Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Designing genetic perturbation experiments for model selection under uncertainty ⋆

Eve Tasiudi * **Claude Lormeau** * **Hans-Michael Kaltenbach** *
Jörg Stelling *

* *Department of Biosystems Science and Engineering, and SIB Swiss Institute of Bioinformatics, ETH Zurich, 4058 Basel, Switzerland (e-mails: eve.tasiudi@bsse.ethz.ch, claude.lormeau@bsse.ethz.ch, michael.kaltenbach@bsse.ethz.ch, joerg.stelling@bsse.ethz.ch).*

Abstract: Deterministic dynamic models play a crucial role in elucidating the function of biological networks. However, the underlying biological mechanisms are often only partially known, and different biological hypotheses on the unknown molecular mechanisms lead to multiple potential network topologies for the model. Limitations in generating comprehensive quantitative data often prevent identification of the correct model topology and additionally leave substantial uncertainty about a model's parameter values. Here, we introduce an experiment design method for model discrimination under parameter uncertainty. We focus on genetic perturbations, such as gene deletions, as our possible experimental interventions. We start from an initial dataset and a single model whose topology includes all different hypotheses. We obtain the set of models compatible with the initial dataset, their posterior probabilities, and the distribution of compatible parameter values using our previously published topological filtering approach. We then employ a fully Bayesian approach to identify the genetic perturbation that yields the maximal expected information gain in a subsequent experiment. This approach explicitly accounts for parameter uncertainty; it also naturally allows comparing an arbitrary number of candidate models simultaneously. In contrast to previous approaches, our intervention alters the topology of the dynamic system rather than selecting optimal inputs, observables, or time-points for measurements. We demonstrate its applicability with an in-silico study based on a published real-world biological example.

Keywords: Systems biology; Experimental design; Model selection; Uncertainty quantification; Gene networks.

1. INTRODUCTION

Compared to engineered systems, models of biological systems are associated with substantially higher uncertainties, both regarding mechanisms and their quantitative characteristics. For example, high uncertainties arise because experimental characterization methods in biology are usually imprecise or not comprehensive, such that systems become often non-identifiable; quantitative characteristics of biological components can also vary significantly depending on the context, such as the cell type they operate in. Model development for biological systems therefore relies critically on uncertainty quantification and model selection to deal with uncertain model parameters and model structures, respectively (Hug et al., 2016; Sunnåker and Stelling, 2016).

To reduce such uncertainties by conducting experiments with high information content, various optimal experimental design (OED) methods have been developed and applied in systems biology. Classical methods rely on local approximations of the effect of parameter variations on system behavior, usually via first-order sensitivities and the Fisher Information Matrix for dynamic systems models (Chakrabarty et al., 2013). However, for highly non-linear biological systems, these approximations are often ill-suited; in addition, corresponding experimental design methods nearly exclusively address the parameter identification, and not the model selection problem.

Alternative, more recent methods therefore aim at a full characterization of uncertainties by relying on the Bayesian framework. It allows for systematic updating of prior distributions on parameters and models through the integration of new (observational or *in-silico*) data, and for quantification–and thus optimization–of the information gain due to the update (Liepe et al., 2013). Because of its computational efficiency, Approximate Bayesian Computation (ABC; Beaumont (2019); Vyshemirsky and Girolami (2008); Ryan et al. (2016)) has gained increased attention as a basis for developing experimental design methods. When combined with guarantees on near-optimal design results that allow for greedy searches of the design space, ABC-based methods can also become computationally feasible for detailed, medium-scale models of biological systems (Busetto et al., 2013).

Current ABC-based OED for biological systems, however, is limited in two important aspects. First, as for classical

⋆ This publication is supported by the Swiss National Science Foundation as part of the NCCR Molecular Systems Engineering.

approaches, only few methods exist that perform OED for model selection (Toni et al., 2009; Vanlier et al., 2014; Busetto et al., 2013) and OED methods designed for improving parameter estimation are known to perform poorly for model selection (Ryan et al., 2016). Second, the experimental design space of current OED methods includes selection of inputs, observables, or parametric perturbations, but not structural perturbations to the network topology. While the former clearly have relevance due to technological advances, for example, in microfluidics to generate complex dynamic inputs (Braniff and Ingalls, 2018), the latter are powerful tools for inferring the structure of biological networks, for example, by studying the effects of single or multiple gene deletions (Costanzo et al., 2016).

