

# Maximizing the Information Content of Experiments in Systems Biology

Juliane Liepe<sup>1,9</sup>, Sarah Filippi<sup>1,9</sup>, Michał Komorowski<sup>2</sup>, Michael P. H. Stumpf<sup>1,3,\*</sup>

**1** Centre for Integrative Systems Biology and Bioinformatics, Imperial College London, London, United Kingdom, **2** Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland, **3** Institute of Chemical Biology, Imperial College London, London, United Kingdom

## Abstract

Our understanding of most biological systems is in its infancy. Learning their structure and intricacies is fraught with challenges, and often side-stepped in favour of studying the function of different gene products in isolation from their physiological context. Constructing and inferring global mathematical models from experimental data is, however, central to systems biology. Different experimental setups provide different insights into such systems. Here we show how we can combine concepts from Bayesian inference and information theory in order to identify experiments that maximize the information content of the resulting data. This approach allows us to incorporate preliminary information; it is global and not constrained to some local neighbourhood in parameter space and it readily yields information on parameter robustness and confidence. Here we develop the theoretical framework and apply it to a range of exemplary problems that highlight how we can improve experimental investigations into the structure and dynamics of biological systems and their behavior.

**Citation:** Liepe J, Filippi S, Komorowski M, Stumpf MPH (2013) Maximizing the Information Content of Experiments in Systems Biology. *PLoS Comput Biol* 9(1): e1002888. doi:10.1371/journal.pcbi.1002888

**Editor:** Andrey Rzhetsky, University of Chicago, United States of America

**Received:** July 27, 2012; **Accepted:** November 30, 2012; **Published:** January 31, 2013

**Copyright:** © 2013 Liepe et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors gratefully acknowledge funding from the Wellcome Trust (through a PhD studentship to JL; www.wellcome.ac.uk), MRC (through a Biocomputing Research Fellowship to SF; www.mrc.ac.uk) and BBSRC (to MK and MPH; BB/G020434/1, www.bbsrc.ac.uk). MPH is a Royal Society Wolfson Research Merit Award Holder (www.royalsociety.org). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: m.stumpf@imperial.ac.uk

<sup>9</sup> These authors contributed equally to this work.

## Introduction

Mathematical models of biomolecular systems are by design and necessity abstractions of a much more complicated reality [1,2]. In mathematics, and the theoretical sciences more generally, such abstraction is seen primarily as a virtue which allows us to capture the essential features or defining mechanisms underlying the workings of natural systems and processes. But while qualitative agreement between even very simple models and very complex systems is easily achieved, formally assessing whether a given model is indeed good (or even just useful) is notoriously difficult. These difficulties are exacerbated in no small measure for many of the most important and topical research areas in biology [3–5]. The regulatory, metabolic and signalling processes involved in cell fate and other biological decision-making processes are often only indirectly observable; moreover, when studied in isolation their behavior can often be markedly altered compared to the experimentally more challenging *in vivo* contexts [6]. The so-called “inverse problem” — to learn, construct or infer mathematical or mechanistic models from experimental data — is often considered (see e.g. Brenner [7]) as one of the major problems facing systems biologists.

These challenges have prompted the development of novel statistical and inferential tools, required to construct (or improve) mathematical models of such systems. We can loosely group these methods into (i) those aimed at reconstructing network models [8–10] (using correlations or statistical dependencies in observed datasets), (ii) methods to estimate (biochemical reaction) rate

parameters of models describing the dynamics of biological systems [11–13], and (iii) approaches that allow us to rank or discern between different candidate models/hypotheses [14,15]. The first set of challenges is typically faced when dealing with new systems where little information is known, and where network-inference algorithms offer a convenient way of generating novel mechanistic hypotheses from data. Here we address the second point. In particular, we start from a model that describes how the abundances of a set of molecular entities,  $x$ , change with time,  $t$ ; the rate of change in  $x(t)$  over time is typically described in terms of (ordinary, partial or stochastic) differential equation systems,

$$\frac{dx(t)}{dt} = F_q(x, \theta);$$

where  $x$  is a  $k$ -dimensional vector describing the system's state and  $\theta = (\theta_1 \dots \theta_l)$  is an  $l$ -dimensional vector containing the model parameters. Finally,  $q$  denotes the particular experimental setup under which data are collected. This dependence is generally tacitly ignored but, as we will show below, explicitly incorporating the experimental approach (and the fact that different experimental choices are typically available) into the model and any downstream statistical analysis allows us to develop strategies that yield more detailed insights into biological systems, and better models thereof. The aims of the present study are to develop experimental design strategies that allow us to infer the (unknown or only poorly known) model parameter,  $\theta$ , and to reduce the uncertainty in the predicted model behavior.

## Author Summary

For most biological signalling and regulatory systems we still lack reliable mechanistic models. And where such models exist, e.g. in the form of differential equations, we typically have only rough estimates for the parameters that characterize the biochemical reactions. In order to improve our knowledge of such systems we require better estimates for these parameters and here we show how judicious choice of experiments, based on a combination of simulations and information theoretical analysis, can help us. Our approach builds on the available, frequently rudimentary information, and identifies which experimental set-up provides most additional information about all the parameters, or individual parameters. We will also consider the related but subtly different problem of which experiments need to be performed in order to decrease the uncertainty about the behaviour of the system under altered conditions. We develop the theoretical framework in the necessary detail before illustrating its use and applying it to the repressilator model, the regulation of Hes1 and signal transduction in the Akt pathway.

Inferential tools have been developed that, given some observed biological data and a suitable mathematical candidate model, provide us with parameters that best describe the system's dynamics. Unfortunately obtaining reliable parameter estimates for dynamical systems is plagued with difficulties [16,17]. Usually sparse and notoriously noisy data are fitted using models with large number of parameters [18]. As a result, over-parameterized models tend to fit to the noisy data but may lose confidence in predictive behavior. Conventional fitting approaches to such data routinely fail to capture this complexity by underestimating the uncertainty in the estimated parameters, which substantially increases the uncertainty in prediction of model behavior.

We use  $\mathcal{Q} = \{q_1, \dots, q_{|\mathcal{Q}|}\}$  to denote the total set of available experimental assays that could be used to probe a system in a given situation. These might, for example, include knock-out or knock-down mutants, transcriptomic or proteomic assays, different time-courses or different environmental conditions (or both), etc. Here we remain very flexible as to which type of experimental setup is included in  $\mathcal{Q}$ , but merely acknowledge that it is rarely possible to probe all important aspects of a system simultaneously. Instead different techniques require different sample preparations etc., and therefore separate experiments. Here we account also for the possibility, that for two different experimental set-ups,  $q$  and  $q'$ , the mathematical model may differ; therefore the dependence on  $q$  is made explicit in our notation  $F_q$ . For example, if species  $x_k$  is knocked out in experiment  $q^*$ , we can ignore any terms referring to it when modelling  $F_{q^*}$ .

Performing different experiments is costly, however, in terms of both money and time, and not all experiments are equally informative. Ideally we would like to perform only those experiments which yield *substantial* and *relevant* information. We regard any information that decreases our uncertainty about model parameters or model predictions as *relevant*. As we will show below, what is *substantial* information is then easily and naturally resolved. We will show, for example, that experimental interventions differ in the amount of information they provide e.g. about model parameters. Equally some experiments provide insights that are more useful for making predictions about system behavior than others. It may seem surprising that we consider parameter inference and prediction of output separately, but this merely reflects the fact that not all parameters contribute equally to system

output: varying some parameters will have huge impact on the output, while varying other parameters will lead to negligible changes in the output. By making the reduction in uncertainty of predicted model behavior the target of experimental design we explicitly acknowledge this.

Experimental design in systems biology is different from classical experimental design studies. The latter theory was first developed at a time when the number of alternative hypotheses was smaller than the amount of available data and replicates [19]. Systems biology, on the other hand is hypotheses rich and data rarely suffice to decide clearly in favour of one model unambiguously. Moreover for dynamical systems, as a host of recent studies have demonstrated, generally less than half of the parameters are tightly confined by experimental data [16,17]. Together these two challenges have given rise to a number of approaches aimed at improving our ability to develop mechanistic models of such systems. Here we meld concepts from Bayesian inference and information theory to guide experimental investigations into biological systems to arrive at better parameter estimates and better model predictions.

Several authors have used the information theoretical framework, in particular the expected gain in Shannon information to assess the information content of an experiment [20–24]. Although the methodology of Bayesian experimental design is well established, its applicability has been computationally limited to small models involving only several free parameters. Recently their use for systems biology becomes possible as a result of increased computational resources. Vanlier *et al* proposed an approach that uses the Bayesian predictive distribution to assess the predictive power of experiments [25]. Huan and Marzouk used a framework, which is similar to ours, in that it maximizes mutual information via Monte Carlo approximation to find optimal experiments [26]; but they only focus on parameter inference and ignore prediction. Furthermore they only apply their method to systems with small number of parameters. Here we demonstrate how such an approach can be utilized to analyse multi-parameter models described by ordinary differential equations (ODEs) regarding both, parameter inference and prediction of system behavior. The latter is especially useful when one aims to predict the outcome of an experiment which is too laborious or impossible to perform. Our approach improves on previous methods [25–32] in a number of ways: first we are able to incorporate but do not require preliminary experimental data; second, it is a global approach that is not limited to some neighbourhood in parameter space unlike approaches solely based on e.g. the Fisher information [29,33]; third, we obtain comprehensive statistical predictions (including confidence, sensitivity and robustness assessments if desired); and we are very flexible in the type of information that we seek to optimize.

Below we first develop the theoretical concepts before demonstrating the use (and usefulness) of the Bayesian experimental design approach in the context of a number of biological systems that exemplify the set of problems encountered in practice. In order to demonstrate the practical applicability of our approach we investigate two simple models (repressilator and Hes1 systems), as well as a complex signalling pathway (AKT) with experimentally measured dynamics.

