



Phenotype-centric modeling for elucidation of biological design principles[☆]



Miguel A. Valderrama-Gómez^a, Rebecca E. Parales^a, Michael A. Savageau^{a,b,*}

^a Department of Microbiology & Molecular Genetics, University of California, Davis, CA 95616 USA

^b Department of Biomedical Engineering, University of California, Davis, CA 95616 USA

ARTICLE INFO

Article history:

Received 1 May 2018

Accepted 9 July 2018

Available online 20 July 2018

Keywords:

Gene circuit architecture

Dynamic phenotypes

System design space

Global robustness and evolvability

ABSTRACT

A recently developed ‘phenotype-centric’ modeling strategy combines four innovations with the potential to advance our understanding of complex biological systems: (1) a rigorous mathematical definition of biochemical phenotypes, (2) a method for enumerating the phenotypic repertoire based on the biomolecular network architecture, (3) an integrated suite of computational algorithms for the efficient prediction of parameter values and analysis of the phenotypic repertoire, and (4) a user-focused environment for navigating the resulting space of phenotypes and identifying biologically relevant features and system design principles. These innovations will facilitate deterministic and stochastic simulations that require parameter values, will accelerate both hypothesis discrimination in systems biology and the design cycle in synthetic biology. Here we first review the fundamental definition of biochemical phenotype that enables this new modeling strategy and give an overview of the strategy using a simple system from phage λ to provide an example of a global design principle. Second, we illustrate this approach in more detail with an application to a common network architecture involving positive and negative feedback. We report system design principles related to the global tolerances of this system’s phenotypes. Finally, we apply the phenotype-centric strategy to a logic network and compare the results with those obtained from a Boolean approach. Mechanistic and Boolean models have well-documented complementary advantages and disadvantages. Mechanistic models have the advantage of being biologically realistic; however, they also are limited by the large number of kinetic parameters whose values are largely unknown. Boolean models have the advantage of being parameter free; however, they also are limited by the absence of well-known physical and chemical constraints. We show that the phenotype-centric modeling strategy combines advantages of both.

© 2018 Elsevier Ltd. All rights reserved.

Personal preface

I first met Rene back in the 1970s. We were often participants at the same conferences, and engaged in a number of stimulating discussions. We shared an interest in phage λ , Rene beginning with his experimental work in the field, and me as a result of my interactions with the λ community through my colleagues David Friedman at the University of Michigan and John Little at the University of Arizona. We also shared a more abstract interest in the underlying design principles not only of phage λ but of biological systems in general. We pursued these interests with complementary approaches, Rene favoring Boolean models whereas I favored mechanistic models. Because of both these shared interests, in this paper my colleagues and I

have selected a simple system from phage λ to introduce some of the basic concepts in our phenotype-centric approach and to provide an example of a global design principle. There has been a long history of comparisons involving Boolean and mechanistic models; our new phenotype-centric approach, and the design principles that it has elucidated, provide a new perspective on the advantages and disadvantages of these complementary approaches. I wish Rene were here today; I am sure we would have a wonderful lively discussion.

Michael Savageau

1. Introduction

Relating the genotype and environment to the phenotype exhibited by a biological system is one of the ‘Grand Challenges’ in Biology (Brenner, 2000). Advances in high-throughput DNA sequencing has given us the complete genome sequence for numerous organisms. As a result, we have a well-defined generic concept of *genotype* as the repertoire of genes encoded in the digital

[☆] This article is further included in a special issue of JTB dedicated to the memory of Prof. René THOMAS.

* Corresponding author.

E-mail address: masavageau@ucdavis.edu (M.A. Savageau).

sequence. However, there has been no corresponding concept of *phenotype*. These tend to be ad hoc and descriptive – size, shape, color, etc. Without a comparable generic definition of phenotype there can be no deep understanding of the relationship between genotype and phenotype; one cannot “predict” a phenotype that has not already been seen! To address this problem, we have proposed a rigorous definition of phenotype in terms of the biochemistry that mechanistically links genotype and environment to the phenotype (Savageau et al., 2009).

Before we consider this definition, it will be helpful first to provide some context. We start with the premise that organisms are biochemical systems: There is a common chemical basis for all forms of life as we know it; organisms deal with many forms of energy, but the basic unit of exchange is chemical; cellular functions are typically catalyzed by enzymes. There are a number of fundamental constraints on these systems, including microscopic reversibility of chemical kinetics, Haldane relations of biochemical kinetics, conserved moieties, stoichiometry of reactions, precursor-product relationships, molecular crowding and solubility limits. The most quantitative and scalable descriptions of these biochemical systems involve rate law functions – stochastic, deterministic or Boolean – each with their advantages and disadvantages. Our focus, although not exclusive, is on deterministic rate laws. These include the power-law functions of chemical kinetics and the rational functions of biochemical kinetics. They have advantages of naturally incorporating fundamental constraints and analytically determining design principles, but the disadvantage of requiring numerous, typically unknown, parameter values. Thus, the foundation that provides our modeling context is *fundamental biochemical kinetics*, which has broad general applicability as indicated by the vast majority of biochemical models that are of this type (Chelliah et al., 2013).

Given this context, we consider the scope of biochemical systems theory to include mechanistic models governed by rate laws. These rate laws are the power functions of chemical kinetics and the rational functions of biochemical kinetics. Functions of these rate laws are integrated into a network by means of Kirchhoff's Node Law. The result is a system of differential-algebraic equations. *Without loss of generality*, the differential-algebraic equations consisting of power-laws and rational functions can be recast trivially into Generalized Mass Action (GMA) equations consisting only of sums and products of power-law functions (Savageau and Voit, 1987):

$$\begin{aligned} \frac{dX_i}{dt} &= \sum_{k=1}^{P_i} \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} - \sum_{k=1}^{Q_i} \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}}, \quad i = 1, \dots, n_c \\ 0 &= \sum_{k=1}^{P_i} \alpha_{ik} \prod_{j=1}^{n+m} X_j^{g_{ijk}} - \sum_{k=1}^{Q_i} \beta_{ik} \prod_{j=1}^{n+m} X_j^{h_{ijk}}, \quad i = (n_c + 1), \dots, n \end{aligned} \quad (1)$$

The m independent, n dependent, n_c chemical and $(n - n_c)$ auxiliary variables X_i are all non-negative real. The rate constants α_{ik} and β_{ik} are non-negative real, and the kinetic orders g_{ijk} and h_{ijk} are integer. P_i and Q_i are the number of positive and negative terms in each equation.

2. Definition of phenotypes

We start by defining phenotypes in terms of the fixed points of the system. Each constituent of the system will in general have several processes described by the positive terms in the GMA equations and several described by the negative terms. Imagine a snapshot of a system in steady state. For each constituent, one of its positive terms will be larger than the others; similarly, one of its negative terms will be larger than the others. Call these the *dominant input process* and *dominant output process* for

the constituent pool. Construct a dominant sub-system (S-system) consisting only of the dominant processes for each constituent:

$$0 = \alpha_{ip} \prod_{j=1}^{n+m} X_j^{g_{ijp}} - \beta_{iq} \prod_{j=1}^{n+m} X_j^{h_{ijq}}, \quad i = 1, \dots, n \quad (2)$$

where p is the dominant input process among the P_i terms in Eq. (1), and q is the dominant output process among the Q_i terms. If this S-system has a solution, then test to see that it is self-consistent by substituting the solution into all of the other terms in the original equations and demonstrate that the dominant terms are indeed the largest. If it satisfies this test, then it defines an *elemental phenotype* of the system. There is a finite number of such combinations that define a dominant S-system, and these define the repertoire of qualitatively distinct phenotypes.

Note that Eq. (2) is a linear system of equations in logarithmic coordinates, that the test for validity involves a system of linear inequalities in logarithmic coordinates, and that the boundaries of the phenotype in the parameter space of the original system are rigorously defined by linear hyper-planes. Thus, all of this involves well-known linear mathematics. It should be clearly understood that this approach involves approximations to the actual system. Experience to date shows that overall the accuracy is very good, with errors concentrated in the neighborhood of the boundaries, where by definition there is no dominance. The important point here is not the inaccuracies near the boundaries but that the boundaries separating qualitatively distinct phenotypes are rigorously defined. Although this approach has a strong foundation based on well-known linear mathematics (in log space), there are still challenges in the software implementation that automates subsequent analysis. An obvious issue that arises in all modeling approaches is how it scales with problem size. For some approaches the scaling is straightforward and the relevant metric is the number of system variables whereas for others it is the number of parameters. This is not the case for the phenotype-centric approach for which the number of combinations of terms is key. There is a bound given by the total number of combinations, but this is a very poor bound because many of the combinations lead to mathematical impossibilities and these can be ignored (see the concrete example in Section 4). The other most important, but difficult to specify, issue with regard to scaling is problem structure. In the phenotype-centric approach each phenotype involves a tractable linear analysis that is independent of that for all the other phenotypes. This represents what computer scientists call an embarrassingly parallelizable problem, and suggests that this will greatly improve the scaling to larger problems. Portability across platforms and version updates are currently being addressed, and some of the other challenges have been discussed elsewhere (Lomnitz and Savageau, 2016a; Savageau, 2013) and are of no concern for the material presented here.

To summarize, we have the following definitions: an **elemental phenotype** is the set of concentrations and fluxes corresponding to a valid combination of dominant processes functioning within an intact system, a **qualitatively distinct phenotype** is the characteristic phenotype that exists throughout a region of validity (polytope) in parameter space, and a **phenotypic repertoire** is the collection of qualitatively distinct phenotypes integrated into a space-filling structure in parameter space. These rigorously defined biochemical phenotypes can be combined in various ways to generate complex composite phenotypes. We will illustrate some simple examples of overlapping and clustered composites.

