Research

Improved methods for the mathematically controlled comparison of biochemical systems

John H Schwacke and Eberhard O Voit*

Address: Department of Biometry, Bioinformatics, and Epidemiology Medical University of South Carolina 135 Cannon Street, Suite 303 Charleston, SC 29425, U.S.A

Email: John H Schwacke - schwacke@musc.edu; Eberhard O Voit* - voiteo@musc.edu

* Corresponding author

Abstract

The method of mathematically controlled comparison provides a structured approach for the comparison of alternative biochemical pathways with respect to selected functional effectiveness measures. Under this approach, alternative implementations of a biochemical pathway are modeled mathematically, forced to be equivalent through the application of selected constraints, and compared with respect to selected functional effectiveness measures. While the method has been applied successfully in a variety of studies, we offer recommendations for improvements to the method that (1) relax requirements for definition of constraints sufficient to remove all degrees of freedom in forming the equivalent alternative, (2) facilitate generalization of the results thus avoiding the need to condition those findings on the selected constraints, and (3) provide additional insights into the effect of selected constraints on the functional effectiveness measures. We present improvements to the method and related statistical models, apply the method to a previously conducted comparison of network regulation in the immune system, and compare our results to those previously reported.

Background

Metabolic and signal transduction pathways in biological systems are typically complex networks that necessitate the application of mathematical modeling and computer simulation in efforts to understand their behavior. Mathematical models, developed through these efforts, have value both as tools for predicting system behavior and as descriptions of the system that facilitate the study of the embodied design principles [1,2]. A design principle, as defined by Savageau, is a rule that characterizes a feature of a class of systems and thus facilitates understanding the entire class. As these rules are identified and characterized a catalog of patterns will be developed for use in the identification of additional instances of these patterns within biological systems [3]. To gain a greater understanding of

the benefits of one design over another and to understand the selection criteria driving an evolutionary design choice we need methods by which objective comparisons of alternative designs can be performed.

To perform these comparisons we first require a mathematical framework with which we describe the designs of interest and compare those designs with respect to functional effectiveness measures. The framework chosen here is based on the form of canonical nonlinear modeling referred to as synergistic or S-systems. S-systems, developed as part of Biochemical Systems Theory (BST), are systems of nonlinear ordinary differential equations with a well-defined structure [4-6]. The time rate of change of each quantity in the system is described by a differential

equation of the form given in Equation 1 where $\dot{X}_i$ indicates the first derivative of quantity X_i with respect to time and V_i^+ and V_i^- are positive-valued functions representing the influx and efflux respectively.

$$\dot{X}_i = V_i^+ - V_i^-$$
$$= \alpha_i \prod_{j=1}^{n+m} X_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{i,j}} \text{ for } i \in 1...n \qquad (1)$$

These quantities may represent, for example, substrate, enzyme, metabolite, cofactor, or mRNA concentrations and are referred to generically as pools. The system consists of n equations of this form, one for each of the n dependent variables in the system. The remaining m variables, $X_{n+1} \ldots X_{n+m}$, represent independent quantities. The right-hand side of each equation consists of two terms, one describing the influx or production of the pool of interest (V_i^+) and one describing its degradation or efflux (V_i^-). Both terms are in power-law form. Any other pool (independent or dependent) in the system that influences production or degradation appears as a factor in the appropriate power-law term of the effected pool's differential equation. The exponential coefficient of the factor, referred to as its kinetic order, determines the direction and degree to which the change is influenced. Positive kinetic orders indicate that the influence increases or activates the flux and negative kinetic orders indicate that the influence decreases or inhibits the flux. Kinetic orders associated with the influx term are typically given the label $g_{i,j}$ where the indices i and j denote the influence of variable X_j on the influx to X_i. The label $h_{i,j}$ is typically given to kinetic orders associated with the efflux term. The multiplicative factors α_i and β_i are positive quantities referred to as rate constants. They scale the influx or efflux rate and thus control the time scale of the reaction. The validity of this power-law representation has been analyzed extensively and demonstrated in a variety of biological system modeling applications [7-9].
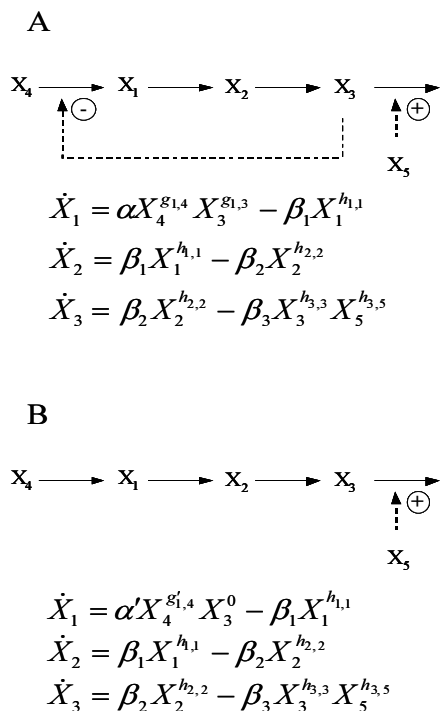
The S-system representation offers two key advantages in the performance of controlled comparisons. First, S-systems have a form that allows for the algebraic determination of the system's steady state by solution of a system of linear equations under logarithmic transformation of the variables (see Appendix). From this steady-state solution, it is possible to determine the local stability of the steady state, the sensitivity of the steady state with respect to parameter changes, and the sensitivity of the steady state with respect to variation in the independent variables. The S-system representation is also advantageous in that it provides a direct mapping from the regulatory structure of
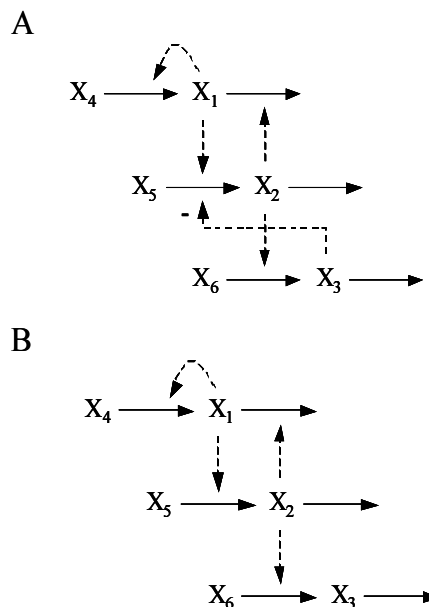
Figure 1
Reference and alternative systems. Biochemical maps for the reference system (with suppression) and the alternative (without suppression) are given in A and B respectively. Adapted from Irvine and Savageau [21].

the system under study to the parameters of the system. If, for example, the influx to a variable of interest, X_i, is regulated by some other variable X_j then the parameter $g_{i,j}$ will be non-zero. If the regulation inhibits the influx, the parameter takes on negative values and if the regulation activates the influx, the parameter takes on positive values. This property of S-systems is particularly useful when performing a controlled comparison of two structures that differ in their regulatory interactions. The alternative structure, without a particular regulatory interaction, can be determined from the reference by forcing the value of the appropriate kinetic order to 0 (Figure 2).

S-systems provide a convenient method for the characterization of systemic performance local to the steady state. System gains, parameter sensitivities, and the margin of local stability are easily determined and often form the basis of functional performance measures used in controlled comparisons. Logarithmic gains represent the change

A

$$\dot{X}_1 = \alpha X_4^{g_{1,4}} X_3^{g_{1,3}} - \beta_1 X_1^{h_{1,1}}$$

$$\dot{X}_2 = \beta_1 X_1^{h_{1,1}} - \beta_2 X_2^{h_{2,2}}$$

$$\dot{X}_3 = \beta_2 X_2^{h_{2,2}} - \beta_3 X_3^{h_{3,3}} X_5^{h_{3,5}}$$

B

$$\dot{X}_1 = \alpha' X_4^{g'_{1,4}} X_3^{0} - \beta_1 X_1^{h_{1,1}}$$

$$\dot{X}_2 = \beta_1 X_1^{h_{1,1}} - \beta_2 X_2^{h_{2,2}}$$

$$\dot{X}_3 = \beta_2 X_2^{h_{2,2}} - \beta_3 X_3^{h_{3,3}} X_5^{h_{3,5}}$$

Figure 2
Example mapping: pathways to S-systems. The S-system framework provides for a straightforward mapping of biochemical pathway maps into systems of equations. The pathway and equations for cases A and B differ only in the feedback inhibition of the first step in the process. This inhibition is represented by a single parameter, $g_{1,3}$.

in the log value of the steady state of a dependent variable or flux as a result of a change in the log value of an independent variable (see Appendix). A log gain of $L_{i,j} = L(X_i, X_j)$ can be interpreted as an indication that a 1% change in independent variable j will result in an approximate $L_{i,j}$% change in the steady-state value of dependent variable i. Logarithmic gains provide a measure of the effect or "gain" of an independent variable on the steady state of the system. A related measure, referred to as system sensitivity, measures the robustness or the degree to which changes in the system parameters (kinetic orders and rate constants) affect the steady state of the system (see Appendix). A sensitivity of $S = S(X_k, g_{i,j})$ indicates that a 1% change in parameter $g_{i,j}$ will result in an approximate S% change in the steady-state value of dependent variable X_k.

