Designing Optimal Experiments to Discriminate Interaction Graph Models

Sven Thiele, Sandra Heise, Wiebke Hessenkemper, Hannes Bongartz,
Melissa Fensky, Fred Schaper, and Steffen Klamt

Abstract—Modern methods for the inference of cellular networks from experimental data often express nondeterminism by proposing an ensemble of candidate models with similar properties. To further discriminate among these model candidates, new experiments need to be carried out. Theoretically, the number of possible experiments is exponential in the number of possible perturbations. In praxis, experiments are expensive and usually there exist several constraints limiting which experiments can be performed. Limiting factors may exist on the combinations of perturbations that are technically possible, which components can be measured, and limitations on the number of affordable experiments. Further, not all experiments are equally well suited to discriminate model candidates. Therefore, the goal of optimal experiment design is to determine those experiments that discriminate most of the candidates while minimizing the costs. We present an approach for experiment planning with interaction graph models and sign consistency methods. This new approach can be used in combination with methods for network inference and consistency checking. The proposed method determines experiments which are most suitable to deliver results that reduce the number of candidate models. We applied our method to study the Erythropoietin signal transduction in human kidney cells `HEK293`. We first used simulated experiment data from an ODE model to demonstrate *in silico* that our experimental design results in the inference of the gold standard model. Finally, we used the approach to plan *in vivo* experiments that enabled us to discriminate model candidates for the Erythropoietin signal transduction in this cell line.

Index Terms—Signaling networks, interaction graph models, sign consistency, experiment design, erythropoietin signal transduction, HEK293

✦

---

1 INTRODUCTION

MODERN experimental technologies allow the observation and targeted perturbation of different cellular components. This ability allows us to gain insight into the internal functioning of cellular signaling processes and to infer unknown interactions between them. Interaction or influence graphs (IG) are a simple and widely used type of model that capture interesting and relevant behaviors [1], [2], [3], [4], [5], [6].

In our previous work we presented qualitative methods that use constraints on the signs of node activity to relate measurements and model. We proposed methods for consistency checking of models and measurements, for the inference of models, and for the prediction of system responses towards external perturbations. Due to the limitations of todays perturbation and measurement techniques as well as the uncertainty in data, inference methods often describe indeterminacy through an ensemble of candidate models with similar properties being consistent with the data at hand. To further discriminate among these models, and to eventually elucidate the network structure, additional experiments need to be carried out.

The goal of experiment planning is to design the most informative experiments in order to identify more accurate models [7]. Previous work on this subject dealt with the design of experiments that give insight into the model structure and parameter values [8], [9], [10], [11], [12]. Bardsley et al. [13] investigated the optimal timing of measurements for time series data. Many statistical approaches [8], [14], [15], [16] have been proposed and various criteria to access the optimality of experiments like the Shannon entropy, Akaike Information Criterion, the likelihood ratio test, and different forms of the sum of squared differences [7] have been applied. Most of the existing research only handles single perturbation experiments. Only few methods for the selection of optimal sets of multiple perturbations have been proposed [17], [18] so far.

With the present study, we extend our previous contributions in the context of sign consistency and influence graph models and close the gap in the reasoning loop of model inference, consistency checking, hypothesis generation, and finally experiment design. We propose a method to guide experiment design, computing optimal sets of perturbations that allow the discrimination of the candidate models. We provide a formal characterization of the combinatorial problem, together with a solution based on Answer Set Programming [19] included within the open source

- S. Thiele, S. Heise, and S. Klamt are with the Max Planck Institute for Dynamics of Complex Technical Systems, 39106 Magdeburg, Sachsen-Anhalt, Germany. E-mail: sthiele78@gmail.com, sunny.heise@web.de, klamt@mpi-magdeburg.mpg.de.
- W. Hessenkemper, H. Bongartz, M. Fensky, and F. Schaper are with the Otto-von-Guericke University Magdeburg, Institute of Biology, Universitätsplatz 2, 39106 Magdeburg, Sachsen-Anhalt, Germany. E-mail: {wiebke.hessenkemper, hannes.bongartz, melissa.fensky, fred.schaper}@ovgu.de.
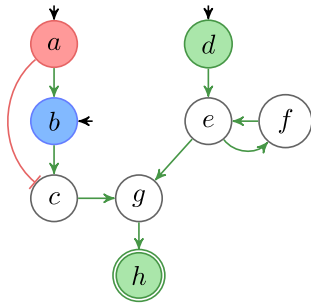
Fig. 1. Example of an interaction graph with experiment data. Input variables are marked with a small black arrow. The color indicates the induced or measured difference between two system states, green (increase), red (decrease), and blue (0-change). The readout node is marked with a double circle.

python package `ExDesi`, which is freely available for download. We applied our method to an *in silico* case study using simulated experiment data from an ODE model of a realistic signaling network and show that an optimal model identical to the gold standard can be inferred. Further, we used the approach to plan *in vivo* experiments to discriminate possible models of the Erythropoietin signal transduction in human embryonal kidney cells `HEK293`.

2   METHODS

Interaction Graph. An *interaction graph* (IG) is a directed graph (V, E, σ), where V is a set of nodes, E a set of edges, and $\sigma : E \to \{+, -\}$ a labeling of the edges. Every node in V represents a species in the modeled system and an edge $i \to j$ means that the change of i in time influences the activity or concentration level of j. Every edge $i \to j$ of an influence graph can be labeled with a sign, either $+$ or $-$, denoted by $\sigma(i, j)$, where $+$ resp. $-$ indicates that i tends to increase resp. decrease the level or activity of j. An example influence graph is given in Fig. 1.

Interaction graphs are an abstraction of dynamic quantitative systems where a quantitative state of the system at time point t is a mapping $s_t : V \to \mathbb{R}^+$. Sign consistency methods use the signs $\{+, -, 0\}$ to denote changes in the variables of the modeled system in the transient phase or after steady-state shift experiments. Examples for such changes could be increased or decreased metabolite concentrations or expression levels of genes. The signs $+$ and $-$ denote increase and decrease and 0 signifies no-change. Sign consistency methods relate the IG model of the system and the changes between system states by representing the differences as labels on the nodes in the graph. For example, the changes between two states of the system can be represented as a sign labeling of the IG. Given two system states s_i and s_j the differences between these states can be represented as the labeling $\mu_{ij} : V \to \{+, -, 0\}$ with $\mu_{ij}(x) = sign(s_j(x) - s_i(x))$.

Inputs and Perturbations. Biological systems react to changes in their environment and to internal manipulations. The variables of a system which are controlled externally are called *inputs* and we denote the set of nodes that represent input variables with $I \subseteq V$. Because input nodes are controlled externally, they are excluded from the sign consistency rules.

Perturbations are externally induced changes of the input variables (e.g., by gene knockouts, down regulation, or use of inhibitors). The sign of a perturbation signifies the difference to a reference state of the system. The signs of the

perturbation are defined as $pert : I \to \{+, -, 0\}$, where an increase perturbation $pert(i) = +$ (resp. decrease perturbation $pert(i) = -$) means that i is controlled such that it has a significant higher (resp. lower) value than in the reference state. A special role play so called 0-perturbations with $pert(i) = 0$. These are perturbations where the value of a variable is kept constant. Such perturbations are particular useful as one can practically ignore the outgoing edges of 0-perturbed nodes because these nodes can not be responsible for any downstream changes. It is easy to see that for complex models with many cyclic interactions such perturbations are very useful to investigate the influence of single components while excluding the influence of others.

