

Parallel Tempering with Lasso for Model Reduction in Systems Biology

Sanjana Gupta¹, Robin E.C. Lee¹, James R. Faeder¹

¹ Department of Computational and Systems Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260, USA,
Email: {sag134,robinlee,faeder}@pitt.edu

Abstract

Systems Biology models reveal relationships between signaling inputs and observable molecular or cellular behaviors. The complexity of these models, however, often obscures key elements that regulate emergent properties. We use a Bayesian model reduction approach that combines Parallel Tempering with Lasso regularization to identify minimal subsets of reactions in a signaling network that are sufficient to reproduce experimentally observed data. The Bayesian approach finds distinct reduced models that fit data equivalently. A variant of this approach based on Group Lasso is applied to the NF- κ B signaling network to test the necessity of feedback loops for responses to pulsatile and continuous pathway stimulation. Taken together, our results demonstrate that Bayesian parameter estimation combined with regularization can isolate and reveal core motifs sufficient to explain data from complex signaling systems.

Introduction

Cells use complex networks of proteins and other biomolecules to translate environmental cues into various cell fate decisions. Mathematical and computational models are increasingly used to analyze the nonlinear dynamics of these complex biochemical signaling systems [1–4]. As our knowledge of the biochemical processes in a cell increases, reaction network models of cell signaling have been growing more detailed [4–6]. Detailed models are a useful summary of knowledge about a system but they suffer from several drawbacks. First, the complexity may obscure simpler motifs that govern emergent cellular functions [7–9]. Second, the large number of parameters creates a high-dimensional search problem for parameter values where the model fits the data. To mitigate these problems, it is useful to reduce the number of reactions in a model, provided that the reduced model is still able to reproduce a given set of experimental observations. In this work we pose model reduction as a constrained Bayesian parameter estimation (BPE) problem to simultaneously calibrate and reduce models. Given a prior reaction network model, our method finds all possible minimal subsets of non-zero parameters that fit the data.

Previous studies have addressed model reduction for biochemical systems [10]. Some examples include reductions by topological modifications to resolve non-identifiability in models [11, 12] and reductions by timescale partitioning [13, 14]. Non-identifiability arises when multiple unique parameterizations of a model give the same model output. Quaiser et al. [11] and Raue et al. [12] developed methods to find non-identifiable parameters and used this analysis to resolve non-identifiability by model simplifications such as lumping or removal of reactions. The simplification step, however, is not automated and requires a skilled modeler. Timescale partitioning methods use timescale

separations in the reaction kinetics to apply model reduction based on quasi-steady-state and related approximations [10, 14]. Both of these methods generate reduced models but do not carry out parameter estimation to fit experimental data. Another approach to model reduction that includes parameter estimation [15] uses mixed-integer nonlinear optimization to combine parameter estimation with model reduction by reaction elimination, a technique common in the field of chemical engineering [16, 17]. Drawbacks of this approach are that it requires an additional binary parameter for every reaction in the model and that the genetic algorithms used for the optimization only provide point estimates of the parameters.

Here, we develop reaction elimination in a Bayesian framework that combines parameter estimation and model reduction without requiring additional parameters. BPE has been shown to be useful to characterize high-dimensional, rugged, multimodal parameter landscapes common to systems biology models [1, 18–21], but it suffers from the drawback that the Markov Chain Monte Carlo (MCMC) methods commonly used to sample model parameter space are often slow to converge and do not scale well with the number of model parameters. We recently showed that Parallel Tempering (PT), a physics-based method for accelerating MCMC [22], outperforms conventional MCMC for systems biology models with up to dozens of parameters [18]. Here, we apply Lasso (also known as L1 regularization), a penalty on the absolute values of the parameters being optimized, to carry out model reduction. In the fields of statistics and machine learning, Lasso is widely used for variable selection to identify a parsimonious model – a minimal subset of variables required to explain the data [23]. In the context of biology, Lasso has been widely applied to gene expression and genomic data typically in combination with standard regression techniques [24–28] and less commonly in Bayesian frameworks [29, 30]. In the mechanistic modeling context, Lasso regression has been used to predict cell type specific parameters in ODE reaction network models [31], but to our knowledge it has not been implemented to reduce such models.

Our method, PTLasso, combines PT with Lasso regularization to simultaneously calibrate and reduce models. The core idea is that every reaction in the model is governed by a rate constant parameter that, when estimated as zero, removes the reaction from the model simulation. Since the approach is Bayesian, PTLasso extracts all possible minimal subsets of reactions, which provides alternate mechanisms to explain the data. Using both synthetic and real biological data, we demonstrate that PTLasso is an effective approach for model reduction. We also apply PT coupled with Group Lasso [32] in a larger model of NF- κ B signaling to select over reaction-network modules instead of individual reactions. Group Lasso can test mechanistic hypotheses about the necessity of signaling modules, such as feedback loops, to explain data from particular experimental conditions. Overall, our results demonstrate that BPE combined with regularization is a powerful approach to dissect complex systems biology models and identify core reactions that govern cell behavior.

The remainder of this paper is organized as follows. In Methods we provide an overview of the PTLasso approach with in-depth descriptions of PT, Lasso regularization, Group Lasso and the setup of computational experiments. In Results we demonstrate PTLasso on synthetic examples of increasing complexity followed by an application of the Group Lasso approach to address mechanistic questions in NF- κ B signaling. Finally, in Discussion we highlight advances as well as limitations of the method and present the implications of this study for the broader context of biological modeling and analysis.

Methods

In this work we use Bayesian parameter estimation (BPE) for model reduction. Following [18], BPE methods aim to estimate the probability distribution for the model parameters conditioned on the data. The probability of observing the parameter vector, $\vec{\theta}$, given the data, Y , is given by Bayes' rule

$$p(\vec{\theta}|Y) \propto p(Y|\vec{\theta})p(\vec{\theta}). \quad (1)$$

Here, $p(Y|\vec{\theta})$ is the conditional probability of Y given $\vec{\theta}$, and is described by a *likelihood model*. For the ODE models in this study, we assumed Gaussian experimental measurement error, in which case the likelihood of a parameter vector, $\vec{\theta}$, is given by

$$L(\vec{\theta}) = e^{-\sum_S \sum_T (Y_{\text{sim}}(\vec{\theta}) - Y_{\text{expt}})^2 / 2\sigma^2}, \quad (2)$$

where S is a list of the observed species, T is a list of the time points at which observations are made, σ is the standard deviation of the likelihood model, $Y_{\text{sim}}(\vec{\theta})$ is the model output for parameter vector, $\vec{\theta}$, and Y_{expt} is the corresponding experimental data. $p(\vec{\theta})$ is the independent probability of $\vec{\theta}$, often referred to as the *prior distribution*, which represents our prior beliefs about the model parameters. It can be used to restrict parameters to a range of values or even to limit the number of nonzero parameters, as discussed further below.

MCMC sampling

MCMC methods sample from the posterior distribution, $p(\theta|Y)$, by constructing a Markov chain with $p(\theta|Y)$ as its stationary distribution. Following the notation of Metropolis *et al.* [33], we define the energy of a parameter vector $\vec{\theta}$ as

$$E(\vec{\theta}) = -\log L(\vec{\theta}) - \log p(\vec{\theta}), \quad (3)$$

where L and p are the likelihood and prior distribution functions defined above. In this section we will briefly describe the Metropolis-Hastings and Parallel Tempering algorithms for MCMC sampling.

Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm is a commonly-used MCMC algorithm for BPE [34]. At each step, n , the method uses a proposal function to generate a new parameter vector, $\vec{\theta}_{\text{new}}$, given the current parameter vector, $\vec{\theta}_n$. A common choice of proposal function is a normal distribution centered at $\vec{\theta}_n$:

$$f(\vec{\theta}_{\text{new}}, \vec{\theta}_n) \propto e^{-\alpha|\vec{\theta}_{\text{new}} - \vec{\theta}_n|^2}. \quad (4)$$

For any f that is symmetric with respect to $\vec{\theta}_{\text{new}}$ and $\vec{\theta}_n$, the move $\vec{\theta}_{n+1} = \vec{\theta}_{\text{new}}$ is accepted with probability $\min(1, e^{-\Delta E})$, where $\Delta E = E(\vec{\theta}_{n+1}) - E(\vec{\theta}_n)$. If the move is not accepted $\vec{\theta}_{n+1}$ is set to $\vec{\theta}_n$.

