Phenotypic deconstruction of gene circuitry

Jason G. Lomnitz¹ and Michael A. Savageau^{1,2,a)}

¹Department of Biomedical Engineering, University of California, Davis, California 95616, USA

²Microbiology Graduate Group, University of California, Davis, California 95616, USA

(Received 17 January 2013; accepted 21 May 2013; published online 11 June 2013)

It remains a challenge to obtain a global perspective on the behavioral repertoire of complex nonlinear gene circuits. In this paper, we describe a method for deconstructing complex systems into nonlinear sub-systems, based on *mathematically defined phenotypes*, which are then represented within a *system design space* that allows the repertoire of qualitatively distinct phenotypes of the complex system to be identified, enumerated, and analyzed. This method efficiently characterizes large regions of system design space and quickly generates alternative hypotheses for experimental testing. We describe the motivation and strategy in general terms, illustrate its use with a detailed example involving a two-gene circuit with a rich repertoire of dynamic behavior, and discuss experimental means of navigating the system design space.
 © 2013 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4809776>]

Throughout the pre-genomic era, there was sustained interest in the relationship between genotype and phenotype. However, the announcement of the draft sequence for the human genome revealed the true magnitude of the difficulties. Two of the many challenges are defining a generic concept of “phenotype” and achieving predictive understanding of the complex nonlinear systems that link these levels. Although the genome sequence provides a generic concept of “genotype,” there is no corresponding concept of “phenotype.” Providing a rigorous framework for a deep understanding of their relationship will require concepts that are in some sense comparable. Providing predictive understanding of complex systems will require general methods for their deconstruction into tractable nonlinear sub-systems whose analysis can provide insight into the original system. In this paper, we provide a generic definition of phenotype and a general method of deconstructing complex biochemical systems into qualitatively distinct phenotypes. We describe the motivation for this method and provide a flow diagram of the strategy that is made concrete by an example. The example involves a two-gene circuit with a repertoire that includes monostable, bistable, and oscillatory phenotypes. The results, which are computationally generated, are presented graphically and methods for navigating the system design space experimentally are discussed.

I. INTRODUCTION

Although we now have a generic concept of “genotype” provided by the detailed DNA sequence,⁵ there is no corresponding concept of “phenotype.” Without a generic concept of phenotype, there can be no rigorous framework for a deep understanding of the complex nonlinear systems that link genotype to phenotype. The concept of phenotype must ultimately be grounded in the underlying biochemistry. The vast

majority of biochemical models are represented in terms of the power functions of chemical kinetics^{7,10} or the rational functions of biochemical kinetics,^{1,3} which result from chemical kinetics plus constraints. Achieving predictive understanding of complex nonlinear systems, such as those manifested at various levels of biological organization, is at the heart of the “Genotype to Phenotype Problem” and represents an enormous challenge.⁶ The task could be facilitated if such systems could be deconstructed into a set of tractable nonlinear sub-systems and the results of their analysis reassembled to provide insight into the original system.

The system design space methodology described in this paper provides two innovations that show promise for dealing with the genotype-phenotype challenge. First, it provides a mathematically rigorous definition of *phenotype* as a combination of dominant processes with boundary conditions within which the phenotype is valid. Second, it provides for the *deconstruction of complex systems* into a finite number of mathematically tractable nonlinear sub-systems representing qualitatively distinct phenotypes. This deconstruction based on phenotypic differences differs from the traditional approaches to deconstruction based on differences in space,²⁸ time,^{14,33} or function.^{15,16}

II. MOTIVATION AND STRATEGY FOR CONSTRUCTING A SYSTEM DESIGN SPACE

In this section, we describe a novel approach to system deconstruction based on *differences in phenotype* followed by integration into a *system design space* that allows qualitatively distinct phenotypes of the complex system to be rigorously defined and counted, their relative fitness to be analyzed and compared, their global tolerance to be measured, and their biological design principles to be identified.^{25,27} Here, we describe the approach in general terms with an accompanying flowchart (Fig. 1) for orientation before we go into technical details of a concrete example.

We begin with a conceptual design representing the system of interest. The key components and their interactions

^{a)}Author to whom correspondence should be addressed. Email: masavageau@ucdavis.edu

FIG. 1. A flowchart illustrating the major steps in the phenotypic decomposition and analysis of a nonlinear system. These steps will be made concrete by the example in Sec. III.

are displayed explicitly. We then formulate a mathematical model based on common assumptions about the mechanisms that are represented by the design. The most common models of biochemical systems include the elementary rate laws of chemical kinetics and the rational function rate laws of biochemical kinetics. The result is a system of nonlinear ordinary differential equations that is analytically intractable. Although there are various local methods for analytically examining particular fixed points and numerical approaches as well, each system is typically treated in an ad hoc fashion

and it remains a challenge to obtain a global perspective on the behavioral repertoire of such systems.

We have shown some time ago that a broad class of nonlinear functions and systems of nonlinear ordinary differential equations can be recast *exactly* into the power-law formalism.²⁶ The recasting procedure consists of a few simple steps repeated a finite number of times to yield the representation in Eq. (1), which is called a *generalized mass action system*, or GMA-system, within the power-law formalism. Although there are other representations within the power-law formalism and one of these will become important later, we first focus on the GMA representation. Equation (1) differs from the traditional mass action representation in two ways: (a) the exponential parameters need not be small integer values but can have real values (positive or negative) and (b) there are specific algebraic constraints among the initial conditions for the recast equations, alternatively the original differential equations become a set of differential-algebraic equations. Although perhaps less obvious initially, the rational function representation also can be considered a special case of Eq. (1) as a result of recasting, as we will see later.

$$\frac{dX_i}{dt} = \sum_{k=1}^r \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} - \sum_{k=1}^r \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}} \quad (1)$$

$$X_i(0) = X_{i0} \quad i = 1, \dots, n.$$

The power-law formalism combines nonlinear elements having a very specific structure (products of power laws) with linear operators (addition and differentiation) to form a set of ordinary differential equations that are capable of representing a broad range of nonlinear functions. The derivatives of the state variables with respect to time t are given by dX_i/dt . The α and β parameters are rate constants; whereas the g and h parameters are kinetic orders. There are in general n dependent variables, m independent variables, and a maximum of r terms of a given sign. In any realistic biochemical system, there are several processes that typically contribute to the production and consumption of any given biochemical entity. This corresponds to a GMA-system with several positive terms and several negative terms in each equation. Although the recast representation provides a general framework for the global representation, classification, and systematic comparison of nonlinear systems, it is not the most useful representation for general symbolic analysis because these equations, like the equations in their original form, are analytically intractable.

