

Research

Open Access

Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering

Filippo Menolascina*^{1,2,5}, Domenico Bellomo^{3,6,7}, Thomas Maiwald⁴, Vitoantonio Bevilacqua¹, Caterina Ciminelli¹, Angelo Paradiso⁵ and Stefania Tommasi⁵

Address: ¹Department of Electrical Engineering and Electronics, Technical University of Bari, Via E. Orabona 4, 70125, Bari, Italy, ²Systems and Synthetic Biology Laboratory, Telethon Institute of Genetics and Medicine, Via Pietro Castellino 111, 80131, Naples, Italy, ³Information and Communication Theory Group, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, the Netherlands, ⁴Department of Systems Biology, Harvard Medical School, 200 Longwood Ave, Alpert 536, Boston, MA 02115, USA, ⁵Clinical Experimental Oncology Laboratory, National Cancer Institute, Via F. Hahnemann 10, 70126, Bari, Italy, ⁶Bioprocess Technology Section, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, Julianalaan 67, 2628 BC Delft, The Netherlands and ⁷Kluyver Centre for Genomics of Industrial Fermentation, Julianalaan 67, 2628 BC Delft, The Netherlands

E-mail: Filippo Menolascina* - f.menolascina@ieee.org; Domenico Bellomo - d.bellomo@tudelft.nl;

Thomas Maiwald - thomas_maiwald@hms.harvard.edu; Vitoantonio Bevilacqua - bevilacqua@poliba.it;

Caterina Ciminelli - c.ciminelli@poliba.it; Angelo Paradiso - a.paradiso@oncologico.bari.it; Stefania Tommasi - s.tommasi@oncologico.bari.it

*Corresponding author

from Bioinformatics Methods for Biomedical Complex Systems Applications (NETTAB2008)
Varenna, Italy 19–21 May 2008

Published: 15 October 2009

BMC Bioinformatics 2009, 10(Suppl 12):S4 doi: 10.1186/1471-2105-10-S12-S4

This article is available from: <http://www.biomedcentral.com/1471-2105/10/S12/S4>

© 2009 Menolascina et al; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Mechanistic models are becoming more and more popular in Systems Biology; identification and control of models underlying biochemical pathways of interest in oncology is a primary goal in this field. Unfortunately the scarce availability of data still limits our understanding of the intrinsic characteristics of complex pathologies like cancer: acquiring information for a system understanding of complex reaction networks is time consuming and expensive. Stimulus response experiments (SRE) have been used to gain a deeper insight into the details of biochemical mechanisms underlying cell life and functioning. Optimisation of the input time-profile, however, still remains a major area of research due to the complexity of the problem and its relevance for the task of information retrieval in systems biology-related experiments.

Results: We have addressed the problem of quantifying the information associated to an experiment using the Fisher Information Matrix and we have proposed an optimal experimental design strategy based on evolutionary algorithm to cope with the problem of information gathering in Systems Biology. On the basis of the theoretical results obtained in the field of control systems theory, we have studied the dynamical properties of the signals to be used in cell stimulation. The results of this study have been used to develop a microfluidic device for the automation of the process of cell stimulation for system identification.

Conclusion: We have applied the proposed approach to the Epidermal Growth Factor Receptor pathway and we observed that it minimises the amount of parametric uncertainty associated to the identified model. A statistical framework based on Monte-Carlo estimations of the uncertainty ellipsoid confirmed the superiority of optimally designed experiments over canonical inputs. The proposed approach can be easily extended to multiobjective formulations that can also take advantage of identifiability analysis. Moreover, the availability of fully automated microfluidic platforms explicitly developed for the task of biochemical model identification will hopefully reduce the effects of the 'data rich-data poor' paradox in Systems Biology.

Background

Our understanding of molecular basis of complex diseases is being dramatically changed by systems investigation supported by the most advanced tools and techniques developed by the scientific community. In particular, cancer investigation has greatly benefited by systems level approaches since tumor development and progression are believed to be among those system trajectories that arise from abnormal working states. The work by Hornberg and colleagues [1] pointed out the relevance of Systems Biology approaches in the study of dynamics leading to cancer. Epidermal Growth Factor Receptor (EGFR) pathway is one of those biochemical reaction networks believed to play a central role in cancer development. As a matter of fact EGFR and receptors in the same family (ErbB2, ErbB3 and ErbB4) mediate cell to cell interactions both in organogenesis and in adult tissues [2]. The 40-year long study of this pathway led to associate overexpression of the EGFR family members to several types of cancer [3]. Because of the high clinical relevance, several efforts have been spent in the last decades in unravelling the complex dynamics of this biochemical network, as well as in finding potential targets of therapeutic intervention [4-6]. Although global models of EGFR pathway exist [7-12], many questions still remain open both in terms of model accuracy [13-15], parameter identifiability [16] and driving input design [17,18]. In this context we put the pioneering works by Arkin and colleagues [19-22], van Oudenaarden and colleagues [23] and Steuer and colleagues [24]. Other recent works have focused on the connections between optimal experimental design strategies and *structural* and *experimental identifiability* analysis of biochemical pathways; this is the case of [16,25-28].

Structural identifiability refers to the possibility of finding the mathematical model of the true system (see [29,30] for references in biological systems investigation), after having applied a specific search strategy in the space of the solutions.

Experimental identifiability [31], on the other hand, is related to the possibility of finding the mathematical

representation of the true model given a predetermined set of observations. This is a central aspect of this class of identifiability problems since it is more focused on the available data and, in particular, on information content. This aspect establishes an interesting bridge between System Identification Theory and Experimental Design. The Design of Experiments (DOE) is a well developed methodology in statistics [32] focusing on the design of all information-gathering exercises where variation is present, the main objective of the whole task being the maximisation of the information obtained from experiment and the minimisation of the number of experiments. This specific task is commonly referred to as 'Optimal Experimental Design' (OED). This discipline quickly gained a significant interest among researchers mostly in natural and social sciences but became an active research field in engineering only with the pioneering work by Lennart Ljung and his standard model for dynamical system identification oriented experiments [33]. This model has been recently modified by Phair et al. [34] and Cho et al. [35].

Nevertheless the main idea behind system identification in Systems Biology remained unchanged [36]. In line with Fisher's criteria, Ljung's scheme [33] suggests to define a detailed plan of the experiments to be carried out before starting to collect input-output data from the system to be modeled. Specifications like data sampling strategies and driving inputs should be fixed in order to optimise the information yield of each experiment and to address the cost minimisation task OED is aimed at. These issues gain an even stronger relevance, if we consider the so-called 'data rich-data poor paradox' [37] resulting from the difficulties and costs involved in Systems Biology related assays. For these reasons and in order to develop a comprehensive framework for system identification in Systems Biology, we will describe how a specific issue of OED, namely Optimal Input Design (OID), can be addressed using optimality criteria and microfluidics-based experiments. As a matter of fact, microfluidic platforms have been shown to provide a powerful tool for the development of data-rich experimental strategies able to fill the gap of the previously cited paradox. Signals obtained in this stage are used as

templates for the development of a microfluidic device for a flexible and automated platform for affordable single-cell experiments in Systems Biology.

In the following paragraphs, we go through a brief introduction of the EGFR model then we analyse OED and OID criteria. We review current approaches to cell stimulation in the 'Methods' section and compare them with optimality criteria derived ones. An analysis of the experimental results follows in 'Results', where we introduce a feasible design of the microfluidic device thought to speed up the process of data collection in Systems Biology by lowering the costs associated to experiments. A discussion of the results presented herein and final cues for further research are given in the last paragraph.

Results and discussion

In order to model and understand the functionality of the EGFR signaling cascade a quantitative description of the signal dynamics is of major relevance. For this reason, we discuss the computational results obtained from *in-silico* simulations carried out using POTTERS WHEEL [38]. POTTERS WHEEL is a multi-experimental fitting MATLAB package intended to allow researchers to ease model analysis and experimental setup steps. In particular, this package is one of the few in Systems Biology providing a simple interface to external input based simulation of biological pathways behavior.