The method of mathematically controlled comparison provides a structured approach for the comparison of design alternatives under controlled conditions much like a controlled laboratory experiment [10]. The approach, as currently applied, is implemented in the following steps. (1) Mathematical models for the reference design and one or more alternatives are developed using the S-system modeling framework described above. The alternatives are allowed to differ from the reference at only a single process that becomes the focus of the analysis. (2) The alternative design is forced to be *internally equivalent* to the reference by constraining the parameters of the alternative to be equal to those of the reference for processes other than the process of interest. (3) Using the mathematical framework, selected systemic properties or functions of those properties are identified and used to form constraints which fix the, as yet, unconstrained parameters in the alternative design. Typically, steady-state values and selected logarithmic gains are forced to be equal in the reference and alternative. Parameters for the process of interest in the alternative are then determined as a function of the parameters in the reference so as to satisfy these constraints. The application of these constraints forces the reference and alternative to be *externally equivalent* with respect to the selected properties. The term "external equivalence" refers to the fact that the alternative and reference are equivalent to an external observer with respect to the constrained systemic properties. Constraints are imposed until all of the free parameters in the alternative are determined. (4) Finally, measures of functional effectiveness relevant to the biological context of these designs are determined and used to compare the reference and its internally and externally equivalent alternative through algebraic methods.

In many cases the comparison of these functional effectiveness measures cannot be determined independent of the parameter values. To improve the applicability of the method in these cases, Alves and Savageau extended the method of controlled comparisons through the incorporation of statistical techniques [11,12]. Under this extension, parameter values are sampled from distributions representing prior knowledge about the likely ranges for those parameters. An instance of the reference design is constructed from the sampled parameters and an instance of the alternative is then constructed from the reference by applying the constraint relationships. Functional effectiveness measures are then computed for the each sampled reference and its equivalent alternative ($M_{R,i}$ and $M_{A,i}$). The ratio of the performance measure of the reference relative to that of the alternative is computed for all of the samples and plotted as $M_{R,i}/M_{A,i}$ versus a property P of the reference design. A moving median plot is then prepared by plotting the median of $M_{R,i}/M_{A,i}$ versus the median of P in a sliding window to reveal both the

median of the relative measure and its variation across the range of P. If M is defined such that smaller values indicate greater functional effectiveness, ratios of $M_{R,i}/M_{A,i} < 1$ indicate that the reference is preferred to the alternative according to the given measure. Examination of the density of ratios and moving median plots allows determination of preference for the reference over the alternative (or visa versa) and how that preference varies with the selected property. These extensions have been applied to the analysis of preferences for irreversible steps in biosynthetic pathways [13] and to the comparison of regulator gene expression in a repressible genetic circuit [14].

Rationale for Improvements
While the Method of Mathematically Controlled Comparisons has been successfully applied in many cases [13-20], we offer for consideration enhancements to the method that extend the application of sampling and statistical comparison given by Alves and Savageau [11,12]. These enhancements are offered primarily to (1) allow for the incremental incorporation of constraints in the model, (2) provide evidence for the generalization of comparisons, and (3) provide additional insight into the effects of the selected constraints on our interpretation of the results. The enhancements also address two concerns with the method as presently applied. First, the current approach requires that we identify a number of constraints sufficient to numerically fix all free parameters. An objective of our approach is to relax this requirement for cases where the identification of a sufficient number of constraints is not practical or not desired. Second, the enhanced approach incorporates a step that excludes the use of unrealistic alternatives resulting from the application of constraints.

The existing method currently requires the identification of enough constraints to remove all degrees of freedom associated with parameters of the alternative model not fixed by internal equivalence. The construction of an alternative pair for a given reference in a controlled comparison is similar to the process of matching in an epidemiological study in that both attempt to prevent confounding by restricting comparisons to pairs that have been matched on the confounding variable. The key difference is that in an epidemiological study cases and controls or treatment groups are drawn from the sample population and then matched whereas in a controlled comparison the reference is drawn and the alternative is constructed from the reference to enforce the match. In both cases we become unable to make statements with regard to differences in the systemic properties (confounding variable) that we have matched on. Since both the reference and alternative system were matched at a constraint of our choosing the observation that the matched property or any function of the matched prop-
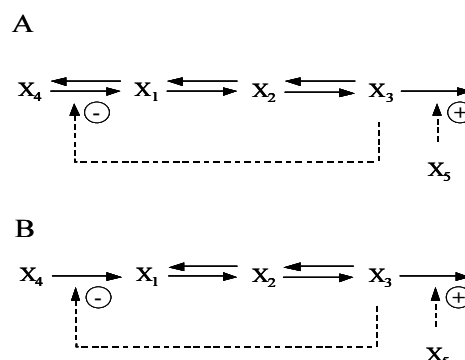
Figure 3
Biosynthetic pathway alternatives. Biosynthetic pathways similar to that illustrated were compared using the method of mathematically controlled comparison by Alves and Savageau [13]. These biosynthetic pathways differ only in the reversibility of the first step.

erty is equal in both systems adds no information to the comparison. Unlike the epidemiological study, a controlled comparison requires us to identify constraints sufficient to eliminate all of the free parameters in the alternative. If we cannot identify a sufficient number of constraints with meaningful interpretations, we may be forced to select constraints for mathematical convenience. Since our observations are conditioned on the constraints imposed in the analysis, the choice of mathematically convenient constraints may lead to complications in interpreting the results.

The application of constraints in forming instances of the alternative design has the potential of producing systems that are unreasonable with respect to their parameter values and thus alternative systems constructed through the application of these constraints must be evaluated for reasonableness. Clearly, these parameter values are related to the kinetic parameters of the underlying biological process and thus are expected to fall within ranges representative of the physical limits of the modeled process. In some cases, the application of constraints can yield alternatives with parameter values far from those expected in a realizable system. Unlike the epidemiological study, the alternative is constructed so as to satisfy the given constraints without concern for the reasonableness of the alternative. Under these conditions, we might mistakenly compare a reference that matches our prior belief about realistic parameter ranges to an unrealistic alternative. In

the existing approach, there is no explicit evaluation of the likelihood or reasonableness of an alternative formed from a given reference. Constraints on resulting kinetic orders have been imposed in some previous applications of controlled comparisons [14,17] but the step has not been applied in methods using statistical extensions. Consider, for example, the analysis of irreversible step positions in unbranched biosynthetic pathways presented in [13]. The structure of the reference and alternative are illustrated in Figure 3. As part of the numerical comparisons, parameter values for kinetic orders and rate constants were drawn from uniform and log-uniform distributions respectively. Kinetic orders were drawn from Unif(0,5) for positive or Unif(-5,0) for negative kinetic orders and log (base 10) rate constants were drawn from Unif(-5,5). Constraint relationships were applied and reference models with irreversible steps at each position were constructed. We repeated the described sampling process and constructed 4-step alternatives with an irreversible reaction at the first step. The following parameter values were drawn for one of the reference systems in our sampling:

$$
\begin{aligned}
&g_{1,1} = -1.3865 && g_{2,2} = -3.8822 && g_{3,3} = -3.5399 && g_{4,4} = -3.0146 \\
&g_{1,0} = 0.4422 && g_{2,1} = 1.1487 && g_{3,2} = 2.4753 && g_{4,3} = 0.1503 \\
&g_{5,4} = 3.2397 && g_{1,4} = -1.8842 && g_{5,5} = 0.4619 && \alpha_1 = 1.0103 \times 10^{-2} \\
&\alpha_2 = 473.31 && \alpha_3 = 6.8966 \times 10^{-2} && \alpha_4 = 8.2053 \times 10^3 && \alpha_5 = 1
\end{aligned}
\tag{2}
$$

Applying the constraints from [13] yields the following alternative:

$$
\begin{aligned}
&g_{1,1} = 0 && g_{2,2} = -3.8822 && g_{3,3} = -3.5399 && g_{4,4} = -3.0146 \\
&g_{1,0} = 0.4422 && g_{2,1} = 1.1487 && g_{3,2} = 2.4753 && g_{4,3} = 0.1503 \\
&g_{5,4} = 3.2397 && g_{1,4} = -290.7305 && g_{5,5} = 0.4619 && \alpha_1 = 4.8260 \times 10^{174} \\
&\alpha_2 = 473.31 && \alpha_3 = 6.8966 \times 10^{-2} && \alpha_4 = 8.2053 \times 10^3 && \alpha_5 = 1
\end{aligned}
\tag{3}
$$

As required by the defined constraints, the steady-state values, log gains with respect to supply, and sensitivity with respect to α_1 are equivalent in the reference and alternative. However, the application of these constraints resulted in a kinetic order ($g_{1,4}$ = -290.7) and a rate constant ($\alpha_1 = 4.8 \times 10^{174}$) that are well beyond the range of reasonable values. Since our prior belief is that kinetic orders should have magnitudes less than 5, this finding gives rise to concern that the sampled reference is being compared to an unrealistic alternative in the cases studied. We therefore recommend that references resulting in unrealistic alternatives be eliminated from consideration in statistical comparisons and that the rate of occurrence of unrealistic alternatives be evaluated as part of the method.