However, in any given condition, one term in each sum is larger than the others in that sum; in other words, one process dominates the net rate of production and another the net rate of consumption of each entity. For example, if the p th term is dominant among the positive terms of the i th equation, then

$$\alpha_{ip} \prod_{j=1}^{n+m} X_j^{g_{ipj}} > \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} \quad k = 1, \dots, r; \quad k \neq p, \quad (2)$$

which corresponds to $r - 1$ linear inequalities in log space that must be satisfied. The total number of potential

combinations of dominant processes is equal to the product of the number of terms of a given sign in each of the equations. However, only some of these combinations are valid. Each valid combination must satisfy a set of dominance conditions [Eq. (2)]. In addition, there must be a steady-state solution that also satisfies these conditions. This intuitive picture corresponds to a reduction of the GMA-system in Eq. (1) to the following representation within the power-law formalism known as an S-system.

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \quad X_i(0) = X_{i0} \quad i = 1, \dots, n. \quad (3)$$

In contrast to the problematic utility of the GMA representation, this local representation within the power-law formalism is most important because it combines the advantages of capturing essential nonlinear behavior while remaining analytically tractable. In the nonlinear realm of biological systems, it has many of the advantages associated with conventional linear systems. Indeed, Taylor's theorem (in log space) gives a rigorous justification for the S-system representation and specific error bounds within which it will provide an accurate representation. Moreover, it is typically accurate over a wider range of variation than the corresponding linear representation, and it has a very desirable structure from the standpoint of general theory and symbolic analysis.²⁰

The steady-state equations are obtained by setting $dX_i/dt = 0$ for all i in Eq. (3). When none of the X_j 's and none of the rate constants is equal to zero, these equations can be divided by $\alpha_i \prod_{j=1}^{n+m} X_j^{h_{ij}}$ resulting in the following linear system:²⁰

$$[A]y = b, \quad (4)$$

where $[A]$ is a matrix with elements representing differences in kinetic orders $a_{ij} = (g_{ij} - h_{ij})$, y is a column vector with elements representing logarithms of variables $y_i = \ln X_i$, and b is a column vector with elements representing differences in logarithms of rate constants $b_i = \ln \beta_i - \ln \alpha_i$.

The arrays in Eq. (4) can be partitioned into dependent and independent elements, $[A]_d y_d = -[A]_i y_i + b$, and solved to obtain the dependent state variables *explicitly* in terms of the independent variables and parameters of the system

$$\begin{array}{ccc} y_d = [L]y_i + [M]b \\ \uparrow \quad \quad \uparrow \\ \text{slope} \quad \text{intercept}, \end{array} \quad (5)$$

where $[M] = [A]_d^{-1}$ and $[L] = -[M][A]_i$. This solution for the logarithms of the dependent concentrations y_d (y_j , $j = 1, \dots, n$) is divided into two parts. The first exhibits the linear dependence on the logarithms of the independent variables y_i (y_j , $j = n+1, \dots, n+m$); the second exhibits the linear dependence on the logarithms of the rate constant parameters b ($b_j = \ln(\beta_j/\alpha_j)$, $j = 1, \dots, n$). The elements of

$[M]$ and $[L]$ are rational functions involving only kinetic order parameters.

The flux through any pool X_k in steady state is obtained by a simple secondary calculation involving the aggregate rate law for the influx or efflux of X_k and the known values for the concentration variables in steady state. For example, starting with the rate law

$$V_{+k} = \alpha_k \prod_{j=1}^{n+m} X_j^{g_{kj}} \quad k = 1, 2, \dots, n, \quad (6)$$

taking logarithms and expressing the results in matrix notation yields

$$(\ln V_{+}) = (\ln \alpha) + [G]y. \quad (7)$$

Thus, the steady-state solution for the flux variables also is a linear function of the independent variables, which can be readily demonstrated by separating dependent and independent components of the solution and substituting the explicit solution for the dependent state variables.

The S-system equations for an n -variable system [Eq. (3)] can be linearized and expressed as follows:²¹

$$du/dt = F^T A u \quad u(0) = u_0, \quad (8)$$

where $a_{ij} = g_{ij} - h_{ij}$ and $u_i = y_i - y_{is} = (X_i - X_{is})/X_{is}$ (i.e., percent variation in X_i about a steady state indicated by the additional subscript s). The elements of the premultiplier

$$F_i = \alpha_i \prod_{j=1}^{n+m} X_{js}^{g_{ij}} / X_{is} = V_{is} / X_{is} \quad i = 1, 2, \dots, n \quad (9)$$

are always positive. The F_i may be viewed as a pseudo-first-order rate constant or the reciprocal of the turn-over time for the pool X_i .

The explicit solution in Eqs. (5) and (7) gives the complete relationship between the steady-state values of the dependent state variables on the one hand and the values of the independent variables and parameters of the system on the other. The independent variables may be thought of as those that are determined by factors outside the system of interest, the environmentally determined variables. The parameters, which characterize the relatively fixed aspect of the system itself, may be thought of as genetically determined parameters. The separation of these two types of influences is important for a clear understanding of system behavior.

There are a large number of potentially dominant terms and steady-state solutions for the corresponding S-systems; however, not all combinations of dominant terms are valid. To be valid, a particular combination must meet two requirements. First, the resulting S-system must have a steady-state solution. Second, given that solution, all of the other terms in a given sum must be smaller than the presumed dominant term. As noted above, both involve linear systems in logarithmic coordinates. Thus, finding the valid combinations is a tractable linear programming problem^{9,32} that involves solving the S-system equations (set of linear equations in log space) along with the dominance conditions (a set of linear

inequalities in log space).^{11,27} The resulting solutions provide a rigorous mathematical definition of the boundaries between regions of qualitatively distinct behavior for the system. These boundaries involve all of the environmentally determined variables and all of the genetically determined parameters of the system.

