Statistical and Mechanistic Approaches to Study Cell Signaling Dynamics

by

Sanjana Gupta

B.Tech., Indian Institute of Technology Guwahati, 2014

Submitted to the Graduate Faculty of

the School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Sanjana Gupta

It was defended on

July 9, 2020

and approved by

Ipsita Banerjee, PhD, Department of Chemical and Petroleum Engineering, University of

Pittsburgh

Tuomas W. Sandholm, PhD, Computer Science Department, Carnegie Mellon University

Dissertation Director: Robin E.C. Lee, PhD, Department of Computational and Systems

Biology, University of Pittsburgh

Dissertation Director: James R. Faeder, PhD, Department of Computational and Systems

Biology, University of Pittsburgh

Statistical and Mechanistic Approaches to Study Cell Signaling Dynamics

Sanjana Gupta, PhD

University of Pittsburgh, 2020

Cells use complex signaling systems to constantly detect environmental changes, relay extracellular information from the cell membrane to the nucleus, and drive cell responses, such as transcription. The ability of each single cell to dynamically respond to changes in its environment is the basis for healthy, functioning, multicellular beings. Diseases often arise from dysregulated signaling, and our ability to manipulate cell responses, that stems from our growing understanding of signaling processes, is often the basis for disease treatments.

Computational approaches can complement experimental studies of cellular systems, allowing us to formalize our growing body of knowledge of cellular biochemistry. Mechanistic modeling provides a natural framework to describe and simulate complex systems with many system components and causal interactions that often lead to non-intuitive emergent behavior, lending itself well to the analysis of signaling systems. Statistical approaches can complement mechanistic modeling by enabling an analysis of complex input-output relationships in the data, providing insight into how cells translate input environmental cues into output responses, even when the underlying mechanisms are only partially understood.

In this thesis, we explore both mechanistic and statistical approaches and address several challenges in modeling signaling processes within a cell, and signaling heterogeneity between cells, using the NF-κB pathway as a model system. First, we evaluate methods to efficiently determine numerical values of model parameters, enabling model simulations that are comparable to experimental data. Second, we develop methods to identify reduced submodels that are sufficient for the data, highlighting simple mechanisms that drive emergent behavior. Third, switching gears to study signaling heterogeneity, we use information-theoretic analyses to evaluate the capabilities of the NF-κB pathway to effectively transduce cytokine dosage information in the presence of biochemical noise. Finally, we develop a framework to calibrate mechanistic models to heterogeneous signaling data, enabling simulation-based analyses of single-cell signaling capabilities.

Table of Contents

List of Tables

List of Figures

Preface

I would like to start by thanking my advisors, Dr. James Faeder and Dr. Robin Lee. Your mentorship and guidance over the years has helped me become the scientist I am today, and this dissertation would not have been possible without your support. I am grateful to have had the opportunity to work closely with and learn so much from both of you.

I feel lucky to have been a part of two labs during my PhD, and I would like to thank all the current and previous members of both the Faeder and Lee Labs. You have all been wonderful colleagues, and I have really enjoyed my time working with you. I would like to especially thank Gabriel Kowalczyk, Chaitanya Mokashi, David Schipper, Yue Guo, Dr. J. Agustin Cruz, Kunal Aggarwal, Dr. Ali Saglam, Dr. Cihan Kaya, Dr. John Sekar, Dr. Jose-Juan Tapia, Dr. Robert Sheehan, Neha Cheemalavagu and Steven Smeal, both for your friendship over the years, and for many interesting scientific discussions.

I would like to thank my committee members, Dr. Tuomas Sandholm and Dr. Ipsita Banerjee, for taking the time to serve on my committee and provide valuable feedback on my dissertation. I would also like to thank all the CMU-Pitt Computational Biology (CPCB) program staff and administrators, especially Kelly Gentille and Nancy Linderman, for always making my graduate school experience so much smoother.

I would like to extend a deep thanks to all of my friends, especially Jocelyn Sunseri, Michael Kleyman, Bhavya Balu, Dr. Raghavendran Partha, Dr. Wayne Mao and Dr. Kevin Kaczorowski. Your friendship has made my graduate school experience so much more enjoyable and has helped me whenever things got difficult, and I am so grateful for that.

Finally, my biggest thanks is to my family, especially my sister Sneha Gupta, and my parents, Dr. Debjani Gupta and Dr. Satish Gupta. Your constant and unconditional love and support has made everything possible, and this thesis is dedicated to you.

1.0   Introduction

1.1   Cell signaling pathways and their computational representations

Cells are fundamental units of all living systems [1]. Signaling underlies the basic ability of an individual cell to interact with its environment, communicate with other cells, and effectively organize into complex hierarchical multicellular organisms [2,3]. Since dysregulation of cell signaling is the basis for many human diseases [3,4], an increased understanding of signaling mechanisms will lead to the ability to artificially manipulate cell responses and design effective treatment strategies for disease [5–9]. Therefore, a key question in Systems Biology, a discipline that studies the behavior of complex biological systems by considering the many interactions between system components [10], is "How do cells signal?".

Cell signaling can be described as the process by which cells use complex networks of interacting proteins and other biomolecules to translate environmental cues into responses and cell fate decisions [11,12]. While these environmental cues may be mechanical, electrical, or biochemical in nature [11], we restrict the scope of this thesis to biochemical signal transduction, which begins with a ligand that is detected by binding to a cell-surface receptor [3]. This usually triggers a cascade of biochemical reactions culminating in a cell response, such as transcription of a particular gene [3,11]. The sequence of biochemical events starting from a particular ligand binding its cognate receptor and ending in a cellular response, is referred to as a signaling pathway [3]. Because signaling pathways are usually very complex and nonlinear [13], and often include network components and interactions that cannot directly be experimentally observed and confirmed [14], we have increasingly been turning towards mathematical and computational models to understand signaling systems [15–17].

Models of biological systems include both statistical models, that aim to characterize input-output relationships in biological data, and mechanistic models, that include explicit descriptions of causal biological mechanisms [18,19]. In this thesis, we see instances of both of these types of models applied to cell signaling. Mechanistic models describe both known and hypothesized biological mechanisms, and enable predictions that test our hypotheses

1

[20]. There are a variety of methods for simulations of mechanistic biochemical models, and the exact choice of method often depends on the desired degree of spatial and molecular resolution [21]. If the system is assumed to be well-mixed, implying that each biochemical species is distributed uniformly throughout the cell (or cellular compartment being modeled), then a commonly employed method is Ordinary Differential Equation (ODE) modeling [21, 22].

In an ODE model of signaling, every biochemical reaction proceeds at a rate that is determined by a rate law. There is a differential equation for each biochemical species in the signaling pathway that describes its evolution over time, as dictated by the rate laws for reactions that include the species. Since the same biochemical species may participate in multiple reactions, the equation describing its evolution may have contributions from multiple rate laws. A common example of a rate law, given by the mass action assumption, is the product of a rate constant and the abundances or concentrations of each reactant [23]. ODE models are associated with several parameters, such as the initial species abundances that provide the initial conditions to solve the ODEs and the rate constants used in the rate laws. The first set of challenges that we address in this thesis is, given a model of signaling and some experimental data, how can we efficiently determine the numerical values of these parameters such that the model output matches the data, a problem called *parameter estimation*, and can we reduce the number of model parameters that are needed to explain the data, a problem called *model reduction*.

1.1.1   Estimating parameters of cell signaling models

While there are several methods in the parameter estimation literature that span various fields [24, 25], there are certain challenges that are more unique to mechanistic models in systems biology. These challenges include the sparse availability of data compared to the large number of model parameters that are needed to accurately express biochemical mechanisms [26], temporal correlations in the time-series data used for fitting, and correlations in model parameters [27]. Mechanistic models of cell signaling therefore often possess a property called *non-identifiability* [28], where many unique parameter sets may fit the data

equally well, and it becomes desirable to estimate not just best-fit parameters, but also the uncertainty in the parameter estimates [20, 29]. Bayesian parameter estimation methods achieve this by estimating the posterior probability distributions of model parameters given the data [20,29–31]. In Chapter 2, we systematically investigate Parallel Tempering (PT), a physics-based Bayesian sampling method that has gained widespread success in the field of molecular dynamics simulations [30], for its utility and efficiency for estimating parameters of cell signaling models [32].

1.1.2 Reducing cell signaling models

While having more reactions and parameters increases the expressiveness of mechanistic biochemical models, the resulting large reaction networks may obscure simpler network structures that can drive specific cell behaviors [33, 34]. Detailed models more accurately represent our knowledge of a system, however the difficulty in parsing complex networks of all known components and interactions has often led to them being described as "hairballs" [35–37]. As advances in experimental technology are allowing us to probe more detailed aspects of a cell and our knowledge of cellular biochemistry is rapidly increasing, there is an increased need for model reduction methods that can bridge the gap between detailed models that can serve as libraries [38] for everything we know about a cell signaling system, and simpler mechanistic explanations that are sufficient for particular emergent properties, for example an incoherent feedforward loop that leads to fold-change detection [33]. As a step in this direction, in Chapter 3, we develop a framework that combines Bayesian Parameter Estimation using PT with Lasso, a sparsity-promoting penalty commonly used in Machine Learning [39], for simultaneous parameter estimation and model reduction [12].

1.2 NF-κB signaling as a model signal transduction pathway

While we develop or use analyses that are broadly applicable to cell signaling studies, in this thesis we focus on applications to the TNF-NF-κB signaling pathway, which is the

activation of transcription factor NF-κB mediated by the inflammatory cytokine, Tumor Necrosis Factor (TNF). NF-κB signaling is a key regulator of innate immunity and controls the transcription of several inflammatory genes [40]. In a resting cell, NF-κB is sequestered in the cytoplasm by binding to its inhibitor, IκB [40]. TNF on binding to its cognate receptor, TNFR, stimulates the recruitment, and subsequent activation, of IκB kinases (IKKs) to a protein complex at the receptor [41]. Active IKK phosphorylates IκB, targeting it for degradation and freeing NF-κB to translocate into the nucleus where it drives transcription of its own negative regulators, A20 and IκB, as well as various inflammatory genes. A20 and IκB create negative feedback loops that drive NF-κB out of the nucleus [42].

A recurring theme in cell signaling is that dynamics are important. It has been shown that different temporal patterns of TNF stimulation are capable of eliciting distinct cell responses [43, 44], and that the dynamics of NF-κB control expression of distinct genes that can be classified as early, intermediate, or late responding [45]. To study the dynamics of the TNF-NF-κB response in single cells, we use live-cell microscopy data [46]. A fluorescent protein is fused with NF-κB, and live-cell microscopy is used to observe the translocation of NF-κB in and out of the nucleus, measured as the change in nuclear fluorescence intensity, allowing us to track the single-cell NF-κB response over time in several cells that are exposed to the same TNF stimulus condition [46].

While NF-κB signaling is a relatively well-studied pathway, there are still open questions surrounding mechanisms of regulation, such as the requirement of various feedback elements under different temporal stimulus conditions, mechanisms of heterogeneity or noise, and the impact of noise on the capabilities of the pathway to accurately transduce information about the cellular environment, such as information about cytokine dose [47, 48]. In this thesis, we use both mechanistic approaches, with Bayesian parameter estimation and model reduction, and statistical tools from information theory, to address these outstanding questions.

In Chapter 3, we demonstrate an application of PT with Lasso to test the requirement of NF-κB signaling feedback loops in single cells exposed to either pulsatile or continuous TNF stimulation [12]. In Chapters 4 and 5, we switch gears from studying mechanisms of NF-κB signaling within independent single cells, to studying the heterogeneity in NF-κB signaling that is commonly observed, even in genetically identical cell populations.

1.3 Heterogeneity in cell signaling systems

In studying mechanisms of cell signaling, with the goal of manipulating cell behavior, it is unavoidable to consider the commonly observed heterogeneity between cells. Heterogeneity is a common property in many biological signaling systems [49], and even genetically identical cells often have markedly different responses to the same external stimulus [49–51]. This is somewhat surprising as dysregulation of signaling is associated with many human diseases [3, 4], for example dysregulation of TNF-NF-κB signaling is associated with chronic inflammation [52], and so one might expect signaling to be strictly regulated and responses across cells to be highly consistent. Understanding the mechanistic basis for cells responding differently to the same stimulus, and understanding how the variability or "noise" in the responses affects the signal processing capabilities of a cell, may provide valuable clues to design optimal treatment strategies to manipulate cell responses.

In studying cell heterogeneity, the first question we consider is – does noise limit the signal processing capabilities of individual cells [53]? An interpretation of the significant heterogeneity observed in TNF-NF-κB signaling that arose in the field several years ago, using an information-theoretic analysis, is that this signaling system is too noisy to accurately resolve environmental cues such as the concentration of extracellular TNF, and can instead, simply behave as a switch that can detect the presence or absence of TNF [54]. However, given the many roles of this pathway as a central mediator of inflammatory responses [52], an inability of single cells to grade NF-κB responses to TNF dose would be rather surprising. In Chapter 4, we combine more recent advances in the field of information theory for biochemical systems [53, 55], advances in knowledge about NF-κB signaling [44, 56], and hundreds of dynamic single-cell TNF-NF-κB responses measured in response to diverse experimental conditions using live-cell imaging, to investigate this question of dose-discrimination in TNF-NF-κB signaling [46]. Understanding single-cell processing capabilities is a key step in eventually being able to control cell responses. Knowing whether a single cell is acting as a switch, or is able to grade its response to dose, may later help guide treatment strategies to control single-cell behavior.

While in Chapter 4 we consider the *effect* of cell heterogeneity, in Chapter 5 we address

the complementary question of understanding the *cause* of heterogeneity. An approach that has been successfully applied in the literature is to hypothesize that variation arises from systematic differences in signaling elements, that are expressed as differences in model parameters [22, 57–59]. Often only a few individual-level parameters are assumed to drive the heterogeneity, while the rest are shared and common to all the individuals [22, 58, 59]. In particular, protein expression variation has commonly been reported as a source of signaling variability in cellular systems [50, 60–62]. Building closely on a previously described approach [59], we present a workflow that uses PT and other sampling approaches to calibrate the individual parameters, assumed to be abundances of key signaling proteins, and shared parameters of a model [12] to a large set of heterogeneous single-cell NF-κB responses from live-cell imaging data [46]. The calibrated model of heterogeneity is then further used to simulate single-cell dose-response behavior and examine differences in signaling capabilities of individual cells that can arise from differences in protein abundances.

1.4   Thesis goals

Overall, this thesis presents several computational strategies, both mechanistic and statistical, to improve our understanding of cell signaling and demonstrates applications to the NF-κB signaling pathway as a model system. In Chapters 2 and 3 we address two technical challenges in modeling complex signaling systems *within cells*, namely parameter estimation and model reduction. Going further, in Chapters 4 and 5 we address the challenge of studying the heterogeneity in signaling systems *between cells*, specifically applied to NF-κB signaling. Finally, in Chapter 6 we summarize our conclusions and present our findings in the broader context of the field of systems biology and cell signaling.

The work in Chapters 2, 3 and 4 has been published in articles [32], [12] and [46] respectively. The work in Chapter 5 is preliminary, and we discuss several avenues for further investigation. Additional graduate work, not included in this thesis, includes a book chapter on spatial stochastic modeling using the software package, MCell, and graphical user interface, CellBlender [21].

2.0 Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology

2.1 Introduction

Mathematical and computational models have been gaining widespread use as tools to summarize our understanding of biological systems and to make novel predictions that can be tested experimentally [63,64]. Doing this requires a model to be correctly parameterized. Parameter estimation, the process of inferring model parameters from experimental data, typically involves defining a cost function that quantifies the discrepancy between the model output and the data, and then performing a search for parameterizations that minimize the cost [20,65].

There are many commonly-used methods for finding parameter sets that minimize the model cost. These can broadly be divided into gradient-based and gradient-free methods. Gradient-based methods are local optimization methods that iteratively use the gradient of the cost function to compute a search direction and step length, followed by updating the parameters and checking for convergence [65]. Popular gradient-based methods in systems biology include gradient descent, Newton's method, the Gauss-Newton algorithm, and the Levenberg-Marquardt algorithm [24]. However, these methods can fail to find the global minimum when landscapes are discontinuous or multi-modal, as is frequently the case for large biological models, which can have many more parameters than independent data points to constrain the model [66].

Gradient-free methods have the advantage that the landscape need not be smooth, but local search methods, such as the Nelder-Mead simplex, become inefficient for high-dimensional problems [25]. Gradient-free global optimization methods such as genetic algorithms and particle swarm optimization can be effective at finding optimal solutions in high-dimensional spaces [67]. However, the combination of high-dimensional parameter spaces and the limited amount of data available from typical biological experiments often means that multiple parameter combinations equivalently describe the experimental data, which is referred to as the parameter identifiability problem [68]. When parameters are non-identifiable, a single parameter set is insufficient to describe the feasible space of parameters associated with a model.

Bayesian methods solve this problem naturally by attempting to estimate the probability distribution of the model parameters given the experimental data [20], which allows simultaneous determination of best-fit parameters and parameter sensitivities, while also providing a framework to introduce prior information that the modeler may have about the parameters. Bayesian methods include likelihood-based approaches, such as Markov Chain Monte Carlo (MCMC) methods [31], and likelihood-free approaches, such as Approximate Bayesian Computation (ABC) [20].

MCMC is commonly used in systems biology, but slow convergence is often a major bottleneck for standard sampling algorithms, such as Metropolis-Hastings (MH) [31]. The development of modular and rule-based software for model construction and simulation [69–71], allows for the construction of increasingly complex models (e.g., [72]), which combined with the increasing availability of single-cell data [57] motivates the need for accelerated methods for Bayesian parameter estimation. Parallel tempering (PT) is a physics-based MCMC method that efficiently samples a probability distribution and can accelerate convergence over conventional MCMC methods [30]. This method has been widely used for molecular dynamics simulations to sample the conformational space of biomolecules [73, 74], but is less common in systems biology [66, 75, 76]. Here, we describe key algorithmic elements of the method, provide a software implementation, and evaluate its performance on a series of biological models of increasing complexity.

The remainder of this chapter is organized as follows: In Sec. 2.2, we describe the MH

and PT algorithms as well as the ABC and ABC-SMC methods used by the software ABC-SysBio [20], which we will later use for comparison. We also include a brief description of the PTEMPEST software for Bayesian parameter estimation. In Sec. 2.3 we present a series of examples of increasing complexity to test the performance of PT relative to MH with regards to quality of fit, convergence speed, and sampling efficiency. We include a comparison with ABC-SysBio and further show an application of using Bayesian methods with Laplace priors to achieve model reduction. Finally, in Sec. 2.4 we discuss our main findings, limitations, and areas for future work.

2.2   Methods

Bayesian parameter estimation methods infer the posterior distribution that describes the uncertainty in the parameter values that remains even after the data is known [20]. The probability of observing the parameter set θ given the data Y is given by Bayes' rule

$$p(\theta|Y) \propto p(Y|\theta)p(\theta),$$

where $p(Y|\theta)$ is the conditional probability of Y given θ, which is described by a *likelihood model*, and $p(\theta)$ is the independent probability of θ, often referred to as the *prior distribution* on model parameters. This distribution represents our prior beliefs about the model parameters, and can be used to restrict parameters to a range of values or even to limit the number of nonzero parameters, as discussed further below.

2.2.1   MCMC Methods

MCMC methods for parameter estimation sample from the posterior distribution, $p(\theta|Y)$, by constructing a Markov chain with $p(\theta|Y)$ as its stationary distribution. The key required elements are:

- *A likelihood model* that gives $p(Y|\theta)$. Assuming the model is continuous (e.g., an ordinary differential equation (ODE) model) and Gaussian experimental measurement error, the likelihood function is given by

$$L = e^{-\Sigma_S \Sigma_T (Y_{\text{sim}} - Y_{\text{expt}})^2 / 2\sigma^2},$$

  where S is a list of the observed species and T is a list of the time points at which observations are made. PTEMPEST allows other likelihood models, such as the built-in t-distribution [77], or any user-supplied function.

- *Prior distributions* on the parameters to be estimated. Uniform priors are a common choice when little is known about the parameters except for upper and lower limits. Priors can also be introduced to simplify a model by reducing some of its parameters to zero, a process called regularization. For example Lasso regularization [78] penalizes the sum of absolute values of parameters (the L1 norm), and Ridge regression [79] penalizes the sum of the squared parameter values (the L2-norm).

- A *proposal function* to define the probability distribution for the next parameter set to sample given the current set. A common choice is a normal distribution centered at the current value with a user-specified variance, which determines the effective step size. PTEMPEST uses a single adaptive step-size to determine the change in all parameters, but there are other MCMC implementations which permit different step sizes to govern changes in different directions in parameter space [31].

Following Metropolis *et al.* [80], we define the energy of a parameter set θ as

$$E(\theta) = -\log L(\theta) - \log p(\theta),$$

where L and p are the likelihood and prior distribution functions defined above.

2.2.1.1 Metropolis-Hastings algorithm The Metropolis-Hastings (MH) algorithm is one of the most popular MCMC methods [81]. If we assume a symmetric proposal function, i.e., the probability of moving from a parameter set θ_i to θ_j equals that of moving from θ_j to θ_i, then the algorithm to sample from $p(\theta|Y)$ is as follows:

1. Select an initial parameter vector θ_0 that has energy $E(\theta_0)$ and set $i = 0$.
2. For each step i until $i = N$
    a. Propose a new parameter vector θ_{new} and calculate the $E(\theta_{\text{new}})$.
    b. Set $\theta_i = \theta_{\text{new}}$ with probability $\min(1, e^{-\Delta E})$, where $\Delta E = E(\theta_{\text{new}}) - E(\theta_{\text{i-1}})$ (acceptance). Otherwise, set $\theta_i = \theta_{i-1}$ (rejection).
    c. Increment i by 1.

2.2.1.2 Parallel Tempering One of the key differences between MH and PT is the existence of a temperature parameter, β, that scales the effective "shallowness" of the energy landscape. Several Markov chains are constructed in parallel, each with a different β. A Markov chain with a β value of 1 samples the true energy landscape, while higher temperature chains have lower values of β and sample shallower landscapes with the acceptance probability now given by $\min(1, e^{-\beta \Delta E})$. Higher temperature chains accept unfavorable moves with a higher probability and therefore sample parameter space more broadly. Tempering refers to periodic attempts to swap 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 [30]. 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} MCMC 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 $\Delta E = E_j - E_{j-1}$, and $\Delta \beta = \beta_j - \beta_{j-1}$, and E_j and β_j are the energy and temperature parameter respectively of the j chain.

Adapting the step size and the temperature parameter can further increase the efficiency of sampling [30]. However, varying parameters during the construction of the chain violates the assumption of a symmetric proposal function (also referred to as "detailed balance"), and it is advisable to do this during a "burn-in" phase prior to sampling.

2.2.1.3 Implementation In this work we present PTempEst, a MATLAB-based tool for parameter estimation using PT that is integrated with the rule-based modeling software BioNetGen [71]. Models specified in the BioNetGen language (BNGL) can be exported as ODE models that are called as MATLAB functions by PTempEst. The BioNetGen commands `writeMfile` or `writeMexfile` are used to export models in MATLAB's M-file format, which uses MATLAB's built-in integrators, or as a MATLAB MEX-file, which encodes the model in C and invokes the CVODE library [82], which is usually much more efficient in our experience. For additional compatibility, models can be imported into BioNetGen in the System Biology Markup Language (SBML) [83], or the user can write their own cost function in MATLAB. The Bayesian parameter estimation capabilities of PTempEst complement those of another tool for performing parameter estimation on rule-based models, BioNetFit [84].

PTempEst uses adaptive step sizes and temperatures. The user provides the following hyper-parameters to control sampling: initial step size, initial temperature, and adaptation intervals and target acceptance probabilities for steps and swaps. At given intervals, the step acceptance probabilities and swap acceptance probabilities are calculated, and the step sizes and chain temperatures are adjusted to bring the step and swap acceptance probabilities closer to their target values respectively. For example, if the step acceptance rates are too high, the step size will be increased and vice versa. Similarly, if the swap acceptance rates are too high, the chain temperatures will be increased and vice versa. Although there are a considerable number of hyper-parameters associated with this method, we have found that the default values provided in PTempEst generally work well in practice.

The MATLAB source code for PTempEst along with model and data files used in the experiments described below are available at `http://github.com/RuleWorld/ptempest`.

2.2.2 Approximate Bayesian Computation methods

2.2.2.1 ABC rejection The simplest ABC algorithm is a rejection algorithm [85], which involves repeatedly sampling a parameter vector θ_i from the prior distribution, simulating the model with the sampled parameters, and calculating the discrepancy (often in the form of a distance function) between the simulated data Y_{sim} and the experimental data Y_{expt}. If the discrepancy is below a threshold, ϵ, θ_i is accepted as a member of the posterior distribution; otherwise, it is discarded and another θ_i is drawn. This process continues until the number of samples reaches a specified number, resulting in an approximation of the distribution $p(\theta|Y_{\text{sim}} - Y_{\text{expt}} < \epsilon)$, which in the limit of $\epsilon \to 0$, will approach the true posterior distribution $p(\theta|Y_{\text{expt}})$.

2.2.2.2 Approximate Bayesian Computation-Sequential Monte Carlo (ABC-SMC) The ABC rejection algorithm can suffer from low acceptance rates [85]. The ABC-SMC algorithm uses a tolerance schedule to decrease ϵ, and sequentially constructs approximate posterior distributions of increasing accuracy, which eventually converge to the true posterior distribution [20, 85]. We use ABC-SMC to generate the results shown in Sec. 2.3.2

2.2.3 Metrics for algorithm performance comparisons

In our analyses we fit ODE models to synthetic data generated using fixed parameter values. For the comparison to ABC presented in Sec. 2.3.2 we used synthetic data with additional noise, as was provided in the ABC-SysBio example files.

For models containing 3–6 parameters, both the MH and PT algorithms find the global minimum, and we compared the performance using convergence time and sampling efficiency. The *convergence time* is defined as the number of MCMC steps before the energy drops below a specified threshold, determined empirically [86]. For PT convergence time is based on the number of MCMC steps in the lowest temperature chain. With uniform priors and data simulated without noise, the negative log likelihood approaches zero when the chain converges to the global minimum.