To estimate the effects of different inputs on parameters estimates uncertainties, we carried out 1000 identification experiments for each of the three classes of stimuli, namely: a step input, a persistently exciting input and the time profile of the stimulation obtained from the optimisation task. Therefore we plotted a bivariate distribution of both the V_{max} and K_m of the first Michaelis-Menten based reaction ($v_0 = \frac{V_{max}[S]}{K_m + [S]}$ where v_0 is the initial reaction rate and $[S]$ the substrate concentration, see 'Methods' for a detailed description of the mathematical modelling step) in Kholodenko's model accounting for the dephosphorylation of the EGF-EGFR dimer. It should be noted that parameter correlation can greatly affect our ability to successfully recover real parameters. This is one of the main issues arising in the field of parameter identifiability. In particular parameters that are structurally correlated cannot be uniquely identified from experiments. In order to investigate such peculiarities of our dynamical model we carried out an identifiability analysis based on the 'Alternating Conditional Expectation' algorithm (ACE) method described in [16] and implemented in the *Mean Optimal Transformation Approach* (MOTA) package. We

present herein the computational results so as to provide a tool for comparing the different approaches to OED in Systems Biology.

Identifiability analysis

The statistical investigation of the properties of OED should be a primary goal when the time profile of the input is computed. Previous works in this field have focused on the comparison among alternative designs on the FIM cost values and confidence intervals [39]. Nevertheless it has been noticed that the FIM is derived from a linearisation of the least squares thus it may be unreliable in cases of considerably extended non linearities. The non identifiability of one of the parameters directly implies the functional relationships among at least two of them [16]. This phenomenon can be easily observed by plotting the joint probability distribution of each of them which will show statistically significant differences if compared with the expected multivariate normal one. From an algebraic point of view this results in the loss of rank of the covariance matrix of the multivariate distribution or, alternatively, a condition number of the same matrix asymptotically tending to infinity. Then, in order to cope with identifiability issues that may arise in identification tasks, we propose to investigate this property by using the ACE method proposed by Hengl and colleagues [16]. MOTA package is included in POTTERS WHEEL toolbox and is applied together with a linear fit sequence analysis. MOTA detects groups of two or more linearly or non-linearly related parameters. It revealed some major non-identifiabilities in the parameter space (results not shown) whose nature should be certainly investigated in order to understand their causes and possible solutions. We should remark that an integrated approach using Monte-Carlo methods for both experimental design optimisation and parameter correlation investigation can be a feasible choice. However, we should consider that this would imply a major rise in computational costs of the approach resulting from the high number of parameter estimation tasks to be accomplished. Not to mention the issues arising from large scale models, noise and potential multimodality that would certainly imply using a robust global optimisation algorithm.

Monte-Carlo based analysis

Several alternative choices to dynamic optimisation methods have been used in the context of OED, the most widely employed are direct methods such like complete parameterisation, control vector parameterisation and multiple shooting approach [39]. These methods are based on the transformation of the original infinite dimension optimisation problem into a non linear programming problem through the discretisation of the state and the stimuli or only the stimuli and the approximation of the time dependencies using locally

defined function. In this work, we employed a slightly different approach with similar discretisation strategies but based on an Genetic Algorithm (GA); for this purpose we used the implementation provided by *MATLAB* through the GA routine. It should be noted that no formal proof of the convergence of GAs can be derived for the problem at hand. However, some property of this kind of approaches, just like computational complexity and efficiency can be studied in a more formal context [40]. We set the population to be composed by 200 individuals and we used the tournament system as selection criterion; crossover and mutation operators were set to 'uniform' and 'heuristic' respectively. All the other options were left at the default values while constraints on signal amplitude reflecting technological limitations were coded in the appropriate arrays (namely A and b for which $Ax \leq b$ should be satisfied). As previously outlined the fitness function of this GA encodes for the FIM associated to the specific experimental design under investigation. In order to investigate if any statistically significant difference existed between parametric uncertainties estimated from classical and OED based experiments, we developed a Monte-Carlo based analysis with $N = 1000$ repetitions. We collected parameter estimates for each identification experiment. At this stage we carried out an intermediate analysis to find the parameters with the highest relative uncertainty; we selected the highest two, namely V_4 and K_4 . A χ^2 goodness-of-fit test confirmed that the probability density function for these parameters can be well approximated by a normal curve. In order to compare the three experimental design selected we performed a Gaussian Mixture Model (GMM) fitting of the identified parameters starting from a bivariate distribution arising from K_4 and V_4 variables. Estimates were normalised and then plotted. Figure 1 shows the distribution of parameters estimates couples from sustained input experiments; Fig. 2 and 3 show the same plot for persistently exciting and OED based experiments, respectively. The plots show the 95% confidence interval of each distribution computed as the ellipsoid centered in the mean of the bivariate distribution and having:

- major semiaxis equal to $\lambda_{max}(Cov)$
- minor semiaxis equal to $\lambda_{min}(Cov)$
- rotational offset with respect to the the x axis equal to $arctan\left(\frac{y_{eigv}}{x_{eigv}}\right)$

where Cov is the covariance matrix estimated from GMM fitting and $\lambda_{max}/\lambda_{min}$ its max and min eigenvalues

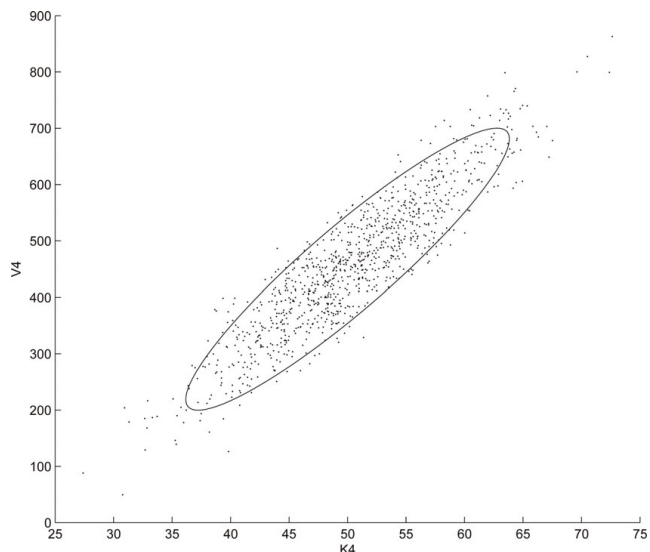


Figure 1
Scatter plot of the step input-based experiment estimations. The 95% confidence intervals for the parameters V_4 (on the y axis) and K_4 (on the x axis) in the case of estimation based on the step input driven system. Mean vector and covariance matrix are fitted on the data in order to obtain the best bivariate gaussian distribution approximating data from *in-silico* experiments.

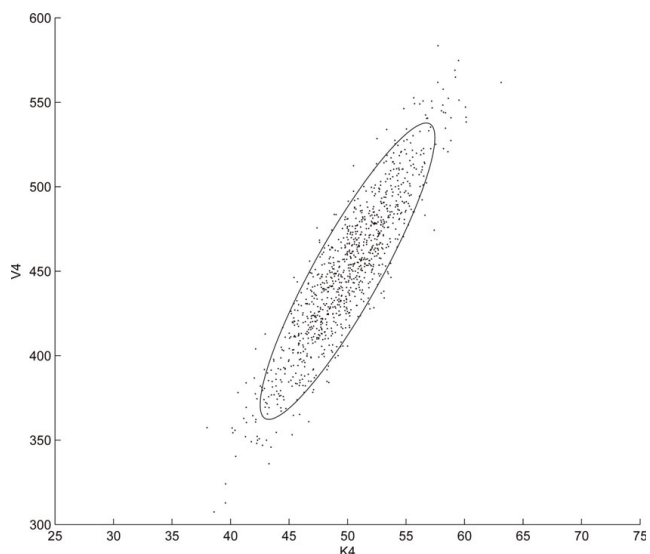


Figure 2
Scatter plot of the PE input-based experiment estimations. The 95% confidence intervals for the parameters V_4 and K_4 in the case of estimation based on persistently exciting input driven system. Mean vector and covariance matrix are fitted on the data in order to obtain the best bivariate gaussian distribution approximating data from *in-silico* experiments.