Predicting Model Behaviors. A prediction function in the sign consistency approach is a function $pred : V \times (I \to \{+, -, 0\}) \to 2^{\{+, -, 0\}} \setminus \emptyset$. Given a vertex and a perturbation function it returns a nonempty set of possible signs. The here presented approach is similar to the *dependency matrix* [4] to make behavioral predictions for initial responses. We consider pair-wise dependencies among the variables in the IG model. The relations between two nodes are:

- i is an activator of j if there exists a positive elementary path from i to j, and
- i is an inhibitor of j if there exists a negative elementary path from i to j.

A pair of nodes i, j can be in one, both, or none of these relations. Using these relations our prediction function is defined as follows. Given a perturbation function $pert : I \to \{+, -, 0\}$ the dependency matrix defines a prediction function $pred : V \times (I \to \{+, -, 0\}) \to 2^{\{+, -, 0\}} \setminus \emptyset$ such that:

$+ \in pred(j, pert)$ if
- $pert(j) = +$, or
- $j \notin I$ and $\exists i \in I$ and $pert(i) = +$ and i is an activator of j, or
- $j \notin I$ and $\exists i \in I$ and $pert(i) = -$ and i is an inhibitor of j.

$- \in pred(j, pert)$ if
- $pert(j) = -$, or
- $j \notin I$ and $\exists i \in I$ and $pert(i) = -$ and i is an activator of j, or
- $j \notin I$ and $\exists i \in I$ and $pert(i) = +$ and i is an inhibitor of j.

$0 \in pred(j, pert)$ if
- $pert(j) = 0$, or
- $j \notin I$ and $+ \in pred(j, pert)$ and $- \in pred(j, pert)$, or
- $j \notin I$ and $+ \notin pred(j, pert)$ and $- \notin pred(j, pert)$.

In the following we simply write $pred(j)$ instead of $pred(j, pert)$. This prediction function can predict the behaviors increase = $\{+\}$, decrease = $\{-\}$, 0-change = $\{0\}$, or undetermined=$\{+, -, 0\} = *$. In principle sign consistency based methods can also make predictions as no-increase = $\{-, 0\}$, no-decrease = $\{+, 0\}$, and increase-or-decrease = $\{+, -\}$, for example when backward propagation is applied [20]. Here we apply only forward propagation (from the inputs to the readouts) and therefore do not make such predictions. Fig. 2 shows three models and their predicted reaction with respect to perturbations. It contains examples for each type of prediction. Model A predicts
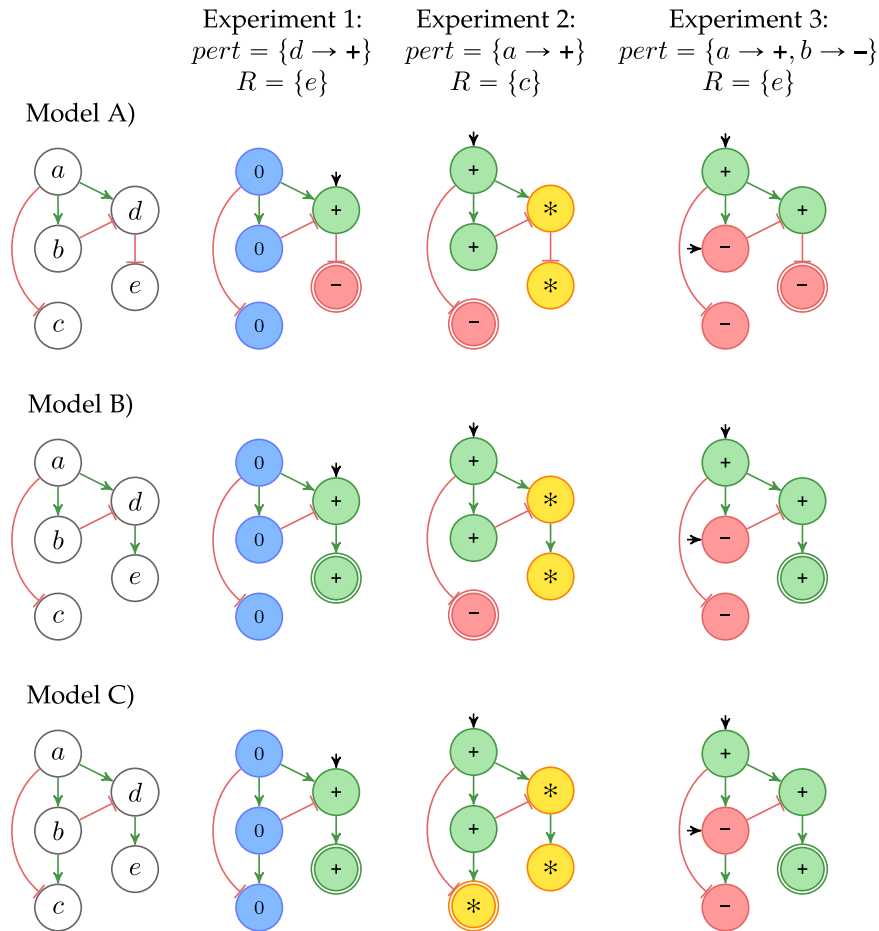
Fig. 2. Three interaction graph models which can be distinguished experimentally. Predicted behaviors and experimental conditions are represented as colors on the nodes, green (increase), red (decrease), blue (0-change), and yellow (undetermined). Perturbations are indicated with a small black arrow and readouts are shown as double circles. For Experiment 1, Model A predicts a decrease in node e, while the Models B and C predict an increase. If one measures in Experiment 1 a decrease in e one must discard Model B and C. Should the experiment result in an increase in e one must discard Model A. If the Experiment 1 results in a 0-change in e, all three models are invalidated and new network candidates must be computed. For Experiment 2, Models A and B predict a decrease in c, while Model C makes no prediction for c. If Experiment 2 results in an increase or 0-change in c, one must discard Models A and B. Should Experiment 2 result in a decrease of c one cannot discriminate any of the three models. For Experiment 3, all models predict the same behavior as for Experiment 1, but Experiment 3 uses a different set of perturbations to achieve this readout behavior.

0-change in a wrt. Experiment 1: $pred(a) = 0$, because there exists no activator nor inhibitor of a in I. With respect to Experiment 2, Model A predicts an increase in b: $pred(b) = +$, and a decrease in c: $pred(c) = -$, because a is an activator of b and an inhibitor of c and $pert(a) = +$. Finally, with respect to Experiment 2, Model A makes no prediction for d: $pred(d) = *$, because a can be an activator as well as an inhibitor of d.

Experiment. An experiment consists of two parts, the experiment setup and the measurements. The experiment setup describes which perturbations are performed, and which species will be measured. Formally, given an IG model (V, E, σ), an *experiment setup* is defined as a tuple $(I, pert, R)$ where $pert : I \to \{+, -, 0\}$ describes the performed perturbations, and $R \subseteq V$ is the set of readouts (species which are measured).

