

# Phenotypes and tolerances in the design space of biochemical systems

Michael A. Savageau<sup>a,1</sup>, Pedro M. B. M. Coelho<sup>a,b</sup>, Rick A. Fasani<sup>a</sup>, Dean A. Tolla<sup>a</sup>, and Armindo Salvador<sup>b,c</sup>

<sup>a</sup>Department of Biomedical Engineering, University of California, One Shields Avenue, Davis, CA 95616-5294; <sup>b</sup>Biological Chemistry Group, Chemistry Department, University of Coimbra, 3004-535 Coimbra, Portugal; and <sup>c</sup>Center for Neurosciences and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal

Edited by John Ross, Stanford University, Stanford, CA, and approved January 22, 2009 (received for review January 17, 2009)

One of the major unsolved problems of modern biology is deep understanding of the complex relationship between the information encoded in the genome of an organism and the phenotypic properties manifested by that organism. Fundamental advances must be made before we can begin to approach the goal of predicting the phenotypic consequences of a given mutation or an organism's response to a novel environmental challenge. Although this problem is often portrayed as if the task were to find a more or less direct link between genotypic and phenotypic levels, on closer examination the relationship is far more layered and complex. Although there are some intuitive notions of what is meant by phenotype at the level of the organism, it is far from clear what this term means at the biochemical level. We have described design principles that are readily revealed by representation of molecular systems in an appropriate design space. Here, we first describe a generic approach to the construction of such a design space in which qualitatively distinct phenotypes can be identified and counted. Second, we show how the boundaries between these phenotypic regions provide a method of characterizing a system's tolerance to large changes in the values of its parameters. Third, we illustrate the approach for one of the most basic modules of biochemical networks and describe an associated design principle. Finally, we discuss the scaling of this approach to large systems.

biological design principles | piecewise power-law representation | robustness | biochemical systems theory | metabolic network motifs

The molecular revolution that swept through biology in the latter half of the 20th century made it apparent that we would soon be completing the parts catalog for a few of the simplest and best-studied organisms. However, it has become increasingly clear that our knowledge of even the best-studied organisms is still incomplete and fragmentary. We lack the ability to predict the organism's response to a novel mutation in its gene sequence or to a novel (e.g., man-made) compound in its environment. This is the central problem of relating the genotype to the phenotype of the organism. For many, what bridges the enormous divide between genotype and phenotype is gene circuitry, interpreted in the broadest sense to include all of the molecules and interactions that link genes to one another. The function of this circuitry is obvious on one level. The genotype is determined by the information encoded in the DNA sequence, the phenotype by the context-dependent expression of the genome, and the circuitry is there to interpret the context and orchestrate appropriate responses for the organism. This is all true, but it is not very helpful in relating the genotype to the phenotype.

If we look deeper, we find a multilevel hierarchy of systems between the genotype and the phenotype. At each level of the hierarchy, one finds a diversity of designs for achieving what often appears to be the same function. This raises the fundamental question: are these differences in design the result of accident or rule (1)? Are they the result of frozen accidents in the evolutionary process that happen to work well enough to survive natural selection? Alternatively, are they the result of selection to perform subtly different functions, and, if so, can we

discover rules that predict when a given design might evolve to perform a particular function in a specific context? The various designs at a given level can be analyzed and compared within a common framework of representation to reveal functional differences; however, relating designs at different levels requires detailed knowledge of the mapping between the levels.

**Representations and Mappings.** Questions of representation arise at every level, and at many levels the challenges are grand. Clearly, there are many different representations for the same object, and one must decide which is the most appropriate for a given purpose [see [supporting information \(SI\) Dataset S1](#)]. If we are to relate genotype to phenotype, we must consider representations that are appropriate for these two levels and for the various intervening levels.

Relating phenotype and genotype from first principles requires at least two major and qualitatively different mappings. First, the digital representation of the genotype space must be related to the analog representation of the parameter space for the environment and subsystems comprising the global system that is the organism, at present a fundamental unsolved problem. For example, knowing the DNA sequence does not allow one to determine the kinetic parameters of an enzyme. Second, the parameter space must be identified for the specific environment and subsystems, which in many cases remain to be discovered, and then related to the phenotype space, another fundamental unsolved problem. For example, knowing the parameter values does not tell one how many qualitatively distinct phenotypes are in the organism's repertoire or the relative fitness of the phenotypes in different environments.