To address this gap, here we therefore suggest a Bayesian method for the design of genetic perturbation experiments, such as gene deletions, that aims at model selection. Specifically, we present a computationally efficient ABC-based design method that determines the expected information gain of such experiments. We evaluate the approach for a gene regulatory network in budding yeast, for which alternative models have been established and selected previously (Milias-Argeitis et al., 2016). This example application demonstrates feasibility of our proposed approach for perturbation design that accounts for model topology and parameter uncertainties.

2. PERTURBATION DESIGN METHOD

In the following, we first give a precise definition of the experimental design problem, and then describe the theory (and some aspects of computation) underlying our proposed method.

2.1 Problem definition

Assuming that molecular copy numbers are sufficiently high to neglect stochastic effects in biochemical reactions, we can capture the dynamic behavior of cellular networks by a parametric system of nonlinear ordinary differential equations (ODEs). We consider a set $\mathcal{M}$ of such models, with each model $M \in \mathcal{M}$ of the form:

$$M: \frac{dx(t)}{dt} = f(x(t), u(t), \theta),\ x(t_0) = x_0\ , \tag{1}$$

where $x(t) \in \mathbb{R}^{n_x}_{\geq 0}$ is the vector of the dynamic, non-negative concentrations of the n_x molecular species, $f(x(t), u(t), \theta)$ is a system of n_x functions that define the rate of change of the species concentrations depending on the current system state $x(t)$ and on the parameter vector $\theta \in \Theta \subseteq \mathbb{R}^{n_p}_{\geq 0}$ capturing the n_p physical constants associated with the biochemical reactions in parameter space Θ, and on the n_u inputs $u(t) \in \mathbb{R}^{n_u}$. We will write, for example, $n_p(M)$ and $\theta(M)$ if the model is not clear from the context.

Experimental observations then relate to system states via a measurement model. For simplicity (and motivated by noise sources in microarray experiments as in our application study; see below), we assume it to be of the form:

$$y(t_j) = Hx(t_j)(1+\epsilon_m)+\epsilon_a, \epsilon_m \sim N(0, \Sigma_m),\ \epsilon_a \sim N(0, \Sigma_a) \tag{2}$$

where the matrix H maps system states to observations $y(t_j)$ at time points t_j, and we assume multiplicative measurement noise ϵ_m and additive measurement noise ϵ_a drawn from multivariate Normal distributions with mean zero and covariance matrices Σ_m and Σ_a, respectively. A dataset $\mathcal{D}$ is a collection of such measurements at one or more timepoints.

Here, we consider experimental design for genetic interventions, such as deletions of genes, that structurally change the biochemical reaction network. For each model M, an intervention Δ is given as a subset $\Delta(M) \subseteq \{1, \ldots, n_p(M)\}$ of parameters set to zero, and we define the resulting parameter vector with elements $\theta_\Delta(i)$ as:

$$\theta_\Delta(i) := \begin{cases} 0, & \text{if } i \in \Delta(M) \\ \theta(i), & \text{else.} \end{cases} \tag{3}$$

For example, by setting the maximal transcription rate parameter for a gene to zero, the model's steady-state represents the effect of a gene deletion (mRNA and protein for this gene are not present).

Given a set $\mathcal{M}$ of candidate models together with known initial conditions (x_0), an input function ($u(t)$), a set of already acquired experimental observations $\mathcal{D}$ that each of the models can describe, and a set of possible interventions $\mathcal{I}$, we aim to find the intervention that maximizes the expected information gain for model selection if this intervention experiment was performed next.