Although the application to complex developmental, physiological and behavioral traits exhibited by higher organisms is well beyond current capabilities and our purposes here, we have a basis for addressing these challenges based on comparable concepts for relating genotype to phenotype. Namely, the *genotypic repertoire* is

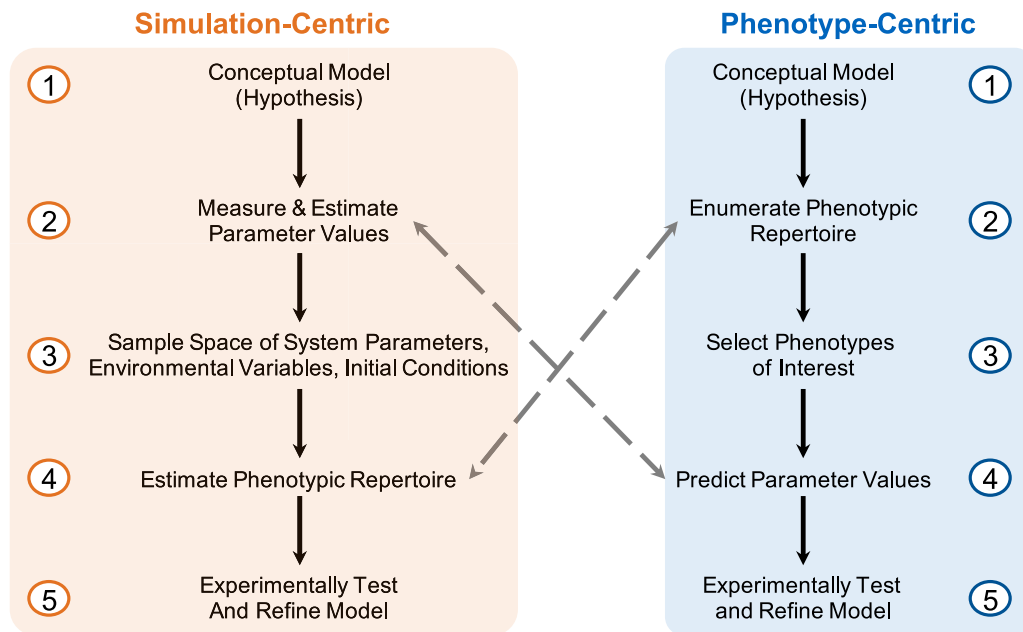


Fig. 1. Simulation- and Phenotype-centric Modeling Strategies. While parameter values have a central role in the traditional simulation-centric modeling strategy, experimentally observable phenotypes are the focus of our novel modeling approach. Instead of using parameter values as input, the phenotype-centric approach predicts a region in the parameter space for the realization of a phenotype of interest. By exhaustively enumerating the phenotypic repertoire of the model at an early stage, our modeling strategy allows for the rapid elimination of wrong hypotheses (models) that are not able to describe the experimental phenotype of interest.

the collection of genes for a system and the phenotypic repertoire is the collection of qualitatively distinct phenotypes for a system.

In exploring the implications of our phenotype definitions, we initially focused on well-characterized systems for which a nominal set of parameter values was available (Reviewed in Savageau, 2013). This allowed us to characterize the phenotypic repertoire in a highly structured ‘system design space’, which consists of a finite number of space-filling ‘chunks’ (irregular polytopes) corresponding to the qualitatively distinct phenotypes of the system. Given the nominal set of parameter values and its location in this design space, we suggested a new definition of robustness that we call *global tolerance*; namely, the fold change in a parameter value that the system can tolerate before there is a change in phenotype (Coelho et al., 2009). Until recently (Lomnitz and Savageau, 2016a), we were still thinking in terms of the conventional modeling strategy in which one first had to start with values for the parameters; since then, we have discovered deeper implications that enable a very different modeling strategy.

3. A phenotype-centric modeling strategy

The experimental and computational challenges in modeling complex biological systems are hard to over-estimate. These are complex, nonlinear, stochastic systems with rough fitness landscapes. They involve large numbers of variables, parameters, inputs, and initial conditions. This gives rise to a combinatorial explosion involving experiments and simulations. Time, cost and technical limitations lead to noisy data and to sparse sampling of experiments and simulations, and many of these challenges exist even for modest-sized systems.

Our definition of phenotypes helps to address some of these issues by enabling a novel phenotype-centric modeling strategy that largely inverts the conventional strategy, which we might call simulation-centric. An overview of the differences is given in Fig. 1. In either case the starting point is a conceptual model (hypothesis). This leads to a mathematical model, which as noted above, is typically a complex nonlinear system with many unknown parameters that is analytically intractable.

Table 1

Biochemical Systems Have Relatively Fixed Parts and Variable Parts. Our novel phenotype-centric modeling approach exploits architectural features of the system to determine its phenotypic repertoire. The parametric component of the system, which is rarely known, can be predicted for a system’s phenotype of interest.

Fixed architecture	Variable parameters
<i>Topology of interconnections</i>	<i>Rate constants</i>
- Relatively easy to determine	- Difficult to determine in situ
- High throughput methods available	- No high throughput methods
<i>Signs of interactions</i>	<i>Binding Constants</i>
- Relatively easy to determine	- Difficult to determine in situ
- High throughput methods available	- No high throughput methods
<i>Numbers of Binding sites for the interactions</i>	<i>Environmental inputs</i>
- Small number of possibilities	- Many and difficult to know
- Sampling is feasible	- No high throughput methods
- High throughput methods available	

In the conventional approach the focus is *first* on measuring, estimating or sampling parameter values and fitting known experimental data. Only when a set of parameter values is in hand can one proceed to simulate the nonlinear system. This provides validation of the parameter set by demonstrating agreement with experimental data and, when there is disagreement, refinement of the model in an iterative fashion as part of the usual scientific method. Having a validated set of nominal parameter values, a parameterized model then allows one to explore parameter space and predict new phenotypes that were not used in the initial parameterization. Success is obtained if the predictions are subsequently confirmed by experimental tests. In this conventional strategy, simulating a system with a given set of parameter values is easy; obtaining the parameter values in the first place is hard.

In the phenotype-centric strategy the focus is first on *analytically* enumerating the phenotypic repertoire of the model and later *predicting* parameter values for phenotypes of interest. This is possible because biochemical kinetic models consist of both fixed and variable features (Lomnitz and Savageau, 2015) (Table 1). The fixed, or *architectural*, features include connectivity (e.g., protein-DNA

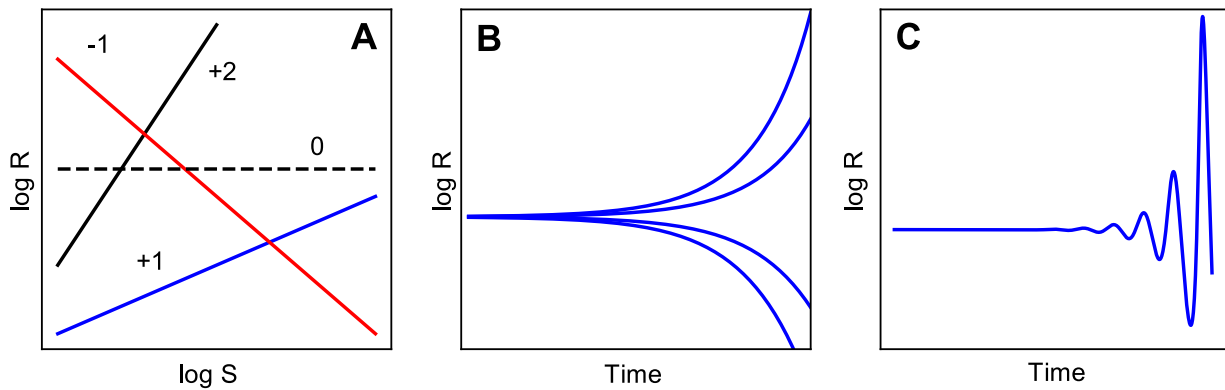


Fig. 2. Parameter-Independent Phenotypes of a System's Fixed Points. R represents an output variable (response) of an arbitrary system, while S represents a parameter or an input variable (signal) of that system. (A) Fixed steady-state logarithmic gains resulting from a (quasi-)steady state titration of the signal. (B) Exponential instability with one positive real eigenvalue, which typically leads to hysteretic bistability. (C) Oscillatory instability with a pair of complex conjugate eigenvalues having a positive real part, which typically leads to sustained limit-cycle oscillations.

binding), signs (e.g., activation/repression) and numbers (e.g., kinetic orders) of interactions for a given class of models. The variable, or *parametric*, features include kinetic and thermodynamic parameters and environmental variables that quantitatively distinguish the members of this class.

The architectural features give rise to important 'parameter-independent' phenotypes associated with the system's fixed points (Fig. 2). These characteristics, which are analytically determined using linear algebra, exist throughout a polytope region in parameter space that defines the phenotype. They can be matched to phenotypes of biological interest. Moreover, a linear program can then be used to predict a nominal set of parameter values. Thus, in this novel phenotypic-centric strategy, the initial effort is focused on analytically enumerating the phenotypic repertoire and then on predicting parameter values. This inverts the conventional strategy in which the initial focus is on estimating parameter values and then enumerating the phenotypic repertoire using simulation (Fig. 1). The differences between the two strategies for a model of phage λ induction are illustrated in Fig. 3 by the workflow and results obtained.

In the conventional simulation-centric strategy, as noted above, effort is focused initially on measuring, estimating or sampling values and fitting experimental data (Fig. 3C) and the result is a set of nominal values for the parameters (Fig. 3D). Once a parameterized model has been obtained, the effort then turns to an exploration of parameter space by dense sampling and simulation to predict new phenotypes (Fig. 3E) that were not used in the initial parameterization (Fig. 3F). Success is obtained if the predictions are subsequently confirmed by experimental tests.