## Results

### Information content of experimental data

To achieve their full functionality mathematical models require parameter values that generally need to be inferred from experimental data. The extraction of this information is, however,

a nontrivial task and is further compounded by the need to assess the statistical confidence of parameter estimates. In the Bayesian framework for example, we seek to evaluate the conditional probability distribution,  $p(\theta|x)$ , which relates to the prior knowledge  $p(\theta)$  and the distribution of data,  $x$ , given parameters,  $p(x|\theta)$ , via Bayes' formula

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)d\theta}. \quad (1)$$

The posterior probability density function,  $p(\theta|x)$ , describes the probability of finding a parameter  $\theta$  in the volume element  $d\theta$  of parameter space, given the data, the model and the prior information. From this distribution we can obtain all relevant information about the parameters: how sensitive the solution is to varying them individually or together; how correlated different parameters are with one another; if there are non-linear dependencies among the parameters; and the level of precision with which each parameter is known in light of the data. A highly readable and enlightening account of the Bayesian formalism is given e.g. in [34]. Finding the posterior,  $p(\theta|x)$ , is usually achieved by means of powerful (if costly) computational algorithms such as Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC) methods, which also exist in the approximate Bayesian computation (ABC) framework [15,35,36].

Rather than providing a single parameter estimate the posterior distribution allows us to assess how well a parameter is constrained by data (see Figure 1 A). More formally, we measure the uncertainty about a parameter information-theoretically in terms of the Shannon entropy [37],

$$H(\Theta) = - \int p(\theta) \log(p(\theta)) d\theta, \quad (2)$$

for the prior and

$$H(\Theta|x) = - \int p(\theta|x) \log(p(\theta|x)) d\theta, \quad (3)$$

for the posterior. The information gained by collecting data  $x$  can then be expressed as  $H(\Theta) - H(\Theta|x)$ . The output of the experiment, however, is in turn "random" with distribution  $p(x)$ , and therefore the average posterior uncertainty is

$$H_X(\Theta|X) = \int H(\Theta|x) p(x) dx, \quad (4)$$

which leads to the average information gain called mutual information between  $X$  and  $\Theta$ ,

$$I(X, \Theta) = H(\Theta) - H_X(\Theta|X). \quad (5)$$

When faced with different experimental setups,  $q$ , and hence different datasets,  $X_q$ , choosing the set(s) which maximize  $I(X_q, \Theta)$  will provide the best insights into the system via improved parameter estimates [20,38,39]. This observation is the basis of our experimental design methodology which consists of computing the mutual information  $I(X_q, \Theta)$  for every experiment  $q$  and selecting the experiment resulting in the highest mutual information (see *Methods* for computational details). Once the chosen experiment has been carried out, the new data are used to update the model and the posterior distribution of the parameters (see Figure 1 B).

Given the importance of the predictive role of mathematical modelling it is also of interest to reduce the uncertainty of model predictions; intriguingly and perhaps counterintuitively — but demonstrably and provably (see below and *Supplementary Material*) — better parameter estimates are not necessarily required for better, more certain model predictions. Therefore, instead of focussing on parameter inference we can directly seek to identify the experimental condition  $q^*$  which minimizes the uncertainty in the predicted trajectories  $Y$ . Analogously to the previous case minimizing uncertainty in predictions of  $Y$  means to maximize mutual information between  $X$  and  $Y$  (see *Methods*):

$$I(X, Y) = H(Y) - H(Y|X). \quad (6)$$

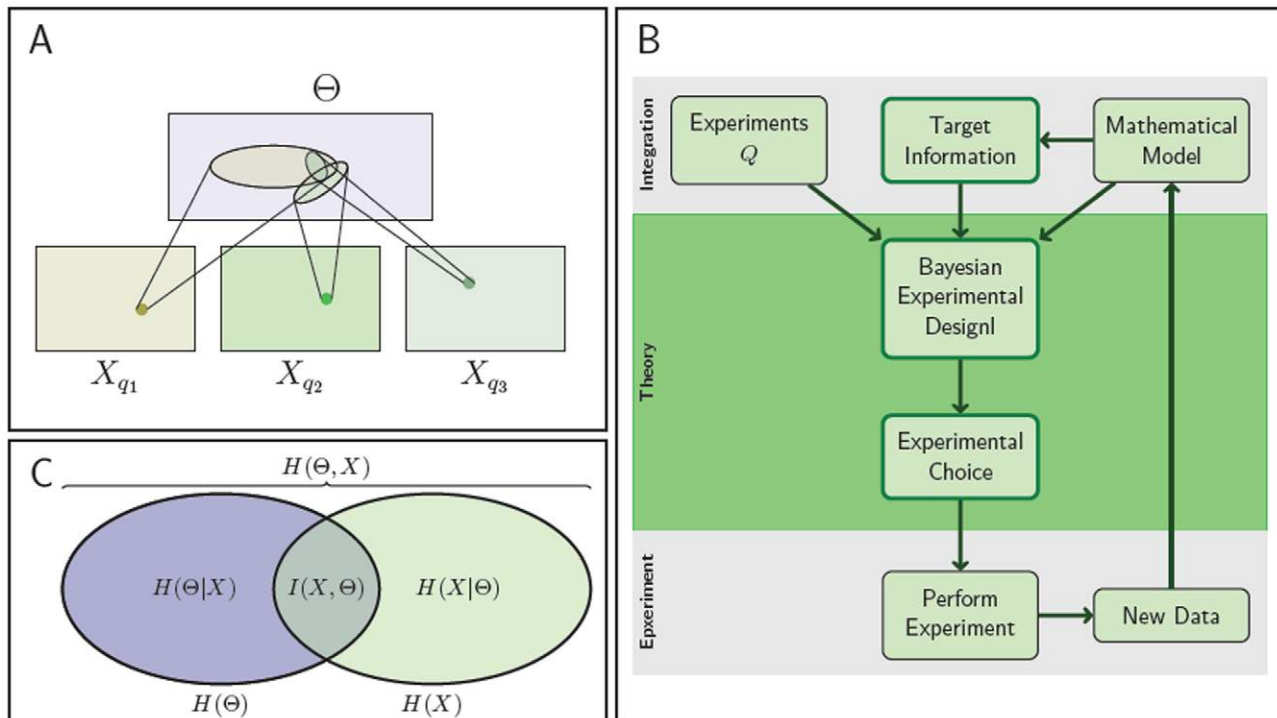
Below we use three examples of different complexity to show how this combination of rigorous Bayesian and information theoretical frameworks allows us to design/choose optimal experimental setups for parameter/model inference and prediction, respectively.

### Experiment selection for parameter inference

To investigate the potential of our experimental design method for parameter estimation we first apply it to the repressilator model, a popular toy model for gene regulatory systems [40]. It consists of three genes connected in a feedback loop, where each gene transcribes the repressor protein for the next gene in the loop (see Figure 2 A and B).

To infer the parameters of this model,  $h$ ,  $\alpha$ ,  $\alpha_0$ , and  $\beta$ , we propose 5 sets of possible experiments: the original repressilator model (set 1) which is described in Figure 2 A and corresponds to the ordinary differential equations in Figure 2 B, and 4 modifications of the original model, see Figure 2 C. The suggested 4 experiments are hypothetical, but can be linked to potential experiments. For example, a decrease in the parameter  $\alpha_0$  corresponds to a decrease of the basal transcription rate, which could be achieved with inhibitors or site-directed mutation of the corresponding transcription factor binding site. The proposed modifications can lead to different dynamics, and this in turn can lead to a higher mutual information between the parameters and the resulting mRNA and protein trajectories, which are here the output of the system. The information content increases as differences in the outputs resulting from different parameter values increase. In Figure S1, we illustrate the link between the increase in mutual information and the dynamics of the system for three different regimes.

To determine which experiment to carry out we compute the mutual information between the parameter prior distribution and the system output via Monte-Carlo estimation. We use uniform priors over  $[1,10]$  for  $h$ , over  $[0,20]$  for  $\alpha_0$ , over  $[500,2000]$  for  $\alpha$ , and over  $[0,10]$  for  $\beta$ . Figure 3 shows that experiment 2 and 5 have highest mutual information, i.e. carrying out those experiments will decrease the uncertainty in the parameter estimates most. To confirm this we simulate data for the 5 experiments using the parameter  $(h^*, \alpha^*, \alpha_0^*, \beta^*) = (2, 10, 1000, 5)$ . The simulated data are shown in Figure S2. Based on these data we perform parameter inference using an approximate Bayesian computation approach [41] for each experiment separately and compare the posterior distributions shown in Figure 3. We observe that using the data generated from set 1 (original repressilator system) only 2 parameters can be inferred with confidence:  $h$  and  $\alpha_0$ . By contrast, the data generated by set 2 and set 5 allow us to estimate all 4 parameters. In addition, for each experiment we compute the



**Figure 1. Information content of experimental data and flow chart of the experimental design method.** (A) The regions of plausible parameters values for three different experiments. Each ellipse defines the set of parameters which are commensurate with the output  $X_q$  of an experiment  $q$ . In this example, the data  $X_{q_3}$  leads to the most precise inference of the parameters. The parameters which explain the output of all the three experiments are at the intersection of the three ellipsoids. (B) Flowchart of the experimental design method. Given a mathematical model of the biological system, a set of experiments and the target information—which can be either a set of parameters to infer or a description of the experiment to predict—the Bayesian Experimental Design method determine the experiment to carry out. Once the experiment has been performed, the experimental data are then used to provide target information and to improve the model. Thereafter, the process can be iterated to select other experiments in order to improve the accuracy of the target information. (C) Link between the total and conditional entropies and the mutual information of experimental data  $X$  and parameters  $\theta$ . doi:10.1371/journal.pcbi.1002888.g001

reduction of uncertainty from the prior to the posterior distribution. The results are consistent with the results using mutual information and confirm that we should choose experiment 2 or 5 for parameter inference. In practice not all molecular species may be experimentally accessible and it is therefore also of interest to decide which species carries most information about the parameters. We can estimate the mutual information between the parameter and each species independently, and, for example, for experimental set 5 we observe that mRNA  $m_1$  and  $m_2$  as well as protein  $p_1$  carry equally high information. This mutual information for each mRNA and each protein is plotted in Figure S3.

