to both DnaK and FtsH. All three regulatory mechanisms are shown in Fig. 15.

The dynamics of the regulatory unit shown in Fig. 15 are described as

$$\dot{\boldsymbol{x}} = F(\boldsymbol{u}) - \alpha\boldsymbol{x}, \tag{50}$$

$$\dot{\boldsymbol{y}} = G(\boldsymbol{x}) - \beta\boldsymbol{y}, \tag{51}$$

$$\dot{u} = H(s, \boldsymbol{y}, u) - \gamma u, \tag{52}$$

$$\dot{s} = K(e, \boldsymbol{y}, s) - \delta s, \tag{53}$$

where $\mathbf{x}(= [x_1 x_2]^{\mathrm{T}})$, $\boldsymbol{y}(= [y_1 y_2]^{\mathrm{T}})$, e, s, and u are the concentrations of mRNA (x_1 and x_2 are for *dnaK* and *hflB*, respectively), produced proteins (y_1 and y_2 are for DnaK and FtsH, respectively), external temperature, total σ_{32}, and free σ_{32}, respectively, and $\alpha(= \mathrm{diag}\{\alpha_1, \alpha_2\})$, $\beta(= \mathrm{diag}\{\beta_1, \beta_2\})$, γ, and δ are the degradation rates (together with the growth rate) for $\boldsymbol{x}$, $\boldsymbol{y}$, u, and s, respectively.

El-Samad et al. (9) proposed a simple mathematical model for the heat shock response system and used it to analyze how well the designed regulatory mechanisms fit the system requirements for robustness, response speed, noise rejection, efficiency, etc., by clarifying the role of each regulatory mechanism. The dynamics of their simpler model (reduced-order model), shown in Fig. 16, are described as

$$\dot{y}_1 = bu - \beta y_1, \quad y_2 = \alpha y_1, \tag{54}$$

$$\dot{u} = 0, \tag{55}$$

$$\dot{s} = K(e) - H_{\mathrm{s}}(s, \boldsymbol{y}) G_1(y_1) G_2(y_2) - \delta s, \tag{56}$$

corresponding to the general description, Eqs. 24–26. The y_1 and y_2 stand for the concentrations of DnaK and FtsH, respectively. State $\boldsymbol{x}$ does not explicitly appear in the model based on the assumption that $y_{\mathrm{i}} = a_{\mathrm{i}} x_{\mathrm{i}}$ ($i = 1, 2$), with constants a_{i}. The α is a constant, and β and δ represent the degradation rates for y_1 and s, respectively. A static relation, $u = H_{\mathrm{s}}(s, \boldsymbol{y})$, is given for Eq. 55. The $K(e)$ in Eq. 56 denotes feedforward regulation in response to temperature upshifts, and the second term represents degradation feedback by which free σ_{32} factors are degraded when they are bound to both DnaK (y_1) and FtsH (y_2), so that the resulting term is given by the product of free σ_{32} factors (H_{s}), free DnaK ($G_1(y_1)$), and free FtsH ($G_2(y_2)$).

λ-system

As the last example, we consider the bacteriophage λ-system, which has been studied since the 1950s (31) as a paradigm for developmental genetic networks (32). A λ-infected bacterium follows either of two pathways: lysis, in which λ actively produces viral copies using the bacterial molecular machinery, and lysogeny, in which all but one of the phage genes are turned off, and one phage chromosome integrates itself into the host chromosome so that the viral copies are produced passively as part of the host chromosome. The regulatory mechanism responsible for the lysogeny/lysis decision, the λ-switch, has been extensively studied, and its detailed mechanism has been revealed, although not yet completely (33–35).