Parallel Tempering

In PT (also referred to as replica exchange Monte Carlo [22]), several Markov chains are constructed in parallel, each with a different temperature parameter, β , which scales the acceptance probability from the MH algorithm, which is now given by $\min(1, e^{-\beta\Delta E})$. A Markov chain with $\beta = 1$ samples the true energy landscape as in MH. Higher

temperature chains have $\beta < 1$ and accept unfavorable moves with a higher probability, sampling parameter space broadly. Tempering refers to periodic attempts to swap parameter configurations between high and low temperature chains. These moves allow the low temperature chain to escape from local minima and improve both convergence and sampling efficiency [22]. Following [18], the PT algorithm is as follows:

1. For each of N swap attempts (called “swaps” for short)
 - (a) For each of N_c chains (these can be run in parallel)
 - i. Run N_{MCMC} MH steps
 - ii. Record the values of the parameters and energy on the final step.
 - (b) For each consecutive pair in the set of chains in decreasing order of temperature, accept swaps with probability $\min(1, e^{\Delta\beta\Delta E})$, where ΔE and $\Delta\beta$ are the differences in the energy and β of the chains, respectively.

Adapting the step size and the temperature parameter can further increase the efficiency of sampling [22], but varying parameters during the construction of the chain violates the assumption of a symmetric proposal function (also referred to as “detailed balance”). It is therefore advisable to do this during a “burn-in” phase prior to sampling.

Regularization with Lasso

Lasso regularization penalizes the L1-norm (sum of absolute values) of the parameter vector, which biases all model parameters towards a value of zero [23]. In a Bayesian framework, the Lasso penalty is equivalent to assuming a Laplace prior on each parameter θ_i given by

$$p(\theta_i) = \frac{1}{2b} \exp\left(\frac{|\theta_i - \mu|}{b}\right), \quad (5)$$

where b is the width and μ is the mean, which is set to zero for variable selection in linear parameter space. The energy function is then

$$E(\vec{\theta}) = -\log L(\vec{\theta}) - \sum_{i=1}^{n_{\text{par}}} \frac{|\theta_i - \mu|}{b}, \quad (6)$$

where n_{par} is the number of model parameters. Since PT only uses energy differences, the normalization constant $\frac{1}{2b}$ in Eq. 5 does not appear in Eq. 6. b is inversely proportional to the regularization strength.

For efficiency we usually perform parameter estimation in log parameter space, so instead of regularizing by setting μ to zero, we set it to a large negative value, such that the parameter value is small enough that it does not affect the dynamics of the model variables on the timescale of the simulation.

Regularization with Group Lasso

To account for modularity in complex signaling networks [35], we use Group Lasso, previously described for regression problems [32], to perform selection at the level of reaction modules instead of individual reactions. All reactions in a module share a common penalty parameter that is multiplied with a reaction-specific parameter to get the full reaction rate constant.

For every reaction i in module m , the reaction rate constant is given by

$$\theta_i = k_i \lambda_m, \quad (7)$$

where λ_m is the penalty parameter for module m and k_i is a reaction-specific parameter. Defining $\theta'_i = \log(\theta_i)$, $k'_i = \log(k_i)$, and $\lambda'_m = \log(\lambda_m)$, we have

$$\theta'_i = k'_i + \lambda'_m. \quad (8)$$

The energy function is then

$$E(\vec{\theta}') = -\log L(\vec{\theta}') - \sum_{m=1}^{n_{\text{mod}}} \frac{|\lambda'_m - \mu|}{b} - \sum_{i=1}^{n_{\text{par}}} f(k'_i), \quad (9)$$

where

$$f(k'_i) = \begin{cases} 0, & \text{if } k'_i \in (LB_i, UB_i) \\ \infty, & \text{otherwise.} \end{cases} \quad (10)$$

Here, n_{mod} is the number of modules and LB_i and UB_i are parameters that restrict the reaction-specific parameters. UB_i is chosen such that when λ'_m is within the Laplace prior boundaries, i.e., $\lambda'_m \approx \mu$, the maximum value of θ'_i , $\approx UB_i + \mu$, is small enough that it does not affect the dynamics of the model variables on the timescale of the simulation. For the application to NF- κ B signaling we chose $\mu = -25$, $LB_i = -5$ and $UB_i = 10$ for all i .

Synthetic data sets used in model calibration

For the two examples presented in Results that used synthetic data, we generated the sets labeled “true data” by simulating the model with a single set of parameter values and sampling with a fine time resolution. We then generated 10 noisy replicates of this data at a coarser set of time points by adding Gaussian noise with mean of zero and variance of either 10% or 30% of the true value at each point. The mean and variance of the replicates then defined the “observed data” used for fitting.

Constraining the model

We use two kinds of constraints in fitting, soft constraints and hard constraints. Soft constraints can be violated, but are associated with a finite penalty [36]. For example, the energy function penalizes parameter vectors for producing model outputs that deviate from the data. Hard constraints, on the other hand, cannot be violated because they are associated with an infinite penalty. We used hard constraints in the NF- κ B signaling model to enforce certain known properties of the NF- κ B system, such as that the exit rate of NF- κ B -I κ B complex from the nucleus is greater than that of free NF- κ B [37]. A full list of constraints applied to the NF- κ B signaling model is listed in Table S4.

MCMC chain initialization

All MCMC chains must be initialized with a starting parameter vector. For simple examples, such as the pulse-generator motif and linear dose-response models, chains were initialized by randomly sampling from the prior until a parameter vector with energy below a threshold is found. For more complex examples, to avoid long burn-in periods when starting from unfavorable start points, parameter vectors obtained from short PT runs were used to initialize longer PT chains. For example, parameter sets obtained for one NF- κ B trajectory could be used as a start point for fitting a different NF- κ B trajectory, or a PTLasso chain with a small value of b (more constrained) could be initialized from a parameter set obtained from PTLasso with a large value of b (less constrained). The exact procedures used to generate the starting configurations used in

all computational experiments are provided in the Supplemental Code that will be provided on GitHub upon publication of this manuscript.

Convergence testing

To check for convergence, PT or PTLasso was run twice for each computational experiment, and the two parameter chains were used to calculate the Potential Scale Reduction Factor (PSRF) for each model parameter. The PSRF compares intra-chain and inter-chain variances for model parameters and serves as a measure of convergence [38]. In keeping with the literature, we consider a PSRF less than 1.2 [19, 20, 38] as indicating convergence. We also calculated the stricter Multivariate PSRF (MPSRF), which extends PSRF by checking for convergence of parameter covariation. Third-party MATLAB libraries used for the MPSRF calculations are available at <https://research.cs.aalto.fi/pml/software/mcmcdiag/>.

For models with a large number of parameters, such as the 26-parameter NF- κ B signaling model, the number of PT samples needed for convergence was large and time consuming to obtain in a single run. Instead of running two long PT chains each of length N , we picked two favorable initial conditions and from each ran a set of M PT chains of length N/M in parallel to reduce wall clock time. We calculated the univariate PSRF of the M energy chains within each group, and if PSRF was less than 1.2, we assumed that the chains were sampling the same energy basin and combined them (Table S5). This gave us two groups of N PT samples that we used to calculate parameter convergence.

PSRF and MPSRF values for each computational experiment are shown in Tables S1 and S2 respectively. We also show in Table S3 that the acceptance rates for most chains are close to the optimal value of 0.234 [39].

Hyperparameter selection

The hyperparameters associated with PTLasso are μ and b , the mean and width of the Laplace prior on each parameter that is being regularized. For simplicity, we keep these the same for all model parameters, although they could in principle vary, which would lead to a more difficult inference problem. To select the hyperparameters, we varied b and used the “elbow” in the negative log likelihood vs. b plot to find the smallest value of b (maximum regularization strength) that does not substantially increase the negative log likelihood of the fit [18, 40]. We also checked that the results were insensitive to small variations in μ (Figs. S1–S3).

For more computationally expensive models, we used hyperparameter estimates close to those obtained from the smaller synthetic models and compared the average log likelihoods of the fits from PT and PTLasso. For all of the examples shown, we found that the fit with PTLasso is at least as good as the fit with PT (Figs. 4, S1–S3).

Software

All results reported in this work were obtained using `ptempest` [18], which is a MATLAB package for parameter estimation that implements PT with support for regularization. The source code is available at <http://github.com/RuleWorld/ptempest>.

