MIT Open Access Articles

Sloppy Models, Parameter Uncertainty, and the Role of Experimental Design

Sloppy Models, Parameter Uncertainty, and the Role of Experimental Design

Joshua F. Apgar[1,2,†], **David K. Witmer**[2,3,†], **Forest M. White**[1,4], and **Bruce Tidor**[1,2,3,*]

[1]Department of Biological Engineering

[2]Computer Science and Artificial Intelligence Laboratory

[3]Department of Electrical Engineering and Computer Science

[4]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Summary

Computational models are increasingly used to understand and predict complex biological phenomena. These models contain many unknown parameters, at least some of which are difficult to measure directly, and instead are estimated by fitting to time-course data. Previous work has suggested that even with precise data sets, many parameters are unknowable by trajectory measurements. We examined this question in the context of a pathway model of epidermal growth factor (EGF) and neuronal growth factor (NGF) signaling. Computationally, we examined a palette of experimental perturbations that included different doses of EGF and NGF as well as single and multiple gene knockdowns and overexpressions. While no single experiment could accurately estimate all of the parameters, experimental design methodology identified a set of five complementary experiments that could. These results suggest optimism for the prospects for calibrating even large models, that the success of parameter estimation is intimately linked to the experimental perturbations used, and that experimental design methodology is important for parameter fitting of biological models and likely for the accuracy that can be expected for them.

Keywords

epidermal growth factor; model calibration; neuronal growth factor; parameter estimation; sloppy models

Introduction

One of the goals of systems biology is the construction of computational models that can accurately predict the response of a biological system to novel stimuli.[1-3] Such models serve to encapsulate our current understanding of biological systems, can indicate gaps in that understanding, and have the potential to provide a basis for the rational design of experiments,[4,5] clinical interventions,[6,7] and synthetic biological systems.[8] There are many varieties of computational models ranging from abstracted data-driven models to highly detailed molecular-mechanics ones. In this report we focus on the popular class of ordinary differential equation (ODE) models[9-16] typically used to describe systems at the biochemical and pharmacokinetic level but which are also appropriate at more abstract levels.

*Corresponding Author: Bruce Tidor, MIT Room 32-212, Cambridge MA 02139-4307 USA, Phone: (617) 253-7258, tidor@mit.edu.
†These authors contributed equally to this work

Constructing an ODE model is comprised of writing kinetic rate equations that describe the time rate of change of the various chemical species (representing the model topology), and determining the unknown parameters in those equations (typically rate constants and initial concentrations). Unknown parameters are estimated from a variety of data that often includes time-course measurements of concentration or activity. In this study, we have focused on the estimation of parameters, which is often referred to as model calibration. Using computational modeling and experimental design methodology, we have found that the selection of a set of experiments whose members provide complementary information can lead to efficient model calibration.

It should be noted that the problem of model calibration is different from model construction, where increasing numbers of parameters can be used to improve the fit to any given set of measurements, although parameter uncertainty may remain large. There is a considerable body of work focused on the problem of model complexity as it relates to parameter uncertainty.[17-20] In general these methods attempt to balance the ability of a more complex model to reduce fitting errors against the increased likelihood that a more complicated model will be able to fit the data by chance. Here we fix the model structure and number of parameters and vary only the measurements taken to develop a strategy for fitting the constant number of parameters with as little uncertainty as possible.

A detailed treatment of the theory for the current study is present in the Theory section. Here we provide a framework treatment of that theory. The quality of fit between measurements and a model can be expressed as the weighted sum-of-squares of the disagreement between them, which is a chi-squared (χ^2) metric. Finding parameter values for a fixed model topology that minimizes χ^2 gives the best-fit parameter values, but because of measurement uncertainty, different sets of parameter values may be consistent with any given set of measurements. A common approximation of this parameter uncertainty is to expand χ^2 as a function of the parameters and truncate after the second-order term. The linear term vanishes because the first derivative of χ^2 with respect to parameters is zero when the expansion is carried out about the best-fit parameter values. χ^2 is thus approximated as a constant plus the second-order terms involving the second derivative of the χ^2 quality of fit with respect to the parameters, known as the Hessian. A given amount of measurement uncertainty leads to an ellipsoid shaped envelope of constant χ^2 in an appropriately scaled parameter space. Sets of parameters within the envelope are consistent with the measurements and their associated uncertainty.

Longer axes of the ellipsoid correspond to parameter combinations of greater uncertainty(i.e., that are less well determined by the measurements), whereas shorter ellipsoidal axes correspond to parameter combinations of less uncertainty. The mathematics is such that the aces directions of the ellipsoid are given by the eigenvectors (v_i's) of the Hessian, and their associated uncertainty is given by the reciprocal of the square root of the corresponding eigenvalues ($\lambda_i^{-1/2}$). Thus, a set of parameters is well determined by a collection of measurements when the eigenvalues of the corresponding Hessian are all sufficiently large that they correspond to small relative parameter uncertainty.

Recently Gutenkunst *et al.*[21] examined parameter uncertainty for 17 models in the EMBL BioModels Database.[22] In their study, the authors assumed noise-free measurements of every model species sampled continuously in time. The study found that the eigenvalues of the Hessian spanned a large range ($> 10^6$). From this they suggested that, while it may be possible to estimate some parameters from system-wide data, in practice it would be difficult or impossible to estimate most of the parameters even from an unrealistically high-quality data set.[21,23,24] Moreover, they pointed out that due to the high eccentricity and skewness of uncertainly ellipses in parameter space, system-wide data can define system

behavior better than independent measurements of each parameter and may also produce better predictions in some circumstances.

Here we extend the previous work by more fully considering the effect of experimental perturbations on the parameter estimation problem and use experimental design to probe for particularly effective perturbation experiments. The χ^2 goodness of fit metric depends on both the model and the set of experimental conditions. Some experiments may be more helpful in calibrating the model than others. In the current work we use effectively continuous-time data, but many experiments require the selection of discrete time points for measurements to be taken.[25,26] It is well established in the systems biology literature that optimal experimental design can have an impact on the parameter estimation problem for a single experiment.[23,25,27-29] For example, work by Faller *et al.* has shown for a small model of a mitogen activated protein kinase (MAPK) cascade that the application of time-varying stimulation significantly improved the parameter estimation problem.[29] Essentially this corresponds to finding the time-varying input signal that gives the best shaped error ellipsoid.

In this work, we apply a related approach and examine the extent to which multiple complementary experiments can be combined to improve the overall parameter estimation problem. Figure 1B,C shows the result of combining data from two separate experiments. The parameter estimates from the individual data sets (blue and red ellipses) tightly constrain one parameter direction and weakly constrain the other. In Figure 1B, the weakly constrained parameter directions are very similar, so the parameter estimates from the combined data set are about the same as the estimates from the individual experiments (green ellipse); by contrast, in Figure 1C the experiments are complementary and together dramatically constrain the parameter estimates.

Because complementary experiments can constrain parameter estimation space, we have developed an approach to identify sets of complementary experiments to optimally minimize parameter uncertainty and tested it in a pathway model of signaling in response to EGF and NGF.[30] We have selected this model so that our results may be directly compared to the previously published analysis of this model performed by Gutenkunst *et al.*[21] For consistency, where possible we have used their methods and formalisms. In selecting sets of complementary experiments, we have explored a palette of candidate experiments consisting of overexpression or knockdown of single and multiple genes combined with different doses of EGF and NGF, either alone or in combination.