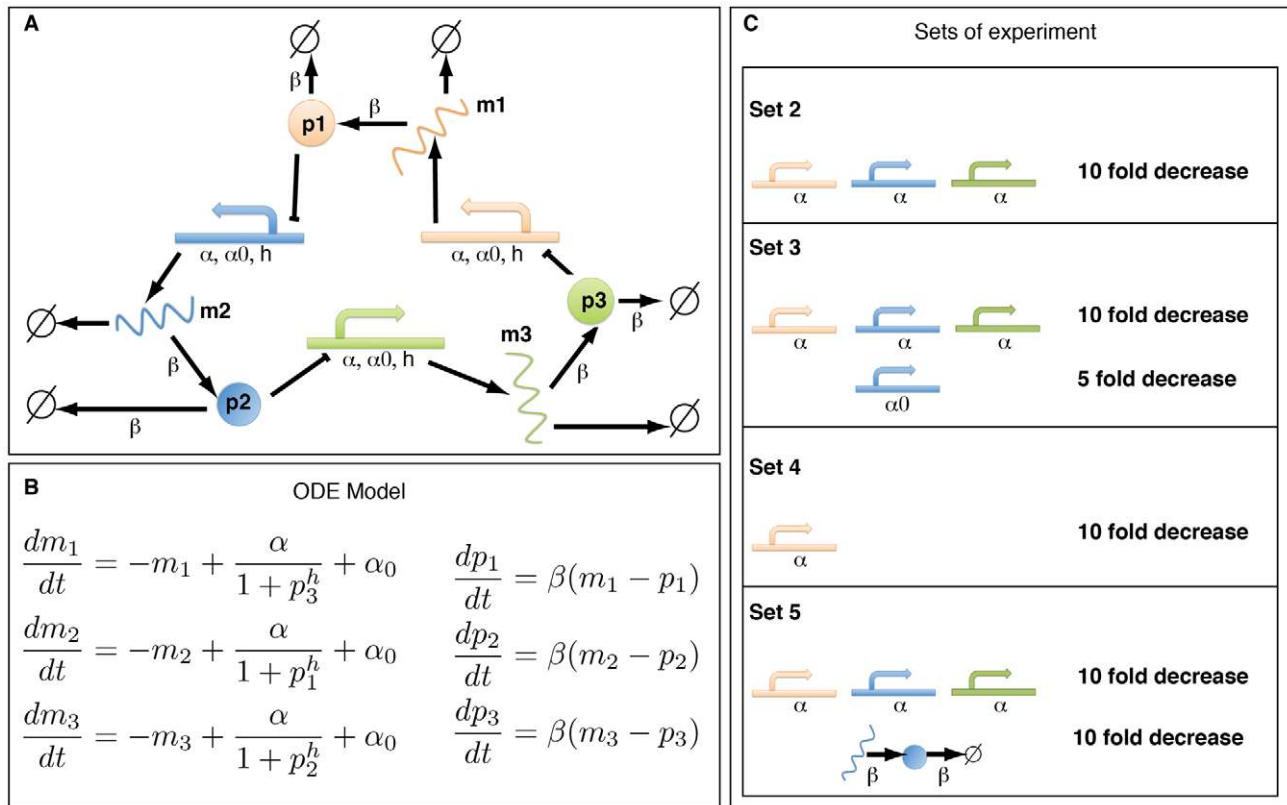
Sometimes we are interested in estimating only some of the parameters, e.g. those that have a direct physiological meaning or are under experimental control. To investigate this aspect we consider the Hes1 transcription factor that plays a number of important roles, including in the cell differentiation and segmentation of vertebrate embryos. Oscillations observed in the Hes1 system [42] might be connected with formation of spatial patterns during development. The Hes1 oscillator can be modelled by a simple three-component ODE model [43] as shown in Figure 4 A. This model contains 4 parameters,  $k_1$ ,  $P_0$ ,  $v$ , and  $h$ , and 3 species: Hes 1 mRNA,  $m$ , Hes 1 nuclear protein,  $p_1$ , and Hes 1 cytosolic protein,  $p_2$ . It is possible to measure either the mRNA (using real-time PCR) or the total cellular Hes 1 protein concentration  $p_1 + p_2$  (using Western blots). We investigate whether protein or mRNA measurements provide more information about the model parameters. Thus we estimate the mutual information between

mRNA and parameters, and between protein and parameters. Figure 4 B shows that mRNA measurements carry more information about all of the parameters.

This can again be further substantiated by simulations. We perform parameter inference based on such simulated data (simulated data are shown in Figure S5) and compute the difference between the entropy of the prior and that of the resulting posterior distribution. The results shown in Figure 4 C are consistent with the predictions based on mutual information: mRNA measurements carry more information for parameter inference. Interestingly, however, although the mutual information computation indicates that the protein measurements should contain more information about parameter  $k_1$  than about the other parameters, this is not confirmed by the difference in entropy result for this simulated data set. This divergence is due to the fact that the mutual information measures the amount of information contained *on average* over all the possible behaviours of the system, whereas Figure 4 C represents the decrease in entropy from the prior to the posterior distribution given *specific data*. The differences in entropy for other data sets simulated using different parameter regimes are thus in better agreement with the mutual information results. We confirm this in Figure S6 where we show the results of the same analysis based on a different simulated data set.

#### Experiment selection for prediction

We next focus on a scenario where we aim to predict the behaviour of a biological system [44] under conditions for which it



**Figure 2. The repressilator model (as described in [40]) and the set of possible experiments.** (A) Illustration of the original repressilator model. The model consists of 3 mRNA species (coloured wavy lines, labeled  $m_1$ ,  $m_2$  and  $m_3$ ) and their corresponding proteins (circles shown in the same color with labels  $p_1$ ,  $p_2$  and  $p_3$ ). The 3 regulatory DNA regions are not modeled explicitly but included in the mRNA production process. They are shown for illustration purpose only. (B) The ordinary differential equations which describe the evolution of the concentration of the mRNAs and proteins over time. (C) Four potential modifications of the wild-type model. For each experimental intervention the modified parameters are listed (colours are as in A). The modifications of the wild-type model consist of decreasing one or several of the parameters of the system: in sets 2, 3 and 5, the regime of the parameter  $\alpha$  is changed; in sets 3, 4 and 5, respectively, parameters  $\alpha_0$ ,  $\alpha$  and  $\beta$  are modified for only one gene which breaks the symmetry of the system.

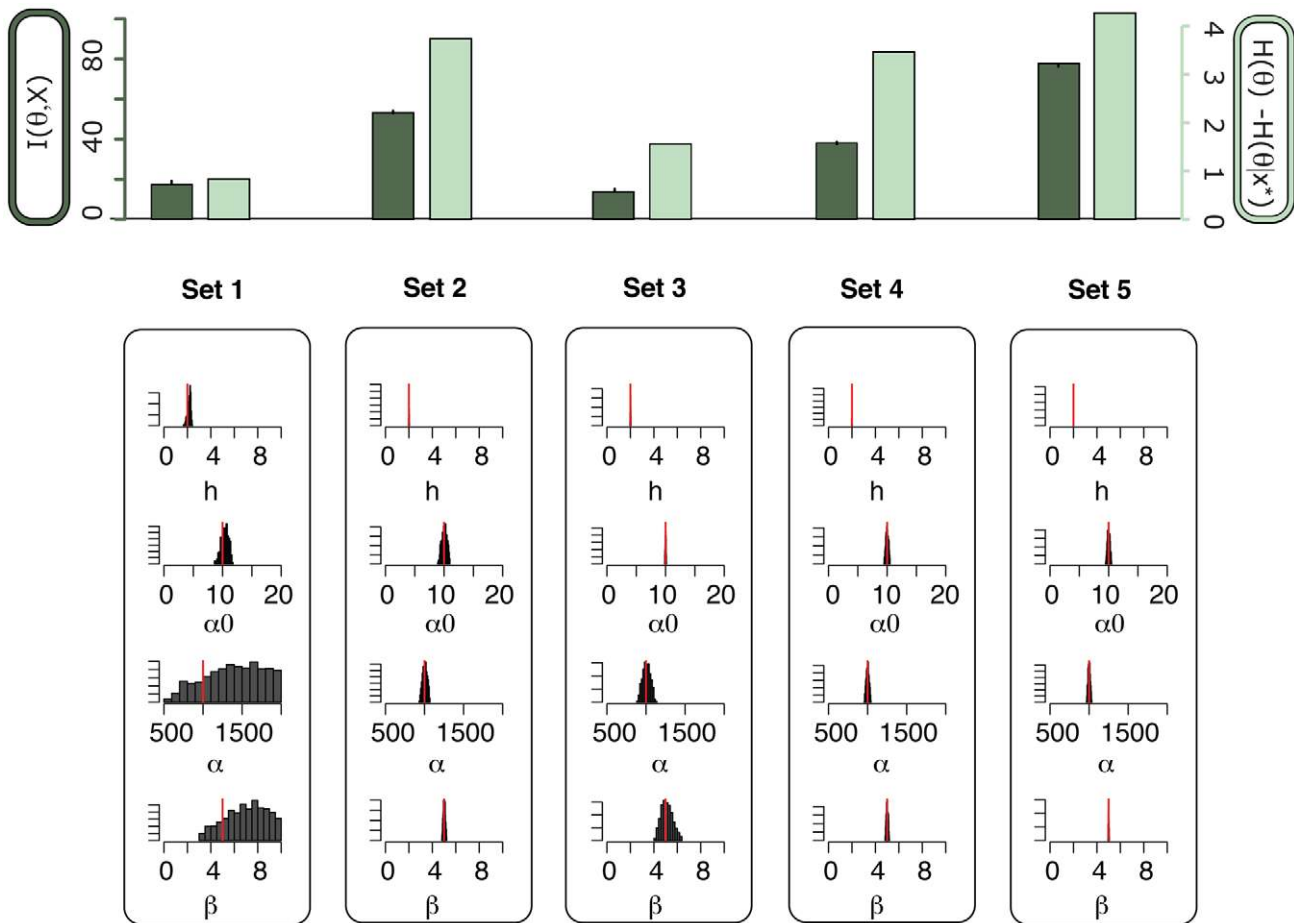
doi:10.1371/journal.pcbi.1002888.g002

is not possible to obtain direct measurements. We consider as an example the phosphorylation of Akt and ribosomal binding protein S6 in response to a epidermal growth factor (EGF) signal. Figure 5 A shows the pathway of interest: the EGF growth factor binding to the activated receptor EGFR leads to phosphorylation of EGFR and a signal cascade which results in the phosphorylation of Akt (pAkt) which in turn can activate downstream signalling cascades and leads to the phosphorylation of S6 (pS6); a corresponding mathematical model is shown in Figure S7 [45].