Results

Reduced motifs can be inferred from dense reaction-networks in the absence of a prior architecture

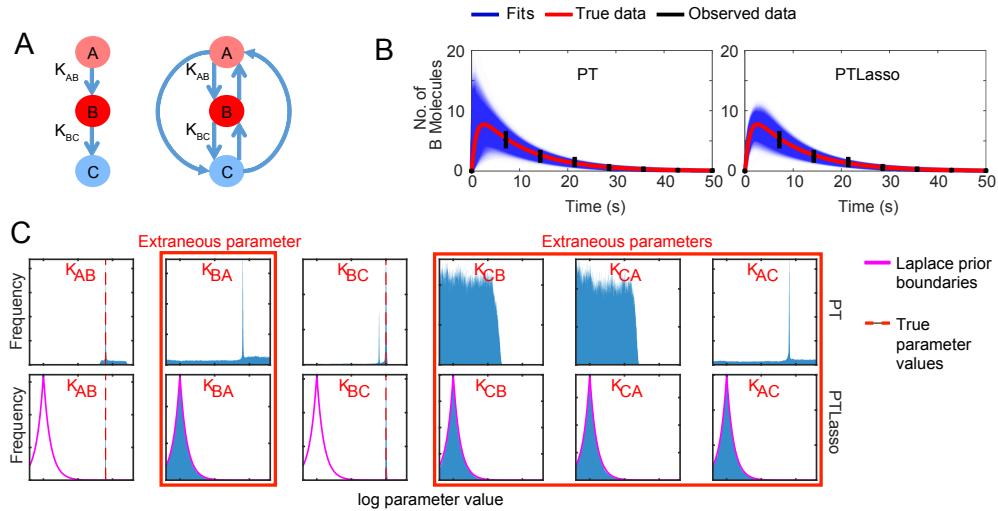
To demonstrate that PTLasso can recover a minimal model architecture without prior knowledge of the reaction network, we used synthetic time-course data to infer a pulse-generator motif from a fully connected 3-node network of unimolecular reactions. The motif $A \rightarrow B \rightarrow C$ (Fig. 1A, left), modeled as a system of ODE's, was used to generate a time course for species B after initializing the system with 100 molecules of species A at time $t = 0$ (red curves in Fig. 1B labeled 'true data'). To simulate the effects of experimental noise and cell-to-cell variability, Gaussian noise was added to generate ten noisy trajectories that were sampled at eight time points (Fig. S1A). The mean and standard deviation of these synthetic trajectories formed the 'observed data' (black points and error bars in Fig. 1B) used for subsequent parameter estimation and model reduction.

PT and PTLasso were then used to fit this data using the fully-connected 3-node network comprised of six reactions (Fig. 1A, right). Time courses from PT and PTLasso (Fig. 1B) both fit the observed data (Fig. S1C), but the PTLasso curves fit the true data better at times before the first observed data point. PT finds parameter probability distributions (Fig. 1C, top row) that exhibit sharp peaks near the exact values of the two nonzero parameters, K_{AB} and K_{BC} , but finds significant probability for other values of these parameters and non-zero values for the other rate constants in the complete network that should have zero value (labeled 'extraneous'). By contrast, PTLasso (Fig. 1C, bottom row) recovers tight distributions on the values of the two nonzero parameters that lie well outside the Laplace prior, while the probability distributions for the extraneous parameters all conform tightly to the prior distribution, indicating that the corresponding reactions can be removed from the network. Taken together, these results demonstrate that PTLasso can recover network architecture and parameter values that are not inferred by PT alone.

To determine if the method scales to larger networks, we applied PT and PTLasso to a fully connected 5-node network (Fig. 2A, Fig. S2A). As with the 3-node example, PTLasso fits for a complete 5-node network are more similar to the true data than fits with PT alone (Fig 2B). Similarly, rate constant parameter distributions with PT are all broad (Fig. 2C), whereas the extraneous parameters for the PTLasso fits were within the Laplace prior (Fig 2D). In addition to a correct reaction rate constant distribution for K_{AB} , PTLasso recovered bimodal distributions for K_{BC} , K_{BD} , and K_{BE} , suggesting that the essentiality of each of the reactions $B \rightarrow C$, $B \rightarrow D$ and $B \rightarrow E$ depends on which of the other two are included. This is because the model $A \rightarrow B \rightarrow C$ is indistinguishable from $A \rightarrow B \rightarrow D$ and $A \rightarrow B \rightarrow E$ without more information about the system. Even though the marginal posterior distributions show all three parameters playing a role, parameter covariation (Fig. S2D, right) reveals that only one of the reactions $B \rightarrow C$, $B \rightarrow D$, $B \rightarrow E$ is simultaneously active and rate constant distributions for the other two are centered at 10^{-10} (proxy for 0 when sampling in log-scale). The same covariation plot obtained without Lasso does not show any meaningful structure (Fig. S2D, left). Taken together, these results show that PTLasso correctly identifies network parameters and suggests that $A \rightarrow B \rightarrow C$, $A \rightarrow B \rightarrow D$, and $A \rightarrow B \rightarrow E$ are equivalent parsimonious models.

Overall, our results show that PTLasso is a global approach that can extract correct parameter estimates and architectures of all equivalent reduced models that fit the data from fully connected networks of varying sizes. This is especially useful in the context of complex cell signaling systems that often have redundant elements in which case the method can be used to identify alternate signaling mechanisms that fit the data.

Fig 1. Model reduction using Lasso with a fully connected 3-node graph. **A)** Motif used to generate the observed data (left) and reaction network diagram of the fully connected 3-node network used as the starting point for PTLasso (right). The initial concentration of A (light red) is 100 molecules. The initial concentrations of B and C are 0. The concentration of B (red) is observed at multiple time points, but the concentration of C (blue) is not observed. Each reaction has an associated rate constant parameter. $K_{AB} = 0.1$ and $K_{BC} = 1$ are the rate constant values used to generate the observed data. **B)** Fits of the model to the data with PT (left) and PTLasso (right). Blue lines show ensemble fits (4e3 samples, 100 time points per sample), red line shows the true data (100 time points), and the black error bars show the mean \pm standard deviation of the observed data (8 time points). **C)** Frequency histograms showing probability distributions of the parameters (from 4e5 PT samples) for fits with PT (top row) and PTLasso (bottom row). The y-axis for each panel is scaled from 0 to the max value of the distribution to emphasize differences in the shapes of the distribution. The pink lines show the Laplace prior boundaries, the dashed red lines (panels for K_{AB} and K_{BC}) show the parameter values of the known model used to generate the true data, and the x-axis range, [-12,3], represents the sampling range. Parameter distributions confined within the Laplace prior boundaries are extraneous.

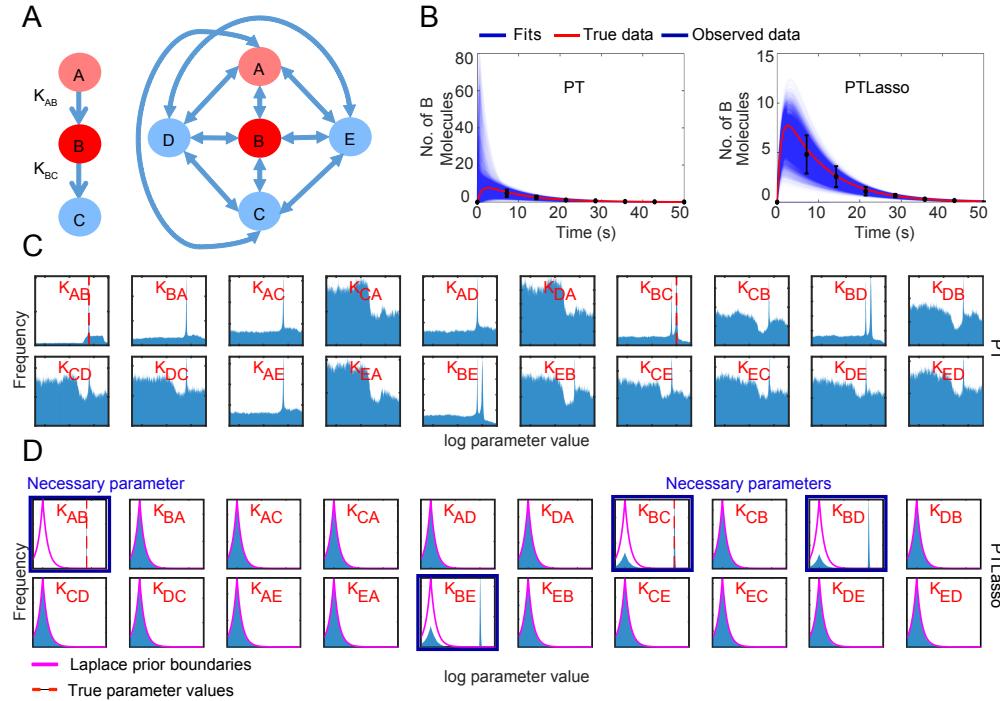


Motifs with specific dose-response relationships can be inferred from a prior network

In the previous section we assumed no prior knowledge of a reaction-network and fitted a simple model output. To demonstrate the extraction of motifs with more complex behaviors in the more likely scenario where there is some prior network of hypothesized molecular interactions, we used PTLasso to extract parts of the prior network required to produce specific dose-response relationships.

