Dynamic Pathway Modeling

Feasibility Analysis and Optimal Experimental Design

THOMAS MAIWALD,[a] CLEMENS KREUTZ,[a] ANDREA C. PFEIFER,[b] SEBASTIAN BOHL,[b] URSULA KLINGMÜLLER,[b] AND JENS TIMMER[a]

[a] *Freiburg Center for Data Analysis and Modeling, Freiburg University, 79104 Freiburg, Germany*

[b] *German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany*

ABSTRACT: A major challenge in systems biology is to evaluate the feasibility of a biological research project prior to its realization. Since experiments are animals-, cost- and time-consuming, approaches allowing researchers to discriminate alternative hypotheses with a minimal set of experiments are highly desirable. Given a null hypothesis and alternative model, as well as laboratory constraints like observable players, sample size, noise level, and stimulation options, we suggest a method to obtain a list of required experiments in order to significantly reject the null hypothesis model M_0 if a specified alternative model M_A is realized. For this purpose, we estimate the power to detect a violation of M_0 by means of Monte Carlo simulations. Iteratively, the power is maximized over all feasible stimulations of the system using multi-experiment fitting, leading to an optimal combination of experimental settings to discriminate the null hypothesis and alternative model. We prove the importance of simultaneous modeling of combined experiments with quantitative, highly sampled *in vivo* measurements from the Jak/STAT5 signaling pathway in fibroblasts, stimulated with erythropoietin (Epo). Afterwards we apply the presented iterative experimental design approach to the Jak/STAT3 pathway of primary hepatocytes stimulated with IL-6. Our approach offers the possibility of deciding which scientific questions can be answered based on existing laboratory constraints. To be able to concentrate on feasible questions on account of inexpensive computational simulations yields not only enormous cost and time saving, but also helps to specify realizable, systematic research projects in advance.

KEYWORDS: systems biology; mathematical modeling; experimental design; null hypothesis; multi-experiment fitting

Address for correspondence: Thomas Maiwald, Freiburg Center for Data Analysis and Modeling, Freiburg University, Eckerstrasse 1, 79104 Freiburg, Germany. Voice: +49-761-1528-703; fax: +49-761-203-5967.
maiwald@fdm.uni-freiburg.de

Ann. N.Y. Acad. Sci. 1115: 212–220 (2007).
doi: 10.1196/annals.1407.007

INTRODUCTION

Mathematical modeling of cellular processes deepens the functional understanding of biochemical interaction at the microbiological level.[1–5] However, successful modeling requires high-quality time-resolved quantitative measurements of protein concentrations, which are still difficult to obtain.[6,7] Hence, for a given noise level and set of observable proteins, the question arises as to which biological hypotheses can be investigated with a reasonable effort. Several methods were suggested to improve the experimental setting in order to optimize the estimation of model parameters.[8,9] We developed an experimental design approach which suggests an optimal set of required experiments in order to discriminate competing biological hypotheses for given laboratory constraints.

METHODS

Statistical Background: Hypothesis Testing

According to the Neyman–Pearson paradigm, a test statistic T(x) with sample values x can be used to decide whether or not to reject a null hypothesis H_0 in favor of an alternative H_A.[10,11] The sets of values of T for which H_0 is accepted is called the acceptance region or confidence interval with size of $1 - \alpha$ and those for which H_0 is rejected are called the rejection region of the test. The critical value determines the border between the two regions. In applying the paradigm, a type 1 error (i.e., a false positive event) occurs when H_0 is rejected although it is true. Its probability is designated by α. Type 2 errors (i.e., false negative events) occur when H_0 is accepted although it is false. Its probability is β. The power of the test is defined as the probability that H_0 is rejected when it is false and equals $1 - \beta$. An ideal test would have $\alpha = \beta = 0$. Hence, by designing optimal experiments we try to minimize the overlap between null and alternative distributions in order to improve the model discrimination power. The interpretation of a test is as follows: If H_0 is not rejected, it is either really true or the test had no power to detect the violation. Otherwise, if H_0 is rejected, one can only conclude that H_0 is not true, but the validity of the alternative is not proved.