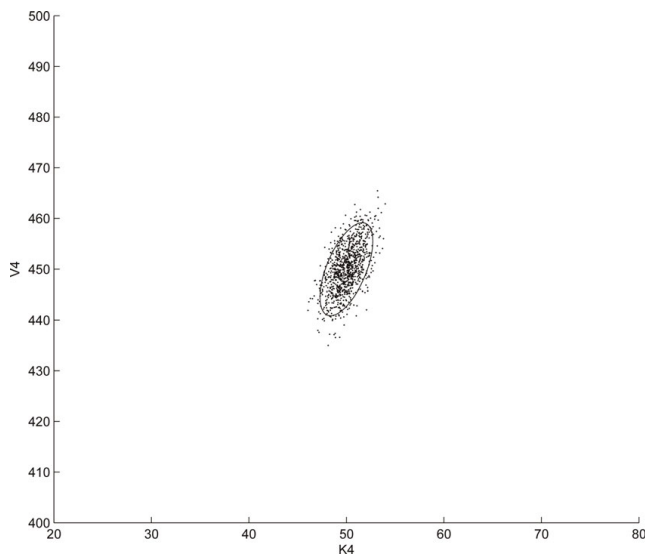


Figure 3
Scatter plot of the optimal input-based experiment estimations. The 95% confidence intervals for the parameters V_4 and K_4 in the case of estimation based on optimal input driven system. Mean vector and covariance matrix are fitted on the data in order to obtain the best bivariate gaussian distribution approximating data from *in-silico* experiments.

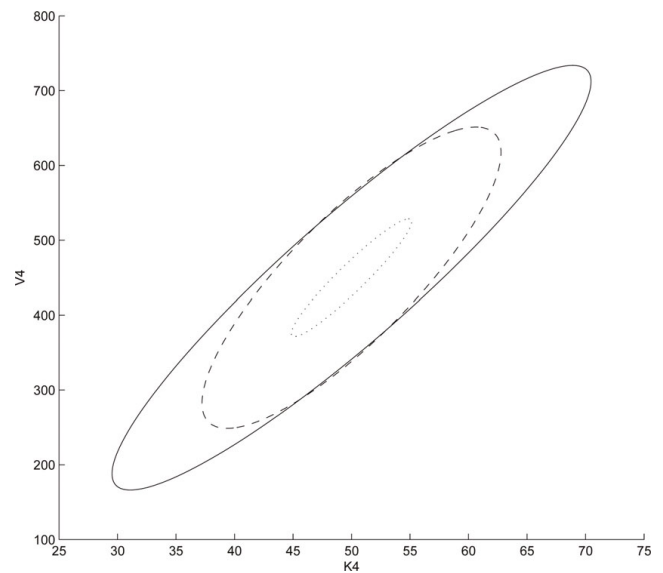


Figure 4
Comparison of the 95% confidence intervals. The three 95% confidence intervals compared; continuous line (step input), dashed line (persistently exciting input) and dotted line (optimal input). A comparison of the boundaries and positions of the ellipsoids puts in evidence that OID-based experiments are characterised by the lowest uncertainty (smallest ellipsoid area) and therefore provide the greatest amount of information on the model.

respectively, while γ_{eigv} and x_{eigv} are the γ and x component of the eigenvectors of Cov matrix. Figure 4 shows the 95% ellipsoids of the three experimental designs compared. It is evident that the volume of the uncertainty ellipsoids gets minimised by more appropriate designs. Moreover the OID based strategy proves to be the one providing the best experimental conditions for accurate parameter estimation and system identification. In order to obtain a more quantitative estimation of the information gain provided by OED based experiments we performed an Ansari-Bradley test [41] on the estimated parameter values; the Ansari-Bradley test checks the hypothesis that two independent samples come from the same distribution, against the alternative that they come from distributions that have the same median and shape but different variances. Pairwise tests carried out on OID vs PE and OID vs step experiments returned p-values smaller than 0.01 thus supporting the rejection of the null hypothesis and then suggesting evidence of statistically meaningful advantages of OED based experiments over both PE and step input based ones.

Conclusion

The intrinsic quantitative nature of Systems Biology poses new issues in everyday laboratory practice. Modelling, in this context, has long suffered from data shortage; the

'data rich-data poor' paradox greatly influenced the pace towards a comprehensive understanding of molecular mechanisms governing biological systems. Nevertheless the potential of novel experimental techniques seems to promise new groundbreaking innovations thus increasing the versatility of new laboratory protocols and keeping experiment-associated costs low. Among these major limitations we should certainly mention the ability to stimulate cells in chemostats with input having very limited harmonic content. Microfluidic technology currently allows us to go beyond step like stimulation and to generate complex time-varying signals whose modulation can be achieved using control engineering strategies [42]. The availability of such tools and devices will allow us overcome the limits of indicial response for highly complex and fed-back dynamical systems identification as outlined in [43,44]. In this framework, the ability to optimally control and take advantage of the new methods and devices will be a major focus of the scientific community. This contribution presents, then, a mathematical formulation of the problem of optimal experimental design in Systems Biology by considering a case study of one of the most relevant biological pathways for cancer development. Formal derivation of problem definition results and heuristic solutions to a

highly non-linear optimisation problem have been both provided. In particular we formulated the problem of OED in Systems Biology as a non-linear optimisation task in which the amount of information per experiment, quantified in the Fisher Information Matrix, is optimised by varying the stimulus time profile here representing the concentration of extracellular EGF ligand. We set up an evolutionary optimisation task aimed at finding the time sequence of the input signal that maximises the amount of information associated to the experiment. Moreover we proposed a statistical framework based on Monte-Carlo estimates for the computation of the uncertainty regions for the parameter values; identifiability analysis, on the other hand, has been carried out using the ACE approach integrated in POTTERS WHEEL package. The results shown clearly indicate that dynamic experiments outperform canonical experiments based on sustained or persistently exciting inputs. Nonetheless we should consider that the approach presented herein depends on the starting model; a sequential experimental design should be investigated in order to overcome this issue. Moreover we should consider that all the simulations reported should be validated in a series of experiments. For this reason, we proposed the microfluidic device described in the 'Methods' section. 'Labrys' goes beyond the specific context of EGFR and, associated to a Hardware Abstraction Layer like Biosteram [45], is thought to provide researchers with a fully automated platform for complex experiment development and implementation. Future work in this field will certainly require a more tight collaboration among the different competences in the field of Systems Biology aimed at the full integration of both hardware and software findings for the development of a common, powerful and versatile platform for *systems* oriented experiments.

Methods

Model definition

In this work, we consider the EGFR signaling network model proposed by Kholodenko and colleagues in [7]. This model explores the short term pattern of cellular responses to epidermal growth factor (EGF) in isolated hepatocytes and predicts how the cellular response is controlled by the relative levels and activity states of signaling proteins and under what conditions activation patterns are transient or sustained. BioModels database [46] provides a selection of the most common file formats of this model. For our purposes, we will use the SBML version [47] featuring 25 molecular species, 23 reactions and 50 parameters. The set of ODEs describing this model can be extracted using COPASI [48] and translated in a MATLAB SIMULINK® model for further simulations and analyses. Both mass-action and

Michaelis-Menten kinetics have been used by Kholodenko and colleagues, resulting in a non-linear model. In order to elicit this pathway with a driving input, we slightly modified it so as to include an external control for the ligand species, i.e. the Epidermal Growth Factor. For further experiments we selected POTTERS WHEEL [38] platform; this software provides highly powerful tools for the investigation of biological models' properties (just like parameter fitting and identifiability analysis) and, to the best of our knowledge, is one of the few ones in Systems Biology allowing researchers to easily define time evolution of forcing inputs without using complex formulations based on events and rules provided by SBML specifications.