Computational experimental design methods determined all 48 free parameters to within 10% of their value using just five complementary experiments. Selection of *complementary* experiments was essential, as the same level of model calibration could not be achieved with arbitrary experiments or even with a larger number of "highly informative" experiments. Moreover, we argue that predictions that are sensitive to information complementary to that used to parameterize a model could be significantly in error. Experimental design methods can provide sufficient coverage for all parameter directions and thus guide model calibration for a given topology to maximize predictive accuracy. As systems biology models are applied to target identification and clinical trial design, the use of experimental design approaches to improve model prediction quality could be of crucial importance.

Previous work on the model calibration problem has focused on optimization within the scope of a single experiment.[31,32] Examples include selecting optimal time points,[33,34] species,[35-38] or stimulus conditions[5,39,40] that would be most effective in reducing parameter uncertainty. However, even highly optimized single experiments are generally insufficient for model calibration. For this reason, such methods have largely been applied to

smaller scale problems. The current work is different in spirit in that it addresses the question of how improved model calibration might result from combinations of experiments that could collectively define all of the parameters. By design, the individual experiments may be easier to implement, yet relatively small combinations of simple experiments can determine all parameters in a medium-sized pathway model.

Theory

In this work we formulate the model calibration problem as a nonlinear least squares optimization problem, where the goal is to find the set of parameters that minimizes the fit metric,[21]

$$\chi^2(p) = \frac{1}{2n_s n_c} \sum_s \sum_c \frac{1}{T_c} \int_0^{T_c} \frac{[y_{s,c}(p, t) - y_{s,c}(p^*, t)]^2}{\sigma_{s,c}^2(p^*, t)} dt \quad (1)$$

where n_c is the number of experimental conditions, n_s is the number of species for which measurements are available, the indices c and s run over the conditions and species, respectively, T_c is the length of the time course for condition c, $y_{s,c}(\mathbf{p}, t)$ is the model output for species s and condition c at time t with parameter set $\mathbf{p}$, $y_{s,c}(\mathbf{p^*}, t)$ is the corresponding output for the true model parameterization, and $\sigma_{s,c}^2(p^*, t)$ is a weighting factor that is often taken as proportional to the uncertainty of the experimental measurement.

There is a significant amount of work devoted to how best to solve this optimization problem for biological models.[41-44] However, in any experimental system, there will always be uncertainty in the data, which means there will be some range of parameter values that, while not optimal, cannot be excluded based on the data. Given a maximum acceptable fitting error, χ_{max}^2 the calibration problem becomes that of finding all parameter sets such that the error is less than this threshold. In a neighborhood around the optimum parameterization $\mathbf{p}$ *, the least squares cost function can be approximated by its Taylor series expansion

$$\left[\chi_{max}^2 - \chi^2(p^*)\right] \geq \frac{1}{2}[\log(p) - \log(p^*)]^T H[\log(p) - \log(p^*)] \quad (2)$$

Equation 2 describes an n_p-dimensional ellipsoid in parameter space (n_p being the number of fitted parameters), where all of the parameterizations inside the ellipsoid are feasible. The size and shape of this ellipsoid describe the multidimensional parameter uncertainty. For example, the longest axis of the ellipsoid corresponds to the parameter direction (that is, the linear combination of parameters) with the worst error. Likewise, the axis-parallel bounding box defines the error range for individual parameters. Figure 1 shows an example for a two-dimensional system. An important distinction illustrated in Figure 1A is that some parameter directions can have very small uncertainty, while the individual parameters can be quite uncertain.[21] For example, forward and reverse kinetic constants for a binding reaction may be poorly constrained, yet the equilibrium constant (given by their ratio), can be well defined.

H, the matrix of second derivatives of the fit metric is known as the Hessian, where $H_{i,j}$ is the derivative of $\chi^2(\mathbf{p})$ with respect to $\log(p_i)$ and $\log(p_j)$.

$$H_{i,j}(p) = \frac{1}{n_s n_c} \sum_s \sum_c \frac{1}{T_c} \int_0^{T_c} \frac{1}{\sigma_{s,c}^2(p^*, t)} \left(\frac{dy_{s,c}(p, t)}{d\log(p_i)} \times \frac{dy_{s,c}(p, t)}{d\ \log(p_j)} + [y_{s,c}(p, t) - y_{s,c}(p^*, t)] \frac{d^2 y_{s,c}(p, t)}{d\ \log(p_i) d\ \log(p_j)} \right) dt \quad (3)$$

The derivative is taken with respect to the natural logarithm of p_i to obtain a relative uncertainty. We can dissect the parameter uncertainty in terms of the eigenvalues λ_i and eigenvectors ν_i of **H**. The eigenvectors form a natural coordinate system for the ellipsoid, pointing along the axes. The lengths of the axes are proportional to $\lambda_i^{-1/2}$, meaning that smaller eigenvalues correspond to larger relative parameter error.[41]

Results

Experimental design and computer simulations were applied in tandem to select a collection of experiments that together could most directly establish each of the rate constant parameters for the EGF/NGF signaling pathway modeled here. To define all 48 rate parameters, experimental design procedures must select a set of experiments that together exercise the model in complementary and sufficiently different ways, rather than simply choosing multiple different experiments that exercise the model in similar ways. For this work we chose a palette of experimental perturbations consisting of stimulation with EGF (10^7, 10^5, 10^3, 10, or 0 molecules/cell-volume) or NGF (4.52×10^7, 4.52×10^5, 4.52×10^3, 45.2, or 0 molecules/cell-volume) individually, or combined treatment with both ligands. We supplemented this choice of ligand stimulation with a panel of experiments in which protein expression levels could be modulated by 100-fold overexpression or knockdown for individual proteins in the network. We then constructed candidate experiments from the combination of ligand choice and protein expression level changes; specifically, each experiment was allowed to comprise one stimulation pattern and up to three simultaneous changes in protein expression level. This experimental set-up resulted in a trial perturbation set of 164,500 individual computational experiments.

All parameters can be determined to high accuracy

The experiments were evaluated and the number of rate parameter directions determined to within 10% of their nominal value for each experiment was recorded. The best individual experiment in this set defined only 29 of the 48 rate parameter directions to this high level of accuracy. In order to improve on this result, each single experiment was re-evaluated to determine how many new rate parameter directions could be defined to within 10% of their nominal value when combined with the best individual experiment. In this manner, a greedy algorithm was applied to select sequentially sets of experiments based on the ability to generate tighter bounds on parameter estimates. The results of this greedy algorithm are shown in Figure 2A. The parameter uncertainties are expressed as an eigenspectrum for each set of experiments, with increasingly larger experimental sets displayed along the abscissa and eigenvalues displayed along the ordinate. The horizontal dashed line indicates the 10% error level, and the number of eigenvalues above the dashed line represents the number of parameter directions determined to the 10% error level. In Figure 2D, the number of parameters estimated to the 10% level is shown as a function of the number of experiments within the experimental set. It is striking to observe that, by properly choosing the correct combination of experiments, only five total experiments are sufficient to determine all 48 directions (and, indeed, all 48 actual parameters) to within 10% accuracy. This result indicates that parameter uncertainty, rather than being inherent to biological models, can be progressively reduced by perturbation experiments.

The five experiments determined here to elucidate all 48 parameters are listed in Table I. The selected experiments include a tendency for dual stimulation with both EGF and NGF, combined with multiple protein-expression changes, and a preference for overexpression as opposed to knockdown of given proteins. Interestingly, the experiments do not appear to systematically explore all regions of the perturbation space. For instance, four of the five experiments have a low dose of EGF stimulation, and three of the five experiments have a

high dose of NGF stimulation. Additionally, NGF stimulation occurs in the absence of EGF stimulation in one experiment, but EGF stimulation alone is never utilized. The combination of experiments and the manner in which they explore different aspects of the model in order to adequately define all parameters is not readily apparent from the chosen experiments. However, as discussed below, there are some general trends and insights to be gained from an analysis of the results.