The measurements describe the observed difference in the readouts between the reference state and the measured state. Given a set of readout nodes $R \subseteq V$, the measurements define a mapping $m : R \to \{+, -, 0\}$.[1] Following the sign

consistency approach, perturbations and measurements define constraints on the labeling of the influence graph. An illustration of an experiment setup is given in Fig. 1.

Network Inference as Optional Preprocessing. The input for ExDesi is always a set of candidate models. One has either given an ad hoc set of candidate models or, what is mostly (and also in the realistic application described in Section 3) the case, an initial (text book) model and some set of (a priori) measurements from previous experiments. For the latter case, the given data are often inconsistent with the initial model structure ($m(i) \notin pred(i)$). Here, based on suitable network inference methods, these data are then used to generate consistent candidate models which then serve as input for ExDesi.

Although ExDesi is independent from the model inference method that is used for this purpose, we here generate the initial set of candidate models from prior experimental data using OPT_GRAPH [21] a sign-consistency based method for model inference. This allows us to implement the complete workflow of experiment planning using sign consistency methods. OPT_GRAPH intends to resolve inconsistencies by allowing edge removals and insertions in parallel. However, as the insertion of new interactions increases the

1. For our approach to discretization and the handling of uncertain observations, see [20].

solution space dramatically in large networks, OPT_GRAPH implements a greedy strategy which determines, in each iteration, the optimal (single) edge whose inclusion in combination with the edge removals, decreases the fitting error the most. OPT_GRAPH then adds this edge permanently and repeats until no further significant improvement can be obtained by inserting a new edge. If more than one optimal single edge is found during an iteration, then all alternative solution are tracked and explored by OPT_GRAPH.

Discriminative Power of an Experiment. If for a given experiment, two models M_1 and M_2 have in a readout $i \in R$ different predicted behaviors $(pred_{M_1}(i) \neq pred_{M_2}(i))$, we call this a *distinctive readout*. The readout in an experiment is *strongly distinctive* for two models if both models predict contradictory behaviors $(pred_{M_1}(i) \cap pred_{M_2}(i) = \emptyset)$. For example, in Fig. 2 in Experiment 1, readout e is strongly distinctive, Model A predicts a decrease while Models B and C predict an increase. When the experiment is performed the measured behavior of e will invalidate at least one of the models. In Experiment 2 in Fig. 2, the readout c is distinctive but not strongly distinctive, Models A and B predict a decrease while Model C gives an undetermined prediction. If Experiment 2 is performed the measurements may discard Models A and B if an increase or 0-change in c is measured, but if a decrease is measured then all three models are consistent.

While Experiment 1 will in any case invalidate some of the candidate models, Experiment 2 only has the potential to discriminate some candidates. We say an experiment is *definitely distinguishing* two models if at least one *strongly distinctive* readout exists (Experiment 1), and an experiment *potentially distinguishing* two networks if at least one *distinctive* readout exists but no *strongly distinctive* readout (Experiment 2). We say an experiment can not distinguish two networks if no distinctive readout exist.

We define two functions to approximate the discriminative power of an experiment E wrt. a set of candidate models.

- $def\_dis(E)$ denotes the number of pairs of candidate models which are definitely distinguished by an experiment E.
- $pot\_dis(E)$ denotes the number of pairs of candidate models which can potentially be distinguished by E.

Used as optimization criteria, we want to maximize the number of distinguished candidate models.

Experiment Costs. Apart from the discriminative power a second optimization criterion is the costs of the experiments. This means in the first place minimizing the number of perturbations. For an experiment E, we denote number of perturbations with $cost(E)$. Revisiting our example in Fig. 2, we can see that the Experiment 3 with $pert = \{a \rightarrow +, b \rightarrow -\}$ has the same discriminative power as Experiment 1, but demands more perturbations. Hence, we might prefer Experiment 1 over Experiment 3.

The Optimal Experiment Design Problem. The main goal of optimal experiment design is to find the experiments that are most likely to discard the most models. We want to maximize the number of *definitely distinguishable* models and with a lower priority maximize the number of *potentially distinguishable* models, and as a third objective we want to minimize the costs.

Formally, given a set *Cand* of candidate models, a set *Done* of already performed experiments, a function $ppert : V \rightarrow$ $2^{\{+,-,0\}}$ indicating the possible perturbations for a species, and a set R_p of possible readouts. We want to find experiments $E = (I, pert, R) \notin Done$, with $pert(i) \in ppert(i)$ for all $i \in I$, and $R \subseteq R_p$ that

1. maximize $def\_dis(E)$,
2. maximize $pot\_dis(E)$, and
3. maximize $cost(E)$.

The optimization problem can be performed using solvers that natively support hierarchical optimization. Alternatively, hierarchical optimization could be emulated with a single linear optimization function using factors α, β, γ:

$$\text{maximize } \alpha \cdot def\_dis(E) + \beta \cdot pot\_dis(E) + \gamma \cdot cost(E).$$

If the single objective function is used, the factors α, β, γ can always be chosen such that the hierarchical ordering among the solutions is ensured.

Note that the problem definition for optimal single experiments can be straightforwardly extended to optimal sets of experiments. Then we are looking for a set of experiments which maximizes the discriminative power while minimizing the number of experiments and perturbations.

While for *definite distinctions* one strongly distinctive readout is sufficient to distinguish two models, for *potential distinctions* more distinctive readouts can increase the probability of an experimental result that allows to distinguish two models. Therefore, another goal might be to maximize the overall number of *distinctive* readouts.

Relevant Perturbations. Because the measurements of a potentially distinguishing readout might not lead to the invalidation of a model, experiments containing irrelevant perturbations might be proposed in a subsequent planning step. For example, in Fig. 3, Experiment 1 aims to distinguish the Models A and B, using the measurement of readout c, but a measurement $m(c) = +$ does invalidate none of the models. Therefore, a subsequent planning step might propose Experiment 2. Although Experiment 2 leads to different predictions as Experiment 1, we want to discard Experiment 2 because the perturbation in d does not effect any distinctive readout. Therefore, we use an additional constraint to exclude perturbations that have in no candidate model a path to a distinctive readout.

Implementation. To solve the optimal experiment design problem we formulated it by means of Answer Set Programming (ASP) [22]. ASP is a declarative problem solving paradigm from the field of logic programming and knowledge representation, that offers a rich modeling language [23] along with highly efficient inference engines [24] based on Boolean constraint solving technology. We provide a python program called ExDesi that uses the ASP solver through the PyASP library. ExDesi is open source and available as part of the BioASP software collection at https://github.com/bioasp/exdesi.