2.2 Theory

To systematically represent uncertainties, we cast model selection and experimental design in a Bayesian framework. Using Bayes' theorem, we first determine the posterior probability of each model $M \in \mathcal{M}$ given $\mathcal{D}$ as

$$Q(M) := P(M|\mathcal{D}) = P(\mathcal{D}|M)\, P(M)\ /\ P(\mathcal{D})\ , \tag{4}$$

where $P(\mathcal{D}|M)$ and $P(M)$ are the model's likelihood and prior probability, respectively, and $P(\mathcal{D})$ is the Bayesian evidence. We obtain this posterior by marginalizing over the parameters θ:

$$P(\mathcal{D}|M) = \int_{\Theta(M)} P(\mathcal{D}|M, \theta)\, P(\theta|M)\, d\theta\ . \tag{5}$$

To determine these quantities computationally, we use approximate Bayesian computation (Toni et al., 2009), which relies on defining a viable region(s) in parameter space

$$\mathcal{V}(M, \mathcal{D}) = \{\theta \mid P(\mathcal{D}|M, \theta) \geq \tau\} \tag{6}$$

using a threshold τ derived, for example, by a χ^2-test. We approximate the viable region(s) of M using samples from its parameter space and we say a sample is viable if it fulfills the condition on the likelihood (Eq. 6). We then use an efficient Monte-Carlo method (Zamora-Sillero et al., 2011) to generate N viable parameter samples $\theta^{(l)}$, with $l \in [1, N]$, distributed uniformly over the region(s), which allows us to approximate the probabilities in Eq. 5 based on a uniform prior on the parameters. We used the estimated Monte Carlo integration accuracy to choose N for accurate approximation of the viable region(s). The normalizing constant $P(\mathcal{D})$ (from Eq. 4) is determined by the relation $\sum_{M \in \mathcal{M}} P(M|\mathcal{D}) = 1$.

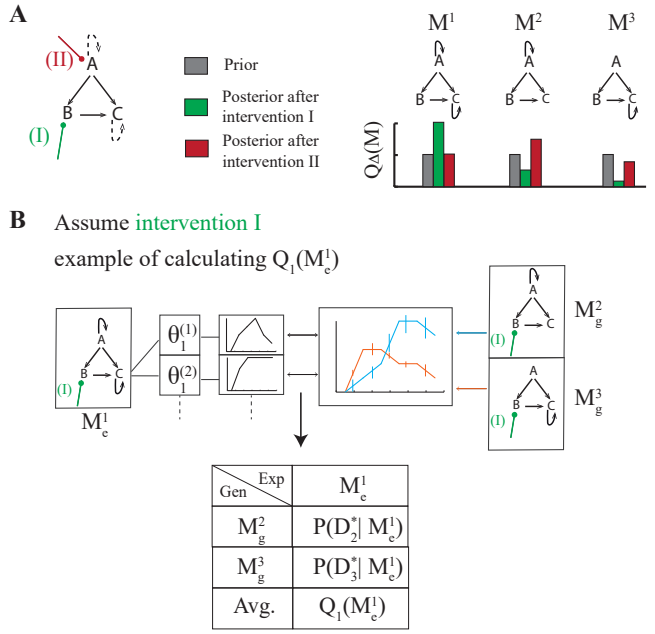

Fig. 1. Schematic overview of obtaining a model posterior probability for an intervention. (**A**) Starting from a prior model probability distribution, we want to identify the intervention that yields a highly informative update such as intervention I in this example. Here we use a simple model with two hypothesized interactions (dashed lines leading to 3 candidate models), well defined interactions are represented by solid lines. Interventions are represented with node-headed colored lines. (**B**) For a given intervention, the explanatory model M_e is simulated for each parameter sample (left). The remaining candidate models in the set are treated as generator models M_g (right). For each generator model the mean and standard deviation at specific time points for the assumed observable is obtained (center) using its parameter samples. The likelihood of each explanatory model M_e with a given parameter sample of generating the data from M_g is then calculated. The marginal likelihoods of M_e are averaged over all generating models (bottom) and Q_Δ is calculated by Eq. 8.

In analogy to Busetto et al. (2013), we quantify the information gain of an intervention $\Delta \in \mathcal{I}$ by the Kullback-Leibler divergence between the updated posterior distribution Q_Δ on $\mathcal{M}$ after applying the intervention and a reference distribution π:

$$D_{KL}\left[Q_\Delta||\pi\right] = \sum_{M\in\mathcal{M}} Q_\Delta(M)\log_2\left(\frac{Q_\Delta(M)}{\pi(M)}\right). \quad (7)$$

The KL-divergence corresponds to the expected number of additional bits of information that would be lost if the intervention experiment was not performed, and informative experiments for model selection should have high divergence (see Fig. 1A). For $\pi(M)$, we use either (i) the posterior distribution Q to favor exploration of the model space, or (ii) a uniform distribution $\mathcal{U}$ to favor convergence to one model.