In most cases, a parameterized model, defined by its parameter values and implied structure, is but a sample from a population of models that might all represent the given design. In these cases one must question the gener-

alizability or robustness of statements made when point estimates for these parameter values are used in a controlled comparison. Consider, for example, the immune response model described in [8,21]. The referenced study compares the functional effectiveness of systems with and without suppressor lymphocyte regulation of effector lymphocyte production (Figure 1). Antigen and effector step responses to a four-fold increase in systemic antigen were included as functional effectiveness measures in this study. The authors developed time courses for both the reference (with suppression) and alternative system (without suppression) for a specific set of kinetic orders and rate constants determined to be reasonable based on prior knowledge of the system being studied. They compared time courses and concluded that the system with suppression was superior to one without suppression with respect to the peak antigen and effector levels in response to the step challenge. We repeated their calculations and reproduce the time courses in Figure 4A. As they observed, the peak levels are lower in the reference system. Next we examined the step response for models drawn from a narrow neighborhood about the selected parameters and found that the conclusion does not hold in general. Figure 4B illustrates the step response for one such case. We see that for this case the system without suppression is superior with respect to peak effector level. The analysis described by Irvine and Savageau, which however preceded the extensions of Alves and Savageau by 15 years, requires statistical methods to fully explore the regulatory preferences of the immune system. We provide this example as reinforcement to the recommendations of Alves and Savageau and for reference as we repeat the comparison of regulatory preferences in the immune system model in the sections that follow.

Methods
Below we describe the proposed enhancement to the method of mathematically controlled comparisons. We set the following requirements in the development of this method. (1) In the limit, as the alternative is forced to be fully equivalent, the conclusions of the improved method must match those of the currently defined method for cases in which the current method provides unambiguous conclusions and the alternatives are reasonable with respect to our prior knowledge of the parameter ranges. (2) The improved method should allow for various levels of equivalence ranging from alternatives independent of the reference to alternatives that are both internally and externally equivalent to the reference. (3) The improved method must avoid comparisons of unreasonable alternatives. (4) Finally, the improved method must provide a statistically meaningful measure comparable across various levels of equivalence and must allow for a test of homogeneity of conclusions across those levels. The statistical model and the procedure for implementation of
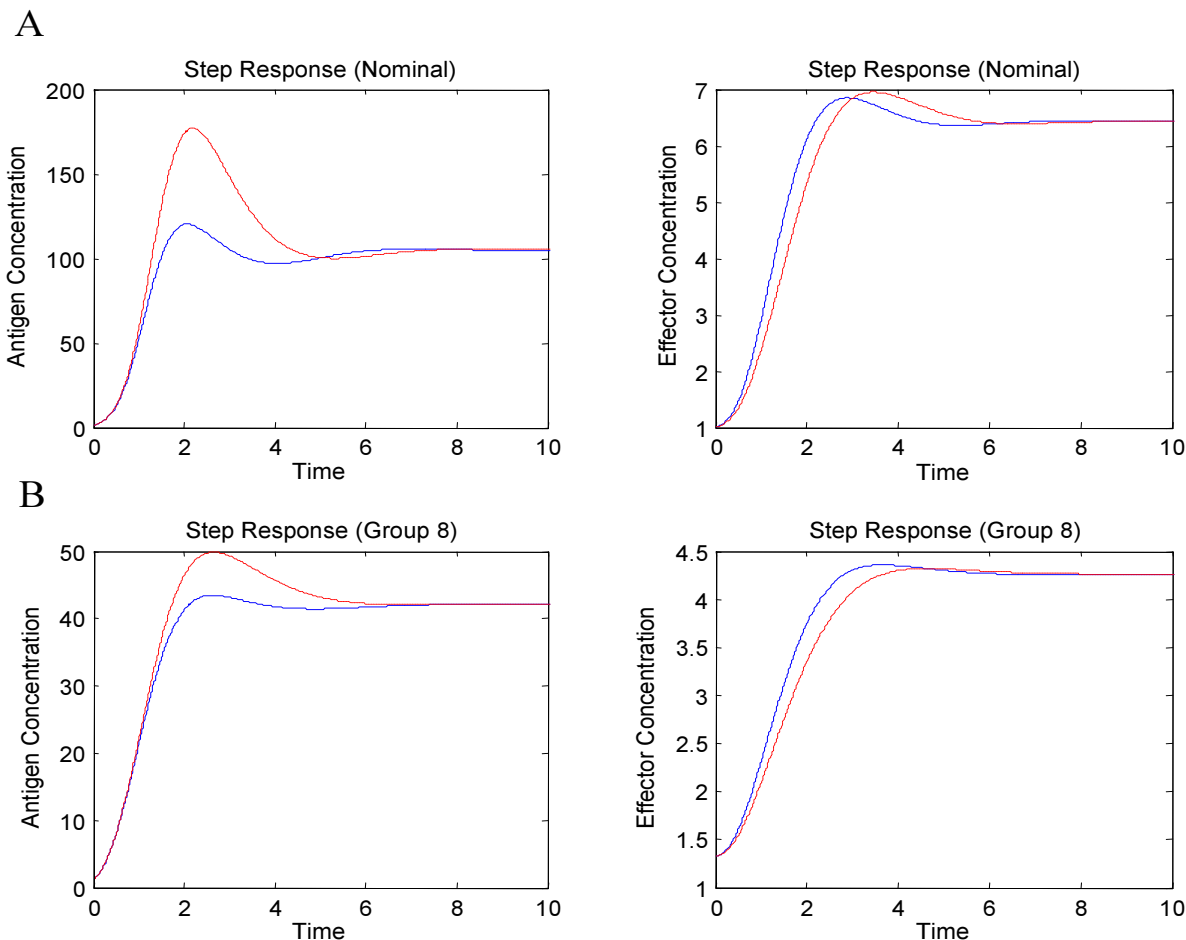
Figure 4
Step responses to source antigen increase. Step responses to a four-fold increase in source antigen are presented for both the nominal values (panel A) (from Irvine and Savageau) and for a case in which the values were drawn from a narrow distribution about those nominal values (panel B). Systemic antigen responses are shown on the left and effector on the right. Solid lines indicate the response for the reference system (with suppression) and dashed lines are used for the alternative. Step responses for the nominal values indicate a preference for the system with suppression. The step responses for the sampled case indicate a preference for the alternative when considering dynamic peaks for effector concentration.

the method are described below. An example of its application is given in the Results section.

Statistical Methods for Comparison of Alternatives
As described above, a controlled comparison under the extensions of Alves and Savageau is similar to a prospective study in epidemiology. In both cases we sample from a population, construct comparison groups, observe the frequency of outcomes for a given measure of effectiveness, and estimate a relative magnitude of effect that indicates the preference for one group over the other with

respect to that outcome. In epidemiological studies, these comparisons are supported by the methods of categorical data analysis where observations are separated into groups based on common traits (reference and alternative in this study). Categorical data analysis has a strong theoretical basis, has been applied extensively, provides meaningful measures of preference in the form of odds or odds ratios, and allows for the assessment of statistical significance in those measures. For these reasons we have chosen to employ the methods of categorical data analysis in performing controlled comparisons [22].

We begin by defining the categories of observations important to our analysis. In this analysis we wish to compare a reference design to an alternative design at K levels of equivalence. Each level of equivalence defines a set of constraints on the alternative that make it equivalent to the reference with respect to one or more properties. There are, therefore, $K + 1$ comparison groups in this analysis where the first group includes all instances of the reference design and the $k + 1$st group contains all instances of alternative designs at equivalence level k. Instances of alternative designs at level k are equivalent to their paired references with respect to the same set of constraints. Although not a requirement of the method, we generally order the application of constraints to form increasing levels of equivalence. At the lowest level, the model parameters of the alternative design instance and those of the paired reference are independent. The reference and the alternative share only the values of the independent variables and thus are subjected to the same external environment. The next level constrains the alternative instance to be internally equivalent to its paired reference in addition to sharing common values for the independent variables. Increased levels of equivalence successively apply constraints eventually resulting in full external equivalence, the highest level of equivalence. The number of constraints applied determines the number of levels of equivalence and thus the number of comparison groups.