Tyson et al. [9] previously described two simple biochemical models that individually produce linear or perfectly adapting dose-response relationships. We constructed a prior network of a signal, S, response, R, and intermediate, X, by combining the linear and adaptive dose-response models into a single shared architecture (Fig. 3A). We show that PTLasso correctly identifies the linear and adaptive submodels when the combined model is fit to different simulated data. To begin with, the linear dose-response submodel was used to generate synthetic time courses for R in response to 4 increasing levels of S (S=1,2,3,4). As earlier, Gaussian noise was added to each trajectory to simulate experimental noise and cell-to-cell variability, and the mean and standard deviation for each time course was calculated at 4 distinct time points (including t=0), creating 16 data points that constitute the observed data. When fit to the observed data, PT produced fitted time courses for R that go through the observed data points but have fast time scale deviations from the true data time courses (Fig. 3B, left). PTLasso again produced fits closer to the true data (Fig. 3B, right). Parameter distributions inferred with PTLasso show that the 2-parameter model reduced from the combined model correctly recovers the architecture and parameterization of the

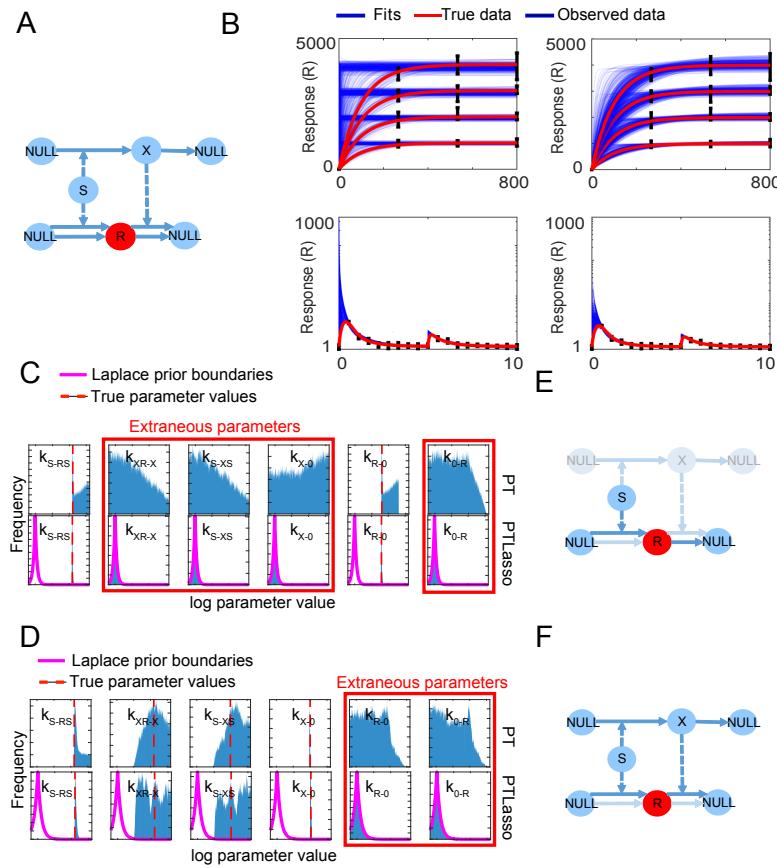
Fig 2. Model reduction using Lasso with a fully connected 5-node graph. **A)** Motif used to generate the observed data (left) and reaction network diagram of the fully connected 5-node network used as the starting point for PTLasso (right). The initial concentration of A (light red) is 100 molecules. The initial concentrations of B,C and D are 0. The concentration of B (red) is observed at multiple time points, but the concentration of C, D and E (blue) are not observed. $K_{AB} = 0.1$ and $K_{BC} = 1$ are the rate constant values used to generate the observed data. **B)** Fit of the model to the data with PT (left) and PTLasso (right). Blue lines show ensemble fits (7e3 samples, 100 time points per sample), red line shows the true data (100 time points), and the black error bars show the mean \pm standard deviation of the observed data (8 time points). **C)** Frequency histograms showing probability distributions of the parameters (from 7e5 PT samples) for fits with PT and **D)** PTLasso. The y-axis for each panel is scaled from 0 to the max value of the distribution to emphasize differences in the shapes of the distribution. The pink lines show the Laplace prior boundaries, the dashed red lines (panels for K_{AB} and K_{BC}) show the parameter values of the known model used to generate the true data, and the x-axis range, [-12,3], represents the sampling range. Parameter distributions that deviate from the prior are necessary.



linear-dose response submodel used to generate the synthetic time courses (Fig 3C,E).

When the perfectly adaptive dose-response submodel was similarly used to generate observed data in response to two successive increasing signal concentrations, PTLasso reduced the combined system to a 4-parameter model (Fig 3D,F). In this case, parameters k_{s-x_s} and k_{x_s-x} in the reduced model are unidentifiable, however, PTLasso captures their linear correlation (Fig. S3A) even though these parameters are not individually constrained, providing a complete picture of the system. While signaling systems are complex and can involve large numbers of reactions, not every reaction is relevant for every function. Taken together our results demonstrate that distinct elements of a large reaction-network may be responsible for different complex behaviors and can be successfully isolated using PTLasso.

Fig 3. Motif inference from a prior network constrained with dose-response data. **A)** Reaction network diagram of the prior network. The value of the signal S is known, response R (red) is observed at multiple time points, but intermediate X (blue) is not observed. Solid lines show species conversions and dashed lines show influences. $k_{s-rs} = 10$, $k_{r-0} = 0.01$ are the rate constant values used to generate observed data for the linear dose-response model. $k_{s-rs} = 10$, $k_{xr-x} = 10$, $k_{s-xs} = 1$, $k_{r-0} = 1$ are the rate constant values used to generate observed data for the perfectly adapting dose-response model **B)** Fit of the model to the linear dose-response data (top, y-axis is in linear scale) and perfectly adapting dose-response data (bottom, y-axis is in log scale) with PT (left) and PTLasso (right). Blue lines show ensemble fits (400 samples with 1000 time points for linear dose-response, 800 samples with 2000 time points for perfectly adapting dose response), red line shows the true data (1000 time points for linear dose-response, 2000 time points for perfectly adapting dose response), and the black error bars show the mean \pm standard deviation of the observed data **C)** Frequency histograms showing probability distributions of the parameters for linear dose response fits (from 4e5 PT samples) and **D)** perfectly adapting dose-response (from 8e5 PT samples) with PT (top) and PTLasso (bottom). The pink lines show the Laplace prior boundaries, the dashed red lines show the parameter values of the known model used to generate the true data, and the x-axis range, [-12,6], represents the sampling range. Parameter distributions that deviate from the prior are necessary, while those that are confined within the Laplace prior boundaries are extraneous. **E)** Reduced model corresponding to linear dose-response and **F)** perfectly adapting dose-response highlighted in prior network. Faded nodes and arrows are extraneous and are removed from the model. Solid lines show species conversions and dashed lines show influences.



A reduced model of NF- κ B signaling without A20 feedback explains single-cell NF- κ B responses to a short TNF pulse

Complex biological signaling networks are frequently modular [41, 42] with distinct motifs such as feedback loops that operate on separate time scales [35]. To account for the modular structure of signaling we extended our Lasso approach to Group Lasso [32], a technique that applies a module-specific Lasso penalty to all reactions within a particular module (see Methods). PT combined with Group Lasso finds minimal sets of reaction modules that explain experimental data. We used this method to test the

requirement of A20 feedback to explain previously published single-cell NF- κ B responses to a short TNF pulse [43].

A prior model of NF- κ B signaling was created by combining simplified elements of models from [37] and [2]. The network was divided into three biologically motivated network modules (Fig. 4A). The I κ B and A20 modules describe negative feedback mediated by the inhibitor I κ B and negative regulator A20, respectively. The activation module includes all remaining reactions that describe the path from TNF binding to its cognate TNF-receptor (TNFR) to the eventual translocation of NF- κ B into the nucleus. The reaction rate constants within a module are constrained by a common Lasso penalty parameter (see Methods). If the penalty parameter for a module is estimated as 0, (here, 10^{-25} is used as a proxy for 0 when sampling in log scale), the entire module is removed from the simulation. To test which of the three modules are necessary to explain NF- κ B responses to a single TNF pulse, PTLasso was used to fit the model to three previously published, experimentally obtained, single-cell NF- κ B responses (Fig. 4B, Column 1) [43]. In addition to the NF- κ B data, other constraints were applied to make the system behave consistently with known biology. These constraints are listed in Table S4, and Fig. S4 demonstrates that PTLasso correctly followed the imposed parameter covariation.

The probability distributions for the module penalty parameters (Fig. 4B, Columns 2-4) show that the A20 penalty parameter is confined within the prior boundaries while the others have deviated, suggesting that to fit these particular single-cell NF- κ B trajectories, the A20 module is dispensable, whereas the I κ B and activation modules are not. The A20 module might still be essential for other biology of the system, but the model does not require the A20 module to produce these single-cell NF- κ B responses under the given experimental condition and network constraints. The fits with PTLasso were as good as the fits with PT alone (Fig. 4B, Column 1), as is demonstrated by comparing the average log-likelihoods (Fig. 4C).