The Jak/STAT3 Signaling Pathway

We exemplify our approach to the IL-6/Jak/STAT3 signaling pathway of primary hepatocytes. The null hypothesis model M0 is depicted in FIGURE 1: The gp130 receptor gets activated via IL-6 ligands and phosphorylates STAT3 monomers. Two pSTAT3 molecules build dimers, are phosphorylated a second time, and enter the nucleus, where they act as transcription factors before

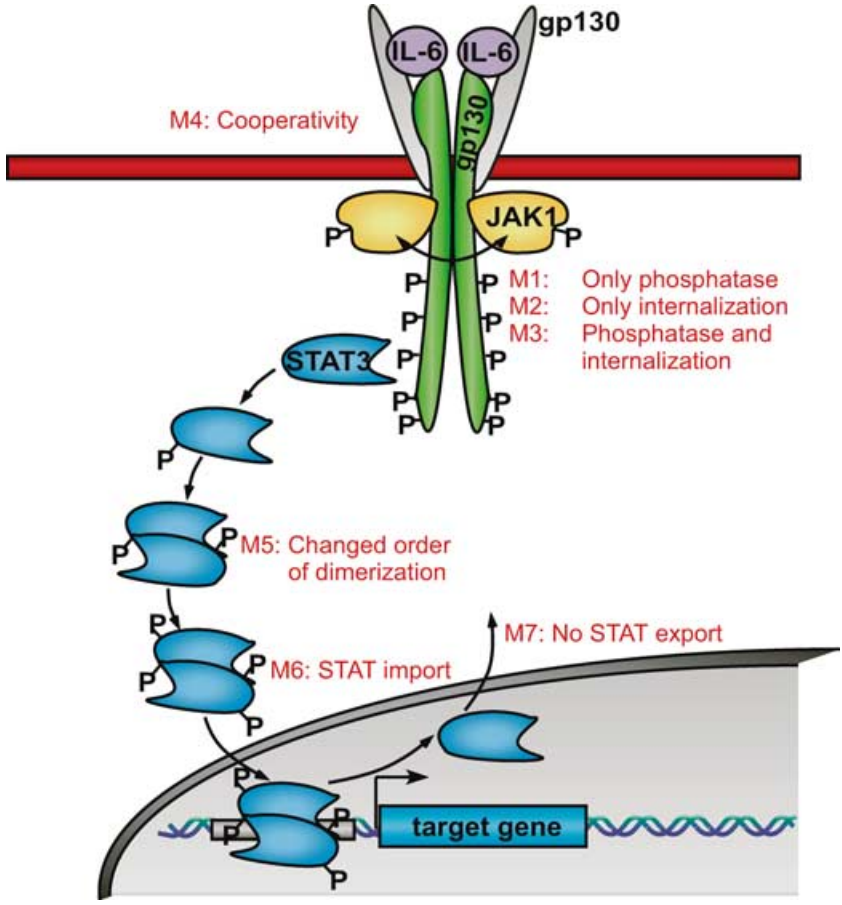

FIGURE 1. Null-hypothesis and alternative models of the Jak/STAT3 pathway. Under the null hypothesis, the gp130 receptor is activated via IL-6 and phosphorylates STAT3 monomers. Two pSTAT3 molecules build dimers, get phosphorylated a second time, and enter the nucleus. where they act as transcription factors before leaving the nucleus. M_1–M_7 represent possible extensions or modifications of the null hypothesis model.

leaving the nucleus again. The corresponding chemical reactions are translated into ordinary differential equations via mass-action kinetics and fitted to quantitative measurements in primary hepatocytes. Discrepancies between model and data and screening of the literature led to several alternative models, M_1–M_7. The question arose as to which alternative could be successfully distinguished from the null hypothesis model for a given set of observable proteins and a given noise level.

Laboratory Constraints, Stimulus Control, and System Characteristics

Our iterative experimental design approach requires specification of laboratory constraints, characteristic times, and stimulation doses. Depending on the investigated cell system, technical facilities, and experimental skill, several constraints are formulated:

- maximum number of possible time points per experiment;
- expected measurement noise level;
- set of observable proteins; and
- realizable stimulations (e.g., dose responses, pulses, ramps).

These constraints have a strong influence on the optimal combination of experiments and the feasibility of the whole research agenda.

As in control engineering, the properties of a dynamical system can only be characterized satisfactorily if the system can be controlled or stimulated with a variety of different input functions. Oscillating inputs—the most often used input function in control engineering—are currently not applicable to the biological systems. However, continuous, pulsed, and ramped stimulations can be applied. This enhances model selection dramatically if methods for multi-experiment fitting are available.

Different signal transduction pathways may possess a strongly different temporal behavior and may need different stimulation doses. Our approach generates a set of basic stimulation experiments which are scaled in time and dosage to meet a realistic stimulus. Therefore a short, medium, and long time period must be specified and similarly a low, medium, and large stimulation dose.