Supplementary Figure S1 shows the location of parameters in the model determined after each new experiment was added in sequence to the set. The thickness of each reaction arrow in the Figure indicates whether 0 (thin arrow), 1 (medium arrow), or 2 (thick arrow) of the parameters associated with the arrow are known to within 10% at that point in the experimental sequence. Note that parameters closer to the top of the pathway tend to be determined first, while parameters toward the bottom of the pathway, further from the application of ligand stimulation, tend to be determined only after multiple experiments are combined.

The selection of complementary experiments resulting from our experimental design procedure is non-trivial. For example, if experiments were added sequentially based on their ability to determine a large number of parameters when applied on their own (Figure 2B) or if experiments were added randomly (Figure 2C), the performance was much worse; neither procedure could determine the full 48 parameters with up to 20 experiments, whereas the greedy experimental design procedure required only five (Figure 2D). Interestingly, the "random" procedure was about as effective as the "best singles" one beyond the initial few experiments, which suggests that selecting "good" or "bad" experiments isn't nearly as important as choosing complementary ones For example, most of the "best singles" experiments tended to involve stimulation with EGF and NGF simultaneously (Supplementary Figure S2), but the type of complementarity required to tease apart all the parameter directions can be achieved by more subtle variation (Table I). Together these results show that selection of complementary experiments can very efficiently lead to full parameterization of complex models and that experimental design procedures are important to choose an appropriately complementary set.

Relative resolving power of different experiment types

There should be a cost–benefit relationship between the complexity of experimental perturbations used and the amount of information obtained. Intuitively, we expect it may take more simple experiments to obtain the same knowledge gleaned from a smaller number of more complex experiments. The above results included experiments that simultaneously modified three protein concentrations. To probe whether simpler experiments could achieve similar results, we repeated the search but limited the experiments to EGF/NGF doses only (Figure 3A,B), EGF/NGF doses and single knockdowns and overexpressions (Figure 3C,D), or EGF/NGF doses and up to double knockdowns and overexpressions (Figure 3E,F). For comparison, results for the full set of experiments with up to triple expression changes are shown in Figure 3G,H. An experimental set with up to forty experiments was constructed for each using our greedy experimental design approach, and the maximum number of parameters determined to within 10% is shown. For each pair of figure panels, the left panel shows the location in the network of parameters established by the experimental set, with the same arrow thickness scheme as in Supplementary Figure S1 of thicker arrows indicating more parameters defined. The eigenvector matrix, a representation of best determined to least well-determined directions in rate-constant space, is displayed in the right panel for each sub-section of Figure 3. The eigenvector matrix shows which individual parameters contribute to each parameter direction. All rate constants could be determined with up to triple expression changes (as shown above, with five experiments) and 47 of 48 parameters with up to double expression changes (requiring eight experiments, with no improvement

from the next 32 most complementary experiments). Most rate constants could be determined using single changes in protein expression (45 of 48, but requiring 17 experiments), but doses of EGF/NGF alone were only able to establish just over half of the rate constants (25 of 48, requiring just 2 experiments with no improvement resulting from the remaining 23 experiments in the class). Thus, more complex perturbation experiments improved model calibration through establishing a greater number of parameters and generally doing so with fewer, albeit more difficult, experiments. The tradeoff is such that in many cases the greater complexity may be justified by the reduced number of total experiments required. Modifications to future versions of the optimization could be biased towards reuse of genetic modifications in experiments with different dosage treatments, so that the greater effort of the former might be better leveraged.

Biochemical basis for complementarity of experiments

Analysis of the five experiments sufficient for determining all the parameters suggests that one role for some selected experiments is to specifically adjust conditions of enzymatic reactions so that k_{cat} and K_M could be independently determined. Because the calculations were done in log parameter space, the $(+k_{cat}, +K_M)$ subspace direction corresponds to $\log(k_{cat})+\log(K_M) = \log(k_{cat} \times K_M)$. Interestingly, the $(+k_{cat}, -K_M)$ parameter direction, corresponding to $\log(k_{cat}/K_M)$, is easily estimated for many reactions here because in the wild type many enzymes operate under k_{cat}/K_M conditions with substrate concentration well below K_M. Some selected experimental perturbations appear to have been chosen because they drive substrate concentration sufficiently high as to move outside of k_{cat}/K_M conditions to determine k_{cat} and K_M independently (Figure 4).

More stringent parameterization can be achieved with greater experimental effort

It is important to consider whether there are fundamental limits to how accurately biological models can be parameterized—essentially whether there exist model parameters that are essentially unknowable. As one step to addressing the question of knowability, we repeated the optimal experimental design calculations using up to three expression changes, but with different values of the error threshold. Our previous results were computed with the requirement that parameters be established to within 10% of their nominal values. The full set of results is shown in Figure 5. While five experiments were required to establish all 48 parameters to within 10% error, only four experiments were required for the less stringent 37% error. Likewise, as the stringency was increased, greater experimental effort was required to establish a given number of parameters. Together these results suggest that more stringent parameterization can generally be achieved with greater experimental effort, although experimental design procedures may be necessary to define how best to apply additional effort towards determining new knowledge. Additional experimental complexity may be necessary to dissect particularly difficult parameter combinations. Whether such effort is worthwhile may depend on the sensitivity of predictions made by the model to such difficult parameter combinations.

Discussion

In this study computer simulations and experimental design methods were used to probe the relationship between experimental perturbations applied to a complex biological system and the relative certainty with which model parameters describing that system can be established. These *in silico* experiments resolve an important question in computational systems biology. They imply that uncertainty is not inherent to biological network models— rather, uncertainty can be progressively diminished through sequential addition of perturbation experiments to a cumulative data set used for model calibration. In the case studied here, a cellular network activated in response to stimulation by EGF and NGF

important for cell growth and differentiation, all 48 rate constant parameters could in principle be fit to within 10% of their nominal value using concentration time courses from five multi-perturbation experiments. While the accuracy of the measurements and the desired accuracy of the parameters affect the number of required experiments and their complexity, all parameters could be estimated to very high accuracy.

An important characteristic of minimal sets of experiments that together define all model parameters is that the members are mutually complementary. That is, while they may contain some overlapping information, each experiment should also contain information that is not provided by the others. The overall parameter uncertainty of a set of experiments is related to the intersection of the parameter uncertainty associated with each individual experiment. Thus, the intersection diagrams of Figures 1 and 4 indicate that the directions of large uncertainty in one experiment should correspond to a direction of smaller uncertainty in at least one other experiment. This form of complementarity is non-obvious and non-trivial, but experimental design methodology can efficiently identify sets possessing this property. In the current example, running enzymes under both k_{cat}/K_M and k_{cat} regimes was one important form of complementarity. Single experiments were generally incapable of spending sufficient time in each regime, and different experiments were required for each. The design of single experiments that visit both regimes could be beneficial.

The current case also exhibited a tendency to select protein overexpression experimentsas opposed to knockdowns, although the generality of this result remains to be seen. Operationally, in many cases it may be possible and convenient to alter the activity of proteins with selective inhibitors while leaving the expression level constant. Small molecule inhibitors or activators also enable time-dependent perturbation, thereby providing important new degrees of freedom that may permit improved model calibration with fewer experimental manipulations.[5] However, we found certain parameter directions for which overexpressions were important for their exploration. In the work described here, we used full trajectories of all concentrations in fitting the model; however, current experimental technology generally probes only a subset of species and time points. The experimental design framework described here can be used to determine the most productive species and time point measurements to make in order to most expeditiously calibrate model parameters.