In the phenotype-centric strategy, effort is focused initially on enumerating the phenotypic repertoire, without specifying parameter values, and the phenotypic repertoire that results is filtered for the phenotype(s) of experimental interest (Fig. 3G). A major advantage at this point is rapid model discrimination; if the phenotypic repertoire does not include the phenotypes of interest, then the model (hypothesis) can be rejected (Lomnitz and Savageau, 2016b). Alternatively, if the phenotypic repertoire does include the phenotypes exhibited by the system, then a representative set of parameter values can be predicted for the realization of each phenotype within an *appropriately localized region* of parameter space (Fig. 3H). The relationships among phenotypes for this set of parameters can then be visualized in 2-D slices through the system design space (Fig. 3I). A predicted progression of phenotypes resulting from the steady-state titration of a given parameter can generate a variety of composite phenotypes, such as activation followed by repression (Fig. 3J).

As this example shows, starting only with the architectural features and no parameter values, within minutes the phenotype-centric strategy obtained the results shown in Fig. 3G–J for a model of phage λ induction. Not only do these results qualitatively match those from the 'simulation-centric' strategy based on decades of experimental work to estimate model parameters (Savageau and Fasani, 2009), but the same repertoire of phenotypes is found with both strategies (Fig. 3E & I). The quasi-steady state concentration of *CI* mRNA (Fig. 3A) in response to increasing levels of RecA activity (a proxy for DNA damage) also is qualitatively similar for the two strategies (Fig. 3F & J).

These results suggest that a phased combination of the two strategies offers distinct advantages. The first phase, provided by the phenotype-centric strategy, is the most efficient when parameter values are unknown; it quickly yields qualitatively appropriate phenotypes and a full set of analytically predicted parameter values. The second phase, consisting of focused experiments and deterministic and stochastic numerical simulations (e.g., Fasani and Savageau, 2013), can then be used to verify and refine parameter values; it yields quantitative as well as qualitative results. Anytime definitive values are available for any of the parameters, these should be incorporated into the model before starting, as this will improve the efficiency of either strategy.

An important result of the previous analysis was the prediction of a system design principle for phage λ to maintain its biphasic life style (Savageau and Fasani, 2009); it consists of two inequalities that involve constellations of values for all the parameter of the model (Fig. 3B):

$$\frac{\gamma_{M \max}^2 \gamma^2 c \gamma_D}{\delta_M^2 \delta_C^2 \max \delta_D} \frac{\delta_D}{(\beta_D + \delta_D)} < K_D < \frac{\sqrt{\gamma_{M \max} \gamma_{M \min}}}{\delta_M} \frac{\gamma_C}{2 \delta_D} \quad (3)$$

Moreover, when the experimentally determined (Savageau and Fasani, 2009) and predicted parameter values are substituted into the inequalities one finds that the design principle is satisfied in both cases.

4. System design principles for a common regulatory architecture

In this section we examine a very common molecular architecture involving positive and negative feedback (Fig. 4). Bistable-hysteretic switches are generated by **positive feedback**. They are common features of commitment in cell-fate determination in viruses including prophage induction (Dodd et al., 2001) and restriction-modification systems (Williams et al., 2013), in bacteria including toxin-antitoxin systems of *Escherichia coli* (Fasani and

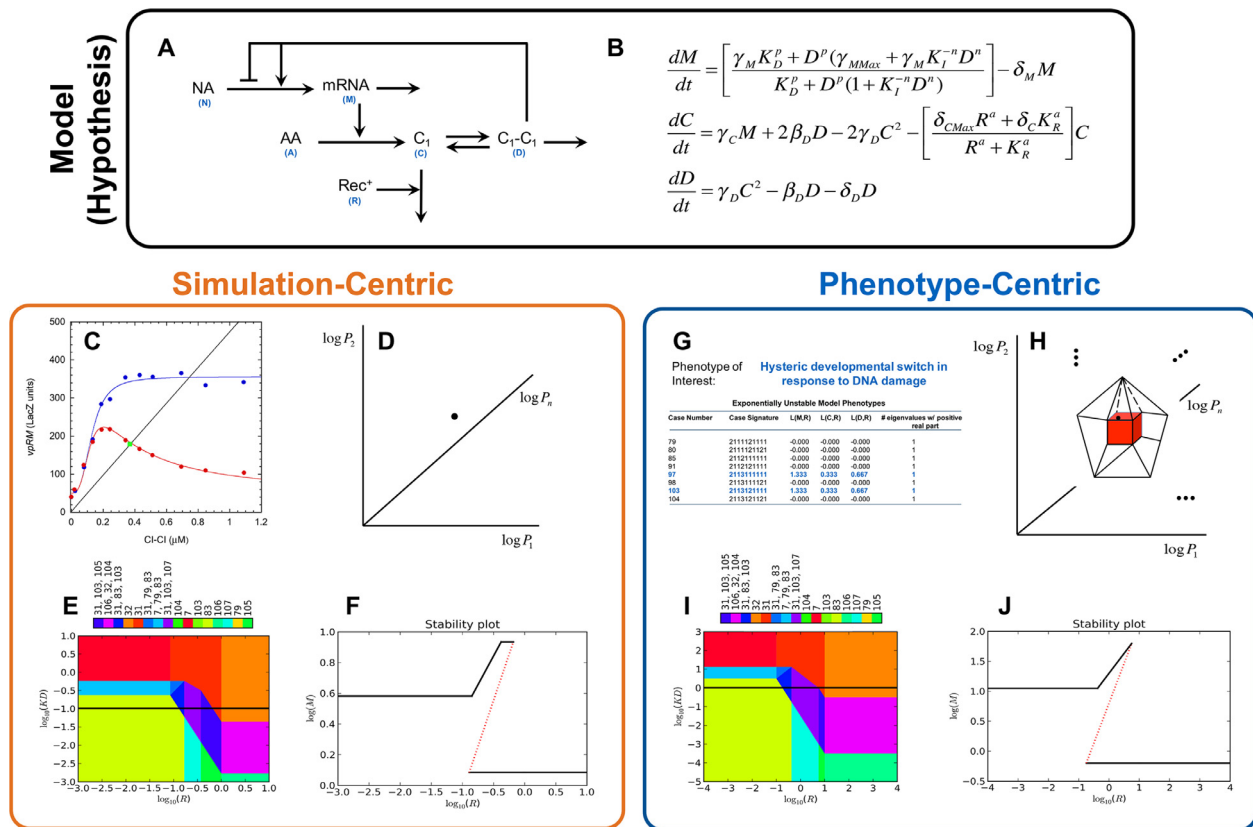


Fig. 3. Workflow from Conceptual Model to Predicted Induction Characteristic. (A) Model architecture includes the interaction network, signs of interactions, and number of binding events in the interactions. Blue characters in parenthesis are used to construct the mathematical model shown in the neighboring panel. (B) Mathematical model consisting of chemical and biochemical kinetic equations. Simulation-centric strategy: (C) Decades of work experimentally measuring and computationally estimating values for the parameters of the model in (A). (D) The resulting nominal set of parameter values. (E) The phenotypic repertoire can be obtained by dense sampling the parameter space and simulation; however, for our purposes here, we use the DST2 software with the experimentally determined values for the parameters to visualize the distinct phenotypic regions (See also Fig. 5). (F) The steady-state induction characteristic is predicted for various values of the input variable, RecA activity in the case of phage λ induction. Phenotype-centric strategy: (G) The phenotypic repertoire is enumerated without specifying values for the kinetic and thermodynamic parameters, and the list can be filtered to obtain only the phenotypes of interest (blue). (H) Parameter values are predicted automatically for each qualitatively-distinct phenotype of interest to localize estimates within a 'chunk' of parameter space. (I) The phenotypic repertoire can be visualized without sampling by taking slices through the high-dimensional object in the system design space (See also Fig. 5). (J) The steady-state induction characteristic is predicted for various values of the input variable, which is RecA activity.

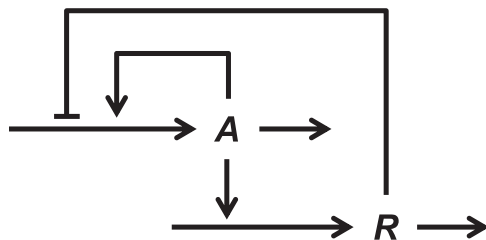


Fig. 4. Genetic Network Involving an Activator and a Repressor. The synthesis of the activator molecule A undergoes an autocatalytic activation and a repression by the repressor molecule R, whose synthesis is in turn activated by A.

Savageau, 2013) and induction preferences in catabolically diverse *Pseudomonas putida* (Nichols and Harwood, 1995; Rojo, 2010), in plants including asymmetric stem cell division of *Arabidopsis thaliana* (Cruz-Ramírez et al., 2012), and in animals including tracheal cell specification from a field of progenitor cells of *Drosophila melanogaster* (Metzger and Krasnow, 1999; Zelzer and Shilo, 2000) and neural progenitor cells switching to oligodendroglia in the brains of *Rattus norvegicus* (Lai et al., 2004). Homeostatic regulation is generated by **negative feedback**, and under certain conditions it also can generate oscillations (Elowitz and Leibler, 2000);

however, more robust oscillations are obtained with a combination of **positive and negative feedback** (Lomnitz and Savageau, 2014; Novák and Tyson, 2008; Purcell et al., 2010; Tsai et al., 2008). This architecture is at the core of circadian clocks found in organisms including cyanobacteria (Tomita et al., 2005), flies (Hardin, 2011), plants (Nohales and Kay, 2016), and mammals (Papazyan et al., 2016). Additionally, it is at the core of many synthetic gene oscillators (Atkinson et al., 2003; Stricker et al., 2008; Tigges et al., 2009) that provide a simplified and experimentally tractable context for study.

The examination of the network illustrated in Fig. 4 will provide a more detailed treatment of the various steps in a design space analysis and demonstrate how this type of analysis can be used to elucidate underlying design principles that would otherwise be difficult if not impossible to discover by intuition or tractable experiments.