We imported Kholodenko's model in POTTERS WHEEL and we edited the m-file so as to force EGF to be an input for our system and downstream species as our observables or outputs. This is a Single Input-Multi Output (SIMO) formulation of the EGFR and will prove to be an interesting model for the stimulation of interesting behaviors (e.g. transmission blocking zeros elicitation which is a counter-intuitive behavior of some dynamical systems which show a null output even if they are stimulated by non-trivial inputs with specific harmonic content). For our purpose, however, we will select only one downstream species to be observed just like in Single Input-Single Output systems whose study drove the field of dynamical system identification.

Dynamical systems identification

Biological systems, just like any other dynamical system, can be described by a number of mathematical tools being ODEs an easy way to investigate systems' properties and Stochastic Differential Equations (SDE) more appropriate when small copy numbers of molecules drive system's dynamics. We will focus on the former representation approach since it's more appropriate in case of cell population-based studies (the most common ones in current practice and in many of current microfluidics-based experiments). In ODEs based models we distinguish state variables \bar{x} , inputs \bar{u} and outputs \bar{y} ; from now on bold notation will be used in place of the vector one for readability sake. We define time evolution of these entities by adding to them time dependence, obtaining $x_i(t)$ $i = 1, \dots, n$, $u_j(t)$ $j = 1, \dots, m$ and $y_k(t)$ $k = 1, \dots, p$ respectively. We can use a set of ODEs to represent the state change through time:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \boldsymbol{\theta}, \mathbf{u}(t)) \quad (1)$$

Time evolution of the system state $\mathbf{x}(t) \in \mathbb{R}^n$ can be easily derived by solving the system of differential equations in Eq. 1 and imposing constraints on initial conditions $\mathbf{x}(0)$. Notice that the rate of change of x_i depends, in

general non-linearly, on state variables $x_j, j = 1, \dots, n$, on input trajectories u_i and on parameters vector θ .

In order to explain this representation choice we will go through a brief example. Consider a simple isomerisation reaction



if we were to model this reaction using mass-action kinetics we would model the time evolution of these species using the following set of ODEs (using A for the concentration $[A]$)

$$\frac{dA}{dt} = -k_1 \cdot A \tag{2}$$

$$\frac{dB}{dt} = k_1 \cdot A \tag{3}$$

Notice that, in order to solve these ODEs, at least two quantities are required: k_1 and the initial concentration of A species, $A(0)$. The parameters vector θ is usually intended to collect these quantities.

Accurate identification of parameters governing the dynamics of biochemical reaction networks is currently considered a major challenge in Systems Biology. In fact, even though some control on initial concentration of species can be obtained with experimental protocols (e.g. starvation), rate coefficients are driven by several external factors (temperature, PH etc.) in a very complex way. Moreover accurate parameters estimation is a key step for the elicitation of interesting behaviors in cellular pathways [49].

Unfortunately the number of observable species is usually lower than the experimenter would want. For this reason we define the $\gamma^M(t, \theta) \in \mathbb{R}^p$ as the vector of measurable molecular species in an assigned experiment and we write

$$\gamma^M(t_i, \theta, \mathbf{u}) = \mathbf{g}(\mathbf{x}(t_i, \theta, \mathbf{u})) \quad i = 1, \dots, n \tag{4}$$

The observations $\gamma^O(t_i) \in \mathbb{R}^p$ are then given by

$$\gamma^O(t_i) = \gamma^M(t_i, \theta_0, \mathbf{u}) + \varepsilon_i \quad i = 1, \dots, n \tag{5}$$

with

$$\varepsilon_{ij} = N(0, \sigma_{ij}^2) \quad i = 1, \dots, n \quad j = 1, \dots, p \tag{6}$$

where we denoted the true parameter vector with θ_0 . Here $\varepsilon_i \in \mathbb{R}^p$ describes the gaussian component of the error at time t_i . We notice that the observation function

$\mathbf{g}(\cdot)$ together with the input function $\mathbf{u}(\cdot)$ and the set of sampling times t fully defines the *experimental design*. In this we glimpse the triple nature of OED which aims at establishing optimal strategies for (a) sampling time [50], (b) species to be measured [17] and (c) input selection [18].

Optimal experimental design in systems biology

As previously stated OED has its main objective in maximising the information yield returned by an experiment. This is a central aspect in everyday practice in Systems Biology since experiments can be both highly expensive and time-consuming, limiting practical feasibility of otherwise promising protocols. Applications of OED in Systems Biology have been described in [17,51-55]; in particular [54,56-60] have focused on model discrimination by OED. Experimental designs are usually categorised as *starting* and *sequential designs*.

In *starting designs* no data have been previously collected and the experimenter is interested in drawing the maximum amount of information from the experiment to be planned. This is done by minimising (or maximising) a specified objective function. Within this category we identify two subcategories: *exact* and *continuous designs*.

Exact designs have their own objective in the optimal placement of a finite number of design points [61]; on the other hand *continuous designs* deal with the selection of a design measure, η , which is equivalent to a probability density over the design space.

Sequential designs try to develop optimal strategies for model refinement of a pre-existing model [62]. In this paper we will focus on a semi-sequential approach that can be considered a *sequential design* in that it starts from a compiled dynamic model of the EGFR pathway, but we do not use the results of this design to carry out further identification experiments since this would require non standard technological platforms. Given the potential impact of OED strategies on Systems Biology research, some researchers proposed software packages providing the user with significant opportunities for optimal experiments planning [63-65]. All of these packages are built on the principles of optimality and based on metrics being defined on the Fisher Information Matrix. This is a quite general framework for OED; unfortunately none of them currently provides a solution for Optimal Input Design. We will analyse this and other issues related to biochemical pathway stimulation in the next sections.

Optimal input design

Optimal Input Design for system identification provides several alternative measures of the identified model

being used for optimal design [66]. Here we start by reporting some of the main results in the field [33] and then we discuss their implications in the specific case. Several contributions in this field focused on the minimisation of some measure of the variance of the estimated parameters like Fisher Information Matrix which can be used to estimate variance-uncertainty associated to parameter estimates. This process will be analysed in the next section; in this paragraph we will focus on a theoretical study of the OID for dynamic systems identification. Identification processes start with data collected on Input-Output behavior of the system under investigation.

Let the true system be described as:

$$S: \gamma(s) = \underbrace{G(s, \theta_0)}_{G_0(s)} u(s) + \underbrace{H(s, \theta_0)}_{\nu(s)} e(s) \tag{7}$$

for some initially unknown parameter vector $\theta_0 \in \mathbb{R}^k$, where $e(s)$ is white noise of variance σ_e^2 , while $G(s, \theta_0)$ and $H(s, \theta_0)$ are stable transfer function (a simple frequency-based representation of the input-output behavior of linear systems), with $H(s, \theta_0)$ monic and minimum-phase. In most of the literature concerned with identification issues it is assumed that the system is identified with a model structure

$$M = \{G(s, \theta), H(s, \theta)\}, \quad \theta \in \mathbb{R}^k \tag{8}$$

In general if Z^N is our source data set composed by the observation data we would want to fit these data to the model structure M . M describes a set of models M^* within which the best one is sought for. In this framework we could argue that identifiability of model structure, as it was previously defined, concerns the question whether different parameter vectors may describe the same model in the set M^* . However a strictly related question is whether the data set Z^∞ allows us to distinguish between different models in the set. In this context we say that a data set can be defined 'informative' if it allows us to distinguish among different models. We say that a quasi-stationary data set Z^∞ is *informative enough with respect to the model set M^** if, for any two models $W_1(q)$ and $W_2(q)$ in this set:

$$\bar{E}[(W_1(q) - W_2(q)z(t))]^2 = 0$$

which implies that:

$$W_1(e^{j\omega}) \equiv W_2(e^{j\omega})$$

for almost all ω . On the other hand a quasi-stationary data set Z^∞ is *informative* if it is informative enough with respect to the model set ℓ^* consisting of all linear, time

invariant (LTI) models. The concept of informative dataset is tightly related to concept of *persistently exciting inputs*. This can be seen easily by observing that a quasi-stationary dataset Z^∞ is *informative* if the spectrum matrix for $z(t) = [u(t) \gamma(t)]^T$ is strictly positive definite for almost all ω . In fact, if we consider two models $W_1(q)$ and $W_2(q)$ and denote $\bar{W}(q) = W_1(q) - W_2(q)$, applying a well known theorem on signal filtration [33], we can write:

$$\int_{-\infty}^{+\infty} \bar{W}(e^{j\omega}) \Phi_z(\omega) \bar{W}^T(e^{-j\omega}) d\omega = 0$$

with $\Phi_x(\omega)$ spectrum of the signal $x(t)$ and where

$$\Phi_z(\omega) = \begin{bmatrix} \Phi_{uu}(\omega) & \Phi_{uy}(\omega) \\ \Phi_{yu}(\omega) & \Phi_{yy}(\omega) \end{bmatrix}.$$

Since $\Phi_z(\omega)$ is positive definite, this implies that $\bar{W}(e^{j\omega}) \equiv 0$ almost everywhere that proves the previous statement. Moreover we can observe that, given the $\Phi_z(\omega)$, for the Schur's Lemma we can assure algorithm convergence only if $\Phi_{uu}(\omega) > 0$ and $\Phi_{uu}(\omega) - \Phi_{uy}(\omega) \Phi_{yy}(\omega) \Phi_{yu}(\omega) > 0$. Evidently the only block of this array we have control on is the one representing the spectrum of input signal which directly depends on dynamical properties of the driving input signal we design. It is therefore convenient to reintroduce the concept of *persistently exciting signal of order n* for a quasi-stationary stimulus $u(t)$: we say that a similar signal, with spectrum $\Phi_u(\omega)$ is persistently exciting of order n is, for all filters of the form $M_n(q) = m_1q^{-1} + \dots + m_nq^{-n}$ the relation $|M_n(e^{j\omega})|^2 \Phi_u(\omega) \equiv 0$ implies that $|M_n(e^{j\omega})| \equiv 0$. Evidently the function $M_n(z) M_n(z^{-1})$ can have $n - 1$ different zeros on the unit circle (since one zero is always at the origin) taking symmetry into account. Hence $u(t)$ is persistently exciting of order n if $\Phi_u(\omega)$ is different from zero on at least n points in the interval $-\pi \leq \omega \leq \pi$. This is a direct consequence of the definition.

Signals that show such properties have been investigated and include:

- Pseudo Random Binary Signal
- Generalised Binary Noise (or Random Binary Signal)
- Sum of Sines and Filtered Noise
- Coloured Noise

Notice that this is a rather general result for the class of systems considered herein. Nevertheless a similar argumentation can be carried out by considering an

information metric directly tied to experimental data and to the model to be identified. These results are commonly derived from the analysis of some metric on the Fisher Information Matrix which are commonly referred to as 'Optimality criteria'.

Optimality criteria

As we previously outlined, we can estimate the information content of a measurement by the covariance matrix Σ of the estimated parameters. In order to illustrate how this can be done we will consider a simple estimation problem based on a widely used estimator: Maximum Likelihood Estimator (notice that these results can be extended to the Least Squares Estimator [67]). If we assume a normally distributed noise the identification process can be reduced to find θ that minimises:

$$\chi^2(\theta) = \sum_{j=1}^L \sum_{i=1}^N \left(\frac{y_j^D(t_i) - y_j^M(t_i, \theta, u)}{\sigma_{ij}} \right)^2 \tag{9}$$

The asymptotic distribution of the least squares estimate $\hat{\theta}$ can be computed analytically; then for a large number of observed samples the difference between real and estimated parameters tends to 0. For this reason, rather than minimising Eq. 9 directly, we linearise the function $y_j(t_i, \cdot)$ around θ_0 and minimise this simpler function. Using the Taylor series expansion of $y_j^M(t_i, \theta)$ around θ_0 we obtain

$$\begin{aligned} y_j^M(t_i, \theta, u) &= y_j^M(t_i, \theta_0, u) + \nabla_{\theta} y_j|_{t_i, \theta_0} (\theta - \theta_0) \\ &= y_j^M(t_i, \theta_0, u) + \left\{ \nabla_{\theta} g_j \nabla_{x_j} g_j \left([\nabla_{\theta} x_1]^T \dots [\nabla_{\theta} x_k]^T \right)^T \right\} |_{t_i, \theta_0} (\theta - \theta_0) \\ &= y_j^M(t_i, \theta_0, u) + \nabla_{\theta} y_j^M(t_i, \theta_0, u)^T \Delta\theta + o(\Delta\theta) \end{aligned}$$

where $\Delta\theta = \theta - \theta_0$ and ∇ is the so called nabla differential operator defined as

$$\Delta\theta = \left(\frac{\partial}{\partial\theta_1}, \dots, \frac{\partial}{\partial\theta_{n+k}} \right)^T \tag{10}$$

and $o(h^k)$ refers to the family of functions $w(h)$ for which

$$\lim_{h \rightarrow 0} \frac{w(h)}{h^k} = 0 \tag{11}$$

Substituting these results in the minimisation of the functional, we obtain

$$\chi^2(\theta) = \sum_{j=1}^L \sum_{i=1}^N \left[\frac{\varepsilon_{ij}^2}{\sigma_{ij}^2} - \frac{2\varepsilon_{ij}}{\sigma_{ij}^2} \nabla_{\theta} y_i(t_i, \theta) \Delta\theta + [\Delta\theta]^T \left(\frac{1}{\sigma_{ij}^2} [\nabla_{\theta} y_j]^T [\nabla_{\theta} y_j] \right) |_{t_i, \theta_0} \Delta\theta \right]$$

The minimisation of $\chi^2(\theta)$ with respect to θ brings to the following equation for the estimated deviation of the parameter vector $\Delta\theta$

$$\begin{aligned} \left\{ \sum_{j=1}^P \sum_{i=1}^N \frac{1}{\sigma_{ij}^2} [\nabla_{\theta} y_j]^T [\nabla_{\theta} y_j] \right\} \Delta\theta &=: F \Delta\theta \\ &= \sum_{j=1}^P \sum_{i=1}^N \frac{\varepsilon_{ij}}{\sigma_{ij}^2} [\nabla_{\theta} y_j]^T \end{aligned}$$

where we identified in $F = \sum_{j=1}^P \sum_{i=1}^N \frac{1}{\sigma_{ij}^2} [\nabla_{\theta} y_j]^T [\nabla_{\theta} y_j]$

the so called *Fisher Information Matrix*. Then we can solve the last equation

$$\Delta\theta = F^{-1} \sum_{j=1}^P \sum_{i=1}^N \frac{\varepsilon_{ij}}{\sigma_{ij}^2} [\nabla_{\theta} y_j]^T \tag{12}$$