RESULTS

Improved Model Discrimination through Multi-Experiment Fitting

The enormous increase of model discrimination power by use of multi-experiment fitting is shown for a stimulation experiment with fibroblasts. The cells were stimulated with erythropoietin (Epo) in one experiment continuously over 180 minutes and in a second experiment for only 5 min (FIG. 2A,B). Three models, M_1–M_3, with increasing complexity were used to describe the receptor kinetics. Fitting of the simplest model M_1 to the activated receptor time courses shows a strong discrepancy for the continuous stimulation (FIG. 2C). Model M_2 can describe both data sets separately (FIG. 2E,F), but only with different, *locally* fitted parameter values. Since a model should be able to describe a system with the same parameter values for different stimulations, a multi-experiment fit is required where both data sets are fitted simultaneously. Figure 2G and H show that model M_2 is not flexible enough to explain the

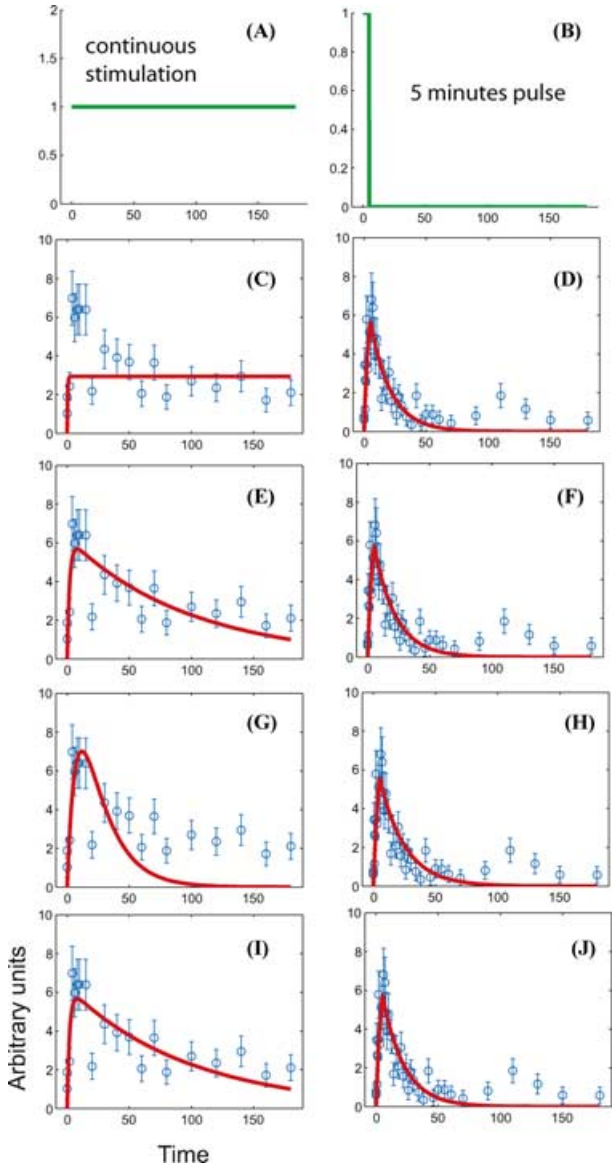

FIGURE 2. Single and multi-experiment fitting. Fibroblasts are stimulated in two experiments with Epo continuously (**A**) and for 5 min (**B**). The activated receptor is measured and fitted with models M_1–M_3 in single and multi-experiment manner, that is, with parameter values depending on each data set or globally unique values. M_1 can describe the fast decay after the pulsed stimulation (**D**), but not the slow decrease of the continuous one (**C**). M_2 can explain both data sets, but only with locally fitted parameters (**E, F**). The multi-experiment fit explains the pulsed stimulation well (**H**), but shows strong discrepancies for the continuous stimulation (**G**). Only M_3 is able to be fitted simultaneously to both time courses (**I, J**).

slow decrease for the continuous stimulation and the fast decrease after the pulsed stimulation at once. Only model $M_{3,}$ which comprises models M_1 and $M_{2,}$ can explain the data sets for the identical set of parameters (FIG. 2I, J).