A typical model with the equations for mRNA dynamics assumed to be fast and their quasi-steady state values incorporated into the slower equations for protein dynamics involves rational functions for the synthesis of the repressor (R) and activator (A) in the following equations:

$$\frac{1}{\beta_R} \frac{dR}{dt} = \gamma_R \frac{1}{1 + \left(\frac{A}{K_A}\right)^n} - R \quad \rho_R > 1 \quad (4)$$

$$\frac{1}{\beta_A} \frac{dA}{dt} = \gamma_A \frac{1 + \left(\frac{A}{K_{AA}}\right)^n + \frac{1}{\rho_A} \left(\frac{R}{K_R}\right)^n}{1 + \left(\frac{A}{K_{AA}}\right)^n + \left(\frac{R}{K_R}\right)^n} - A \quad \rho_A > 1 \quad (5)$$

Note that the synthesis of **R** is modeled as an activator-primary process (Lomnitz and Savageau, 2014), meaning that in the absence of **A**, **R** is constitutively synthesized and exhibits a minimum steady state concentration of γ_R/ρ_R . In the presence of the activator molecule **A**, the steady state concentration of **R** can be increased to reach a maximum value of γ_R , as shown in Eq. (4). On the other hand, the synthesis of **A** is modeled as a repressor-primary process (Lomnitz and Savageau, 2014), meaning that in the absence of both **A** and **R**, **A** is constitutively synthesized and exhibits a maximum steady state concentration of γ_A . At a sufficiently high concentration of **R**, the synthesis of **A** is repressed, and its steady state concentration reaches a minimum value of γ_A/ρ_A . The autocatalytic activation of **A** reduces the repression by **R**, as shown in Eq. (5).

Recasting Eqs. (4) and (5) into the GMA form yields:

$$\frac{1}{\beta_R} \frac{dR}{dt} = \gamma_R \rho_R^{-1} D_R^{-1} + \gamma_R A^n K_{AA}^{-n} D_R^{-1} - R \quad (6)$$

$$\frac{1}{\beta_A} \frac{dA}{dt} = \gamma_A D_A^{-1} + \gamma_A A^n K_{AA}^{-n} D_A^{-1} + \gamma_A \rho_A^{-1} R^n K_R^{-n} D_A^{-1} - A \quad (7)$$

$$0 = 1 + A^n K_{AA}^{-n} - D_R \quad (8)$$

$$0 = 1 + A^n K_{AA}^{-n} + R^n K_R^{-n} - D_A \quad (9)$$

Note that Eqs. (8) and (9) are algebraic constraints introduced during the recasting process and define auxiliary variables D_A and D_R , respectively.

Eqs. (6)–(9) can be automatically analyzed within the design space formalism. To that end, we use the Design Space Toolbox V2 (DST2), software that allows for the automatic enumeration of the phenotypic repertoire, the prediction of phenotype-specific parameter values and the characterization of model phenotypes using analytical and numerical methods (Lomnitz and Savageau, 2016a). Depending on the focus of the analysis being performed, a number of analytical workflows are possible. Here, we will show how the DST2 can be used to identify parameter values corresponding to a phenotype with a desired dynamic behavior and elucidate its design principles.

4.1. Parameter-independent characteristics of the phenotypic repertoire can guide the identification of regions in the parameter space with desired dynamic behavior

The first operation typically is to enumerate the full repertoire of phenotypes along with some of their phenotypic characteristic of interest by using the ‘create cases table’ command. Table 2 lists the phenotypic repertoire corresponding to the genetic network depicted in Fig. 4 along with the number of eigenvalues with positive real part for each S-system. There is a total of 36 potential S-systems, from which only 15 are valid. As an example of a potential S-system that is invalid consider cases involving the four combinations of terms in Eqs. (6) and (8). Combinations in which the first term in Eq. (6) and the second term in Eq. (8) are the dominant positive terms would require $1 < A^n K_{AA}^{-n} < \rho_R^{-1} < 1$, which is mathematically impossible.

A closer inspection of Table 2 reveals the existence of two systems with the potential to exhibit bistability (Cases 9 and 30) and one system with the potential to exhibit an oscillatory behavior (Case 27). When the repertoire is very large, it is useful to filter the list with various criteria that are part of the enumeration command. For example, in this case we could have filtered the list for

Table 2

Phenotypic Repertoire. The DST2 allows for the automatic enumeration of all potential S-systems of a given network. In the case of the genetic network shown in Fig. 4, there are a total of 36 potential S-systems, of which 15 are valid. Each S-system has a case number and a uniquely defined case signature that identifies the dominant terms in each equation (Fasani and Savageau, 2010). As indicated by the number of eigenvalues with positive real part, the S-systems with Case numbers 9 and 30 have the potential to exhibit bistability, whereas the S-system with Case number 27 has the potential to exhibit oscillatory behavior.

Case number	Case signature	Number of eigenvalues with positive real part
1	111111	0
...	...	...
9	11211131	1
...	...	...
27	21211131	2
29	21212121	0
30	21212131	1
...	...	...

only those phenotypes that have 2 eigenvalues with positive real part; the list that is returned would then contain only Case 27.

Note that each case (S-system) is associated with a specific high dimensional polytope in parameter space, whose boundaries can be readily calculated by means of linear programming. These polytopes are then fit together to fill the parameter space for visualization as two-dimensional slices (Fig. 5A). Various characteristics of the phenotypes can then be plotted as a heat-map in the third dimension (Fig. 5C).

Since DST2 allows for the prediction of a complete parameter set representative of each S-system, the dynamical behavior of the whole system parameterized with this parameter set can be compared with the dynamic behavior of the corresponding S-system. This procedure is shown in Fig. 5 for S-system 27 (Fig. 5B), which should exhibit oscillatory behavior, and for S-system 30 (Fig. 5D), which should exhibit bistability. Phenotypes having two complex conjugate eigenvalues with positive real part need not exhibit sustained oscillations throughout the associated polytope region, as the conditions are partly dependent on parameter values; however, with the architecture of the model in Fig. 4 the oscillations are particularly robust. Necessary conditions for sustained oscillations have been described in detail elsewhere (Lomnitz and Savageau, 2014). In contrast, phenotypes having one eigenvalue with positive real part exhibit exponential instability throughout the associated polytope region, regardless of parameter values.

Thus, a useful initial strategy for design space analysis, as we have seen, involves using the ‘create cases table’ command to enumerate the phenotypic repertoire, identify phenotypes with desired steady-state and/or dynamic properties, and predict parameter values for their realization.

4.2. Visual inspection of the parameter space reveals qualitative design principles

Visualization of a high-dimensional design space remains a challenge, but it offers opportunities to identify system design principles that would be difficult if not impossible to achieve by other means. These issues are currently addressed by using the ‘create plot’ command to visualize 2D slices of the space with the parameters on the axes selected for their particular biological interest. The dimensions of the design space (number of parameters) can be reduced by introducing dimensionless parameters, which group the system parameters into a lower number of different terms (Savageau et al., 2009). In the specific case of the network shown in Fig. 4, and mathematically described by Eqs. (4) and (5), the numerical values for three parameters, namely K_A , K_{AA} and K_R are required to fully characterize the strength of interactions