To estimate Q_Δ, we in turn consider each model as the 'correct' model, and generate a corresponding *in-silico* dataset from it as follows (see Fig. 1A): let $M_g \in \mathcal{M}$ be the generator model; we modify its N parameter samples $\theta^{(l)}$ to $\theta^{(l)}_\Delta$ by applying intervention Δ according to Eq. 3. For each of these modified parameter samples, we first simulate the system to obtain trajectories $x(t)$ and apply a measurement model (Eq. 2), leading to predicted observations $y(t_j)$. The elements of the dataset $\mathcal{D}^*_g$ are then constructed by averaging the resulting N observations for each timepoint and using their standard deviation as 'measurement noise'. This accounts for both measurement uncertainty and parametric uncertainty in M_g.

Next, we evaluate how well $\mathcal{D}^*_g$ is captured by each of the remaining models (see Fig. 1B). For each such explanatory model $M_e \in \mathcal{M}\backslash\{M_g\}$, we calculate the likelihood $P\left(\mathcal{D}^*_g|M_e\right)$ by marginalizing the likelihood $P\left(\mathcal{D}^*_g|M_e,\theta_\Delta\right)$ similar to Eq. 5. Let $\theta^{(l)}$ be N uniform samples of the original viable region(s) $\mathcal{V}(M_e,\mathcal{D})$, and let $\theta^{(l)}_\Delta$ be the corresponding samples after applying the intervention Δ. Further, let $\theta^{*(l)}_\Delta$ be the N^* samples of $\theta^{(l)}$ that are also viable in the new dataset $\mathcal{D}^*$ with appropriately adapted τ to account for different dimensionalities. Then,

$$P\left(\mathcal{D}^*_g|M_e\right) = \frac{1}{N^*}\sum_{l=1}^{N^*} P\left(\mathcal{D}^*_g|M_e,\theta^{*(l)}_\Delta\right) P\left(\theta^{*(l)}_\Delta\right),$$

with $P\left(\theta^{*(l)}_\Delta\right) = N^*/N$ being an estimate of the marginal likelihood. Note that N^*/N is an under-approximation: samples that are not viable in one of the datasets could be viable in the joint dataset.

To determine the updated posterior model probabilities after the perturbation experiment, we use Bayes' formula (Eq. 4) to determine the posterior for each pair of generating and explanatory models, and average over the generating models (see Fig. 1B). Specifically, we calculate for each model M_e the posterior probability

$$Q_\Delta(M_e) \propto \sum_{M_g\in\mathcal{M}} P\left(\mathcal{D}^*_g|M_e\right) P\left(M_e|\mathcal{D}\right) P\left(M_g|\mathcal{D}, M_g \neq M_e\right). \quad (8)$$

Here, the last two terms account for the prior (with respect to $\mathcal{D}$) probabilities of the explanatory and generating models, respectively. Again, we can derive the normalizing constant from $\sum_{M\in\mathcal{M}} Q_\Delta(M) = 1$.

3. APPLICATION STUDY

3.1 GATA factor network

We evaluated our approach for a model of ordinary differential equations (ODEs) representing the gene regulatory network of the GATA factors in the yeast *Saccharomyces cerevisiae* from (Milias-Argeitis et al., 2016). The GATA factor network is composed of several feedback loops between the four transcription factors (Fig. 2) and responds to changes in nitrogen availability. The presence of a poor nitrogen source in the cells' environment results in transport of the activators (Gln3 and Gat1) into the nucleus. This enables the transcription of genes that will help utilize non-preferred nitrogen sources, a phenomenon known as nitrogen catabolite repression.