3 CASE STUDY: ERYTHROPOIETIN SIGNAL TRANSDUCTION

We tested our approach to optimal experiment design using the model of the Erythropoietin signal transduction in the cell line HEK293 derived from human embryonic kidney cells. Erythropoietin (Epo) is a cytokine which initiates red blood cell generation and thus adjusts the capacity of the
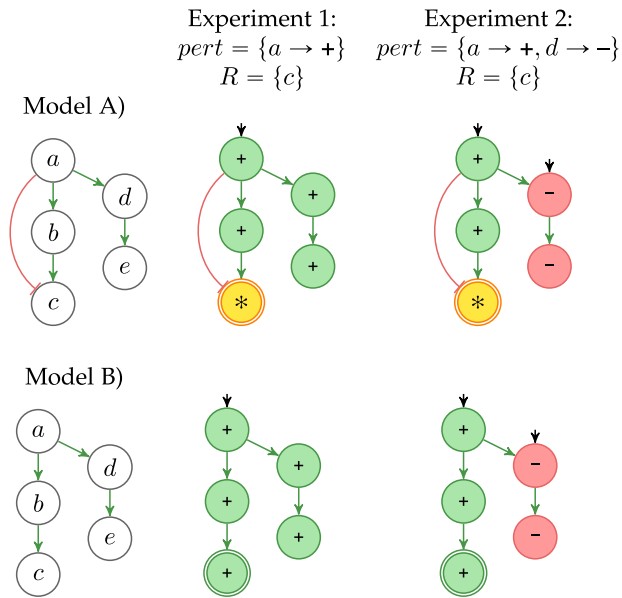
Fig. 3. Two interaction graph models which are potentially distinguished by Experiments 1 and 2. The predicted behaviors and experimental conditions are represented as colors on the nodes, green (increase), red (decrease), blue (0-change), and yellow (undetermined). Perturbations are indicated with a black incoming edge. For Experiments 1 and 2, Model A predicts an undetermined behavior in node c while Model B predicts an increase in c. Both experiments may distinguish Model A and B, but for the minimal number of perturbations, one would prefer Experiment 1. Experiment 2 is redundant, even if Experiment 1 fails to distinguish the two models (because c is measured $+$). Experiment 2 is not capable of causing a different behavior in any readout. The only difference between Experiments 1 and 2 is a perturbation in d which has in none of the models a path to a distinctive readout.

blood to transport oxygen. The Epo-encoding gene is activated in case of hypoxia in the kidney and liver by the transcription factor hypoxia induced factor (HIF). Epo acts on erythroid progenitor cells which finally develop to reticulocytes [25]. This differentiation depends on binding of Epo to the transmembrane Epo receptor. The receptor does not harbour intrinsic kinase activity but is constitutively associated with the tyrosine kinase Janus kinase 2 (JAK2). After binding of Epo to the receptor, JAK2 is tyrosine phosphorylated and thus activated. JAK2 subsequently phosphorylates and activates the transcription factors signal transducer and activator of transcription (STAT) 3 and 5. A negative feedback is implemented by the induction of the STAT-dependent feedback inhibitors SOCS3 and CIS [26]. The regulation of the STAT-independent signalling modules involves the multisite docking proteins of the Grb2 associated binder (Gab) family [27]. Gab-proteins contribute to the coordination and activation of the PI3K/AKT pathway, the MAPK-cascade, and the PLC-cascade. The detailed interconnection of these signalling modules is not fully understood, yet. In brief, Gab proteins are recruited to the plasma membrane and there act as a signalling platform by binding Grb2/SOS, SHP2, PI3K, PLC and other signalling components. Several regulatory loops are predicted as e.g., Gab binding to PIP3 in the plasma membrane depends on PI3K- and MAPK-activity [28]. The original model that served as our gold standard was given in form of a Boolean network (BN). The hypergraph representation of the Boolean network contains 23 species and 28 hyperarcs. In Boolean networks the underlying interaction graphs can easily be obtained [4].

A realistic experimental setup involves the following list of possible perturbations and readouts.

- nine possible down-regulations (e.g., by use of suitable inhibitors): akt, mek1, mtorc1, pi3k, gab1_ps, gab1_bras_py, gab1_bpi3k_py, shp2, shp2_ph,
- three possible up-regulations: gab1_ps, shp2, shp2_ph, and
- eight possible readouts: akt, erk, gab1_bshp2_ph_py, gab1_ps, jak2_p, plcg, shp2, stat5ab_py.

In the first part of this case study, we used *in silico* experiments to validate our approach by creating a distorted version of the gold standard and then planning experiments that restored the gold standard. In the second part, we planned *in vivo* experiments to refine our knowledge of the Erythropoietin signaling pathway in HEK293 cells.

3.1 *In Silico* Experiments

We tested our approach *in silico* by restoring the gold standard model of the Erythropoietin signal transduction from a distorted version of the model. We performed quantitative simulations of the proposed experiments using an ODE system which was created from the gold standard. A detailed description of the *in silico* study is given in the following:

- First, we created an ODE system from the gold standard BN using the software ODEfy [29].
- This ODE model of the gold standard was used to simulate three initial experiments and subsequently the experiments that were proposed by our experiment design procedure.

 We identified a steady state in the ODE system, which we used as initial state of the system in our simulated experiments. In order to simulate the perturbation at time point 0, we fixed the variables for the perturbed species to constants such that increased species had a value significantly higher than in the initial state, decreased species a value significantly lower than in initial state, and 0-perturbed species were kept constant as in the initial state. We restarted the simulation using these perturbed values and tracked the sign of the first change of the dependent system variables as initial response.

 The three initial experiments were 1) an increase of mek1, 2) a decrease of pi3k, and 3) the simultaneous decrease of mek1 and pi3k.

- As starting point for experiment planning and model reconstruction, we transformed the gold standard BN into an influence graph which consists of 23 nodes and 51 edges (Fig. 4a). We then created a distorted version (Fig. 4b) where we removed the edge (erk → gab1_ps) and added two new edges (mek1 ⊣ shp2) and (mek1 → stat5ab_py). These wrong/missing edges needed to be identified by the inference method.
- We used the OPT_GRAPH method for network inference on the distorted model together with the data of the initial experiments. As expected the data of the simulated experiment were consistent with our gold standard but inconsistent with the distorted IG model. OPT_GRAPH proposed six candidate models that were consistent with the initial experiment data.
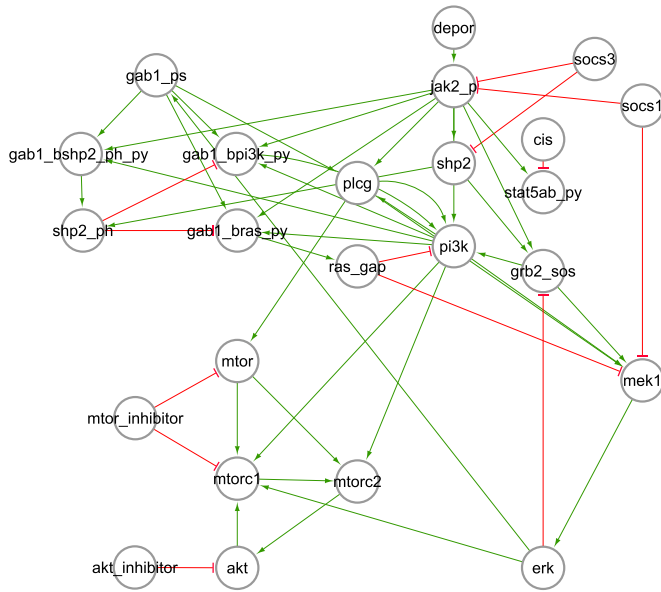
a) gold standard
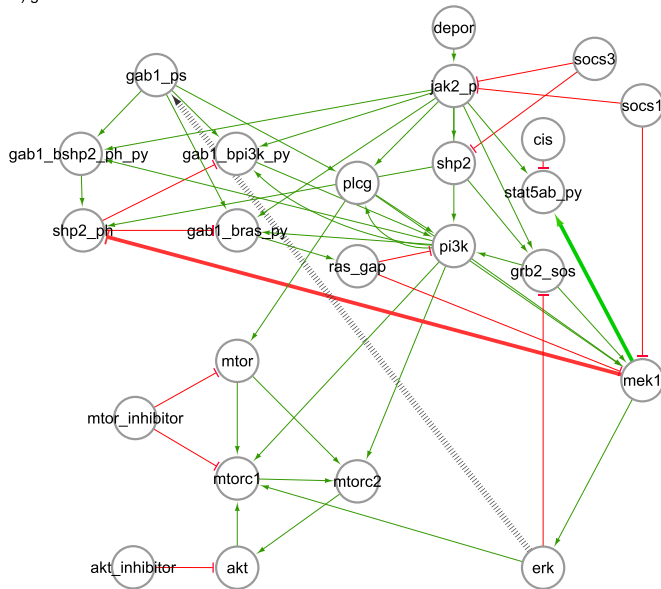
b) distorted version

Fig. 4. The influence graph model of the gold standard (a) and the distorted version (b). Added edges are drawn thicker and removed edges are striked out.