We are interested in predicting the dynamics under multiple pulsed stimuli with EGF in the presence of background noise, as shown in Figure 6 A. We consider 5 pulses of intensity 1 ng/ml and length 60 seconds spaced by 400 seconds with additive background noise. This input is difficult to realize in an experimental system (let alone an animal or clinical setting). Using an initial data set, see Figure 5 B, we can infer system parameters using ABC SMC. The resulting fit to the data using the inferred parameters are shown in Figure S8. From the resulting posterior distribution we then sample 1000 parameter combinations and simulate the model with the 5-pulse-stimulus in order to predict the time courses of phosphorylated EGF receptor (EGFR), phosphorylated Akt and phosphorylated S6; based on just the estimated parameters these predictions are, however, associated with high uncertainty, see Figure 6 B.

To obtain better predictions we can use data from other experiments measuring the time course of the 3 species of interest for a experimentally more straightforward input signals chosen from among 12 possible stimuli: impulse, step or ramp stimuli with different EGF concentrations (see Figure 6). To determine which of those inputs would result in the most reliable predictions we compute the mutual information between the time courses for the different potential experimental inputs and the time-course of the target 5-impulse noisy stimulus. We incorporate initial information about model parameters by computing the mutual information based on the posterior distribution inferred above. Figure 6 C shows that a step stimulus of intensity 3 ng/ml has the highest mutual information and therefore reduces the uncertainty of the predicted behavior of our target stimulus pattern.

In Figure 6 D we show that this does indeed yield much improved predictions compared to the initial prediction. This reduction in uncertainty about predicted model behaviour results from the difference in the posterior distributions obtained under different stimulus regimes; by focussing on predictive ability we focus implicitly on data that is informative about those parameters that will affect the system behaviour most under the target (5-pulse) stimulus. The posterior distributions are represented in Figure 6 E for two parameters, the EGFR turn over and the EGFR initial concentration, which appear to be essential for the



**Figure 3. Experiment choice for parameter inference in the repressator model.** Top: The mutual information  $I(\theta, X)$  between the parameters  $\theta$  and the output of each set of experiment (in dark green), and the entropy difference between the prior distribution and the posterior distribution. The posterior distribution is based on data obtained from simulation of the system for each experiment (in light green). The error bars on the mutual information barplots show the variance of the mutual information estimations over 3 independent simulations. Bottom: For each set of experiment we show the histogram of the marginal of the posterior distribution of every parameters. The red line indicates the true parameter value. doi:10.1371/journal.pcbi.1002888.g003

prediction of the evolution of Akt/S6 phosphorylation patterns under the 5-pulse stimulus. Those two parameters were not inferred using the initial dataset alone, whereas the addition of the outcome of the step stimulus experiment suggested by our methods infers these parameters at the required precision.

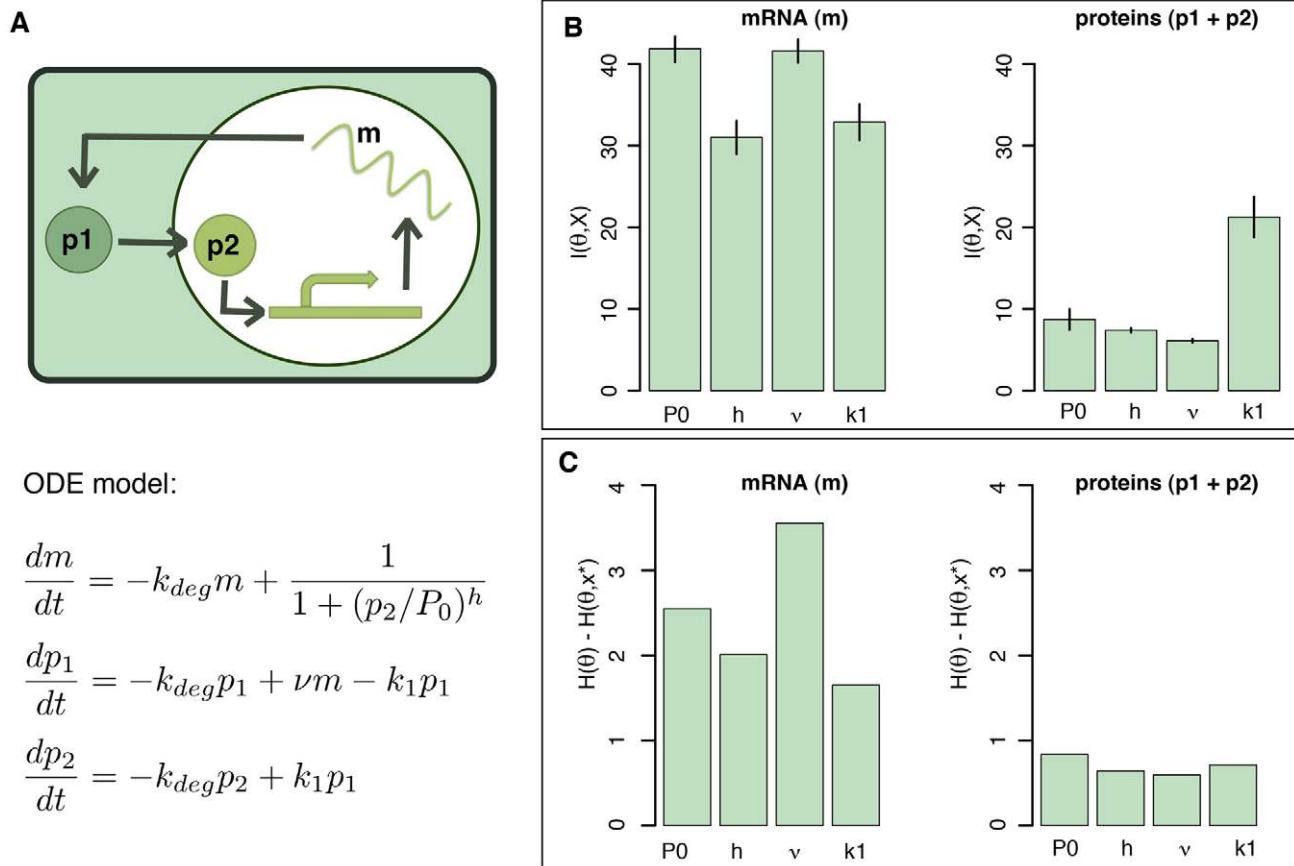
This ability to predict the time courses extends to much greater signal distortion and even with a noise level of 100 percent of the signal intensity we find that our experimental design method yields similar improvements in the predictions. This additional analysis of the new input signal is shown in Figure S10 and S11. We observe that the direct target (EGFR receptor) as well as activated AKT (pAKT) efficiently filter out the noise but capture the 5 pulses; EGFR activation quickly returns to base level in response to the higher frequency background noise. This indicates that there might be a constant low concentration of activated EGFR (pEGFR), but the activation of S6 has very different characteristics and is far less robust to noise. The level of noise is amplified as can be observed in the pS6 time course. This might suggest that the downstream molecule pS6 has a longer time delay to react to a signal. Moreover, pS6 does not have time to relax to its baseline between the 5 pulses, which leads to incremental signal amplification. This behavior fits with the low-pass filter characteristics previously described [45]. In further support earlier studies

[46] found that a downstream molecule can be more sensitive to an upstream activator than the direct target molecule of the activator. This might explain that the activation of EGFR and AKT is more robust to noise than the downstream molecule, S6.

## Discussion

We have found that maximizing the mutual information between our target information — here either model parameter values or predictions of system behaviour — and the (simulated) output of potentially available experiments offers a means of arriving at optimally informative experiments. The experiments that are chosen from a set of candidates are always those that add most to existing knowledge: they are, in fact, the experiments that most challenge our current understanding of a system.

This framework has a number of advantages: First, we can simulate cheaply any experimental set-up that can in principle be implemented; second, using simulations allows us to propagate the model dynamics and to quantify rigorously the amount of (relevant) information that is generated by any given experimental design; third, our information measure gives us a means of meaningfully comparing different designs; finally, our approach can be used to design experiments sequentially — our preferred



**Figure 4. Experiment selection for parameter inference in the Hes 1 model.** (A) Diagram of the Hes 1 model and the associated ordinary differential equations. The model shows a cell (green box) with its nucleus (white circle). Hes 1 mRNA (wavy line with label  $m$ ) is produced inside the nucleus and translated into Hes 1 cellular protein ( $p_1$ ), which in turn is then transported into the nucleus and becomes nuclear Hes 1 protein ( $p_2$ ).  $p_2$  is regulating the production of  $m$ . (B) The mutual information between each parameter and the output of the system for respectively mRNA measurement (left) and total cellular Hes 1 measurement (right). The barplot represents the mean and the variance over 3 repetitions of the Monte-Carlo estimation. (C) Estimation of the difference between the entropy of the parameter prior and the entropy of the posterior distribution given one dataset. The dataset results from simulation of the Hes 1 system. The parameter  $k_{deg}$  was set to be  $0.03 \text{ min}^{-1}$  as experimentally determined by [42]. doi:10.1371/journal.pcbi.1002888.g004

route as this will enable us to update iteratively our knowledge of a system along the way — or in parallel, i.e. selecting more than one experiment. Previous approaches had taken a more local approach [25,29,30,47] that relied on initial parameter guesses and often data; our approach also readily incorporates different stimulus patterns [48].