Once the boundaries have been determined for all the valid combinations of dominant terms, the results are integrated into a *system design space*. This space, with axes that represent the environmentally determined variables and the genetically determined parameters of the system, consists of space-filling irregular polytopes. We define each point in this gene-by-environment space as a distinct *phenotype* of the system, the characteristic phenotype of the system throughout a valid region as a *qualitatively distinct phenotype*, and the sum of the qualitatively distinct phenotypes as the *phenotypic repertoire* of the system.

The system design space provides a global perspective that links genotype, environment, and phenotype. The system representation for each qualitatively distinct phenotype is always a simple S-system for which determination of local nonlinear behavior reduces to conventional linear analysis.^{22,23} Thus, these phenotypes are completely determined, and their relative fitness can be compared on the basis of relevant performance criteria. These criteria can be quantified by a variety of conventional means including the eigenvalues or the Routh conditions for local stability,¹³ which often characterize the local stability for an entire qualitatively distinct phenotype, logarithmic gains for signal amplification, parameter sensitivities for local robustness, and dominant eigenvalues for response times. In addition, the boundaries that delineate a given phenotype can be used to quantify the tolerance (*global robustness*) to large changes in genotype and environment.⁸

The system design space provides a means of graphically illuminating the relationship between genotypically determined parameters, the environmentally determined variables, and the qualitatively distinct phenotypes of the system. For this reason, we typically choose to plot a genotypically determined parameter on the vertical y-axis, an environmentally determined variable on the horizontal x-axis, and a phenotypic character of interest as a heat map on the z-axis. However, in some circumstances, such as the following example, other views of the space are most illuminating.

III. APPLICATION TO A GENE CIRCUIT

A detailed application will make the motivation and strategy outlined in Sec. II more concrete. Consider as an example, the design for a gene circuit in Fig. 2.

A. Mathematical model

A conventional model for this class of circuitry involves rational function nonlinearities for mRNA synthesis and linear functions for the other processes. These equations can be written as follows:

FIG. 2. Conceptual design for a relaxation oscillator involving a two-gene circuit with both positive and negative feedback loops.

$$\frac{dX_1}{dt} = \alpha_1 \left[\frac{(\rho_1)^{-1} + (X_2/K_{2A})^{g_{12}} + (\rho_1)^{-1} (X_4/K_{4A})^{g_{14}}}{1 + (X_2/K_{2A})^{g_{12}} + (X_4/K_{4A})^{g_{14}}} \right] - \beta_1 X_1, \quad (10)$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 - \beta_2 X_2, \quad (11)$$

$$\frac{dX_3}{dt} = \alpha_3 \left[\frac{(\rho_3)^{-1} + (X_2/K_{2R})^{g_{32}}}{1 + (X_2/K_{2R})^{g_{32}}} \right] - \beta_3 X_3, \quad (12)$$

$$\frac{dX_4}{dt} = \alpha_4 X_3 - \beta_4 X_4. \quad (13)$$

The interpretation of the parameters in these equations is the following: the α 's and β 's are first-order rate constants, the ρ 's are the regulatory capacities, the g 's are kinetic orders representing the co-operativity of repression or activation, the K_{2A} and K_{2R} are concentrations of activator for half-maximal influence on the rate of activator and repressor transcription, and K_{4A} is the concentration of repressor for half-maximal influence on the rate of activator transcription.

B. Recast GMA-system

The recasting of rational function models into the GMA form is a particularly simple process. Although there are alternative strategies, the simplest is to define each polynomial in a denominator as a new variable. The equivalent differential-algebraic system in GMA form of Eq. (10) through Eq. (13) can then be written following the introduction of two new variables as

$$\begin{aligned} \frac{dX_1}{dt} = & \alpha_1 (\rho_1)^{-1} (X_5)^{-1} + \alpha_1 (X_2/K_{2A})^{g_{12}} (X_5)^{-1} \\ & + \alpha_1 (\rho_1)^{-1} (X_4/K_{4A})^{g_{14}} (X_5)^{-1} - \beta_1 X_1, \end{aligned} \quad (14)$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 - \beta_2 X_2, \quad (15)$$

$$\frac{dX_3}{dt} = \alpha_3 (\rho_3)^{-1} (X_6)^{-1} + \alpha_3 (X_2/K_{2R})^{g_{32}} (X_6)^{-1} - \beta_3 X_3, \quad (16)$$

$$\frac{dX_4}{dt} = \alpha_4 X_3 - \beta_4 X_4, \quad (17)$$

$$0 = X_5 - 1 - (X_2/K_{2A})^{g_{12}} - (X_4/K_{4A})^{g_{14}}, \quad (18)$$

$$0 = X_6 - 1 - (X_2/K_{2R})^{g_{32}}. \quad (19)$$

For purposes of illustration, we shall arbitrarily select the nominal steady-state operating point for the system to be at the geometric mean of the regulatable region for both of the rational functions and choose values for the parameters based on the following information for the *lac* operon of *Escherichia coli*. The co-operativity for induction is 2,^{18,19} the capacity for regulation is about 100,^{18,19} the steady-state ratio of regulator to mRNA is 30,^{31,35} the rate constant for mRNA decay is 20.8 h⁻¹,^{4,17} and the rate constant for loss of the stable repressor has a value of (ln2)/T h⁻¹ in cells growing exponentially with a doubling time of T h. The resulting values for the parameters in Eq. (14) through Eq. (19) become: $g_{12} = g_{32} = g_{14} = 2$ for the kinetic order parameters, $\rho_1 = \rho_3 = 100$ for the regulatory capacities, $K_{2A} = 0.3$, $K_{2R} = 9.5$ and $K_{4A} = 0.95$ for the half-maximal concentrations, and $\alpha_1 = 20.82$, $\alpha_3 = 208.2$, $\alpha_2 = \alpha_4 = 30 * \ln(2)/T$, $\beta_1 = \beta_3 = 20.82$, and $\beta_2 = \beta_4 = \ln(2)/T$ for the rate constant parameters in cells growing with a doubling time of T hours.