The *sampling efficiency* is defined as the ratio of the range of the posterior distribution

to the range of the prior distribution, either for a model parameter that is known to be uniformly distributed, or for an added control parameter that does not contribute to the model output and therefore should be uniformly distributed.

For more complex models (11-25 parameters), we do not always obtain parameter sets that fit the data. In this case we compare the algorithms in terms of the negative log likelihood of the best fit parameter sets. In the case of uniform priors, this directly corresponds to the minimum energy attained by the Markov chain.

To compare disparate algorithms in terms of the total amount of computational resource used, we allowed each to perform a specified number of model integrations. For MH the number of model integrations is the number of MCMC steps, while for PT it is the number of MCMC steps times the number of chains run in parallel. For ABC algorithms, which use rejection sampling, the number of model integrations equals the total number of parameter sets evaluated to generate the desired number of samples.

2.3   Results

2.3.1   Michaelis-Menten Kinetics

We start with a simple model to demonstrate how Bayesian methods can identify constrained parameter relationships even when individual parameters are unidentifiable. The Michaelis-Menten model describes enzyme substrate kinetics using the following scheme:

$$E + S \underset{k_\mathrm{r}}{\overset{k_\mathrm{f}}{\rightleftharpoons}} ES \overset{k_\mathrm{cat}}{\rightarrow} E + P$$
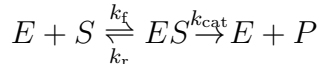
When the total enzyme concentration, $[E]_T$ is much smaller than that of the substrate, the rate of product formation is given by

$$\frac{d[P]}{dt} = \frac{k_\mathrm{cat}[E]_\mathrm{T}[S]}{(K_\mathrm{M} + [S])},$$

where $K_\mathrm{M} = (k_\mathrm{cat} + k_\mathrm{r})/k_\mathrm{f}$ is the Michaelis constant. The product trajectory only constrains k_cat and K_M, while the individual forward and backward rates k_f and k_r are unidentifiable. We generated a synthetic product trajectory using parameters $k_\mathrm{f} = 10^{-2.77}, k_\mathrm{r} = 10^{-1}, k_\mathrm{cat} =$

10^{-2}, and constructed a likelihood function assuming 1% Gaussian error. The 3 model parameters are sampled in log-space, with uniform priors on the intervals $[-3, 1]$, $[-1, 3]$ and $[-3, 3]$ for k_f, k_r and k_{cat} respectively. The fit is repeated 100 times using MH, and PT with 4 chains, starting from an initial parameter set of $[-1, 1, 0]$, corresponding to the midpoints of the priors. Both algorithms were run for 250,000 MCMC steps.

The quality of fit produced by MH and PT is comparable (Figure 2.1A-B). However, on average MH required 4203 MCMC steps to reach convergence, while PT required 369 (Figure 2.1C). Thus, even though each PT step needs 4 times as many model integrations, the total number of model integrations is smaller than for MH. This is consistent with the observation made in [30], that PT with M chains of length N can be more efficient than a single-chain Monte Carlo search of length MN. PT also has higher sampling efficiency for k_r and k_f compared to MH (Figure 2.1D).

As we would expect from the non-identifiability of k_f and k_r, the posterior distributions of $\log_{10}(k_r)$ and $\log_{10}(k_f)$ are uniform across the prior (Figure 2.1E), but their ratio is constrained (Figure 2.1F). $\log_{10}(k_{cat})$ is an identifiable parameter and has a constrained distribution centered at -2 (Figure 2.1E). The distributions shown in Figures 2.1E,F were obtained using PT with 4 chains run for 1,000,000 MCMC steps.

2.3.2 mRNA self-regulation

In this section we compare the efficiency of ABC-SMC, PT and MH for parameter estimation on a simple model of mRNA self-regulation (Figure 2.2A). The ABC-SysBio software is distributed with example files to estimate the parameters of this model assuming uniform priors using the ABC-SMC algorithm. The model has 5 parameters, one of which is fixed [20]. The quality of fit is defined as the Euclidean distance between the fitted trajectory and the data. For the ABC-SMC algorithm we extended the default 18-step tolerance schedule provided in ABC-SysBio from 50-15 to a 23-step schedule from 50-5 and set the ensemble size as 100. We ran ABC-SMC 50 times, and found that each run used an average of 6.7×10^4 model integrations. We then ran 50 repeats of 4-chain PT for 16750 MCMC steps, and of MH for 6.7×10^4 MCMC steps, using a likelihood function with 1% Gaussian error. The
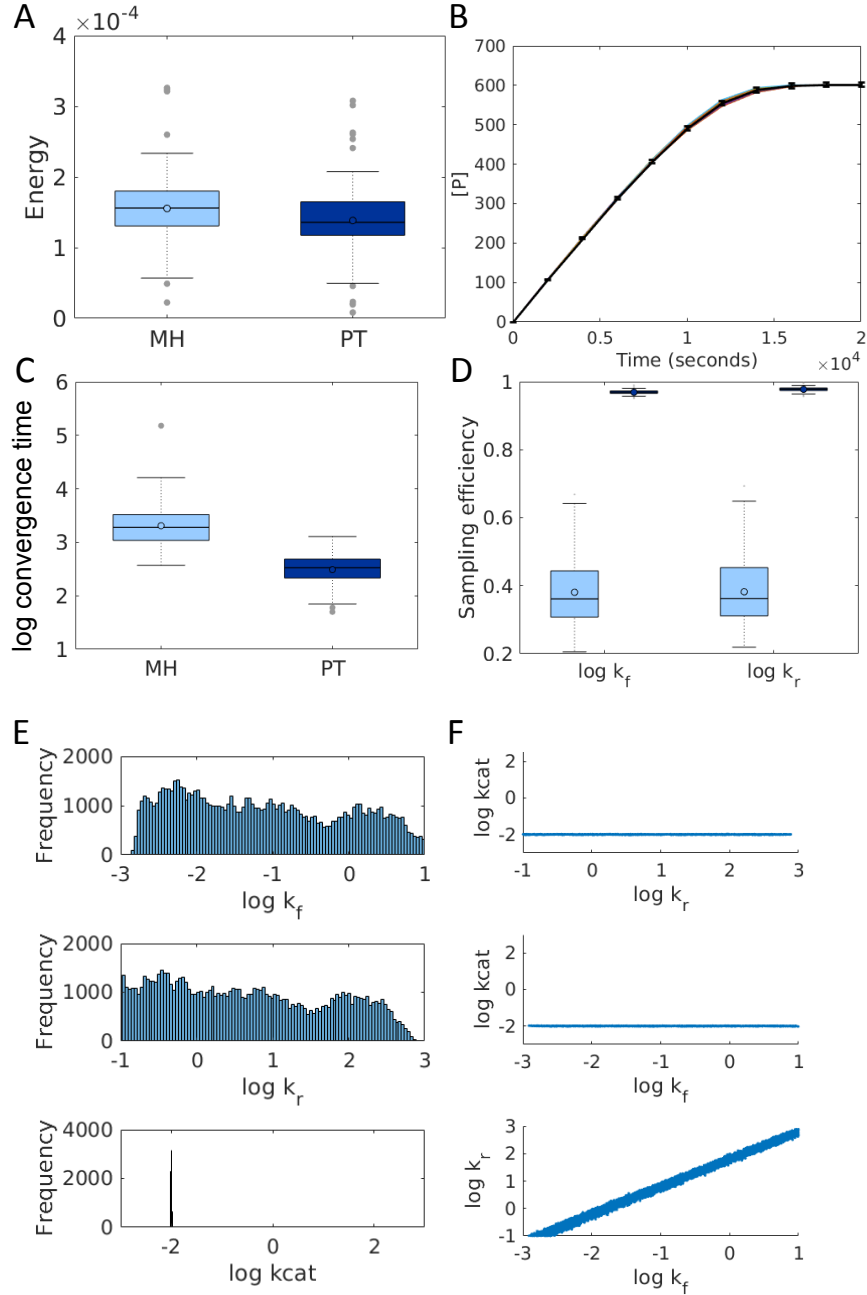
Figure 2.1: Parameter estimation for the Michaelis-Menten model. (A) Distribution of minimum energy values obtained via MH and PT. (B) Example of a fitted ensemble (colored lines) obtained for the synthetic data (black lines with error bars) using PT. (C) Distribution of convergence times for PT vs. MH with an energy threshold of 1. (D) Sampling efficiency for parameters k_r and k_f over 100 repeats using MH (light blue) and PT (dark blue). (E) Estimated posterior distributions for each of the model parameters. The x-axis limits are the uniform prior boundaries. (F) Scatter plots of sampled parameter sets for each pair of model parameters. Axis limits reflect prior boundaries.

16

sampling efficiency of PT and MH was compared using a control parameter as described above. The quality of fit produced by the MCMC-based algorithms is substantially higher than what we get from ABC-SMC (Figure 2.2B-D). PT takes fewer steps to reach convergence (Figure 2.2E), and has higher sampling efficiency than MH given the same number of model integrations (Figure 2.2F).

2.3.3 Model reduction with Lasso

In this section we demonstrate the use of MCMC approaches to perform model reduction, by coupling parameter estimation with regularization. Lasso regularization penalizes the L1-norm of the parameter vector while minimizing the cost function during parameter estimation. This performs variable selection by finding the minimum number of non-zero parameters required to fit the data [39]. The Lasso penalty is equivalent to assuming a Laplace prior on the parameters [78], and the width of the prior is inversely related to the regularization parameter that governs the strength of the penalty. Here, we present an example of using the Bayesian Lasso for model reduction, and compare the use of PT and MH for this problem.

A core model of negative feedback regulation with three processes (blue arrows in Figure 2.3A) was simulated to get a synthetic trajectory for species A using a value of 10 for all three rate constants (Figure 2.3C). Three extraneous process were added to the model (red arrows in Figure 2.3A), so that only a subset of the reactions in the reaction scheme are required to fit the data. We constructed a likelihood function assuming 2% Gaussian error, and assumed Laplace priors of width b on each of the 6 model parameters, where b is the regularization parameter that needs to be tuned. High values of b, i.e., wide priors, will not impact the log likelihood but will not achieve much variable selection. Conversely for low values of b most of the parameters will go to 0 at the cost of degrading the log likelihood.

Here, we tested a range of b values. For each we ran PT with 500,000 MCMC steps 50 times to obtain a distribution of negative log-likelihood values (Figure 2.3B). Figure 2.3C shows examples of fitted ensembles obtained with different regularization strengths. We chose the smallest value of b, 0.5, that does not significantly increase the negative log likelihood

17
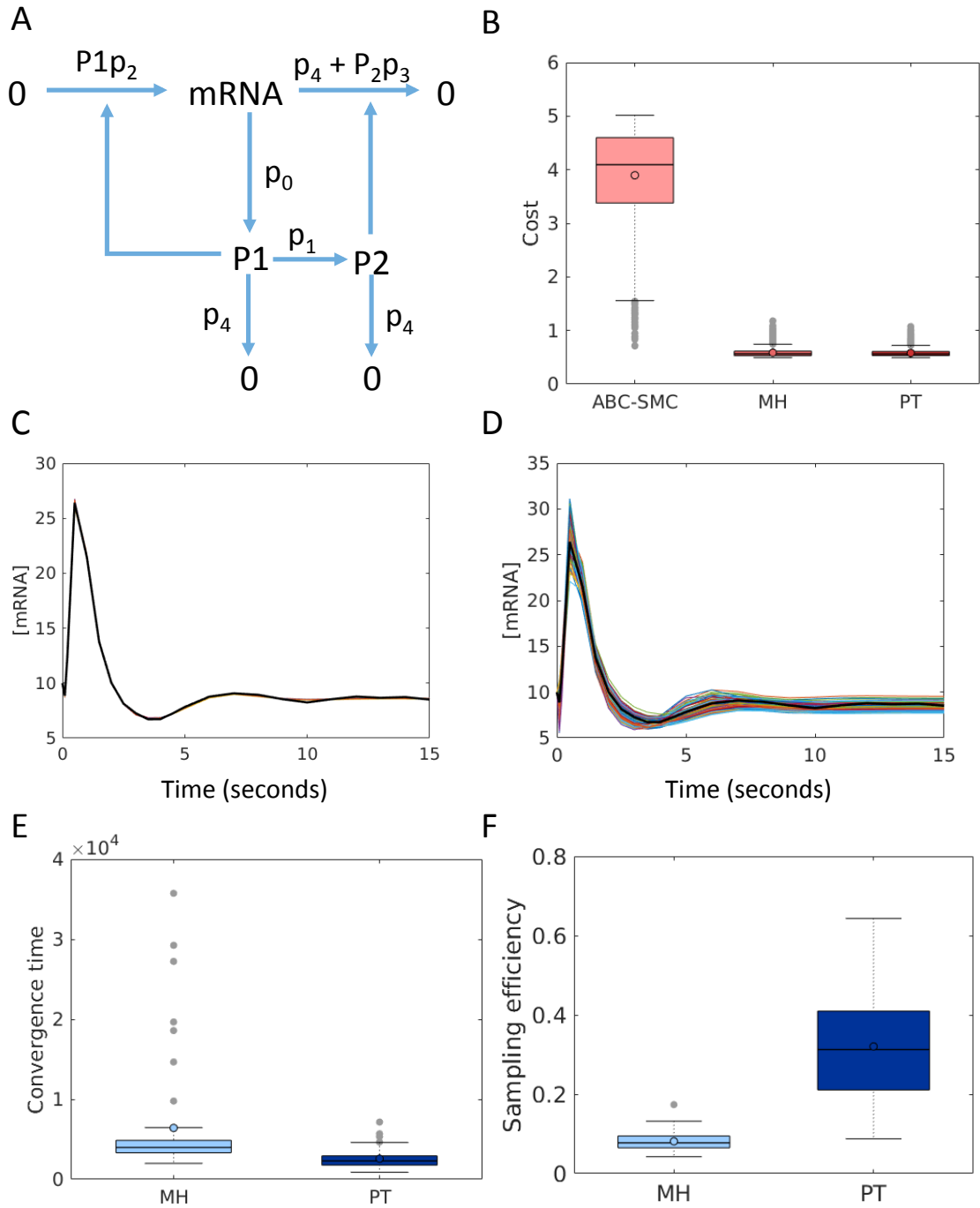
Figure 2.2: Parameter estimation for the model of mRNA self regulation. (A) Reaction network diagram of the mRNA self regulation model from [20] (B) Quality of fit of the final ensemble obtained from ABC-SysBio, PT and MH. The box plots show the distribution of the Euclidean distances of the 100 members of each of 50 fitted ensembles from the synthetic data. Example of a typical fitted ensemble obtained from (C) PT and (D) ABC-SysBio. Black lines show the synthetic data, and the colored lines show the fitted trajectories. [mRNA] refers to the number of mRNA molecules. (E) Distribution of convergence times for PT vs. MH with an energy threshold of 20. PT takes on average ∼ 2-fold fewer steps to reach convergence. (F) Comparison of sampling efficiency of MH vs. PT.
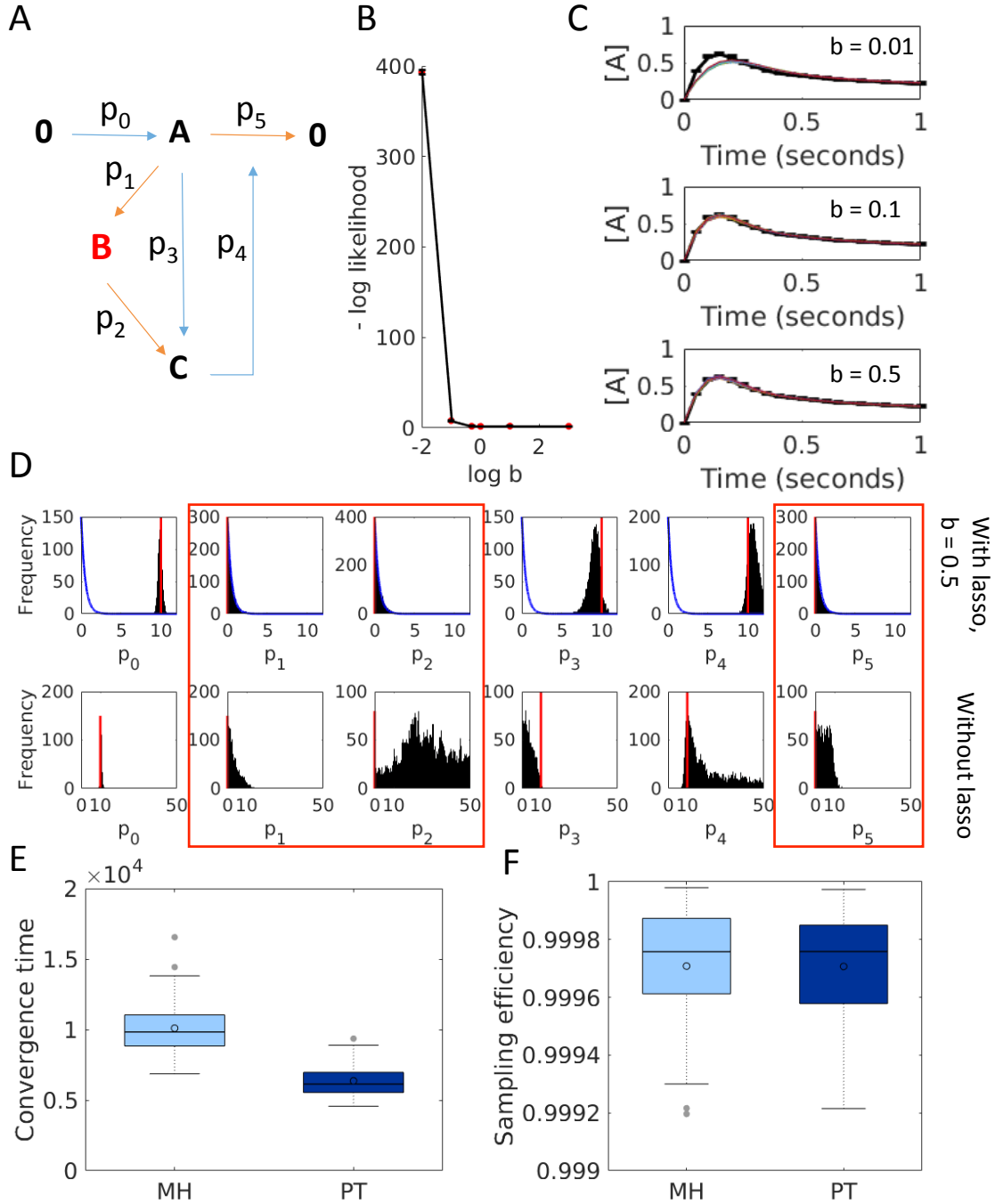
Figure 2.3: Model reduction with Lasso. (A) Reaction network diagram of a toy negative feedback model. The core model used to obtain the synthetic data for the fit is in blue, and extraneous elements are in red (B) Tuning the regularization parameter, i.e. the width of the Laplace prior, w.r.t the negative log likelihood of the fitted ensembles (C) Examples of fitted ensembles corresponding to different regularization strengths. Error bars show synthetic data. Solid lines show the simulated fits. (D) Posterior distributions with lasso (top row) for parameters show extraneous parameters peaking at 0. Red lines indicate true parameter values, and the blue lines show the Laplace prior ($b = 0.5$). The bottom row shows the posterior distributions obtained without Lasso. Red boxes indicate extraneous parameters. (E) Distribution of convergence times for PT vs. MH with an energy threshold of 65. (F) Distributions of of sampling efficiency for PT and MH across 50 repeats.

(Figure 2.3B), and used this for further analysis.

The posterior distributions for the model parameters obtained with regularization show the extraneous parameters peaking at 0, while the essential parameters have well defined distributions that peak close to their true values (Figure 2.3D, top row). Without regularization the extraneous parameters p_1, p_2 and p_5 (red boxes in Figure 2.3D) take on non-zero values and make the other parameters unidentifiable (Figure 2.3D, bottom row). PT converges faster than MH (Figure 2.3E), but the sampling efficiencies calculated over 200,000 MCMC steps are comparable (Figure 2.3F).

2.3.4   Calcium signaling

The models considered so far have a relatively small number of parameters and both MH and PT achieve convergence readily. For models with more parameters and more complex dynamics, convergence becomes difficult to achieve. As an example, we consider a four-species model of calcium oscillations that has 12 free parameters [87]. The model describes the dynamics of G_α subunits of the G-protein, active PLC, free cytosolic calcium, and calcium in the endoplasmic reticulum. We generated synthetic data for free cytosolic calcium (Figure 2.4A), and constructed a likelihood function assuming 20% Gaussian error. The 12 free parameters were sampled in log-space with uniform priors, 6 units wide and centered at the true values. We generated 100 random initial parameter sets, and from each starting point sampled using MH, PT with 4 chains (PT-4) and PT with 6 chains (PT-6). Only a fraction of the chains converged to the global minimum in 500,000 MCMC steps. Figure 2.4A shows an example of an MH chain that has converged to a local minimum with high energy, and another of a PT chain that has converged to the global minimum. The distributions of minimum energy for chains obtained from each algorithm (Figure 2.4B) show that PT-6 found better fits than PT-4, which in turn did better than MH, which returned highly variable results and frequently did not reach the global minimum.

2.3.5 Negative feedback oscillator

We also considered the three species negative feedback oscillator from Tyson et al. [88], to evaluate the more difficult case of fitting a model to complex dynamics of multiple species. We generated synthetic data for all three model species under conditions where all three species undergo sustained oscillations. 11 model parameters are sampled in log-space, with uniform priors that are 10 units wide and centered at the true values. The likelihood function is a t-distribution with 10% error. We generated 15 random initial parameter sets, and from each starting set ran MH, PT with 4 chains (PT-4) and PT with 6 chains (PT-6) for 500,000 MCMC steps.

Figure 2.5A shows examples of chains converging to different minima. The top row shows an example of convergence to a high energy. As in the case of calcium signaling, PT with 6 chains outperforms PT with 4 chains, which in turn outperforms the MH algorithm in finding the global minimum (Figure 2.5B). Interestingly, the data that we generated did not sufficiently constrain the frequency of oscillations exhibited by the model, and we find parameter sets corresponding to different frequencies that all fit the data. Figure 2.5C shows the posterior distributions of the 11 model parameters corresponding to the fit shown in the middle panel of Figure 2.5A, obtained using PT with 6 chains. The first parameter shows 3 clear peaks, one of which is centered at the true value. Separating the parameter sets corresponding to these peaks shows that they correspond to specific differences in oscillation frequencies that are all part of the fitted ensemble (Figure 2.5D), reinforcing the need to use Bayesian methods with such problems.
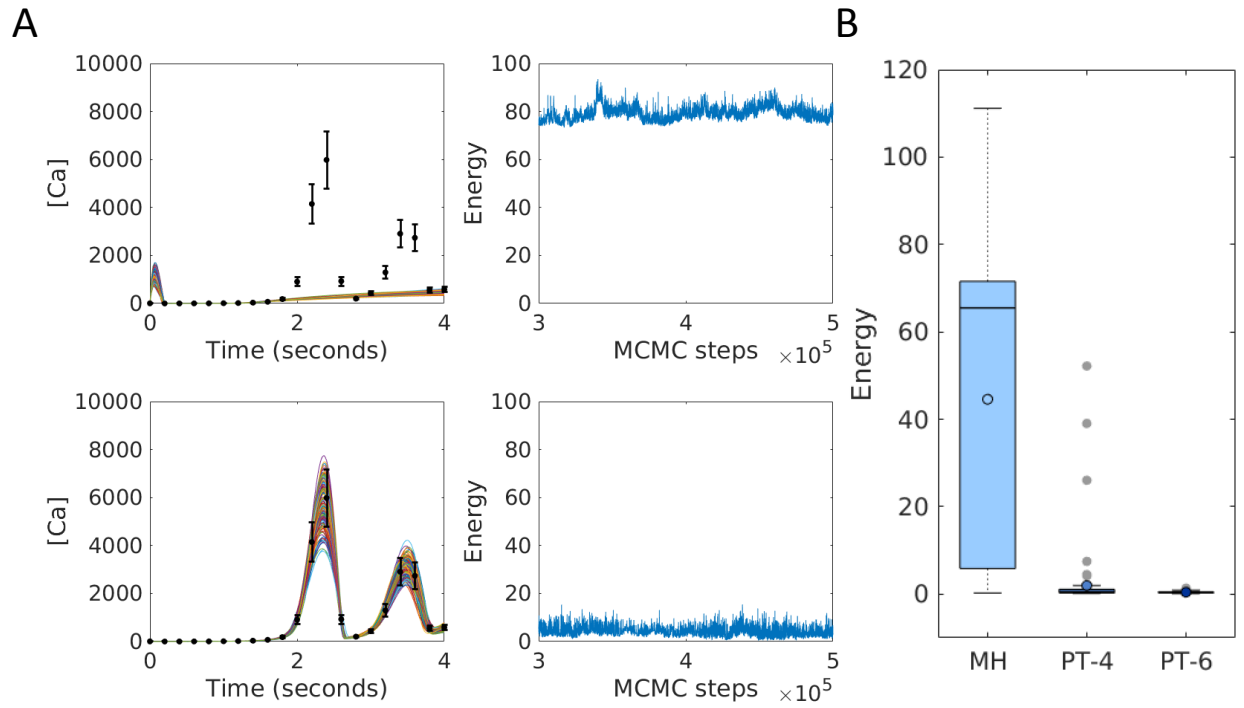
Figure 2.4: Parameter estimation for the model of calcium signaling. (A) Examples of convergence to a local minimum (top) and to the global minimum (bottom). Error bars show synthetic data. Solid lines show the simulated fits. The right column shows the energy chains corresponding to the fits on the left. (B) Distributions of the minimum energy from MH, PT-4 and PT-6 over 100 repeats.
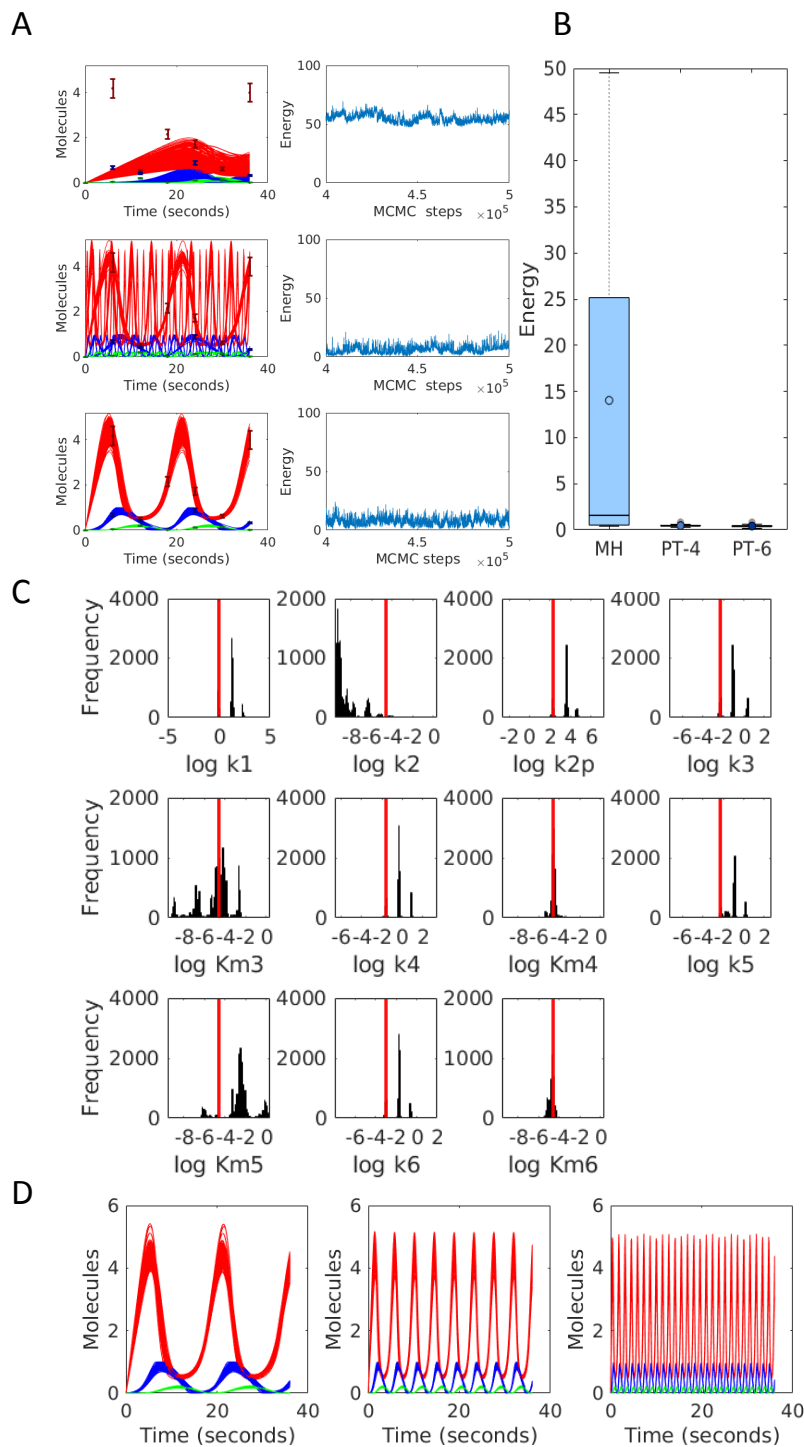
Figure 2.5: Parameter estimation for the negative feedback oscillator. (A) Examples of convergence to different minima. Error bars show synthetic data. Solid lines show the simulated fits. The three colors correspond to the three different model species. The right column shows energy chains corresponding to the fits shown on the left. (B) Distributions of the minimum energy by MH, PT-4 and PT-6 over 15 repeats. (C) Posterior distributions corresponding to the middle panel in (A). (D) Simulated fits corresponding to each of the three peaks in the posterior distribution of the first parameter shown in (C).