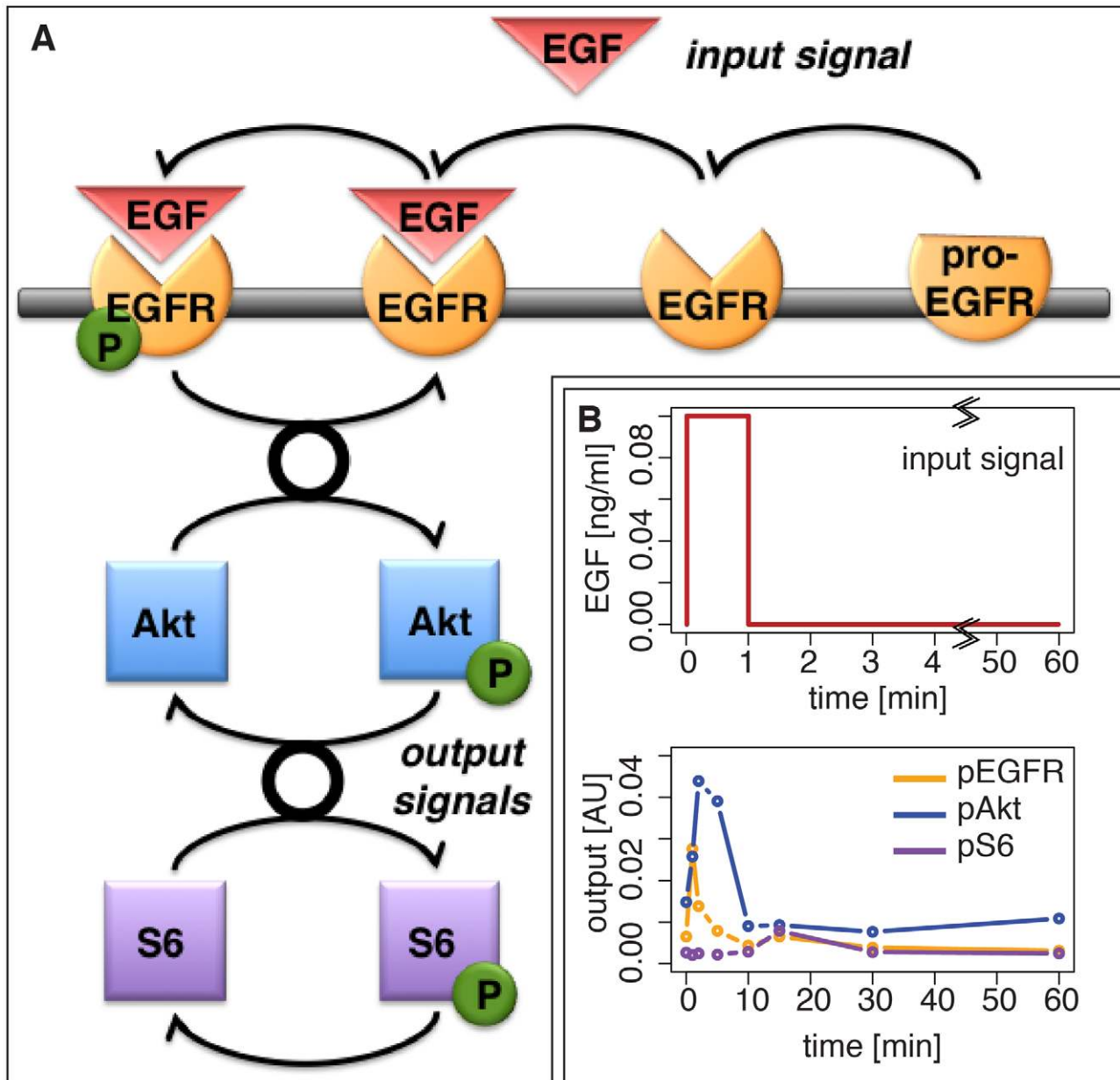
Here we have focussed on designing experiments that increase our ability to estimate model parameters and to predict model behaviour. The latter depends on model parameters in a very subtle way: not all parameters affect system output equally and under all conditions. Target conditions could, for example, include clinical settings which are generally not experimentally amenable (at least in early stage research); here the current approach offers a rationale for designing [49] therapeutic interventions into complex systems based on investigations of suitable model systems. In this study we provide an approach to chose the optimal experiment out of a finite and discrete set of possible experiments. Experimental design with a continuous set of experiments requires a different approach, as for example shown in [50].

With an optimal design we can overcome the problems of sloppy parameters [16] (which are, of course, dependent on the experimental intervention chosen [17]) and can narrow down the posterior probability intervals of parameters. We would like to reiterate, however, the importance of considering joint distributions

rather than merely the marginal probabilities of (or the confidence regions associated with) individual parameters: parameters of dynamical systems tend to show high levels of correlation (i.e. we can vary them simultaneously in a way that does not affect the output of the system — at least in some areas of parameter space) and their posterior probability distributions often deviate from normality (which also motivated the use of information theoretical measures which can deal with non-linearities and non-Gaussian probability distributions).

We tested and applied our approach in three different contexts: while the repressilator serves as a toy model with hypothetical data and experiments, the Hes1 transcription regulatory system and the EGF induced Akt pathway are relevant biological systems. The question to answer in the Hes 1 system, which molecular species to measure, is a common question in laboratories. The presented study of the Akt system does not only demonstrate how to chose experiments for prediction of system behaviours, but it is also an example for stimulus design.

The approach we presented here yields the potential for model discrimination or checking the target of our analysis [48,51], and, for example, choose experimental designs that maximize our ability to distinguish between competing alternative hypotheses or models. All of this is straightforwardly reconciled in the Bayesian framework, which also naturally lends itself to such iterative procedures where



**Figure 5. The EGF-dependent AKT pathway and an initial dataset.** (A) Diagram of the model of the EGF-dependent AKT pathway. Epidermal growth factor (EGF, red triangle) is a stimulus for a signalling cascade, which results in the phosphorylation (green circle) of Akt (blue square) and S6 (purple square). EGF binds to the EGF membrane receptor EGFR (orange), which is generated from a pro-EGFR. The binding results in the phosphorylation of the receptor, which consequently leads to the activation of downstream cascades (thick black circle). This simplified model was shown to capture the experimentally determined dynamics [45]. (B) A impulse input of EGF over 60 seconds with an intensity of 0.1 ng/ml (top) and the resulting time course of phosphorylated EGF receptor (pEGFR), phosphorylated Akt (pAkt) and phosphorylated S6 (pS6) in response to this stimulus (bottom). Data were provided by the authors of [45].  
doi:10.1371/journal.pcbi.1002888.g005

“today’s posterior” is “tomorrow’s prior” and models are understood increasingly better as new, more informative data are systematically being generated.

## Methods

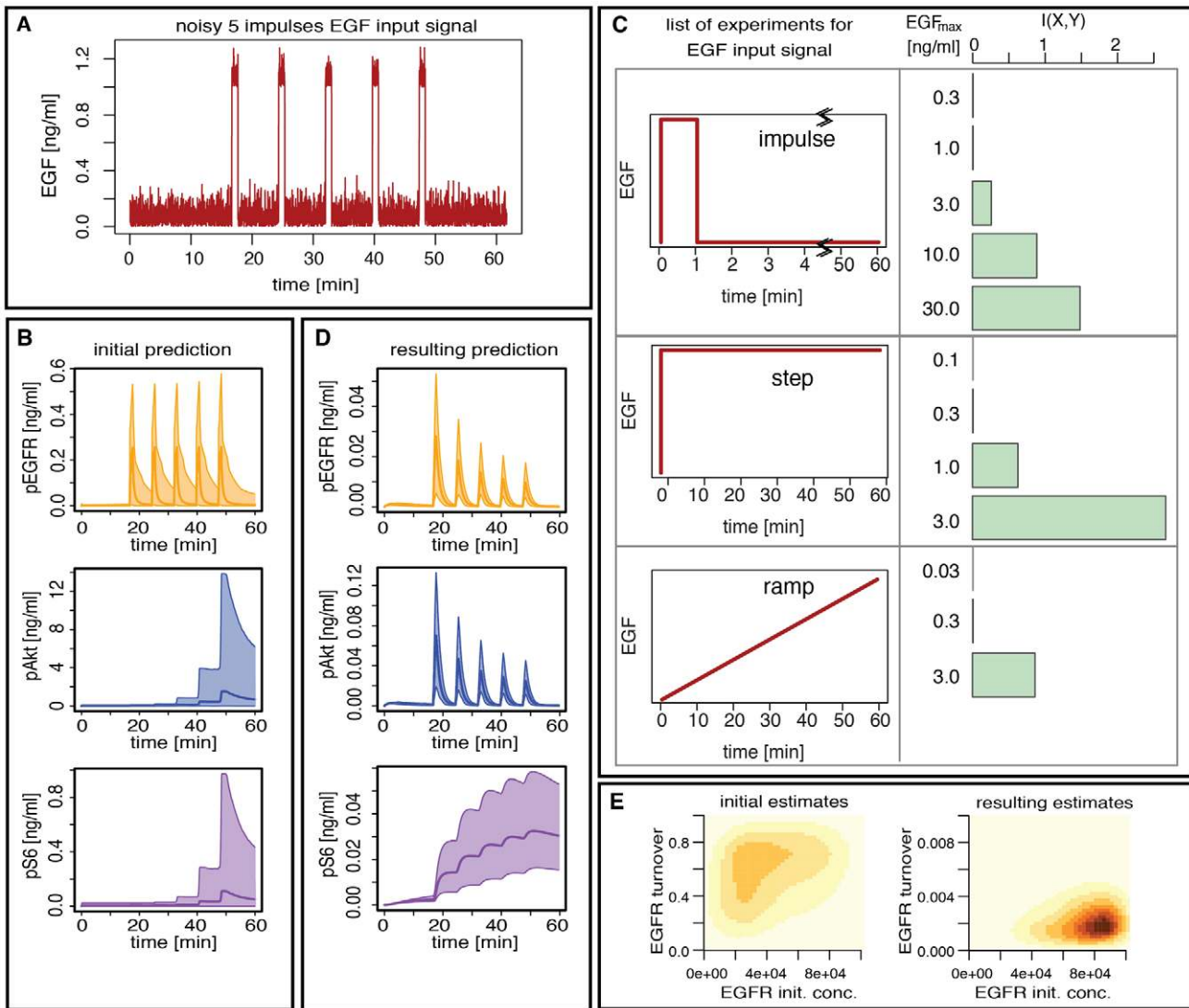
### Information theoretic design criteria