Applying constraints in the construction of alternatives causes the alternative to be statistically dependent on the paired reference because its parameters are determined from those of the reference and they share a common set of values for the independent variables. When comparing the alternative and reference designs we must, in our statistical model, account for systematically high or low functional effectiveness resulting from this dependence. As such we define a second dimension of grouping to account for this effect. An instance of the reference design and all alternative instances derived from that reference are considered to be part of a matched group. If we sample J instances of the reference design and construct K alternatives from each reference instance we generate a population of $J \cdot (K + 1)$ samples in J matched groups. The resulting set of instances can then be viewed as being part of a J by $K + 1$ table where the $K + 1$ columns associated with the comparison groups and the J rows with the matched groups. We label a sample with the indices of this table, thus $S_{k + 1,j}$ is an instance of the alternative design at the kth equivalence level derived from the jth reference instance and $S_{1,j}$ is that paired reference instance.

Let $M(S_{k,j})$ be a measure that can be determined from the reference and alternative instances' parameter values and that orders their functional effectiveness. This measure is taken to represent the true merit of the design. We cannot,

however, directly measure the true merit of the design and must infer it from the measurement of M for samples from a population of instances that represent the design. We compute M for many instances of the reference design and its associated alternatives and compare those results to determine preference for one design over the other. Estimation of these preferences requires us to define an outcome that indicates the direction of preference. We can either independently compare the effectiveness measures for each instance to a common threshold and represent the resulting frequency of occurrences as an odds ratio or we can perform pairwise comparisons of each reference and its paired alternative and measure the frequency of occurrence as an odds. Each method has its advantages. Consider a comparison in which the reference is always better than the alternative but only by an infinitesimally small amount. In the first approach we would probably detect no difference between the two designs because when compared to a common threshold both groups would demonstrate about the same odds (an odds ratio of 1) of exceeding the threshold. In the second approach we would find the odds of preferring the reference design to be infinite as it is always better than the alternative even though only infinitesimally so. As with most applications of statistics, the key to the appropriate choice is in the question to be answered. For applications of controlled comparisons we recommend inclusion of both methods of comparison as they provide both a measure of the magnitude of the difference and allow us to detect strict but small differences that may have biological significance.

Method 1

Let W be a threshold such that systems for which $M > W$ are taken to be part of a functionally desirable class. Membership in this desirable class is therefore represented by a dichotomous variable given by the outcome of such a test. We formally define this as follows.

$$Y_{k,j} = \begin{cases} 1 & \text{if} \quad M(S_{k,j}) > W \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

All alternatives in the same group j are derived from the same reference instance, $S_{1,j}$, and therefore the $Y_{k,j}$ within a matched group are correlated. We wish to compare the odds of an instance of the reference design being a member of the desirable class to the odds of an instance of the alternative design, at equivalence level k. The following log-linear model is used.

$$\text{logit}\left(p_{k,j}\right) = \theta_1 x_{1,j} + \theta_2 x_{2,j} + \theta_3 x_{3,j} + \ldots + \theta_K x_{K,j} + \sum_{q=1}^{J} \gamma_q \omega_{q,j}$$
$$p_{k,j} = \Pr\left(Y_{k,j} = 1\right) \tag{5}$$

where

- $Y_{k,j}$ is the outcome for $S_{k,j}$ (1 = member of the desirable class, 0 = not a member of the desirable class) with respect to M and threshold W,

- $X_{k,j}$ are indicators taking value 1 if the instance is an alternative at equivalence level k formed from the jth reference instance.

- $\omega_{q,j}$ are a collection of J indicator variables where $\omega_{q,j}$ takes value 1 if $q = j$ and 0 otherwise.

The parameters (θ_k) are estimated by conditioning out the nuisance variables (γ_q) using conditional logistic regression. The $\exp(\theta_k)$ then give the odds ratios for desirable class membership comparing alternative structure at equivalence level k to the reference structure after controlling for group effects. The methods of categorical data analysis and logistic regression are described in many texts on statistics, for example [22].

This method allows us to address structural preference with respect to M by independently comparing both the population of reference systems and the population of alternative systems to a common threshold to determine odds of membership in the desirable class after controlling for group effects. The odds of membership for the reference are compared to the odds for the alternative in the odds ratios estimated in the regression. Ratios found to be significantly different from 1 indicate a preference with respect to measure M. For this method to be applied we must choose threshold W. For consistency of comparison with Method 2 we choose W to be the median of the observed values of M for instances of the reference. Although this selection for W is somewhat arbitrary, it has the desirable effect of making the odds of class membership for the reference system equal to 1.

Method 2
The method above provides us with a comparison of the alternative design and reference design based on a common threshold test. In Method 2 we perform a pairwise comparison of each alternative design instance and its paired reference and compute the odds that the reference is better than its paired alternative with respect to the measure of comparison. For this assessment we consider the general linear model for paired comparison [23]. Under this model the probability that design D_i is preferred over design D_j is then given by

$$\pi_{i,j} = F(M(D_i) - M(D_j)) \quad (6)$$

Where $F(\cdot)$ represents a symmetric cumulative distribution function centered at 0, M measures the true merit of the design, and $\pi_{i,j}$ is the probability that D_i is preferred over D_j with respect to measure M. When the logistic dis-

tribution is assumed for $F(\cdot)$, the linear model is equivalent to the Bradley-Terry Model for paired comparisons (see description in [23]). The Bradley-Terry model is most often associated with analysis of orderings of objects in paired comparisons such as paired competitions in sports or in subjective pairwise comparisons like wine tasting. In our application we compare, pairwise, the reference design to several alternative designs under various levels of equivalence. Each new reference and its associated alternative instances yields a new set of observations from matched comparisons of computed measures of effectiveness. Currently we consider only one reference and one alternative design under various levels of equivalence. We can, however, extend the model to include multiple designs which could be compared simultaneously. Such a model would be useful in Alves and Savageau's study of preferred irreversible step positions in biosynthetic pathways [13]. Each possible irreversible step location could be included as another alternative in the statistical model. For our purposes, we continue with the model comparing two designs which we describe as follows:

$$\pi_{1,k} = F\left(M\left(S_{1,.}\right) - M\left(S_{k,.}\right)\right)$$
$$\mathrm{logit}\left(\pi_{1,k}\right) = \beta_R x_R - \beta_A x_A + \gamma_1 e_1 + \cdots + \gamma_K e_K \quad (7)$$
where

x_R is an indicator variable taking value 1 if the reference is used in the comparison (always 1).

x_A is an indicator variable taking value 1 if the alternative is used in the comparison (always 1).

e_k are indicator variables taking value 1 if the comparison is being made at equivalence level k.

The indicators, e_k, representing the equivalence levels of the comparisons are treated as covariates in the model. The indicators x_R and x_A take fixed values for our example as we are comparing only two designs. A more general form of the model can be constructed to compare several design alternatives. For the reference instance and each paired alternative instance we compute the effectiveness measure $M(\cdot)$. We perform pairwise comparisons between the reference and each associated alternative to yield K outcomes per group and the data is then fit by logistic regression (without intercept). Under the given parameterization, the design matrix does not have full rank and so we employ the constraint $\beta_R - \beta_A = 0$. In this way, the regression parameter γ_k gives the log odds of preference for the reference versus the alternative at the kth level of equivalence. Performing pairwise comparisons only within matched groups eliminates within group dependencies. This method allows us to detect a preference for the reference (or alternative) independent of the

magnitude of the difference as measured by *M* as it depends only on the frequency with which the effectiveness of a reference exceeds that of a paired alternative.

Procedure for Controlled Comparisons
This section provides a step-by-step procedure for controlled comparisons under the proposed enhancements. Primary differences between the enhanced method and prior applications of controlled comparisons occur in steps 6 through 10.

Step 1 – Model Development
Using the chosen mathematical framework we develop a mathematical model for the designs being compared, identifying each of the dependent and independent variables and the differential equations describing the behavior of the dependent variables. The mathematical representation is derived from the biochemical map of the system under study using the procedures described in [9]. We identify the parameters associated with the process or step of interest and identify the parameters fixed by the definition of the alternative (e.g., fixing a kinetic order at 0 for an influence we wish to eliminate in the alternative).

Step 2 – Identification of Functional Effectiveness Measures
Based on our knowledge of the system's function we identify functional effectiveness measures. This step is dependent on the system under study. Previous studies have employed measures of margin of stability [13,14,17], sensitivity [14,17,21], aggregated sensitivity [13], logarithmic gains [13-15,21], response time [13,14,20,21], and step response overshoot [21]. These measures are computed through either steady-state or dynamic analysis using the mathematical framework.

Step 3 – Determination of Sampling Space
We identify distributions representing our prior knowledge for each of the parameters. These sampling distributions represent the population of models being studied. The sampling space is chosen based on estimated variability in the model parameters (based on regression results) or on uncertainty in our prior opinion about the parameters. In cases where the parameter value distributions are not known, a uniform distribution is employed. All conclusions of the analysis are conditioned on the chosen sampling space.