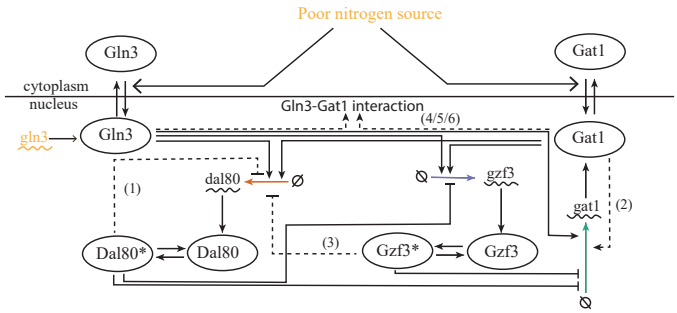

Fig. 2. The GATA factor network under low nitrogen availability. Gln3 and Gat1 mediate the signal by translocation from the cytoplasm to the nucleus. They act as activators by transcriptionally activating the repressors Dal80, Gzf3, among others. The active forms of Dal80 and Gzf3 (represented with asterisks) can repress the transcription of gzf3, gat1 and dal80 genes (wiggled lines). Black solid lines represent well-established interactions and dashed lines represent hypotheses, with numbers in brackets identifying the individual hypotheses for use in model nomenclature. The three inputs to the model are shown in yellow. The colored arrows indicate the interventions used in the application of our experiment design method (and represent that the transcription of the gene is prohibited).

The model with 13 dynamical states and 37-42 sampled parameters used in our application can capture, for all species involved, the mRNA, protein, and protein complex dynamics and it has three input variables reflecting nitrogen quality. The well-established GATA factor interactions characterize the core model (shown as solid black lines in Fig. 2). Interactions between species that are not conclusively justified throughout the literature are represented as extensions of the core model (Fig. 2, black dashed arrows). These hypothesized interactions are not mutually exclusive and create a set of 2^6-1 candidate model topologies. The main difference to Milias-Argeitis et al. (2016) is that we do not evaluate the repression of the Gzf3 transcription factor by Dal80, since their analysis demonstrated very low posterior probabilities for the candidate topologies missing this interaction. Another difference concerns the cooperativity between the transcription factors Gat1 and Gln3 targeting the promoter sequences of the gat1, dal80, and gzf3 genes. Specifically, the additional hypotheses 4, 5 and 6 evaluate the interaction of Gln3 and Gat1 targeting the promoter sequence of gat1, dal80 and gzf3, respectively. We adapted this as three different potential hypotheses (instead of one), considering that cooperation between Gat1 and Gln3 for one promoter region does not imply cooperation for another promoter sequence. In a similar manner as in the original publication (Milias-Argeitis et al., 2016), we designate the various models by the interactions they are missing. For example, model 12 is missing the hypotheses number 1 and 2 (see Fig. 2 for corresponding interactions).

For this application, we considered three genetic perturbations as possible interventions that alter the structure of the model topologies: gene deletions of the transcription factors Gat1, Dal80 and Gzf3. Each gene deletion is defined by setting the corresponding production rate constant for its mRNA to zero (Fig. 2), eliminating the corresponding protein levels simultaneously. Our *in silico* experiment of gene deletions represents a shift, at timepoint 0 (when cells are in steady state), from a good nitrogen source to a poor one (conditions mimicked experimentally by rapamycin inhibition when yeast cells are growing on glutamin. Finally, we assume that the measurements we can obtain are the mRNA fold-change values of the transcription factors at specific time points under conditions of low nitrogen availability.

3.2 Results

To evaluate the candidate topologies we generated *in silico* experimental data (45 datapoints) by simulating the model 13456 that has highest predictive ability in Milias-Argeitis et al. (2016) under poor nitrogen conditions to generate a ground truth. Similar to the original publication, we used $\Sigma_m = \text{diag}(0.1^2)$ and $\Sigma_a = 0$ in Eq. 2.

The initial round of discriminating between candidate topologies was implemented by employing our previously published method of Bayesian topological filtering (Sunnåker et al., 2013) using a uniform prior on the candidate models. In brief, the topological filtering algorithm starts by defining a model that incorporates all the candidate hypotheses (root model). Through a combination of out-of-equilibrium adaptive Monte Carlo and multiple ellipsoid based sampling (Zamora-Sillero et al., 2011) it explores the parameter space of the root model. For each parameter set found, the algorithm advances by evaluating if projections of a single parameter still pass the viability criterion given above, until no further projections are feasible. The result is a set $\mathcal{M}$ of reduced models that are compatible with the experimental data $\mathcal{D}$ initially used, and their respective viable regions $\mathcal{V}(\cdot, \mathcal{D})$.