2.3.6 Growth factor signaling model

Finally we apply MH and PT to a substantially larger model that has 24 parameters — a rule-based model of Shp2 regulation in growth factor signaling [89] that generates 149 species and 1032 reactions. We generated synthetic data for the micromolar concentration of phosphorylated receptors (pYR), an observable that combines the time courses of 136 model species, and constructed a likelihood function assuming 2% Gaussian error. The parameters were sampled in log-space with a uniform prior on the interval $[-6, 6]$. We generated 25 random initial parameter sets, and from each starting point we obtain Markov chains with 200,000 MCMC steps using MH and PT with four chains. Figure 2.6A shows chain convergence to different minima, and Figure 2.6B shows that PT more consistently finds good fits than MH.
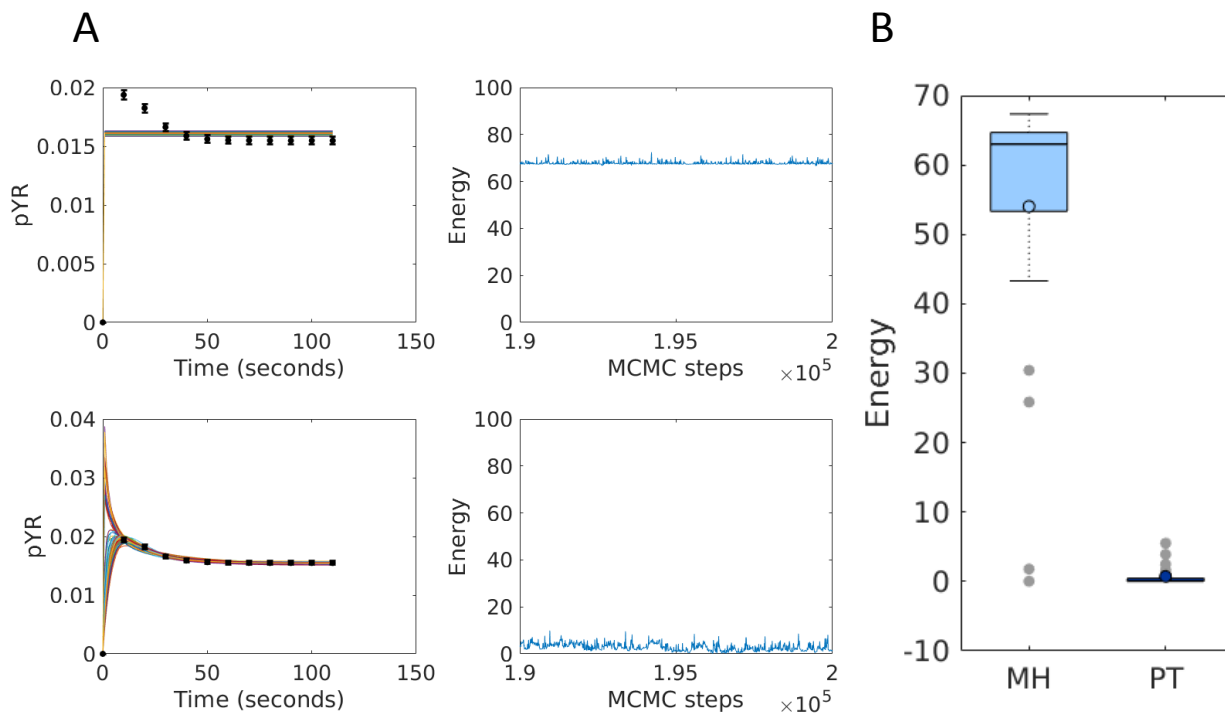
Figure 2.6: Parameter estimation for the growth factor signaling model. (A) Examples of convergence to different minima. Error bars show synthetic data. Solid lines show the simulated fits. The right column shows energy chains corresponding to the fits shown on the left. (B) Distributions of the minimum energy obtained via by MH and PT with 4 chains over 25 repeats.

2.4   Discussion

In this study we have shown that even with relatively simple biochemical models, there are significant benefits to using PT over MH in terms of convergence speed and sampling efficiency. With more complex models we found that given a fixed budget of MCMC steps MH often fails to find the global minimum, whereas PT consistently succeeds. We also showed an example in which Bayesian parameter estimation can effectively perform model reduction through the introduction of a regularizing prior. While ABC methods constitute a popular class of alternative methods for Bayesian parameter estimation in cases where likelihood models are expensive or not available (such as for stochastic models), we found that PT outperformed ABC for parameter sampling on a relatively simple ODE model. Our direct performance comparison supports the previous observation [20] that likelihood-based methods are preferable to likelihood-free methods when likelihood models are feasible to compute, such as with ODE models.

One limitation of our evaluation procedure is that we have not attempted to compare wall clock times for the different algorithms. Instead, as a performance metric we have used the number of MCMC steps or the number of model integrations required, which are independent of the implementation. In practice, the fits reported in Sec. 2.3 can all be performed on a typical workstation computer in times ranging from a few minutes for the smallest model (Michaelis-Menten) to a few hours for the largest model (growth factor signaling). However, we found that despite requiring the same number of model integrations per processor per step, the single-chain MH sampling ran significantly faster per step (20–30%) in terms of wall clock time than PT with four or six chains. Preliminary tests showed that these differences likely arise from the requirement in our current implementation for each chain to complete a fixed number of steps before a swap is attempted. Parallel efficiency decreases when trajectories on different processors take different amounts of time to complete. We found that the difference in wall clock time decreased when the PT chains are all run at the same temperature, but so do the algorithmic benefits. When chains are run at different temperatures, the high temperature chains tend to sample parameter space more broadly, which results in greater variability in the model integration time [90] and causes slow down

due to the synchronization requirement. We plan to investigate asynchronous swapping between chains in order to alleviate this problem.

Another limitation of the current work is that the comparisons were made using specific choices for the hyper-parameters that control the PT algorithm, such as those that control step sizes for the moves and temperatures. Adjustment of these may result in further improvements to sampling efficiency and convergence rates. We would also like to investigate the effect of using different proposal functions, such as Hessian-guided MCMC [31], as well as different likelihood models.

While we have restricted our MCMC comparisons to MH, there has been considerable work toward improving the efficiency of MCMC methods, such as Differential Evolution Adaptive Metropolis (DREAM) [91] and Delayed Rejection Adaptive Metropolis (DRAM) [92]. It would be interesting to investigate whether parallel tempering could be fruitfully combined with these approaches.

Finally, PT, as it has been presented and used to this point both here and in the molecular simulation literature, is only a *moderately* parallel algorithm because it uses just a handful of chains. It remains to be seen whether using a much larger number of chains would retain the advantages of sampling simultaneously at multiple temperatures and result in further acceleration.

2.5   Author contributions

Conceptualization, J.R.F., S.G., J.H., L.H., R.E.C.L; Methodology, J.R.F., S.G., J.H; Investigation, S.G., L.H.; Software, S.G. wrote the code for the computational experiments and performance evaluation, J.H developed *ptempest*; Visualization, S.G.; Writing – Original Draft, S.G.; Writing – Review and Editing, S.G., R.E.C.L., and J.R.F.; Funding Acquisition, R.E.C.L. and J.R.F.; Supervision, R.E.C.L. and J.R.F..

3.0 Parallel Tempering with Lasso for Model Reduction in Systems Biology

This chapter is taken from Gupta S, Lee REC, Faeder JR, Parallel tempering with Lasso for model reduction in systems biology, PLOS. Computational Biology, 2020 [12].

3.1 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 [15, 16, 20, 93]. As our knowledge of the biochemical processes in a cell increases, reaction network models of cell signaling have been growing more detailed [16, 17, 72]. 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 [33, 34, 88]. 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 minimal subsets of non-zero parameters that fit the data.

A number of previous studies have addressed model reduction for biochemical systems, as reviewed in [94]. Some examples include reduction by topological modifications to resolve non-identifiability in models [95, 96] and reduction by timescale partitioning [97–99]. Non-identifiability arises when multiple unique parameterizations of a model give the same model output. Quaiser et al. [95] and Maiwald et al. [96] 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 [94,98]. Both of these methods generate reduced models but do not carry out parameter estimation to fit experimental data. Gabel et al. [100] recently developed FaMoS (Flexible and dynamic Algorithm for Model Selection), a method that uses heuristic search algorithms to search the space of submodel topologies within a larger model. However, each proposed submodel has to be individually refit to experimental data, and the heuristic search algorithms used are not guaranteed to return all possible submodel topologies that fit the data. Maurya et al. [101] used mixed-integer nonlinear optimization to combine parameter estimation with model reduction by reaction elimination, a technique common in the field of chemical engineering [102, 103]. This approach requires an additional binary parameter for every reaction in the model, and 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 can be used to characterize high-dimensional, rugged, multimodal parameter landscapes common to systems biology models [20, 22, 29, 31, 32] but 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 [30], outperforms conventional MCMC for systems biology models with up to dozens of parameters [32]. 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 [39]. In the context of biology, Lasso has been widely applied to gene expression and genomic data, typically in combination with standard regression techniques [104–108] and less commonly in Bayesian frameworks [109, 110]. In the mechanistic modeling context, Lasso regression has been used to predict cell type specific parameters in ODE reaction network models [111], 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 can extract multiple minimal subsets of reactions if present, which provides alternate mechanisms to explain the data. We use synthetic data to demonstrate that PTLasso is an effective approach for model reduction. We also use PT with Lasso on groups of parameters (grouped Lasso) with real biological data in a larger model of NF-κB signaling to select over reaction-network modules instead of individual reactions. Grouped 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 chapter is organized as follows. In Methods we provide an overview of the PTLasso approach with in-depth descriptions of PT, regularization with Lasso and grouped Lasso, and the setup of computational experiments. In Results we demonstrate PTLasso on synthetic examples of increasing complexity followed by an application of the grouped Lasso approach to address mechanistic questions in NF-κB 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.

3.2   Methods

In this work we use Bayesian parameter estimation (BPE) for model reduction. Here, we present an overview of BPE using the MH and PT algorithms for MCMC sampling. Our presentation of these algorithms is modified from Gupta et al. [32]. Following this, we describe the application of regularization in the context of MCMC sampling using either Lasso or grouped Lasso. Finally, we describe the basic steps of the computational experiments, including generation of the synthetic data for fitting, choosing the starting parameter

configurations for the MCMC chains, convergence testing and selection of hyperparameters.

Following [32], 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

$$p(\vec{\theta}|Y) \propto p(Y|\vec{\theta})p(\vec{\theta}). \tag{3.1}$$

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

$$L(\vec{\theta}) = e^{-\Sigma_S \Sigma_T (Y_{\text{sim}}(\vec{\theta}) - Y_{\text{expt}})^2 / 2\sigma^2}, \tag{3.2}$$

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 and can be different for different species and time points, $Y_{\text{sim}}(\vec{\theta})$ is the model output for parameter vector $\vec{\theta}$, and Y_{expt} is the corresponding experimental data. $p(Y|\vec{\theta})$ is equal to the normalized $L(\vec{\theta})$. $p(\vec{\theta})$ 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.

3.2.1 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.* [80], we define the energy of a parameter vector $\vec{\theta}$ as

$$E(\vec{\theta}) = -\log L(\vec{\theta}) - \log p(\vec{\theta}), \tag{3.3}$$

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.

3.2.1.1 Metropolis-Hastings algorithm The Metropolis-Hastings (MH) algorithm is a commonly-used MCMC algorithm for BPE [81]. At each step, n, the method uses a proposal function to generate a new parameter vector, $\vec{\theta}_{\mathrm{new}}$, given the current parameter vector, $\vec{\theta}_n$. A common choice of proposal function is a normal distribution centered at $\vec{\theta}_n$:

$$f(\vec{\theta}_{\mathrm{new}}, \vec{\theta}_n) \propto e^{-\alpha|\vec{\theta}_{\mathrm{new}} - \vec{\theta}_n|^2}. \tag{3.4}$$

For any f that is symmetric with respect to $\vec{\theta}_{\mathrm{new}}$ and $\vec{\theta}_n$, the move $\vec{\theta}_{n+1} = \vec{\theta}_{\mathrm{new}}$ 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 $\vec{\theta}_n$.

3.2.1.2 Parallel Tempering In PT (also referred to as replica exchange Monte Carlo [30]), 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 more 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 [30, 32]. Following [32], 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.

Note that in the Results we often refer to a parameter vector obtained from the lowest temperature chain at a particular swap as a sample. Ensemble fits are shown by subsampling parameter vectors from the lowest temperature chain.

Adapting the step size and the temperature parameter can further increase the efficiency of sampling [30], 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. Another way to increase efficiency of sampling for parameters that may be on different scales is to sample in log-space as we have done for all of the examples in this manuscript.

3.2.2 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 [39]. In a Bayesian framework, the Lasso penalty is equivalent to assuming a Laplace prior on each parameter θ_i given by

$$p(\theta_i) = \frac{1}{2b} \exp\left(-\frac{|\theta_i - \mu|}{b}\right), \tag{3.5}$$

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

$$E(\vec{\theta}) = -\log L(\vec{\theta}) + \sum_{i=1}^{n_{\mathrm{par}}} \frac{|\theta_i - \mu|}{b}, \tag{3.6}$$

where n_{par} is the number of model parameters, and we have dropped the constant term arising from the normalization constant $\frac{1}{2b}$ in Eq. 3.5. From Eq. 3.6 it can be seen that the regularization strength is inversely proportional to b. For efficiency we 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.

3.2.3 Regularization with grouped Lasso

To account for modularity in complex signaling networks [112], we use grouped Lasso, to perform selection at the level of reaction modules instead of individual reactions (Note that this differs from the standard Group Lasso penalty [113] that is typically used for regression

problems). 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

$$\theta_i = k_i \lambda_m, \tag{3.7}$$

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

$$\theta'_i = k'_i + \lambda'_m. \tag{3.8}$$

The energy function is then

$$E(\vec{\theta'}) = -\log L(\vec{\theta'}) + \sum_{m=1}^{n_{\mathrm{mod}}} \frac{|\lambda'_m - \mu|}{b} + \sum_{i=1}^{n_{\mathrm{par}}} f(k'_i), \tag{3.9}$$

where

$$f(k'_i) = \begin{cases} 0, & \text{if } k'_i \in (LB_i, UB_i) \\ \infty, & \text{otherwise.} \end{cases} \tag{3.10}$$

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 $\theta'_i, \approx UB_i + \mu$, 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-κB signaling we chose $\mu = -25$, $LB_i = -5$ and $UB_i = 10$ for all i.

3.2.4 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 (labeled as "true 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.

3.2.5 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 [114]. 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-κB signaling model to enforce certain known properties of the NF-κB system, such as that the exit rate of NF-κB-IκB complex from the nucleus is greater than that of free NF-κB [56]. A full list of constraints applied to the NF-κB signaling model is listed in Table A1.

3.2.6 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 PT (or PTLasso) chains that were previously run with similar data or hyperparameter configurations were used to initialize the current PT (or PTLasso) chains. For example, parameter vectors obtained for one NF-κB trajectory could be used as a start point for fitting a different NF-κB 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 available at `https://github.com/RuleWorld/SupplementalMaterials/tree/master/Gupta2019`.

3.2.7 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 (Table A2). The PSRF compares intra-chain and inter-chain variances for model parameter distributions and serves as a measure of convergence [115]. In keeping with the literature, we consider a PSRF less than 1.2 [29, 31, 115] as consistent with convergence. We also calculated the stricter Multivariate PSRF (MP-SRF), which extends PSRF by checking for convergence of parameter covariation (Table A3). Third-party MATLAB libraries used for the PSRF and 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-κB signaling model, the number of PT (or PTLasso) swaps needed for convergence was large and time consuming to obtain in a single run. Instead of running two long PT (or PTLasso) chains each of length N, we picked two favorable initial conditions and from each ran a set of M PT (or PTLasso) 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 A4, A5). This gave us two groups of N PT (or PTLasso) samples that we used to test convergence of parameter distributions.

PSRF and MPSRF values for each computational experiment are shown in Tables A2 and A3 respectively. We also show in Table A6 that the step acceptance rates for most chains are close to the optimal value of 0.234 [116]. Table A7 shows the swap acceptance rates of the two lowest temperature chains for each computational experiment.

3.2.8 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 [32, 117]. We also checked that the results were insensitive to small variations in μ

(Figs. A1B, A2B, A4B,C).

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. 3.4E, A1C, A2C, A4D,E).

3.2.9   Software

All results reported in this work were obtained using `ptempest` [32], which is a MAT-LAB package for parameter estimation that implements PT with support for regularization. `ptempest` uses MATLAB's Mex interface to support the efficient integration of ODEs in C using the CVODE library and is directly compatible with the popular rule-based modeling software BioNetGen [71] which enables use with models built in both Systems Biology Markup Language (SBML) [118] and the BioNetGen Language (BNGL). The source code is available at `http://github.com/RuleWorld/ptempest`.

3.3   Results

3.3.1   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 (Table A8). The motif A→B→C (Fig. 3.1A, 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. 3.1B labeled "true data"). As described in Methods, Gaussian noise (mean = 0, standard deviation = 30% of the true data value) was added to generate ten noisy trajectories that were sampled at eight time points (Fig. A1A) to simulate the effects of experimental noise and cell-to-cell variability. The mean and standard deviation

of these synthetic trajectories formed the "observed data" (black points and error bars in Fig. 3.1B) 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. 3.1A, right). Time courses from PT and PTLasso (Fig. 3.1B) both fit the observed data (Fig. A1C), but the PTLasso fits are more similar to the true data at times before the first observed data point. PT finds parameter probability distributions (Fig. 3.1C, top row) that exhibit sharp peaks near the true values of the two nonzero parameters that were used to generate the data, 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. 3.1C, bottom row) recovers tight distributions near the true 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 (Table A9, Fig. 3.2A, Fig. A2A). 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. 3.2B). Similarly, rate constant parameter distributions with PT are all broad (Fig. 3.2C), whereas the extraneous parameters for the PTLasso fits were within the Laplace prior (Fig. 3.2D). In addition to a tight distribution near the true value 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→C, B→D and B→E depends on which of the other two are included. This is because the model A→B→C is indistinguishable from A→B→D and A→B→E without more information about the system. Even though the marginal posterior distributions show all three parameters playing a role, parameter covariation (Fig. A2D) reveals that only one of the reactions B→C, B→D, B→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 similar

clustering (Fig. A2E). Taken together, these results show that PTLasso correctly identifies network parameters and suggests that A→B→C, A→B→D, and A→B→E are alternate reduced models for the data.

The primary noise model that we have chosen for all the synthetic experiments in this chapter is Gaussian noise added to the data. However, one might ask how the results of PTLasso are affected when noise is added to the true parameter values instead, which is perhaps a more accurate representation of cell-to-cell variability in biological signaling systems. To test this, we perturbed the true parameters ($\log k_{AB} = -1$, $\log k_{BC} = 0$) 10 times with Gaussian noise (mean = 0, standard deviation = 0.05) (Fig. A3A). The mean and standard deviation of the resulting 10 noisy model outputs formed the observed data for fitting (Fig. A3B).The results of PTLasso were qualitatively the same when fitting fully connected three-node and five-node networks to this data (Fig. A3C–F).