Step 4 – Identification of Constraints
We identify constraints that reduce the differences in the reference and alternative design instances. These constraints are defined in terms of steady-state systemic properties that can be computed from the mathematical model (steady-state values of dependent variables, logarithmic gains, sensitivities, etc.). Previous studies have employed steady-state values of dependent variables [13-15,20,21],

specific logarithmic gains [13-15,20,21], combinations of logarithmic gains [21], or specific sensitivities [13] in the definition of constraints. For each constraint, we identify a relationship that fixes remaining free parameters in terms of the parameters of the paired reference instance. Constraint relationships are determined using symbolic steady-state solutions developed with a computer algebra system such as the Matlab Symbolic Toolbox. For this study we have employed BSTLab, a Matlab toolbox capable of developing symbolic solutions for S-system steady states, sensitivities, and logarithmic gains [24].

Step 5 – Sampling of the Reference Design's Population
We construct an instance of a reference design by sampling model parameter values from the distributions defined in Step 3. The model structure and the sampled parameters fully define one instance of the reference system. For this study we sampled 1,000 reference design instances for the main results and an additional 5,000 instances to confirm some of our findings.

Step 6 – Construction of Alternatives
For each sampled reference design we construct one or more alternatives by applying the constraints identified in Step 4. We first construct an independent alternative by sampling parameters from the distributions defined in Step 3 followed by the application of constraints on the parameters that are fixed by the alternative design's structure. We then construct additional alternatives by the application of constraints starting with internal equivalence and ending with full (internal and external) equivalence. The parameters computed through the application of constraints in the alternative are then checked against the range of reasonable parameter values. Sampled references and associated alternatives are discarded when any of their parameters exceed the range of reasonable values. Steps 5 and 6 are repeated until the desired sample size is achieved.

Step 7 – Evaluation of Functional Effectiveness
Functional effectiveness measures, identified in Step 2, are computed for instances of the reference and associated alternatives. Alternatives and references are compared to the common threshold (for Method 1) and each alternative is compared to its associated reference (for Method 2) with respect to each measure. For Method 1 a binary outcome is recorded for each instance and effectiveness measure and the outcomes for Method 2 are recorded as categorical values indicating that the reference is better than, equal to, or worse than the alternative with respect to the given performance measure.

Step 8 – Analysis of Outcomes
We analyze the outcomes for each case using conditional logistic regression (for Method 1) or logistic regression

(for Method 2). The estimated parameters for the regression model can then be interpreted as odds ratios (comparison to a common threshold) or odds (paired comparisons) for preference of the reference system over the alternative given a specified level of equivalence. The analysis also provides confidence intervals on these parameters allowing us to measure the significance of our statements with respect to the given sampling of the reference design population. We perform this analysis using a statistical computing system such as R [25].

Step 9 – Identification of Significant Differences
Odds or odds ratios found to be statistically significant indicate differences between the reference and alternative populations. Odds or odds ratios that are not significantly different from the null value of 1 are taken as an indication that in this sampling there is no evidence of a difference between the reference and alternative design with respect to the given performance measure at the given level of equivalence. The ability to detect small differences in preference depends on the size of the sample used in the analysis. In these studies we have taken between 1,000 and 5,000 randomly constructed groups (one reference and one or more alternatives). We summarize these data in the form of analysis tables giving the odds and odds ratios for these comparisons along with indications of significance and indications of those measures fixed by equivalence.

Step 10 – Generalization of Differences
We next examine the homogeneity of conclusions across the levels of equivalence with respect to the direction of the effect and with respect to magnitude. Where statistically meaningful differences are required, contrasts on the regression parameters are computed.

Results
We illustrate the proposed enhancements by repeating the analysis of network regulation in the immune system performed by Irvine and Savageau [21] and summarized in [8]. In particular, we focus on their comparison of systems that include suppression of effector lymphocyte production and those that do not. The schematic representations of the reference design (with suppression) and the alternative (without suppression) are given in Figure 1. The only difference in the designs occurs in the step associated with the production of effector lymphocytes where, in the reference design, the production is inhibited by the concentration of suppressor lymphocytes. Using the procedures in [9] the system of equations is written as follows.

$$\dot{X}_1 = \alpha_1 X_1^{g_{1,1}} X_4^{g_{1,4}} - \beta_1 X_1^{h_{1,1}} X_2^{h_{1,2}}$$
$$\dot{X}_2 = \alpha_2 X_1^{g_{2,1}} X_3^{g_{2,3}} X_5^{g_{2,5}} - \beta_2 X_2^{h_{2,2}} \qquad (8)$$
$$\dot{X}_3 = \alpha_3 X_2^{g_{3,2}} X_6^{g_{3,6}} - \beta_3 X_3^{h_{3,3}}$$

In this model all of the $g_{i,j}$ and $h_{i,j}$ are greater than 0 except for $g_{2,3}$ which takes values less than 0 in the reference design and is fixed equal to 0 in the alternative.

To facilitate comparison with the results of Irvine and Savageau, we select the same seven functional performance measures. The first two performance measures are the basal levels of systemic antigen and effector lymphocytes, determined by the steady-state values of X_1 and X_2. We also include the antigenic gain and the effector gain determined by $L_{1,4}$ and $L_{2,4}$. Dynamic analysis yields two more measures given by the magnitude of the overshoot of systemic antigen and effector lymphocytes in response to a four-fold step increase in source antigen. These values are determined by integrating the system of equations for each case, initially at steady state, in response to the four-fold increase in source antigen. The difference between the peak value of the time course and the new steady state as a fraction of the new steady-state value are taken as the functional performance measure. Finally we include the sensitivity of the logarithmic gain $L_{1,4}$ with respect to parameter $h_{2,2}$ as a measure of the system sensitivity with respect to parameter variation ($S(L_{1,4}, h_{2,2})$). In all cases, lower values indicate a more desirable design. The rationale for the selection of these measures is given in [21].

Values for each of the parameters are sampled in a neighborhood about the parameter values given in [8] from the following distributions.

$$\log_{10}(\alpha_1) \sim N\left(\log_{10}(2.02), \sigma^2\right)$$
$$\log_{10}(\beta_1) \sim N\left(\log_{10}(2), \sigma^2\right)$$
$$\log_{10}(\alpha_2), \ \log_{10}(\beta_2), \ \log_{10}(\alpha_3) \sim N\left(\log_{10}(1), \sigma^2\right)$$
$$\beta_3 = 1$$
$$g_{1,4}, \ g_{2,5}, \ g_{3,6}, \ h_{1,1}, \ h_{2,2}, \ h_{3,3} \sim N\left(1, \sigma^2\right) \qquad (9)$$
$$g_{1,1} \sim N\left(0.9, \sigma^2\right)$$
$$g_{2,3} \sim N\left(-0.5, \sigma^2\right)$$
$$g_{2,1}, g_{3,2}, h_{1,2} \sim N\left(0.5, \sigma^2\right)$$

The rate constant β_3 is fixed to set the time scale. When sampling instances of the alternative design, the value of $g_{2,3}$ is set to 0. The distributions for kinetic orders are truncated to prevent positive kinetic orders less than 0.1 and negative kinetic orders greater than -0.1. The values of the

independent variables X_4, X_5, and X_6 are sampled from the following distributions.

$$\log_{10}(X_4), \log_{10}(X_5), \log_{10}(X_6) \sim N(0, \sigma^2) \quad (10)$$

The analysis is conducted at $\sigma = 0.1$ for baseline results and again at $\sigma = 0.2$ to address sensitivity with respect to the sampling space.

We define four levels of equivalence for this analysis. The lowest level requires only that the alternative instance adhere to the sampling distributions for parameters described above. Instances drawn as such are representatives of the alternative design but do not exhibit internal or external equivalence with the paired reference system. The only values shared between the alternative and the paired reference are those of the independent variables. We next define an internally equivalent alternative as one in which the values for α_1, α_3, β_1, β_2, β_3, $g_{1,1}$, $g_{1,4}$, $g_{3,2}$, $g_{3,6}$, $h_{1,1}$, $h_{1,2}$, $h_{2,2}$, and $h_{3,3}$ in the alternative are equal to the associated values in the paired reference. We define a partially equivalent alternative as one that is both internally equivalent and that also has steady-state values of dependent variables equal to those of the paired reference. This constraint is used to fix the value of α_2 in the alternative instance. Symbolic solutions for the steady-state values of the dependent variables are computed for both the reference and alternative, expressions for X_2 are set equal and solved for α_2. To construct a partially equivalent alternative we form an internally equivalent alternative and additionally compute the value of α_2 to satisfy the constraint. The fourth level of equivalence, referred to as full equivalence, incorporates constraints sufficient to remove all remaining degrees of freedom in the formation of the alternative. The fully equivalent form is constructed by forcing two additional constraints which fix the values of $g_{2,1}$ and $g_{2,5}$. The value for $g_{2,1}$ is determined by requiring that $L_{1,4}$ be equal in both the reference and alternative systems and the value for $g_{2,5}$ is determined by requiring that $L_{1,5} + L_{1,6}$ be equal in both systems. Symbolic solutions for the log gains are computed, set equal and solved for the desired parameter. Fully equivalent systems are then formed by forcing internal equivalence and by fixing α_2, $g_{2,1}$, and $g_{2,5}$ to satisfy the given constraints. The selected constraints match those employed by Irvine and Savageau [21].