Here, the priors of the parameters were kept in agreement with the original article by Milias-Argeitis et al. (2016). The observables in $\mathcal{D}_g^*$ are the same as for generating $\mathcal{D}$ in this application, but we additionally assume, for numerical reasons, that $\Sigma_a = \text{diag}(0.1^2)$ in Eq. 2; this additive noise can be thought of as the limit of our detection in mRNA microarray experiments. The residuals between the experimental data and the simulations are assumed to be normally distributed and the likelihood is thus calculated based on squared errors as in Toni et al. (2009). We used the 95% quantile of the chi-square distribution with degrees of freedom equal to the number of data points minus the number of parameters to define the threshold τ in Eq. 6. The worst estimated viable region(s) had a relative error of 8%, all others had 1-4%. Interestingly, the true model was within the reduced set of candidate models, but had only second-highest posterior probability after model 1245, with model 12435 having almost equal probability (Fig. 3B). This indicates that, if we were to stop the model selection process at this point, we would have inferred incorrect mechanistic details of the GATA factor network.

This first selection step determined 32 of 63 candidate model topologies as viable, and we next introduced structural changes Δ in a brute force manner to further discriminate between these topologies. For example, the second- and third-ranked models 13456 and 12345 initially showed

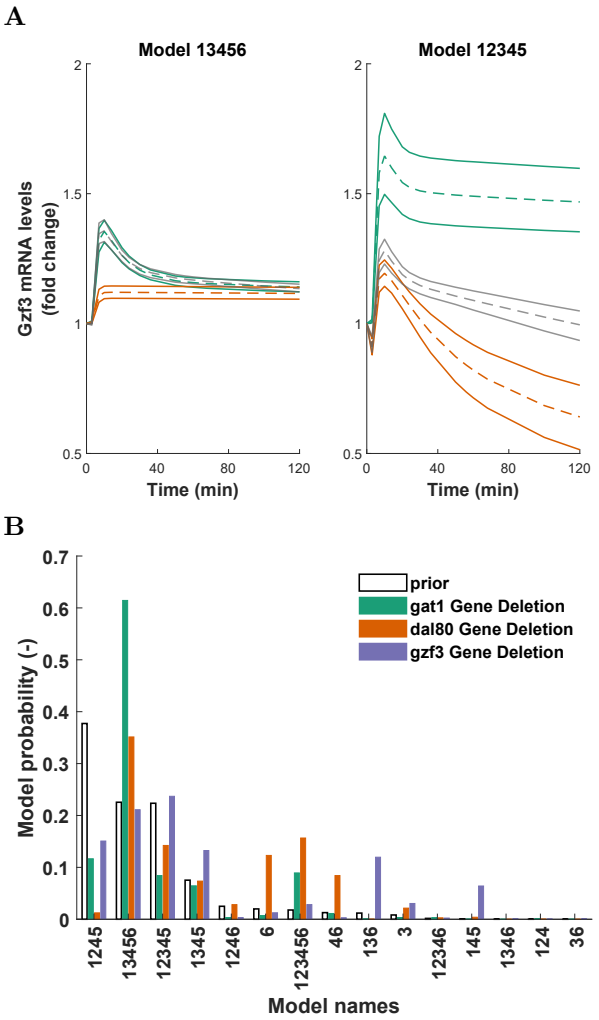

Fig. 3. Results for the GATA factor example network. (**A**) Trajectories of the Gzf3 mRNA for the initially second- and third-ranked models, after a shift from a rich to a poor nitrogen source. Dashed lines represent the mean trajectories, and the continuous lines refer to the 90% and 10% quantiles, respectively. Trajectories for the wild type, using parameter samples from the first model selection round with $\mathcal{D}$, are shown in gray, while trajectories after performing a gat1 or dal80 deletion are shown in green and orange, respectively. (**B**) Posterior model probabilities before and after a gene deletion experiment; color coding for the experiments corresponds to the legend. Only models with prior probability greater than 8.35e-05 are shown for readability.

similar mRNA fold-changes of Gzf3 (Fig. 3A, gray lines). Their dynamic behavior becomes distinct, however, when setting the production constant of Gat1 to zero (green lines) or similarly deleting the dal80 gene (orange lines).

Next, we calculated the updated posteriors for each of the three gene deletion experiments by implementing the theory outlined previously (Eq. 8) using the posteriors $Q(M)$ from the initial round of model discrimination as model priors. Using forward simulations, we evaluated if the parameters that were identified as viable in the first model selection round were still viable under the new

Table 1. Expected information gain for gene deletion experiments (in bits).