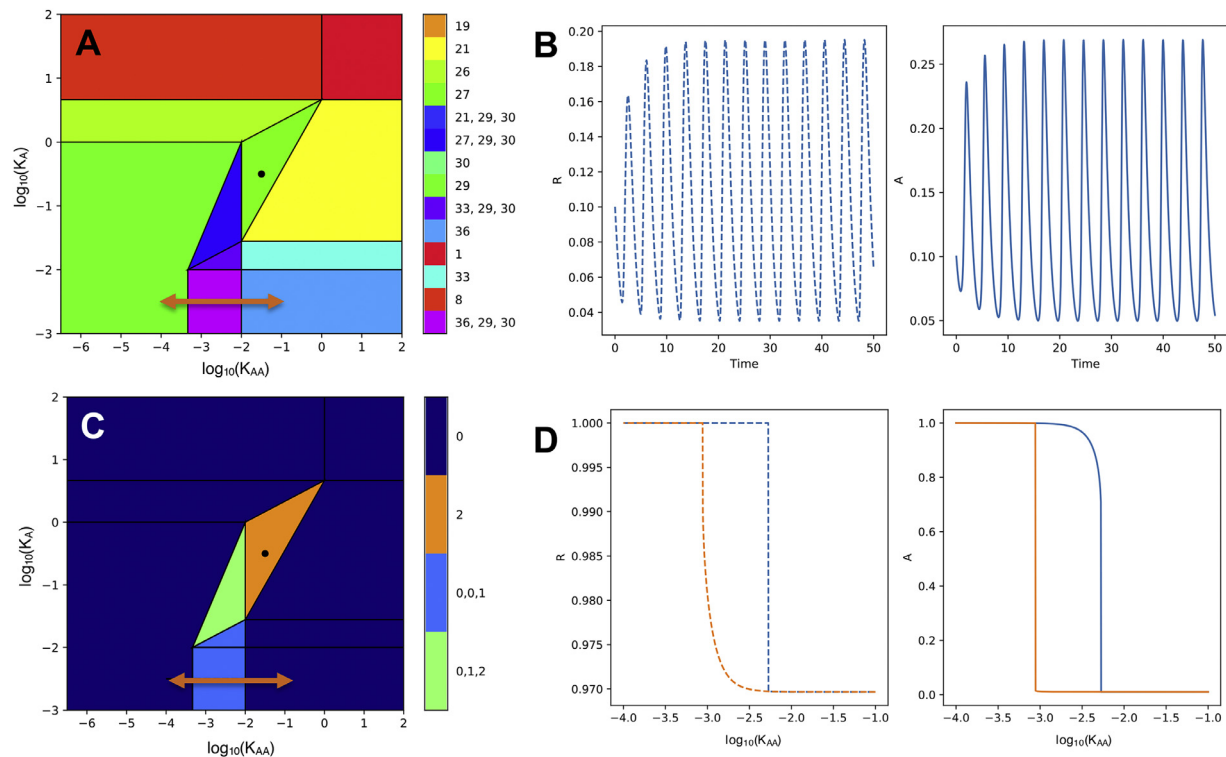


Fig. 5. Design Space and Dynamic Behavior of the System. Panel A is the system design space showing the arrangement and shape of phenotypic regions identified by color and case number when $K_R = 10^{-2}$. Note that instances with three case numbers represent overlapping phenotypic regions that typically signify hysteretic bistable regions. Panel C is a stability plot represented in three dimensions with the values of K_{AA} and K_A on the x- and y-axis, and the number of eigenvalues with positive real part as a color map on the z-axis. The polytopes for Cases 27 and 30 in Panel A have their stability represented in Panel C. The oscillatory behavior in Panel B is generated by the whole system having the parameter set shown as the black dot. The hysteretic bistable response in Panel D, when $K_A = 10^{-2.5}$, is generated by the whole system in response to changes in K_{AA} . Blue lines represent the system response with K_{AA} increasing from low to high values, while orange lines represent the response for decreasing K_{AA} from high to low values. The qualitative behavior is determined by the three binding constants K_R , K_A , and K_{AA} . The other parameters do not affect the qualitative behavior: the constants γ_R and γ_A determine the concentration scales for R and A and are arbitrarily set to one; the capacities for regulation ρ_R and ρ_A are required to be greater than one and are arbitrarily set to 100, which is typical for the best studied transcription factors of *E. coli*, and the kinetic order n was set to a value of 3 to allow the system to exhibit sustained oscillations. Necessary conditions for sustained oscillations involve a trade-off between cooperativity and delay and are described in detail elsewhere (Lomnitz and Savageau, 2014). For the network under study, cooperativity values greater than 2 are necessary for sustained oscillations.

between the repressor **R** and activator **A**. In terms of the design space, this means that the phenotypic repertoire of the network can be assembled into a space-filling, three-dimensional structure. By setting K_R to different constant values, the shape of the phenotypic regions in this three-dimensional structure can be visualized in a plane, as shown in Fig. 6A. Note that the use of K_R to “cut” the design space is arbitrary and either K_A or K_{AA} would represent equally valid alternatives.

The DST2 allows for a convenient, interactive visualization of the design space as the value of a given parameter, in this case K_R , is changed. By visually inspecting the shape of the design space (refer to Figs. 6B–D), one can rapidly identify three zones we might denote **L**, **M** and **H**, each exhibiting a qualitatively different design space and associated stability pattern. While within zone **H** all phenotypes are stable (Fig. 6G), zone **M** is characterized by the presence of an oscillatory phenotype (Case 27, parallelogram composed of green and orange regions in Figs. 6F) and a bistable phenotype (Case 30, composed of blue and green regions in Fig. 6F). In addition to these two dynamic phenotypes, zone **L** exhibits a further bistable phenotype (Case 9, upper blue and green region in Fig. 6E). The effect of K_A , K_{AA} and K_R on the dynamic behavior potentially exhibited by the system can be observed in Figs. 6E–G and can be summarized in Table 3.

4.3. System design principles and global robustness

One could vary the values of K_R by simple bisection to numerically estimate the threshold values that separate the three

Table 3

Three Different Regions of Design Space for Each of the Three Dynamic Phenotypes. Each region is defined by the effect of decreasing values of K_R on the area of each of the phenotypes in a two-dimensional plane defined by $\log K_A$ and $\log K_{AA}$ on the y- and x-axis respectively (see Fig. 6).

Dynamic phenotypes	Region		
	Maximum	Expanding	Non-existent
Lower Hysteresis (Case 30)	$\log K_R \leq -1$	$-1 < \log K_R < 0$	$\log K_R \geq 0$
Upper Hysteresis (Case 9)	$\log K_R \leq -3$	$-3 < \log K_R < -2$	$\log K_R \geq -2$
Oscillation (Case 27)	$\log K_R \leq -3$	$-3 < \log K_R < 0$	$\log K_R \geq 0$

qualitative regions in each case. However, this would not necessarily illuminate the underlying design principles governing this system. Instead, we could use the ‘analyze a case’ command in DST2 to find algebraic expressions for the boundaries of each phenotype. By following this procedure, we can identify system design principles involving all of the system parameters for each phenotype. For example, the result for the lower hysteretic polytope (phenotype 30) is given by

$$\left[\frac{K_R}{\gamma_R} \right]_{\max} \left\{ \left(\frac{K_R}{\gamma_R} \right)^{n-1}, \left(\frac{1}{\rho_A} \right)^{\frac{n-1}{n}}, \left(\frac{K_A}{\gamma_A} \right)^{\frac{n-1}{n}} \right\} < \frac{K_{AA}}{\gamma_A} < \left[\frac{K_R}{\gamma_R} \right] \quad (10)$$

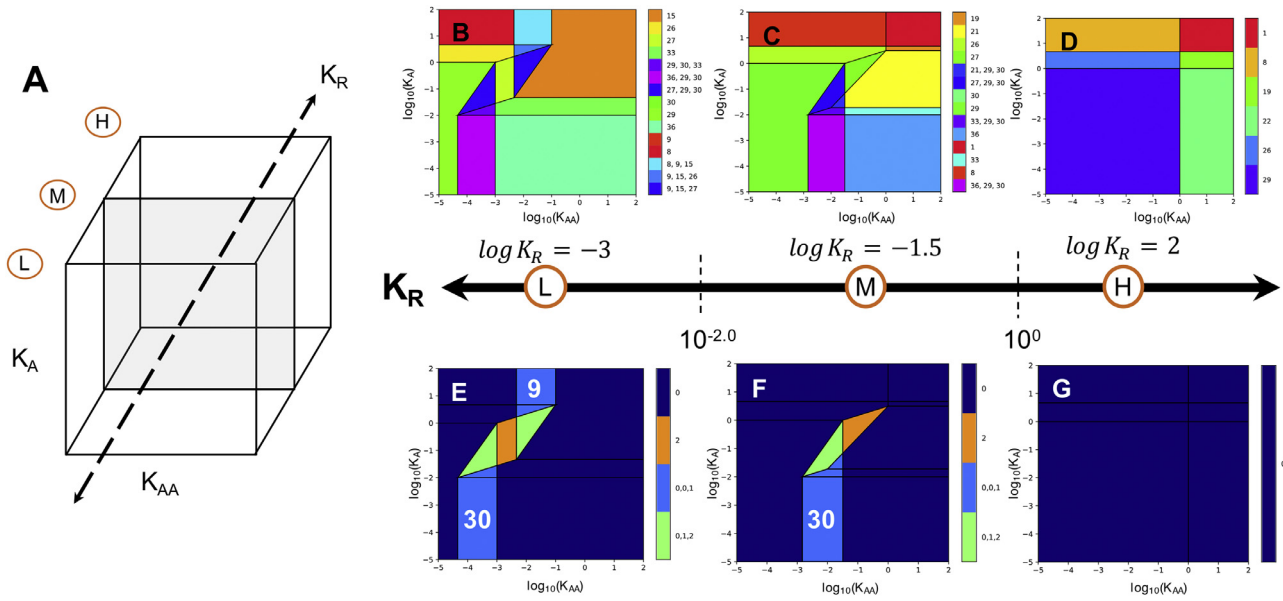


Fig. 6. Representation of the Design Space in a Series of Two-dimensional Plots. The design space is spanned by three axes, each containing numerical values for one parameter. We use K_R to generate slices of the three-dimensional structure (A). Panels B to D show the arrangements and shapes of phenotypic regions identified by color and case number as the parameter K_R takes on values of 10^{-3} , $10^{-1.5}$ and 10^0 , respectively. Panels E to G are stability plots represented in three dimensions with the values of K_{AA} and K_A on the x- and y-axis, and the number of eigenvalues with positive real part as a heat map on the z-axis. Other parameter values are as stated in Fig. 5.

that for upper hysteretic polytope (phenotype 9) is

$$\left[\frac{\rho_R K_R}{\gamma_R} \right]_{\max} \left\{ \left(\frac{\rho_R K_R}{\gamma_R} \right)^{n-1}, \left(\frac{1}{\rho_A} \right)^{\frac{n-1}{n}} \right\} < \frac{K_{AA}}{\gamma_A} < \left[\frac{\rho_R K_R}{\gamma_R} \right]_{\min} \left\{ \left(\frac{1}{\rho_R^{1/n} \gamma_A} \right)^{\frac{n-1}{n}}, 1 \right\} \quad (11)$$

and that for the oscillatory polytope (phenotype 27) is

$$\left[\left(\frac{\gamma_R}{K_R} \right) \left(\frac{K_{AA}}{\gamma_A} \right) \right]_{\max} \left\{ \left(\frac{1}{\rho_R} \right)^{\frac{n^2-n+1}{n-1}} \left(\frac{K_{AA} \gamma_R}{\gamma_A K_R} \right)^{\frac{n^2-n+1}{n-1}}, \left(\frac{K_{AA}}{\gamma_A} \right)^{\frac{n^2-n+1}{n}} \right\} < \left(\frac{K_A}{\gamma_A} \right)^n < \left[\left(\frac{\gamma_R}{K_R} \right) \left(\frac{K_{AA}}{\gamma_A} \right) \right]_{\min} \left\{ \left(\frac{K_{AA} \gamma_R}{\gamma_A K_R} \right)^{\frac{n^2-n+1}{n-1}}, 1 \right\} \quad (12)$$

Note that the parameters representing the maximum steady-state concentration of $\mathbf{R}$ (γ_R) and $\mathbf{A}$ (γ_A) determine the concentration scale, the binding constants (K_R , K_{AA} , K_A) have the units of concentration, and the capacities for regulation (ρ_R and ρ_A) and the kinetic order n are dimensionless. Necessary conditions for hysteresis and oscillation can be obtained from Eqs. (10)–(12) by canceling common terms on both sides of the inequalities. This approach also provides a means to characterize volume and shape of the polytope for each phenotype and thereby address the global robustness and evolvability of each phenotype.