While a variety of experimental interventions can be used to change protein expression, another possibility is to make use of natural variation. An examination of the expression levels of the proteins that correspond to the 19 proteins in the EGF/NGF signaling model as reported in the GNF SymAtlas database[45] is shown in Supplementary Figure S3 and indicates expression ranges of 100 fold or more across tissues and cell types. Natural variation on this scale suggests that collecting data in multiple cell types may be an alternative to using genetic manipulation, but it also suggests that accurate estimation of parameters and quantification of protein expression levels may be necessary for a model to be applicable across different cell and tissue types. However, using multiple cell lines introduces additional complications, as differences in biology and un-modeled effects may dominate the results. In this case, having a well-calibrated model may make it easier to distinguish between calibration issues and real biological differences.

For the current work protein knockdowns and overexpressions were treated with no changes in the expression of other genes, as gene expression is not currently part of the model.[30] Expression regulation, when modeled, can readily be account for with the methods described here.

Many different sets of experiments can be mutually complementary and each can define all parameters equally well. Furthermore, tradeoffs exist between the complexity of

experiments available and the number of experiments required to probe fully all parameters. Multiple genetic changes were particularly effective in the EGF/NGF network studied here, with a larger number of simpler experiments performing less well overall. Here we designed sets of experiments to define the parameters of a pre-existing model, as this would allow us to most directly address the theoretical feasibility of the approach. One result is that rather complex experiments were called for, and we were curious whether this was a necessity of the underlying biological system or an artifact of the model itself. We noticed that a few K_M values in the model were quite large, and we built an alternative model with more standard K_M values that fit the experimental as well as the original model (Supplementary Figure S4). We applied our experimental design methodology but only permitted an experiment to consist of an EGF/NGF and 100-fold expression change of *one* protein. All 48 parameters could be determined using this simpler experimental set in just 15 experiments (Supplementary Figure S3 and Supplementary Table SI). Interestingly, in this case protein overexpression experiments dominated underexpression again. Moreover, three genetic perturbations were re-used, and in two pairs of experiments the same gene was used in both overexpression and underexpression experiments. Thus, it is likely that numerous, simple experiments can be used to quantitatively understand and describe even complex biological systems.

In our analysis we have considered the value of the Hessian only at the optimal parameter set when computing parameter uncertainty. This corresponds to assuming that the log parameter errors are Gaussian, an assumption that may not be true for poorly determined parameter directions. However, as more complementary experiments are added, and as the parameter uncertainty is reduced, this approximation will become increasingly accurate.[46]

Throughout out this work, we have stressed the importance of fully determining the parameters of a model. However, as has been observed by others,[21] this may not be necessary to make a particular prediction. In fact, the corollary to no single experiment determining all of the parameters is that no single prediction depends on all of the parameters. This result suggests a potential variation to our method. If a particular model is intended only to be used in a narrowly defined context where only certain parameters are sensitive, then the method could choose experiments to define the parameter directions that spanned the sensitive parameter directions. For example, if a substrate was not to be expressed at a level above the K_M of any of the enzymes that modify it, one could decide to ignore the parameter errors that pointed in the $(+k_{cat}, +K_M)$ direction. However, the results presented here suggest that the saving in terms of experimental effort may be minimal, as a small number of experiments was able to cover the entire parameter space, and it comes at the cost of decreasing the predictive power of the model.

If a complementary set of experiments is used to fully parameterize a model, then there exists no perturbation to the system for which the parameterization is inadequate. The identification of therapeutic approaches, drug targets, and treatment regimens essentially corresponds to identifying perturbations that produce desirable outcomes. To be meaningfully accurate, such predictions must still be adequately parameterized by the calibration experiments. The construction of a complementary set of experiments sequentially reduces all directions of parameter uncertainty. Intuitively, once all the parameters are well determined, the model will make extremely good predictions whose error can be bounded through propagation of the remaining uncertainty. (These statements are predicated on the model topology being sufficiently correct to accurately describe system dynamics and on the absence of bifurcations and other anomalies within the remaining parameter uncertainty.) On the other hand, if a model is calibrated with an incomplete set of experiments, then there exist parameter directions with large uncertainty. Any experimental prediction whose outcome is sensitive to the undetermined parameter

directions, such as the effects of drug therapy, for instance, could be grossly incorrect. In fact, such an experiment might be a poor therapy but an excellent calibration tool. Our results demonstrate that it is not necessary to modulate the expression level of every protein in the network in order to determine all parameters and thus fully define the behavior of the system. Rather, only a complete set of complementary experiments need be used for model calibration;in principle, excellent predictions should follow.

Methods

The model

In this work we examined a model of EGF and NGF signaling.[30,47-51] The model, which can be found in the BioModels database (BIOMD0000000033),[22] consists of 19 distinct proteins, two extracellular proteins (EGF and NGF, which act as inputs), two cell surface receptors (EGFR and NGFR), and 15 intracellular proteins. The two receptors and 11 of the cytoplasmic proteins can be in either an active or inactive state. The remaining four intracellular proteins are constitutively active, resulting in a total of 32 distinct chemical species, 26 chemical reactions, and 48 parameters (22 Michaelis constants, K_M; 22 catalytic rate constants, k_{cat}; 2 second-order association constants, k_{on}, and 2 first-order dissociation constants, k_{off}; note that all initial concentrations are assumed known). Twenty-two of the reactions are implemented with Michaelis–Menten kinetics, while the remaining four reactions are the mass-action binding reactions of EGF and NGF to their respective receptors. One modification was made to the model, which was to put free EGF and NGF into an extracellular compartment having a volume 1000 times the volume of the intracellular compartment. This was chosen to better reflect a typical experiment, where the extracellular volume greatly exceeds the intracellular volume. Preliminary calculations using a variety of different volumes for the extracellular compartment demonstrated that the results did not depend strongly on the value for this volume (data not shown).

Objective function

The goal of experimental design for model calibration can be expressed as selecting the experimental conditions such that the resulting parameter uncertainty ellipsoid will be prescriptively small. In this work, we focused on maximizing the number of parameters with uncertainty less than a given threshold. Equation 4 shows this metric F_N as functions of the **H**

$$F_N(\lambda_{\text{thresh}}) = \sum_{i=1}^{n_p} \text{thresh}(\lambda_i, \lambda_{\text{thresh}}) \quad \text{where} \quad \text{thresh}(\lambda_i, \lambda_{\text{thresh}}) = \left\{ \begin{array}{ll} 1 & \lambda_i > \lambda_{\text{thresh}} \\ \frac{\lambda_i}{\lambda_{\text{thresh}} n_p} & \text{otherwise} \end{array} \right\}, \quad (4)$$

where n_P is the total number of parameter directions and λ_{thresh} is determined below. We chose to include some contribution to our objective function from parameters that did not meet the threshold to break ties between experiments with the same number of good parameter directions (thus the term $\frac{\lambda_i}{\lambda_{thresh} n_p}$, rather than zero). Essentially $F_N(\lambda_{\text{thresh}})$ counts up the number of eigenvalues λ_i greater than a threshold. By choosing this threshold to correspond to a 10% relative parameter uncertainty, the function counts up the number of parameter directions with uncertainty less than 10%.