At this nominal steady state, the system operates at the geometric mean of the regulatable region for each rational function, the operating point will be centered at the origin of the system design space in Sec. III F; and as a necessary condition for limit-cycle oscillations, this fixed point should be an unstable focus. The exact values are not critical for the region with the potential for oscillation, as will be seen in Sec. III F, since it is estimated that these values can change over roughly an order of magnitude without a change in phenotype.

C. Selecting dominant terms

The total number of potentially valid combinations of dominant terms for this example is 36. The cases are numbered systematically by a method that will become clear in Sec. III F. Four representative cases will illustrate the implications of selecting various combinations of dominant terms.

An example of an invalid combination of dominant terms (processes) is the following (case 14). Selecting the second positive term in Eq. (14), the first positive term in Eq. (16), the first negative term in Eq. (18), and the second negative term in Eq. (19) yields the following S-system:

$$\frac{dX_1}{dt} = \alpha_1 (X_2/K_{2A})^{g_{12}} - \beta_1 X_1, \quad (20)$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 - \beta_2 X_2, \quad (21)$$

$$\frac{dX_3}{dt} = \alpha_3 (\rho_3)^{-1} (X_2/K_{2R})^{-g_{32}} - \beta_3 X_3, \quad (22)$$

$$\frac{dX_4}{dt} = \alpha_4 X_3 - \beta_4 X_4. \quad (23)$$

The conditions for dominance of the second positive term in the first equation are $X_4 < K_{4A}[\rho_1(X_2/K_{2A})^{g_{12}}]^{1/g_{14}}$ and

$X_2 > K_{2A}(\rho_1)^{-1/g_{12}}$, for the first positive term in the third equation $X_2 < K_{2R}(\rho_3)^{-1/g_{32}}$, for the first negative term in the fifth equation $X_2 < K_{2A}$ and $X_4 < K_{4A}$, and for the second term in the sixth equation $X_2 > K_{2R}$. Without going further, it is clear that X_2 cannot satisfy the third and the sixth of these conditions because $\rho_3 > 1$, and so there is no valid phenotypic region corresponding to this particular combination of dominant terms.

An example of a valid combination of dominant terms that represents a stable node is the following (case 22). Select the second positive term in Eqs. (14) and (16) and the second negative term in Eqs. (18) and (19) to yield the following S-system:

$$\frac{dX_1}{dt} = \alpha_1 - \beta_1 X_1, \quad (24)$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 - \beta_2 X_2, \quad (25)$$

$$\frac{dX_3}{dt} = \alpha_3 - \beta_3 X_3, \quad (26)$$

$$\frac{dX_4}{dt} = \alpha_4 X_3 - \beta_4 X_4. \quad (27)$$

The corresponding dominance conditions are the following: $X_4 < K_{4A}[\rho_1(X_2/K_{2A})^{g_{12}}]^{1/g_{14}}$ and $X_2 > K_{2A}(\rho_1)^{-1/g_{12}}$, $X_2 > K_{2R}(\rho_3)^{-1/g_{32}}$, $X_2 > K_{2A}$ and $X_4 < K_{4A}(X_2/K_{2A})^{g_{12}/g_{14}}$, and $X_2 > K_{2R}$.

An example of a valid combination of dominant terms that represents an unstable focus is the following (case 23). Select the second positive term in Eqs. (14) and (16), and the third and first negative term in Eqs. (18) and (19) to yield the following S-system:

$$\frac{dX_1}{dt} = \alpha_1 (X_2/K_{2A})^{g_{12}} (X_4/K_{4A})^{-g_{14}} - \beta_1 X_1, \quad (28)$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 - \beta_2 X_2, \quad (29)$$

$$\frac{dX_3}{dt} = \alpha_3 (X_2/K_{2R})^{g_{32}} - \beta_3 X_3, \quad (30)$$

$$\frac{dX_4}{dt} = \alpha_4 X_3 - \beta_4 X_4. \quad (31)$$

The corresponding dominance conditions are the following: $X_4 < K_{4A}[\rho_1(X_2/K_{2A})^{g_{12}}]^{1/g_{14}}$ and $X_2 > K_{2A}(\rho_1)^{-1/g_{12}}$, $X_2 > K_{2R}(\rho_3)^{-1/g_{32}}$, $X_4 > K_{4A}$ and $X_4 > K_{4A}(X_2/K_{2A})^{g_{12}/g_{14}}$, and $X_2 < K_{2R}$.

An example of a valid combination of dominant terms that represents an unstable node is the following (case 24). Select the second positive term in Eqs. (14) and (16) and the third and second negative term in Eqs. (18) and (19) to yield the following S-system:

$$\frac{dX_1}{dt} = \alpha_1 (X_2/K_{2A})^{g_{12}} (X_4/K_{4A})^{-g_{14}} - \beta_1 X_1, \quad (32)$$

$$\frac{dX_2}{dt} = \alpha_2 X_1 - \beta_2 X_2, \quad (33)$$

$$\frac{dX_3}{dt} = \alpha_3 - \beta_3 X_3, \quad (34)$$

$$\frac{dX_4}{dt} = \alpha_4 X_3 - \beta_4 X_4. \quad (35)$$

The corresponding dominance conditions are the following: $X_4 < K_{4A}[\rho_1(X_2/K_{2A})^{g_{12}}]^{1/g_{14}}$ and $X_2 > K_{2A}(\rho_1)^{-1/g_{12}}$, $X_2 > K_{2R}(\rho_3)^{-1/g_{32}}$, $X_4 > K_{4A}$ and $X_4 > K_{4A}(X_2/K_{2A})^{g_{12}/g_{14}}$, and $X_2 > K_{2R}$.