The parameter space, which can be considered an intermediate in relating different hierarchical levels, can only be defined through the elucidation and parametric characterization of the intervening system. Examples that illustrate this process can be found for many of the hierarchical levels between an organism's genotype and phenotype. These examples also demonstrate the important point that "phenotypes" and "environments" are manifested at all of these levels and not just at the level of the organism.

**Design Space.** Our goal is to construct a "design space" that includes relationships among genotype, phenotype, and environment and that illuminates the function, design, and fitness of the organism. Currently, we must forego the mapping of genotype to system parameters; as noted above, this is a fundamental

Author contributions: M.A.S., P.M.B.M.C., and A.S. designed research; M.A.S., P.M.B.M.C., and D.A.T. performed research; M.A.S., P.M.B.M.C., R.A.F., D.A.T., and A.S. analyzed data; and M.A.S., P.M.B.M.C., R.A.F., D.A.T., and A.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>To whom correspondence should be addressed. E-mail: masavageau@ucdavis.edu.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0809869106/DCSupplemental](http://www.pnas.org/cgi/content/full/0809869106/DCSupplemental).

unsolved problem. Instead, we must use parameter values as a proxy for the genotype.

The representation of such a design space, as distinct from the parameter space, involves a dimensional compression of parameter space by forming dimensionless quantities and grouping parameters and independent variables into aggregate factors. Such a design space is of lower dimension but still includes all of the parameters and independent variables, at least implicitly. This dimensional compression can be seen graphically in the examples of handcrafted design spaces that are described in *Dataset S2*. These design spaces are partitioned into regions by a variety of geometrical relationships representing physical limits, functional constraints, different qualitative dynamics, and other qualitatively distinct aspects of the phenotype. Indeed, one may consider the regions of such a design space as corresponding to different phenotypes.

**Toward a Generic Construction of Design Space.** Such handcrafted examples have motivated us to explore the possibilities of developing a more generic approach to the construction of design spaces based on the power-law formalism (*Dataset S1*). We first describe our general approach in this section and then illustrate it with specific examples, one in the next section and two others in *Dataset S3* and *Dataset S4*. We begin with the most common types of models used to represent biochemical systems.

One of the two most common rate-law models of biochemical systems can be described in terms of mass action equations,

$$\frac{dX_i}{dt} = \sum_k \alpha_{ik} \prod_j X_j^{g_{ijk}} - \sum_k \beta_{ik} \prod_j X_j^{h_{ijk}} \quad i = 1, 2, \dots, n \quad [1]$$

in which  $X_i$  represents a concentration. In traditional chemical kinetics, the sums of products of power laws involve coefficients ( $\alpha_{ik}$  and  $\beta_{ik}$ , rate constants) that are positive real and exponents ( $g_{ijk}$  and  $h_{ijk}$ , kinetics orders) that have small positive integer values: 1, 2, or very rarely 3. In the power-law formalism (2), the exponents are not restricted to positive integer values, but can have real values, thus producing a generalized mass action (GMA) model.

The other most common rate-law model of biochemical systems can be described in terms of rational function equations,

$$\frac{dX_i}{dt} = \frac{\sum_k \alpha_{ik} \prod_j X_j^{g_{ijk}}}{\sum_k \delta_{ik} \prod_j X_j^{p_{ijk}}} - \frac{\sum_k \beta_{ik} \prod_j X_j^{h_{ijk}}}{\sum_k \gamma_{ik} \prod_j X_j^{q_{ijk}}} \quad i = 1, 2, \dots, n \quad [2]$$

again, in which the sums of products of power laws involve positive coefficients and exponents that have small positive integer values. By using a straightforward process called recasting (2), one can transform such equations into an equivalent but larger set of GMA equations.