	Experiment 1 pert.	read	Experiment 2 pert.		read	Experiment 3 pert.		read
model id	akt_inhibitor	gab1_ps	gab1_bras_py	shp2	gab1_ps	erk	mek1	gab1_ps
5	+	0	−	−	*	+	−	+
4	+	0	−	−	*	+	−	−
6	+	0	−	−	+			
1	+	−						
2	+	−						
3	+	−						
				simulation results				
gold standard		0			−			+

Fig. 5. Results of the *in silico* study. Predicted behavior for the candidate models and behavior of the gold standard simulation under the proposed experiment. Predictions that are consistent with the simulation results are in green, inconsistent predictions are shown red.

Note, that in our case the gold standard IG was among those candidate models, which is not guaranteed by the OPT_GRAPH method.

- We started the first round of experiment planning, with these six candidate models including the constraints on possible perturbations and readouts. In the first round an experimental setup with one perturbation (decrease of akt via akt_inhibitor) was proposed. Fig. 5 shows the predicted behaviors for the six network candidates in the proposed experiment. The simulation results of the proposed experiment were consistent with three models (4, 5, 6) and could exclude three candidates (1, 2, 3).
- In the second round of experiment planning two experiments with two perturbations distinguishing

two classes of models were proposed. Fig. 5 shows the predicted behaviors for one of the proposed experiments (simultaneous decrease of gab1_bras_py and shp2). The simulation results of the proposed experiment were consistent with two models (4, 5) and could exclude one candidate (6).

This point marks the limit of what is currently technically possible. We have had exhausted all the perturbation and measuring methods available to the wet lab. It would not be possible to distinguish the remaining models. However our ODE model does not succumb to these restrictions as we can observe and perturb all components.

- We continued the third round of experiment planning dropping all restrictions on possible perturbations. Two experiment setups distinguishing the maximal number of two classes of models with the minimum of two perturbations were proposed. The Fig. 5 shows the predicted behaviors for one of the proposed experiments (simultaneous increase of erk and decrease of mek). Finally, the predictions of only one model is consistent with the results of the simulated experiment, model 5 which is identical to our gold standard.

3.2 *In Vivo* Experiments

For *in vivo* experimentation, we started with 26 model candidates based on the gold standard and existing experimental data.[2] Our experiment planning proposed the experiment shown in Fig. 6 in which the inhibition of akt would allow us to distinguish two classes of models and thus to invalidate at least six model candidates. These model classes are illustrated in Fig. 7. All model candidates suggest to introduce a regulation of erk that was not present in the gold standard. The models of class *A* contain regulations upstream of akt via gab1_bshp2_ph_py, gab1_bpi3k_py, plcg, gab1_bras_py, ras_gap, pi3k or mtor. If a model from class *A* reflects the biological reality, then an inhibition of akt should have no effect on erk. The models of class *B* contain a regulation of erk downstream of akt, via mtorc1, mtorc2 or akt directly. If a model of

	pert.	readout					
model id	akt_inhibitor	erk	jak2_p	plcg	gab1_ps	stat5ab_py	shp2
Class A 20 models	+	0	0	0	0	0	0
Class B 6 models	+	+	∗	∗	+	∗	∗

Fig. 6. Predicted behavior for the candidate models of the Erythropoietin signal transduction in `HEK293` cells in Experiment 1 of the *in vivo* study.

class B reflects reality we should see an increase in erk and gab1_ps.

The experiment was set up and performed according to the computed specification. `HEK293` cells expressing JAK2 wild type and the Epo receptor were pre-treated with the Akt inhibitor MK-2206 (2 μM, 30 min) or solvent control before Epo stimulation (3 U/ml, 15 min). Cell lysates were prepared and proteins were separated by SDS–PAGE. After Western blotting, the membranes were stained for phosphorylated forms of Gab1, JAK2, STAT5, PLCγ, SHP2, Erk1/2 and Akt. Three independent experiments were performed. The experimental results showed that the inhibition of akt had no effect on erk and gab1_ps. Therefore, we were able to exclude six of the model candidates and perform a second round of experiment planning trying to discriminate among the remaing 20 model candidates.

In the second round our experiment planning proposed the experiment shown in Fig. 8. The proposed inhibition of mtor distinguished two classes of models. Promising to invalidate at least two of the model candidates. The models of class C contain regulations upstream of mtor via gab1_bshp2_ph_py, gab1_bpi3k_py, plcg, gab1_bras_py, ras_gap or pi3k. If a model from class C reflects the biological reality, then an inhibition of mtor should have no effect on erk. The models of class D contain a regulation of erk via mtor. If a model of class D reflects reality we should see an increase in erk and gab1_ps.

The experiment was set up and performed according to the computed specification. `HEK293` cells expressing JAK2 wild type and the Epo receptor were pre-treated with the mTOR inhibitor Rapamycin (10 nM, 30 min) or solvent control before Epo stimulation (3 U/ml, 15 min). Cell lysates were prepared and proteins were separated by SDS–PAGE. After Western blotting, the membranes were stained for phosphorylated forms of Gab1, JAK2, STAT5, PLCγ, SHP2, Erk1/2 and Akt. Three independent experiments were performed. The experimental results showed that the inhibition of mtor had no effect on our readout species. Therefore, we were able to exclude another two of the model candidates reducing the candidate set to 18.

Continued experiment planning proposed further experiments that contained multiple perturbations which, however, were too complex to conduct with our facilities.