D. Steady-state solution of the S-system

The steady-state solution for case 14 [Eqs. (20) through (23)] is obtained by setting the derivatives to zero, transforming the equations into logarithmic coordinates, solving the linear algebraic equations, and then transforming back to Cartesian coordinates

$$\begin{aligned} X_1 &= \left(\frac{\beta_1}{\alpha_1}\right)^{\frac{1}{g_{12}-1}} \left(\frac{\beta_2 K_{2A}}{\alpha_2}\right)^{\frac{g_{12}}{g_{12}-1}}, \\ X_2 &= \left(\frac{\beta_1 \beta_2 (K_{2A})^{g_{12}}}{\alpha_1 \alpha_2}\right)^{\frac{1}{g_{12}-1}}, \\ X_3 &= \frac{\alpha_3 (K_{2R})^{g_{32}}}{\beta_3 \rho_3} \left(\frac{\beta_1 \beta_2 (K_{2A})^{g_{12}}}{\alpha_1 \alpha_2}\right)^{\frac{1}{g_{12}-1}}, \\ X_4 &= \frac{\alpha_3 \alpha_4 (K_{2R})^{g_{32}}}{\beta_3 \beta_4 \rho_3} \left(\frac{\beta_1 \beta_2 (K_{2A})^{g_{12}}}{\alpha_1 \alpha_2}\right)^{\frac{1}{g_{12}-1}}. \end{aligned} \quad (36)$$

Although there is an easily determined steady-state solution for this S-system, there is no phenotypic region in which it is valid because the dominance conditions cannot be satisfied.

The corresponding steady-state solution for case 22 [Eq. (24) through Eq. (27)] is the following:

$$X_1 = \frac{\alpha_1}{\beta_1}, \quad X_2 = \frac{\alpha_1 \alpha_2}{\beta_1 \beta_2}, \quad X_3 = \frac{\alpha_3}{\beta_3}, \quad X_4 = \frac{\alpha_3 \alpha_4}{\beta_3 \beta_4}. \quad (37)$$

Similarly for case 23 [Eq. (28) through Eq. (31)],

$$\begin{aligned} X_1 &= \left\{ \left(\frac{\alpha_1}{\beta_1}\right) \left(\frac{\alpha_2}{\beta_2}\right)^{g_{14}g_{32}-g_{12}} \left(\frac{\beta_3 \beta_4 K_{4A}}{\alpha_3 \alpha_4}\right)^{g_{14}} \left(\frac{1}{K_{2A}}\right)^{g_{12}} (K_{2R})^{g_{14}g_{32}} \right\}^{\frac{1}{1-g_{12}+g_{14}g_{32}}}, \\ X_2 &= \left\{ \left(\frac{\alpha_1 \alpha_2}{\beta_1 \beta_2}\right) \left(\frac{\beta_3 \beta_4 K_{4A}}{\alpha_3 \alpha_4}\right)^{g_{14}} \left(\frac{1}{K_{2A}}\right)^{g_{12}} (K_{2R})^{g_{14}g_{32}} \right\}^{\frac{1}{1-g_{12}+g_{14}g_{32}}}, \\ X_3 &= \left\{ \left(\frac{\alpha_1 \alpha_2}{\beta_1 \beta_2}\right) \left(\frac{\beta_3}{\alpha_3}\right)^{\frac{g_{12}-1}{g_{32}}} \left(\frac{\beta_4 K_{4A}}{\alpha_4}\right)^{g_{14}} \left(\frac{1}{K_{2A}}\right)^{g_{12}} (K_{2R})^{g_{12}-1} \right\}^{\frac{g_{32}}{1-g_{12}+g_{14}g_{32}}}, \\ X_4 &= \left\{ \left(\frac{\alpha_1 \alpha_2}{\beta_1 \beta_2}\right) \left(\frac{\beta_3 \beta_4}{\alpha_3 \alpha_4}\right)^{\frac{g_{12}-1}{g_{32}}} (K_{4A})^{g_{14}} \left(\frac{1}{K_{2A}}\right)^{g_{12}} (K_{2R})^{g_{12}-1} \right\}^{\frac{g_{32}}{1-g_{12}+g_{14}g_{32}}} \end{aligned} \quad (38)$$

and for Case 24 [Eq. (32) through Eq. (35)]:

$$\begin{aligned} X_1 &= \left\{ \left(\frac{\beta_1}{\alpha_1}\right) \left(\frac{\beta_2}{\alpha_2}\right)^{g_{12}} \left(\frac{\alpha_3 \alpha_4}{\beta_3 \beta_4 K_{4A}}\right)^{g_{14}} (K_{2A})^{g_{12}} \right\}^{\frac{1}{g_{12}-1}}, \\ X_2 &= \left\{ \left(\frac{\beta_1 \beta_2}{\alpha_1 \alpha_2}\right) \left(\frac{\alpha_3 \alpha_4}{\beta_3 \beta_4 K_{4A}}\right)^{g_{14}} (K_{2A})^{g_{12}} \right\}^{\frac{1}{g_{12}-1}}, \\ X_3 &= \frac{\alpha_3}{\beta_3}, \quad X_4 = \frac{\alpha_3 \alpha_4}{\beta_3 \beta_4}. \end{aligned} \quad (39)$$

Clearly, the S-system that determines the phenotype of the various cases results in qualitatively different steady-state solutions. For example, with the assumed values for the parameters in Sec. III B, the dependence of the activator concentration X_2 on the repressor lifetime $1/\beta_4$ is zero-order in case 22, inverse 2/3 power in case 23, and second-order in case 24.

E. Boundary conditions for a valid phenotype

Inserting the steady-state solution for case 22 into the corresponding dominance conditions yields the following boundary conditions within which this phenotype is valid:

$$\begin{aligned} \log \beta_4 &> \log \left[\frac{\alpha_3 \alpha_4}{\beta_3 K_{4A}} \left(\frac{\beta_1 K_{2A}}{\alpha_1 \alpha_2}\right)^{g_{12}/g_{14}} \right] + \left(\frac{g_{12}}{g_{14}}\right) \log \beta_2 \\ \text{and } \log \beta_2 &< \left[\frac{\alpha_1 \alpha_2}{\beta_1 K_{2R}} \right] \end{aligned} \quad (40)$$

or with the numerical values for parameters from Sec. III B

$$\log \beta_4 > 0.5 + \log \beta_2 \quad \text{and} \quad \log \beta_2 < 0.340. \quad (41)$$

The boundary conditions for case 23 can be obtained in a similar fashion. Their expression in symbolic form analogous to Eq. (40) is linear although very complicated. However, the corresponding expressions in numerical form

$$\begin{aligned} -1.82 - \log \beta_2 < \log \beta_4 < 1.18 - \log \beta_2 \text{ and} \\ -0.830 + 0.5 \log \beta_2 < \log \beta_4 < 0.669 + 0.5 \log \beta_2 \end{aligned} \quad (42)$$

are more useful for making comparisons in Sec. III F. Similarly, for case 24, the boundary conditions in numerical form are

$$\begin{aligned} 0.5 + \log \beta_2 < \log \beta_4 < 1.5 + \log \beta_2 \text{ and} \\ \log \beta_4 < 0.669 + 0.5 \log \beta_2. \end{aligned} \quad (43)$$

As seen in Eq. (40) through Eq. (43), all of the boundaries are linear in logarithmic coordinates (see also Fig. 3(a)).