At steady state, the equivalent set of GMA equations within the power-law formalism is given below. Traditional mass action models are simply special cases of the GMA representation.

$$\sum_k \alpha_{ik} \prod_j X_j^{g_{ijk}} = \sum_k \beta_{ik} \prod_j X_j^{h_{ijk}} \quad i = 1, 2, \dots, n \quad [3]$$

Rational function models can be recast into an equivalent GMA representation in a number of ways. The following is a simple example recasting the steady state of Eq. 2 by the introduction of two new variables for each equation.

$$\sum_k \alpha_{ik} \prod_j X_j^{g_{ijk}} X_{ai}^{-1} = \sum_k \beta_{ik} \prod_j X_j^{h_{ijk}} X_{bi}^{-1}$$

$$X_{ai} = \sum_k \delta_{ik} \prod_j X_j^{p_{ijk}} \quad i = 1, 2, \dots, n \quad [4]$$

$$X_{bi} = \sum_k \gamma_{ik} \prod_j X_j^{q_{ijk}}$$

Thus, without loss of generality, we can focus on models in the GMA representation within the power-law formalism.

Each side of the steady-state equations is a sum of several terms. When one term on each side is dominant (i.e., is the largest term on its side), the system of equations can be approximated locally by an S-system (*Dataset S1*) in steady state.

$$\alpha_i \prod_j X_j^{g_{ij}} = \beta_i \prod_j X_j^{h_{ij}} \quad i = 1, 2, \dots, n \quad [5]$$

These equations have a single analytical solution that is linear in the logarithms of the concentration variables and rate constants (2, 3). The conditions for any given term to be dominant are provided by a set of linear inequalities in log space.

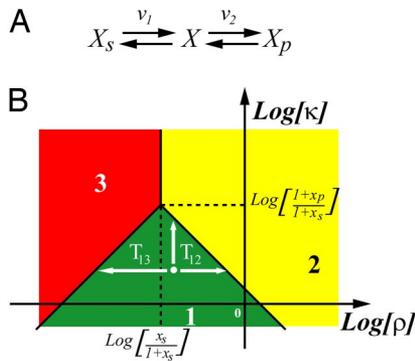
Because each term on each side is potentially dominant, there are as many potential solutions as there are combinations of terms in the GMA system; hence a bound on the number of steady-state solutions is provided. However, not all potential solutions are necessarily valid. A test of each potential solution against the inequalities necessary for its validity will determine whether or not a potential solution is in fact a valid solution. The pathway example discussed in the next section will make these ideas more concrete.

The inequalities and the corresponding solution define the boundaries for a region in design space. Within each region, there is a qualitatively distinct solution that can be characterized (2) with respect to signal amplification (logarithmic gains), robustness (parameter insensitivity), and stability (eigenvalues) involving local (small) changes in variables and parameters. These characteristics can be compared against a set of quantitative performance criteria to determine the relative fitness of each region. These qualitatively distinct solutions can be defined as molecular phenotypes.

In addition to characterizing local (small) changes in performance, this approach of partitioning design space provides a natural means, based on the boundaries, for calculating tolerances to global (large) changes in parameter values. These tolerances are defined for each parameter as the ratio of its value at the nominal steady state (normal operating point for the system) to its value on the boundary to an adjacent phenotypic region (or the inverse, depending on which value is larger). Although comparison of the actual and the piecewise solutions reveals discrepancies, which are greatest near the boundaries, accuracy throughout the design space is not the primary concern here. Rather, it is the ease with which boundaries can be analytically defined based on the underlying equations and can be used to quantify global tolerances.

The strategy outlined above will be illustrated by constructing design spaces for the most widespread elements of biochemical networks: pathways (in the next section), cycles (*Dataset S3*), and branches (*Dataset S4*).

**Generic Construction of the Design Space for Pathways.** Although there are a few important cases in which a single biochemical pathway (or portion of a pathway) undergoes a reversal of net flux, most pathways in cells operate in a unidirectional fashion. If a metabolite is synthesized under one set of conditions and catabolized under another set of conditions, these processes are seldom carried out by reversal of the same pathway. For example,



**Fig. 1.** Metabolic pathway. (A) Reversible pathway with substrate  $X_S$ , intermediate  $X$ , and product  $X_P$ . (B) Design space in which the nominal operating point (white dot) is depicted in phenotypic region 1, and the tolerances to phenotypic regions 2 ( $T_{12}$ ) and 3 ( $T_{13}$ ) are indicated by arrows. See *Results* for discussion.

in *Salmonella typhimurium* there is a histidine biosynthetic pathway (4) and a separate histidine utilization pathway (5).