3.3 Testing Scalability and Sensitivity

To assess the sensitivity and scalability of our approach with respect to different settings we created 165 distorted models from the gold standard (Fig. 4a). In each distorted model, 1 edge from the gold standard has been deleted and 2 edges not contained in the gold standard have been added.
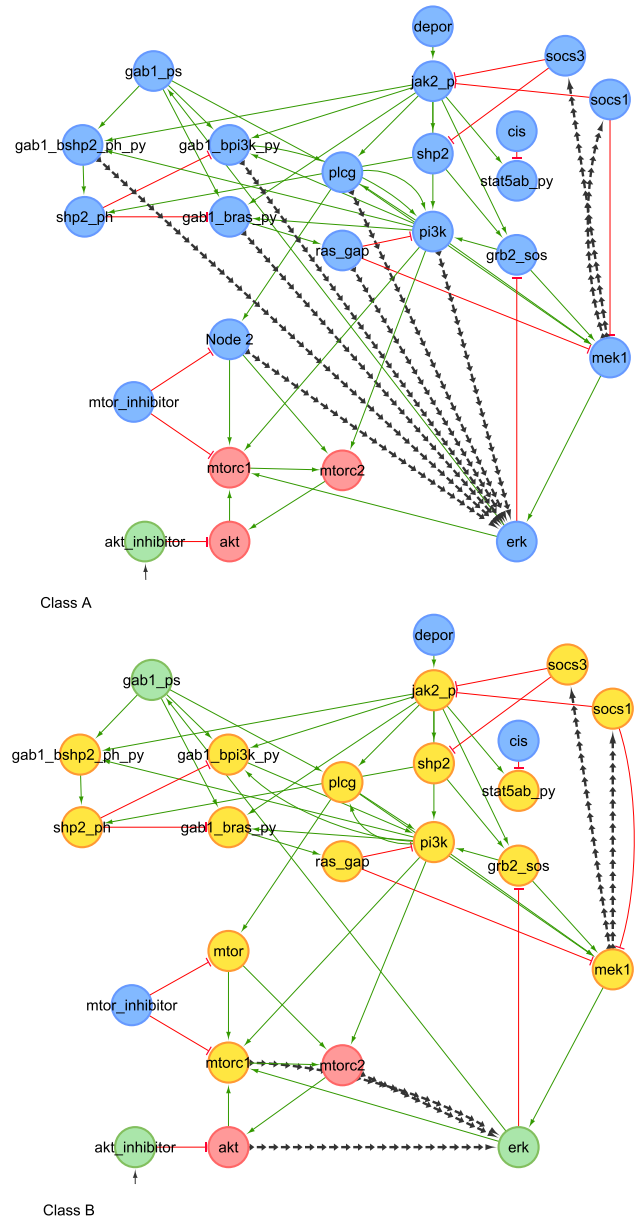

Fig. 7. Two model classes for the Erythropoietin signal transduction network in `HEK293` cells. Alternative regulations in different models are shown with black arrows. Class A contains regulations of erk upstream of akt. Class B contains regulations of erk downstream of akt. Models of class A predict no effect of an inhibition of akt, while models of class B predict an increase of erk and gab1_ps.

For each distorted model we used the OPT_GRAPH method together with the existing real experimental data and computed candidate models that correct the inconsistencies with the experimental data. Fig. 9 shows a histogram for the

	pert.	readout					
model id	mtor_inhibitor	erk	jak2_p	plcg	gab1_ps	stat5ab_py	shp2
Class C 18 models	+	0	0	0	0	0	0
Class D 2 models	+	+	∗	∗	+	∗	∗

Fig. 8. Predicted behavior for the candidate models of the Erythropoietin signal transduction in `HEK293` cells in Experiment 2 of the *in vivo* study.
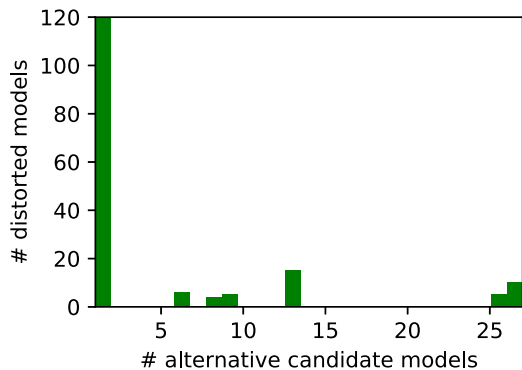
Fig. 9. Number of candidate models proposed for each of the 165 distorted models. For 120 distorted models, OPT_GRAPH proposed a single candidate model while for the remaining 45 distorted models up to 27 repair candidates were identified.

number of candidate models that have been proposed for the distorted models.

42 of the 165 distorted models remained consistent with the data, and another 78 distorted models were found inconsistent but had only a single optimal repair candidate proposed. These singleton candidate sets required no experimental design for discrimination. The remaining 45 distorted models exhibit inconsistencies with the experiments and OPT_GRAPH proposed more than one (between 6 and 27) alternative candidate models that could then be treated with our experimental design approach.

We tested our method with different parameter settings. Starting with the realistic setup of 8 possible readouts and 13 possible perturbations we created further scenario combinations, where all 23 nodes are considered as readouts or/and with increased number (20 instead of 13) of possible perturbations (the additional perturbations were randomly selected).

For each computed optimal experiment, we recorded how many perturbations and how many alternative experiments were proposed and how good the experiment could distinguish among the candidates. For the latter, we computed for each proposed experiment the portion of pairs of candidate networks that could be distinguished. Given the number n of candidates, $all\_pairs = n \times (n-1)/2$ is the number of all possible candidate pairs. With the number def_dis of pairs that could be definitely distinguished with the proposed optimal experiments we computed the ratio def_dis/all_pairs. Note that in practice an experiment that

can distinguish one network from the rest can be enough to discriminate $n - 1$ candidate pairs. Such an experiment has thus a discrimination ratio of $2(n - 1)/(n \times (n - 1))$. Hence, with increasing n the ratio of such experiments approaches zero. On the other hand, the more network candidates one has the *easier* it will be to discriminate at least some of them.

Table 1 shows the effect of the different settings on the ratio of distinguishable candidate pairs, the number of necessary perturbations, and the number of proposed alternative experiments.

Increasing the number of possible readouts had no significant effect on the runtime of the optimal experiment computation but, as expected, resulted in better experiments in the sense that they allow distinguishing more model candidates (sometimes requiring more perturbations). Overall, less alternative optimal experiments are proposed when the number of readouts increased. Hence, the optimal experimental design strategies become more constrained.

Increasing the number of possible perturbation points from 13 to 20 resulted in experiments by which the model candidates can be better distinguished, however, the effect was much less pronounced compared to the scenario where all nodes are considered as readout nodes. Moreover, in contrast to a higher number of readouts, increasing the number of perturbations has an effect on the size of the search space of optimal solutions and thus also on the runtime of the computation. In fact, further increasing the number of possible perturbations in the benchmarks (beyond 20 as used herein) resulted partially in computation times that exceeded our time limit (1 day). These results suggest, at least for our case study, investments in readouts are more rewarding than increasing the number of perturbations. Importantly, for all 45 sets of candidate models, at least one discriminating experiment could be proposed.

Further conclusions regarding the benchmark study are given in the following Discussion section.