Therefore we can compute the covariance matrix of the parameter vector as

$$\Sigma = \langle \Delta\theta \Delta\theta^T \rangle = F^{-1} \tag{13}$$

In order to compute this matrix we need the derivations of the observation function with respect to the parameters, $\nabla_{\theta} y_j(t_i)$. We then need to compute the derivative of g with respect to θ and x . In addition the derivation of $\nabla_{\theta} x_k(t_i)$ has to be computed from the system of ordinary differential equations,

$$\frac{\partial}{\partial t} (\nabla_{\theta} x_k) = \sum_{r=1}^k \frac{\partial f_k(\mathbf{x}, \theta)}{\partial x_r} \nabla_{\theta} x_r + \nabla_{\theta} f_k(\mathbf{x}, \theta) \tag{14}$$

with the initial conditions $(\nabla_{\theta} x_k)(0) = \nabla_{\theta} x_k(0)$. Given the Fisher Information Matrix (and thus the covariance matrix of the estimated parameters) the asymptotic confidence intervals for the estimates can be computed from the multivariate normal distribution

$$p(\theta) = \frac{\sqrt{\text{Det}(F)}}{P} e^{-\frac{1}{2} \theta^T F \theta} \tag{15}$$

$[2\pi] \frac{1}{2}$

We should also notice that, given the Cramer-Rao bound

$$E[\theta^2] \geq \frac{1}{F} \tag{16}$$

we can easily derive a lower bound for the efficacy of a general and unbiased estimator that directly depends

only on the Fisher Information Matrix. Evidently, the smaller the joint confidence intervals for the estimated parameters are, the more information the experiment carries with it. We can summarise the information about the variability in the covariance matrix into a single number by using metrics like $Det(F)$, $max(\lambda_i)$ (with λ_i representing the i^{th} eigenvalue of F). This is where the alternative choice of optimality criteria arises. We can distinguish four major measures of the information content [32]:

- A-Optimal design: maximising $trace(F)$
- D-Optimal design: minimising $Det(\Sigma)$
- E-Optimal design: minimising $\lambda_{max}(\Sigma)$
- Modified E-Optimal design: minimising $\frac{\lambda_{max}(\Sigma)}{\lambda_{min}(\Sigma)}$

A-Optimal designs are rarely used since they can lead to non-informative experiments [68]. D-Optimal designs can be interpreted as geometric means minimisation of the errors in the parameter estimates. E-Optimal and Modified E-Optimal designs try to minimise the largest uncertainty and the ratio of the largest and smallest uncertainties among parameter estimates respectively. Given the characteristics of each of these criteria and the computational efforts required for the specific problem we selected D-Optimality as driving criterion for our input design task.

Computational implementation

In order to carry out the optimisation of the input time profile we set up an optimisation routine based on a Genetic Algorithm (GA) thought to minimise an objective function encoding the D-Optimality metric on the FIM. Here we present a pseudocode of the proposed approach.

Algorithm 1 *Genetic_Algorithm(population, Fitness_Function)* returns individual

while {an individual has a fitness value higher(lower)than threshold} **do**

$new_population \leftarrow \emptyset$

for i from 1 to $Dim(population)$ **do**

$x \leftarrow Random_Selection(population, Fitness_Function)$

$y \leftarrow Random_Selection(population, Fitness_Function)$

$descendant \leftarrow Reproduction(x, y)$

if $small_random_probability \leftarrow new_population$ **then**

$new_population \leftarrow new_population \cup descendant$

end if

$population \leftarrow new_population$

end for

end while

return best individual evaluated on the Fitness_Function

Algorithm 2 *Reproduction(x, y)* returns individual

$x \leftarrow Length(x)$

$c \leftarrow random\ number\ in\ the\ range\ \{0, n\}$

return $Append(Substring(x, 1, c), Substring(y, c + 1, n))$

Algorithm 3 *Fitness Function x(x)* returns fitness value

$input \leftarrow x$

$time_evolution \leftarrow Simulate(EGFR\ pathway, input, \theta)$

$FIM \leftarrow Fisher_Information_Matrix(time_evolution)$

return FIM

Algorithm 1 shows how the optimisation task is carried out: here an individual encodes the time profile of the ligand concentration outside the cell. While the Fitness Function (FF) is used to estimate the quality of the single individual, *mutation* and *crossover* operators boost the search space exploration of the GA. This approach should help the algorithm returning the best solution (individual) to the input optimisation problem by optimising, generation after generation, the fitness value of the individuals in the population.

Microfluidic device design

Implementing complex time-varying signals is quite simple from a computational point of view; however obtaining realisations of signals with such properties is an active area of research in current microfluidics. Developing geometries that satisfy physical conditions for the generation of signals compliant with the specifications imposed by the theoretical results is not a trivial task. Several alternative solutions have been proposed for signal modulation in microfluidic channels [69-74] being [42,75,76] the most recent and advanced contribution in this fields; they are based on diverse physical principles like boundary diffusion controlled by relative velocity (like in H filters [74]), by exciting cells with diverse laminar flows that affect different parts of the cell etc.

In particular two main areas of research arose in this context: on the one hand interesting phenomena in fluid dynamics have been investigated [77,78] in order to address the problem of cell stimulation via constrained signals [79,80]; on the other hand the area of Digital Microfluidics has found one of its most active fields of research. In order to implement the signals obtained from the previously described input optimisation task we propose a polydimethylsiloxane (PDMS) based platform for cell stimulation which exploits the signal modulation at the microliter level. This platform has been designed to implement spectral properties of the signals that have been characterised during the theoretical study of the system under investigation: in this way it will become part of the optimal experimental design for cell stimulation.

The platform we propose is thought to be controlled via an *Hardware Abstraction Layer* (HAL) that should allow the user to design his experiment and let the system operate with fluids and devices (like pumps, latches etc). Bio-stream is a suitable example of such an architecture [45]. The design of 'Labrys', this is the name selected for the device, is reported in Fig. 5. The proposed architecture is based on the studies carried out in [42] and by Dr. Thomson [79,80]. *Labrys* features two distinct layers being the upper one in charge of the control mechanisms. This control layer has been designed using the max-flow-min-cost principle implemented in the MICADO package (v0.5) developed by Nada Amin. MICADO can also extract the microfluidic ISA the chip will be based on. The second layer has been designed using the *Mix-Store and Use* principle: this layer is composed of Inputs/Outputs (red boxes in Fig. 5), Muxes-Demuxes (blue boxes in Fig. 5), Registers (green boxes in Fig. 5), Mixing Units (brown boxes in Fig. 5), Flow Chamber (the 'processing unit' in violet in Fig. 5) and for this reason can be thought of as a sort of unconventional computational architecture for generating, storing and using previously characterised inputs to be used for cell stimulation. Testing such a platform, given the computational results discussed above is a simple task. A feasible approach to carry out such a process can employ an off-line generation of the discrete levels of the inducer molecule (the EGF concentration in our case) in the medium and its sequential and automated supply to the cells through a suitable valves actuation strategy.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FM has conceived the idea, has designed the device, has carried out the experiments and has written the

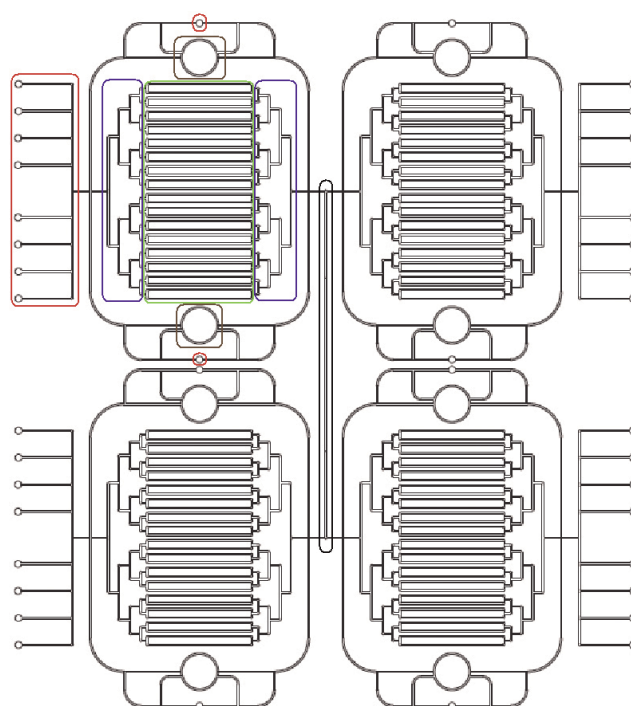


Figure 5

Labrys design. Outlook of the microfluidic device designed. The chip is composed of four functional units each featuring an input-output section (red boxes), two rotary mixers (brown box), two sets of demultiplexers and multiplexers (blue boxes) to drive the fluids in the appropriate registers (green box). The flow channel (violet box) is intended to host cells to be stimulated by compound mixes stored in the registers and moved by peristaltic valve actuation.

manuscript. TM and DB have partly supervised the work and assisted in manuscript preparation. VB, CC, AP and ST supervised the project and inspired the integration of the computational and device engineering sides of this work. All authors read and approved the final manuscript.