**Kinetic model.** Consider the simplest example of such a pathway that can be used to illustrate the construction of the design space (Fig. 1A). The two reactions follow reversible Michaelis–Menten mechanisms (6). The substrate and product concentrations are fixed such that the flux through the pathway is driven to the right, far from equilibrium, and the reverse reactions can be neglected. The resulting equation for the intermediate  $X$  is

$$\frac{dX}{dt} = \frac{V_{\text{Max},1} \frac{X_S}{K_S}}{1 + \frac{X_S}{K_S} + \frac{X}{K_I}} - \frac{V_{\text{Max},2} \frac{X}{K}}{1 + \frac{X}{K} + \frac{X_P}{K_P}} \quad [6]$$

or, in dimensionless form,

$$\frac{dx}{d\tau} = \frac{x_s}{1 + x_s + x/\kappa} - \frac{\rho x}{1 + x + x_p} \quad [7]$$

where the dimensionless variables and parameters are given by  $x = X/K$ ,  $x_s = X_S/K_S$ ,  $x_p = X_P/K_P$ ,  $\rho = V_{\text{Max},2}/V_{\text{Max},1}$ ,  $\kappa = K/K_I$ ,  $\tau = (V_{\text{Max},1}/K)t$ . This effectively reduces the number of parameters and independent variables from 8 to 4.

**Recasting.** The recast GMA version of Eq. 7, in steady state (and letting  $x_1 = x$ ), is obtained by the procedure described in the previous section in which two new variables [ $x_2 = (1 + x_s) + x_1/\kappa$  and  $x_3 = (1 + x_p) + x_1$ ] are introduced.

$$\begin{aligned} 0 &= x_s x_2^{-1} - \rho x_1 x_3^{-1} \\ 0 &= (1 + x_s) + x_1/\kappa - x_2 \\ 0 &= (1 + x_p) + x_1 - x_3 \end{aligned} \quad [8]$$

In addition to the normalization described above, we have formed aggregate parameters  $(1 + x_s)$  and  $(1 + x_p)$  to promote dimensional compression of the design space that is eventually generated.

The number of combinations of terms in Eq. 8 gives a bound on the total number of potential phenotypic regions ( $T$ )

$$T = \prod_{i=1}^3 P_i * N_i = (1*1)*(2*1)*(2*1) = 4 \quad [9]$$

where  $P_i$  and  $N_i$  are the number of positive and negative terms in the  $i$ th equation.

**Breakpoint conditions.** The breakpoints between the two positive terms in the second and third of Eq. 8 are obtained by equating the two terms and taking logarithms to generate linear expressions.

$$\begin{aligned} \text{Log } x_1 &= \text{Log}[\kappa(1 + x_s)] \\ \text{Log } x_1 &= \text{Log}(1 + x_p) \end{aligned} \quad [10]$$

The need for logarithms may not be obvious in this simple example because each term only involves a single dependent variable, but in more complex examples, taking logarithms generates linear equations that greatly simplify the construction of design space.

These equations give rise to four linear inequality conditions that are based on the dominant terms in Eq. 8 and involve the aggregate parameters  $\kappa$ ,  $(1 + x_s)$ , and  $(1 + x_p)$ . Consider in detail one of the four potential phenotypic regions, call it phenotypic region 1, for which the breakpoint conditions are given by

$$\begin{aligned} \text{Log } x_1 &> \text{Log}[\kappa(1 + x_s)] \\ \text{Log } x_1 &< \text{Log}(1 + x_p) \end{aligned} \quad [11]$$

**Local solution.** In this region, the dominant terms from Eq. 8 comprise an S-system that has the following linear form in logarithmic space.

$$\begin{aligned} \text{Log } x_1 + \text{Log } x_2 - \text{Log } x_3 &= \text{Log}(x_s/\rho) \\ \text{Log } x_1 - \text{Log } x_2 &= \text{Log } \kappa \\ \text{Log } x_3 &= \text{Log}(1 + x_p) \end{aligned} \quad [12]$$

The unique solution for the intermediate is given by

$$\text{Log } x_1 = 0.5 \text{Log}[\kappa x_s(1 + x_p)/\rho] \quad [13]$$

**Phenotypic boundaries.** The boundaries of the phenotypic region in which this solution is valid are obtained by inserting the linear solution (Eq. 13) into the corresponding linear breakpoint conditions (Eq. 11). The result is the following set of boundaries that are linear in log space.

$$\begin{aligned} \text{Log } \kappa &< \text{Log}[x_s(1 + x_p)/(1 + x_s)^2] - \text{Log } \rho \\ \text{Log } \kappa &< \text{Log}[(1 + x_p)/x_s] + \text{Log } \rho \end{aligned} \quad [14]$$

The same procedure can be applied to each of the other three combinations of breakpoint inequalities. The results for the three meaningful steady states are summarized graphically in the design space of Fig. 1B.