$\pi(M)$	Deleted gene	45 Datapoints	90 Datapoints	135 Datapoints
Q	gat1	0.7481	2.0371	2.0913
	dal80	1.159	1.7665	1.7979
	gzf3	0.809	2.1687	2.7163
$\mathcal{U}$	gat1	3.1026	4.8301	4.9151
	dal80	2.2986	3.3938	3.9713
	gzf3	2.1114	2.4590	2.5562

threshold established by the relation between the corresponding M_e and M_g for a given perturbation experiment, indicating that the modified model M_e can demonstrate both behaviors of the initial dataset and the prospective perturbation experiment.

We found the highest posterior probabilities for the (true) 13456 model when applying the gat1 and dal80 gene deletions, while this model has equal probability with the 12345 model for the Gzf3 experiment. In addition, the posterior probability of the previous top-ranked model 1245 dropped substantially for all perturbation experiments. This demonstrates that our perturbation design experiments can further discriminate between candidate model topologies. The information gains (Table 1) indicate: (i) with increasing data (45 datapoints were the minimum for the χ^2-test), the information gain increases, but appears to saturate (the gain was 2.629 bits for the initial dataset); (ii) for the reference distribution Q based on prior data, 45 datapoints with low information gain may be insufficient for selecting the intervention (yet higher data density is not realistic for microarray measurements), (iii) in contrast to the uniform distribution favoring model selection.

Our results partially confirm those of Milias-Argeitis et al. (2016): we found that hypotheses 1 and 3 (inhibition of dal80 transcription by Dal80, respectively Gzf3) are absent in most of our high-ranked models. We cannot directly compare results for the original hypothesis 5 (Gln3-Gat1 interactions), which we split into three individual interaction hypotheses. We found, however, that our partial hypothesis 4 is absent from most of the highly-ranked models. This means that the transcription factors Gln3 and Gat1 do not interact, but rather compete for the same sequence promoter of Gat1, a new result supported by the experimental biology literature (Stanbrough et al., 1995).

4. DISCUSSION

We introduced an experimental design method that is distinct from existing methods by the combination of two main features: first, it uses the Bayesian framework to systematically represent parametric and structural uncertainties in biological networks due to limited biological knowledge, relatively sparse data, and frequent non-identifiability. The Bayesian updating of model probabilities also implies that the method can consider sequential experiments. Second, our method relies on structural interventions such as gene knockouts to define the experimental design space. This aligns well with the practical implementation of designed experiments: genetic interventions have become easy to conduct. In addition, we assume only a limited set of specific measurements to be performed, which essentially obviates the need to establish new measurement or stimulation techniques in the laboratory for conducting experiments proposed by other design meth-

ods. In principle, the method is also not limited to gene deletions; other types of interventions such as applications of small interfering RNA, protein inhibitors, or protein over-expression could be represented by appropriate mappings to perturbations of model parameters.

The application to the GATA factor network of *S. cerevisae* demonstrates that our design method can be applied to models of a size that is typical for detailed, dynamic systems biology models; we are only aware of Busetto et al. (2013) addressing models of similar or larger size. This application also indicates how designed genetic intervention experiments can reduce experimental effort: with one gene deletion it was possible to identify the correct model structure in our example.

However, the proposed approach comes with conceptual and computational limitations: First, we assume that the true model is in the model set; if observations from the proposed experiment differ significantly from model predictions, we should reconsider the biological hypotheses. Second, to establish prior model probabilities for experimental design, we rely on topological filtering (Sunnåker et al., 2013) to initially generate a reduced model set. Due to its greedy approach to model selection, we may miss viable models given the initial experimental observations $\mathcal{D}$. If the number of models is fairly small, however, this step can be eliminated and we can use standard methods for Bayesian model selection to obtain the corresponding posteriors and parameter distributions. Third, for computational efficiency we require that parameter vectors are viable with respect to both $\mathcal{D}$ and $\mathcal{D}^*$, and not $\mathcal{D} \bigcup \mathcal{D}^*$, which may be too restrictive. This approach, fourth, can lead to a low number of viable parameter vectors after an *in silico* intervention, which reduces the accuracy of estimates for the expected model posteriors. Re-sampling of parameter vectors, either from a larger parameter space or within the ellipsoid approximations used, could solve this problem, yet with increased computational efforts.