Overall, these results show that PTLasso is a global approach that can extract correct parameter estimates and architectures of alternate 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.
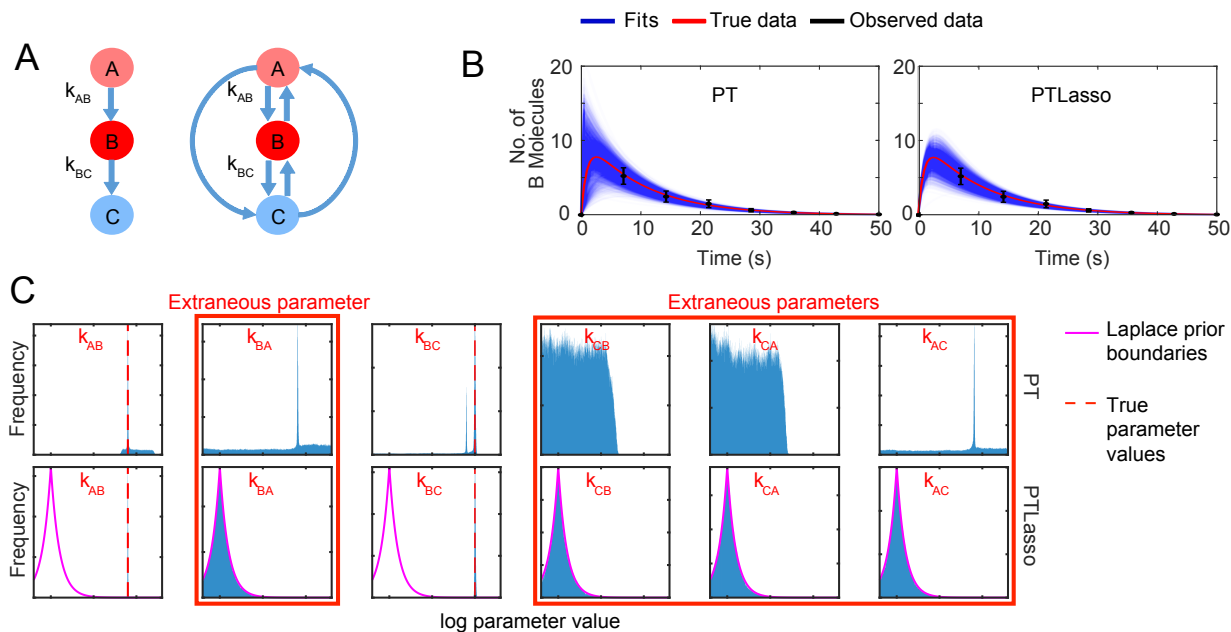
Figure 3.1: Model reduction using PTLasso with a fully connected 3-node graph. A) Motif used to generate the observed data (left) and reaction network diagram of the fully connected 3-node network used as the starting point for PTLasso (right). The initial concentration of A (light red) is 100 molecules. 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. $k_{AB} = 0.1\,\mathrm{s}^{-1}$ and $k_{BC} = 1\,\mathrm{s}^{-1}$ are the true parameter values (rate constants not specified were set to zero). B) Fits of the model to the data with PT (left) and PTLasso (right). Transparent blue lines show ensemble fits (from 4,000 parameter samples, 100 time points per trajectory), 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 400,000 parameter samples) for fits with PT (top row) and PTLasso (bottom row). The range of log parameter values on each x-axis is -12 to 3, which covers the full range over which parameters were allowed to vary. The y-axis of each panel is scaled to the maximum value of the corresponding distribution to emphasize differences in shape. The pink lines show the boundaries of the Laplace prior with $\mu = -10$, $b = 1$ and the dashed red lines in panels for k_{AB} and k_{BC} show the true parameter values. A parameter distribution confined within the Laplace prior boundaries indicates that the parameter is extraneous (panels with red border).
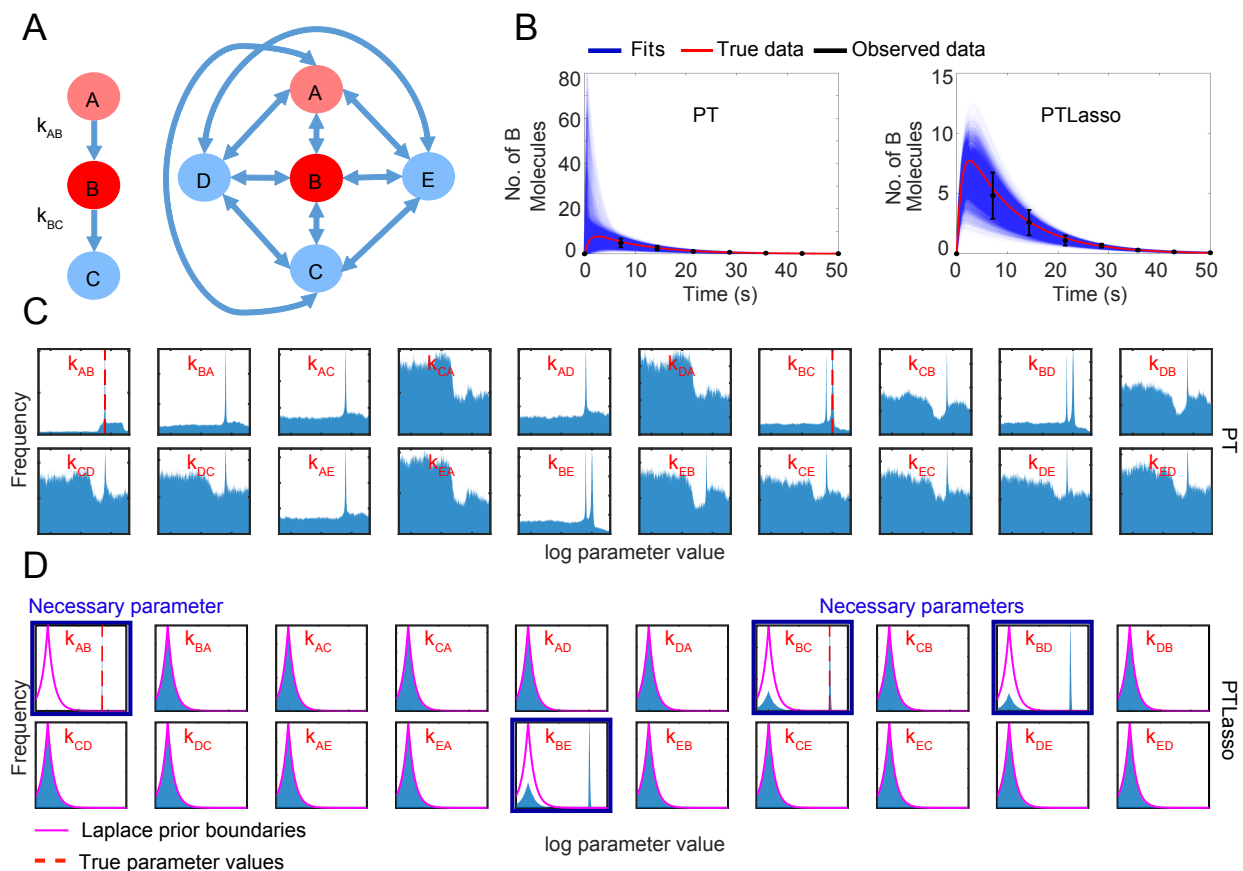
Figure 3.2: Model reduction using PTLasso with a fully connected 5-node graph. A) Motif used to generate the observed data (left) and reaction network diagram of the fully connected 5-node network used as the starting point for PTLasso (right). $k_{AB} = 0.1\,\mathrm{s}^{-1}$ and $k_{BC} = 1\,\mathrm{s}^{-1}$ are the true parameter values (rate constants not specified were set to zero). The initial concentration of A (light red) was set to 100 molecules, while the initial concentrations of B, C, D and E were set to 0. The concentration of B (red) is observed at multiple time points, but the concentrations of C, D and E (blue) are not observed. B) Fit of the model to the data with PT (left) and PTLasso (right). Transparent blue lines show ensemble fits (from 7,000 parameter samples, 100 time points per trajectory), 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 700,000 parameter samples) for fits with PT and D) PTLasso. The range of log parameter values on each x-axis is -12 to 3, which covers the full range over which parameters were allowed to vary. The y-axis of each panel is scaled to the maximum value of the corresponding distribution to emphasize differences in shape. The pink lines show the boundaries of the Laplace prior with $\mu = -10$, $b = 1$ and the dashed red lines in panels for k_{AB} and k_{BC} show the true parameter values. A parameter distribution that deviates from the prior is necessary (panels with blue border).

3.3.2 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 subnetworks required to produce specific dose-response relationships.

Tyson et al. [88] 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 six-parameter model (Table A10, Fig. 3.3A). We show that PTLasso correctly identifies the linear and adaptive submodels when the combined model is fit to different simulated data. The linear dose-response submodel was used to generate synthetic time courses for R in response to increasing levels of S (Fig. 3.3B, top row). As earlier, Gaussian noise (mean $= 0$, standard deviation $= 10\%$ of the true data value) 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 four distinct time points (including $t = 0$), creating 16 data points that constitute the observed data. As in the previous example of fully connected networks, PTLasso fits of the prior network to the observed data are more similar to the true data than fits from PT alone (Fig. 3.3B, top row). PTLasso recovers tight distributions for k_{s-rs} and k_{r-0}, which are the only model parameter values that lie outside the Laplace prior, providing a reduced two-parameter model that is sufficient to produce the synthetic data (Fig. 3.3C,D).

When the perfectly adaptive dose-response submodel was similarly used to generate observed data in response to two successive increasing values of S, PTLasso reduced the prior network to a four-parameter model (Fig. 3.3E,F) that fits the data (Fig. 3.3B, bottom row). In this case, parameters k_{s-xs} and k_{xr-x} in the reduced model have broad distributions and are unidentifiable (Fig. 3.3E, bottom row), but PTLasso captures their linear correlation (Fig. A4A), which may provide further avenues for model reduction [96]. 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.
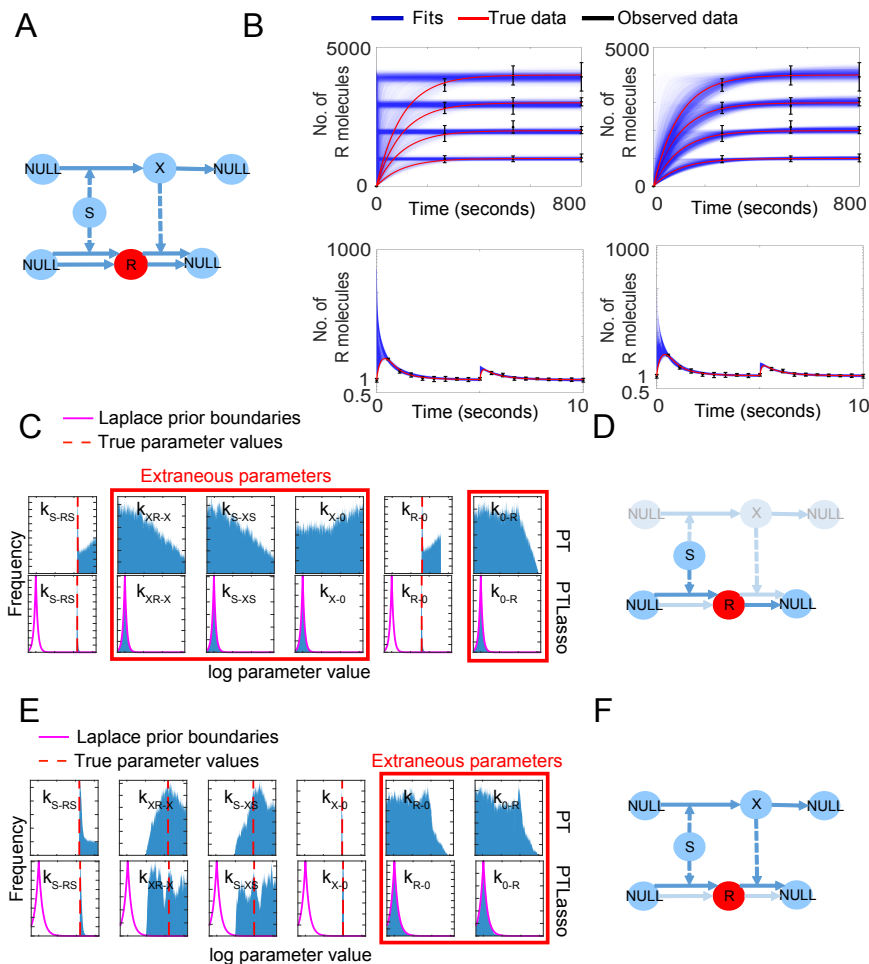
Figure 3.3: Motif inference from a prior network constrained with dose-response data. A) Reaction network diagram of the prior network. The value of the signal S is known, response R (red) is observed at multiple time points, but intermediate X (blue) is not observed. Solid lines show species conversions and dashed lines show influences, where a species affects the rate of the corresponding reaction without being consumed. B) Fit of the model to the linear dose-response data (top row, linear scale y-axis) and perfectly adapting dose-response data (bottom row, log scale y-axis) with PT (left) and PTLasso (right). $k_{s-rs} = 10\,\mathrm{s}^{-1}$, $k_{r-0} = 0.01\,\mathrm{s}^{-1}$ are the true parameter values for the linear dose-response data and $k_{s-rs} = 10\,\mathrm{s}^{-1}$, $k_{xr-x} = 10\,\mathrm{molecule}^{-1}\,\mathrm{s}^{-1}$, $k_{s-xs} = 1\,\mathrm{s}^{-1}$, $k_{x-0} = 1\,\mathrm{s}^{-1}$ are the true parameter values for the perfectly adapting dose-response data (rate constants not specified were set to zero in each case). Transparent blue lines show ensemble fits (from 4,000 parameter samples with 1,000 time points per trajectory for linear dose-response, and from 8,000 parameter samples with 2,000 time points per trajectory for perfectly adapting dose response), red lines show the true data (1,000 time points for linear dose-response, 2,000 time points for perfectly adapting dose response), and the black error bars show the mean $\pm$ standard deviation of the observed data. The four increasing linear dose-response values correspond to S values of 1, 2, 3 and 4, and the two successive perfectly adapting dose response responses corresponding to S values of 1 and 2. C) Frequency histograms showing probability distributions of the parameters for linear dose response fits (from 400,000 parameter samples) with PT (top) and PTLasso (bottom). D) Reduced model corresponding to linear dose-response highlighted in prior network. Faded nodes and arrows are extraneous and are removed from the model. E) Frequency histograms showing probability distributions of the parameters for perfectly adapting dose-response fits (from 800,000 parameter samples) with PT (top) and PTLasso (bottom). For panels C and E, the range of log parameter values on each x-axis is -12 to 6, which covers the full range over which parameters were allowed to vary. The y-axis of each panel is scaled to the maximum value of the corresponding distribution to emphasize differences in shape. The pink lines show the boundaries of the Laplace prior with $\mu = -10$, $b = 0.5$ for the linear dose-response model, and $\mu = -10$, $b = 1$ for the perfectly adapting dose-response model, and the dashed red lines show the true parameter values. A parameter distribution confined within the Laplace prior boundaries indicates that the parameter is extraneous (panels with red border). F) Reduced model corresponding to perfectly adapting dose-response highlighted in prior network. Faded nodes and arrows are extraneous and are removed from the model.

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

Complex biological signaling networks are frequently modular [119, 120] with distinct motifs such as feedback loops that operate on separate time scales [112]. To account for the modular structure of signaling we extended our Lasso approach to grouped Lasso, a technique that applies a module-specific Lasso penalty to all reactions within a particular module (see Methods). PT combined with grouped 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-κB responses to a short TNF pulse [46]. A prior model of NF-κB signaling was created by combining simplified elements of models from [56] and [15] (Table A11). The network was divided into three biologically motivated network modules (Fig. 3.4A). The IκB and A20 modules describe negative feedback mediated by the inhibitor IκB 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-κB 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 is necessary to explain NF-κB responses to a single TNF pulse, PTLasso was used to fit the model to three previously published, experimentally obtained, single-cell NF-κB responses (Fig. 3.4B) [46]. In addition to the NF-κB data, other constraints were applied to make the system behave consistently with known biology. These constraints are listed in Table A1, and Fig. A5 demonstrates that PTLasso correctly followed the imposed parameter covariation.

The probability distributions for the module penalty parameters (Fig. 3.4C) show that the A20 parameter is confined within the prior boundaries while the others have deviated, suggesting that to fit these particular single-cell NF-κB trajectories, the A20 module is dispensable, whereas the IκB 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-κB responses under the given experimental condition and network constraints. The fits with PTLasso were as good as the fits with PT alone (Fig. 3.4B,D), as is demonstrated by comparing the average log-likelihoods (Fig. 3.4E), though the fits with PT produced broader distributions for the A20 parameters (Fig. 3.4F, Fig. A6). The posterior distributions estimated for many of the reaction rate constants showed overlap with values previously reported in the literature [15, 56, 121] (Fig. A6).

To test the requirement of A20 feedback under different experimental conditions and network constraints, we also fit the model to a published single-cell NF-κB response to continuous TNF stimulation [46] (Fig. 3.4G). A soft constraint that IKK responses are transient was added for consistency with published observations [41, 122]. For responses to a TNF pulse, IKK activity naturally adapts back to its baseline abundance without additional negative regulation (Fig. A7). In this case, all three module penalty parameters deviate from the prior (Fig. 3.4H), indicating that the A20 mediated negative regulation of IKK is essential for responses to continuous TNF stimulation. Taken together, the results for the NF-κB signaling model provide an example where PTLasso isolates reaction modules sufficient for responses to specific experimental conditions and time scales.
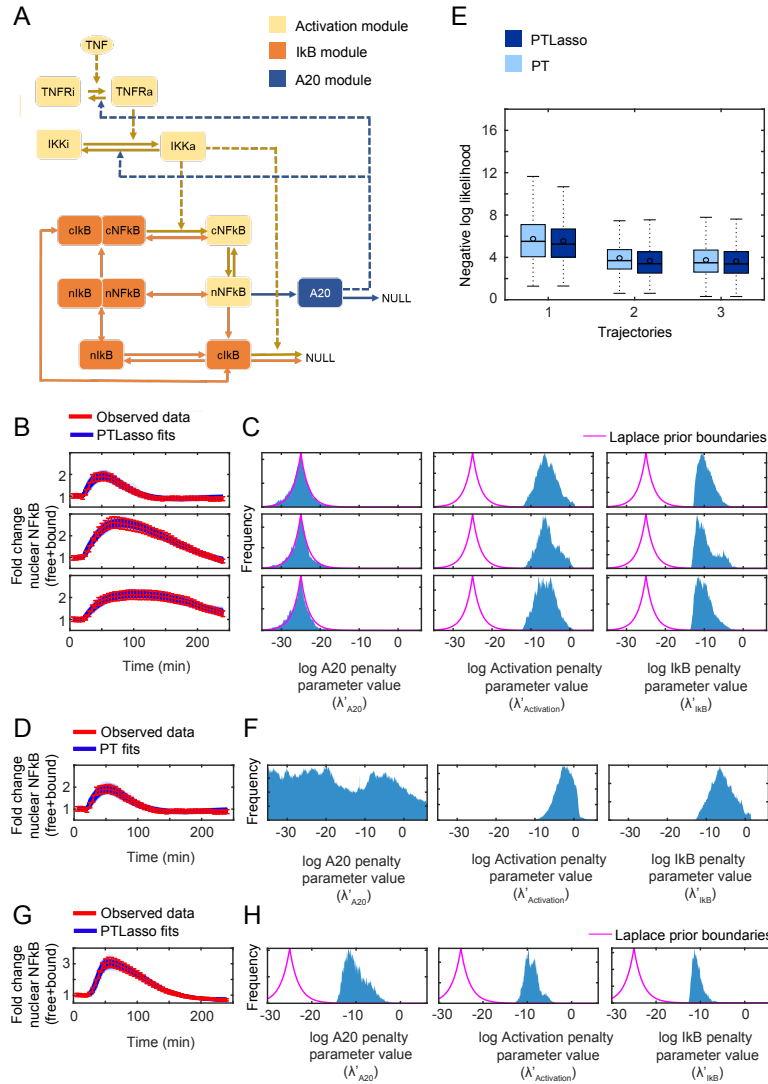
Figure 3.4: Model reduction using grouped Lasso with a model of NF-κB signaling. A) Reaction network diagram of a simplified model of TNF-NF-κB signaling. The colors indicate the different modules. Suffix "a" or "i" refer to active and inactive versions of the species respectively. Prefix "n" and "c" distinguish between nuclear and cytoplasmic versions of the species respectively. Solid lines indicate transformations and dashed lines indicate influences, where a species affects the rate of the corresponding reaction without being consumed. B) PTLasso fits of the model to three distinct single-cell NF-κB responses to pulsatile TNF stimulation. C) Frequency histograms showing probability distributions of the penalty parameters (from 5,640,000 parameter samples) from model fits of single-cell NF-κB responses (from panel B) to pulsatile TNF stimulation with PTLasso. D) PT fits of the model to the single-cell NF-κB response from the first row of panel B to pulsatile TNF stimulation. E) Box plots comparing the log likelihood of the fits for NF-κB responses to pulsatile TNF stimulation (from 5,640 parameter samples) with PT and PTLasso. Trajectories 1–3 correspond to the three trajectories in rows 1–3 of panel B. Boxes show data in the 25th–75th percentile and the circles show the mean. F) Frequency histograms showing probability distributions of the penalty parameters (from 5,640,000 parameter samples) from model fits of the single-cell NF-κB response in panel D to pulsatile TNF stimulation with PT. G) PTLasso fits of the model to a single-cell NF-κB response to continuous TNF stimulation. H) Frequency histograms showing probability distributions of the penalty parameters (from 3,200,000 parameter samples) from model fits of the single-cell NF-κB response to continuous TNF stimulation with PTLasso. For panels B, D and G, transparent blue lines show ensemble fits (from 288 parameter samples for the fits with pulsatile TNF and 168 parameter samples for the fits with continuous TNF stimulation) of the model to single-cell NF-κB responses. An NF-κB response is calculated as the fold change of the sum of the abundances of bound and free NF-κB in the nucleus. Red lines show the experimental data. Error bars show the 10% standard deviation assumed for the likelihood function during fitting and represent measurement error. Pulsatile TNF stimulation is 5 ng/ml for 5 minutes. Continuous TNF stimulation is 0.1 ng/ml. For panels C, F and H, the range of log parameter values on each x-axis covers the full range over which parameters were allowed to vary (-35 to 6 for pulsatile TNF stimulation and -30 to 6 for continuous TNF stimulation). The y-axis of each panel is scaled to the maximum value of the corresponding distribution to emphasize differences in shape. The pink lines show the boundaries of the Laplace prior with $\mu = -25$, $b = 2$. When fitting with PT, the penalty parameter distributions have uniform priors.

3.4 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 dose-response relationships. In a model of NF-κB signaling, PT with grouped 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.

Energy landscapes for systems biology models are often multimodal [123], which raises the possibility that multiple minimal models will fit the data. The fully connected five-node network example demonstrates that PTLasso can identify multiple minima when present, but we also note that the posterior distributions for the parameters k_{BC}, k_{BD}, k_{BE} were not identical as would be expected from symmetry (Fig 3.2D). These differences show that even when the PSRF and MPSRF values are below the standard thresholds, MCMC sampling methods may not obtain the correct probabilities for each possible solution. In a worst case, an apparently converged sample might miss a plausible mechanism.

Here, we have only used PTLasso to reduce ODE models of biochemical kinetics, but the method is in principle agnostic to the modeling formalism used. The grouped Lasso approach to select over reaction modules can even be adapted to select over abstract representations of signaling mechanisms, for example, coarse-grained nonlinear input-output functions, when detailed molecular reaction networks are not known. This may be useful to highlight pathways sufficient for certain experimental data in large multi-pathway models, such as whole-cell models [124] or models of signaling crosstalk [125], where it may not be possible or desirable to accurately represent each biochemical pathway in full mechanistic detail.

Another 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 be explained by differences in parameter values [111], but comparing cell-type specific parameter distributions in high dimensional space may be difficult when the models are non-identifiable. 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 swaps 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 the chains when convergence is reached would prevent unnecessary sampling and reduce the overall execution time. Approaches such as APT-MCMC [126] and Hessian-guided MCMC [31] 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 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 [32] 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 accurately reflect existing knowledge of a biochemical system. But detailed models may obscure our ability to identify underlying mechanisms.

47

PTLasso serves as a bridge between these detailed models and simpler mechanistic explanations that are sufficient to account for system behavior under specific conditions.

4.0 NF-κB Dynamics Discriminate between TNF Doses in Single Cells

This chapter is taken from Zhang Q*, Gupta S*, Schipper DL, Kowalczyk GJ, Mancini AE, Faeder JR, Lee REC, NF-κB dynamics discriminate between TNF doses in single cells, Cell Systems, 2017, *co-first authors [46]. Details of the author contributions can be found in Section 4.5

4.1 Introduction

To mediate cell fate decisions, cytokines such as tumor necrosis factor (TNF) initiate dynamic alterations in the abundance and activity of signaling proteins in each cell [127, 128]. IkB kinases (IKKs) are recruited to ubiquitin structures near the plasma membrane and activated through induced proximity with other signaling proteins [122, 129]. Nuclear Factor κB (NF-κB) inhibitor proteins (IkB) are phosphorylated by active IKKs and rapidly degraded, leading to nuclear accumulation of NF-κB transcription factors and expression of anti-apoptotic and pro-inflammatory regulators [130–132]. In addition to canonical roles, NF-κB responses are context sensitive and, in certain settings, also drive transcription of pro-apoptotic and anti-inflammatory cytokines [133–135]. Since deregulation of NF-κB contributes to disease [136–138], responses to cytokines must be robust to sources of biochemical noise [139, 140] that limit the accuracy of information transmission in each cell [54, 55].

Temporal properties of activated NF-κB in the nucleus control diverse transcriptional programs that can be clustered based on the timing of their expression [56, 141–144]. Early-response genes encode secreted cytokines and negative regulators of the NF-κB pathway such as A20 and IκBα that peak after 1 hour, whereas expression of mid- and late-response genes does not peak until hours later [142, 145]. When measured in a cell population, the duration and concentration of cytokine stimulation differentially regulates the expression of gene clusters [143, 145–147] giving the impression of a tunable signaling system where graded responses regulate different physiological functions.

When viewed in single cells, the abundance and timing of nuclear NF-κB localization show substantial cell-to-cell variability [45, 56, 145, 146]. Models parameterized to single cells exposed to limiting concentrations of TNF or other NF-κB agonists suggest that pathway activation may be switch-like, in which the strength of an inducing signal changes the probability of a stochastic all-or-nothing response in each cell [45, 146, 148, 149]. Switch-like mechanisms for pathway activation reduce sensitivity to noise by filtering low-level environmental fluctuations and preventing spurious inflammatory responses [146, 149]. However, at the single-cell level, a switch cannot transmit dosage information and restricts the range of responses in each cell to either 'on' or 'off' when exposed to time-varying stimuli. It is difficult to reconcile how cytokines control diverse cell fate decisions using a mechanism that transmits little information in each cell.

Here, we use models to show how mechanisms of cytokine responses can be discerned from single-cell data. We measure time courses of nuclear NF-κB in 1000s of single cells exposed to TNF over a range of concentrations and durations, and use an information theory formalism to quantify the channel capacity of the system. Although we find evidence of a threshold for pathway activation that varies between cells, we also find that single-cell dynamics encode sufficient information for multiple states of responses to cytokine conditions. Despite significant heterogeneity when comparing between single cells in the same condition, repeat stimulation experiments reveal dose-dependence when different conditions are compared in the same cell. Our results reveal that single cells can discriminate and grade multiple levels of responses to cytokine doses across orders of magnitude in concentration.