Groups of cases were constructed by sampling the parameter distributions defined above. For each group we drew one set of values for the parameters of the reference, one set of values for the independent variables, and for each level of equivalence we drew a set of values for those parameters not fixed by the associated constraints. Five cases were constructed for each of 1,000 groups consisting of one reference and four alternatives, one for each of the

four levels of equivalence. Fourteen cases were identified having values of $g_{2,5}$ less than 0.1 (the lower bound on our accepted range for $g_{2,5}$). Two of those 14 cases had values of $g_{2,5}$ less than 0 demonstrating that application of the constraints can lead to a change in the alternative design (an activation becomes and inhibition) unless checks for reasonableness of constrained parameters are employed. Groups were also checked for the existence of stable steady states. A total of 160 groups were eliminated because of unreasonable parameters or because one or more cases in the group did not exhibit a stable steady state at either the baseline conditions (X_4, X_5, and X_6) or at the steady state for a four-fold increase in source antigen. Summary statistics were computed for each of the constrained parameters (Table 1). Eliminating the groups identified above resulted in 4200 cases in 840 groups that were further analyzed.

Functional performance measures were computed for each case and analyzed according to both Method 1 and Method 2 using the statistical package R [25]. Density plots for each of the effectiveness measures and associated medians for the reference design are given in Figure 5. For Method 1, functional effectiveness measures for all cases were compared to the median over all reference cases. The outcome was defined such that odds ratios greater than 1 indicate a preference for the reference design. For Method 2, functional effectiveness measures for each case were compared to the functional effectiveness measures of the paired reference. Cases in which the performance measures were equal were eliminated prior to the regression as they contribute no information to an estimate of preference. Outcomes were again defined such that odds greater than 1 indicate a preference for the reference design. The conditional logistic regressions for Method 1 were computed using the clogit function and the logistic regressions for Method 2 were computed using glm with binomial family and logit link. Regression parameters were exponentiated to yield odds ratios (Method 1) and odds (Method 2). Results with $\sigma = 0.1$ for Method 1 are given in Table 2 and for Method 2 in Table 3. Comparisons that were fixed by constraint are indicated in the table. Regression parameters were considered to indicate a preference if they were determined to be statistically different from the null value at the $\alpha = 0.05$ level. The process was repeated for $\sigma = 0.2$ to examine sensitivity with respect to the sampling distributions. Those results are given in Tables 4 and 5.

We find that the first two measures (basal levels of antigen and effector lymphocytes) show no statistically significant difference between the reference and alternative design even when the reference and alternative instances are independent. This finding holds for both methods of analysis. Examining the pairwise scatter plots for these

Table 1: Summary statistics for model parameters. Summary statistics for all of the parameters fixed by equivalence were computed across the three alternatives for which equivalence constraints were applied after elimination of cases which violated the requirement for steady state or lower bounds for $g_{2,1}$ or $g_{2,5}$.

Property	Parameter		
	α_2	$g_{2,1}$	$g_{2,5}$
Minimum	0.065	0.137	0.100
1st Quartile	0.830	0.395	0.501
Median	1.001	0.468	0.928
Mean	1.074	0.470	0.804
3rd Quartile	1.203	0.544	1.032
Maximum	22.965	0.843	1.398
Standard Deviation	0.682	0.107	0.300
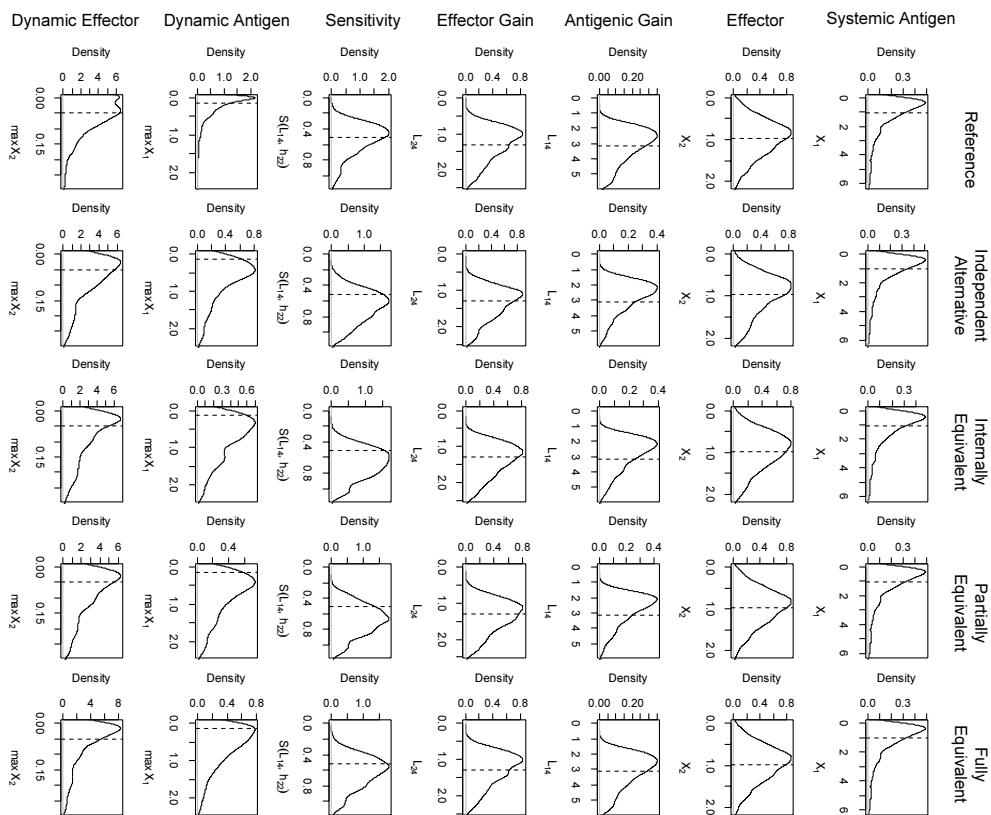
Figure 5
Performance measure distributions. Smoothed histograms were computed using R (density function) and are presented for each of the 7 functional performance measures for each of the equivalence levels. The dashed vertical line indicates the median of the distribution for the reference system. The alternative is preferred when more of its probability mass is distributed to the left of the dashed line.

Table 2: Odds ratios for comparison using Method 1. Odds ratios for preference of system with suppression using conditional logistic regression (Method 1). The data consisted of 840 groups of 5 cases each (reference and four alternatives) sampled as described. Values shown give the odds that the reference system is preferred over the alternative. Values marked with asterisks are significant at the $\alpha =$ 0.05 level. Kinetic orders and rate constants were sampled with $\sigma = 0.1$. Cells marked with an odds ratio of 1 or = indicate cases that were fixed by equivalence.

Functional Performance Objective	Independent Alternative	Internally Equivalent	Partially Equivalent	Fully Equivalent	Irvine and Savageau
Minimize Basal Level of Systemic Antigen	0.76	0.94	1	1	=
Minimize Basal Level of Effector Lymphocytes	0.85	1.1	1	1	=
Minimize Antigenic Gain	0.34*	0.39*	0.36*	1	=
Minimize Effector Gain	1.4*	1.5*	1.8*	1	=
Minimize Dynamic Levels of Antigen	32*	28*	34*	11*	+
Minimize Dynamic Levels of Effector	2.3*	1.9*	2.4*	0.97	+
Minimize Sensitivity to Parameter Variation	11*	12*	13*	5.5*	+

Table 3: Odds ratios for comparison using Method 2. Odds for preference of system with suppression using paired comparisons and logistic regression (Method 2). The data consisted of 840 groups of 5 cases each (reference and four alternatives) sampled as described. Values shown give the odds that the reference system is preferred over its paired alternative. Values marked with asterisks are significant at the $\alpha = 0.05$ level. Kinetic orders and rate constants were sampled with $\sigma = 0.1$. Cells marked with an odds ratio of 1 or = indicate cases that were fixed by equivalence. [†] Odds of ∞ shown for measures were reference was found to be better than the alternative in every case sampled.

Functional Performance Objective	Independent Alternative	Internally Equivalent	Partially Equivalent	Fully Equivalent	Irvine and Savageau
Minimize Basal Level of Systemic Antigen	0.91	0.98	1	1	=
Minimize Basal Level of Effector Lymphocytes	1.0	0.98	1	1	=
Minimize Antigenic Gain	0.63*	0.31*	0.25*	1	=
Minimize Effector Gain	1.2*	2.1*	2.2*	1	=
Minimize Dynamic Levels of Antigen	3.9*	69*	55*	55*	+
Minimize Dynamic Levels of Effector	1.6*	1.9*	2.2*	1.2*	+
Minimize Sensitivity to Parameter Variation	2.2*	22*	21*	∞[†]	+

cases (Figure 6) we see little correlation at this level of equivalence. The introduction of internal equivalence creates a much stronger correlation between the values but the correlation follows the null line (same basal level in both reference and alternative) and thus still shows no sta-

tistically significant difference for paired comparisons. The introduction of partial and full equivalence requires these measures to be the same and thus contributes no information to the comparison.