Acknowledgements

This work was partially supported by progetto finalizzato ACC, 'Rete Nazionale di Bioinformatica Oncologica' 2007.

This article has been published as part of *BMC Bioinformatics* Volume 10 Supplement 12, 2009: Bioinformatics Methods for Biomedical Complex System Applications. The full contents of the supplement are available online at <http://www.biomedcentral.com/1471-2105/10?issue=S12>.

References

1. Hornberg JJ, Bruggeman FJ, Westerhoff HV and Lankelma J: **Cancer: A Systems Biology disease.** *Biosystems* 2006, **83**:81-90.
2. Burden S and Yarden Y: **Neuregulins and their receptors: a versatile signaling module in organogenesis and oncogenesis.** *Neuron* 1997, **18**:847-855.
3. Yarden Y and Sliwkowski MX: **Untangling the ErbB signalling network.** *Nature Reviews. Molecular Cell Biology* 2001, **2**:127-37.

4. Eastman A and Perez RP: **New targets and challenges in the molecular therapeutics of cancer.** *British Journal of Clinical Pharmacology* 2006, **62**:5–14.
5. Sebastian S, Settleman J, Reshkin SJ, Azzariti A, Bellizzi A and Paradiso A: **The complexity of targeting EGFR signalling in cancer: from expression to turnover.** *Biochimica Et Biophysica Acta* 2006, **1766**:120–39.
6. Mendelsohn J and Baselga J: **The EGF receptor family as targets for cancer therapy.** *Oncogene* 2000, **19**:6550–65.
7. Kholodenko BN, Demin OV, Moehren G and Hoek JB: **Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor.** *J Biol Chem* 1999, **274**:30169–30181.
8. Oda K, Matsuoka Y, Funahashi A and Kitano H: **A comprehensive pathway map of epidermal growth factor receptor signaling.** *Mol Syst Biol* 2005, **1**.
9. Wiley HS, Shvartsman SY and Lauffenburger DA: **Computational modeling of the EGF-receptor system: a paradigm for systems biology.** *Trends in Cell Biology* 2003, **13**:43–50.
10. Schoeberl B, Eichler-Jonsson C, Gilles ED and Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors.** 2002, **20**:370–5.
11. Hornberg JJ, Binder B, Bruggeman FJ, Schoeberl B, Heinrich R and Westerhoff HV: **Control of MAPK signalling: from complexity to what really matters.** *Oncogene* 2005, **24**:5533–42.
12. Breiting R and Hoeller D: **Current challenges in quantitative modeling of epidermal growth factor signaling.** *FEBS Letters* 2005, **579**:6289–94.
13. Zhang Y, Wolf-Yadlin A, Ross PL, Pappin DJ, Rush J, Lauffenburger DA and White FM: **Time-resolved mass spectrometry of tyrosine phosphorylation sites in the epidermal growth factor receptor signaling network reveals dynamic modules.** *Molecular & Cellular Proteomics: MCP* 2005, **4**:1240–50.
14. Olsen JV, Blagoev B, Gnäd F, Macek B, Kumar C, Mortensen P and Mann M: **Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.** *Cell* 2006, **127**:635–48.
15. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ and Mann M: **A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling.** *Nature Biotechnology* 2003, **21**:315–8.
16. Hengl S, Kreutz C, Timmer J and Maiwald T: **Data-based identifiability analysis of non-linear dynamical models.** *Bioinformatics* 2007, **23**:2612–2618.
17. Casey FP, Baird D, Feng Q, Gutenkunst RN, Waterfall JJ, Myers CR, Brown KS, Cerione RA and Sethna JP: **Optimal experimental design in an EGFR signaling and down-regulation model.** *q-bio/0610024* 2006 <http://arxiv.org/abs/q-bio/0610024>.
18. Faller D, Klingmüller U and Timmer J: **Simulation Methods for Optimal Experimental Design in Systems Biology.** *SIMULATION* 2003, **79**:717–725.
19. Vance W, Arkin A and Ross J: **Determination of causal connectivities of species in reaction networks.** *Proc Natl Acad Sci U S A* 2002, **99**:5816–5821.
20. Samoilov M, Arkin A and Ross J: **On the deduction of chemical reaction pathways from measurements of time series of concentrations.** *Chaos (Woodbury, N.Y.)* 2001, **11**:108–114.
21. Vlad MO, Arkin A and Ross J: **Response experiments for nonlinear systems with application to reaction kinetics and genetics.** *Proc Natl Acad Sci U S A* 2004, **101**:7223–7228.
22. Flaherty P, Jordan MI and Arkin AP: **Robust design of biological experiments.** *Proceedings of the Neural Information Processing Symposium* 2005, **2005**.
23. Mettetal JT, Muzzey D, Gómez-Urbe C and van Oudenaarden A: **The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*.** *Science (New York, N.Y.)* 2008, **319**:482–4.
24. Steuer R, Kurths J, Fiehn O and Weckwerth W: **Observing and interpreting correlations in metabolomic networks.** *Bioinformatics* 2003, **19**:1019–1026.
25. Yue H, Brown M, Knowles J, Wang H, Broomhead DS and Kell DB: **Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF-[small kappa]B signalling pathway.** *Molecular BioSystems* 2006, **2**:640–649.
26. Quaiser T, Marquardt W and Monnigmann M: **Local identifiability analysis of large signalling pathway models.** *Proc. of FOSBE 2007* 2007, **465**–470.
27. Anguelova M and Wennberg B: **Identifiability of the Time-lag Parameter in Delay Systems with Applications to Systems Biology.** *Proc. of FOSBE 2007* 2007, **533**–538.
28. Geffen Dara, S M, A F. Findeisen Rolf and Guay M: **The Question of Parameter Identifiability for Biochemical Reaction Networks Considering the NF-kappaB Signal Transduction Pathway.** *Proc. of FOSBE 2007* 2007, **509**–514.
29. Cobelli C and Romanin-Jacur G: **Structural identifiability of strongly connected biological compartmental systems.** *Med Biol Eng* 1975, **13**:831–838.
30. Cobelli C, Lepschy A and Romanin-Jacur G: **Structural identifiability of biological compartmental systems. Digital computer implementation of a testing procedure** 1976, **88**–102 http://dx.doi.org/10.1007/3-540-07622-0_463.
31. Bellu G, Saccomani MP, Audoly S and D'Angiò L: **DAISY: a new software tool to test global identifiability of biological and physiological systems.** *Comput Methods Programs Biomed* 2007, **88**:52–61.
32. Pronzato L: **Optimal experimental design and some related control problems.** *0802.4381* 2008 <http://arxiv.org/abs/0802.4381>. [Automatica/Automatika] IFAC; Automatika; Automatica IFAC J 44 (2008) 303–325].
33. Ljung L: *System Identification: Theory for the User* Prentice Hall PTR; 21999.
34. Phair RD and Misteli T: **Kinetic modelling approaches to in vivo imaging.** *Nature Reviews. Molecular Cell Biology* 2001, **2**:898–907, [PMID: 11733769].
35. Cho KH, Shin SY, Kolch W and Wolkenhauer O: **Experimental Design in Systems Biology, Based on Parameter Sensitivity Analysis Using a Monte Carlo Method: A Case Study for the TNFalpha-Mediated NF-kappa B Signal Transduction Pathway.** *SIMULATION* 2003, **79**:726–739.
36. Lipschultz CA, Li Y and Smith-Gill S: **Experimental design for analysis of complex kinetics using surface plasmon resonance.** *Methods (San Diego, Calif.)* 2000, **20**:310–8.
37. Sontag E: **Molecular Systems Biology and Control: A Qualitative-Quantitative Approach.** *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on 20052314–2319*.
38. Maiwald T and Timmer J: **Dynamical Modeling and Multi-Experiment Fitting with PottersWheel.** *Bioinformatics* 2008, **btn350** <http://bioinformatics.oxfordjournals.org/cgi/content/abstract/btn350v1>.
39. Balsa-Canto E, Alonso A and Banga J: **Computational procedures for optimal experimental design in biological systems.** *IEE Systems Biology* 2008, **2(4)**:163–172.
40. Rylander BI: **Computational complexity and the genetic algorithm.** *PhD thesis* University of Idaho; 2001 <http://portal.acm.org/citation.cfm?id=933869>.
41. Hollander M and Wolfe DA: *Nonparametric Statistical Methods* Wiley-Interscience; 21999.
42. Andrew N, Craig D, Urbanski JP, Gunawardena J and Thorsen T: **Microfluidic temporal cell stimulation.** *µTAS* 08 2008.
43. Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV and Hoek JB: **Untangling the wires: a strategy to trace functional interactions in signaling and gene networks.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99(20)**:12841–6.
44. Sontag E, Kiyatkin A and Kholodenko BN: **Inferring dynamic architecture of cellular networks using time series of gene expression, protein and metabolite data.** *Bioinformatics* 2004, **20(12)**:1877–1886.
45. Thies W, Urbanski J, Thorsen T and Amarasinghe S: **Abstraction layers for scalable microfluidic biocomputing.** *Natural Computing* 2006 <http://dx.doi.org/10.1007/s11047-006-9032-6>.
46. **BioModels Database.** <http://www.ebi.ac.uk/biomodels-main/static-pages.do?page=home>.
47. **BioModels Database – Kholodenko1999-EGFRsignaling.** <http://www.ebi.ac.uk/biomodels-main/publ-model.do?mid=-BIOMD0000000048>.
48. Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P and Kummer U: **COPASI-a Complex Pathway Simulator.** *Bioinformatics* 2006, **22**:3067–3074.
49. Gunawardena J: **Signals and Systems: Towards a Systems Biology of Signal Transduction.** *Proceedings of the IEEE* 2008, **96**:1386–1397.
50. Hu S: **Optimal time points sampling in pathway modelling.** *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE* 2004, **1**:671–674.
51. Steinke F, Seeger M and Tsuda K: **Experimental design for efficient identification of gene regulatory networks using sparse Bayesian models.** *BMC Systems Biology* 2007, **1**:51.