4.2   Results

4.2.1   Single-cell data are necessary to discriminate between models for switch-like and graded signal transduction

To determine how switch-like and graded mechanisms for cytokine responses differ when viewed at the levels of single cells and cell populations, we developed a set of heuristic models

to simulate a heterogeneous population of cells exposed to a range of cytokine doses. Each model represents one of three distinct mechanisms for pathway activation (Figure 4.1A). For all models we simulated 100 single cells, each initialized with a distinct 'cell state' (integers between 1 and 100), which determines how strongly it will respond to stimulus, and a random threshold for activation (See Figures 4.1B and B1). All models included terms for basal and inducible noise to better approximate biological systems (Figure B1). Each model describes a plausible yet distinct scenario, with differences apparent at the single-cell level (Figure 4.1B). For Model 1 (M1), pathway activation is switch-like, and a cell exposed to any dose greater than its activation threshold will respond fully as determined by its cell state. Response distributions overlap between doses in M1 for responsive and non-responsive (NR) cells respectively. For M2, there are no non-responder cells at any non-zero dose (Figure 4.1B, right column), and single-cell responses are graded with 'dose' and cell state. M3 combines M1 and M2, where cell responses increase with dose and cell state, but only for doses greater than the cell's activation threshold. For M3, NR cells exist with overlapping distributions for all non-saturating doses. When each cell's response is normalized to its cellular state in M2 and M3, the response strength to a single input dose is nearly invariant across activated cells (Figure 4.1B, middle column). In contrast for M1, there is no dosage information transmitted beyond sub- or supra-threshold cytokine concentration. Remarkably, when viewed as population-level averages, for example as may be observed for western blots, these models are indistinguishable (Figure 4.1C).
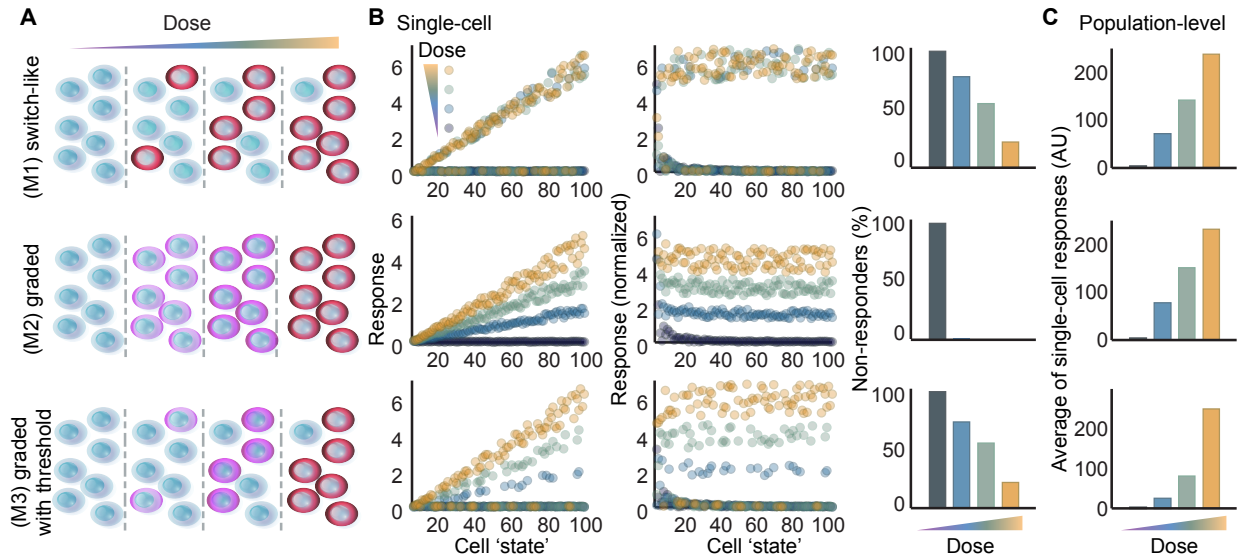
Figure 4.1: Population-Level Data Cannot Distinguish between Switch-like and Graded Response Mechanisms (A) Schematic for (M1) a switch-like mechanism for activation where a cytokine dose increases the probability of an all-or-nothing response in each cell; (M2) a graded mechanism for single-cell activation in which each cell's response is graded in proportion with increasing cytokine dose; and (M3) a combined threshold with graded mechanism where single-cell responses are graded in proportion with cytokine dose only if the dose is greater than the cell's threshold for activation. (B) Simulated responses of single cells modeled with extrinsic noise (described in Figure B1). Each cell is initialized in a unique "cell state" that approximates its responsiveness to cytokine. Responses to four doses are compared in each model plotted as a raw response (left), or a normalized response that divides each cell's response by its cell state (middle). The fraction of non-responder cells is quantified for each dose (right). (C) Three mechanistically distinct models are indiscernible when using the average of single-cell responses to approximate a population-level measurement.

4.2.2 Cell lines show differing sensitivity and responsiveness to cytokine stimulation

To investigate mechanisms of dose discrimination, we set out to identify a cell line that responds strongly to TNF. We used fixed-cell immunofluorescence to characterize subcellular NF-κB localization in a panel of human cancer cell lines. Because the duration of exposure to cytokines can distinctly coordinate downstream signal transduction pathways [44,141,146, 147], cells were exposed to a range of TNF concentrations as a single pulse or continuously. The average nuclear abundance of the RelA subunit of NF-κB was measured from fixed-cell images (Figure 4.2A, thick lines) to compare cell lines and rank their relative sensitivities.

The intensity of nuclear RelA varied among single cells of the same type when exposed to

TNF (Figure B2A. To quantify pathway activation for each cell line, time courses of nuclear RelA averages were decomposed into a series of descriptors (Figure 4.2B; [56, 150, 151]) and plotted as heat maps (Figure 4.2C). Using the area under the curve (AUC) and the maximum fold change (F_{max}/F_i), KYM1 and A549 cells showed the strongest responses to TNF. Although both cell lines had comparable rates of nuclear change ($Rate_{in}$ and $Rate_{out}$) and deviations from adaptation (Figures 4.2C and B2B), KYM1 cells were selected for further study because they are sensitive to the broadest range of TNF concentrations (0.1 ng/mL; Figure B2C).

Temporal properties of molecules within single cells can be quantified over multiple time points using live-cell experiments [55, 56, 141, 145, 152]. To further characterize TNF-induced dynamics of RelA, KYM1 cells that stably express fluorescent protein-RelA (FP-RelA) fusions (Figure B2D-F) were imaged by time-lapse in growth medium with trace amounts of a live-cell dye to assist with nuclear segmentation [147]. Compared between single cells, FP-RelA showed transient and variable translocation into the nucleus (Figures 4.2D,E). Consistent with previous results in HeLa [56], coefficient of variation (CV) for descriptors for nuclear RelA that approximate response dynamics (such as fold change F_{max}/F_i, or adaptation F_f/F_i) showed significantly less variability than others (Figure B2G; non-overlapping 95% confidence interval).

If response dynamics are less variable between cells and also encode condition-specific signals [153], then they may more accurately convey information about the environment. Indeed, calculations for the information transmission using multivariate vectors to describe dynamic responses of single cells, such as RelA dynamics in cells exposed to lipopolysaccharides, yield significantly greater values than those calculated from single time points [55]. Although TNF-induced pathway activation was previously characterized as a <1-bit system using measurements of nuclear RelA at a single time point [54], more information may be encoded by dynamics.
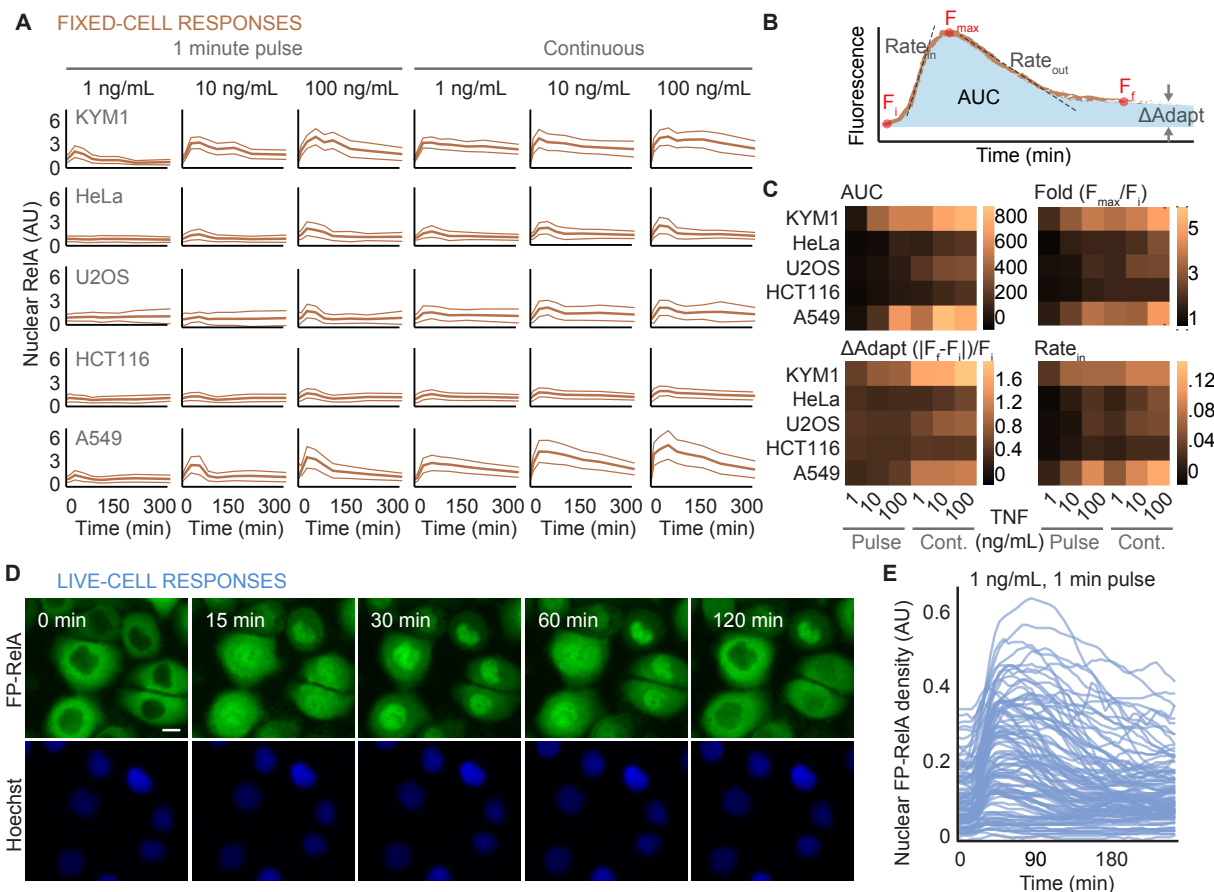
Figure 4.2: Heterogeneity of Responses to TNF between Cell Lines and Single Cells. (A) Time courses for average nuclear RelA from fixed cells are shown for a panel of human cell lines exposed to indicated concentrations of TNF continuously or as a single 1-min pulse (solid orange curve). On average, 11,374 single cells were measured across the time points for each cytokine condition. Light colored lines indicate the SD of approximately 1,274 cells measured at each time point. Time points for fixation included 0, 10, 30, 60, 90, 120, 180, 240, and 360 min following exposure to TNF. (B) Descriptors used to quantify the response of cells to a cytokine. F_i, F_{max}, and F_f, respectively, describe the initial, maximal, and final amount of nuclear RelA fluorescence. AUC, the area under the curve for the cytokine response. ΔAdapt quantifies the deviation from a perfect adaptive response. $Rate_{in}$ and $Rate_{out}$ quantify the maximal rate of nuclear entry and exit, respectively, for average of fixed-cell data. (C) Heatmaps for each descriptor quantified in a panel of cell lines exposed to indicated cytokine conditions. Formulae used to calculate ΔAdapt and fold change (Fold) are shown. See also Figure B2. (D) Time-lapse images of FP-RelA stably expressed in KYM1 cells exposed to a single 1-min pulse of 1 ng/mL TNF. Scale bar, 10 μm. (E) Time courses of nuclear FP-RelA density measured in single cells treated with a 1-min pulse of 1 ng/mL TNF.

4.2.3 Dynamics of TNF-responses carry more information but are confounded by populations of NR cells

To determine if NF-κB dynamics contain more information about TNF dose, we measured single-cell responses to a broad range of TNF concentrations. Cells were also exposed

to a single pulse of TNF with a 30-s, 1-min, or 5-min duration using the same range of concentrations (Figure 4.3A, in addition to an untreated control in Figure B3A) to reflect conditions that cells may encounter in vivo [44, 146]. Time courses for each cell's nuclear FP-RelA were described by a vector, and a k-nearest neighbors approach was used to calculate the information transmission capacity [55]. Because experimental data are subject to measurement noise, the resulting channel capacity estimates a lower bound for mutual information between single-cell responses and TNF conditions or, alternatively, the number of distinguishable dose-response pairs [53]. Using a subset of 26 conditions (Figure 4.3A, where the number of responsive cells is labeled in red in addition to untreated control; see Section 4.4), the information transmission capacity converged to 1.2 or 1.4 bits when representing single-cell time courses in a.u. ("Raw") or as fold change ("Fold"), respectively (Figures 4.3B, B3A, B3B, and B3E).

NF-κB was previously characterized as a system with digital properties, having fewer cells that respond to low strength stimulus [44, 45, 146, 149]. Since the channel capacity estimates the number of distinct response distributions a system can achieve, we asked whether calculations are affected by distributions of NR cells that are likely to exist and overlap in all conditions. We developed a statistical model of an eight-state system (S_0–S_7) with a channel capacity of 3 bits (Figures B3C and B3D; see Section 4.4). Simulations compared scenarios where each of the excited states (S_1–S_7) contained a predefined fraction of non-responders sampled from the ground state distribution (S_0). Even though the channel capacity of the excitable subpopulation remained constant, simulations showed that bit depths for the whole system drop in proportion with the non-responder fraction (Figure B3C) suggesting that information transmitted accurately by responders may be concealed by cells that do not respond to stimulation.

We set out to determine whether NR cells in our dataset impact the channel capacity of TNF-induced signals. Previously, we demonstrated that cells exposed to TNF with less than 1.2-fold change in nuclear FP-RelA are unlikely to activate transcription [44], so we applied the same threshold to categorize cells as TNF responsive and NR (listed in Figure 4.3A) in each condition. The fraction of NR varies widely between TNF conditions, with fewer cells responding to low-concentration or short-duration TNF exposure. By removing these

trajectories from our data, we created a dataset with non-responder cells removed (NRR). Although channel capacity of the Raw-NRR showed a marginal yet significant increase in bit depths over unfiltered data ($p < 0.05$; Figure 4.3B), the Fold-NRR had significantly greater channel capacity (1.7 bits) than the Fold dataset ($p < 10^{-12}$) in addition to a subset from the Fold-NRR with only continuous treatment conditions (1.3 bits; $p < 10^{-16}$; bottom row of Figure 4.3A), which shows that cells transmit information about pulse treatment that is distinct from continuous conditions.

Since descriptors can summarize dynamic properties of a cell's response, we asked whether certain descriptors transmit more information about TNF. Single-cell responses in the Fold-NRR were fit to two sigmoid functions (Figure B3F; [154]), used to compute descriptors described above (Figure 4.2B), in addition to two temporal descriptors for the time of maximal rates for nuclear entry or exit (t_{in} and t_{out}, respectively). We calculated the channel capacity for each descriptor (Figure 4.3C) and found that the AUC, a summary statistic that approximates the whole time course of a cell's response, carries the most information (1.4 bits), followed closely by the maximum fold change.

Our results show that a subpopulation of cells can accurately resolve at least 1.7 bits of information about TNF-exposure conditions. Although the channel capacity using vectors of single-cell dynamics are higher than scalars by definition [155], we also find the system transmits approximately 1.4 bits of information using scalar descriptors. These results suggest that single cells are capable of more than on or off responses to TNF, but do not ensure that single cells can achieve multiple distinct levels of response that grade with cytokine dose.
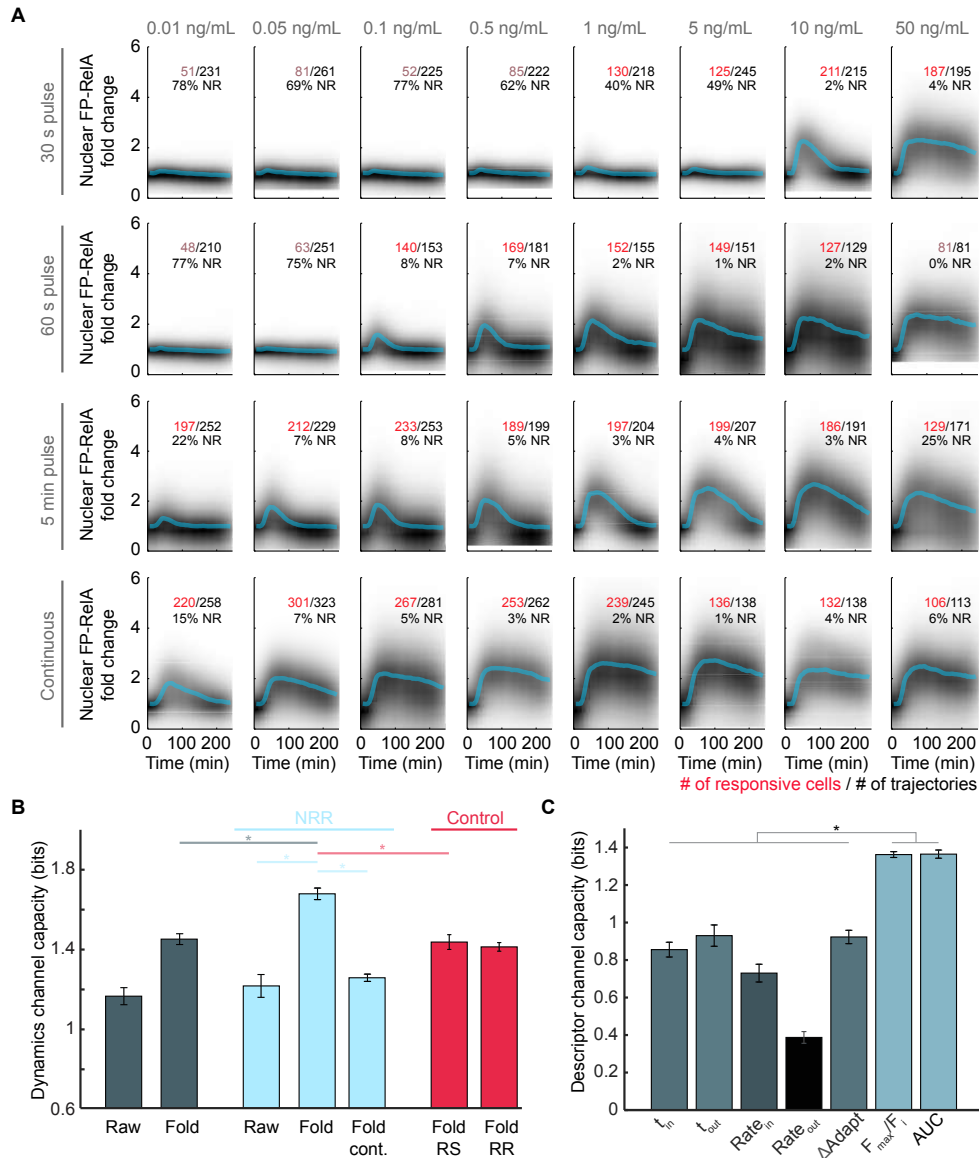
Figure 4.3: Information Transmission Capacity of the TNF-NF-κB Pathway. (A) Density plots of single-cell FP-RelA time courses for responses to TNF with indicated concentration and duration. Median of single-cell responses for each condition is shown in blue. Inset numbers indicate the total number of single-cell time courses collected (black), the number of cells with a significant amount of FP-RelA translocation (red or pink), and the fraction of non-responders (NR) for each condition. (B) Channel capacity values calculated for each dataset: (dark blue) Raw and Fold datasets where each single-cell time course is represented in a.u. or fold change (Figure B3A); (light blue) NRR, datasets where time courses for non-responder cells are removed, the "Fold cont." dataset only includes conditions from the Fold-NRR with continuous exposure to TNF, bottom row of (A); (red) average of 20 subsample control datasets where the same number of cell trajectories are removed from the Fold dataset as in the NRR, but cells were either "Randomly Selected" (Fold RS) or "Responding cells were targeted for Removal" (Fold RR) (see Section 4.4). For all datasets, conditions with fewer than 100 responder cells, pink numbers in (A) were removed from channel capacity calculations; *p < 10⁻¹², t test. (C) Channel capacity values for scalar descriptors of FP-RelA dynamics (*p ≪ 10⁻¹³, t test). Error bars represent SD.

4.2.4 Single cells can grade multiple levels of responses with TNF dose

We used repeat TNF stimulation to determine whether graded responses can be observed in the same cell. From our matrix of live-cell responses (Figure 4.3A), we reasoned that a 1-min pulse of 0.2 ng/mL TNF should induce a detectable yet weak response in most cells, leading to minimal perturbation. Following exposure to a pulse of TNF, we measured expression for A20 and IκBα and observed that the abundance for mediators of negative feedback return to nearly baseline after 3 hr (Figures B4A and B4B; compare 0.2 and 10 ng/mL). Subsequently we refer to a 1-min pulse of 0.2 ng/mL as a "reference," used as a common condition to compare same-cell responses to other TNF conditions. AUC1 and AUC2 (Figure 4.4B, inset) were measured from a manually curated dataset of single cells exposed to a reference, followed by a 3-hr recovery period and then a 1-min "test pulse" of indicated concentration (Figures 4.4A–C). Only cells that were responsive to reference stimulation were included in the dataset. Results from control experiments comparing a 6-hr recovery period were similar, but there were clear indicators of tolerance when using a high concentration reference (Figures B4C–E).

Comparing test pulse only, histograms for AUC2 showed overlap, with a significantly smaller bit depth than values calculated above for response dynamics ($p < 10^{-10}$; Figures B4F and B4G). By contrast, when AUC2 was normalized by AUC1 for the same cell (AUC2/AUC1), and plotted against AUC1, we saw clearer separation between conditions in proportion with TNF concentration (Figure 4.4D; $p < 10^{-84}$ one-way ANOVA; pairwise comparisons in Table B2) analogous to normalized response patterns of M2 and M3 in Figure 4.1B. To compare cells with similar responsiveness to TNF, data were binned by AUC1, and within each bin sigmoid fits were calculated for AUC2 plotted against dose (Figure 4.4E). Although there was a trend that cells with large AUC1 in response to the reference also have proportionately larger AUC2 for the same test pulse condition, all datasets were well-fit by sigmoid functions with comparable Hill coefficients much smaller than 1 (Figure 4.4E, inset). These results demonstrate that the dose-response curve of single cells is not switch-like [156]. Similarly, even though responses to reference are independent of the test pulse, channel capacity values calculated for [AUC1 and AUC2] are significantly larger ($p < 10^{-11}$)

and comparable with the channel capacity calculated from the matrix of live-cell responses (Figure B4G), suggesting that the accuracy of responses increase when conditioned on the cell state. Together, our results demonstrate that single cells are capable of multiple levels of responses that grade in proportion with TNF concentration.
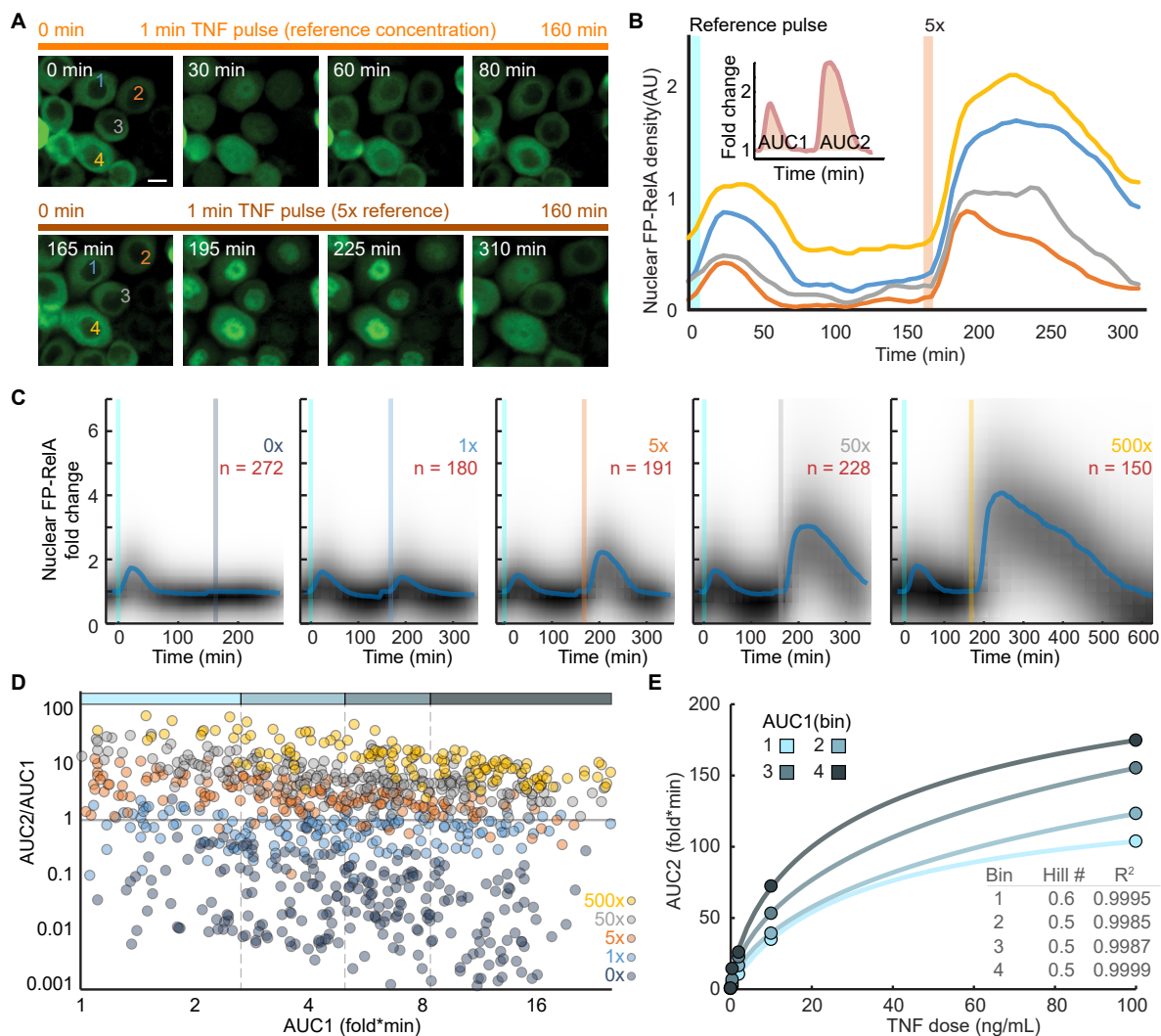
Figure 4.4: Repeat TNF Stimulation Reveals a Graded Mechanism of Dose Discrimination in Single Cells (A) Time-lapse images of FP-RelA in live cells stimulated with a 1-min reference pulse of 0.2 ng/mL TNF, followed by a 1-min pulse with 5× the reference dose (1 ng/mL TNF). Same cells are marked. Scale bar, 10 μm. (B) Representative single-cell time courses for the cells numbered 1–4 in (A). Inset shows schematic of AUC calculations for single-cell responses to the first (AUC1) and second (AUC2) TNF pulse. (C) Density plots of single-cell FP-RelA time courses for cells exposed to a reference TNF pulse followed by a range of increasing test doses. Median of single-cell responses is shown in blue and inset numbers indicate the number of single-cell time courses collected in each condition. (D) Scatterplots showing AUC2/AUC1 stratified along AUC1 across the range of test conditions. Colored bar along top depicts bins of single cells based on AUC1 into an approximately equal number of cells per condition. (E) Average response of single cells (AUC2) to increasing test pulse concentrations (0, 0.2, 1, 10, or 100 ng/mL TNF) for each bin in (D); inset describes Hill coefficients and goodness of fit for logistic regression.