The task of the λ-system is efficient and adaptive switching between lysogeny/lysis pathways in response to environmental changes such as in the nutritional state and in UV

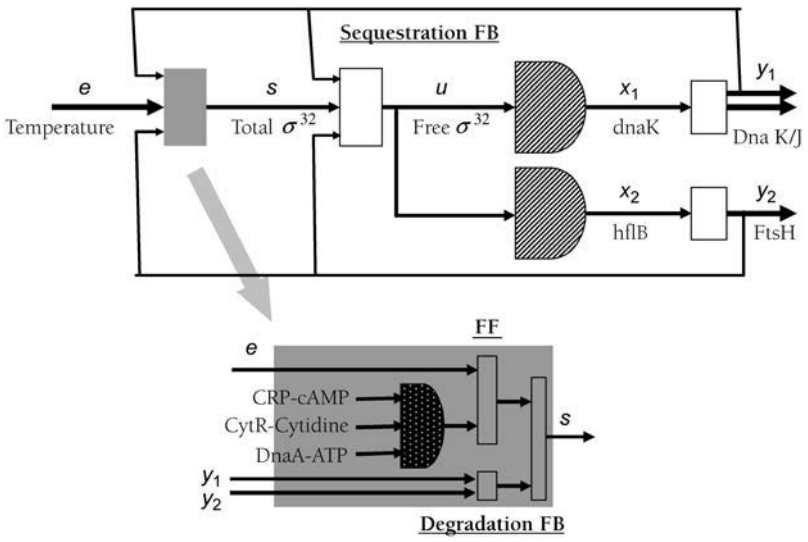

FIGURE 15 Regulatory unit for the heat shock response system.

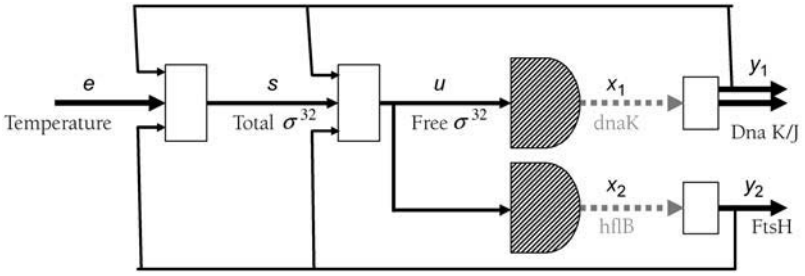

FIGURE 16 El-Samad et al.'s model of the heat shock response system (9).

irradiation. It is done by appropriate production of related proteins such as CI and Cro: CI, the λ-repressor, turns off all phage genes except its own gene and thus leads to stable lysogenic growth, whereas Cro promotes and is required for lytic growth. The corresponding mRNA for *cI* and *cro* are both transcribed from the right operator (O_R).

The regulatory unit of the λ-system is shown in Fig. 17, with O_R as the central entity responsible for the task. The regulators of O_R are CI_2 and Cro_2 dimers, whereas the outputs of the regulatory unit are CI and Cro monomers. We consider the nutritional state and UV irradiation as environmental changes that affect the regulatory molecules via the second messengers of CII and activated RecA proteins. CII is responsible for kickstarting CI production, and the nutritional state affects the production of CII. For example, the mean and peak of the CII concentration levels are reduced in well-fed cells, so there is less probability of CI production being kickstarted, resulting in the reduced probability of lysogeny. The RecA protein becomes activated when DNA is damaged, for example, by UV irradiation. The activated protein cleaves the CI monomers, which cannot dimerize anymore, and repressors fall off the operator. As a result, the rate of repressor synthesis drops and transcription of *cro* begins. Although there are more complex mechanisms for regulation of CII and RecA production, we consider a simplified scheme for the regulatory unit that omits other proteins such as N, Q, HflA, HflB, and CIII and terminators such as T_{L_1} and T_R in order to focus on the effects of environmental changes on the transcription of main mRNAs from the compound control viewpoint.