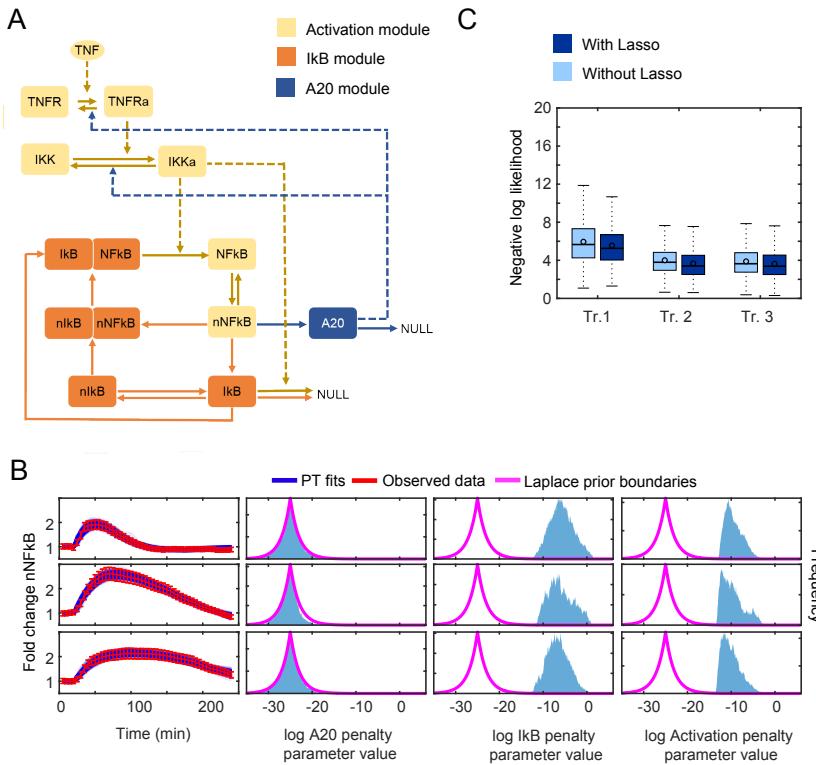
To test the requirement of A20 feedback under different experimental conditions and network constraints, we fit the model to a published single-cell NF- κ B response to continuous TNF stimulation [43]. A soft constraint that IKK responses are transient was added for consistency with published observations [44, 45]. For responses to a TNF pulse, IKK activity naturally adapts back to its baseline abundance without additional negative regulation (Fig. S6). In this case, all three module penalty parameters deviate from the prior (Fig. S7), indicating that the A20 mediated negative regulation of IKK is essential for responses to continuous TNF stimulation. Taken together, the results for the NF- κ B signaling model provide an example where PTLasso isolate reaction modules sufficient for responses to specific experimental conditions and time scales.

Discussion

In this work we have demonstrated that PT combined with Lasso is an effective approach to learn reduced models from a prior model with a larger number of reactions. Even when starting from a complete graph without prior knowledge of the underlying signaling network, PTLasso correctly identified reduced model architectures and reaction rate constants. PTLasso also correctly isolated subnetworks that are necessary for distinct concentration and temporal dose-response relationships. In a model of NF- κ B signaling, PT with Group Lasso found that in the absence of other network constraints, A20 feedback was not required to explain single-cell responses to a short TNF pulse, but is required when TNF treatment was continuous. Model reduction using PTLasso can therefore highlight aspects of the reaction network that are important for specific experimental conditions and timescales and not others.

A potential application of model reduction arises in fitting a model to data from different cell types. Differences in responses to the same experimental condition might

Fig 4. Model reduction using Group Lasso with a model of NF- κ B signaling **A)** Reaction network diagram of a simplified model of TNF-NF- κ B signaling. The colors indicate the different modules. ‘a’ or ‘n’ refer to active and nuclear versions of the species respectively. Solid lines indicate transformations and dashed lines indicate influences. **B)** Ensemble fits of the model (in blue, 288 PTLasso samples, $\mu=-25, b=2$) to a single-cell NF- κ B response to pulsatile TNF stimulation (in red). Error bars show the 10% standard deviation assumed for the likelihood function during fitting and represent measurement error (Column 1). Frequency histograms (from 5.64e6 samples) for the penalty parameter distributions corresponding to the different network modules (Columns 2-4). The pink lines show the Laplace prior boundaries, and the x-axis range, [-35,6], represents the sampling range. **C)** Box plots (5640 samples) comparing the log likelihood of the fits with PT and PTLasso ($\mu=-25, b=2$) for the three trajectories (Tr. 1-3)



be explained by differences in parameters values [31], but comparing cell-type specific parameter distributions in high dimensional space may be difficult. Reducing the number of model parameters lowers the dimensionality of the space and makes this problem easier.

A limitation of PTLasso is the large number of samples required to reach convergence, which can lead to long execution times. For the simplest examples presented here, convergence happens on the order of hours on a standard workstation computer, but for the more complex signaling systems, convergence can take several days. Most of the execution time is dedicated to converging the joint parameter distribution. Currently PT and PTLasso are both run for fixed chain lengths followed by convergence testing at the end, often generating more samples than were required to pass convergence tests. Testing convergence on-the-fly and terminating PT chains when convergence is reached would prevent unnecessary sampling and reduce the overall execution time. Approaches such as APT-MCMC [46] and Hessian-guided MCMC [20] that account for the shape of the parameter landscape during sampling could also reduce the number of samples required for convergence.

Along with working to reduce the amount of sampling, we are also investigating algorithmic modifications to reduce the execution time of individual samples or PT swaps. Synchronous swapping in our current implementation of PT requires each chain

to complete a fixed number of steps before attempting a swap. Because high temperature chains sample parameter space broadly and encounter regions where stiffness leads to long integration times, lower temperature chains often have to wait for the higher temperature chains to complete before swaps can be attempted. Asynchronous swapping [18] may therefore reduce execution times. Overall, there are still many opportunities for future PTLasso implementations to increase efficiency and applicability to larger systems biology models.

In this study we have presented a Bayesian framework that systematically dissects mechanistic ODE models of biochemical systems to identify minimal subsets of model reactions that are sufficient to explain experimental data. Technology now enables the building and simulating of highly detailed models that provide accurately reflect existing knowledge of a biochemical system. But detailed models may obscure our ability to identify underlying mechanisms. PTLasso serves as a bridge between these detailed models and simpler mechanistic explanations that can account for system behavior under specific conditions.

Supporting information

Fig S1. Hyperparameter tuning for PTLasso with a fully connected 3-node graph. **A)** Data generated for fitting. Red dashed lines show the model simulation at 8 time points with the known parameter values. Each colored line represents a noisy trajectory obtained by adding Gaussian noise to the true simulation. The black error bars show the mean and standard deviation of the 10 repeats, and is used for fitting. **B)** Hyperparameter tuning plot showing variation in the negative log likelihood distribution with μ and b (red points show the mean, and black lines show mean \pm standard deviation for 4e3 samples). The hyperparameters selected ($\mu = -10$, $b = 1$) provide the most regularization while not substantially increasing the negative log likelihood. **D)** Comparison of the log likelihood distributions (4000 samples) of the fits with PT and PTLasso ($\mu = -10$, $b = 1$). Box plots are obtained using a third party MATLAB library, aboxplot*, with outliers not shown. **D)** Example of PTLasso fits (4e3 samples) where b is too small ($\mu = -10$, $b = 0.1$) and the negative log likelihood of the fit is increased, and **E)** the corresponding parameter distributions (4e5 samples). Since the regularization strength was too high, none of the parameters deviated from the prior. *http://alex.bikfalvi.com/research/advanced_matlab_boxplot/

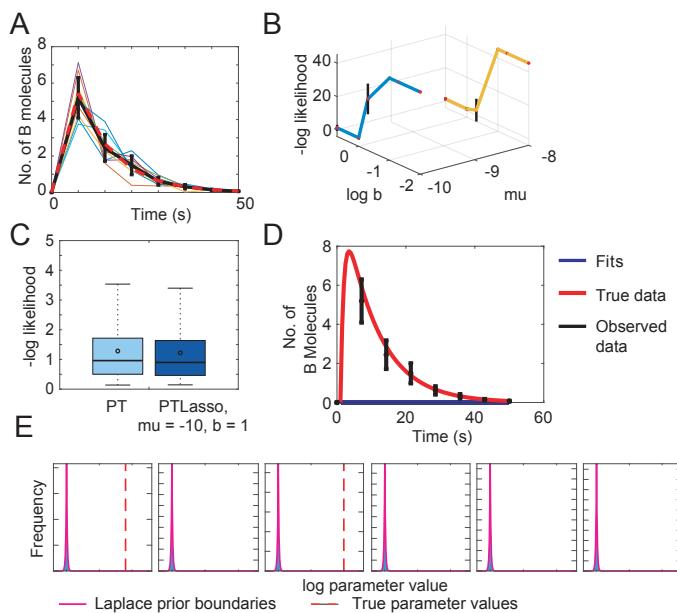


Fig S2. Hyperparameter tuning for PTLasso with a fully connected 5-node graph. A
Observed data generated for fitting. Red dashed lines show the model simulation at 8 time points with the known parameter values. Each colored line represents a noisy trajectory obtained by adding Gaussian noise to the true data. The black error bars show the mean and standard deviation of the 10 repeats, and is the observed data used for fitting. **B)** Hyperparameter tuning plot showing variation in the negative log likelihood distribution with μ and b (red points show the mean, and black lines show mean \pm standard deviation for 7e3 samples). The hyperparameters selected ($\mu = -10$, $b = 1$) provide the most regularization while not substantially increasing the negative log likelihood. **C)** Box plots comparing the log likelihood distribution (4000 samples) obtained with and without Lasso for the chosen values of hyperparameters. Box plots are obtained using a third party MATLAB library, aboxplot*, with outliers not shown. **D)** Parameter co-variation (400000 samples from the distribution) of the three selected parameters with PT and **E)** with PTLasso. *http://alex.bikfalvi.com/research/advanced_matlab_boxplot/