In this work we assume that our data has Gaussian relative error with zero mean and 10% standard deviation. We weighted concentration differences in Equation 1 by the variance in the measured value, so $\sigma_{s,c}(\mathbf{p}^*,t) = f \times y_{s.c}(\mathbf{p}^*,t)$, which makes our least squares estimator a maximum likelihood estimator. Here f is the relative measurement error, which we set to 0.10 (10%). Equation 1 describes the information from an average data point, where the

average is over species, conditions, and time. Following the method of Gutenkunst *et al.*,[21] we scaled the cost function by n_d. This corresponds to collecting n_d independent measurements of the system, where n_d is large enough so that the sample mean approaches the mean. In this work we assumed 100 times the number of species, to correspond to discretely sampling the system at 100 time points. Note that as the number of conditions increase we do not increase the number of data points as this provides useful normalization.

The variance of the parameter uncertainty can be computed by the Cramer–Rao bound.[52] The maximum likelihood estimator is asymptotically normal, unbiased, and efficient, so in the limit of a large number of data points the variance $\sigma^2_{v_i^T \log p}$ of the log parameter uncertainty in the ith log parameter eigendirection ψ_i can be approximated as

$$\sigma^2_{v_i^T \log p} = \frac{f^2}{n_d \lambda_i}. \quad (5)$$

We define the relative parameter error to be the ratio of the upper and lower values of the parameter corresponding to the 95% or two standard deviation confidence interval.[53] Computing this involves exponentiation of the bounds of log parameter uncertainty.

$$\Sigma_{v_i^T \log p} = \frac{\exp\left(v_i^T \log p^* + 2\sigma^2_{v_i^T \log p}\right)}{\exp\left(v_i^T \log p^* - 2\sigma^2_{v_i^T \log p}\right)} - 1 = \exp\left(4\sigma^2_{v_i^T \log p}\right) - 1. \quad (6)$$

As an example, $\Sigma_{v_i^T \log p} = 0.1$ means that the ratio of the maximum parameter value to the minimum parameter value is 1.1, which corresponds to ±10% relative error at the 95% confidence interval. We solved for a λ_{thresh} that corresponded to a desired parameter error by solving Equations 5 and 6; for $f = 0.1$ and $\Sigma_{v_i^T \log p} = 0.1$, $\lambda_{\text{thresh}} = 0.005$.. We used this value, except where noted otherwise.

Experimental design

Equation 1 shows that the fit metric is a sum over conditions. Likewise, the Hessian of the fit metric can also be constructed as a sum of the Hessians from each individual condition

$$H(C, p) = \frac{1}{n_c} \sum_{c \varepsilon C} H_c(p) \quad (7)$$

where C is a set of experimental conditions, c indexes individual experimental conditions in the set, and n_c is the number of experiments in C. Thus, we formulated the experimental design problem as the process of selecting the subset of all possible experiments such that the sum of the Hessians has the desired properties. Operationally, we simulated each individual experimental condition in the trial space and computed the Hessian. We then performed a greedy search to find the best set of n_c experiments. At each step of the algorithm, the Hessian from the previous step was added to the Hessian for each candidate experiment, and the objective function was evaluated. The trial experiment that led to the best new Hessian was added to the set. The search terminated if the goal was met or if the maximum number of experiments was reached. The choice of a greedy search does not necessarily produce the optimal subset of a given size, indicating that these results are an upper bound on the number of experiments required to achieve a particular goal.

Computing Hessians

We computed the individual Hessians of this fit metric for each species in each condition using the M~ATLAB~ SimBiology Toolbox v2.4 (The Mathworks, Natick, MA) following the method of Gutenkunst *et al.*[21] The ODE system was integrated using $_{ODE}15_s$[54] and sensitivities were computed by complex-step finite differencing.[55] The entries in the Hessian were computed by numerical integration with the trapezoidal rule. Because we evaluated the Hessian at $\mathbf{p} = \mathbf{p}^*$, then $y_{s,c}(\mathbf{p},t) - y_{s,c}(\mathbf{p}^*,t) = 0$ for all time points, species, and conditions; thus, the second term in the integrand of Equation 3 (containing the second derivatives of $y_{s,c}$) was also zero. For the purpose of this study we dropped this term. Note that even with imperfect data it may be possible to approximate the Hessian using only first derivatives if the fit to the data is sufficiently good that the second derivative term is negligible. If this is not the case then the full Hessian equation should be used, and the rest of the analysis remains unchanged. We normalized the Hessians as calculated above using $f = 0.1$, which represents 10% relative measurement error. The computations were performed in parallel on 128 processors using approximately 800 cpu hours in total.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

(1). Kitano H. Science. 2002; 295:1662–1664. [PubMed: 11872829]

(2). Kremling A, Saez-Rodriguez J. J. Biotechnol. 2007; 129:329–351. [PubMed: 17400319]

(3). Wiley HS, Shvartsman SY, Lauffenburger DA. Trends Cell Biol. 2003; 13:43–50. [PubMed: 12480339]

(4). Chen BH, Asprey SP. Ind. Eng. Chem. Res. 2003; 42:1379–1390.

(5). Apgar JF, Toettcher JE, Endy D, White FM, Tidor B. PLoS Comput. Biol. 2008; 4:e30. [PubMed: 18282085]

(6). Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Drug Discov. Today. 2006; 11:806–811. [PubMed: 16935748]

(7). Butcher EC, Berg EL, Kunkel EJ. Nat. Biotechnol. 2004; 22:1253–1259. [PubMed: 15470465]

(8). Endler L, Rodriguez N, Juty N, Chelliah V, Laibe C, Li C, Le Novere NJR. Soc. Interface. 2009; 6(Suppl 4):S405–S417.

(9). Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. Science. 1996; 271:1582–1586. [PubMed: 8599114]

(10). Huang CY, Ferrell JE Jr. Proc. Natl. Acad. Sci. USA. 1996; 93:10078–10083. [PubMed: 8816754]

(11). Fallon EM, Lauffenburger DA. Biotechnol. Progr. 2000; 16:905–916.

(12). Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. Nat. Biotechnol. 2002; 20:370–375. [PubMed: 11923843]

(13). Kaplan EH, Craft DL, Wein LM. Proc. Natl. Acad. Sci. USA. 2002; 99:10935–10940. [PubMed: 12118122]

(14). Haugh JM. Biotechnol. Progr. 2004; 20:1337–1344.

(15). Flaherty P, Radhakrishnan ML, Dinh T, Rebres RA, Roach TI, Jordan MI, Arkin AP. PLoS Comput. Biol. 2008; 4:e1000185. [PubMed: 18818727]

(16). Dahari H, Layden-Almer JE, Kallwitz E, Ribeiro RM, Cotler SJ, Layden TJ, Perelson AS. Gastroenterology. 2009; 136:1402–1409. [PubMed: 19208338]

(17). Akaike, H. Second International Symposium on Information Theory. Petrov, B.; Csaki, F., editors. Akademiai Kiado; Budapest: 1973. p. 267-281.

(18). Schwartz G. Ann. Stat. 1978; 6:461–464.

(19). Bozdogan H. J. Math. Psychol. 2000; 44:62–91. [PubMed: 10733858]

(20). Haunschild MD, Freisleben B, Takors R, Wiechert W. Bioinformatics. 2005; 21:1617–1625. [PubMed: 15604106]

(21). Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. PLoS Comput. Biol. 2007; 3:1871–1878. [PubMed: 17922568]

(22). Le Novere N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep JL, Hucka M. Nucl. Acids Res. 2006; 34:D689–D691. [PubMed: 16381960]

(23). Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, Myers CR, Brown KS, Cerione RA, Sethna JP. IET Sys. Biol. 2007; 1:190–202.