Future improvements of the method could therefore include re-sampling schemes that balance accuracy (evaluated by standard approaches for Monte-Carlo integration) and computational effort with respect to the method's aim, experimental design for model selection. We believe, however, that more important advances could be possible by exploiting the nested structure of models in the topological filtering framework. For our application here, the sets of hypotheses (generating models as subsets of the root model) and of interventions were orthogonal to each other. In general, reduced models may correspond to the application of interventions to non-reduced models, and this relation could be used to avoid re-estimation of relevant quantities for the reduced models by an optimized order of model evaluations. Ultimately, however, experimental validations of model selection with predicted experimental designs will be essential to evaluate the performance of our experimental design method.

REFERENCES

Beaumont, M.A. (2019). Approximate Bayesian computation. *Annual Review of Statistics and Its Application*, 6, 379–403. doi:10.1146/annurev-statistics-030718-105212.

Braniff, N. and Ingalls, B. (2018). New opportunities for optimal design of dynamic experiments in systems and synthetic biology. *Current Opinion in Systems Biology*, 9, 42–48.

Busetto, A.G., Hauser, A., Krummenacher, G., Sunnåker, M., Dimopoulos, S., Ong, C.S., Stelling, J., and Buhmann, J.M. (2013). Near-optimal experimental design for model selection in systems biology. *Bioinformatics (Oxford, England)*, 29, 2625–2632. doi: 10.1093/bioinformatics/btt436.

Chakrabarty, A., Buzzard, G.T., and Rundell, A.E. (2013). Model-based design of experiments for cellular processes. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 5(2), 181–203.

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science*, 353(6306), aaf1420.

Hug, S., Schmidl, D., Li, W.B., Greiter, M.B., and Theis, F.J. (2016). Bayesian model selection methods and their application to biological ODE systems. In *Uncertainty in Biology*, 243–268. Springer, Cham.

Liepe, J., Filippi, S., Komorowski, M., and Stumpf, M.P.H. (2013). Maximizing the information content of experiments in systems biology. *PLoS Computational Biology*, 9, e1002888. doi:10.1371/journal.pcbi.1002888.

Milias-Argeitis, A., Oliveira, A.P., Gerosa, L., Falter, L., Sauer, U., and Lygeros, J. (2016). Elucidation of genetic interactions in the yeast GATA-factor network using Bayesian model selection. *PLoS Computational Biology*, 12, e1004784. doi:10.1371/journal.pcbi.1004784.

Ryan, E.G., Drovandi, C.C., McGree, J.M., and Pettitt, A.N. (2016). A review of modern computational algorithms for Bayesian optimal design. *International Statistical Review*, 84(1), 128–154.

Stanbrough, M., Rowen, D.W., and Magasanik, B. (1995). Role of the GATA factors Gln3p and Nil1p of Saccharomyces cerevisiae in the expression of nitrogen-regulated genes. *Proceedings of the National Academy of Sciences*, 92(21), 9450–9454.

Sunnåker, M. and Stelling, J. (2016). Model extension and model selection. In *Uncertainty in Biology*, 213–241. Springer, Cham.

Sunnåker, M., Zamora-Sillero, E., Dechant, R., Ludwig, C., Busetto, A.G., Wagner, A., and Stelling, J. (2013). Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key Msn2 control mechanism. *Science Signaling*, 6, ra41.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M.P.H. (2009). Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society, Interface*, 6, 187–202. doi:10.1098/rsif.2008.0172.

Vanlier, J., Tiemann, C.A., Hilbers, P.A., and van Riel, N.A. (2014). Optimal experiment design for model selection in biochemical networks. *BMC Systems Biology*, 8(1), 20.

Vyshemirsky, V. and Girolami, M. (2008). Bayesian ranking of biochemical system models. *Bioinformatics*, 24(6), 833–839. doi:10.1093/bioinformatics/btm607.

Zamora-Sillero, E., Hafner, M., Ibig, A., Stelling, J., and Wagner, A. (2011). Efficient characterization of high-dimensional parameter spaces for systems biology. *BMC Syst Biol*, 5, 142.