F. Constructing the system design space

All the boundaries of the valid phenotypes, which are linear hyper-planes, can be determined and the system design space constructed computationally (for small systems using a MATLAB tool box¹¹). All but 8 of the 36 potential phenotypes for this example are valid. The results are plotted in Fig. 3(a) with the rate of removal of the activator on the x-axis and that of the repressor on the y-axis.

The *global tolerance* to a change in phenotype by parameter variation is estimated by the ratio of values at the extremes that remain within a phenotypic region. For example, the global tolerances to a change in the oscillatory phenotype by variation in a regulator half-life are estimated to be 10 fold, as is evident from the slice of system design space in Fig. 3(a). The global tolerances for the other parameters can be readily estimated in a similar fashion, and the values are all at least an order of magnitude.

G. Analysis

The analysis of each phenotype is a straightforward application of linear analysis in the logarithmic space. Any phenotypic characteristic of interest can be determined by analysis and plotted as a heat map on the z-axis. We have already shown how the steady-state solutions are obtained in the process of determining phenotype validity (Sec. III D).

The steady-state solution for the activator X_2 over all the phenotypic regions is plotted for increasing values of β_2 in Fig. 3(b) and for decreasing values in Fig. 3(c). The phenotypes with different values in the two plots represent cases of bistability and hysteresis.

The local dynamics can be verified analytically and the results often apply to an entire region of the design space. The behavior in regions with three overlapping phenotypes is of particular interest because it represents regions of hysteresis in which two of the phenotypes are stable and the third unstable. As an example, the stability for the case 22 phenotype is obtained following application of Eq. (8):

$$\begin{aligned} du_1/dt &= F_1[-u_1] \\ du_2/dt &= F_2[u_1 - u_2] \\ du_3/dt &= F_3[-u_3] \\ du_4/dt &= F_4[u_3 - u_4]. \end{aligned} \quad (44)$$

The F-factors are always positive, so this S-system is clearly stable throughout the region for case 22, regardless of parameter values. Similarly, for the case 35 phenotype,

$$\begin{aligned} du_1/dt &= F_1[-u_1] \\ du_2/dt &= F_2[u_1 - u_2] \\ du_3/dt &= F_3[g_{32}u_2 - u_3] \\ du_4/dt &= F_4[u_3 - u_4]. \end{aligned} \quad (45)$$

This S-system, like that for case 22, is stable throughout its phenotypic region. For the third overlapping phenotype (case 24),

$$\begin{aligned} du_1/dt &= F_1[-u_1 + g_{12}u_2 - g_{14}u_4] \\ du_2/dt &= F_2[u_1 - u_2] \\ du_3/dt &= F_3[-u_3] \\ du_4/dt &= F_4[u_3 - u_4]. \end{aligned} \quad (46)$$

The characteristic equation in this case is $[\lambda^2 + (F_1 + F_2)\lambda + F_1F_2(1 - g_{12})](\lambda + F_3)(\lambda + F_4) = 0$, and

FIG. 3. (a) System design space for a relaxation oscillator involving both positive and negative feedback loops. The color bar represents the case number for each of the phenotypes. (b)–(d) Phenotypic characteristics represented as a heat map on the z-axis of the system design space: (b) logarithm of activator concentration X_2 with increasing values or (c) decreasing values of the parameter β_2 , and (d) number of eigenvalues with positive real part. Color bar represents the value of the phenotypic character in each example.

this S-system is clearly unstable throughout its phenotypic region since $g_{12} = 2 > 1$.

A global view of these dynamic traits, another phenotypic characteristic, can be obtained by plotting the number of eigenvalues with positive real part on the z-axis. The results are shown in Fig. 3(d), in which the stable regimes with zero are shown in blue, the bistable regimes with one (for the unstable phenotype) are in red, and the oscillatory regimes with two are in yellow.

IV. COMPARISONS WITH CONVENTIONAL BIFURCATION ANALYSIS

The system design space provides a global view of the qualitatively distinct phenotypes of the system. The boundaries in system design space are not arbitrary, rather they are completely determined by the model itself. In some cases, these boundaries correspond to conventional bifurcations. In others, they signify a qualitative change in phenotype represented by the solution of the different forms of the S-system equations governing the adjacent phenotypes. For example, as was seen in Sec. III D, the dependency of the activator concentration X_2 on the repressor lifetime $1/\beta_4$ is zero-order in case 22, inverse 2/3 power in case 23, and second-order in case 24.

The individual phenotypes can be further analyzed in detail using quantitative methods. In this particular case, involving the example in Sec. III, the focus is on the dynamic characteristics. Fig. 3(d) exhibits a number of stable regions, several hysteretic regions, and a central oscillatory region. The regions representing hysteresis or oscillation tend to be overestimates due to the lack of dominance at the boundaries.²⁴

The system design space methodology provides for the rapid identification of regions of potential interest that can then be explored in more detail by quantitative methods if necessary. In the case of the circuit in Fig. 2, the method identified regions with three distinct dynamic behaviors. The results from the analysis in system design space are compared with the results from conventional bifurcation analysis³⁰ in Fig. 4. The bifurcation diagram in Fig. 4(a) is obtained with the lifetime of the repressor fixed at a value corresponding to $\beta_4 = 100/\text{h}$. Superimposed is the corresponding steady-state value of X_2 and its local stability as determined from the analysis in system design space. A comparison of the two results shows that the hysteretic region, as determined by bifurcation analysis, is entirely within that determined by the design space analysis, as expected. The bifurcation diagram in Fig. 4(b) is obtained with the lifetime of the repressor fixed at a value corresponding to $\beta_4 = \log(2)/\text{h}$. In this case, the hysteretic and oscillatory regions as determined by the bifurcation analysis also are entirely within those determined by the design space analysis.