Chi-Square Test Statistic

The discrepancy between a fitted model and a data set—the goodness-of-fit—can be quantified by the chi-square value,

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - y(x_i; a_1, \ldots, a_M)}{\sigma_i} \right)^2,$$

for N data points y_i measured at x_i with standard deviation σ_i and modeled by $y(x_i; a_1, \ldots, a_M)$ with M parameters a_j.[12] For a valid model and Gaussian distributed errors, the chi-square values are chi-square distributed with k degrees of freedom with k between N-M and N, depending on the model structure. If the calculated chi-square value is not compliant with the corresponding chi-square distribution, the used model can either be significantly rejected to sufficiently describe the data or the used error model is wrong. Since we are interested in discriminating the null hypothesis model from an alternative one, we simulate data from an alternative model and fit the null hypothesis model to the data. Afterwards, the chi-square value is calculated and compared with the significance level of the corresponding chi-square distribution.

Iterative Experimental Design

A basic set S_B of stimulation experiments is generated based on user-defined feasible experimental techniques and characteristics of the current cell system. After specifying an initial set S_0 of stimulation experiments (e.g. the empty set or already applied experiments), new experiments are added iteratively:

1. In iteration i, the current set S_{i-1} is consecutively combined with one stimulation experiment e_k of the basic set of the feasible stimulations S_B.
2. Data from M_A are simulated for $\{S_{i-1}, e_k\}$ and the null hypothesis model M_0 is fitted to the data, leading to the chi-square value d_k.
3. The optimal next experiment e_{opt} is determined by the maximum chi-square value over all d_k, since it has the highest power to discriminate the false null hypothesis from the true alternative.
4. The iteration is stopped if the null-hypothesis can be rejected significantly or if the maximum iteration number is reached. In the last case, the null hypothesis cannot be distinguished from the alternative within the maximal specified experimental effort.

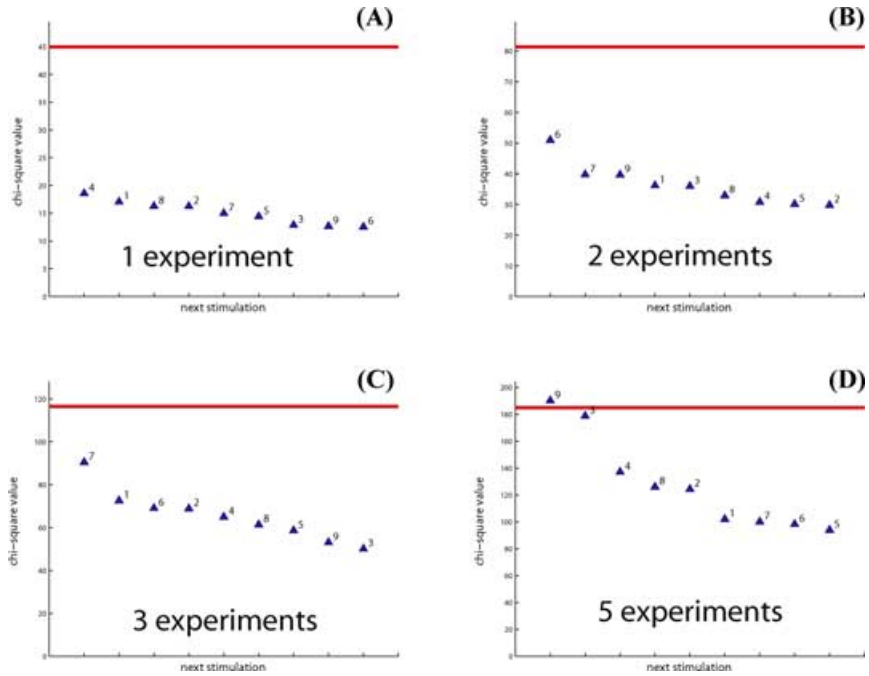

FIGURE 3. Iterative model discrimination. chi-square distance between data simulated with alternative model M_4 and fitted with null hypothesis model M_0. If only one experiment can be accomplished, it should be setting 4 yielding the highest but not significant chi-square value (**A**). In combination with further experiments (**B, C**), the chi-square value increases but is still not large enough, to successfully reject the wrong null hypothesis. M_0 would be statistically compliant with the data produced by M_4. Only an optimal combination of five experiments leads to a significant result (**D**).

The reaction schemes of all models are translated into sets of ordinary differential equations. By means of an implicit Runge–Kutta Fortran integrator the systems are integrated during simulation or fitting.[13] Parameters are adapted during the fitting procedure with the trust-region approach, as implemented in the MATLAB optimization toolbox.[14–18] All simulations and fits were applied within our new developed modeling framework, PottersWheel.[19]

Application to the Jak/STAT3 Signaling Pathway

Alternative M_4 describes a cooperative behavior of the activated receptor. We will determine the optimal experimental design in order to distinguish M_4 from the null-hypothesis M_0. FIGURE 3A shows the ordered chi-square values for nine different stimulations after one hypothetical experiment. Stimulation 4 yields the highest discrimination power, but does not suffice to reject the

wrong null hypothesis. Iterative combination with further stimulations up to four experiments is still not sufficient: M_4 cannot be distinguished from M_0 within four experiments (FIG. 3B, C). Only by combination of five experiments can a significant rejection be achieved (FIG. 3D).