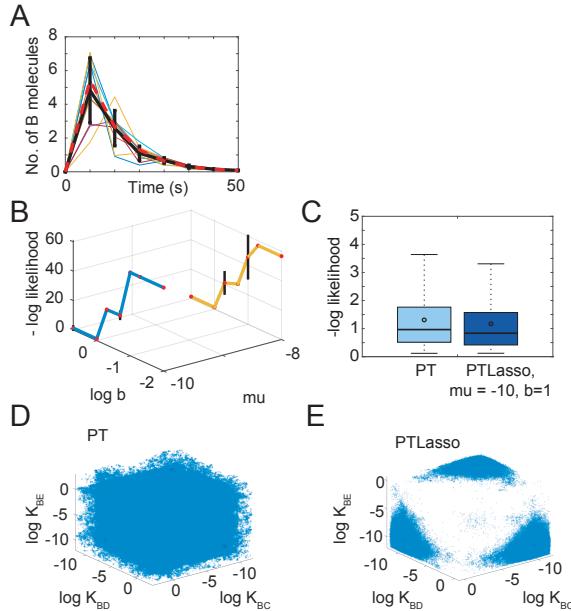


Fig S3. Hyperparameter tuning for PTLasso with dose-response motifs inferred from a prior network. **A)** Linear correlation of non identifiable parameters in the reduced perfectly adapting model shown as a scatter plot (8e5 PTLasso samples). **B)** Hyperparameter tuning plot for the linear dose response model and **C)** the perfectly adapting dose response model. The hyperparameter tuning plot shows variation in the negative log likelihood distribution with μ and b (red points show the mean, and black lines show mean \pm standard deviation, 4e2 samples for the linear dose response model and 8e2 samples for the perfectly adapting dose response model). The hyperparameters selected ($\mu = -10$, $b = 0.5$ for linear dose-response and $\mu = -10$, $b = 1$ for perfectly adapting dose-response) provide the most regularization while not substantially increasing the negative log likelihood. **D)** Box plots comparing the log likelihood distribution obtained with PT and PTLasso for the chosen values of hyperparameters for the linear dose response model (4e2 samples) and **E)** the perfectly adapting dose response model (8e2 samples). Box plots are obtained using a third party MATLAB library, aboxplot*, with outliers not shown.

*http://alex.bikfalvi.com/research/advanced_matlab_boxplot/

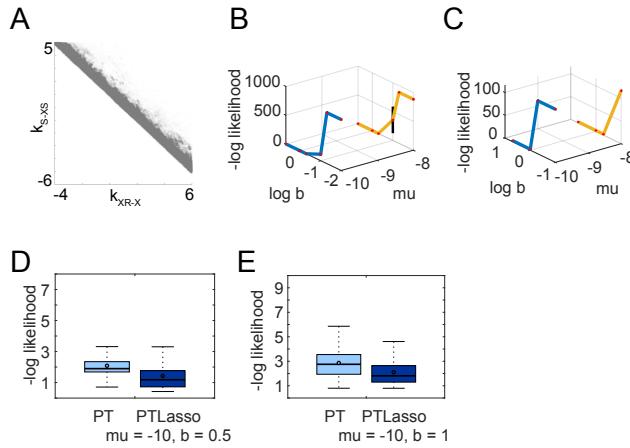


Fig S4. Constraints on parameter covariation in NF- κ B signaling. Binned scatter plots (MATLAB function binscatter with 9.4e5 samples from one of the PTLasso repeats) show the joint distributions for the pairs of parameters for which covariance was constrained during fitting.

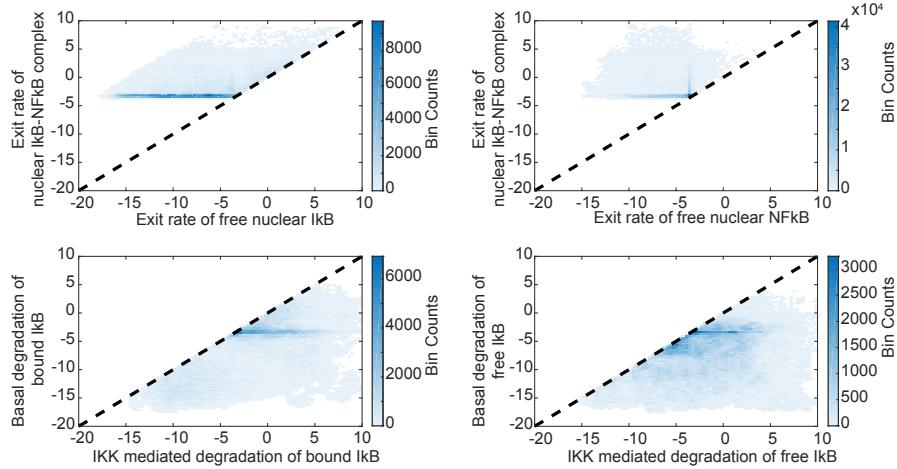


Fig S5. Marginal distributions of model parameters shown with the corresponding published values for a representative NF- κ B trajectory. Distributions of protein abundance parameters and rate constant parameters (from 5.64e6 samples) from the A20 module (blue), Activation module (yellow) and I κ B module (orange). All parameters are in logscale. Protein abundance parameters have uniform priors, and the x-axis range indicates the sampling range. Rate constant parameters are sums of the module penalty parameters and reaction-specific parameters. The pink line corresponds to a best-fit parameter set from one of the PTIlasso repeats. The dashed lines correspond to published values of parameters – Pekalski et al. [2] (red), Lee et al. [37] (black), and Kearns et al. [47] (blue). A published parameter value for a model is only included if the corresponding reaction maintained the same structure in both the published and current models. For unit conversions we used the values mentioned in Lee et al., 1 μ M of NF- κ B = 50,000 molecules/cell and applied this to other species in the model. In the parameter labels ‘import’ refers to translocation from the cytoplasm into the nucleus and ‘export’ is the reverse. ‘Complex’ refers to the NF- κ B -I κ B complex.

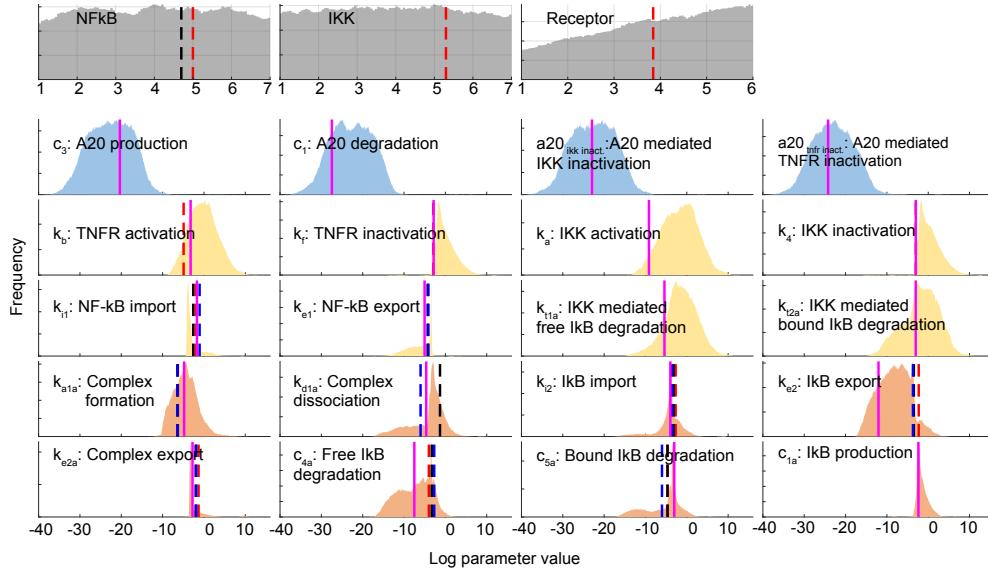


Fig S6. NF- κ B signaling model predictions. Model predictions for non-fitted variables for a representative NF- κ B trajectory (1881 samples from one of the PTIlasso repeats). Time courses are shown for 240 minutes. Suffix ‘a’ and ‘i’ refer to active and inactive versions of a species respectively. Prefix ‘c’ and ‘n’ refer to cytoplasmic and nuclear versions of a species respectively. ‘Complex’ refers to the NF- κ B -I κ B complex