Our aim is to choose an experiment  $q$  from a set of candidate experimental setups,  $Q$ , which either reduces uncertainty about model parameters or uncertainty of an outcome of a particular condition  $q^*$  for which data are impossible or difficult to obtain. In

the information theory framework, these two goals boil down to determining an experiment  $q \in Q$  which contains maximal information about the parameter or the desired predictions for condition  $q^*$ . In order to present in more details these two goals and for better understanding we revise some concepts of information theory [34]. We define the entropy  $H(X)$  of a random variable  $X$ , which measures the uncertainty of the random variable,

$$H(X) = -E_X(\log(p(X))) = -\int \log(p(x))p(x)dx \quad (7)$$





**Figure 6. Experiment selection for prediction in the EGF-dependent AKT pathway.** (A) The noisy 5-impulses EGF input signal: the 5 pulses are of intensity 1 ng/ml and length 60 seconds spaced by 400 seconds with an additive background noise which is the absolute value of a gaussian white noise of variance  $\tau^2 = 0.1$ . (B) The predicted time course of the proteins pEGFR, pAKT and pS6 under the noisy 5-impulses EGF input signal based on the initial dataset. (C) The mutual information between the time course of the 3 species of interest under the noisy 5-impulses EGF input signal and the time course of the species under each of the following 12 possible experiments: an impulse stimulus of length 60 seconds with 5 possible intensity (0.3, 1, 3, 10 and 30 ng/ml), a step stimulus of length 60 minutes with 4 possible intensity (0.1, 0.3, 1 and 3 ng/ml) and a ramp stimulus of length 60 minutes with 3 possible final intensity (0.03, 0.3 and 3 ng/ml). (D) The predicted time course of the proteins pEGFR, pAKT and pS6 under the noisy 5-impulses EGF input signal based on the outcome of the step stimulus with intensity 3 ng/ml, which is the experiment with the highest mutual information. The scale of the y-axis is different for Figures (B) and (D). (E) The posterior distribution of two parameters (EGFR turnover and EGFR initial condition) when using the initial dataset alone (left) and when using the initial dataset and the outcome of the step stimulus with intensity 3 ng/ml (right). The scale of the EGFR turnover is the prior range for the figure on the left panel whereas it is 100 times smaller for the figure in the right panel.  
doi:10.1371/journal.pcbi.1002888.g006

and the mutual information  $I(X, Y)$  between two random variables  $X, Y$ , which is the reduction of the uncertainty that knowing  $Y$  provides about  $X$ ,

$$I(X, Y) = H(X) - E_Y(H(X|Y)) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (8)$$

where  $p(x, y)$  is the joint probability density function of  $X$  and  $Y$  while  $p(x)$  and  $p(y)$  are the marginal probability density functions. We denote by  $E_X$  the expectation with respect to the probability distribution of  $X$ . Here we follow the convention where capital

letters stand for random variables while lower-case stand for a particular realization of a random variable.

### Reducing uncertainty in model parameters

We first consider the task of choosing an experiment that will on average provide most information about model parameters measured through the reduction in their respective uncertainties. In the information theoretic language, as by Lindley [20] and later by Sebastiani and Wynn [52] the initial (prior) uncertainty is given by the entropy  $H(\Theta)$  of the prior distribution  $p(\theta)$ , which after data  $x_q$  have been collected (in experimental setup  $q$ ) gives rise to

the entropy  $H(\Theta|x_q)$  of the posterior distribution  $p(\theta|x_q)$ . The information gained about the parameter by collecting the data  $x_q$  is then  $H(\Theta) - H(\Theta|x_q)$ . On average, however, the decrease of uncertainty about  $\Theta$  after data are collected in an experiment  $q$  is given by  $I(X_q, \Theta)$ . Therefore, in order to reduce parameters' uncertainties one should choose an experiment that maximizes the mutual information between  $X_q$  and  $\Theta$ .

Here we specifically consider models such that the output is of the form

$$x_q = \mu(\theta, q) + \epsilon \tag{9}$$

where  $\mu$  is a deterministic function and  $\epsilon$  an uncorrelated, zero mean, gaussian random variable with variance  $\sigma^2$  (our approach is readily extended to stochastic systems). In such a model maximization of mutual information  $I(\Theta, X_q)$  is equivalent to maximization of the entropy  $H(x_q)$ . This observation described first in [52] results directly from the fact that the mutual information  $I(\Theta, X_q)$  can be written as the difference between  $H(X_q)$  and  $E_{\Theta}(H(X_q|\Theta))$  and that

$$E_{\Theta}(H(X_q|\Theta)) = - \iint p(\theta)p(x_q|\theta) \log(p(x_q|\theta)) dx_q d\theta$$

does not depend on the experiment  $q$ . Indeed, Equation (9) implies that  $p(x_q|\theta)$  is the probability of the experimental noise  $\epsilon$ . Therefore maximization of  $I(X_q, \Theta)$  is equivalent to maximization of  $H(X_q)$ . However, this is only the case for the mutual information between the output  $x_q$  of an experiment  $q$  and the parameter of the system  $\Theta$ . Whenever we are interested in the increase of information about only one component of the parameter vector, or in reducing uncertainty about an experimental outcome we need to use the mutual information and not the entropy.

### Reducing uncertainty in an experimental outcome

Similar reasoning leads us to a criterion for selecting an experiment  $q$  that reduces uncertainty about predictions for the system output under a different set of conditions or experiment  $q^*$ . Choosing  $q$  that maximizes  $I(X_q, X_{q^*})$  leads to an experiment that on average reduces the uncertainty of predictions for condition  $q^*$  most. This can be seen by rewriting (8)

$$I(X_q, X_{q^*}) = H(X_{q^*}) - E_{X_q}(H(X_{q^*}|X_q)). \tag{10}$$

### Estimation of the mutual information

The mutual information for models of type (9) can be estimated using Monte Carlo simulations [26,53]. We first focus on the mutual information between parameters  $\Theta$  and the output  $X_q$  of an experiment  $q$ , which can be written as a function of the prior distribution  $p(\theta)$ , the probability of the output given the parameter  $p(x_q|\theta)$  and the evidence  $p(x_q)$  as follows

$$\begin{aligned} I(\Theta, X_q) &= \iint p(\theta, x_q) \log \frac{p(\theta, x_q)}{p(\theta)p(x_q)} d\theta dx_q \\ &= \iint p(\theta)p(x_q|\theta) \log \frac{p(x_q|\theta)}{p(x_q)} d\theta dx_q. \end{aligned} \tag{11}$$

Drawing a sample  $\{\theta^{(i)}\}_{1 \leq i \leq N_1}$  from the prior distribution  $p(\theta)$  we obtain a Monte-Carlo estimate,

$$I(\Theta, X_q) \approx \frac{1}{N_1} \sum_{i=1}^{N_1} \log \frac{p(x_q^{(i)}|\theta^{(i)})}{p(x_q^{(i)})}, \tag{12}$$

where for all  $1 \leq i \leq N_1$ ,  $x_q^{(i)}$  is an output of the system for the parameter  $\theta^{(i)}$ . For models of type (9)  $p(x_q|\theta)$  is the probability density function of a Gaussian distribution with mean  $\mu(\theta, q)$  and covariance  $\sigma^2 \mathbf{I}$  taken at  $x_q$ . To compute the quantity in (12) we have to estimate the evidence  $p(x_q^{(i)})$ , which can be done via Monte Carlo simulation: given a  $N_2$ -sample  $\{\theta^{(j)}\}_{N_1+1 \leq j \leq N_1+N_2}$  drawn independently from the prior distribution  $p(\theta)$  with  $\{\theta^{(i)}\}_{1 \leq i \leq N_1}$  we have

$$p(x_q^{(i)}) = \int p(x_q^{(i)}|\theta)p(\theta)d\theta \approx \frac{1}{N_2} \sum_{j=N_1+1}^{N_1+N_2} p(x_q^{(i)}|\theta^{(j)}). \tag{13}$$

Combining equations (12) and (13), we obtain the following estimate of the mutual information between the parameter  $\theta$  and the output  $x_q$ ,

$$I(\Theta, X_q) \approx \frac{1}{N_1} \sum_{i=1}^{N_1} \left[ \log(p(x_q^{(i)}|\theta^{(i)})) - \log\left(\frac{1}{N_2} \sum_{j=N_1+1}^{N_1+N_2} p(x_q^{(i)}|\theta^{(j)})\right) \right]. \tag{14}$$

Similarly, we can estimate the mutual information between any single component  $\theta_c$  of the  $d$ -dimensional parameter vector,  $\Theta$ , and the output  $X_q$ . We denote by  $\Theta_c$  the  $d-1$ -dimensional vector containing all the components of  $\Theta$  except the  $c$ -th one. Integrating over  $\theta_c$ , the mutual information between  $\Theta_c$  and  $X_q$  is equal to

$$\begin{aligned} I(\Theta_c, X_q) &= \iint p(\theta_c, x_q) \log \frac{p(\theta_c, x_q)}{p(\theta_c)p(x_q)} d\theta_c dx_q \\ &= \iint p(\theta, x_q) \log \frac{p(\theta_c, x_q)}{p(\theta_c)p(x_q)} d\theta dx_q \\ &= \iint p(\theta)p(x_q|\theta) \log \frac{p(x_q|\theta_c)}{p(x_q)} d\theta dx_q \end{aligned}$$

and can be estimated through Monte Carlo simulation by

$$I(\Theta_c, X_q) \approx \frac{1}{N_1} \sum_{i=1}^{N_1} \log \frac{p(x_q^{(i)}|\theta_c^{(i)})}{p(x_q^{(i)})} \tag{15}$$

As previously the evidence  $p(x_q^{(i)})$  can be estimated from a sample drawn from the prior without great difficulty. On the other hand, the estimation of the numerator  $p(x_q^{(i)}|\theta_c^{(i)})$  requires an additional integration over  $\theta_c$ . We then have

$$p(x_q^{(i)}|\theta_c^{(i)}) = \int p(x_q^{(i)}|\theta_c^{(i)}, \theta_c)p(\theta_c|\theta_c^{(i)})d\theta_c \approx \frac{1}{N_3} \sum_{j=1}^{N_3} p(x_q^{(i)}|\theta_c^{(i,j)}) \tag{16}$$

where for each  $1 \leq i \leq N_1$ ,  $\{\theta_c^{(i,j)}\}_{1 \leq j \leq N_3}$  is a sample drawn from  $p(\theta)$  under the constraint that  $\theta_c^{(i,j)} = \theta_c^{(i)}$ . Putting all the terms together, we obtain the following estimation of the mutual

information between  $\Theta_c$  and  $X_q$ : given a sample  $\{\theta^{(i)}\}_{1 \leq i \leq N_1+N_2}$  drawn from  $p(\theta)$  and a  $N_1 \times N_3$ -sample  $\{\theta^{(i,j)}\}_{1 \leq i \leq N_1, 1 \leq j \leq N_3}$  such that for all  $i, j$ ,  $\theta^{(i,j)}$  is drawn from  $p(\theta|\theta_c^{(i)})$ ,

$$I(\Theta_c, X_q) \approx \frac{1}{N_1} \sum_{i=1}^{N_1} \left[ \log \left( \frac{1}{N_3} \sum_{j=1}^{N_3} p(x_q^{(j)}|\theta^{(i,j)}) \right) - \log \left( \frac{1}{N_2} \sum_{k=N_1+1}^{N_1+N_2} p(x_q^{(k)}|\theta^{(k)}) \right) \right]. \quad (17)$$