As the bifurcation parameter β_2 increases from the left, oscillations emerge with gradually increasing amplitude as the result of a typical super-critical Hopf bifurcation; as it decreases from the right, large-amplitude oscillations emerge full blown as the result of a sub-critical Hopf bifurcation.

FIG. 4. Comparison of saddle-point (open circles) and Hopf bifurcations (filled circles) as determined by system design space (gray lines) and conventional bifurcation (black lines) analyses. (a) Saddle-point bifurcation predicted with the lifetime of the repressor fixed at a value corresponding to $\beta_4 = 100/\text{h}$. (b) Saddle-point and Hopf bifurcations predicted with the lifetime of the repressor fixed at a value corresponding to $\beta_4 = \log(2)/\text{h}$.

Although the bifurcation diagram in Fig. 4(b) suggests a hysteretic switch with the oscillations forming and dissolving at different threshold values of the bifurcation parameter, the hysteresis is not seen in this case because the amplitude of the oscillation is so large that the system is thrown into the basin of attraction for the stable steady state without maintaining the oscillation. The envelope for the amplitude of the oscillations is shown in Fig. 5(a) and the corresponding frequency of the oscillations is shown in Fig. 5(b), as a function of the bifurcation parameter β_2 .

V. DISCUSSION

The system design space methodology in this paper provides two innovations that show promise for dealing with the genotype-phenotype challenge. First, it provides a mathematically rigorous definition of phenotype as a combination of dominant processes with linear hyper-plane boundaries, which define an irregular polytope within which the phenotype is valid. Second, it provides for the deconstruction of complex systems into a finite number of mathematically tractable nonlinear sub-systems representing the qualitatively distinct phenotypes.

The S-systems that characterize the qualitatively distinct phenotypes are rigorously defined by the underlying mechanisms of the system itself, and all of the parameters are involved in determining the geometrical landmarks in design space. Moreover, the S-systems typically involve distributed aspects of the entire system and are not a localized “module”

FIG. 5. Expanded view of the sub-critical Hopf bifurcation in Fig. 4(b) at higher values of the bifurcation parameter β_2 : (a) Envelope of the amplitude of the oscillations. (b) Frequency of the oscillations.

clearly separable from the remainder of the system. Thus, this method of deconstruction is based on *differences in phenotype*, rather than the traditional approaches to deconstruction based on space, time, or function. The integration of phenotypes into a system design space allows the qualitatively distinct phenotypes to be identified, enumerated, characterized, and compared. The system design space provides an efficient means to characterize system behavior over a broad range of parameter values, and the landmarks in this space represent particular constellations of parameters that define relevant design principles. These characteristics facilitate the rapid generation of alternative hypothesis for experimental testing; in particular, they identify designs that are more promising because the range of realizable values for the parameters is larger.

This methodology provides an efficient means of obtaining a global perspective on the behavioral repertoire of a system. It identifies regions of system design space that are of particular interest and that can be examined in further detail as needed. In the example of Sec. III, the method identified regions of potential hysteresis and limit-cycle oscillation. Although these behaviors were not found throughout the regions identified, the method never failed to find these behaviors somewhere within the identified regions. This has been true in other applications as well.^{12,24} In this sense, the method is conservative, overestimating the potential such that relevant behaviors are not overlooked.

The boundary between qualitatively distinct phenotypes signifies a qualitative change in phenotype represented by a different form of solution for the S-system equations governing the adjacent phenotypes. Some, but not all, of the

boundaries correspond to conventional bifurcations. The comparison with conventional bifurcation analysis in Sec. IV shows quantitative agreement over regions far from the boundaries and qualitative agreement in the regions near the boundaries. The selection of a proper trajectory for the bifurcation parameter to reveal the Hopf bifurcations was facilitated by having the global perspective of the system design space as a guide.

The system design space method provides insights that would be useful in the context of synthetic biology. It is widely known that most efforts in synthetic biology are initially unsuccessful. Indeed, this should be expected, given the decades of experience attempting to characterize and understand natural gene circuitry. Even in synthetic constructs where the relevant interactions have been clearly identified and realized, there is still uncertainty in setting parameter values. The challenges can be amplified if the relevant interactions may not have been identified and/or parameter values are so far off that sampling of the parameter space is difficult. The global perspective of the system design space could shorten the time of the design cycle (design, construct, test, redesign) by narrowing the search space for a desired phenotype and designing for the largest region of realizability.

The insights provided by the example in Sec. III would have increased the probability for success of a previous synthetic project aimed at the realization of a robust sustained oscillator with two transcriptional regulators.² The easily tunable parameters in such a system are the lifetimes of the two regulator proteins (via gratuitous inducers) and the copy number of their transcriptional units. The results in Fig. 3 represent a particular slice through the design space that was selected to emphasize two of these parameters that can be readily tuned to achieve the different phenotypic characteristics of the model. In an appropriate experimental implementation utilizing the Lac repressor and the Ara activator, the lifetime of the Lac repressor can be tuned by altering the concentration of the gratuitous inducer IPTG whereas that of the Ara activator can be tuned by altering the concentration of arabinose.²⁹ Another slice that achieves essentially the same result is to plot the copy number of the activator gene (proportional to α_1) on the x-axis and that of the repressor gene (proportional to α_3) on the y-axis. These parameters can be adjusted by placing the two transcriptional units on separate plasmids with variable copy-number control.³⁴ Any combination of these four parameters could be used to navigate in the system design space. Indeed, any number of slices can be readily visualized, but the choice is determined by the particular view that is most informative in a given context and the difficulty of experimental realization.

In principle, the system design space method can be applied to the vast majority of biochemical models that are represented in terms of the power functions of chemical kinetics or the rational functions of biochemical kinetics. It provides a means of graphically illuminating the relationship between genotypically determined parameters, environmentally determined variables, and the qualitatively distinct phenotypes of the system. Software is already available for analyzing small systems¹¹ and is currently under

development for larger systems. The method is scalable by virtue of the transformation into a tractable linear programming problem that involves solving the S-system equations (set of linear equations in log space) along with the dominance conditions (a set of linear inequalities in log space).