52. van Riel NA: **Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments.** *Brief Bioinform* 2006, **7**:364–374.
53. Asprey S and Macchietto S: **Designing robust optimal dynamic experiments 2002.** <http://www.ingentaconnect.com/content/els/09591524/2002/00000012/00000004/art00020>.
54. Apgar JF, Toettcher JE, Endy D, White FM and Tidor B: **Stimulus Design for Model Selection and Validation in Cell Signaling.** *PLoS Computational Biology* 2008, **4**:e30.
55. Bernaerts K and Impe JV: **Optimal dynamic experiment design for estimation of microbial growth kinetics at sub-optimal temperatures: Modes of implementation.** *Simulation Modelling Practice and Theory* 2005, **13**:129–138.
56. Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, Bullinger E, Allgower F and Gilles ED: **A Benchmark for Methods in Reverse Engineering and Model Discrimination: Problem Formulation and Solutions.** *Genome Res* 2004, **14**:1773–1785.
57. Brik Ternbach M, Bollman C, Wandrey C and Takors R: **Application of model discriminating experimental design for modeling and development of a fermentative fed-batch L-valine production process.** *Biotechnol Bioeng* 2005, **91**:356–368.
58. Franceschini G and Sanro Macchietto: **Model-based design of experiments for parameter precision: State of the art.** *Chemical Engineering Science* 2007.
59. Chen B and Asprey S: **On the Design of Optimally Informative Dynamic Experiments for Model Discrimination in Multi-response Nonlinear Situations.** *Industrial & Engineering Chemistry Research* 2003, **42**:1379–1390.
60. Cooney M and McDonald K: **Optimal dynamic experiments for bioreactor model discrimination.** *Applied Microbiol Biotechnol* 1995, **43**:826–837.
61. Atkinson AC and Donev AN: **Optimum experimental designs.** *Optimum Experimental Designs*.
62. Casey F: **Prediction and Optimal Experimental Design in Systems Biology.** *PhD in Physics* Cornell University, LASSP, Physics Department, Clark Hall, Cornell University, Ithaca, NY 14853-2501; 2006.
63. **UCSB Biosens – Bio-SPICE Dashboard.** <http://www.chemengr.ucsb.edu/~ceweb/faculty/doyle/biosens/BioSens.htm>.
64. **UCSB Biosens – Bio-SPICE Dashboard.** <http://cbbl.imim.es:8080/ByoDyn>.
65. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR and Sethna JP: **Universally Sloppy Parameter Sensitivities in Systems Biology Models.** *PLoS Comput Biol* 2007, **3(10)**:1871–1878.
66. Mehra R: **Optimal input signals for parameter estimation in dynamic systems-Survey and new results.** *Automatic Control, IEEE Transactions on* 1974, **19(6)**:753–768.
67. Rodriguez-Fernandez M, Egea J and Banga J: **Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems.** *BMC Bioinformatics* 2006, **7**:483.
68. Goodwin G: **Identification: Experiment Design.** *Systems and Control Encyclopedia* 1987, **4**:2257–64.
69. Weibel DB and Whitesides GM: **Applications of microfluidics in chemical biology.** *Current Opinion in Chemical Biology* 2006, **10**:584–591.
70. Walker GM, Zeringue HC and Beebe DJ: **Microenvironment design considerations for cellular scale studies.** *Lab on a Chip* 2004, **4**:91–97.
71. Takayama S, Ostuni E, LeDuc P, Naruse K, Ingber DE and Whitesides GM: **Selective Chemical Treatment of Cellular Microdomains Using Multiple Laminar Streams.** *Chemistry & Biology* 2003, **10**:123–130.
72. Takayama S, Ostuni E, LeDuc P, Naruse K, Ingber DE and Whitesides GM: **Subcellular positioning of small molecules.** *Nature* 2001, **411**:1016.
73. Breslauer DN, Lee PJ and Lee LP: **Microfluidics-based systems biology.** *Molecular BioSystems* 2006, **2**:97–112.
74. Squires TM and Quake SR: **Microfluidics: Fluid physics at the nanoliter scale.** *Reviews of Modern Physics* 2005, **77**:977–50.
75. Ye L, Zhang M, Alexopoulos L, Sorger P and Jensen K: **Microfluidic devices for studying the response of adherent cells under short time stimuli treatment.** *μTAS 07* 2007.
76. King KR, Wang S, Jayaraman A, Yarmush ML and Toner M: **Microfluidic flow-encoded switching for parallel control of dynamic cellular microenvironments.** *Lab on a Chip* 2008, **8**:107–116.
77. Urbanski JP: **Application of Microfluidic Emulsion Technology to Biochemistry, Drug Delivery and Lab-on-a-Chip Programmability.** 2005.
78. Thorsen T: **Microfluidic Technologies for High-Throughput Screening Applications.** 2003.
79. Thomson TM: **TMT Thesis Project – OpenWetWare.** 2005 http://openwetware.org/wiki/TMT_Thesis_Project.
80. Thomson TM: **Stimulator Project – OpenWetWare.** 2005 <http://openwetware.org/wiki/Stimulator>.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