4   DISCUSSION

We presented a new approach to experiment planning with multiple perturbations in the context of interaction graph models. We demonstrated *in silico* that our planning approach proposes experiments that are suitable to restore a gold standard model from a distorted model. Further, we showed *in vivo* that our approach proposes experiments that allowed us to systematically reduce our space of possible

TABLE 1
Results from the Randomized Benchmark Calculations for Four Different Scenarios Which Are a Combination of Either 13 Realistic Perturbations or an Extended Set of 20 Perturbations with Either Eight Actually Used Readouts or All 23 Nodes as Readouts

scenario		a) % definitely/% potentially distinguishable candidate pairs	b) # necessary perturbations	c) # alternative experiments
# readouts	# pert.			
real (8)	real (13)	12%/64%	2.34	1.8
real (8)	more (20)	14%/54%	3.02	1.76
all (23)	real (13)	53%/27%	3.36	1
all (23)	more (20)	64%/18%	5.42	1.62

All numbers present the average over all 45 candidate sets under the corresponding readout/perturbation scenario. Column a) shows the percentage of definitely distinguishable pairs of candidates/the average percentage of (only) potentially distinguishable candidate pairs. Column b) shows the number of necessary perturbations (per experiment) proposed by the optimal experiments. Column c) shows the average number of alternative optimal experiments found (per candidate set).
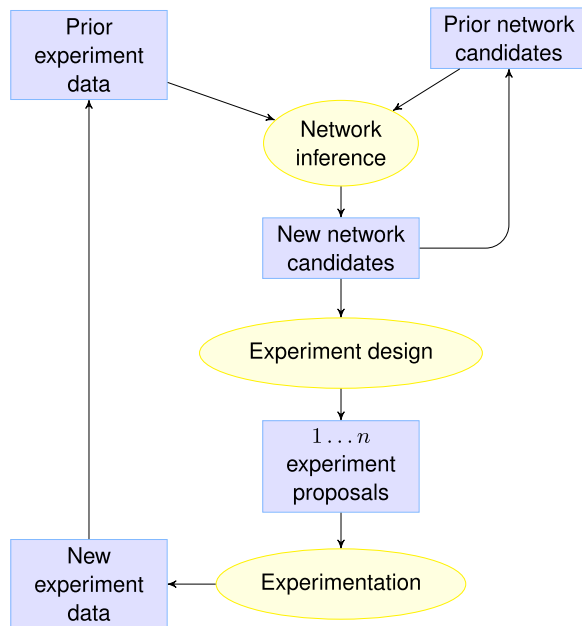
Fig. 10. Workflow of experiment planning.

models for the Erythropoietin signal transduction in `HEK293` cells and to exclude certain interactions.

In an empirical study, we showed that while the approach is sensitive to the number of possible perturbations, it works robustly on problems of realistic sizes. For all the tested candidate sets at least one discriminating experiment could be proposed. `ExDesi` is easily scalable in the number of possible readouts which, as expected, also leads to much better results. In contrast, the approach has only limited scalability in the number of possible perturbations since increasing those leads to an exponential growth of the search space. We are confident that future research and the progress in the used solver technology will allow us to scale this approach to problems with larger search spaces.

Generally, as holds true for many model discrimination methods, models which are underconstrained (e.g., because there are too many candidate edges in the initial graph) can usually not be falsified because the model explains all observed behaviors. Thus, it is very important to start with a sparse model where (most) included interactions have a relatively high confidence.

Only few methods for the selection of optimal sets of multiple perturbations have been proposed so far. `MEED` [17] is an approach that uses ternary logic networks and microarray data. `Caspo` [18] on the other hand uses Boolean logic networks and Boolean data (on and off) within a single state of the biological system. In contrast, our method is based on interaction graphs and uses data on (signed) changes (ups and downs between two states) as response to perturbations. `ExDesi` is to a certain degree related to `Caspo` as both work on topological similar models (Boolean networks transformed to interaction graphs), but one cannot easily extract a Boolean state from data expressing signed changes. For example the sign for an increased value in a variable can be interpreted as a Boolean variable that switches from an OFF state to an ON state, but depending on the thresholds used to mark the border between ON and OFF it could also refer to a variable that remains in its initial OFF/ON state.

The presented approach to experiment planning integrates nicely in the systems biology work-flow of experimentation, data analysis, hypothesis generation and model inference (Fig. 10). One usually starts the process with some prior knowledge and experimental data. Using network inference methods one obtains one or more candidate models. These network candidates together with information about already performed experiments and possible perturbations and readouts are fed into the experiment design process, which proposes new experiments that are suitable to discriminate among these candidates. The resulting experiments are performed and the experimental results are compared with the predicted model behaviors. Models whose predicted behaviors are inconsistent with the experimental result must be discarded and the new experimental data is added to the network inference process to produce new model candidates. Ideally, this process can be re-iterated until one is left with a single model candidate. In the less optimistic scenario one may be left with a set of indistinguishable model candidates, then other methods must come into play.

Looking at the broader context of experiment planning we see further optimization potential, for example the optimal utilization of resources which so far have not been considered. When the resources to perform multiple experiments in parallel exist and if the experiments are time-consuming it will be desirable to conduct multiple experiments at the same time.

ACKNOWLEDGMENTS

REFERENCES

[1] C. Soulé, "Graphic requirements for multistationarity," *Complexus*, vol. 1, no. 3, pp. 123–133, 2003.
[2] C. Soulé, "Mathematical approaches to differentiation and gene regulation," *Comptes Rendus Biologies*, vol. 329, pp. 13–20, 2006.
[3] É. Remy, P. Ruet, and D. Thieffry, "Graphic requirements for multistability and attractive cycles in a Boolean dynamical framework," *Advances Appl. Math.*, vol. 41, no. 3, pp. 335–350, 2008.
[4] S. Klamt, J. Saez-Rodriguez, J. Lindquist, L. Simeoni, and E. Gilles, "A methodology for the structural and functional analysis of signaling and regulatory networks," *BMC Bioinf.*, vol. 7, no. 1, 2006, Art. no. 56.
[5] M. Gebser, C. Guziołowski, M. Ivanchev, T. Schaub, A. Siegel, S. Thiele, and P. Veber, "Repair and prediction (under inconsistency) in large biological networks with answer set programming," in *Proc. 12th Int. Conf. Principles Knowl. Representation Reasoning*, 2010, pp. 497–507. [Online]. Available: http://aaai.org/ocs/index.php/KR/KR2010/paper/view/1334
[6] M. Gebser, T. Schaub, S. Thiele, and P. Veber, "Detecting inconsistencies in large biological networks with answer set programming," *Theory Practice Logic Program.*, vol. 11, no. 2/3, pp. 323–360, 2011.
[7] C. Kreutz and J. Timmer, "Systems biology: Experimental design," *FEBS J.*, vol. 276, no. 4, pp. 923–942, 2009.
[8] B. Mélykúti, E. August, A. Papachristodoulou, and H. El-Samad, "Discriminating between rival biochemical network models: Three approaches to optimal experiment design," *BMC Syst. Biol.*, vol. 4, no. 1, pp. 1–16, 2010.
[9] X. j. Feng and H. Rabitz, "Optimal identification of biochemical reaction networks," *Biophysical J.*, vol. 86, no. 3, pp. 1270–1281, 2004.
[10] C. L. Barrett and B. O. Palsson, "Iterative reconstruction of transcriptional regulatory networks: An algorithmic approach," *PLoS Comput. Biol.*, vol. 2, 2006, Art. no. 6.

[11] R. Gunawan, K. G. Gadkar, and F. J. Doyle, "Methods to identify cellular architecture and dynamics from experimental data," in *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. Cambridge, MA, USA: MIT Press, 2006.

[12] S. Bandara, J. P. Schlöder, R. Eils, H. G. Bock, and T. Meyer, "Optimal experimental design for parameter estimation of a cell signaling model," *PLoS Comput. Biol.*, vol. 5, 2009, Art. no. 1000558.

[13] W. G. Bardsley, R. M. W. Wood, and E. M. Melikhova, "Optimal design: A computer program to study the best possible spacing of design points," *Comput. Chemistry*, vol. 20, pp. 145–157, 1996.