ACKNOWLEDGMENTS

We thank R. A. Fasani and D. Nicklas for fruitful discussions regarding the challenges and practical applications of system design space. This work was supported in part by a grant from the US Public Health Service (RO1-GM30054) and by a Stanislaw Ulam Distinguished Scholar Award from the Center for Non-Linear Studies of the Los Alamos National Laboratory to MAS.

- ¹R. A. Alberty, *Enzyme Kinetics* (Wiley & Sons, Hoboken, N.J., 2011).
- ²M. R. Atkinson, M. A. Savageau, J. T. Myers, and A. J. Ninfa, "Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*," *Cell* **113**, 597–607 (2003).
- ³H. Bisswanger, *Enzyme Kinetics* (Wiley-VCH, Weinheim, Germany, 2008).
- ⁴M. Blundell and D. Kennell, "Evidence for endonucleolytic attack in decay of *lac* messenger RNA in *Escherichia coli*," *J. Mol. Biol.* **83**, 143–161 (1974).
- ⁵F. R. Blattner, G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, "The complete genome sequence of *Escherichia coli* K-12," *Science* **277**, 1453–1462 (1997).
- ⁶S. Brenner, "Genomics: The end of the beginning," *Science* **287**, 2173–2174 (2000).
- ⁷C. Capellas and B. H. J. Bielski, *Kinetic Systems* (Wiley-Interscience, New York, 1972).
- ⁸P. M. B. M. Coelho, A. Salvador, and M. A. Savageau, "Global tolerance of biochemical systems and the design of moiety-transfer cycles," *PLOS Comput. Biol.* **5**(3), e1000319 (2009).
- ⁹G. B. Dantzig, *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ, 1963).
- ¹⁰P. Erdi and J. Toth, *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models* (Manchester University Press, Manchester, UK, 1989).
- ¹¹R. A. Fasani and M. A. Savageau, "Automated construction and analysis of the design space for biochemical systems," *Bioinformatics* **26**, 2601–2609 (2010).
- ¹²R. A. Fasani and M. A. Savageau, "Molecular mechanisms of multiple toxin-antitoxin systems are coordinated to govern the persister phenotype," *Proc. Natl. Acad. Sci. U.S.A.* (to be published).
- ¹³F. R. Gantmacher, *Applications of the Theory of Matrices* (Interscience Publishers, N.Y., 1959).
- ¹⁴J. Gunawardena, "A linear framework for time-scale separation in nonlinear biochemical systems," *PLOS One* **7**(5), e36321 (2012).
- ¹⁵L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray, "From molecular to modular cell biology," *Nature* **402**, C47–C52 (1999).
- ¹⁶A. Hintze and C. Adami, "Evolution of complex modular biological networks," *PLoS Comput. Biol.* **4**(2), e23 (2008).
- ¹⁷L. Leive and V. Kollin, "Synthesis, utilization and degradation of lactose operon mRNA in *Escherichia coli*," *J. Mol. Biol.* **24**, 247–259 (1967).
- ¹⁸S. Oehler, S. Alberti, and B. Müller-Hill, "Induction of the *lac* promoter in the absence of DNA loops and the stoichiometry of induction," *Nucl. Acids Res.* **34**, 606–612 (2006).
- ¹⁹J. R. Sadler and A. Novick, "The properties of repressor and the kinetics of its action," *J. Mol. Biol.* **12**, 305–327 (1965).
- ²⁰M. A. Savageau, "Concepts relating the behavior of biochemical systems to their underlying molecular properties," *Arch. Biochem. Biophys.* **145**, 612–621 (1971).
- ²¹M. A. Savageau, "Optimal design of feedback control by inhibition: Dynamic considerations," *J. Mol. Evol.* **5**, 199–222 (1975).
- ²²M. A. Savageau, "Design principles for elementary gene circuits: Elements, methods, and examples," *Chaos* **11**, 142–159 (2001).
- ²³M. A. Savageau, *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*, 40th Anniversary Edition (Amazon, 2009) [a reprinting of the original edition published by Addison-Wesley, Reading, MA, 1976].
- ²⁴M. A. Savageau, "Design of the *lac* gene circuit revisited," *Math. Biosci.* **231**, 19–38 (2011).
- ²⁵M. A. Savageau, "Phenotypes and design principles in system design space," in *Handbook of Systems Biology*, edited by A. J. M. Walhout, M. Vidal, and J. Dekker (Elsevier, San Diego, CA, 2013), pp. 287–310.
- ²⁶M. A. Savageau and E. O. Voit, "Recasting nonlinear differential equations as S-systems: A canonical nonlinear form," *Math. Biosci.* **87**, 83–115 (1987).
- ²⁷M. A. Savageau, P. M. B. M. Coelho, R. Fasani, D. Tolla, and A. Salvador, "Phenotypes and tolerances in the design space of biochemical systems," *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6435–6440 (2009).
- ²⁸P. A. Srere and K. Mosbach, "Metabolic compartmentation: Symbiotic, organellar, multienzymic, and microenvironmental," *Ann. Rev. Microbiol.* **28**, 61–83 (1974).
- ²⁹J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, "A fast, robust and tunable synthetic gene oscillator," *Nature* **456**, 516–520 (2008).
- ³⁰S. H. Strogatz, *Nonlinear Dynamics and Chaos* (Westview Press, Cambridge, MA, 1994).
- ³¹Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie, "Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells," *Science* **329**, 533–538 (2010).
- ³²R. J. Vanderbei, *Linear Programming: Foundations and Extensions* (Springer, New York, NY, 2008).
- ³³C. H. Waddington, *The Strategy of the Genes* (Macmillan, New York, 1957).
- ³⁴K. Williams, M. A. Savageau, and R. M. Blumenthal, "A bistable hysteretic switch in an activator-repressor regulated restriction-modification system," *Nucl. Acid Res.* (to be published).
- ³⁵O. Yarchuk, N. Jacques, J. Guillerez, and M. Dreyfus, "Inter-dependence of translation, transcription and mRNA degradation in the *lacZ* gene," *J. Mol. Biol.* **226**, 581–596 (1992).