To finish we consider the estimation of the mutual information between the output of the system for two different experiments  $q$  and  $q^*$ . We have

$$\begin{aligned} I(X_q, X_{q^*}) &= \iint p(x_q, x_{q^*}) \log \frac{p(x_q, x_{q^*})}{p(x_q)p(x_{q^*})} dx_q dx_{q^*} \\ &= \iiint p(x_q|\theta)p(x_{q^*}|\theta)p(\theta) (\log p(x_q, x_{q^*}) - \log p(x_q) - \log p(x_{q^*})) d\theta dx_q dx_{q^*} \end{aligned}$$

where we used the fact that  $X_q$  and  $X_{q^*}$  are independent conditionally to a parameter  $\theta$ . This equation leads to a Monte Carlo estimation from a  $N$ -sample  $\{\theta^{(i)}\}_{1 \leq i \leq N}$  drawn from the prior given by

$$\begin{aligned} I(X_q, X_{q^*}) \approx & \frac{1}{N_1} \sum_{i=1}^{N_1} \left[ \log \left( \frac{1}{N_2} \sum_{j=N_1+1}^{N_1+N_2} p(x_q^{(j)}|\theta^{(i)}) p(x_{q^*}^{(j)}|\theta^{(i)}) \right) \right. \\ & - \log \left( \frac{1}{N_3} \sum_{k=N_1+N_2+1}^{N_1+N_2+N_3} p(x_q^{(k)}|\theta^{(k)}) \right) \\ & \left. - \log \left( \frac{1}{N_4} \sum_{l=N_1+N_2+N_3+1}^N p(x_{q^*}^{(l)}|\theta^{(l)}) \right) \right] \end{aligned}$$

where  $N_1 + N_2 + N_3 + N_4 = N$ .

**Technical details.** We implemented the algorithms in *Python*. The numerical solutions of the models were obtained using the solver for ordinary differential equations of the package *cuda-sim* [54], which allows for parallel implementation on graphical processing units (GPUs). The algorithm to obtain the Monte Carlo estimates was also parallelized using GPUs. For the repressilator, we assume measurement noise  $\epsilon$  with variance  $\sigma^2 = 5$ . To estimate the mutual information between the output and the parameter, we use  $N_1 = 100000$  and  $N_2 = 4500000$ . The mutual information between the output of the system and each parameter in the Hes1 example is computed using  $\sigma^2 = 0.5$ , and  $N_1 = N_2 = N_3 = 5000$ . For the AKT model, the variance of the measurement noise is equal to 10,  $N_1 = 1000$ , and  $N_2 = N_3 = N_4 = 4500$ .

### Approximate Bayesian Computation (ABC)

Once data  $x^*$  have been collected, we use an Approximate Bayesian Computation framework [55,56] to infer the posterior parameter distribution  $p(\theta|x^*)$ . This is a simulation-based method which mainly consists in sampling the parameter space from a prior distribution  $p(\theta)$ , simulating the system for each sampled parameter (often called particle) and selecting the particles such that the simulated data are less than some maximal distance away from the observed data. Those particles define an estimate of the

posterior distribution given the observed data:

$$p(\theta|x^*) \approx \frac{p(\theta) \int_x p(x|\theta) \mathbb{1}_{\Delta(x, x^*) \leq \delta} dx}{p(x^*)}$$

Specifically we use an ABC scheme based on sequential Monte Carlo, which has been developed for likelihood-free parameter inference in deterministic and stochastic systems [41]. We use the implementation of this method of the package called *ABC-SysBio* [57].

### Estimation of the entropy

The estimation of entropy has been performed only to test and confirm our experimental choice, which is based on Monte Carlo estimation of mutual information. For each experiment  $q$  we compute the difference between the entropy  $H(\Theta)$  of the prior distribution  $p(\theta)$  and the entropy  $H(\Theta|x_q)$  of the posterior distribution  $p(\theta|x_q)$ . The entropies  $H(\Theta)$  and  $H(\Theta|x_q)$  are approximated using a histogram-based estimators. This discretization of the parameter space leads to a change of scale in the entropy measure. This explains why the scales of the differences between estimated entropies and the estimated mutual information differ despite the fact that the mutual information  $I(\Theta, X_q)$  is the expectation over all possible data  $x_q$  of the difference between  $H(\Theta)$  and  $H(\Theta|x_q)$ . It is also well known that such histogram approach leads to a biased estimate of the entropy [58]. However, since the bias only depends on the number of bins and the sample size, we can compare the estimation results among experiments as long as these algorithm parameters are kept the same.

**Technical details.** To compute the entropy  $H(\Theta|x_q)$  for each experiment  $q$  in the repressilator example we compute a 4-dimensional histogram to discretise the posterior distribution (for all 4 model parameters) using 100 bins for each dimension resulting in a total of  $10^8$  bins. We use the *R* package *entropy* [59] to estimate the entropy. For the Hes1 model we computed histograms over the marginals posterior distribution, to measure the entropy of each parameter separately. Here we used 1000 bins.

### Experimental data

The experimental data sets used to investigate the Akt model were collected and published by the lab of S. Kuroda. The data are normalised Western blot measurements as described in [45].

### Supporting Information

**Figure S1** Information content of different parameter regimes. The mutual information  $I(x, \theta)$  depends on the dynamics of the system given the prior of the system parameters: the more the dynamics for different parameter values differ from each other, the higher is the information content. To visualize this we compute the mutual information between one parameter and the outcome of the system for three different regimes in the repressilator example. Noting that  $\alpha_0$  is a bifurcation parameter and  $\alpha_0 = 2.55$  is a Hopf bifurcation point, we choose 3 different prior regimes for  $\alpha_0$ :  $\pi_1(\alpha_0) \sim U(0.5, 1.5)$ ,  $\pi_2(\alpha_0) \sim U(3, 4)$  and  $\pi_3(\alpha_0) \sim U(10, 11)$ . We keep the remaining 3 parameters constant:  $h = 2$ ,  $\alpha = 1000$  and  $\beta = 5$ . For these three priors we estimate the mutual information  $I(x, \alpha_0)$  and represent the dynamics of the output of the system. We observe that the dynamics resulting from the first prior regime are most diverse and therefore  $I(x, \alpha_0)$  has the highest value ( $118.78 \pm 2.2710^{-3}$ ) compared to the remaining two parameter regimes ( $113.48 \pm 2.4210^{-3}$  for  $\pi_2(\alpha_0)$  and  $113.41 \pm 4.9210^{-4}$  for

$\pi_3(\alpha_0)$ ). Shown is the bifurcation diagram for parameter  $\alpha_0$  with its stable (solid lines) and unstable (dashed lines) states. Estimation of mutual information was performed for 3 different parameter regimes. For illustration we plot the mean (dark blue), 25 and 75 percentiles (blue) and the 5 and 95 percentiles (light blue) of trajectories simulated with 10000 parameter sets, where  $\alpha_0$  is uniformly sampled and the remaining parameters are kept constant ( $h=2$ ,  $\alpha=1000$  and  $\beta=5$ ).

(TIFF)

**Figure S2** The simulated evolution of the mRNA and protein concentration in the repressilator model for each experimental setup. The parameter vector used for simulations is  $(h^*, \alpha^*, \alpha_0^*, \beta^*) = (2, 10, 1000, 5)$ . The colours correspond to those in Figure 2. The dots represent the simulated data and the lines correspond to the mean of the species for 1000 parameters sampled from the posterior distribution computed using ABC SMC.

(TIFF)

**Figure S3** Mutual information between the parameter and each species (3 mRNA and 3 protein measurements) for experiment 5 in the repressilator model.

(TIFF)

**Figure S4** The posterior distribution given the data represented in Figure S2. Each subfigure (A to E) corresponds to an experiment (1 to 5). In each subfigure, the diagonal represents the marginal posterior distribution for each parameter and the off-diagonal elements show the correlations between pairs of parameters.

(TIFF)

**Figure S5** Simulated trajectories of the mRNA and protein concentrations (dots). The parameter used for simulation is  $(P_0, h, v, k_1) = (1.4, 5.7, 0.02, 0.09)$ . The lines represent the 5% and 95% percentiles of the species abundances for 1000 parameters sampled from the posterior distribution computed using ABC SMC.

(TIFF)

**Figure S6** (A) Simulated trajectories of the mRNA and protein concentration (dots) for the parameter  $(P_0, h, v, k_1) = (1.25, 7.54, 5.36 \times 10^{-2}, 4.02 \times 10^{-3})$ . The lines represent the 5% and 95% percentiles of the species abundances for 1000 parameters sampled from the posterior distribution computed using ABC SMC. (B) Estimates of the differences between the entropies of the prior and posteriors.

(TIFF)

**Figure S7** Ordinary differential equations which describe the dynamics of the 11 species of the AKT model. The model contains 12 parameters denoted  $p_i$ ,  $1 \leq i \leq 12$ . The concentration of the following species (in this order) are denoted by  $y_i$ ,  $0 \leq i \leq 10$ : EGF,

EGFR, pEGFR, pEGFR-AKT, AKT, pAKT, S6, pAKT-S6, pS6, pro-EGFR and EGF-EGFR.

(TIFF)

**Figure S8** The time course of phosphorylated EGF receptor (pEGFR), phosphorylated Akt (pAKT) and phosphorylated S6 (pS6) in response to an impulse input of EGF over 60 seconds with an intensity of 0.1 ng/ml (dots). Data are Western blots measurements, described in [45]. The lines represent the 5% and 95% percentiles of the evolution of the species for 1000 parameters sampled from the posterior distribution computed using ABC SMC.