(24). Waterfall JJ, Casey FP, Gutenkunst RN, Brown KS, Myers CR, Brouwer PW, Elser V, Sethna JP. Phys. Rev. Lett. 2006; 97:150601. [PubMed: 17155311]

(25). Kutalik, Z. n.; Cho, K-H.; Wolkenhauer, O. Biosystems. 2004; 75:43–55. [PubMed: 15245803]

(26). Donckels BM, De Pauw DJ, Vanrolleghem PA, De Baets B. J. Comput. Chem. 2009; 30:2064–2077. [PubMed: 19165773]

(27). Asprey SP, Macchietto S. Comput. Chem. Eng. 2000; 24:1261–1267.

(28). Asprey SP, Macchietto S. J. Process Control. 2002; 12:545–556.

(29). Faller D, Klingmuller U, Timmer J. Simul.-Trans. Soc. Model. Simul. Int. 2003; 79:717–725.

(30). Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, Cerione RA. Phys. Biol. 2004; 1:184–195. [PubMed: 16204838]

(31). Jaqaman K, Danuser G. Nat. Rev. Mol. Cell Biol. 2006; 7:813–819. [PubMed: 17006434]

(32). Kontoravdi C, Asprey SP, Pistikopoulos EN, Mantalaris A. Biotechnology Progress. 2005; 21:1128–1135. [PubMed: 16080692]

(33). Kutalik Z, Cho KH, Wolkenhauer O. Biosystems. 2004; 75:43–55. [PubMed: 15245803]

(34). Donckels BMR, De Pauw DJW, Vanrolleghem PA, De Baets B. Journal Of Computational Chemistry. 2009; 30:2064–2077. [PubMed: 19165773]

(35). Yue H, Brown M, Knowles J, Wang H, Broomhead DS, Kell DB. Molecular Biosystems. 2006; 2:640–649. [PubMed: 17216045]

(36). Cho KH, Shin SY, Kolch W, Wolkenhauer O. Simulation-Transactions of the Society for Modeling and Simulation International. 2003; 79:726–739.

(37). Gadkar KG, Gunawan R, Doyle FJ. BMC Bioinformatics. 2005; 6

(38). King BM, Tidor B. Bioinformatics. 2009; 25:1165–1172. [PubMed: 19261718]

(39). Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgower F, Gilles ED. Genome Res. 2004; 14:1773–1785. [PubMed: 15342560]

(40). Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, Myers CR, Brown KS, Cerione RA, Sethna JP. IET Systems Biology. 2007; 1:190–202. [PubMed: 17591178]

(41). Bard, Y. Nonlinear Parameter Estimation. Academic Press; New York: 1974.

(42). Jaqaman K, Danuser G. Nat. Rev. Molec. Cell Biol. 2006; 7:813–819. [PubMed: 17006434]

(43). Moles CG, Mendes P, Banga JR. Genome Res. 2003; 13:2467–2474. [PubMed: 14559783]

(44). Rodriguez-Fernandez M, Mendes P, Banga JR. Biosystems. 2006; 83:248–265. [PubMed: 16236429]

(45). Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. Proc. Natl. Acad. Sci. USA. 2004; 101:6062–6067. [PubMed: 15075390]

(46). Marsili-Libelli S, Guerrizio S, Checchi N. Ecol. Model. 2003; 165:127–146.

(47). Birchmeier C. Exp. Cell Res. 2009; 315:611–618. [PubMed: 19046966]

(48). Britsch S. Adv. Anat. Embryol. Cell Biol. 2007; 190:1–65. [PubMed: 17432114]

(49). Bublil EM, Yarden Y. Curr. Opin. Cell Biol. 2007; 19:124–134. [PubMed: 17314037]

(50). Yarden Y. Eur. J. Cancer. 2001; 37(Suppl 4):S3–S8. [PubMed: 11597398]

(51). Yarden Y, Sliwkowski MX. Nat. Rev. Mol. Cell Biol. 2001; 2:127–137. [PubMed: 11252954]

(52). Seber, GAF.; Wild, CJ. Nonlinear Regression. Wiley; New York: 2003.

(53). Bertsekas, DP.; Tsitsiklis, JN. Introduction to Probability. 2nd ed.. Athena Scientific; Belmont, Mass: 2008.

(54). Shampine LF, Reichelt MW. SIAM J. Sci. Comput. 1997; 18:1–22.

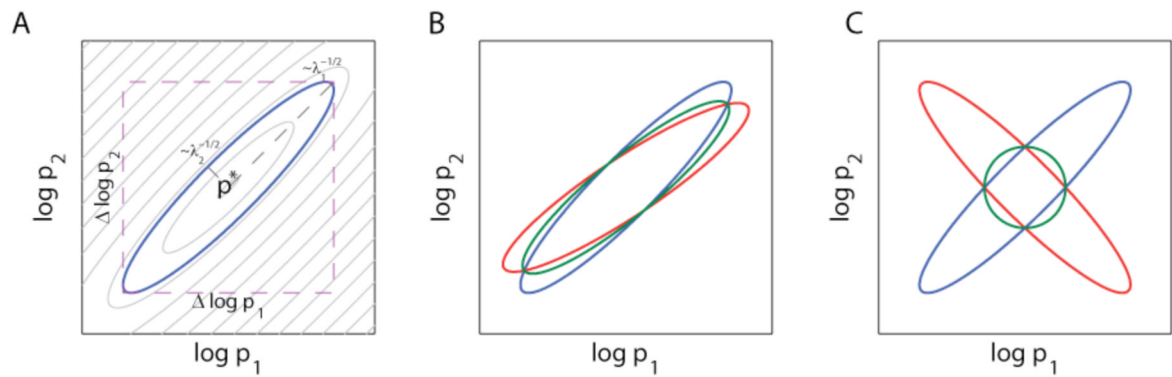(55). Martins JRRA, Sturdza P, Alonso JJ. ACM Trans. Math. Softw. 2003; 29:245–262.