The dynamics of the regulatory unit shown in Fig. 17 are described as

$$\dot{\boldsymbol{x}} = F(\boldsymbol{u}) - \alpha\boldsymbol{x}, \tag{57}$$

$$\dot{\boldsymbol{y}} = G(\boldsymbol{x}) - \beta\boldsymbol{y}, \tag{58}$$

$$\dot{\boldsymbol{u}} = H(\boldsymbol{s},\boldsymbol{y},\boldsymbol{u}) - \gamma\boldsymbol{u}, \tag{59}$$

$$\dot{\boldsymbol{s}} = K(\boldsymbol{e},\boldsymbol{y},\boldsymbol{s}) - \delta\boldsymbol{s}, \tag{60}$$

where $\boldsymbol{x}(=[x_1x_2]^T)$, $\boldsymbol{y}(=[y_1y_2]^T)$, $\boldsymbol{e}(=[e_1e_2]^T)$, $\boldsymbol{s}(=[s_1s_2]^T)$, and $\boldsymbol{u}(=[u_1u_2]^T)$ are the concentrations of mRNA (x_1 and x_2 are for *cI* and *cro*, respectively), produced proteins (y_1 and y_2 are for CI and Cro monomers, respectively), environmental changes (e_1 and e_2 are for the nutritional state and UV irradiation, respectively), second messengers (s_1 and s_2 are for CII and RecA, respectively), transcription factors (u_1 and u_2 are for CI_2 and Cro_2 dimers, respectively), and $\alpha(=\mathrm{diag}\{\alpha_1, \alpha_2\})$, $\beta(=\mathrm{diag}\{\beta_1, \beta_2\})$, $\gamma(=\mathrm{diag}\{\gamma_1, \gamma_2\})$, and $\delta(=\mathrm{diag}\{\delta_1, \delta_2\})$ are degradation rates (together with the growth rate) for $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{u}$, and $\boldsymbol{s}$, respectively.

Based on the development of experimental studies of λ-systems, several theoretical models have been proposed, e.g., (34–37). While two of them (34,35) consider most of the detailed processes related to the λ switch, many others focus only on the transcriptional regulation of O_R. Santillán and Mackey (37) proposed a mathematical model of the λ-system and used it to investigate the stability of the lysogenic state. They focused on the dynamics of O_R, as shown in Fig. 18. The dynamics are described as

$$\dot{x} = F(\boldsymbol{u}) - \alpha\boldsymbol{x}, \tag{61}$$

$$\dot{\boldsymbol{y}} = G(\boldsymbol{x}) - \beta\boldsymbol{y}, \tag{62}$$

$$\dot{\boldsymbol{u}} = 0, \tag{63}$$

corresponding to Eqs. 57–59, whereas second messengers $\boldsymbol{s}$, and thus Eq. 60, are not taken into consideration. Function $F(\boldsymbol{u})$ for the transcription rate is calculated from the probability of different compatible binding states of O_R. The probability of each binding state is calculated under the assumption of thermodynamic equilibrium using the binding

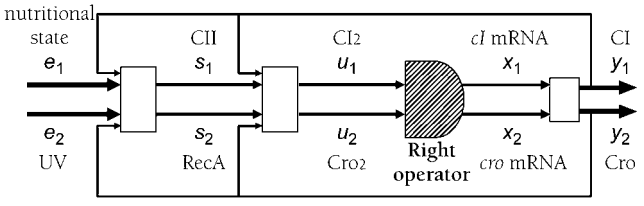

FIGURE 17 Regulatory unit of the λ-system.

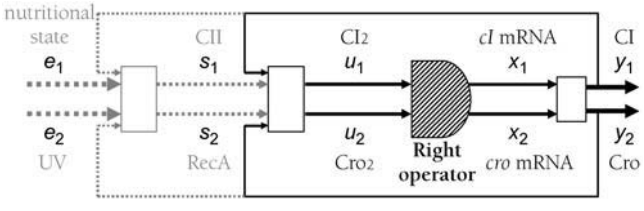

FIGURE 18 Santillán and Mackey's model of the λ-system (37).

energies of corresponding molecules (38). The translation rate $G(\boldsymbol{x})$ is given by

$$G(\boldsymbol{x}) = \begin{bmatrix} b_1 x_1(t-\tau_1) \\ b_2 x_2(t-\tau_2) \end{bmatrix}, \quad (64)$$

for y_1 and y_2, respectively, with τ_1 and τ_2 being the time delay required for translation to begin. The static relationship for $\boldsymbol{u}$ and $\boldsymbol{y}$ is derived for Eq. 63 under a quasi steady-state assumption for the dimerization reactions.