There have been various attempts to define system robustness. Local parameter insensitivity is used most often, but this is unsatisfactory for nonlinear systems in which global behavior is not captured by the local derivative. It is more important to know how large a change in system parameters can be tolerated without a change in the qualitative phenotype of the system. This global robustness is more important but difficult to characterize. It depends critically on the underlying structure of the model in parameter

space; e.g., it can often exhibit long curvilinear shallow valleys that make parameter estimation difficult. The “volume” in parameter space that is characteristic of a particular behavior has often been proposed as a measure of global robustness. A given point in such a volume might exhibit large variations and still remain within the volume; the larger the volume, the more robust the behavior. However, there has been no generic method for determining these volumes. Dense sampling of parameter space is a brute force method, but it becomes impractical in high-dimensional spaces and visualization is always a problem.

The system design space strategy provides an analytical approach to the characterization of global robustness. The boundaries and vertices of the polytopes in design space allow for the calculation of polytope shape and volume. As seen in Table 3, the phenotypes can have smaller, more variable parts, as well as larger more fixed parts. In questions concerning robustness, the large fixed volumes are of more interest, particularly those that avoid the possibility of bifurcation to qualitatively different phenotypes; whereas for questions of evolvability, the smaller, more variable parts with easy access to these bifurcations might be of more interest. Although all parts of the phenotypic regions can be calculated, here we will focus on the large fixed parts that avoid bifurcations as a means to characterize the volumes most relevant for global robustness.

The volume of the polytope for phenotype 30 in its maximum region (Table 3) is given by

$$Volume = \left(\frac{n-1}{n} \right) (\log \rho_A) \left(\log \frac{\gamma_A}{\rho_A K_A^{\min}} \right) \left(\log \frac{\gamma_R}{\rho_A^{1/n} K_R^{\min}} \right) \quad (13)$$

that for phenotype 9 in its maximum region is

$$Volume = \left(\frac{n-1}{n} \right) (\log \rho_A) \left(\log \frac{\rho_A K_A^{\max}}{\rho_R^{1/n} \gamma_A} \right) \left(\log \frac{\gamma_R}{\rho_R \rho_A^{1/n} K_R^{\min}} \right) \quad (14)$$

and that for the oscillatory phenotype 27 in its maximum region, minus that portion of the volume that has the possibility to undergo a bifurcation and lose oscillation, is

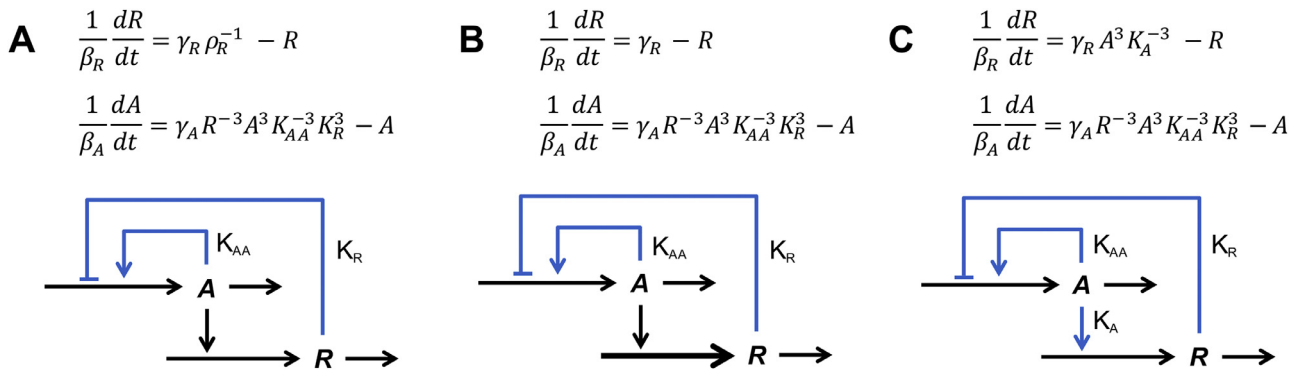


Fig. 7. Elucidating Structural Features of the Network and Their Effect on Dynamic Stability. Phenotypes 9 (A), 30 (B), and 27 (C). Mathematical expressions defining a given S-system can be interpreted in terms of network structure. By doing so, a desired dynamic behavior can be linked to certain structural properties of the network. For the genetic network depicted in Fig. 4 to exhibit bistability, the synthesis of repressor **R** must be independent of the activator **A** (Panels A and B). By contrast, this interaction is required for the same network to exhibit oscillation (C).

$$Volume = \left(\frac{n^2 - n + 1}{n^2} \right) (\log \rho_A) \left[(\log \rho_R) - \left(\frac{n-1}{n} \right) (\log \rho_A) \right] \left(\log \frac{\gamma_R}{\rho_R \rho_A^{1/n} K_R^{min}} \right) \quad (15)$$

The values for all the parameters that would locate the system's operating point furthest from the boundaries whose crossing would lead to a change in phenotype, in some sense the "optimal" parameter set for robustness, can be calculated as well (See Supplemental Material).

It also is interesting that the volume of all three phenotypes have one long major axis involving a linear relationship between K_{AA} and K_R (refer to Figure S3A in the Supplemental Material), which hints at some biophysical constraint involving the DNA binding region and the two transcription factors that must cooperate at this locus.

4.4. Linking structural features of the network with dynamical behavior

The process of phenotype construction by picking one dominant positive and one dominant negative term for each species (variable) can be interpreted in terms of network structure. This connection can be exploited by analyzing structural characteristics of phenotypes exhibiting desired dynamic behaviors, as shown in Fig. 7. Mathematical expressions defining Cases 9 (bistable phenotype, Fig. 7A), 30 (bistable phenotype, Fig. 7B) and 27 (oscillatory phenotype, Fig. 7C) are interpreted in terms of 'active' connections of the network, which are represented in blue. An interaction is considered active if the parameter defining the strength of interaction is contained in the mathematical expression for a given S-system. If the expression of a given dominant positive term does not contain any binding constant (K_R , K_A or K_{AA}), this is interpreted as a constitutive synthesis (refer to Fig. 7A and B). For simplicity and since a single negative term for each species is considered, only terms contributing to the synthesis of each pool are analyzed.

A closer examination of the network structure representing Case 9 and Case 30, both exhibiting bistability, reveals an interesting similarity (compare Fig. 7A and B). In terms of dynamic behavior, this network structure shows that the activation of the repressor molecule **R** by the activator **A** is not a necessary feature for the system to exhibit bistability. This holds for both a low (Case 9) and a high (Case 30) constitutive synthesis of **R**. Note that by contrast, this interaction is needed for the network to exhibit an oscillatory behavior (Case 27, Fig. 7C).

By unravelling the relationship between structural network features and associated dynamical behavior, the DST2 can support the design of synthetic networks that robustly exhibit a desired behavior. For instance, if one is exclusively interested in the hysteric switching capabilities of the network depicted in Fig. 4, the connectivity of the network should be modified so that the repressor molecule **R** is constitutively synthesized, independent from the concentration of the activator molecule **A**.

5. Logical description of gene regulatory networks

Qualitative logical network modeling has been widely applied to successfully describe topological properties of complex biological networks, including steady-state characterization and network robustness (Albert and Othmer, 2003; Li et al., 2004; Wang et al., 2012). While a kinetic description has the potential to fully capture the behavior of such complex systems in a quantitative way, its application to the analysis of gene regulatory networks has been somewhat limited due to lack of knowledge of both the detailed mathematical expressions describing regulatory interactions between network components and associated numerical values for their parameters (Wang et al., 2012).

Here, we present a simplified Boolean description of the network depicted in Fig. 4 and compare it with the kinetic representation offered by our phenotypic-centric modeling approach, which, as described throughout this paper, does not require previous knowledge of parameter values. We start our analysis with a set of Boolean functions representing the regulatory interactions within the network. The following assumptions were used to construct the logical rules: when a single input is involved (**A** influencing **R**), the Boolean function is a simple 'if-then'; when several inputs are involved (**A**, **R** influencing **A**), the Boolean function is typically treated as an 'if OR-then', which amounts to an assumption that the inputs act independently (Steinway et al., 2014). The alternative would be 'if AND-then'; e.g., the (Activator AND (NOT Repressor)) logic of the classical *lac* operon (Savageau, 2001). This set of functions was then used to generate a state transition graph, which represents the system dynamics in Boolean terms. For simplicity, we use a synchronous updating scheme. According to our naïve Boolean representation, the network under study exhibits a single stable fixed point (Fig. 8A).