Note that in this case there is a compression of the 8-dimensional space of the original parameters and independent variables to a 2-dimensional design space. As a natural consequence of constructing the design space, the original 8 parameters and independent variables give rise to 4 aggregate factors [ $\rho$ ,  $\kappa$ ,  $x_s/(1 + x_s)$ , and  $(1 + x_p)/(1 + x_s)$ ] (Fig. 1B).

**Local performance.** The local performance in each of the phenotypic regions can be readily compared on the basis of relevant quantitative criteria because the system representation within each phenotypic region is a simple but nonlinear S-system for which determination of local behavior reduces to conventional linear analysis (2, 3). Thus, the behavior involving local (small) variations is completely determined, and there are criteria that can be defined and evaluated to characterize the performance of the system. These criteria are quantified by using logarithmic gains, parameter sensitivities, and response times.

Logarithmic gains in concentrations and fluxes in response to changes in value for an independent variable are defined by the relative derivative of the explicit steady-state solution. For

**Table 1. Steady-state solutions and boundaries for the four phenotypic regions in the design space of the pathway in Fig. 1A**

Phenotypic region	Steady-state solutions		Boundaries
	$x = X/K$	$v = V_i/V_{max,i}$	
1	$\left[\frac{\kappa(1+x_p)x_s}{\rho}\right]^{1/2}$	$\left[\frac{\kappa\rho x_s}{(1+x_p)}\right]^{1/2}$	$\text{Log } \kappa < \text{Log} [x_s(1+x_p)/(1+x_s)^2] - \text{Log } \rho$ $\text{Log } \kappa < \text{Log} [(1+x_p)/x_s] + \text{Log } \rho$
2	$\frac{x_s(1+x_p)}{\rho(1+x_s)}$	$\frac{x_s}{(1+x_s)}$	$\text{Log } \kappa > \text{Log} [x_s(1+x_p)/(1+x_s)^2] - \text{Log } \rho$ $\text{Log } \rho > \text{Log} [x_s/(1+x_s)]$
3	$\frac{\kappa x_s}{\rho}$	$\rho$	$\text{Log } \kappa > \text{Log} [(1+x_p)/x_s] + \text{Log } \rho$ $\text{Log } \rho < \text{Log} [x_s/(1+x_s)]$
4	Undefined $(1+x_p) < x < \kappa(1+x_s)$	$\rho$	$\text{Log } \kappa > \text{Log} [(1+x_p)/x_s] + \text{Log } \rho$ $\text{Log } \rho = \text{Log} [x_s/(1+x_s)]$

example, using the intermediate concentration  $x$ , pathway flux  $v$ , and independent substrate concentration  $x_s$ , representative log-arithmetic gains for the pathway in Fig. 1A are

$$L(x, x_s) = \frac{\partial \text{Log } x}{\partial \text{Log } x_s} = \frac{\partial x}{\partial x_s} \frac{x_s}{x} \quad L(v, x_s) = \frac{\partial \text{Log } v}{\partial \text{Log } x_s} = \frac{\partial v}{\partial x_s} \frac{x_s}{v} \quad [15]$$

Parameter sensitivities of such state variables in response to changes in value for the parameters that define the structure of the system (e.g., Michaelis constants and maximal velocities) are defined by the relative derivative of the explicit steady-state solution. For example,

$$S(x, \kappa) = \frac{\partial \text{Log } x}{\partial \text{Log } \kappa} = \frac{\partial x}{\partial \kappa} \frac{\kappa}{x} \quad S(v, \rho) = \frac{\partial \text{Log } v}{\partial \text{Log } \rho} = \frac{\partial v}{\partial \rho} \frac{\rho}{v} \quad [16]$$

Response times are determined by the inverse of the real part of the dominant eigenvalue.