Figure 1.
(A) Uncertainty ellipse for a simple two-parameter system. The parameters inside the ellipse are feasible. The major and minor axes of the ellipse are proportional to $\lambda_1^{-1/2}$ and $\lambda_2^{-1/2}$, respectively. The grey contours are lines of equal objective function value. The bounding box (dashed magenta lines) shows the single parameter errors. The length and width of the bounding box is the range of values for the individual parameters. (B) Two non-complementary experiments. (C) Two complementary experiments. When two experiments (blue and red ellipses) are combined the resulting parameter estimate (green ellipses) can be improved as in the case of panel C or not as in the case of panel B depending on the complementarity of the component experiments.
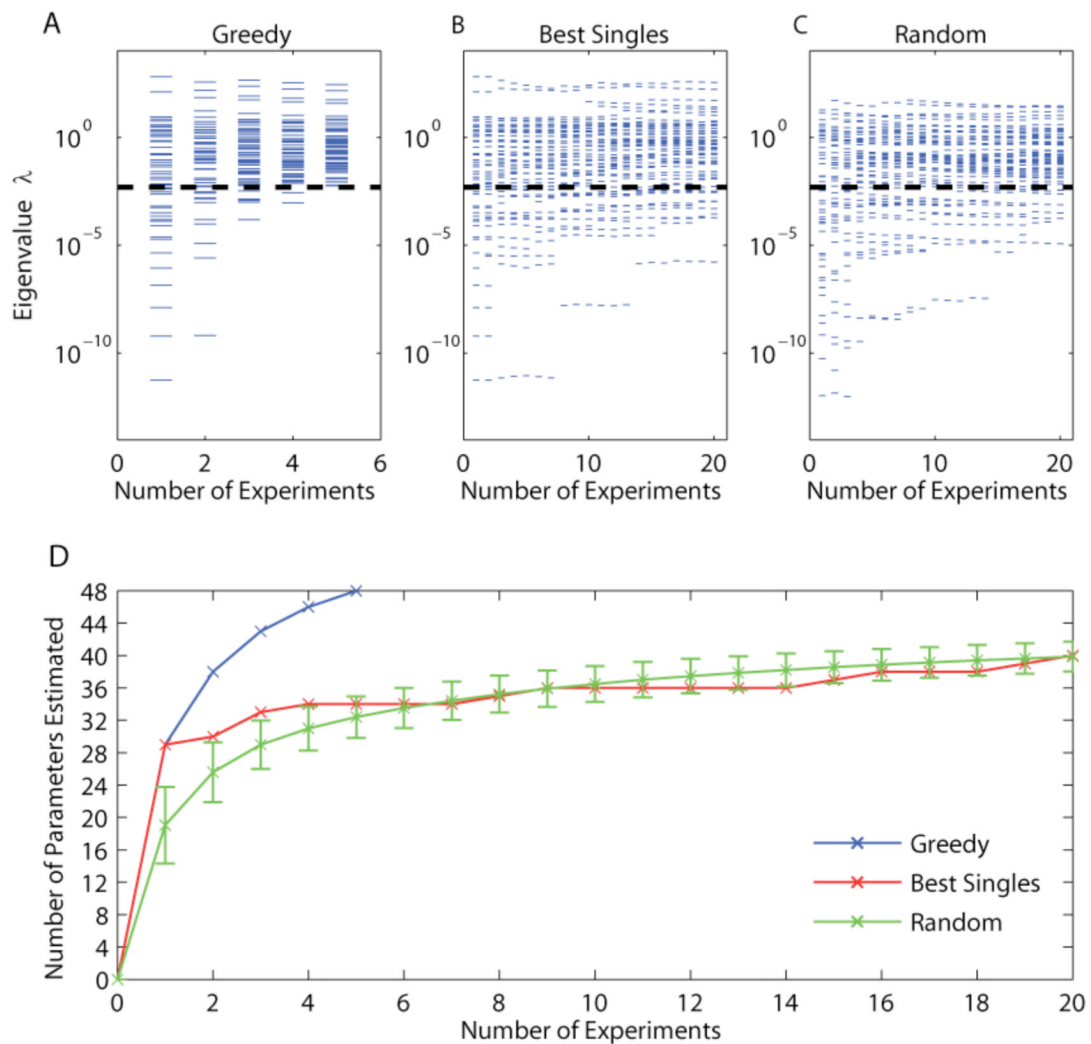
Figure 2.
(A) Search result for sets of experiments that maximize the number of parameters estimated. Each spectrum is the eigenspectrum of estimation problem. The goal of the design is to maximize the number of parameters with errors less than 10% (dashed line). (B) Design based on selecting the best single experiments. (C) Design based on selecting random experiments. (D) The number of parameters estimated by each search method. By the fifth experiment the greedy algorithm is able to estimate all 48 parameters to the desired accuracy. The error bars show the standard deviation for 10000 random searches.
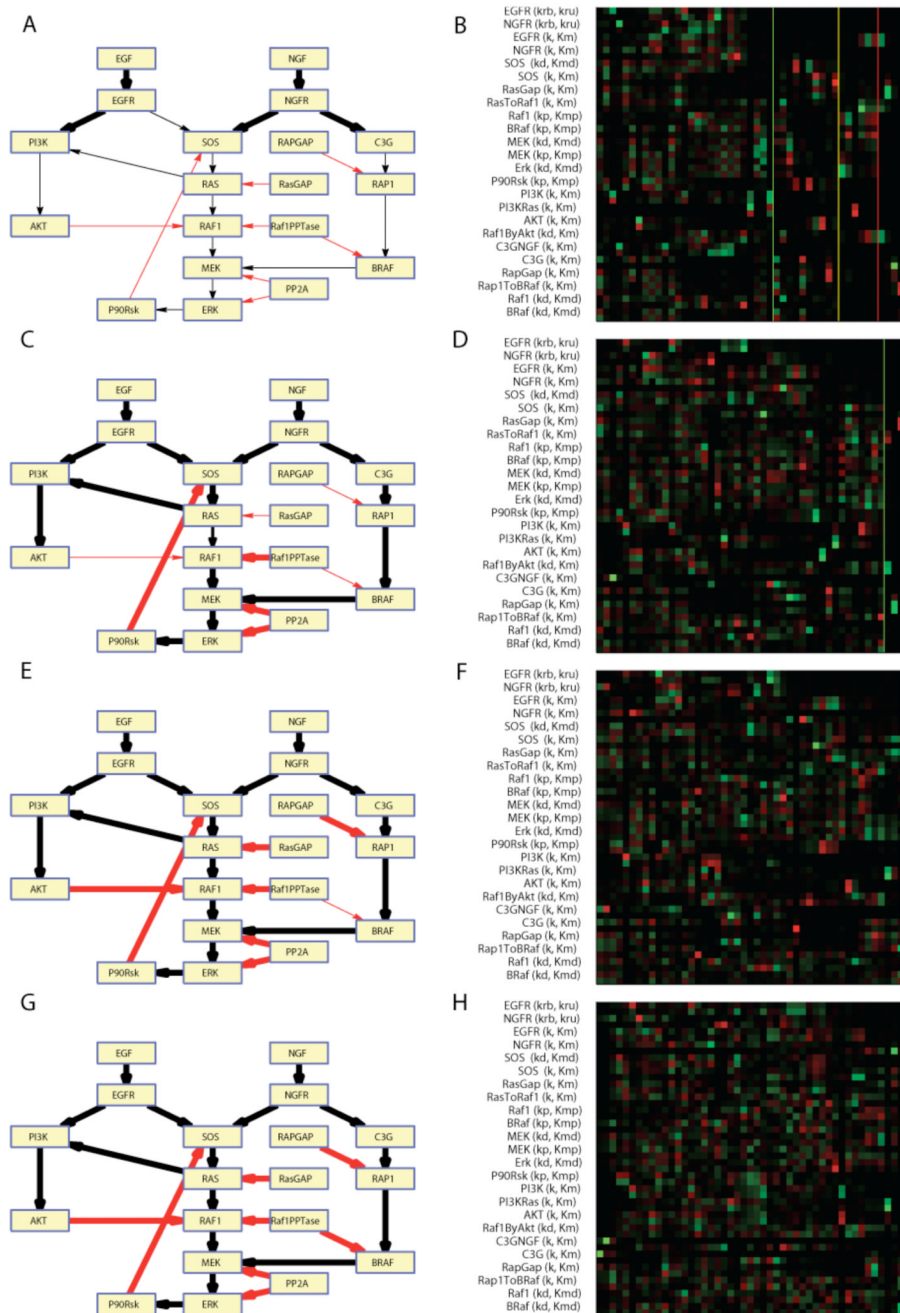
Figure 3.
The subset of parameters that can be estimated with different levels of experiment. (A,B) Doses of EGF and NGF alone. (C,D) Doses and single knockouts/overexpressions. (E,F) Doses and double knockouts/overexpressions. (G,H) Doses and triple knockouts/ overexpressions. The network diagrams show the individual parameter errors. Each arrow represents a reaction (black for activating, red for inhibitory). Each reaction is parameterized by two parameters. If both parameters are estimated to 10% the line is thick, if only one of the two parameters is estimated the line is medium, and if neither parameter is estimated then the line is thin. The eigenvector matrices pictures on the right show the vector perspective. The eigenvalues decrease from left to right. The green lines indicates the cutoff