4.3 Discussion

A key question is how coarse-grained is a single cell's ability to detect and respond to inflammatory cytokines in its environment: do activating signals increase the probability of a stochastic all-or-nothing response, or do cells precisely grade their responses to a range of stimulus levels? In this work, channel capacity calculations and repeat stimulation experiments show that some cells can accurately grade multiple levels of responses to cytokine conditions. Although these observations contradict a purely switch-like model for pathway activation proposed by M1 (Figure 4.1), we also observed in all conditions a subpopulation of cells with a non-responder phenotype. The fractional abundance of non-responder cells varies widely depending on the intensity of cytokine conditions (Figure 4.3A), which contradicts the purely graded model of pathway activation described by M2 (Figure 4.1B, right column). Taken together, our results support M3, in which low-level cytokine fluctuations are filtered by a threshold for pathway activation that varies between cells, and responses to supra-threshold conditions are graded with dose.

Several factors that were not considered in previous studies were critical to observe greater than 1.5 bits of information in the Fold-NRR (Figure 4.3B): (1) by analyzing the subpopulation of responder cells in isolation from non-responders the channel capacity was increased by $\sim$0.3 bits (cf. Fold and Fold-NRR); (2) by representing single-cell time courses as fold change, the channel capacity was increased by $\sim$0.5 bits (cf. Raw-NRR and Fold-NRR). This is consistent with incoherent feedforward network motifs in many signaling systems that more robustly encode and detect fold-change dynamics [56, 151, 157]; and, finally, (3) $\sim$0.5 bits of information was gained by exposing cells to TNF as a pulse in addition to continuous treatment conditions across a range of concentrations (cf. "Continuous-Fold-NRR" and Fold-NRR). Although each factor contributes to the channel capacity, their combined effects are non-linear and synergistic. Above all, repeat stimulation experiments that compare responses to a reference in the same cell were essential to demonstrate the capabilities of single cells.

Cellular information processing can be divided into categories of "cell autonomous," such as the channel capacity of a single cell studied here, and "cell population" in which cells share ligands through paracrine feedbacks to coordinate decisions over longer timescales

[158]. Underlying both is the impression that signal transduction networks have an intrinsic capacity to transmit information that is the same in every cell, yet our results suggest that this may not be the case. Although non-responder cells share a similar "response-phenotype," their abundance varies between TNF conditions, which suggests they are a heterogeneous subpopulation that cannot be defined by a unique network state. For example, cells that do not express a critical component of the TNF-receptor complex may be completely refractory to any TNF stimulus yet biochemically distinct from other NR cells. Whereas stochastic fluctuation in the abundance of regulatory proteins in a cell may broaden or restrict the range of stimulus conditions to which it can respond. It is therefore conceivable that cell subpopulations have distinct information transmission capacities that are conditioned by their biochemical state, and that mixtures of these subpopulations lead to underestimates of the network's capacity to transmit information.

Sources of noise make biological systems appear unpredictable, but that does not mean they are unreliable. Through accounting for heterogeneity in cell states, the information transmission capacity of single cells is better approximated, and it is found that cells can grade multiple levels of responses to cytokines over a range of doses. Although it remains to be determined whether grading is biologically important for cell fate decisions, either cell autonomous or through cell-cell communication, grading provides additional control points for cells to fine-tune and diversify responses to stimuli.

4.4   Methods

See full-text online for key resources table:
`https://doi.org/10.1016/j.cels.2017.10.011`

4.4.1   Experimental model and subject details

4.4.1.1   Cell Culture   Parental KYM1 (female), HeLa (female), U2OS (female), HCT116 (male), and A549 (male) cell lines (ATCC) were cultured in RPMI, DMEM, McCoy's 5A,

McCoy's 5A, and F12K media respectively at 37°C and 5% CO2. Parental KYM1 and KYM1 cells infected with a retroviral construct to express mVenus-RelA under control of the CMV were kindly provided as a gift from Suzanne Gaudet. From these cells, a clonal cell line was selected based on FP-RelA expression, localization, and responsiveness to TNF. All media was supplemented with 10% Corning Regular FBS, 100 U/mL penicillin, 100 μg/mL streptomycin and 0.2 mM L-glutamine (Invitrogen).

4.4.2 Method Details

4.4.2.1 Fixed-cell immunofluorescence For fixed-cell experiments, cells were seeded into plastic bottom 96-well imaging plates (Fisher) at the following densities: KYM1 12000 cell/well 24 hours prior to treatment, HeLa 7000 cell/well 24 hours prior to treatment, U2OS 6000 cell/well 24 hours prior to treatment, HCT116 20000 cell/well 24 hours prior to treatment, and A549 5000 cell/well 48 hours prior to treatment or 6000 cell/well 24 hours prior to treatment for low cytokine concentration test. On the day of the experiment, media containing cytokine (TNF) was prepared at 15X the desired concentration for each well. Timing of cytokine treatment was planned so that fixation for all time points (0, 10, 30, 60, 90, 120, 180, 240, and 360 minutes) occurred simultaneously at time zero. For continuous treatments, 12.3μL of pre-warmed 15X cytokine mixture was spiked into wells and mixed. For pulse treatments, 12.3μL of 15X cytokine was mixed into wells and after 1 minute all the media was removed, cells were washed in PBS, and 185μL of fresh media was added to each well. Between treatments the cells remained in environmentally controlled conditions (37°C and 5% CO2) and 15X cytokine solutions were kept at 4°C to prevent degradation.

At time zero, media was removed from the wells, 185 μL of PBS was used to wash the wells, and wells were incubated at room temp in 120 μL of 4% paraformaldehyde (PFA) in 1X PBS for 10 minutes. Wells were washed 3X three minutes with 185 μL 1X PBS and then incubated in 120 μL 100% methanol for 10 min at room temp. Next, wells were washed 3X three minutes in PBS-T (1XPBS 0.1% Tween 20) followed by 120 μL of primary antibody solution (3% BSA in PBS-T) with a primary antibody (selected from: 1 μg/mL α-RelA [sc-8008; Santa Cruz]; 5 μg/mL α-A20 [ab13597; Abcam]; or, 1:400 α-IκBα [#4818;

CST]). Plates were wrapped in para-film and left to incubate at 4°C overnight. The following morning, wells were washed 3X five minutes in 185 μL PBS-T followed by incubation for 1 hour in 120 μL of the secondary antibody solution (3% BSA PBS-T, 4 μg/mL Goat anti-Mouse IgG Alexa Fluor 647 (Thermo Fisher)). 185 μL PBS-T was used to wash the wells for 5 minutes and they were put into 120ul Hoechst solution (PBS-T, 200ng/mL Hoechst) for 20 min. Finally wells were washed five minutes with PBS-T and then 185 μL PBS was used to fill the wells and keep the cells hydrated during imaging. Cells were imaged using Delta Vision Elite imaging system at 20x magnification with a LUCPLFLN objective (0.45NA; Olympus).

4.4.2.2 Live-cell imaging Live cells were imaged in an environmentally controlled chamber (37°C, 5% CO2) on a DeltaVision Elite microscope equipped with a pco.edge sCMOS camera and an Insight solid-state illumination module (GE). KYM1 cells expressing FP-RelA were seeded at a density of 25000 cells/well 24 hours prior to live-cell imaging experiments on no. 1.5 glass bottom 96 well imaging plates (Matriplate). 2 hours before imaging, live-cells were pre-stained with 300ng/mL Hoechst 33342 for 1 hour. Following this all growth medium contained trace concentrations of Hoechst (60 ng/mL) to maintain the nuclear stain and assist with segmentation in downstream applications. Consistent with previous studies [147] cells continued to grow and divide normally for at least 10 hours while imaged in these conditions. Wide-field epifluorescence and DIC images of KYM1 cells were collected using a 20x LUCPLFLN objective (0.45NA; Olympus). For all treatments, cytokine mixtures were prepared and pre-warmed so that addition of 120μL added to wells results in a final concentration as indicated. For wash-out experiments, cytokine-containing media was aspirated, wells were washed with pre-warmed media, and fresh media was added to each well. FP-RelA, time-lapse images were collected over at least 4 fields per condition with a temporal resolution of 5 minutes per frame before and after exposure to TNF. Cells were imaged for at least 30 minutes prior to addition of TNF.

4.4.2.3 Western blots and analysis Kym1 cells stably expressing FP-RelA were cultured for 24 hours in complete growth medium. After treatments, cells were lysed in SDS-

based lysis buffer consisting of 112 mM Tris-Cl, pH 6.8, 4% SDS supplemented with protease and phosphatase inhibitors at 4°C for 30 min. Protein extracts were clarified by centrifugation at 4°C at 12,000 × g for 10 min. Nuclear extracts were isolated using NE-PER™ Nuclear and Cytoplasmic Extraction Reagents (Pierce). Lysate protein levels were quantified by BCA assay (Pierce). Samples were separated by SDS-PAGE, 25 μg total protein per lane, then transferred to PVDF membranes. Blocking was done in 5% milk in TBS for 1 hour. Primary antibodies directed at RelA, A20, IκBα, β-actin, Lamin A/C(#4764, #5630, #4814, #3700, and #4777 respectively; Cell Signaling Technology), GAPDH (sc25778; Santa Cruz) were diluted in 5% milk in TBS-T and incubated overnight at 4°C. Alexa 680/800-conjugated secondary antibodies (LI-COR) were used in combination with an Odyssey (LI-COR) scanner for detection and quantification of band intensities.

4.4.2.4 Gaussian model An 8-input system (S_0 through S_7) was modeled by sampling 200 points from each of eight 2 dimensional Gaussians with unit variance and means separated by a scalar in one dimension and 0 in the other dimension. Here, S_0 can be thought of as a ground state and each additional state is excited by a scalar multiple to model a dose-response system with a theoretical maximum channel capacity of 3 bits. Models with 6 units (no overlap) and 2 units (partial overlap) of separation were compared (Figure B3C). To simulate the effects of a non-responsive fraction f, 200f points were sampled from the first Gaussian distribution (ground state S_0), and 200(1- f) points sampled from the Gaussian for each of the excited states (S_1 through S_7). Channel capacity calculations were carried out for 7 values of f, ranging from 0 to 0.6 to match the range of experimentally observed fractions of non-responders.

In a complementary set of experiments, we simulated the effects of sub-sampling on channel capacity calculations (Figure B3D). Here, we started with a mixed population that contains 50% non-responders and 50% cells capable of dose discrimination (i.e. f = 0.5). Selecting for removal the fraction of cells as indicated on the x-axis either randomly (red) or by targeting non-responders (blue), mutual information was calculated. We see that random subsampling does not increase the channel capacity, whereas targeting non-responders does.

4.4.3 Quantification and statistical analysis

4.4.3.1 Fixed-cell image analysis Cell Profiler [159] was used to segment the nuclear compartment cells using the Hoechst channel, and quantify median nuclear intensity values from each cell in fixed-cell images. Data were compiled and statistical analysis performed using MATLAB. Cell numbers used to calculate mean and standard deviations for fixed-cell experiments are summarized in Table B1.

4.4.3.2 Live-cell analysis Custom scripts in MATLAB were used to flat-field correct, background correct, and register movies of FP-RelA. Nuclear regions were segmented using the Hoechst channel and single-cell trajectories of nuclear FP-RelA were tracked using Cell Profiler [159]. For each cell at each time point, median nuclear intensity of FP-RelA was reported by Cell Profiler and represented as three-frame running averages to reduce the influence of high-frequency noise. The fold-change transform was carried out by dividing each single-cell time course by its average nuclear fluorescence from four imaging frames before the time point at which TNF was added (Figure B3A). Cells that were observed to die, divide or leave the field of view within the experimental timeframe were excluded from analysis. Trajectories for each condition were shown as a density plot in MATLAB [160,161]. Cell numbers collected in each live-cell experiment are listed in their respective figure panels.

4.4.3.3 Channel capacity calculations All channel capacity calculations were performed using updated MATLAB scripts (2017-08-15) provided by Roy Wollman and used as described previously [55]. Briefly the algorithm uses a k-nearest neighbor density estimation to compute the conditional probability density of each single-cell response. These are used to compute the entropy of the response, and the entropy of the response conditioned on the signal. The difference in these two terms gives the mutual information, which is maximized over all possible probability distributions of the signal, using the MATLAB optimization function fmincon. The method is described in more detail in [55].

A k value of 5 was used for all the k-nearest neighbor probability density estimation steps. The values reported are the mean and standard deviation of 10 jackknife iterations.

At each iteration, the channel capacity was estimated by subsampling from the data set 20 different sample size (SS) fractions ranging from 0.6 to 0.95. To reduce sample size bias and estimate channel capacity variance, linear regression was used to extrapolate the channel capacity value at a sample size of infinity, i.e. $1/SS = 0$ (See red dots in Figure B3B). The flatness of the resulting slopes are consistent with observations from Selimkhanov et al. [55], suggesting that the size of our data set is sufficient to accurately estimate channel capacity values, i.e. values calculated for the finite sample size are similar to the extrapolated values.

Time courses of nuclear RelA were collected for thousands of single cells across conditions and the information transmission capacity was calculated for the dynamics of TNF-driven signaling. We next described single cell responses using vectors of increasing dimension to compare channel capacities (Figure B3E). For these calculations, each single-cell time course was sub-sampled into a multivariate vector of dimension 'd' using a uniform grid centered on the middle time point as described previously [55]. We used jackknife resampling to correct for sample size bias and determine the variance of calculated bit depths as described by Selimkhanov et al [55]. (Figure B3B). Using a subset of 26 of the tested conditions, discussed below in further detail (Figure 4.3A; conditions where the number of responsive cells is labeled in red), calculations for channel capacity converged to a maximum value of 1.2 bits in fewer than 5 dimensions (Figure B3E, orange line for the 'Raw' data set). Jackknifing to estimate sample size bias confirmed that we had sufficient data to accurately estimate the channel capacity at a vector dimension of 5 (Figure B3B). We therefore use d = 5 to compare channel capacity estimates in subsequent experiments.

To maintain adequate sample sizes for channel capacity calculations and ensure that equal numbers of conditions are compared between data sets, we filtered conditions with fewer than 100 trajectories in the NRR data set for all channel capacity calculations (Figure 4.3A; conditions with pink numbers). All data sets therefore include a total of 26 conditions: 25 TNF-response conditions with at least 100 trajectories in the NRR, in addition to the untreated control sample that was not filtered for non-responder cells. The channel capacity was calculated for each data set using uniformly sampled vectors with an increasing number of dimensions. As previously shown [55], we see that that increasing the number of dimensions increases the estimate of the channel capacity at first, but after a certain point convergence is

achieved upon which adding more vector dimensions does not increase the estimated channel capacity.

For the "NRR" datasets, all single-cell trajectories with a maximum fold change of less than 1.2 were removed from the analysis [44]. We found that channel capacity calculations were robust to slight variations in this threshold. To show that the increase in the estimated channel capacity on removing non-responder cells is not a trivial effect of subsampling the data set, we generated two control data sets, "Fold RS" and "Fold RR". To get "Fold RS" we calculate the total number of non-responding cells in our "fold" data set and remove an equivalent number of trajectories randomly distributed throughout the data set. To get "Fold RR", we remove an equivalent number of 'responder' trajectories randomly distributed among responding cells in the data set. The channel capacity estimated from both of these is similar to that obtained using the "fold" dataset. The randomized controls were generated 20 times, and for each we performed 10 jackknife replicates. The values reported are the mean and variance of the channel capacity obtained from each of the 20 randomized data sets.

4.4.3.4 Sigmoid fits to estimate t_{in} and t_{out} To estimate the times of maximum rate of FP-RelA nuclear entry and exit, the following two-sigmoid model was used:

$$y(t) = \frac{1}{p_1 + p_2 e^{-p_3 t + p_4}} + p_5, t \leq t_{max} \tag{4.1}$$

$$y(t) = \frac{1}{p_6 + p_7 e^{-p_8 t + p_9}} + p_{10}, t > t_{max} \tag{4.2}$$

Where t_{max} is the time of max fold change of nuclear NF-κB . The sigmoid is fit to the data using the genetic algorithm function (ga) in MATLAB to minimize the objective function:

$$\Sigma_t \left(\frac{y(t) - y_{true}(t)}{y_{true}(t)}\right)^2. \tag{4.3}$$

4.4.3.5 Sigmoid fits to calculate hill coefficients The bin boundaries for AUC1 were specified as follows: [0.25 2.8 5.5 9.4 50]. This was done to maintain roughly equal numbers of cells per bin. For cells that fall within a certain bin, i.e. they have similar values of AUC1, a 4 parameter logistic equation (L4P) was fit to describe the relationship between AUC2 and dose, i.e.

$$AUC2 = D + \frac{A - D}{1 + (\frac{dose}{C})^B} \tag{4.4}$$

The hill coefficient, i.e. B obtained for each of these fits was less than 1.

4.4.3.6 Extracting descriptors from NF-κB dynamics

- Max fold change, $\dfrac{F_{\max}}{F_{\mathrm{i}}}$: the maximum value calculated from the fold-change transformed FP-RelA time course. The fold-change transform is carried out by dividing the time course by the average of the first four imaging frames leading to the time point at which TNF was added (F_{i}).

- *Time of max, $t_{\max}$: the time at which the nuclear FP-RelA fluorescence is maximum. If $t_{\max}$ is more than 170 minutes after stimulation, the descriptor could not be determined and the time course was discarded from analysis for this descriptor.

- *The time of max rate in, t_{in}: the time at which the rate of FP-RelA entry into the nucleus is maximum as determined from the slope of the first sigmoid fit. If t_{in} could not be determined within 155 minutes of TNF stimulation, then the time course is discarded from analysis for this descriptor.

- *Maximal rate of nuclear entry, $Rate_{\mathrm{in}}$: was calculated from the data as the slope of the straight line fit to three points on the single-cell time course centered at t_{in}.

- *The time of max rate out, t_{out}: the time at which the rate of nuclear FP-RelA exit from the nucleus is maximum as determined from the slope of the second sigmoid fit.

- *Maximal rate of nuclear exit, $Rate_{\mathrm{out}}$: was calculated from the experimental data as the slope of the straight line fit to three points on the single-cell time course centered at t_{out}.

- Adaptation, ΔAdapt: is the ratio of the absolute difference in the final and initial nuclear FP-RelA to the initial amount of nuclear FP-RelA.

$$\Delta\text{Adapt} = \frac{|F_{final} - F_{initial}|}{F_{initial}} \tag{4.5}$$

- F_{final} is calculated by averaging the last 4 frames of the nuclear FP-RelA time course, and F_{initial} is calculated by averaging 4 frames of the nuclear FP-RelA time course preceding TNF stimulation.

- Area under the curve, AUC: is the area bounded by the nuclear FP-RelA time course and the baseline amount of nuclear FP-RelA, i.e. for fold change where initial nuclear FP-RelA value of 1, i.e. $\Sigma_{i=time\ of\ stimulus}^{i=end\ point\ of\ response}(\max(1, X(i)) - 1)$. The end point of a response is marked as the frame at which the second sigmoid becomes flat after t_{out}, i.e. $\frac{\Delta fold}{\Delta t} < 0.0002$, or the last frame, whichever occurs first.

*Note that temporal and rate features (t_{in}, t_{out}, Rate_{in}, Rate_{out}) are not defined for the untreated control conditions. The control condition was not included in calculating the channel capacity of these features.

4.5   Author contributions

Conceptualization, R.E.C.L. and J.R.F.; Methodology, R.E.C.L., Q.Z., D.L.S., S.G., and A.E.M.; Investigation, Q.Z. and D.L.S.; Software, S.G., G.J.K., and R.E.C.L.; Formal Analysis, Q.Z., S.G., D.L.S., G.J.K., and R.E.C.L.; Writing – Original Draft, R.E.C.L. and S.G.; Writing – Review and Editing, R.E.C.L., S.G., Q.Z., D.L.S., G.J.K., and J.R.F.; Visualization, R.E.C.L.; Funding Acquisition, R.E.C.L. and J.R.F.; Supervision, R.E.C.L.

Q.Z. and S.G. are co-first authors, Q.Z. performed experiments, and S.G. performed computational analyses.

5.0 Modeling Heterogeneity in TNF-NF-κB Signaling

5.1 Introduction

Cells use biochemical signaling networks to detect environmental cues and translate them into responses and cell fate decisions [3]. Heterogeneity is increasingly being recognized as a common property in cell signaling systems [49, 50, 162], even in genetically identical cells [50, 51]. Since dysregulation of signaling is often associated with disease [3, 4], it is surprising to observe considerable spread in responses to the same stimulus, with response-distributions often overlapping between treated and untreated conditions [54, 64]. Understanding the molecular mechanisms for this spread will improve our understanding of single-cell signaling processes, enable rational design of perturbations to manipulate single-cell behavior, and have consequences in pharmacology and disease treatment [163], for instance in addressing therapeutic problems such as fractional killing [164–166] and cancer treatment failures that can arise from non-genetic cell heterogeneity [167]. In this work, we specifically address the heterogeneity that is observed in the TNF-NF-κB signaling pathway [46], a central mediator of inflammation and innate immunity [40, 52].

Computational models, calibrated to experimental data, play a key role in studying complex signaling mechanisms [16, 56, 168]. Population-level experimental techniques, such as western blots, have been useful in calibrating computational models to the population-average behavior of signaling variables [16]. However, since the average response may not represent any cell in the population [49, 169, 170], for example a population average measurement may obscure out-of-phase single-cell oscillations [171], and rarer cell responses at the tails of the response distribution may be biologically important [49], there has been a rise in single-cell signaling experiments, e.g. flow cytometry based assays and live-cell imaging [172], and an increased need for computational models that can explain single-cell signaling data.

Snapshot techniques, like flow cytometry, enable single-cell measurements of several intracellular signaling variables [172–174], making it possible to observe the distribution of multiple signaling proteins across a large population of cells and identify subpopulations

where present [60]. While computational methods have been developed to calibrate models to snapshot data [175], the data has the fundamental limitation that a single-cell response cannot be tracked over time. We know from previous studies, that temporal dynamics play a crucial role in cell signaling [46,55]. Live-cell microscopy, while more limited in the number of signaling variables that can be labeled and observed, presents the distinct advantage of being able to track single-cell responses over time [172, 174]. The first challenge that we address in this work, is calibrating a computational model of NF-κB signaling to previously published heterogeneous single-cell data from live-cell microscopy [46].

Heterogeneity in signaling systems can be modeled as intrinsic noise, arising from stochastic fluctuations in gene regulation, or extrinsic noise, arising from differences in the biochemical states of cells [176]. Repeat stimulation experiments have shown that heterogeneous NF-κB responses are more predictable when conditioned on a cell's biochemical state, approximated by its response to a minimally invasive short pulse of TNF [46], suggesting that variability in NF-κB responses is dominated by extrinsic noise. A common way to mathematically express extrinsic noise is through differences in the parameters of an ordinary differential equation (ODE) model [57,58,171,177].

A range of techniques have been developed to model extrinsic variability in signaling responses based on differences in model parameters [178]. Yao et al. modeled heterogeneity in calcium signaling by independently fitting all model parameters to each single-cell trajectory [57]. However, this approach requires as many fitting operations as there are cells, which many not always be desirable with complex models and large imaging experiments since model fitting may be computationally expensive. Cheng et al. modeled heterogeneity in LPS-NF-κB signaling by starting with a population-level parameterization of the model, assuming log-normal distributions on key model parameters, and identifying standard deviations of the parameter distributions for which simulated response distributions would match the experimental data [171]. Dharmarajan et al. describe more general mixed-effects modeling approaches where individual parameters for each cell are constrained by an overall population distribution, and both the cell parameters and overall population distribution parameters are determined through iterative optimization [58]. This approach requires mathematical assumptions on the relationship between individual cell parameters and the overall

population distribution. Kalita et al. developed a flexible approach to model heterogeneity in TNF-NF-κB signaling, by using Bayesian parameter estimation with Markov Chain Monte Carlo (MCMC) to jointly fit individual parameters, that are unique to each cell, and shared parameters, that are common to all cells, to multiple single-cell responses (dataset size is 20 cells) measured with live-cell microscopy [59]. Note that there are several other examples in the literature of jointly fitting individual and shared parameters to explain heterogeneity in responses, e.g. [177, 179].

Here, we follow the approach from Kalita et al. [59], and jointly fit a small number of NF-κB responses to pulsatile TNF, under the hypothesis that the variation can be captured by differences in abundances of key signaling proteins. We extend the approach to include an additional downstream sampling step (similar to Cheng et al. [171], see Section 5.2 for details) to capture all the responses in our dataset (dataset size is $\sim$ 200 cells, data is from Chapter 4, published in [46]), as simultaneously fitting to a large number of trajectories with Bayesian parameter estimation alone would lead to a very high dimensional parameter search problem, and calculating the objective function, which involves an ODE integration for each cell trajectory, would be computationally expensive. Our preliminary results show that while it is possible to capture the full experimental dataset of NF-κB responses using variation in two protein parameters, the resulting model does not predict well for other biochemical species. We discuss possible causes for this issue and avenues for future investigation in Section 5.3.1.1.

We further extend the analysis by demonstrating examples of use cases of the computational model of variability in examining single-cell dose response properties and the relationship between protein abundances and information transmission capacities of single cells. While an increase in extrinsic noise has been shown to decrease the dose discernibility and information transmission capacity of the TNF-NF-κB signaling pathway at a population-level [180], here we use simulation to examine whether single-cell NF-κB signaling capabilities can vary with protein abundances. Overall, we show that computational models of signaling variability enable simulation based analyses that may in the future guide experimental manipulation of single-cell signaling.

5.2 Methods

5.2.1 Data and models

For calibration, we use previously published single-cell NF-κB responses to a 5-minute pulse of 5 ng/ml TNF (Figure 5.1A, [46]). The dataset has 199 time courses. Each time course lasts 220 minutes after stimulation and comprises measurements taken at 5 minute intervals.