### Results

The analytical expressions for the boundaries and steady states for each phenotypic region (see Fig. 1B) are summarized in Table 1. Note that in addition to the three meaningful steady states there is one unrealistic case. The local performance of designs in the three phenotypic regions that are meaningful can be readily compared on the basis of relevant quantitative criteria (Table 2). The results are summarized in Table 3.

The two most important criteria are the ability of a change in substrate (Criterion 1) or product (Criterion 2) to increase pathway flux with a minimal increase in intermediate. The designs in phenotypic region 3 are unable to do either; those in regions 1 and 2 respond equally well to changes in substrate, but those in region 2 are unable to respond to changes in product.

Increases in the concentration of the intermediate should be minimized in response to an increase in either substrate (Criterion 3) or product (Criterion 4). The designs in phenotypic region 1 are best in responding to substrate if  $x_s < 1$ . They also are better than those in phenotypic region 2 in responding to product if  $x_p < 2$ . This means that the best designs in phenotypic region 1 are located down and to the right, near the boundary with region 2. (Although designs in phenotypic region 3 are best in responding to product, this is irrelevant because pathway flux is unresponsive.)

The robustness of pathway flux (Criterion 5) is best for designs in phenotypic region 2, whereas the robustness of the intermediate concentration (Criterion 6) is equally good for designs in phenotypic regions 1 and 2. Finally, the response time for an increase in substrate (Criterion 7), which in these special cases

**Table 2. Criteria for the local performance of the pathway in Fig. 1A**

Criterion number	Definition	Calculation
1	Maximum ratio for gain of pathway flux $v$ relative to gain of intermediate $X$ in response to changes in substrate $X_s$	$\text{Max} \frac{L(v, X_s)}{L(X, X_s)}$
2	Maximum ratio for gain of pathway flux $v$ relative to gain of intermediate $X$ in response to changes in product $X_p$	$\text{Max} \left[ \frac{L(v, X_p)}{L(X, X_p)} \right]$
3	Minimum gain of intermediate $X$ in response to changes in substrate $X_s$	$\text{Min } L(X, X_s)$
4	Minimum gain of intermediate $X$ in response to changes in product $X_p$	$\text{Min } L(X, X_p)$
5	Maximum robustness of flux $v$ to variations in parameters $p_i$	$\text{Min} \sum_i  S(v, p_i) $
6	Maximum robustness of intermediate $X$ to variations in parameters $p_i$	$\text{Min} \sum_i  S(X, p_i) $
7	Minimum response time	$\text{Min } \tau_{1/2} \quad (\text{Min } \lambda)$

can be obtained analytically, is best for designs in phenotypic region 1. In summary, the designs in region 3 clearly exhibit the worst local performance. Thus, designs in regions 1 and 2 are the only ones that merit further comparison.

There are two classes of intermediates that need to be considered. When the intermediate only functions as a route to producing product, we shall call it a “monofunctional intermediate.” When the intermediate has that function and serves as an important signaling molecule that influences other processes, we shall call it a “bifunctional intermediate.” The accumulation of a monofunctional intermediate compromises the cell’s limited solvent capacity, slows the temporal response, and, for the many instances in which the intermediate is toxic, is damaging to the cell. Thus, it is imperative to minimize accumulation. Under

**Table 3. Comparison of local performance in the three phenotypic regions of Fig. 1B**

Criterion	Preference †	Phenotypic Region		
		1	2	3
1	↑	1	1	0
2	↑	1	0	0
3	↓	$\frac{1}{2}$	$\frac{1}{1+x_s}$	1
4	↓	$\frac{1-x_p}{2(1+x_p)}$	$\frac{1}{1+x_p}$	0
5	↓	2	0	2
6	↓	2	2	4
7	↓	$\text{Log} \left[ \frac{4 \left( \frac{x'_s}{x_s} - 1 \right)}{3 \left( \frac{x'_s}{x_s} \right) - 2 \sqrt{\frac{x'_s}{x_s} - 1}} \right]$	$\text{Log} 2$	$\left( \frac{1}{2} \right) \left[ \left( \frac{x'_s}{x_s} \right) (2 \text{Log} 2 - 1) + 1 \right]$

\*  $x_s^0$  and  $x'_s$  are the fixed (initial) pre- and (final) post-disturbance values of the substrate  $x_s$ .