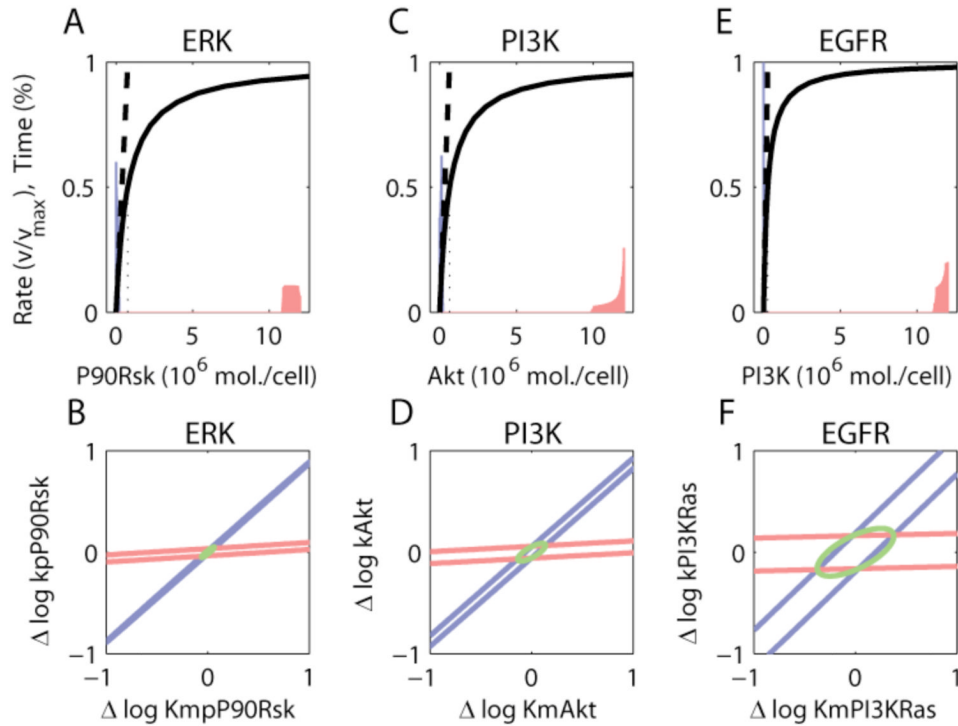for 10% relative error. The yellow and red lines indicate directions with 100% and 1000% error.

Figure 4.

Experimental perturbations push enzymes from operating in pure $\frac{k_{\text{cat}}}{K_M}$ conditions to facilitate estimation of k_{cat} and K_M individually. Experiment #4 calls for the overexpression of the three enzymes P90Rsk, Akt, and PI3K, which are substrates for the enzymes (A,B) ERK, (C,D) PI3K, and (E,F) EGFR, respectively. The top row shows the rate of reaction for each enzyme as a function of its substrate concentration (solid black line), the linearized rate law at low substrate concentration $\frac{k_{\text{cat}}}{K_M}$ (dashed black line), and a histogram of concentration in experiment #1 (blue shading) and in experiment #4 (red shading). The bottom row shows the corresponding parameter uncertainty, with K_M on the abscissa and k_{cat} on the ordinate. Uncertainty is shown for experiment #1 alone (blue ellipse), experiment #4 alone (red ellipse), and the combination of both experiments (green ellipse). In experiment #1 only the linear region of the enzyme rate curve is explored (blue shading in top row). The linear regime specifies $\frac{k_{\text{cat}}}{K_M}$ (represented as $\log(k_{\text{cat}})$–$\log(K_M)$ in log parameter space and corresponds to the blue ellipses in the bottom row of the Figure). The blue ellipses are extended along the $y = x$ direction, indicating very little uncertainty in $\log(k_{\text{cat}})$–$\log(K_M)$ but great uncertainty in $\log(k_{\text{cat}})+\log(K_M)$. In experiment #4 all three substrates are overexpressed and the saturating portion of the rate curve is populated, where $r \approx e_0 k_{\text{cat}}$, which specifies the $\log(k_{\text{cat}})$ but not $\log(K_M)$ direction, as indicated by the red ellipses being aligned nearly along the abscissa. Combining both experiments specifies both rate parameters, as indicated by the small extent of the green ellipses.
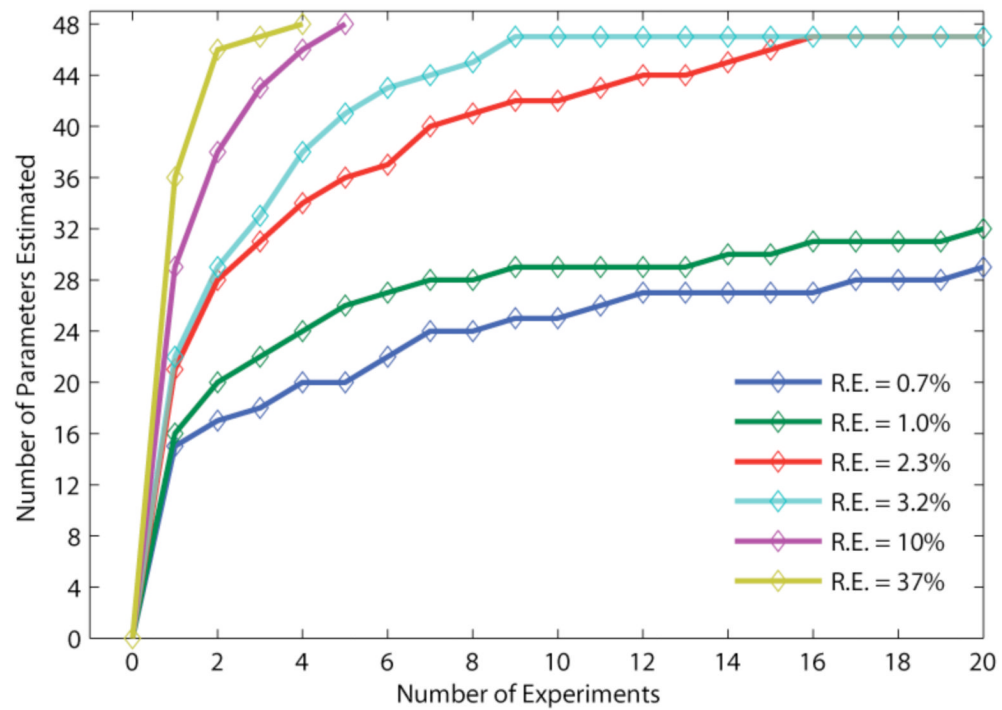
Figure 5.
Experimental design with different relative parameter errors. The violet line is the original design with 10% uncertainty. It is interesting to note that with an objective of 37% uncertainty (gold line) that all but two of the parameters can be estimated with two experiments, and all parameters can be estimated with four experiments.

Table I

Parameter-Defining Experimental Set

Exp	EGF (mol./cell)	NGF (mol./cell)	Overexpressed	Knocked Down
1	1.00×10^5	4.56×10^7	Sos, Ras, C3G	
2	1.00×10^1	4.56×10^1	Mek, Erk	Raf1PPtase
3	0.00	4.56×10^5	BRaf, Rap1	RapGap
4	1.00×10^1	4.56×10^7	P90Rsk, PI3K, Akt	
5	1.00×10^3	4.56×10^3	Raf1	RasGap