Table 4: Odds ratios using Method 1 with increased sampling variance. Results from sensitivity analysis using increased variance in sampled population (σ = 0.2 versus σ = 0.1) giving odds ratios for preference of system with suppression using conditional logistic regression (Method 1). The data consisted of 460 groups of 5 cases each (reference and four alternatives) sampled as described. Values shown give the odds that the reference system is preferred over the alternative. Values marked with asterisks are significant at the α = 0.05 level. Kinetic orders and rate constants were sampled with σ = 0.2. Cells marked with an odds ratio of 1 or = indicate cases that were fixed by equivalence.

Functional Performance Objective	Independent Alternative	Internally Equivalent	Partially Equivalent	Fully Equivalent	Irvine and Savageau
Minimize Basal Level of Systemic Antigen	1.2	0.87	1	1	=
Minimize Basal Level of Effector Lymphocytes	0.92	1.1	1	1	=
Minimize Antigenic Gain	1.2	0.61*	0.48*	1	=
Minimize Effector Gain	2.2*	1.6*	1.4*	1	=
Minimize Dynamic Levels of Antigen	4.9*	4.5*	3.2*	3.8*	+
Minimize Dynamic Levels of Effector	0.62*	0.59*	0.50*	0.44*	+
Minimize Sensitivity to Parameter Variation	5.4*	5.3*	5.2*	2.5*	+

Table 5: Odds ratios using Method 2 with increased sampling variance. Results from sensitivity analysis using increased variance in sampled population (σ = 0.2 versus σ = 0.1) giving odds for preference of system with suppression using paired comparisons and logistic regression (Method 2). The data consisted of 460 groups of 5 cases each (reference and four alternatives) sampled as described. Values shown give the odds that the reference system is preferred over its paired alternative. Values marked with asterisks are significant at the α = 0.05 level. Kinetic orders and rate constants were sampled with σ = 0.2. Cells marked with an odds ratio of 1 or = indicate cases that were fixed by equivalence. † Odds of ∞ shown for measures were reference was found to be better than the alternative in every case sampled.

Functional Performance Objective	Independent Alternative	Internally Equivalent	Partially Equivalent	Fully Equivalent	Irvine and Savageau
Minimize Basal Level of Systemic Antigen	1.2	0.82*	1	1	=
Minimize Basal Level of Effector Lymphocytes	0.98	1.2	1	1	=
Minimize Antigenic Gain	0.89	0.59*	0.58*	1	=
Minimize Effector Gain	1.4	1.5*	1.6*	1	=
Minimize Dynamic Levels of Antigen	1.8*	5.8*	6.8*	30*	+
Minimize Dynamic Levels of Effector	0.98	0.76*	0.69*	0.55*	+
Minimize Sensitivity to Parameter Variation	2.1*	5.5*	6.0*	∞†	+

Functional effectiveness measures of antigenic gain and effector gain show differing preference for the alternative and reference designs. The alternative design shows statistically significant odds of having a lower antigenic gain than the reference for the independent, internal, and partial equivalence cases. The reference system, however, demonstrates statistically significant preference for reduced effector gain. These findings are again consistent
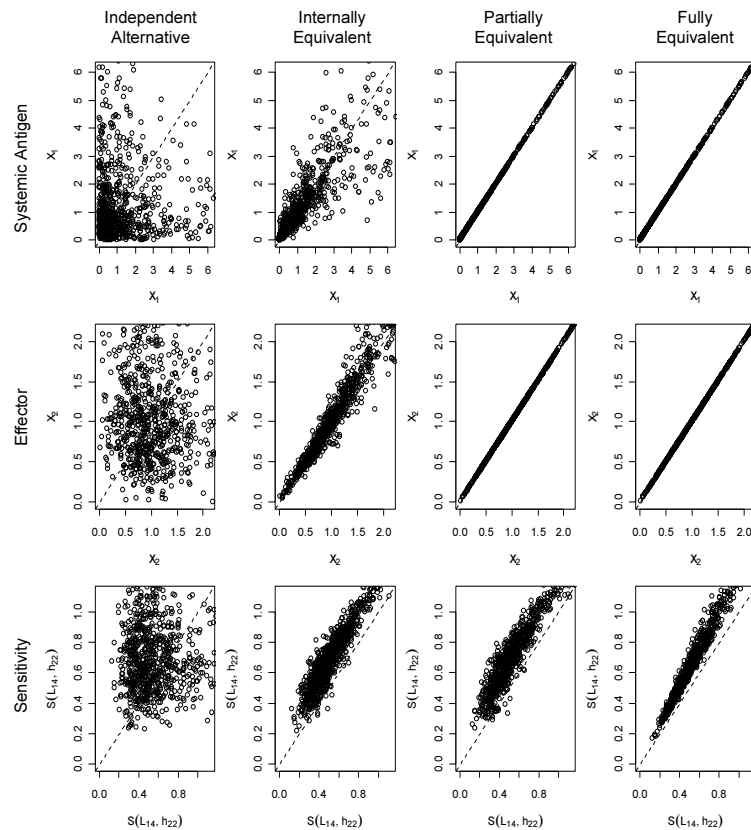
Figure 6
Pairwise comparison of selected performance measures. Scatter plots of 840 pairs were produced in R and show basal systemic antigen level (top), basal effector level (middle), and sensitivity (bottom) in the reference (x-axis) versus the same value in the associated alternative (y-axis) for all equivalence levels. The dashed diagonal line corresponds to the null condition where the reference and the alternative are equal. The reference is preferred over the alternative when more pairs fall above the diagonal. System sensitivity shows a clear preference for the reference with all cases preferring the reference under full equivalence.

across both methods of analysis. Introducing constraints to achieve full equivalence causes these gains to be equal in both the reference and alternative. The introduction of constraints on $L_{1,4}$ and $L_{1,5} + L_{1,6}$ hides the differential preference with respect to effector and antigen gain. Analysis using the incremental introduction of constraints permits the observance of these differences.

Dynamic responses for both antigen peak levels and effector peak levels demonstrate a preference for the reference design. Even for an independent alternative (no internal or external equivalence), the odds of an instance with sup-

pression being preferred over an instance without suppression with respect to dynamic levels of antigen and effector are 32 and 2.3 for Method 1 and 3.9 and 1.6 for Method 2. The strong preference for the design with suppression is seen at all levels of equivalence for both Method 1 and Method 2 when considering dynamic levels of antigen. A significant preference for the design with suppression is also seen for dynamic levels of effector. The finding under this measure does lose significance under full equivalence for Method 1. We conclude that the reference design is better than or equal to the alternative in all cases.

Sensitivity to parameter variation, as measured by $S(L_{1,4}, h_{2,2})$ demonstrates a clear preference for the design with suppression. In population level comparisons (Method 1), the odds of preference for the system with suppression were 11, 12, 13, and 5.5 for the four levels of equivalence. In pairwise comparisons (Method 2) the odds for preference of the reference were 2.2, 22, 21 for the independent, internally equivalent, and partially equivalent cases. Under full equivalence the reference was preferred in every case sampled. Clearly, the design with suppression is preferred, independent of our introduction of constraints.

Overall, the design with suppression is equal to or preferred in six out of the seven functional effectiveness measures and this conclusion can be generalized to the lowest level of equivalence. We also find that these observations apply to both population-based comparisons against a common threshold and controlled for group effects (Method 1) and paired comparisons (Method 2). We conclude that observations of preference for the system with suppression over the system without suppression are very general and are conditioned only on the sampling space used to draw members of the populations studied. We find that these observations differ from those of Irvine and Savageau only with respect to the minimization of antigenic gain and to some extent in the minimization of dynamic effector levels. This difference in finding, as determined by antigenic gain, would be expected as Irvine and Savageau employed a fully equivalent alternative in their comparison that was based on a constraint on $L_{1,4}$. Their study would not have been able to detect a difference in preference based on this measure. Differences in dynamic measures would also be expected as the dynamic data depends on the parameters of the system and a controlled comparison would, as discussed in previously, require a sampling of the parameter space.

Analysis of sensitivity with respect to the sampling distributions indicated that in 6 of the 7 measures, the preferences held as described above. For dynamic levels of effector, the preference shifted from the system with suppression (reference) to the system without (alternative). Under both methods and across all levels of equivalence the preference shifted away from the reference and either became insignificant or shifted slightly in favor of the alternative. To verify these findings we repeated the entire analysis with a larger starting sample (5000 groups versus 1000 groups, results not shown). The same shift in preference was observed for the larger sample size, and we conclude that our observations with respect to this measure should be conditioned on the chosen sample space.