[14] B. H. Chen and S. P. Asprey, "On the design of optimally informative dynamic experiments for model discrimination in multiresponse nonlinear situations," *Ind. Eng. Chemistry Res.*, vol. 42, pp. 1379–1390, 2003.

[15] F. P. Casey, D. Baird, Q. Feng, R. N. Gutenkunst, J. J. Waterfall, C. R. Myers, K. S. Brown, R. A. Cerione, and J. P. Sethna, "Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model," *IET Syst. Biol.*, vol. 1, pp. 190–202, 2007.

[16] E. Balsa-Canto, A. A. Alonso, and J. R. Banga, "Computational procedures for optimal experimental design in biological systems," *IET Syst. Biol.*, vol. 2, no. 4, pp. 163–172, Jul. 2008.

[17] E. Szczurek, I. Gat-Viks, J. Tiuryn, and M. Vingron, "Elucidating regulatory mechanisms downstream of a signaling pathway using informative experiments," *Mol. Syst. Biol.*, vol. 5, no. 1, 2009, Art. no. 287.

[18] S. V. I. Konokotina, L. G. Alexopoulos, J. Saez-Rodriguez, T. Schaub, A. Siegel, and C. Guziolowski, "Designing experiments to discriminate families of logic models," *Front. Bioeng. Biotechnol.*, vol. 3, 2015, Art. no. 131.

[19] M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub, *Answer Set Solving in Practice*. San Rafael, CA, USA: Morgan and Claypool Publishers, 2012.

[20] S. Thiele, L. Cerone, J. Saez-Rodriguez, A. Siegel, C. Guziołowski, and S. Klamt, "Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies," *BMC Bioinf.*, vol. 16, 2015, Art. no. 345.

[21] I. N. Melas, R. Samaga, L. G. Alexopoulos, and S. Klamt, "Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs," *PLoS Comput. Biol.*, vol. 9, no. 9, Sep. 2013, Art. no. e1003204.

[22] C. Baral, *Knowledge Representation, Reasoning and Declarative Problem Solving*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[23] M. Gebser, R. Kaminski, B. Kaufmann, M. Ostrowski, T. Schaub, and S. Thiele, "A user's guide to gringo, clasp, clingo, and iclingo," 2010. [Online]. Available: http://potassco.sourceforge.net

[24] M. Gebser, B. Kaufmann, A. Neumann, and T. Schaub, "Conflict-driven answer set solving," in *Proc. 20th Int. Joint Conf. Artif. Intell.*, 2007, pp. 386–392.

[25] M. Lombardero, K. Kovacs, and B. W. Scheithauer, "Erythropoietin: A hormone with multiple functions," *Pathobiology*, vol. 78, pp. 41–53, 2011.

[26] T. D. Richmond, M. Chohan, and D. L. Barber, "Turning cells red: Signal transduction mediated by erythropoietin," *Trends Cell Biol.*, vol. 15, no. 3, pp. 146–155, 2005.

[27] F. U. Wöhrle, R. J. Daly, and T. Brummer, "Function, regulation and pathological roles of the Gab/DOS docking proteins," *Cell Commun. Signaling*, vol. 7, 2009, Art. no. 22.

[28] A. Wolf, R. Eulenfeld, H. Bongartz, W. Hessenkemper, P. C. Simister, S. Lievens, J. Tavernier, S. M. Feller, and F. Schaper, "MAPK-induced Gab1 translocation to the plasma membrane depends on a regulated intramolecular switch," *Cellular Signalling*, vol. 27, pp. 340–352, 2015.

[29] J. Krumsiek, S. Pölsterl, D. M. Wittmann, and F. J. Theis, "Odefy—From discrete to continuous models," *BMC Bioinf.*, vol. 11, 2010, Art. no. 233.

Sven Thiele received the diploma degree in informatics and the Dr. rer. nat. (PhD) degree from the University of Potsdam, Germany, in 2007 and 2012, respectively. As a postdoctoral researcher, he worked with Inria, the French National Institute for computer science and applied mathematics in Rennes, France. Currently, he works with the Analysis and Redesign of Biological Networks Research Group at the Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg, Germany. His research interests lie in the logic modeling and analysis of biological systems.

Sandra Heise received the diploma degree in biosystems engineering from Otto von Guericke University Magdeburg, Germany, in 2011. She worked from 2011 to 2015 with the Analysis and Redesign of Biological Networks Research Group at the Max Planck Institute in Magdeburg. Her research focuses on development of algorithms for inference gene regulation networks and modeling of signal transduction networks in cells.

Wiebke Hessenkemper received the diploma degree in 2010 and the Dr. rer. nat. (PhD) degree, in 2014 before joining the Department of Systems Biology, Otto von Guericke University Magdeburg, Germany, as a postdoctoral research associate. She studied biology with the Friedrich Schiller University Jena, Germany. She performed the experimental parts of her PhD thesis concerning nuclear hormone receptors at Jena University Hospital in Germany and the Erasmus Medical Center in Rotterdam, The Netherlands. Her present research interests include focused on cytokine-mediated signaling pathways and especially the role and molecular mechanism of the multi-site docking protein Gab1 in coordinating signal transduction.

Hannes Bongartz received the MS degree, in 2013. He is working toward the PhD degree in the Department of Systems Biology elucidating the role of the multi-site docking protein Gab1 in myeloproliferative neoplasm-associated Jak2-V617F-positive cells using biochemical and molecular biological approaches. He studied biosystems engineering with the Otto von Guericke University Magdeburg, Germany.

Melissa Fensky received the MS degree, in 2015. She studied biosystems engineering with the Otto von Guericke University Magdeburg, Germany, and worked for her thesis with the Department of Systems Biology. Her work focused on Gab1 in Jak2-V617F-induced signal transduction. Currently, she holds a position with the pharmaceutical company IDT Biologika in Dessau-Roßlau, Germany, and is working on physico-chemical quality control processes.

Fred Schaper received the diploma degree in biology and the Dr. rer. nat. (PhD) degree from the Technical University Carolo Wilhelmina of Braunschweig, Germany, in 1992 and 1996, respectively. The practical parts of both studies have been realized with the German Research Centre for Biotechnology (GBF) Braunschweig, Germany. He became junior research group leader with the Department of Biochemistry and Molecular Biology, RWTH University, Aachen, Germany in 1996, received the venia legendi for biochemistry and molecular biology in 2002 and became adjunct professor at the same location in 2005. He has been appointed to a full professor (W3) with the Department of Biology, Otto von Guericke University Magdeburg, Germany and chairs the Department of Systems Biology since 2010. His interests include focused on the regulatory and dynamic aspects of IL-6 signaling and the crosstalk of IL-6 with other cytokines and hormones.

Steffen Klamt received the diploma degree in applied systems science from the University Osnabrück, Germany, in 1998, and the PhD degree (Dr.-Ing.) from the University of Stuttgart, in 2005. From 1998-2008, he was research assistant (PhD student) and then postdoc with the Systems Biology Group, Max Planck Institute for Dynamics of Complex Technical Systems in Magdeburg, Germany. Since 2009, he has been the head of the Analysis and Redesign of Biological Networks Research Group at the same institute. His research interests include computational methods for systems biology and systems metabolic engineering and his group develops and applies algorithms for the analysis, inference, and targeted modification of metabolic and signal transduction networks in the cell.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.