To fit the data, we use a simplified 19-parameter model of NF-κB signaling [12] (Figure 5.1B), that uses coarse-grained reactions to describe the TNF mediated activation of its cognate receptor TNFR, leading to the translocation of NF-κB into the nucleus mediated by a central kinase, IKK, and negative feedback mediated by NF-κB driven transcription of its inhibitor, IκB, which binds to NF-κB and shuttles it out from the nucleus. For simplicity, the model omits the longer timescale negative regulation via A20, that in previous work was found to be non-essential for fitting the model to single-cell responses to short pulses of TNF [12]. The fitted observable is the fold change of total nuclear NF-κB, that is, sum of the free nuclear NF-κB and the bound complex NF-κB-IκB.

Following Kalita et al. [59], we fit the model to multiple cell responses simultaneously. Total protein abundances for NF-κB and IKK are unique to each response, while the other 17 parameters (including receptor numbers and rate constant parameters) are shared. The full model to be fitted can be expressed as

$$Y_i = f(\text{NF-}\kappa\text{B}_i, \text{IKK}_i, v), \quad i = 1, \ldots, N, \tag{5.1}$$

where N is the number of cells to be fitted, Y_i is the model output corresponding to the ith cell, uniquely parameterized by protein abundance parameters NF-κB$_i$ and IKK$_i$. v is a vector of shared parameters that includes the receptor abundance and kinetic rate constants. f is the set of ODEs that describe the evolution of each biochemical species in the model over time.

The parameter vector to be fitted is given by

$$[\text{NF-}\kappa\text{B}_i, \text{IKK}_i, v], \quad i = 1, \ldots, N, \tag{5.2}$$

Therefore, the total number of parameters to be fitted is 2N+17, where 17 is the length of v.

5.2.1.1 Constraining shared parameters with PT

For details of the parameter estimation algorithm, PT, see Section 3.2.1.2. Fitting to all 199 trajectories simultaneously is challenging for two reasons: (1) The parameter search problem would be very high dimensional (the dimensionality of the space would be $2*199+17 = 415$), and (2) calculation of the objective function at every step of the PT algorithm would require 199 numerical integrations of the ODE model (one for every cell trajectory).

For these reasons, we use a 2-step heuristic approach of (1) fitting a small subset of our data (4 trajectories), and (2) checking to see if jointly fitting to 4 trajectories sufficiently constrains the shared parameters, enabling a fit to the rest of the dataset simply by uniformly sampling a range of values for total IKK and NF-κB .

For the first step, we manually pick 4 trajectories that visually encompass the range of possible behaviors (see Figure 5.1C). We define $F_{max,t<80}$ as the maximum fold change in nuclear NF-κB within the first 80 minutes after stimulation. The 80-minute cut-off was a heuristic choice that was made to exclude a handful of trajectories that had spuriously high fold changes at later time points, most likely due to technical artifacts. We pick three trajectories that have the maximum, median and lower quartile value of $F_{max,t<80}$ to cover the range of magnitudes of response. To ensure that we have representation of different response shapes, we filter the data for adaptive responses, and include a representative adaptive response as the fourth trajectory. Our heuristic test for adaptation is that the fold change of NF-κB in the nucleus at 105 minutes after stimulation is between 0.9 and 1.1. The resulting representative data set for fitting is labeled $dataset_1$. As a first test for sensitivity to the trajectories chosen, we create an alternate dataset, $dataset_2$, by selecting a different adaptive response. Note that for this preliminary analysis, $dataset_1$ and $dataset_2$ are very similar, and only differ slightly in one trajectory (Figure C2A).

We fit the model in Eq 5.1 and parameter vector in Eq 5.2 to $dataset_1$ (Figure 5.1C), using PT. The fit includes certain hard constraints on known parameter covariation (Figure C1C), the fraction of NF-κB in the nucleus at equilibrium, and a check to make sure that the

model has reached equilibrium before simulating the TNF pulse (see Table C1 for full list of constraints). We use a Gaussian likelihood function for fitting, with a standard deviation, representing measurement error, that is 5% of the data value. We use the Potential Scale Reduction Factor (PSRF) of the energy chain as our convergence diagnostic [115]. A PSRF value less than 1.2 is taken to be consistent with convergence [29].

5.2.1.2 Simulating response distributions by sampling protein abundance parameters The steps described above in Section 5.2.1.1, describe an approach to get an ensemble of shared parameters with the property that varying IKK and NF-κB abundances is sufficient to capture 4 representative trajectories. Extending this further, for any shared parameter set in the ensemble, we uniformly draw a large number of Latin Hypercube (LH) samples, N_{lhs}, of IKK and NF-κB abundances. For any given shared parameter set we can therefore create a simulated response distribution by varying IKK and NF-κB abundances (Figure 5.3A). Note that drawing N_{lhs} parameter samples does not guarantee N_{lhs} simulated NF-κB responses, as parameter samples may be rejected due to integration errors and failure to meet constraints on the equilibrium fraction of NF-κB in the nucleus. However, we check that the number of samples we have drawn is sufficient for the problem by making sure that the results are similar when the number of samples is doubled (Figure C4A,B). Latin Hypercube Sampling is performed using the MATLAB file exchange function *lhsu*, available here: `https://www.mathworks.com/matlabcentral/fileexchange/4352-latin-hypercube-sampling`

To test whether the resulting response distribution adequately explains the experimental data, for each experimental response, we calculate the Euclidean distance to its nearest neighbor in the simulated response dataset [171], which is the error associated with that particular response (Figure 5.3B). For a given error threshold, we calculate the number of experimental trajectories that have an error below the threshold to determine how well the experimental dataset is captured in the simulated data (Figure 5.3C). Empirically, we use error thresholds of 1 (for a good fit) and 2 (for an acceptable fit) (Figure C4C).

Extracting the nearest simulated neighbors for each experimental response is used to create a 1-NN (one nearest neighbor) dataset for downstream analysis. Note that while

the experimental dataset has 199 trajectories, the 1-NN dataset may have less, as multiple experimental trajectories may have the same nearest neighbor.

5.2.2 Analyzing the simulated response distributions

5.2.2.1 Clustering We bin the simulated trajectories by response magnitude, defined as the maximum fold change of nuclear NF-κB , into three equally spaced bins (small, medium and large responses) ranging from 1.2-fold (the threshold of response that has previously been determined to be required for transcription [44]) and $\sim$4.7-fold (the upper limit of the experimental data). Within each bin, we scale the trajectories from 0 to 1 to emphasize differences in shape and degrees of adaptation, and perform k-means clustering (k=3).

5.2.2.2 Simulating dose-response behavior The simulated data was obtained by calibrating the model to an experimental condition of 5 ng/ml 5 minute TNF pulse. To test the dose-response behavior of the model, we simulate the behavior of the 1-NN datasets for a range of TNF concentrations. We have experimental data for a subset of these concentrations, and we validate the model by comparing the mean of the experimental dataset and the mean of the simulated dataset. As before, we use a Euclidean distance measure of error.

We use a scalar statistic of maximum fold change of nuclear NF-κB, to summarize a cell's response to a particular dose. To investigate the relationship between parameters and dose-response behavior, we compute the following scalar dose-response properties:

- *EC50* is the TNF concentration at which the NF-κB response is half maximal
- *EC90* is the TNF concentration at which the NF-κB response is 90% maximal
- *EC10* is the TNF concentration at which the NF-κB response is 10% maximal
- *Bandwidth* is calculated as the difference between the $\log_{10}$ EC90 and $\log_{10}$ EC10 values and reflects the range of responsiveness of a cell to TNF
- *Channel Capacity*, measured in bits, is an information theoretic metric that summarizes the dose discrimination ability of a signaling pathway [47, 48]. We use code from [48] to calculate the channel capacity.

We simulate responses for the following array of 52 doses: $[10^i, 2.5 \times 10^i, 5 \times 10^i, 7.5 \times 10^i]$, where i is an integer that ranges from -6 to 6. To numerically calculate the ECX, that is the TNF concentration at which the NF-κB response is X% maximal, we first use a 1D interpolation to get response values on a dose grid at a higher resolution than what we simulate. The ECX response is calculated as $\min_{response} + \frac{X}{100}(\max_{response} - \min_{response})$. We then find the dose in the high resolution dose grid that has an interpolated response value closest to the ECX response, and return that dose as ECX. $\max_{response}$ is the maximum fold change of NF-κB at the highest administered dose, which is a saturating dose. We also refer to $\max_{response}$ as the *saturating response*. $\min_{response}$ is the maximum fold change of NF-κB at the lowest administered dose, and is typically close to 1.

5.2.3   Software and code availability

The model is written in BioNetGen using the rule-based modeling language, BNGL. MEX code for the coupled differential equations is exported from BioNetGen, allowing the model to be simulated in MATLAB, using the CVODE library [82] for efficient numerical integration of ODEs in C. Third party MATLAB libraries were used to calculate PSRFs (`https://research.cs.aalto.fi/pml/software/mcmcdiag/`), perform LHS (`https://www.mathworks.com/matlabcentral/fileexchange/4352-latin-hypercube-sampling`, [181]), and for plotting [182,183]. Supplementary code for this chapter, including model and data files, is available at

`https://pitt.box.com/s/lx9useuiuxq4r0vjhfgk2cknjiqiwd0d`

5.3   Results

5.3.1   Variation in protein abundance parameters is sufficient to fit a model to heterogeneous single-cell NF-κB responses

Our model of variability parameterizes each single-cell with unique values for total NF-κB abundance and total IKK abundance, while all other model parameters are shared across

all the cells (see Section 5.2.1, [59]). All the parameters are assumed to be unknown and are simultaneously fit to data (Equations 5.1 and 5.2, [59]). On fitting the model (Figure 5.1B, Equation 5.1) to $dataset_1$ (Figure 5.1C), we find shared parameter sets for which varying just the total abundance of NF-κB and IKK is sufficient for the fit (Figure 5.2C, Column 1). 40 shared parameters sets (Figure C4D) are drawn from the fitted ensemble, at regular intervals in the PT chain, and coupled with the Latin hypercube sampling (LHS) step described in Section 5.2.1.2, to get a simulated response distribution (Figure 5.3A) for each shared parameter set. Using the error metric described in Section 5.2.1.2 (see Figures 5.3B,C), we find that the 40 tested simulated response distributions, generated using the 40 shared parameter sets, are able to capture 59-71% of the data within an error of 1 Euclidean distance unit, and 93-96% of the data within an error of 2 Euclidean distance units (Figure 5.3D,G). We note that the experimental dataset contains a handful of noisy trajectories that may reflect technical artifacts in the imaging and image analysis (see last two rows in Figure C4C), and it is not surprising that some of these are not captured by the current model architecture. The procedure was repeated for $dataset_2$ (Figure C2) with a comparable result. 40 tested simulated response distributions, created using shared parameter sets obtained from the fit to $dataset_2$, are able to capture 65-74% of the data within an error of 1 Euclidean distance unit, and 90-95% of the data within an error of 2 Euclidean distance units (Figure 5.3E,H).

Instead of the full $dataset_1$, if only a single trajectory is fit using PT, the shared parameters in the resulting ensemble do not all have the property that varying NF-κB and IKK is sufficient to capture the data (Figure 5.3F,I). One might ask if this issue could be rectified by increasing the range of variation for the protein parameters (previously set to (3–7 $\log_{10}$ molecules/cell). We increased the range to 1–8 $\log_{10}$ molecules/cell, and regenerated simulated response distributions corresponding to five different shared parameter sets (Figure C3A,B). Note that even though the sampling range is 1–8 $\log_{10}$ molecules/cell, not all regions of this space may be valid given our model constraints (Table C1), and parameter sets will be rejected if, for instance, the equilibrium fraction of NF-κB in the nucleus is outside the set range (see Section 5.2.1.2). For two of the five shared parameter sets, increasing the protein variation range enables capturing much more of the experimental data (Figure C3C, shared parameter set index 1,2), for another two the fraction of data captured

after the increase is still modest (Figure C3C, shared parameter set index 3,5) and for one there is minimal change (Figure C3C, shared parameter set index 4). Overall, even after increasing the protein variation range beyond reasonable biophysical limits, fitting to just one trajectory did not produce shared parameter sets that can consistently capture the experimental data by varying NF-κB and IKK expression. Thus fitting jointly, to even a small fraction of the dataset, is a viable approach to systematically obtain an ensemble of shared parameter sets with the property that a desired degree of variation in protein abundance will capture the larger experimental dataset.

5.3.1.1 Summary of limitations and avenues for future investigation   While our variability model is able to jointly fit the NF-κB data well, there are some caveats that merit future investigation.

- First, the number of *active* IKK molecules/cell and *active* Receptor molecules/cell are very small (Figure 5.2C Columns 8,9), and outside of reasonable biophysical limits, even though the total number of IKK and Receptors are reasonable (Figure C1A,B). A related technical observation is that values sampled for the parameter k_{t1a}, the IKK mediated degradation rate of free cytoplasmic IκB, are very large, and the distribution pushes up against the prior boundary. There are several factors that could be controlling this behavior. Here, we discuss three:

    - *Limitations of parameter sampling* – While the PSRF for the energy chain was less than 1.2 (Figure 5.2A), the PSRF for some parameters was still greater than 1.2 (Figure C1B), indicating that while we have converged in terms of quality of fit, we may not have sufficiently sampled parameter space to find equally likely parameters that also give biophysical predictions.

    - *High-temporal correlations in the data could lead to over-fitting* – Our data for nuclear NF-κB is measured at a very high temporal resolution (5 minute intervals). Requiring the model to fit data for one variable closely (the standard deviation of the error model is 5% of the data value, see Section 5.2.1.1), without constraints on others, could lead to over-fitting. Preliminary results suggest that there may be a trade-off between goodness of fit for the NF-κB trajectories and biophysical behavior

for other model species, that can be leveraged to identify parameter sets that fit the data reasonably well while adhering to biophysics. This can be achieved in one of two ways, either by acquiring more experimental data (for instance jointly fitting both NF-κB and IKK in the same cell), or by introducing qualitative constraints in the model on the behavior of other species.

– *Limitations in the variability model structure* – Ultimately, it is also possible that fitting jointly to multiple NF-κB responses reveals limitations in the model structure that may not be apparent when fitting to individual trajectories alone. It is possible that additional regulation at the level of receptors and IKK, or more detailed molecular mechanisms (instead of our coarse-grained reduced representations), or even variation in more parameters are needed to correct the model behavior.

- Second, there is much room for improvement in strategies for picking the representative trajectories. While here we selected the trajectories manually, a potential unbiased approach is to cluster or bin the trajectories and pick the median of each cluster. A more thorough investigation of sensitivity of any model predictions to the initial representative trajectories chosen is another key step in establishing meaningful biological results using the variability model.

- Third, as future work it would be valuable to check for sensitivity to (1) the number of representative trajectories used in the fit, and (2) the choice of error metric. While fits to both $dataset_1$ and $dataset_2$ were successful in obtaining ensembles of shared parameter sets for which simulated NF-κB distributions, obtained by LHS sampling on protein expression, can capture the data, there are other properties of the distributions that differ between the two. For example, using shared parameters from the fit to $dataset_1$, we performed a clustering analysis (see Section 5.2.2.1) on one of the simulated LHS datasets, which revealed clusters in the response distribution that correspond to distinct clusters in protein expression space (Figure C5A,B). However, on repeating the same clustering analysis on one of the simulated LHS datasets from the fit to $dataset_2$, we found that the relationship between protein expression and response space was not conserved (Figure C5C,D). It is not possible to completely rule out pseudoconvergence as a reason for the two fits converging to parameter distributions with different properties.

However, by fitting to a larger number of representative trajectories, and choosing an error metric that is more stringent than Euclidean distance thresholds of 1 and 2, it may be possible to identify more constrained shared parameter distributions, for different sets of representative trajectories, that have conserved properties and predictions.
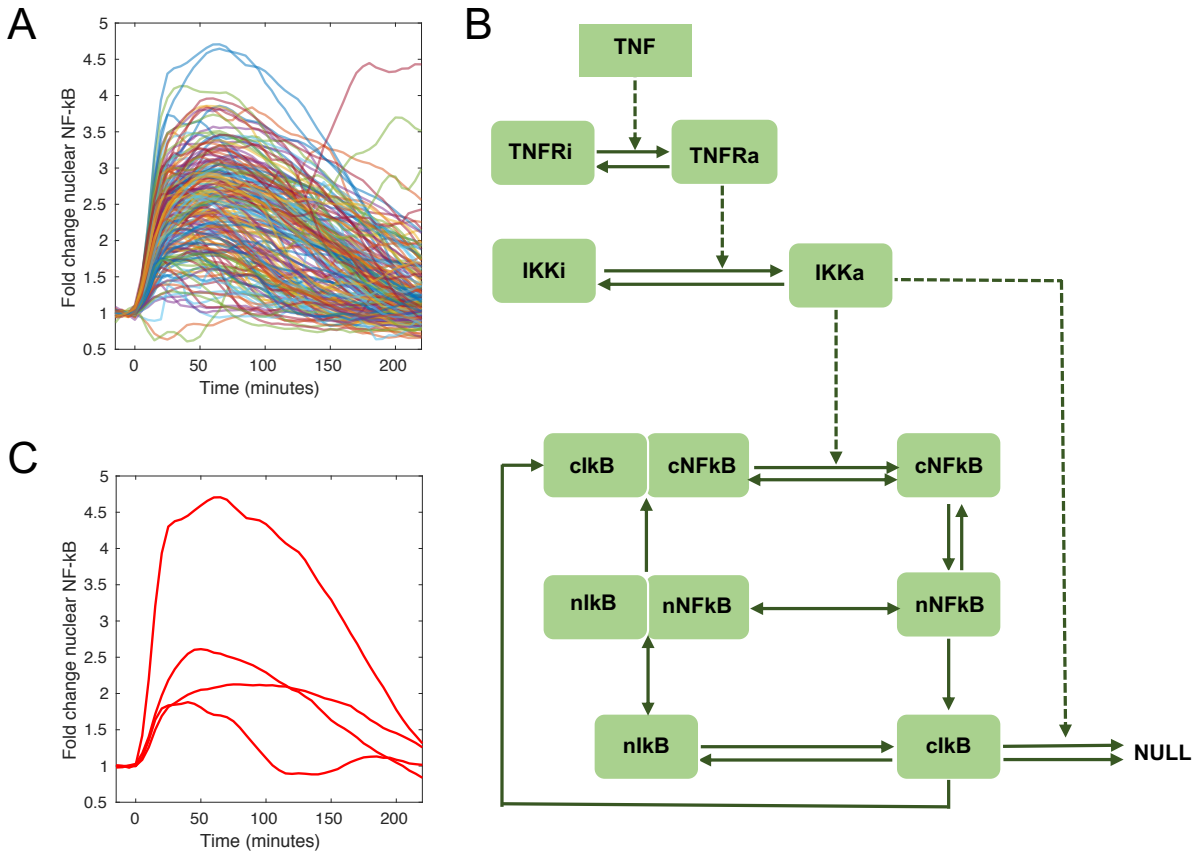
Figure 5.1: NF-κB signaling model and single-cell data. A) Experimental data. Colored lines show time courses for single-cell NF-κB responses to a 5 ng/ml 5 minute pulse TNF stimulation. Data is from [46]. B) Reaction network diagram of a simplified model of TNF-NF-κB signaling [12]. Following [12], suffix "a" and "i" refer to active and inactive versions of the species respectively. Prefix "n" and "c" refer to nuclear and cytoplasmic versions of the species respectively. Solid lines represent transformations and dashed lines represent influences, where a species affects the rate of the corresponding reaction but is not consumed. C) Representative NF-κB responses used for fitting with PT ($dataset_1$).
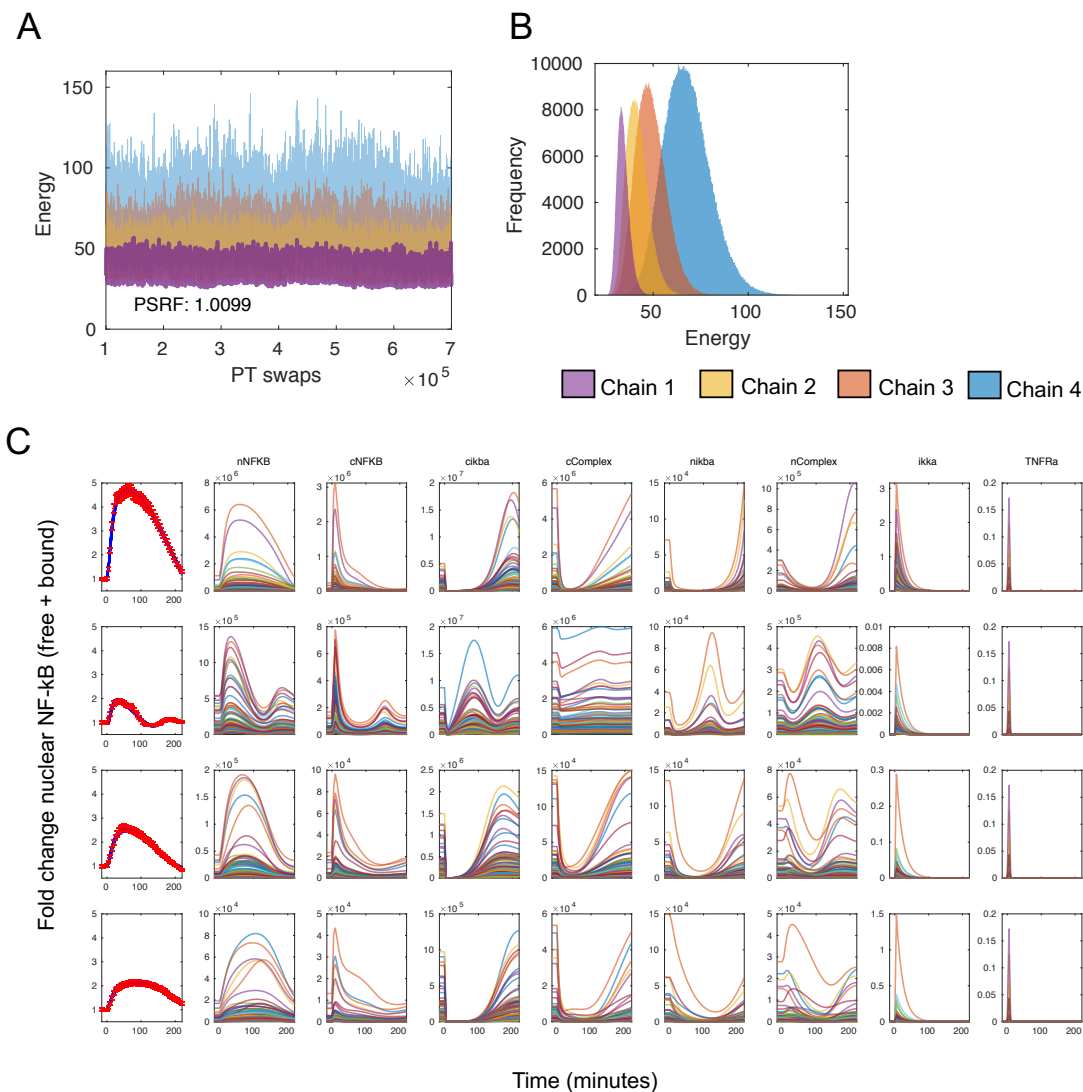
Figure 5.2: Fitting the NF-κB signaling model to multiple single-cell responses with PT. A) PT energy chains corresponding to different temperatures. Burn-in samples have been discarded. The lowest energy chain (in purple) corresponds to a temperature of 1. Chains 1–4 are in increasing order of temperature. B) Energy distributions corresponding to the different temperature chains. C) Ensemble fits to the NF-κB data (Column 1, red error bars show the data, and transparent blue lines show the model fits), and corresponding predictions for other biochemical species in the model (Columns 2–9). The ensemble shows fits from 100 parameters sets drawn at regular intervals from the PT chain.

83
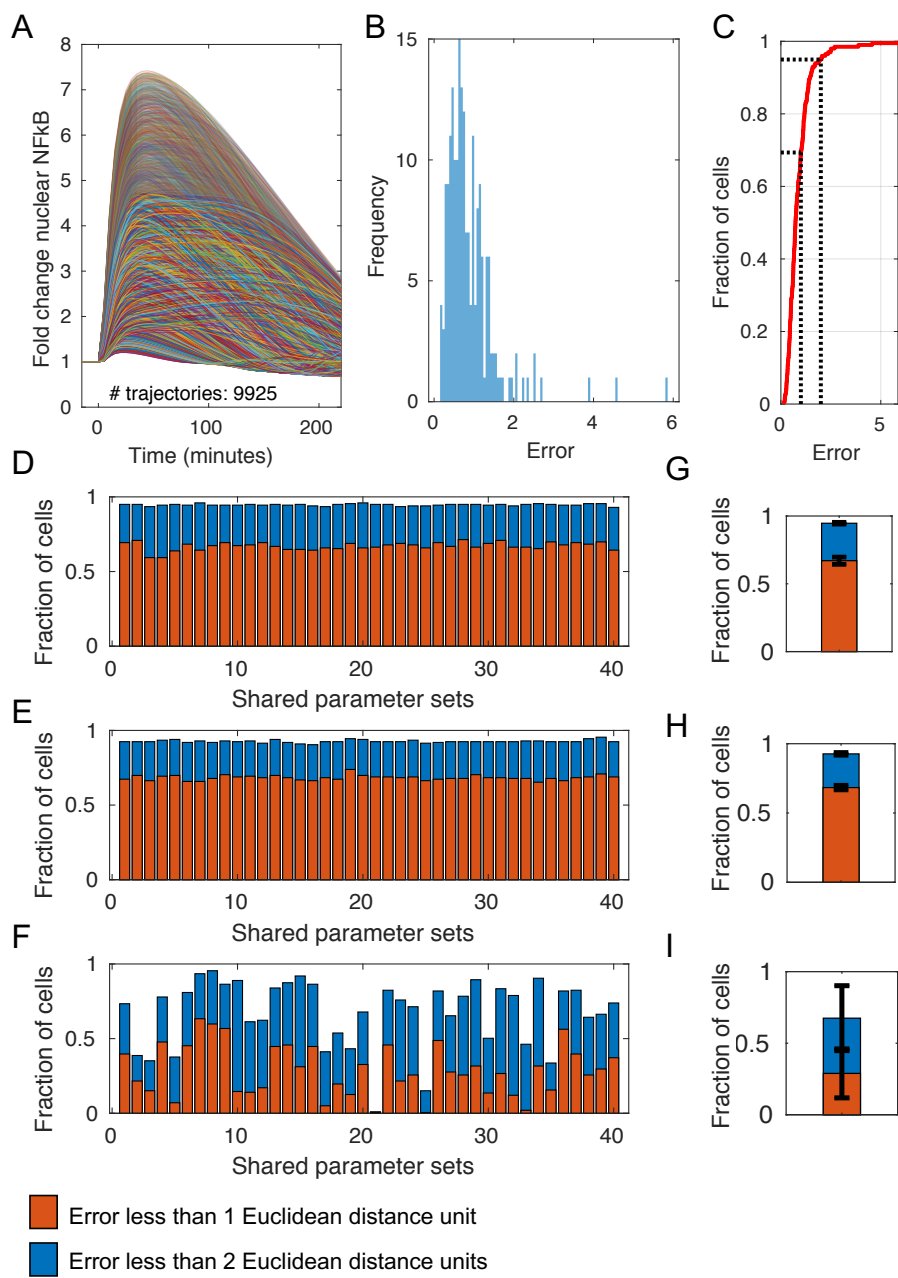
Figure 5.3: Latin hypercube sampling from protein distributions to get simulated response datasets. A) Simulated NF-κB responses obtained by uniformly sampling the $\log_{10}$ values for total NF-κB and total IKK in the range 3–7. Simulated NF-κB responses that have a maximum fold change greater than what is observed experimentally are shown with added transparency. B) The distribution of Euclidean distances between each experimental trajectory and its nearest simulated neighbor. The distribution has 199 points (one for each experimental trajectory) C) Cumulative distribution corresponding to the error distribution in B. For every error threshold, the graph shows the number of trajectories below the error threshold, and the dashed black lines correspond to error thresholds of 1 and 2. D) Fraction of experimental data that falls within an error threshold of 1 (orange) or 2 (blue) for simulated distributions obtained from 40 different shared parameter sets in the PT ensemble for $dataset_1$, E) $dataset_2$, and F) a single trajectory. G) Mean and standard deviation of the data in D. H) Mean and standard deviation of the data in E. I) Mean and standard deviation of the data in F.