Discussion

The primary purpose of this analysis was to demonstrate proposed enhancements to the method of mathematically controlled comparisons. Through this analysis we have demonstrated the incremental introduction of constraints and have identified that in one measure a preference for the alternative was detected consistently across levels of equivalence that would not have been identified through comparison using the fully equivalent alternative. The conclusions reached in this analysis, which qualitatively match those of Irvine and Savageau, are more general in that we demonstrate the existence of preference for the system with suppression with fewer constraints and that these preferences additionally appear in population-level comparisons to a common threshold (Method 1). Our use of methods from categorical data analysis (logistic regression) and odds (or odds ratios) as a measure of preference provides a statistically meaningful method for comparing the results; and estimation of the confidence interval for those measures provides a method for assessing the sufficiency of the sampling used in the analysis. The compilation of these results in a table of increasing equivalence provides a convenient display of the results and facilitates generalization to the fewest set of constraints.

The analysis also indicated that unreasonable alternatives can result from the application of constraints even when sampling in a small region about reasonable nominal values. By using the S-system based framework, we are relying on an equivalence between the canonical S-system representation and the biochemical network structure in this analysis. The application of constraints that result in kinetic orders that change sign or approach 0 is, therefore, equivalent to a change in the structure of the alternative. This would invalidate our assumption that the alternative constructed from the reference by application of constraints results in a representative of the alternative biochemical system design. The incorporation of a step that explicitly checks the reasonableness of constructed alternatives eliminates the potential for such problems.

Our proposed enhancement to the method of controlled comparisons provides results consistent with the currently defined method for the fully equivalent case, allows for the incremental introduction of constraints and generalization of results, eliminates comparison with unrealistic alternatives, and provides a consistent measure of preference that can be compared across levels of equivalence. We find that in this specific analysis, the results of Irvine and Savageau can be generalized and that some differential preference is observed for antigenic and effector gains that would not have been observable in their fully equivalent analysis. In general, we conclude that there are different levels of confidence at which we might declare

one design better than another and that the assessment of these differences requires a more complete exploration of the implications of constraints and sample spaces on the conclusions we reach. As we continue to search for design principles among the many pathways now being studied we need to fully characterize the contexts in which a preference exists. The method of mathematically controlled comparisons coupled with the canonical nonlinear representations of S-systems and well-chosen statistical methods offers significant potential to facilitate these searches.

Competing Interests
None declared.

Author's Contributions
JHS developed the approach, implemented the software, conducted the tests, and prepared the results. EOV provided the theoretical framework, identified the test case, and supported the evaluation and interpretation of the results.

Appendix
S-systems are systems of ordinary differential equations of the following form [26].

$$\frac{dX_i}{dt} = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{i,j}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{i,j}} \text{ for } i \in 1...n$$

The S-system representation offers a particular advantage in that it allows for the determination of the system's steady state by solution of a system of linear equations. Setting the time derivatives to 0 and taking logs results is a linear system of equations from which the log steady-state values of the *n* dependent variables can be determined from the values of the *m* independent variables and the system parameters [26].

$$\mathbf{y}_D = \mathbf{A}_D^{-1}\mathbf{b} - \mathbf{A}_D^{-1}\mathbf{A}_I\mathbf{y}_I$$
$$b_i = \ln\left(\beta_i\right) - \ln\left(\alpha_i\right)$$
$$a_{i,j} = g_{i,j} - h_{i,j} \qquad (11)$$
$$y_{D,i} = \ln\left(X_i\right) \text{ for } i \in 1...n$$
$$y_{I,i} = \ln\left(X_{i+n}\right) \text{ for } i \in 1...m$$

S-systems additionally provide for the convenient characterization of systemic performance local to the steady state. Logarithmic gains that represent the change in the log value of the steady state of a dependent variable or flux as a result of a change in the log value of an independent variable are computed as follows [26].

$$L\left(X_j, X_k\right) = \frac{\partial \ln X_j}{\partial \ln X_k} = \frac{\partial X_j}{\partial X_k}\frac{X_k}{X_j}$$

$$L\left(V_i^+, X_k\right) = \frac{\partial \ln V_i^+}{\partial \ln X_k} = g_{i,k} + \sum_{j=1}^{n} g_{i,j}L\left(X_j, X_k\right) \qquad (12)$$

$$L\left(V_i^-, X_k\right) = \frac{\partial \ln V_i^-}{\partial \ln X_k} = h_{i,k} + \sum_{j=1}^{n} h_{i,j}L\left(X_j, X_k\right)$$

System sensitivities measure the degree to which changes in the system parameters (kinetic orders and rate constants) affect the steady state of the system. Sensitivities are often used as a measure of the robustness of the system.

$$S\left(X_k, \alpha_j\right) = \frac{\partial \ln X_k}{\partial \ln \alpha_i} = \frac{\partial X_k}{\partial \alpha_i}\frac{\alpha_i}{X_k}$$

$$\qquad (13)$$

$$S\left(X_k, g_{i,j}\right) = \frac{\partial \ln X_k}{\partial \ln g_{i,j}} = \frac{\partial X_k}{\partial g_{i,j}}\frac{g_{i,j}}{X_k}$$

Logarithmic gains and sensitivities used in this analysis were computed using the implicit differentiation method described in Chapter 7 of [9].

References
1.   Bhalla US: **Understanding complex signaling networks through models and metaphors.** *Prog Biophys Mol Biol* 2003, **81:**45-65.
2.   Savageau MA: *Biochemical systems analysis : a study of function and design in molecular biology* Reading, Mass.: Addison-Wesley Pub. Co. Advanced Book Program; 1976.
3.   Savageau MA: **Design principles for elementary gene circuits: Elements, methods, and examples.** *Chaos* 2001, **11:**142-159.
4.   Savageau MA: **Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions.** *J Theor Biol* 1969, **25:**365-369.
5.   Savageau MA: **Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation.** *J Theor Biol* 1969, **25:**370-379.
6.   Savageau MA: **Biochemical systems analysis. 3. Dynamic solutions using a power-law approximation.** *J Theor Biol* 1970, **26:**215-226.
7.   Torres NV, Voit EO: *Pathway analysis and optimization in metabolic engineering* New York: Cambridge University Press; 2002.
8.   Voit EO: *Canonical nonlinear modeling : S-system approach to understanding complexity* New York: Van Nostrand Reinhold; 1991.
9.   Voit EO: *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists* New York: Cambridge University Press; 2000.
10. Savageau MA: **A theory of alternative designs for biochemical control systems.** *Biomed Biochim Acta* 1985, **44:**875-880.
11. Alves R, Savageau MA: **Extending the method of mathematically controlled comparison to include numerical comparisons.** *Bioinformatics* 2000, **16:**786-798.
12. Alves R, Savageau MA: **Comparing systemic properties of ensembles of biological networks by graphical and statistical methods.** *Bioinformatics* 2000, **16:**527-533.

13. Alves R, Savageau MA: **Irreversibility in unbranched pathways: preferred positions based on regulatory considerations.** *Biophys J* 2001, **80:**1174-1185.
14. Wall ME, Hlavacek WS, Savageau MA: **Design principles for regulator gene expression in a repressible gene circuit.** *J Mol Biol* 2003, **332:**861-876.
15. Hlavacek WS, Savageau MA: **Completely uncoupled and perfectly coupled gene expression in repressible systems.** *J Mol Biol* 1997, **266:**538-558.
16. Savageau MA: **Optimal design of feedback control by inhibition.** *J Mol Evol* 1974, **4:**139-156.
17. Hlavacek WS, Savageau MA: **Rules for coupled expression of regulator and effector genes in inducible circuits.** *J Mol Biol* 1996, **255:**121-139.
18. Irvine DH, Savageau MA: **Network regulation of the immune response: modulation of suppressor lymphocytes by alternative signals including contrasuppression.** *J Immunol* 1985, **134:**2117-2130.
19. Alves R, Savageau MA: **Effect of overall feedback inhibition in unbranched biosynthetic pathways.** *Biophys J* 2000, **79:**2290-2304.
20. Savageau MA: **Alternative designs for a genetic switch: analysis of switching times using the piecewise power-law representation.** *Math Biosci* 2002, **180:**237-253.
21. Irvine DH, Savageau MA: **Network regulation of the immune response: alternative control points for suppressor modulation of effector lymphocytes.** *J Immunol* 1985, **134:**2100-2116.
22. Agresti A: *Categorical data analysis* New York: Wiley; 1990.
23. David HA: *The method of paired comparisons* New York,: Hafner Pub. Co.; 1963.
24. Schwacke JH, Voit EO: **BSTLab: A Matlab Toolbox for Biochemical Systems Theory.** In: *Eleventh International Conference on Intelligent Systems for Molecular Biology* Brisbane, Australia; 2003.
25. Ihaka R, Gentlemen R: **R: A Language for data analysis and graphics.** *Journal of Computational and Graphical Statistics* 1996, **5:**299-314.
26. Savageau MA: **The behavior of intact biochemical control systems.** *Current Topics in Cellular Regulation* 1972, **6:**63-130.