† To improve performance, one must have either a high (↑) or a low (↓) value for the associated criterion.

**Table 4. Global tolerances to change in parameters and independent variables for a system operating in phenotypic region 1 (see Fig. 1B)**

Parameter	Tolerance	
	Region 2	Region 3
$V_1$	[2.5, $\infty$ ]	[ $\infty$ , 10]
$V_2$	[ $\infty$ , 2.5]	[10, $\infty$ ]
$K_I$	[ $\infty$ , 2.5]	[ $\infty$ , $\infty$ ]
$K$	[2.5, $\infty$ ]	[ $\infty$ , $\infty$ ]
$X_S$	[2.4, $\infty$ ]	[ $\infty$ , 959.6]
$K_S$	[ $\infty$ , 2.4]	[959.6, $\infty$ ]
$X_P$	[ $\infty$ , $\infty$ ]	[ $\infty$ , $\infty$ ]
$K_P$	[ $\infty$ , $\infty$ ]	[ $\infty$ , $\infty$ ]

\*  $x_s^0 = 1/20$ ,  $x_p^0 \approx 0$ ,  $\rho^0 = 1/10$ , and  $\kappa^0 = 1/5$

these circumstances, the local performance of designs in phenotypic region 1 is better than or equal to that of designs in phenotypic region 2 (based on 6 of the 7 criteria). However, the concentration of a bifunctional intermediate must be fairly responsive to changes to promote its signaling function. Under these circumstances, the local performance of designs in phenotypic region 2 is better than or equal to that of designs in phenotypic region 1 (based on 5 of the 7 criteria).

“Global tolerance” to large changes in parameters and independent variables is defined as the value at the boundary between adjacent phenotypic regions relative to the normal operating value (or the inverse if the normal value is greater than the value at the boundary). We will use the expression  $[T_D, T_I]$  to describe the global tolerances, where  $T_D$  = tolerance to a fold decrease and  $T_I$  = tolerance to a fold increase (because boundaries can be crossed either by decreasing or increasing a parameter or independent variable).

For a design with a monofunctional intermediate, the best local performance corresponds to a set of nominal values for the parameters and independent variables that locate the operating point down and to the right in the best phenotypic region (region 1). As a consequence, the global tolerances to changes that would move the design into the poorer phenotypic region (region 2) are smaller than those that would move it into the worst phenotypic region (region 3). Given typical values for such a well-defined system, one can readily determine the global tolerances, which are summarized in Table 4. Thus, a well-designed system has, in addition to good local performance, large global tolerances to changes that degrade performance. This has also been observed for other systems (see *Dataset S3*) but has not been established in general. We postulate that this will be true of well-adapted systems in nature.

As mentioned before, pathways are one of the three most widespread elements in metabolic networks. The second is the moiety-transfer cycle. This form of coupling between reactions is prevalent in metabolism. For example, of all of the enzyme-catalyzed bisubstrate reactions in the reconstructed metabolic networks of *Escherichia coli* (7) and *Saccharomyces cerevisiae* (8), 836 (75%) and 561 (67%), respectively, participate in moiety-transfer cycles. These calculations exclude cycles involving the ubiquitous metabolites  $H_2O$  and  $H^+$ , and pairs of forward-

reverse reactions. Redundant reactions catalyzed by distinct (iso)enzymes were counted as a single reaction. The construction and analysis of the design space for moiety-transfer cycles are given elsewhere and summarized in *Dataset S3*.

The third widespread element in metabolic networks is the branch point, a metabolite on which two or more pathways converge or from which two or more pathways diverge. Important decisions are made at these branch points, such as how much each converging pathway will contribute to the overall synthesis or how much each diverging pathway will draw from a common flux. Even higher-level decisions, such as cell fate differentiation, ultimately involve one or more branch point decisions. Given the importance of branch point decisions, it is not surprising that many regulatory mechanisms have been found to modulate them. An abstracted version of several amino acid biosynthetic pathways (9, 10) includes diverging and converging pathways regulated by a well-established pattern of nested control. The construction and analysis of the design space for this branched model are summarized in *Dataset S4*.