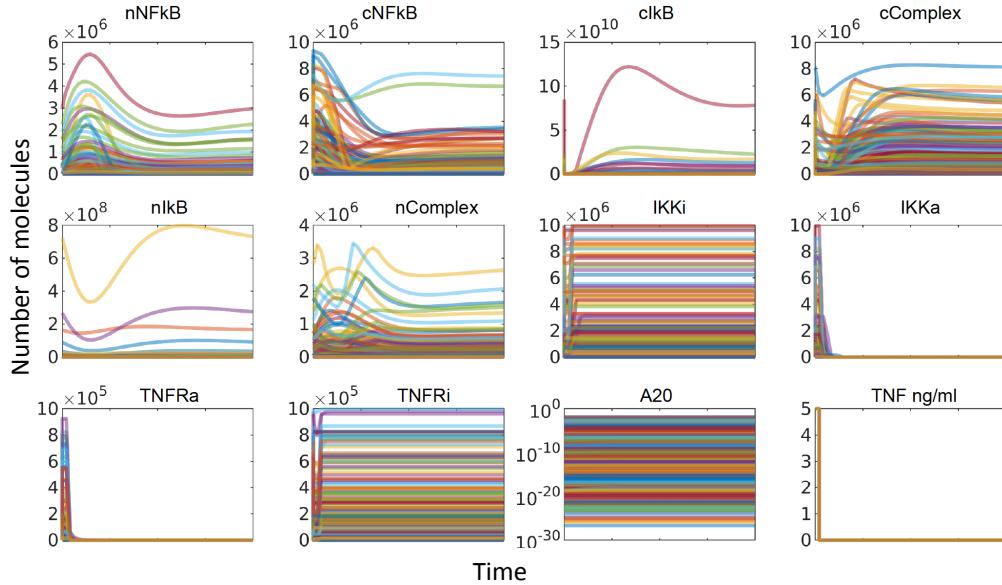


Fig S7. Group Lasso applied to NF- κ B signaling with continuous TNF stimulation. **A)** PTLasso fits of the model (in blue, 172 samples) to a single-cell NF- κ B response to TNF stimulation (in red). Errorbars show mean and standard deviation that represents the assumed 10% measurement uncertainty **B)** Frequency histograms (from 3.4e6 samples) comparing the probability distributions of the penalty parameters corresponding to the different network modules. All the distributions have deviated from the Laplace prior boundaries shown in pink, indicating that all the modules in the network were essential to fit this data.

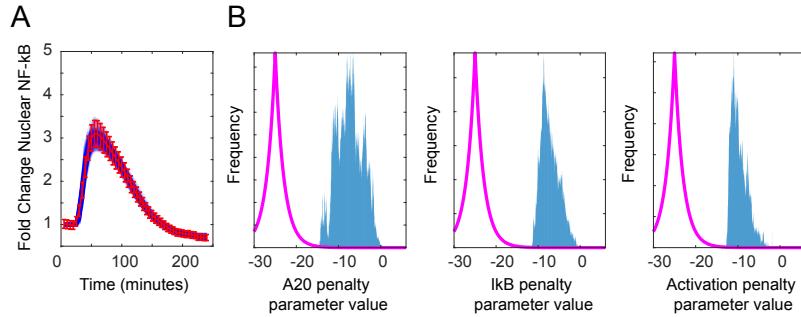


Table S1. Maximum PSRF across all model parameters for each example shown up to 4 significant digits.

Model name, No. of PT samples/chain	With Lasso	Without Lasso
3-node graph, 4e5	1.000	1.002
5-node graph, 4e5	1.032	1.001
Linear Dose-Response, 2e5	1.001	1.001
Perfectly adapting Dose-Response, 8e5	1.015	1.004
Group Lasso with NF- κ B signaling:		
Trajectory 1, 5.6e6	1.025	1.024
Trajectory 2, 5.6e6	1.027	1.057
Trajectory 3, 5.6e6	1.027	1.007

Table S2. MPSRF values for all PT runs in this study shown up to 4 significant digits.

Model name, No. of PT samples/chain	With Lasso	Without Lasso
3-node graph, 4e5	1.001	1.002
5-node graph, 4e5	1.031	1.010
Linear Dose-Response, 2e5	1.003	1.004
Perfectly adapting Dose-Response, 8e5	1.012	1.006
Group Lasso with NF- κ B signaling:		
Trajectory 1, 5.6e6	1.074	1.094
Trajectory 2, 5.6e6	1.120	1.200
Trajectory 3, 5.6e6	1.129	1.014

Table S3. Step acceptance rates for the lowest temperature chain of all PT runs in this study.

Model name, No. of PT samples	With Lasso (repeat 1, 2)	Without Lasso (repeat 1, 2)
3-node graph, 4e5	0.24, 0.25	0.24, 0.25
5-node graph, 4e5	0.24, 0.23	0.24, 0.26
Linear Dose-Response, 2e5	0.24, 0.22	0.24, 0.24
Perfectly adapting Dose-Response, 8e5	0.24, 0.24	0.25, 0.25
Group Lasso with NF- κ B signaling, 5.8e6		
Trajectory 1	0.26, 0.26	0.23, 0.22
Trajectory 2	0.26, 0.25	0.45, 0.26
Trajectory 3	0.24, 0.25	0.26, 0.22

Table S4. Hard constraints in NF- κ B signaling fit

Index	Constraint description	Constraint
1	Exit rate of NF- κ B -I κ B from the nucleus is greater than free NF- κ B	$ke2a > ke1$
2	Exit rate of NF- κ B -I κ B from the nucleus is greater than free I κ B	$ke2a > ke2$
3	IKK mediated degradation of free cytosolic I κ B is greater than the basal rate	$kt1a > c4a$
4	IKK mediated degradation of NF- κ B bound cytosolic I κ B is greater than the basal rate	$kt2a > c5a$
5	At equilibrium there is some NF- κ B in the nucleus	$\frac{nNF\kappa B_{eq} + nNF\kappa B - I\kappa B_{eq}}{\text{TotalNF}\kappa B} \in [0.05, 0.5]$
6	At least 1 TNFR molecule/cell is activated	$\min(TNFRa) > 1$
7	At least 1 IKK molecule/cell is activated	$\min(IKKa) > 1$
8	Equilibrium check before TNF stimulation	Rate of change for each model species $< 1e-4$ molecules/5e5 seconds

Table S5. PSRF to show convergence of energy distributions when combining PT runs for NF- κ B signaling fit.

Trajectory	Lasso group 1, M=6, N=5.6e6	Lasso group 2, M=6, N=5.6e6	Without Lasso group 1, M=6, N=5.6e6	Without Lasso group 2, M=6, N=5.6e6
1	1.041	1.016	1.049	1.021
2	1.139	1.134	1.011	1.003
3	1.028	1.009	1.001	1.003

Table S6. Swap acceptance rates for the two lowest temperature chains of all PT runs in this study.

Model name, No. of PT samples/chain	With Lasso (repeat 1, 2)	Without Lasso (repeat 1, 2)
3-node graph, 4e5	0.24, 0.33	0.25, 0.35
5-node graph, 4e5	0.33, 0.47	0.32, 0.26
Linear Dose-Response, 2e5	0.26, 0.97	0.24, 0.24
Perfectly adapting Dose-Response, 8e5	0.17, 0.37	0.25, 0.26
Group Lasso with NF- κ B signaling, 5.6e5		
Trajectory 1	0.28, 0.25	0.19, 0.25
Trajectory 2	0.17, 0.17	0.29, 0.32
Trajectory 3	0.16, 0.16	0.26, 0.28

Acknowledgments

This work was funded by NIH grant R35-GM119462 to RECL, and by JRF via the NIGMS-funded (P41-GM103712) National Center for Multiscale Modeling of Biological Systems (MMBioS).