According to our phenotypic-centric approach and as shown in Fig. 6, the entire phenotypic repertoire of the network can be graphically represented by three different slices at constant K_R values. Instead of focusing on the eigenvalues as the phenotype characteristic of interest plotted on the z-axis (Fig. 6E–G), one can consider the steady-state concentrations of **R** and **A** as the phenotype

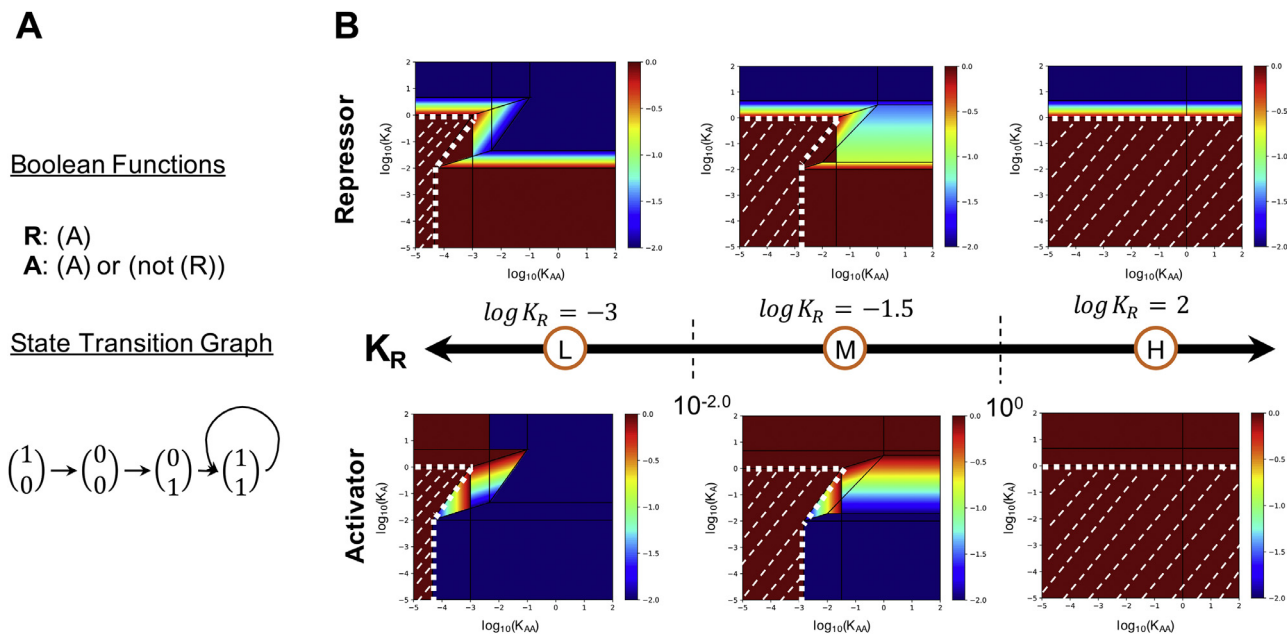


Fig. 8. Characterization of the Steady State Concentrations of the Genetic Network. (A) naïve Boolean characterization for the network depicted in Fig. 4, assuming a synchronous updating scheme. (B) Kinetic characterization offered by our phenotypic-centric approach. All axes are represented in a logarithmic scale, including steady state concentrations for the activator A and repressor R. Red regions correspond to steady state concentrations of 1, (which is represented as zero in the heat map, since $\log(1) = 0$), while blue regions represent steady state concentrations of 0.01 (which is represented as -2.0 in the heat map). Phenotypic regions marked by white dashed lines exhibit a single stable steady state $\begin{pmatrix} R \\ A \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, coinciding with the Boolean characterization of the network. Note that many of the phenotypic regions remain inaccessible to our naïve Boolean characterization. All plots in Figure 8B were generated using the DST2 with the parameter values as stated in Fig. 5.

characteristics of interest and plot them on the z-axis in the design space (Fig. 8B).

It is important to stress that no a priori knowledge of parameter values is needed to generate the phenotypic repertoire of the system. Rather, regions in parameter space for which each phenotype is valid are identified *before parameter values are predicted*. Since the design space is a space-filling, high-dimensional structure, each point in the parameter space can be assigned to a certain phenotype, thus an extensive coverage of properties of interest of the network under study is guaranteed.

As shown in Fig. 8B and depending on the specific values of the parameters, the network under analysis can potentially exhibit all possible mono-stable fixed points as well as bi-stable cases. By contrast, the naïve Boolean description of the same network leads to the identification of a single steady state, missing a variety of different steady states.

The naïve, classical Boolean description presented here can of course be vastly refined. For instance, the logical formalism developed by René Thomas and coworkers involved an asynchronous updating scheme and logical variables (Thomas and Kaufman, 2001). Further extensions of the classical Boolean formalism might include the introduction of logical variables with more than two values and the use of logical parameters (Snoussi, 1989). As the logical formalism becomes more elaborated, richer dynamic behaviors can be described. However, as stated by Thomas and Kaufman, (2001): “Clearly, depending on the values of the logical parameters, one can have a variety of situations, from a single steady state without periodicity to a choice between a stable steady state and a cycle.” In other words, it seems that an increasingly complex logical formalism partly renounces the promise of parameter-free modeling in order to realistically describe complex biological phenomena. The authors elegantly solve this issue by identifying a set of conditions on the logical parameters that are in line with experimental observations for the system being modeled, i.e., number and nature of steady states (Thieffry and Thomas, 1995).

6. Discussion

Living organisms can exhibit a variety of highly complex behaviors. Processes of central importance to life such as cellular growth, division and differentiation are possible thanks to the regulation and coordination provided by gene regulatory networks. Mathematical modeling offers a means to quantitatively describe and understand these processes, whose complexity cannot be fully understood solely by biological intuition. Once the network has been elucidated and represented using a certain mathematical formalism, one can predict previously unobserved phenotypes, design experimental interventions to test these predictions, and gain insights into the underlying biological design principles. Traditionally, regulatory networks have been mathematically represented as a dynamic system using a kinetic or a logical formalism, where the steady states or attractors of the system are interpreted as observable biological states (Li et al., 2004). The decision whether to use a kinetic or logical (Boolean) formalism has been so far mainly determined by the availability of parameter values required by the kinetic (differential) formalism.

Our phenotypic-centric modeling strategy offers a mechanistic, kinetic based modeling framework, while rendering the a priori knowledge of associated kinetic parameters unnecessary. In the specific case of the induction decision made by phage λ , we showed that our novel phenotypic-centric modeling strategy allowed the rapid identification of a region in the system design space that represents a hysteretic bistable switch involved in the commitment to induction. By using a representative set of parameter values within the identified region, we were able to construct a steady-state induction characteristic (one-parameter bifurcation diagram, or signal response curve), which was very similar to the same diagram obtained with the traditional simulation-centric approach using experimentally measured parameter values as input.

We illustrated additional features of our phenotypic-centric modeling strategy by analyzing a common regulatory architecture

consisting of positive and negative feedback. First, we showed how the phenotypic repertoire can be automatically enumerated and filtered using the ‘create cases table’ command to identify phenotypes of specific interest. We filtered for the stability signatures of hysteretic switches and oscillators. In other contexts, one can filter for a combination of logic functions; e.g., the signs of logarithmic gains representing increase (+) or decrease (−) in various output variables in response to a change in specific input variables or parameters (Lomnitz and Savageau, 2016b). Then, we demonstrated how the ‘create plot’ command can be used to deconstruct the system design space by a series of 2D slices, three in this specific case, and the entire repertoire of system phenotypes and their associated characteristics (stability, steady state concentration, etc.) can be visualized using these two-dimensional plots. For this common architecture, we were able to rapidly characterize its dynamic behavior (Fig. 6) and steady state concentrations (Fig. 8) as a function of phenotypic regions covering the design space. Note that instead of just sampling different specific parameter values, our phenotype-centric modeling approach allows an *exhaustive* characterization of the complete design space.

In addition to the qualitative design principles that were immediately visualized by means of an ‘interactive plot’ command, more refined system design principles were uncovered by making use of the ‘analyze a case’ command to calculate boundaries and vertices of the polytope characteristic of each phenotype. This information determines maximal fixed volumes, which can be used as a criterion of global robustness for each phenotype, as well as smaller variable volumes for characterizing evolvability. It is important to note that these volumes are analytically determined, not by sampling parameter values, and that they capture very asymmetrical nonlinear shapes in a tractable linear form in logarithmic coordinates (see Supplemental Material).

We also show that *structural features* of the network are associated with particular dynamic responses (see Fig. 7). For instance, we demonstrated that the activation of the synthesis of the repressor **R** by the activator **A** is not necessary for the network to exhibit bistability. Rather, the repressor should be constitutively synthesized for the network to exclusively exhibit a bistable response. By contrast, oscillatory behavior is only possible when all regulatory interactions contained within the regulatory architecture are active.

Logical formalisms have been widely used to study gene regulatory networks because no parameter values are required. Here, we assessed the performance of a classical, *naïve* Boolean description of the network architecture in Fig. 4 by comparing the characterization of the steady state response offered by this formalism with the parameter-free, kinetic characterization provided by our phenotypic-centric approach. We could show that the coverage of the naïve Boolean description was rather limited, since it failed to describe the rich variety of steady states potentially exhibited by the network (Fig. 8). As identified by René Thomas and co-workers, the analysis of a more elaborated logical formalism, which should allow the description of richer dynamical responses, necessarily involves the consideration of parameter values. Interestingly, the strategy developed by these authors to deal with this issue involved the consideration of all possible dynamic responses of the network, along with the determination of associated parameter values. By constraining this response ‘repertoire’ with known biological behaviors, it is possible to identify consistent values for the logical parameters (Thieffry and Thomas, 1995). In our opinion, this strategy resembles in many respects our phenotype-centric approach, in which the phenotypic repertoire of the model is first enumerated and then a nominal set of parameter values for the realization of a phenotype of interest can be easily calculated. The fact that similar strategies for the analysis of a complex logical mathematical description were independently developed suggests

that the idea of phenotypic-centric modeling might be extended to other mathematical formalisms.

It should be clear that the phenotype-centric modeling strategy does not supplant the traditional simulation-centric approach, particularly in the case of stochastic systems, but rather is a complementary strategy. Indeed, the phenotype-centric approach can be an aid to these other approaches. By making use of the “chunked up” design space to identify relevant regions of parameter space and predict parameter values, the phenotype-centric strategy provides initial estimates of parameter values that can be subsequently refined by a host of well-known numerical methods. For example, Newton’s method for finding fixed points behaves chaotically unless it has good starting values; however, if these are available, then it has nice quadratic convergence properties.

The development of the phenotype-centric approach is still in the early stages and there are several issues related to software development that need to be addressed. However, it is already clear that this approach has a solid theoretical foundation and offers new vistas for the analysis of complex gene regulatory networks.

Acknowledgments

We thank M. Singer and J. Lomnitz for fruitful discussions regarding the challenges and practical applications of design space. This work was supported in part by a grant from the US National Science Foundation Grant number MCB 1716833.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2018.07.009.