CONCLUSION

As a first step toward a general theory of biological control systems, we have described a unified mathematical framework for describing gene regulatory units. Gene regulatory units are defined as an elementary unit dealing with compound control schemes found in different biological systems and thus their main focuses are on the effects of environmental changes on the transcription of the main operons responsible for a cell's task. The proposed system description from the compound-control viewpoint is fairly compatible with that in the classical control theoretical framework and thus should work well as a bridge connecting engineering control theory and biological control mechanisms. We believe that a rich theory can be built on this framework. Control theory is a well-established scientific discipline that enables systematic analysis, identification, and synthesis of engineering control systems. Its important notions related to biological control include, for example, controllability and stability of the systems, optimal control, robust control, and so on.

We have demonstrated the generality of the proposed description through different specific examples. These examples show how the general description provides a unified view of different regulatory units and facilitates systematic comparison of different mathematical models proposed in various literature. We have also proposed a modeling scheme that is not included in the models found in the literature to clarify the intrinsic characteristic of biological control—the simultaneous and effective administration of several appropriate regulations in response to complex and quite often compound environmental changes.

A similar line of studies toward a unified systematic quantification of transcription regulation is found in the work of Bintu et al. (39,40). The objective is systematic derivation of function $F(\boldsymbol{u})$ for operon input-output relationships based on the configuration of binding sites, combinatorial patterns, and thermodynamics. This is consistent with our view of the regulatory units from the control theoretical viewpoint, suggesting that the operon is a plant to be controlled and thus is the first to be identified before proceeding to identification or design of the controller part, which is responsible for achieving appropriate functioning of the regulatory unit in response to compound environmental changes.

Once the plant of the regulatory unit, represented by $F(\boldsymbol{u})$, is systematically identified, it remains for us to investigate the controller part of the regulatory units to clarify the design principles of elaborately regulated cellular systems. The consistency of our general description of regulatory units with the classical control theoretic framework suggests that control theory might provide keys to revealing the fundamental mechanisms in biological control systems. The general description proposed in this article is initially concerned with continuous and deterministic models, as shown in the concrete examples. The matter of stochastic effects is another important subject to be considered.

REFERENCES

1. Bolouri, H., and E. H. Davidson. 2002. Modeling transcriptional regulatory networks. *Bioessays*. 24:1118–1129.
2. Endy, D., and R. Brent. 2001. Modelling cellular behavior. *Nature*. 409:391–395.
3. May, R. M. 2004. Uses and abuses of mathematics in biology. *Science*. 303:790–793.
4. Bower, J. M., and H. Bolouri, editors. 2001. Computational Modeling of Genetic and Biochemical Networks. The MIT Press, Cambridge, MA.
5. Csete, M. E., and J. C. Doyle. 2002. Reverse engineering of biological complexity. *Science*. 295:1664–1669.
6. Stelling, J., U. Sauer, Z. Szallasi, F. J. Doyle, and J. Doyle. 2004. Robustness of cellular functions. *Cell*. 118:675–685.
7. Jacob, F., and J. Monod. 1963. On the regulation of gene activity. *Cold Spring Harb. Symp. Quant. Biol*. 26:193–211.
8. Tomlin, C. J., and J. D. Axelrod. 2005. Understanding biology by reverse engineering the control. *Proc. Natl. Acad. Sci. USA*. 102: 4219–4220.
9. El-Samad, H., H. Kurata, J. C. Doyle, C. A. Gross, and M. Khammash. 2005. Surviving heat shock: control strategies for robustness and performance. *Proc. Natl. Acad. Sci. USA*. 102:2736–2741.
10. Yi, T.-M., Y. Huang, M. I. Simon, and J. Doyle. 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. USA*. 97:4649–4653.
11. Lauffenburger, D. A. 2000. Cell signaling pathways as control modules: complexity for simplicity? *Proc. Natl. Acad. Sci. USA*. 97: 5031–5033.
12. Ito, M. 2002. Controller-regulator model of the central nervous system. *J. Integr. Neurosci*. 1:129–143.
13. Monod, J. 1942. Research on the Growth of Bacterial Cultures. Hermann, Paris, France.
14. Browning, D. F., and S. J. W. Busby. 2004. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol*. 2:57–65.
15. Müller-Hill, B. 1996. The *lac* Operon. Walter de Bruyter & Company, Berlin, Germany.
16. Inada, T., K. Kimata, and H. Aiba. 1996. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes Cells*. 1:293–301.
17. Setty, Y., A. E. Mayo, M. G. Surette, and U. Alon. 2003. Detailed map of a *cis*-regulatory input function. *Proc. Natl. Acad. Sci. USA*. 100: 7702–7707.
18. Yildirim, N., and M. C. Mackey. 2003. Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophys. J*. 84:2841–2851.
19. Ahmadzadeh, A., A. Halasz, S. Prajna, A. Jadbabaie, and V. Kumar. 2005. Analysis of the lactose metabolism in *E. coli* using sum-of-squares decomposition. Proc. 44th IEEE Conf. on Decision and Control. 879–884.
20. Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature*. 427:737–740.