References

1. Liepe J, Kirk P, Filippi S, Toni T, Barnes CP, Stumpf MPH. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*. 2014;9(2):439–456. doi:10.1038/nprot.2014.025.
2. Pękalski J, Zuk PJ, Kochańczyk M, Junkin M, Kellogg R, Tay S, et al. Spontaneous NF- κ B activation by autocrine TNF α signaling: A computational analysis. *PLoS ONE*. 2013;8(11). doi:10.1371/journal.pone.0078887.
3. Hat B, Kochańczyk M, Bogdał MN, Lipniacki T. Feedbacks, Bifurcations, and Cell Fate Decision-Making in the p53 System. *PLoS Computational Biology*. 2016;12(2). doi:10.1371/journal.pcbi.1004787.
4. Faeder JR, Hlavacek WS, Reischl I, Blinov ML, Metzger H, Redondo A, et al. Investigation of Early Events in Fc RI-Mediated Signaling Using a Detailed Mathematical Model. *The Journal of Immunology*. 2003;170(7):3769–3781. doi:10.4049/jimmunol.170.7.3769.
5. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. In: *BioSystems*. vol. 83; 2006. p. 136–151.
6. Chylek LA, Holowka DA, Baird BA, Hlavacek WS. An interaction library for the Fc ϵ RI signaling network. *Frontiers in Immunology*. 2014;5(APR). doi:10.3389/fimmu.2014.00172.
7. Alon U. Network motifs: theory and experimental approaches. *Nature reviews Genetics*. 2007;8(6):450–61. doi:10.1038/nrg2102.
8. Goentoro L, Shoval O, Kirschner MW, Alon U. The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. *Molecular Cell*. 2009;36(5):894–899. doi:10.1016/j.molcel.2009.11.018.
9. Tyson JJ, Chen KC, Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*. 2003;15(2):221–231. doi:10.1016/S0955-0674(03)00017-6.
10. Snowden TJ, van der Graaf PH, Tindall MJ. Methods of Model Reduction for Large-Scale Biological Systems: A Survey of Current Methods and Trends. *Bulletin of Mathematical Biology*. 2017;79(7):1449–1486. doi:10.1007/s11538-017-0277-2.
11. Quaiser T, Dittrich A, Schaper F, Mönnigmann M. A simple work flow for biologically inspired model reduction - application to early JAK-STAT signaling. *BMC Systems Biology*. 2011;5. doi:10.1186/1752-0509-5-30.
12. Raue A, Steiert B, Schelker M, Kreutz C, Maiwald T, Hass H, et al. Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics*. 2015;31(21):3558–3560. doi:10.1093/bioinformatics/btv405.
13. Prescott TP, Papachristodoulou A. Layered decomposition for the model order reduction of timescale separated biochemical reaction networks. *Journal of Theoretical Biology*. 2014;356:113–122. doi:10.1016/j.jtbi.2014.04.007.

14. Ciliberto A, Capuani F, Tyson JJ. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Computational Biology*. 2007;3(3):0463–0472. doi:10.1371/journal.pcbi.0030045.
15. Maurya MR, Bornheimer SJ, Venkatasubramanian V, Subramaniam S. Mixed-integer nonlinear optimisation approach to coarse-graining biochemical networks. *IET systems biology*. 2009;3(1):24–39. doi:10.1049/iet-syb:20080098.
16. Bhattacharjee B, Schwer DA, Barton PI, Green WH. Optimally-reduced kinetic models: Reaction elimination in large-scale kinetic mechanisms. *Combustion and Flame*. 2003;135(3):191–208. doi:10.1016/S0010-2180(03)00159-7.
17. Petzold L, Zhu W. Model reduction for chemical kinetics: An optimization approach. *AIChE Journal*. 1999;45(4):869–886. doi:10.1002/aic.690450418.
18. Gupta S, Hainsworth L, Hogg J, Lee R, Faeder J. Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology. In: *Proceedings - 26th Euromicro International Conference on Parallel, Distributed, and Network-Based Processing, PDP 2018*; 2018. p. 690–697.
19. Klinke DJ. An empirical Bayesian approach for model-based inference of cellular signaling networks. *BMC Bioinformatics*. 2009;10. doi:10.1186/1471-2105-10-371.
20. Eydgahi H, Chen WW, Muhlich JL, Vitkup D, Tsitsiklis JN, Sorger PK. Properties of cell death models calibrated and compared using Bayesian approaches. *Molecular Systems Biology*. 2014;9(1):644–644. doi:10.1038/msb.2012.69.
21. Malkin AD, Sheehan RP, Mathew S, Federspiel WJ, Redl H, Clermont G. A Neutrophil Phenotype Model for Extracorporeal Treatment of Sepsis. *PLoS Computational Biology*. 2015;11(10). doi:10.1371/journal.pcbi.1004314.
22. Earl DJ, Deem MW. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*. 2005;7(23):3910. doi:10.1039/b509983h.
23. Tibshirani [U+FF0C] Robert. Regression Shrinkage and Selection via the Lasso Robert Tibshirani. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2007;58(1):267–288. doi:10.1111/j.1467-9868.2011.00771.x.
24. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, et al. The inferelator: An algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biology*. 2006;7(5). doi:10.1186/gb-2006-7-5-r36.
25. Wu TT, Chen YF, Hastie T, Sobel E, Lange K. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*. 2009;25(6):714–721. doi:10.1093/bioinformatics/btp041.
26. Frost HR, Amos CI. Gene set selection via LASSO penalized regression (SLPR). *Nucleic Acids Research*. 2017;45(12). doi:10.1093/nar/gkx291.
27. Lu Y, Zhou Y, Qu W, Deng M, Zhang C. A Lasso regression model for the construction of microRNA-target regulatory networks. *Bioinformatics*. 2011;27(17):2406–2413. doi:10.1093/bioinformatics/btr410.

28. Vignes M, Vandel J, Allouche D, Ramadan-Alban N, Cierco-Ayrolles C, Schiex T, et al. Gene regulatory network reconstruction using bayesian networks, the dantzig selector, the lasso and their meta-analysis. *PLoS ONE*. 2011;6(12). doi:10.1371/journal.pone.0029165.
29. Li J, Das K, Fu G, Li R, Wu R. The Bayesian lasso for genome-wide association studies. *Bioinformatics*. 2011;27(4):516–523. doi:10.1093/bioinformatics/btq688.
30. Biswas S, Lin S. Logistic Bayesian LASSO for Identifying Association with Rare Haplotypes and Application to Age-Related Macular Degeneration. *Biometrics*. 2012;68(2):587–597. doi:10.1111/j.1541-0420.2011.01680.x.
31. Steiert B, Timmer J, Kreutz C. L1 regularization facilitates detection of cell type-specific parameters in dynamical systems. In: *Bioinformatics*. vol. 32; 2016. p. i718–i726.
32. Meier L, Van De Geer S, Bühlmann P. The group lasso for logistic regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2008;70(1):53–71. doi:10.1111/j.1467-9868.2007.00627.x.
33. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*. 1953;21(6):1087–1092. doi:10.1063/1.1699114.
34. Chib S, Greenberg E. Understanding the metropolis-hastings algorithm. *American Statistician*. 1995;49(4):327–335. doi:10.1080/00031305.1995.10476177.
35. Atay O, Skotheim JM. Modularity and predictability in cell signaling and decision making. *Molecular Biology of the Cell*. 2014;25(22):3445–3450. doi:10.1091/mbc.e14-02-0718.
36. Kautz H, Selman B, Jiang Y. A general stochastic approach to solving problems with hard and soft constraints; 2017. p. 573–585.
37. Lee REC, Walker SR, Savery K, Frank DA, Gaudet S. Fold change of nuclear NF- κ B determines TNF-induced transcription in single cells. *Molecular Cell*. 2014;53(6):867–879. doi:10.1016/j.molcel.2014.01.026.
38. Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*. 1998;7(4):434–455. doi:10.1080/10618600.1998.10474787.
39. Roberts GO, Gelman A, Gilks WR. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*. 1997;7(1):110–120. doi:10.1214/aoap/1034625254.
40. Rosenbaum M, Tsybakov AB. Sparse recovery under matrix uncertainty. *Annals of Statistics*. 2010;38(5):2620–2651. doi:10.1214/10-AOS793.
41. Lauffenburger DA. Cell signaling pathways as control modules: Complexity for simplicity? *Proceedings of the National Academy of Sciences*. 2002;97(10):5031–5033. doi:10.1073/pnas.97.10.5031.
42. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature*. 1999;402(S6761):C47–C52. doi:10.1038/35011540.
43. Zhang Q, Gupta S, Schipper DL, Kowalczyk GJ, Mancini AE, Faeder JR, et al. NF- κ B Dynamics Discriminate between TNF Doses in Single Cells. *Cell Systems*. 2017;5(6):638–645.e5. doi:10.1016/j.cels.2017.10.011.

44. Tarantino N, Tinevez JY, Crowell EF, Boisson B, Henriques R, Mhlanga M, et al. Tnf and il-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures. *Journal of Cell Biology*. 2014;204(2):231–245. doi:10.1083/jcb.201307172.
45. Pabon NA, Zhang Q, Cruz JA, Schipper DL, Camacho CJ, Lee REC. A network-centric approach to drugging TNF-induced NF- κ B signaling. *Nature Communications*. 2019;10(1). doi:10.1038/s41467-019-08802-0.
46. Zhang LA, Urbano A, Clermont G, Swigon D, Banerjee I, Parker RS. APT-MCMC, a C++/Python implementation of Markov Chain Monte Carlo for parameter identification. *Computers and Chemical Engineering*. 2018;110:1–12. doi:10.1016/j.compchemeng.2017.11.011.
47. Kearns JD, Basak S, Werner SL, Huang CS, Hoffmann A. I κ B ϵ provides negative feedback to control NF- κ B oscillations, signaling dynamics, and inflammatory gene expression. *Journal of Cell Biology*. 2006;173(5):659–664. doi:10.1083/jcb.200510155.