CONCLUSION

We presented an approach to iteratively create an optimal set of experiments in order to increase the power to detect a violation of a null hypothesis model when a specific alternative is realized instead. It takes into account cell-system and laboratory-specific constraints like feasible stimulation types, set of observable proteins, and noise levels. The simultaneous fit of one model to several data sets is a key procedure of our approach. Application to the Jak/STAT3 signaling pathway of primary hepatocytes shows that the question of receptor cooperativity cannot be investigated with less than five optimal combined experiments.

ACKNOWLEDGMENTS

This work was supported by the HepatoSys initiative of the German Federal Ministry of Education and Research (BMBF, 0313074D) and the European project of Computational Systems Biology of Cell Signalling (COSBICS, LSHG-CT-2004-512060)

REFERENCES

1. KITANO, H. 2002. Systems biology: a brief overview. Science **295:** 1662–1664.
2. BENTELE, M., I. LAVRIK, M. ULRICH, *et al.* 2004. Mathematical modeling reveals threshold mechanism in CD95-induced apoptosis. J. Cell. Biol. **166:** 839–851.
3. KHOLODENKO, B.N., A. KIYATKIN, F.J. BRUGGEMAN, *et al.* 2002. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proc. Natl. Acad. Sci. USA **99:** 12841–12846.
4. SWAMEYE, I., T.G. MULLER, J. TIMMER, *et al.* 2003. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. Proc. Natl. Acad. Sci. USA **100:** 1028–1033.
5. WESTERHOFF, H.V. & B.O. PALSSON. 2004. The evolution of molecular biology into systems biology. Nat. Biotechnol. **22:** 1249–1252.
6. SCHILLING, M., T. MAIWALD, S. BOHL, *et al.* 2005. Computational processing and error reduction strategies for standardized quantitative data in biological networks. FEBS J. **272:** 6400–6411.
7. SCHILLING, M., T. MAIWALD, S. BOHL, *et al.* 2005. Quantitative data generation for systems biology: the impact of randomisation, calibrators and normalisers. IEEE Syst. Biol. **152:** 193–200.

8. GADKAR, K.G., R. GUNAWAN & F.J. DOYLE, III. 2005. Iterative approach to model identification of biological networks. BMC Bioinformatics **6:** 155.
9. CHO, K.H., S.Y. SHIN, W. KOLCH & O. WOLKENHAUER. 2003. Experimental design in systems biology, based on parameter sensitivity analysis using a Monte Carlo method: a case study for the TNFα-mediated NF-kB signal transduction pathway. Simulation **79:** 726–739.
10. COX, D.R. & D.V. HINKLEY. 1974. Theoretical Statistics. Chapman and Hall. London.
11. RICE, J.A. 1995. Mathematical Statistics and Data Analysis. Duxbury Press. Belmont, MA.
12. PRESS, W.H., W.H. PRESS & W.T. VETTERLING. 1992. Numerical Recipes in C: the Art of Scientific Computing. Cambridge University Press. Cambridge, UK.
13. HAIRER, E. & G. WANNER. 1996. Solving Ordinary Differential Equations II: stiff and Differential Algebraic Problems. Springer Verlag. Berlin.
14. MORÉ, J.J. & D.C. SORENSEN. 1983. Computing a trust region step. SIAM J. Sci. Stat. Comput. **4:** 553–572.
15. COLEMAN, T.F. & A. VERMA. 2001. A preconditioned conjugate gradient approach to linear equality constrained minimization. Computat. Optim. Appl. **20:** 61–72.
16. COLEMAN, T.F. & Y. LI. 1996. An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim. **6:** 418–445.
17. STEIHAUG, T. 1983. The conjugate gradient method and trust regions in large scale optimization. SIAM J. Num. Anal. **20:** 626–637.
18. BYRD, R.H., R.B. SCHNABEL & G.A. SHULTZ. 1988. Approximate solution of the trust region problem by minimization over two-dimensional subspaces. Math. Program. **40:** 247–263.
19. MAIWALD, T. & J. TIMMER. PottersWheel, a Systems Biology Toolbox for Modeling and Multi-Experiment Fitting (www.PottersWheel.de).