## Discussion

The methodology based on the power-law formalism that we have presented has its foundation in algebraic geometry (11). Thus, the boundaries in design space are straight lines in the log space of independent variables and rate constants. The slopes and intercepts of these lines are rational functions of the kinetic orders. As we have seen, these methods provide well-defined bounds on the number of phenotypic regions in design space. In all of the examples given (pathways, cycles, and branch points), there is a single steady state in each phenotypic region, and all of these are locally stable. The same approach applies to other systems that have unstable or multiple steady states. Because construction of the design space involves only linear algebra, the method will scale numerically to larger problems.

One of the two main motivations for developing the design space concept is to facilitate the identification of design principles. The second is to provide a quantitative measure of global tolerance to (large) parameter variation based on well-defined boundaries between qualitatively distinct phenotypic regions in design space.

The construction of a design space for the pathway in Fig. 1A, and the results from the corresponding analysis, suggest the following design principle. First, the worst phenotype (Fig. 1B, region 3) will be avoided if  $\rho > x_S/(1 + x_S)$ . In biological terms, the maximal velocity for the second enzyme should be larger than that of the first (large  $\rho$ ). The condition  $\kappa < [x_S(1 + x_P)/(1 + x_S)^2]\rho$  will produce the optimal monofunctional phenotype (Fig. 1B, region 1), and the reverse will produce the optimal bifunctional phenotype (Fig. 1B, region 2). The results for moiety-transfer cycles (*Dataset S3*) and for branched systems (*Dataset S4*) also suggest design principles that are supported by experimental evidence.

The second of the two main motivations for developing the design space concept is to provide a quantitative measure of global tolerance to (large) parameter variation based on well-defined boundaries between qualitatively distinct phenotypic regions in design space. The results in the previous section and in *Dataset S3* and *Dataset S4* show that these tolerances can be readily calculated and compared. In each case, the results suggest that systems tend to operate in phenotypic regions with good local performance, and that the location of their nominal operating point within these phenotypic regions confers large tolerances to parameter variation, thereby achieving both robust local behavior and large global tolerance.

**ACKNOWLEDGMENTS.** We thank an anonymous reviewer for insightful comments and suggestions that greatly improved the revised manuscript. This work was supported in part by U.S. Public Health Service Grant

R01-GM30054 (to M.A.S.), by Grant PTDC/QUI/70523/2006 and Fellowship SFRH/BPD/945002 from the Portuguese Fundação para a Ciência e Tecnologia (to A.S.), by U.S. Public Health Service Training Grant T32-EB003827

(to D.A.T.), by an Earl C. Anthony fellowship (to R.A.F.), and by Fellowship SFRH/BD/8304/2002 from the Portuguese Fundação para a Ciência e Tecnologia (to P.M.B.M.C.).

1. Savageau MA (1989) Are there rules governing patterns of gene regulation? *Theoretical Biology: Epigenetic and Evolutionary Order*, eds Goodwin BC, Saunders PT (Edinburgh Univ Press, Edinburgh), pp 42–66.
2. Savageau MA (2001) Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* 11:142–159.
3. Savageau MA (1976) *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Addison-Wesley, Reading, MA).
4. Brenner M, Ames BN (1971) The histidine operon and its regulation. *Metabolic Pathways*, 3rd Ed, Vol V, *Metabolic Regulation*, ed Vogel HJ (Academic, New York), pp 349–387.
5. Smith GR, Magasanik B (1971) Nature and self-regulated synthesis of the repressor of the *hut* operons in *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 68:1493–1497.
6. Fromm HJ (1975) *Initial Rate Enzyme Kinetics* (Springer, New York).
7. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA* 97:5528–5533.
8. Forster J, Famili I, Fu P, Palsson BO, Nielsen J (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res* 13:244–253.
9. Stadtman ER (1963) Symposium on multiple forms of enzymes and control mechanisms. II. Enzyme multiplicity and function in the regulation of divergent metabolic pathways. *Bacteriol Rev* 27:170–181.
10. Truffa-Bachi P, Cohen GN (1968) Some aspects of amino acid biosynthesis in microorganisms. *Annu Rev Biochem* 37:79–108.
11. Cox DA, Little JB, O'Shea D (2005) *Using Algebraic Geometry* (Springer, New York), 2nd Ed.