(TIFF)

**Figure S9** The time course of phosphorylated EGF receptor (pEGFR), phosphorylated Akt (pAKT) and phosphorylated S6 (pS6) in response to a step input of EGF over 60 minutes with an intensity of 3.0 ng/ml (dots). Data are Western blots measurements, which have been generated and published by [45]. The lines represent the 5% and 95% percentiles of the evolution of the species for 1000 parameters sampled from the posterior distribution computed using ABC SMC.

(TIFF)

**Figure S10** (A) A noisy 5-impulses EGF input signal: the 5 pulses are of intensity 1 ng/ml and length 60 seconds spaced by 400 seconds with an additive background noise which is the absolute value of a gaussian white noise of variance  $\tau^2 = 1$ . (B) The mutual information between the time course of the 3 species of interest under the noisy input signal represented in (A) and the time course of the species under each of the following 12 possible experiments: an impulse stimulus of length 60 seconds with 5 possible intensity (0.3, 1, 3, 10 and 30 ng/ml), a step stimulus of length 60 minutes with 4 possible intensity (0.1, 0.3, 1 and 3 ng/ml) and a ramp stimulus of length 60 minutes with 3 possible final intensity (0.03, 0.3 and 3 ng/ml).

(TIFF)

**Figure S11** The predicted time course of the proteins pEGFR, pAKT and pS6 under the noisy 5-impulses EGF input signal with a noise of high intensity represented Figure S10 A. In the left panel, the prediction is based on the initial dataset whereas in the right panel in addition to the initial data it is also based on the outcome of the step stimulus with intensity 3 ng/ml, which is the experiment with the highest mutual information. The scale of the y-axis is different for each figure.

(TIFF)

## Author Contributions

Conceived and designed the experiments: JL SF MK MPHS. Performed the experiments: JL SF. Analyzed the data: JL SF. Contributed reagents/materials/analysis tools: JL SF MK. Wrote the paper: JL SF MK MPHS.

## References

- Bruggeman F, Westerhoff H (2007) The nature of systems biology. *TRENDS in Microbiology* 15: 45–50.
- Nurse P, Hayles J (2011) The cell in an era of systems biology. *Cell* 144: 850–854.
- Silver P, Way J (2007) Molecular systems biology in drug development. *Clin Pharmacol Ther* 82: 586–590.
- Spiller D, Wood C, Rand D, White M (2010) Measurement of single-cell dynamics. *Nature* 465: 736–745.
- Del Sol A, Balling R, Hood L, Galas D (2010) Diseases as network perturbations. *Current Opinion in Biotechnology* 21: 566–571.
- Liepe J, Taylor H, Barnes CP, Huvet M, Bugeon L, et al. (2012) Calibrating spatio-temporal models of leukocyte dynamics against in vivo live-imaging data using approximate Bayesian computation. *Integrative biology* 4: 335–345.
- Brenner S (2010) Sequences and consequences. *Phil Trans Biol Sci* 365: 207–212.
- Beal M, Falciani F, Ghahramani Z, Rangel C, Wild D (2005) A bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics* 21: 349–356.
- Sachs K, Perez O, Pe'er D, Lauffenburger D, Nolan G (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* 308: 523–529.
- Lèbre S, Becq J, Devaux F, Stumpf MPH, Lelandais G (2010) Statistical inference of the time-varying structure of gene-regulation networks. *BMC Systems Biology* 4: 130.
- Mendes P, Kell D (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14: 869–883.
- Reinker S, Altman RM, Timmer J (2006) Parameter estimation in stochastic biochemical reactions. *IEE Proc Syst Biol* 153: 168–178.
- Kreutz C, Timmer J (2009) Systems biology: experimental design. *FEBS Journal* 276: 923–942.

14. Vyshemirsky V, Girolami M (2008) Bayesian ranking of biochemical system models. *Bioinformatics* 24: 833.
15. Toni T, Stumpf MPH (2010) Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* 26: 104–110.
16. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, et al. (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol* 3: 1871–1878.
17. Erguler K, Stumpf M (2011) Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol BioSyst* 7: 1593–1602.
18. Wilkinson D (2009) Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics* 10: 122–133.
19. Cox D (2006) *Principles of Statistical Inference*. Cambridge: Cambridge University Press.
20. Lindley DV (1956) On a measure of the information provided by an experiment. *Ann Math Statistics* 27: 986–1005.
21. Stone M (1959) Application of a measure of information to the design and comparison of regression experiments. *The Annals of Mathematical Statistics* 30: 55–70.
22. DeGroot M (1962) Uncertainty, information, and sequential experiments. *The Annals of Mathematical Statistics* 33: 404–419.
23. DeGroot M (1986) Concepts of information based on utility. In: Daboni L, editor. *Recent Developments in the Foundations of Utility and Risk Theory*. Dordrecht, Reidel: Springer. pp. 265–275.
24. Bernardo J (1979) Expected information as expected utility. *The Annals of Statistics* 7: 686–690.
25. Vanlier J, Tiemann C, Hilbers P, van Riel N (2012) A bayesian approach to targeted experiment design. *Bioinformatics* 28: 1136–42.
26. Huan X, Marzouk Y (2012) Simulation-based optimal bayesian experimental design for nonlinear systems. *Journal of Computational Physics* 232: 288–317.
27. Kutalik Z, Cho KH, Wolkenhauer O (2004) Optimal sampling time selection for parameter estimation in dynamic pathway modeling. *Biosystems* 75: 43–55.
28. Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, et al. (2007) Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. *IET Systems Biology* 1: 190–202.
29. Chu Y, Hahn J (2008) Integrating parameter selection with experimental design under uncertainty for nonlinear dynamic systems. *Aiche Journal* 54: 2310–2320.
30. Bandara S, Schloeder J, Eils R, Bock H, Meyer T (2009) Optimal Experimental Design for Parameter Estimation of a Cell Signaling Model. *PLoS Comp Biol* 5: e1000558.
31. Apgar JF, Witmer DK, White FM, Tidor B (2010) Sloppy models, parameter uncertainty, and the role of experimental design. *Molecular BioSystems* 6: 1890.
32. Vanlier J, Tiemann C, Hilbers P, van Riel N (2012) An integrated strategy for prediction uncertainty analysis. *Bioinformatics* 28: 1130–5.
33. Komorowski M, Costa MJ, Rand DA, Stumpf MPH (2011) Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proc Natl Acad Sci USA* 108: 8645–8650.
34. MacKay DJC (2003) *Information theory, inference and learning algorithms*. Cambridge University Press.
35. Marjoram P, Molitor J, Plagnol V, Tavaré S (2003) Markov chain monte carlo without likelihoods. *Proc Natl Acad Sci U S A* 100: 15324–15328.
36. Sisson SA, Fan Y, Tanaka MM (2007) Sequential monte carlo without likelihoods. *Proc Natl Acad Sci U S A* 106: 16889.
37. Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal* 27: 379–423, 623–656.
38. Chaloner K, Verdine I (1995) Bayesian experimental design: A review. *Statist Sci* 10: 273–304.
39. Clyde MA (2001) Experimental design: A bayesian perspective. In: Smelser, editor. *International Encyclopedia of the Social and Behavioral Sciences*. New York: Elsevier Science.
40. Elowitz M, Leibler S, et al. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335–338.
41. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf M (2009) Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface* 6: 187–202.
42. Hirata H, Yoshiura S, Ohtsuka T, Bessho Y, Harada T, et al. (2002) Oscillatory expression of the bhlh factor hes1 regulated by a negative feedback loop. *Science's STKE* 298: 840.
43. Silk D, Kirk P, Barnes C, Toni T, Rose A, et al. (2011) Designing attractive models via automated identification of chaotic and oscillatory dynamical regimes. *Nature Communications* 2: 489.
44. Bazil JN, Buzzard GT, Rundell AE (2012) A global parallel model based design of experiments method to minimize model output uncertainty. *Bull Math Biol* 74: 688–716.
45. Fujita K, Toyoshima Y, Uda S, Ozaki Y, Kubota H, et al. (2010) Decoupling of receptor and downstream signals in the akt pathway by its low-pass filter characteristics. *Science's STKE* 3: ra56.
46. Toyoshima Y, Kakuda H, Fujita K, Uda S, Kuroda S (2012) Sensitivity control through attenuation of signal transfer efficiency by negative regulation of cellular signalling. *Nature Communications* 3: 743.
47. Hengl S, Kreutz C, Timmer J, Maiwald T (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* 23: 2612–2618.
48. Apgar JF, Toettcher JE, Endy D, White FM, Tidor B (2008) Stimulus design for model selection and validation in cell signaling. *PLoS Comp Biol* 4: e30.
49. Barnes C, Silk D, Sheng X, Stumpf M (2011) Bayesian design of synthetic biological systems. *Proc Natl Acad Sci U S A* 108: 15190–15195.
50. Bazil J, Buzzard G, Rundell A (2012) A global parallel model based design of experiments method to minimize model output uncertainty. *Bulletin of mathematical biology* 74: 1–29.
51. Mélykúti B, August E, Papachristodoulou A, El-Samad H (2010) Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Systems Biology* 4: 38.
52. Sebastiani P, Wynn H (2002) Maximum entropy sampling and optimal bayesian experimental design. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62: 145–157.
53. Berg BA (2004) *Markov Chain Monte Carlo Simulations And Their Statistical Analysis*. World Scientific Publishing.
54. Zhou Y, Liepe J, Sheng X, Stumpf M, Barnes C (2011) Gpu accelerated biochemical network simulation. *Bioinformatics* 27: 874–876.
55. Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* 65: 220–228.
56. Beaumont MA, Zhang W, Balding DJ (2002) Approximate bayesian computation in population genetics. *Genetics* 162: 2025–2035.
57. Liepe J, Barnes C, Cule E, Erguler K, Kirk P, et al. (2010) ABC-SysBio—approximate Bayesian computation in Python with GPU support. *Bioinformatics* 26: 1797–1799.
58. Paninski L (2003) Estimation of entropy and mutual information. *Neural Computation* 15: 1191–1253.
59. Hauser J, Srimmer K (2009) Entropy inference and the james-stein estimator, with application to nonlinear gene association networks. *J Mach Learn Res* 10: 1469–1484.