21. Santillán, M., and M. C. Mackey. 2004. Influence of catabolite repression and inducer exclusion on the bistable behavior of the *lac* operon. *Biophys. J.* 86:1282–1292.
22. Doyle, J., and M. Csete. 2005. Motifs, control, and stability. *PLoS Biol.* 3:e392.
23. Savageau, M. A. 2001. Design principles for elementary gene circuits: elements, methods, and examples. *Chaos.* 11:142–159.
24. Neidhardt, F. C., and M. A. Savageau. 1996. Regulation beyond the operon. *In Escherichia coli* and *Salmonella*, 2nd Ed. F.C. Neidhardt, editor-in-chief. ASM Press, Washington, DC.
25. Schleif, R. 2000. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.* 16:559–565.
26. Yanofsky, C. 2003. Using studies on tryptophan metabolism to answer basic biological questions. *J. Biol. Chem.* 28:10859–10878.
27. Sinha, S. 1988. Theoretical study of tryptophan operon: application in microbial technology. *Biotechnol. Bioeng.* 31:117–124.
28. Bliss, R. D., R. P. Painter, and A. G. Marr. 1982. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* 97:177–193.
29. Santillán, M., and E. S. Zeron. 2004. Dynamic influence of feedback enzyme inhibition and transcription attenuation on the tryptophan operon response to nutritional shifts. *J. Theor. Biol.* 231:287–298.
30. Craig, E. A. 1985. The heat shock response. *CRC Crit. Rev. Biochem.* 18:239–280.
31. Lwoff, A. 1953. Lysogeny. *Bacter. Rev.* 17:269–337.
32. Oppenheim, A. B., O. Kobiler, J. Stavans, D. L. Court, and S. Adhya. 2005. Switches in bacteriophage λ-development. *Annu. Rev. Genet.* 39:409–429.
33. Ptashne, M. 1992. A Genetic Switch. Cell & Blackwell Scientific, Cambridge, MA.
34. Arkin, A., J. Ross, and H. H. McAdams. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics.* 149:1633–1648.
35. McAdams, H. H., and L. Shapiro. 1995. Circuit simulation of genetic networks. *Science.* 269:650–656.
36. Reinitz, J., and J. R. Vaisnys. 1990. Theoretical and experimental analysis of the phage-λ genetic switch missing levels of co-operativity. *J. Theor. Biol.* 145:295–318.
37. Santillán, M., and M. C. Mackey. 2004. Why the lysogenic state of phage λ is so stable: a mathematical modeling approach. *Biophys. J.* 86:75–84.
38. Ackers, G. K., A. D. Johnson, and M. A. Sea. 1982. Quantitative model for gene regulation by λ-phage repressor. *Proc. Natl. Acad. Sci. USA.* 79:1129–1133.
39. Bintu, L., N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. 2005. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* 15:116–124.
40. Bintu, L., N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips. 2005. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.* 15:125–135.