References

- Albert, R., Othmer, H.G., 2003. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J. Theor. Biol.* 223, 1–18. [https://doi.org/10.1016/S0022-5193\(03\)00035-3](https://doi.org/10.1016/S0022-5193(03)00035-3).
- Atkinson, M.R., Savageau, M.A., Myers, J.T., Ninfa, A.J., 2003. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* 113, 597–607. [https://doi.org/10.1016/S0092-8674\(03\)00346-5](https://doi.org/10.1016/S0092-8674(03)00346-5).
- Brenner, S., 2000. The end of the beginning. *Science* 287, 2173–2174. <https://doi.org/10.1126/science.287.5461.2173>.
- Chelliah, V., Laibe, C., Le Novère, N., 2013. BioModels database: a repository of mathematical models of biological processes. *Methods Mol. Biol. Clifton NJ* 1021, 189–199. https://doi.org/10.1007/978-1-62703-450-0_10.
- Coelho, P.M.B.M., Salvador, A., Savageau, M.A., 2009. Quantifying global tolerance of biochemical systems: design implications for moiety-transfer cycles. *PLoS Comput. Biol.* 5, e1000319. <https://doi.org/10.1371/journal.pcbi.1000319>.
- Cruz-Ramírez, A., Díaz-Triviño, S., Blilou, I., Grieneisen, V.A., Sozzani, R., Zamioudis, C., Miskolczi, P., Nieuwland, J., Benjamins, R., Dhonukshe, P., Caballero-Pérez, J., Horvath, B., Long, Y., Mähönen, A.P., Zhang, H., Xu, J., Murray, J.A.H., Benfey, P.N., Bako, L., Marée, A.F.M., Scheres, B., 2012. A bistable circuit involving SCARECROW-RETINOBLASTOMA integrates cues to inform asymmetric stem cell division. *Cell* 150, 1002–1015. <https://doi.org/10.1016/j.cell.2012.07.017>.
- Dodd, I.B., Perkins, A.J., Tsemitsidis, D., Egan, J.B., 2001. Octamerization of λ CI repressor is needed for effective repression of PRM and efficient switching from lysogeny. *Genes Dev.* 15, 3013–3022. <https://doi.org/10.1101/gad.937301>.
- Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338. <https://doi.org/10.1038/35002125>.
- Fasani, R.A., Savageau, M.A., 2013. Molecular mechanisms of multiple toxin-antitoxin systems are coordinated to govern the persister phenotype. *Proc. Natl. Acad. Sci.* 110, E2528–E2537. <https://doi.org/10.1073/pnas.1301023110>.
- Fasani, R.A., Savageau, M.A., 2010. Automated construction and analysis of the design space for biochemical systems. *Bioinf. Oxf. Engl.* 26, 2601–2609. <https://doi.org/10.1093/bioinformatics/btq479>.
- Hardin, P.E., 2011. Molecular genetic analysis of circadian timekeeping in *Drosophila*. *Adv. Genet.* 74, 141–173. <https://doi.org/10.1016/B978-0-12-387690-4.00005-2>.
- Lai, K., Robertson, M.J., Schaffer, D.V., 2004. The sonic hedgehog signaling system as a bistable genetic switch. *Biophys. J.* 86, 2748–2757. [https://doi.org/10.1016/S0006-3495\(04\)74328-3](https://doi.org/10.1016/S0006-3495(04)74328-3).
- Li, F., Long, T., Lu, Y., Ouyang, Q., Tang, C., 2004. The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci.* 101, 4781–4786. <https://doi.org/10.1073/pnas.0305937101>.
- Lomnitz, J.G., Savageau, M.A., 2016a. Design space toolbox V2: automated software enabling a novel phenotype-centric modeling strategy for natural and synthetic biological systems. *Front. Genet.* 7, 118. <https://doi.org/10.3389/fgene.2016.00118>.

- Lomnitz, J.G., Savageau, M.A., 2016b. Rapid discrimination among putative mechanistic models of biochemical systems. *Sci. Rep.* 6, 32375. <https://doi.org/10.1038/srep32375>.
- Lomnitz, J.G., Savageau, M.A., 2015. Elucidating the genotype–phenotype map by automatic enumeration and analysis of the phenotypic repertoire. *Npj Syst. Biol. Appl.* 1, 15003.
- Lomnitz, J.G., Savageau, M.A., 2014. Strategy revealing phenotypic differences among synthetic oscillator designs. *ACS Synth. Biol.* 3, 686–701. <https://doi.org/10.1021/sb500236e>.
- Metzger, R.J., Krasnow, M.A., 1999. Genetic control of branching morphogenesis. *Science* 284, 1635–1639.
- Nichols, N.N., Harwood, C.S., 1995. Repression of 4-hydroxybenzoate transport and degradation by benzoate: a new layer of regulatory control in the *Pseudomonas putida* beta-ketoadipate pathway. *J. Bacteriol.* 177, 7033–7040.
- Nohales, M.A., Kay, S.A., 2016. Molecular mechanisms at the core of the plant circadian oscillator. *Nat. Struct. Mol. Biol.* 23, 1061–1069. <https://doi.org/10.1038/nsmb.3327>.
- Novák, B., Tyson, J.J., 2008. Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* 9, 981–991. <https://doi.org/10.1038/nrm2530>.
- Papayyan, R., Zhang, Y., Lazar, M.A., 2016. Genetic and epigenomic mechanisms of mammalian circadian transcription. *Nat. Struct. Mol. Biol.* 23, 1045–1052. <https://doi.org/10.1038/nsmb.3324>.
- Purcell, O., Savery, N.J., Grierson, C.S., di Bernardo, M., 2010. A comparative analysis of synthetic genetic oscillators. *J. R. Soc. Interface* 7, 1503–1524. <https://doi.org/10.1098/rsif.2010.0183>.
- Rojó, F., 2010. Carbon catabolite repression in *Pseudomonas*: optimizing metabolic versatility and interactions with the environment. *FEMS Microbiol. Rev.* 34, 658–684. <https://doi.org/10.1111/j.1574-6976.2010.00218.x>.
- Savageau, M.A., 2013. Chapter 15 - phenotypes and design principles in system design space. In: Walhout, A.J.M., Vidal, M., Dekker, J. (Eds.), *Handbook of Systems Biology*. Academic Press, San Diego, pp. 287–310. <https://doi.org/10.1016/B978-0-12-385944-0.00015-0>.
- Savageau, M.A., 2001. Design principles for elementary gene circuits: elements, methods, and examples. *Chaos Woodbury N* 11, 142–159. <https://doi.org/10.1063/1.1349892>.
- Savageau, M.A., Coelho, P.M.B.M., Fasani, R.A., Tolla, D.A., Salvador, A., 2009. Phenotypes and tolerances in the design space of biochemical systems. *Proc. Natl. Acad. Sci.* 106, 6435–6440. <https://doi.org/10.1073/pnas.0809869106>.
- Savageau, M.A., Fasani, R.A., 2009. Qualitatively distinct phenotypes in the design space of biochemical systems. *FEBS Lett.* 583, 3914–3922. <https://doi.org/10.1016/j.febslet.2009.10.073>.
- Savageau, M.A., Voit, E.O., 1987. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Math. Biosci.* 87, 83–115. [https://doi.org/10.1016/0025-5564\(87\)90035-6](https://doi.org/10.1016/0025-5564(87)90035-6).
- Snoussi, E.H., 1989. Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach. *Dyn. Stab. Syst.* 4, 565–583. <https://doi.org/10.1080/02681118908806072>.
- Steinway, S.N., Zañudo, J.G.T., Ding, W., Rountree, C.B., Feith, D.J., Loughran, T.P., Albert, R., 2014. Network modeling of TGF β signaling in hepatocellular carcinoma epithelial-to-mesenchymal transition reveals joint sonic hedgehog and Wnt pathway activation. *Cancer Res.* 74, 5963–5977. <https://doi.org/10.1158/0008-5472.CAN-14-0225>.
- Stricker, J., Cookson, S., Bennett, M.R., Mather, W.H., Tsimring, L.S., Hasty, J., 2008. A fast, robust and tunable synthetic gene oscillator. *Nature* 456, 516–519. <https://doi.org/10.1038/nature07389>.
- Thieffry, D., Thomas, R., 1995. Dynamical behaviour of biological regulatory networks—II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.* 57, 277–297. <https://doi.org/10.1007/BF02460619>.
- Thomas, R., Kaufman, M., 2001. Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos Woodbury N* 11, 180–195. <https://doi.org/10.1063/1.1349893>.
- Tigges, M., Marquez-Lago, T.T., Stelling, J., Fussenegger, M., 2009. A tunable synthetic mammalian oscillator. *Nature* 457, 309–312. <https://doi.org/10.1038/nature07616>.
- Tomita, J., Nakajima, M., Kondo, T., Iwasaki, H., 2005. No transcription-translation feedback in circadian rhythm of KaiC phosphorylation. *Science* 307, 251–254. <https://doi.org/10.1126/science.1102540>.
- Tsai, T.Y.-C., Choi, Y.S., Ma, W., Pomeroy, J.R., Tang, C., Ferrell, J.E., 2008. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* 321, 126–129. <https://doi.org/10.1126/science.1156951>.
- Wang, R.-S., Saadatpour, A., Albert, R., 2012. Boolean modeling in systems biology: an overview of methodology and applications. *Phys. Biol.* 9, 055001. <https://doi.org/10.1088/1478-3975/9/5/055001>.
- Williams, K., Savageau, M.A., Blumenthal, R.M., 2013. A bistable hysteretic switch in an activator-repressor regulated restriction-modification system. *Nucleic Acids Res* 41, 6045–6057. <https://doi.org/10.1093/nar/gkt324>.
- Zelzer, E., Shilo, B.Z., 2000. Cell fate choices in *Drosophila* tracheal morphogenesis. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 22, 219–226. [https://doi.org/10.1002/\(SICI\)1521-1878\(200003\)22:3\(219::AID-BIES3\)3.0.CO;2-A](https://doi.org/10.1002/(SICI)1521-1878(200003)22:3(219::AID-BIES3)3.0.CO;2-A).