84

5.3.2 Simulating single-cell dose-response behavior using the calibrated model of variability

Here, we demonstrate an application of computational models of variability in studying single-cell dose-response behavior. By extracting the nearest neighbors for every experimental trajectory from within a simulated NF-κB dataset of LH samples, we create a 1-NN dataset of simulated cells that resembles our experimental dataset (Figure 5.4B, Section 5.2.1.2). Using the 1-NN dataset (Figure 5.4B), generated with one of the shared parameter sets from the ensemble fit to $dataset_1$, we simulate the single-cell model parameters with a range of TNF concentrations to study the dose-response behavior (Figure 5.4A). We assess the model's predictive capabilities by measuring the error, as the Euclidean distance between the experimental mean and simulated mean, for doses where experimental data is available (Figure 5.4E). As expected, the error is lowest at the calibration dose (Experimental dose index=6, TNF concentration = 5 ng/ml). There are some systematic differences between the experimental data and corresponding simulated predictions, especially at higher doses (note that at lower doses both the experimental and simulated responses become small, and so the Euclidean distance error declines). However, the error for all the tested TNF concentrations was below 2 Euclidean distance units (our previous heuristic for an acceptable fit). Considering a scalar response of maximum fold change of nuclear NF-κB , we obtain single-cell dose response curves (Figure 5.4C). We calculate scalar properties of the dose-response curves including the EC10, EC50, EC90 and bandwidth values (see Section 5.2.2.2, Figure 5.4F), that play a role in describing a cell's capabilities in responding to environmental cues, and we observe the extent to which these properties vary between single cells (Figure 5.4G).

In testing the sensitivity of the dose-response predictions across 20 shared parameter sets, we find that while the error in the mean responses at the experimentally tested doses is fairly consistent across shared parameters (Figure C6A, though note that now the average error at dose index 8, that is 50 ng/ml, crosses the Euclidean distance threshold of 2), the full dose response behavior varies considerably (Figure C6B, Figure C7). For example, across shared parameters, there are considerable differences in the fraction of cells that have an EC90 dose within the experimental saturating dose of 50 ng/ml, as well as differences in the distributions

of saturating nuclear NF-κB responses (Figure C7), with some distributions containing cells that have a saturating nuclear NF-κB response of greater than 10-fold, which is considerably larger than responses observed experimentally (for example, Figure C7 row 4, column 2). Overall, while fitting to 4 representative NF-κB responses was sufficient to constrain the result of varying NF-κB and IKK at the calibration dose (Figure 5.3D,E,G,H, Figure 5.4D), the resulting shared parameter sets are not constrained in their dose-response predictions and we find a variety of dose-response models that are consistent with our calibration data (Figure C7). More model constraints are required to know which dose-response models are an accurate reflection of the biological system. While some have more constrained distributions for bandwidth, others have bandwidths that vary across orders of magnitude, suggesting that it may be possible for isogenic cells to have considerably different ranges of responsiveness.

5.3.2.1 Summary of limitations and avenues for future investigation

Here, we have demonstrated an important use-case for a variability model in performing single-cell simulations of dose-response behavior, data that is currently not accessible experimentally due to the technical difficulties associated with stimulating the same cell repeatedly while observing long time-scale signaling events. However, there are still some caveats and avenues for future investigation that will further facilitate drawing biological conclusions about our system.

- First, we see considerable parameter sensitivity in dose-response behavior.
- Second, while we expect that the dose-response range measured experimentally will not be exactly true for each single cell, it is also unlikely for an EC90 dose to be orders of magnitude above an experimentally measured saturating dose, or for the saturating NF-κB response to be two to five fold higher than the maximum observed experimentally.

A potential approach to address both of these related concerns is to introduce some dose-response constraints, such as hard thresholds on the EC10, EC90 and saturating response, in the initial PT fitting to further constrain the estimated distributions of shared parameter sets.
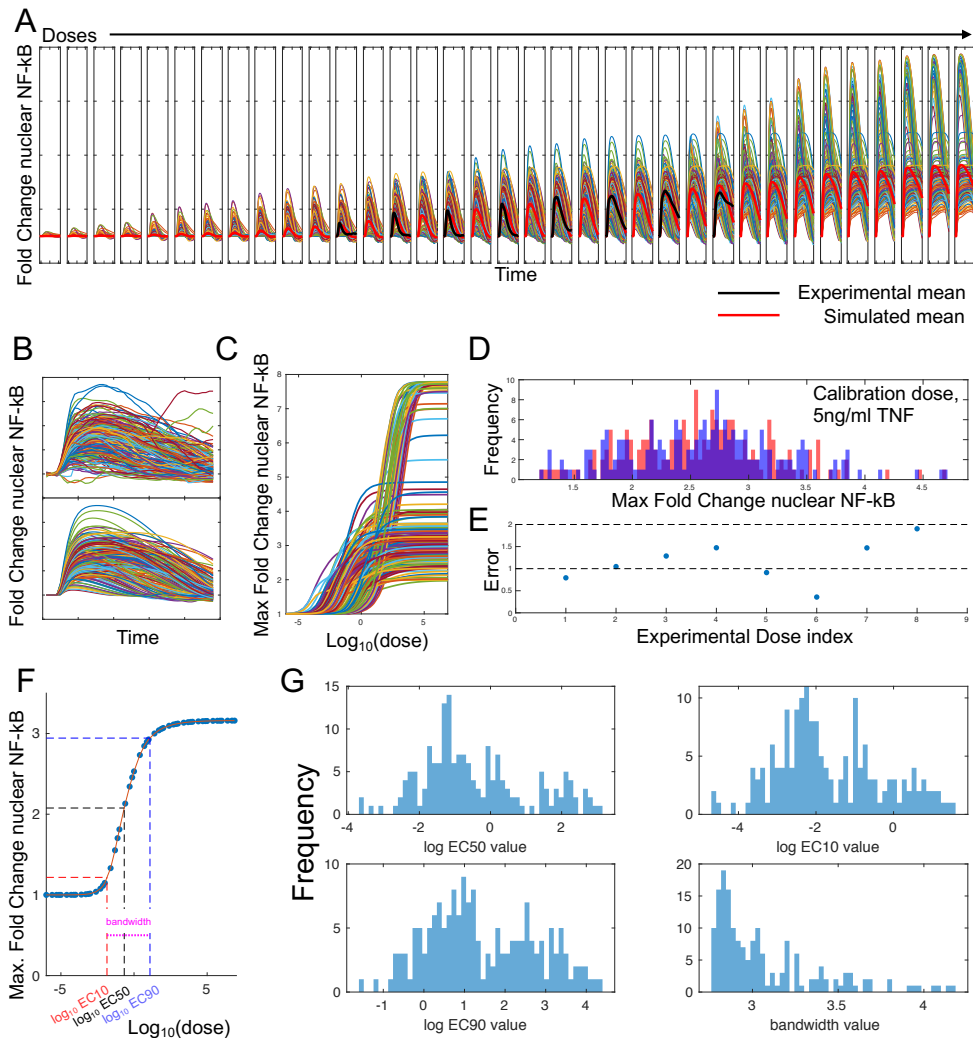
Figure 5.4: Single-cell dose-response simulations. A) Colored lines show simulated single-cell trajectories (parameters taken from a representative 1-NN dataset) to a range of TNF pulse concentrations (pulse length is 5 minutes, concentrations shown range from 2.5×10^{-5} to 7500 ng/ml). Red lines show mean simulated responses. Black lines show the means of the experimental data for available concentrations. y-axis limits are [0,8]. B) Experimental data used for calibration (top) and corresponding 1-NN simulated dataset (bottom). C) Simulated single-cell dose response curves, where the response is the maximum fold change of nuclear NF-kB. D) Comparison of simulated vs. experimental distributions of maximum nuclear NF-kB fold change at the calibration dose. E) Error between simulated and experimental mean, measured as the Euclidean distance, for the 8 doses for which experimental data is available, that is TNF = [0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50] ng/ml. Dashed lines show error thresholds of 1 and 2 that were previously used as heuristics for good and acceptable fits respectively. Dose index 6, is the calibration dose of 5 ng/ml and has the smallest error. F) Dose-response features displayed on a representative dose-response curve. The blue scatter points are the doses at which the response was simulated. The orange line connecting the points is a 1-D interpolation. The red, black and blue dashed lines show the log of the EC10, EC50 and EC90 dose respectively. The pink dashed line represents the bandwidth as the difference between log EC90 and log EC10. G) Distributions of dose response features across the simulated cells in panel B.

5.3.3 A computational model of NF-κB signaling variability supports variable single-cell channel capacities

Finally, we use the simulated single-cell dose response curves in Figure 5.4C, to show an example of calculating single-cell channel capacities. Traditionally, the channel capacity of signaling systems, used to assess the ability of a system to accurately discriminate environmental cues, has been calculated at the population level [46–48]. Recently, in G-protein coupled receptor (GPCR) signaling, the channel capacity has been shown to be highly variable between single cells [184]. However, to experimentally probe single-cell channel capacities, Keshelava et al. [184] had to repeatedly stimulate the same cell with increasing concentrations of calcium, measuring dose-response relationships in the same cell. While this was possible for calcium mediated GPCR signaling, for which the response is measured on the time scale of seconds, this is not possible for NF-κB signaling where the response occurs on the time-scale of hours. Because of the presence of various feedback loops that alter the biochemical state of the cell on the time-scale of the response, and difficulties in imaging a cell for several hours because of photobleaching, phototoxicity and encountering cell division events, in TNF-NF-κB signaling it is a very difficult experiment to even stimulate the same cell twice (see Section 4.2.4), let alone with a range of increasing doses of TNF covering the full range of responsiveness. For this reason this system is an excellent candidate for simulation-based calculations of single-cell channel capacity.

To calculate the channel capacity for a cell, since we are assuming that the cell does not have a perfect one-to-one relationship between dose and response in which case the channel capacity for any responding cell would be infinite, we need to assume a noise model. Since we do not know the correct biological noise model, we test four alternatives:

1. Model 1: Gaussian noise centered on the data with standard deviation = 1% of the data value.

2. Model 2: Gaussian noise centered on the data with standard deviation = 10% of the data value.

3. Model 3: Gaussian noise centered on the data with standard deviation = 0.01

4. Model 4: Gaussian noise centered on the data with standard deviation = 0.1

We combine the dose response curves in Figure 5.4C, generated from the 1-NN dataset in Figure 5.4B, with each of the four noise models described above and calculate the channel capacity for each cell under each noise model. For each noise model, we find variation in the channel capacity calculated across single cells (Figure 5.5A), suggesting that cells can have different capabilities of transmitting TNF dosage information. On relating the protein abundances in each cell to the channel capacity calculated with the second noise model, we find that given this model and shared parameter set, cells with a greater abundance of total NF-κB have higher capacities to transmit TNF dosage information (Figure 5.5B). Focusing on the dose response behavior of the three cells with the maximum, minimum, and median channel capacity (Figure 5.5C–F), we see that the cell with a higher saturating response has a higher capacity. Overall, while there may be many factors and mechanisms that govern a cell's capacity to transmit information about its environment, here we find that one possible mechanism in our model, that is consistent with our calibration data, is for a cell to have a higher total abundance of NF-κB and a higher saturating response to TNF.

5.3.3.1 Summary of limitations and avenues for future investigation   Here, we have demonstrated an example of single cell channel capacity analyses using simulated data. The following are avenues that merit future investigation:

- First, the channel capacity calculation has several hyperparameters, including the choice of 'K' in the K-nearest neighbor density estimation used to estimate conditional probabilities of each response (we use K=20), the sample size, that is the number of points sampled for each simulated cell assuming a noise model (we use a sample size of 500), and the number of simulated doses tested (we tested 52 doses). Future work will investigate the impact of hyperparameter variation.
- Second, we have demonstrated an example of channel capacity analysis using a simulated dataset corresponding to one shared parameter set from the fit to $dataset_1$. In order to better draw biological conclusions, future work will also investigate the sensitivity of the results to (1) the shared parameters (2) the initial set of representative trajectories, for instance $dataset_1$ versus $dataset_2$, and (3) changes to model calibration procedures (see Sections 5.3.1.1 and 5.3.2.1) to ensure more biophysical predictions and reduced

sensitivity to the initial representative trajectories.
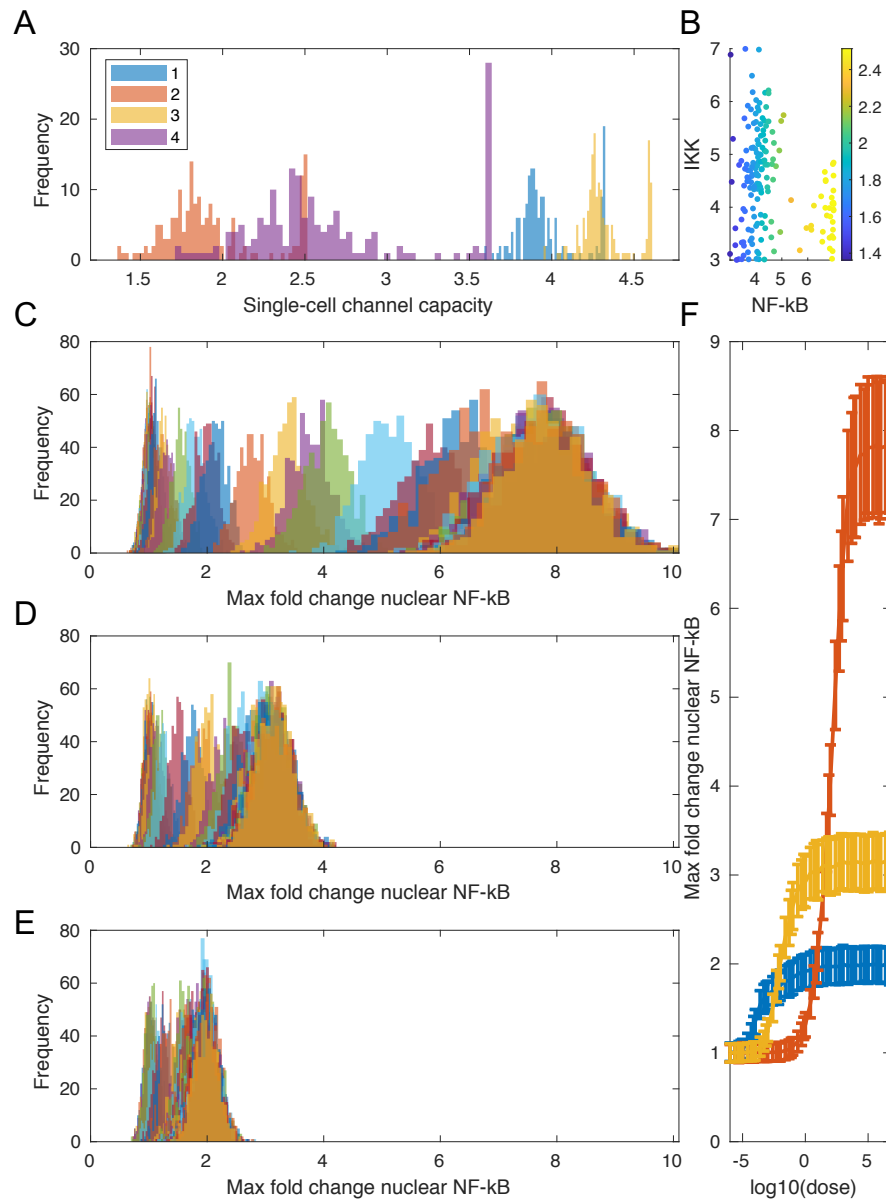
Figure 5.5: Single-cell channel capacities calculated from simulated data A) Single-cell channel capacity distributions with four different error models – Model 1, Gaussian noise centered on the data with standard deviation = 1% of the data value (blue), Model 2, Gaussian noise centered on the data with standard deviation = 10% of the data value (orange), Model 3, Gaussian noise centered on the data with standard deviation = 0.01 (yellow), Model 4, Gaussian noise centered on the data with standard deviation = 0.1 (purple). B) Scatter plot showing NF-κB and IKK protein abundances for the simulated single cells in Figure 5.4B. The color bar indicates the channel capacity of each cell, calculated using Model 2 as the noise model. C) Dose-response distributions for the simulated cell with the highest channel capacity, D) median channel capacity, and E) smallest channel capacity, while assuming Model 2 as the noise model. For panels C, D and E, the different colored distributions correspond to different doses, and each distribution has 500 points obtained by adding 10% Gaussian noise to the single-cell response at any given dose. F) Dose response curves for the cells with the highest (orange), smallest (blue) and median channel capacity (yellow). Plots show the mean and standard deviation of the corresponding responses in panels C, D and E.

5.4   Discussion

In this work, building on previous approaches [59], we have fit a mechanistic ODE model of NF-κB signaling to a large dataset of single-cell NF-κB responses to pulsatile TNF. By assuming that most parameters are shared between cells and that heterogeneity is driven by abundances of key signaling proteins NF-κB and IKK, we were able to use sampling approaches like PT and LHS to calibrate the model. Mechanistic models in systems biology often have many free parameters that are calibrated to a limited amount of data, resulting in non-identifiability, where many parameter sets fit the data equally well. By using Bayesian parameter estimation, we were able to find ensembles of shared parameter sets that can fit a representative subset of the data, all of which possess the property that varying NF-κB and IκB abundances produces simulated NF-κB response datasets that are in good agreement with the data. A caveat currently, however, is that the model predictions for other biochemical species are not within a biophysically feasible realm. Potential strategies to correct this in the future include (1) incorporating more experimental data into the fit, for instance fitting measurements from multiple biochemical species in the same cell, or (2) incorporating qualitative constraints on non-fitted species, based on prior knowledge, to ensure their biophysical behavior [12, 185].

While our hypothesis-driven approach of allowing NF-κB and IKK to vary between cells was informed by the literature, where there are several reported cases of protein variability being a dominant source of heterogeneity in cell signaling systems [50, 60–62], and we were successful in fitting the data, it is possible that variation in other small subsets of parameters is equally likely in fitting the data, and perhaps there is even one key parameter that is sufficient. Steiert et al. [111], used Lasso regularization with maximum likelihood estimation approaches to determine the minimal number of parameters needed to explain the differences in responses between cell types. Similarly, in the future, we would like to incorporate Lasso regularization with parallel tempering, to find alternate minimal subsets of parameters for which variation is sufficient to fit a large number of signaling responses that differ between genetically identical cells.

In addition to fitting the calibration data, our model of heterogeneity had reasonably

good predictions for the average response at other experimentally tested doses. However single-cell dose-response predictions were not constrained across shared parameter sets, and depending on the parameter set, several single-cell dose response models were consistent with the data. Since experimentally we do not have data for a single-cell tracked over several doses, the ability of the model to predict single-cell dose response behavior remains to be evaluated. Finally, by assuming various single-cell noise models, we presented an example of a simulation-based analysis of single-cell channel capacity. For the model simulated with a particular set of shared parameters, we found that it was possible for the capacity to transmit TNF dosage information to be variable across cells. Given our model assumptions, we found a potential determining factor to be the total abundance of nuclear NF-κB and the saturating NF-κB response, with cells that had a higher saturating response, possessing a greater ability to discriminate between TNF doses. Note that here, the saturating response is defined as the maximum fold change of nuclear NF-κB at the saturating dose.

While our models are certainly simplified, and need more investigation to ensure fully biophysical behavior, our work demonstrates several use-cases for computational models of heterogeneity, and supports the idea that genetically identical single cells can have varying dose response properties and information transmission capacities that should be accounted for while designing experiments to rationally manipulate single-cell behavior.

6.0   Conclusion

Mathematical and computational models have become an invaluable complement to experiments for understanding properties of cell signaling systems [186]. It is increasingly being recognized that reductionist approaches to biology, that characterize individual system components, such as genes, proteins or other biomolecules, statically and in isolation, while useful are not sufficient [187–190]. Instead, a comprehensive understanding of biological systems and diseases also require systems biology approaches, that aim to understand how dynamic interactions between system components drive emergent behavior [187–190]. Mathematical and computational modeling present a natural framework to understand how complex nonlinear interactions between system components can give rise to, sometimes non-intuitive, emergent behavior [15, 16, 20, 93, 186].

In Chapter 2 we addressed a crucial technical challenge in connecting models with experimental data, namely parameter estimation. Parameter estimation encompasses computational methods to determine numerical values of model parameters, which cannot always be measured experimentally [168], for which the model output matches the experimental data [32]. We systematically evaluated Parallel Tempering, an MCMC method that has previously gained widespread success in the field of molecular dynamics simulations [30], for its performance in estimating parameters of ODE models of signaling. We showed that on a series of six signaling models of increasing complexity, PT consistently outperformed the Metropolis-Hastings algorithm, a commonly used parameter estimation algorithm, on a variety of performance metrics, such as convergence time, sampling efficiency, and goodness of fit given a fixed computational resource budget. Overall, PT was found to be a reliable method for estimation of parameters and parameter uncertainty in models of cell signaling with up to dozens of parameters.

Models in the field of systems biology exhibit a wide range in complexity [191]. Advances in biochemical modeling and simulation techniques, such as the development of rule-based modeling languages, simulators, and analysis platforms [71, 192–194], have enabled a push towards detailed models that capture many known molecular interactions of biochemical sys-

tems [16, 38]. Detailed models provide more accurate representations of biological systems and enable quantitative predictions that are directly comparable to experimental observables [16, 191, 195]. At the same time it is also recognized that simpler coarse grained models, that capture essential behaviors of the system under specific conditions, have several advantages. Some of these advantages include having fewer parameters, being easier to read and interpret, and enabling identification of key mechanisms that drive emergent behavior [191], for example, incoherent feedforward loops that enable fold change detection [33], or negative feedback loops that drive oscillations [191].

Towards the goal of bridging the gap between detailed and simple models, in Chapter 3 we developed a Bayesian model reduction framework that identifies minimal subsets of reactions, within complex biochemical models, that are sufficient for particular datasets. By combining the PT algorithm that we evaluated in Chapter 2, with Lasso regularization [39], we were able to simultaneously estimate numerical values for model parameters, as well as determine reduced sets of parameters, and by extension reduced sets of reactions, that were sufficient for the data. We demonstrated the approach on synthetic data, and developed an extension of the approach that finds minimal subsets of reaction modules, instead of individual reactions, which we demonstrated on the NF-κB signaling pathway by assessing differences in the requirement of key feedback elements for cell responses to either short time-scale pulses of TNF or longer time-scale continuous TNF stimulation. Overall, the question of whether the guiding principle of Occam's razor, favoring simple model mechanisms, is relevant while studying complex biological systems, is an ongoing philosophical debate [196–198]. However, having tools to extract simple explanatory mechanisms from detailed models may provide useful starting points for target mechanisms for perturbations and enable us to easily tailor our desired model complexity to the question of interest and the resolution of the available data.

In Chapter 4, we evaluated the capabilities of the NF-κB signaling pathway to grade its response to TNF dose, and discriminate between doses in the presence of biochemical noise. With statistical tools from information theory, we used single-cell response data from hundreds of cells exposed to different concentrations and durations of TNF, and calculated the channel capacity of the pathway to be $\sim$1.7 bits after filtering for non-responding cells. The

channel capacity value of greater than 1 bit supports the hypothesis that cells can discriminate multiple doses. Repeat stimulation experiments that use a short, minimally invasive, TNF pulse to measure the cell's responsiveness to TNF, effectively a proxy for the biochemical state of a cell, followed by a test pulse, show that knowing the cell's responsiveness to TNF increases our ability to predict its response to the test pulse. Two interesting conclusions that follow from this set of analyses are that (1) genetically identical subpopulations of cells can have different information transmission capacities (in our case the non-responders do not change their response with TNF dose, and would have an information transmission capacity close to 0, while the responding subpopulation can grade its response to dose and has a higher information transmission capacity); and (2) variability in TNF-NF-κB signaling is dominated by extrinsic noise, that is variation in the biochemical state of a cell, instead of intrinsic noise, that is stochastic fluctuations in transcriptional processes. This is why knowledge of the cell-state, measured by response to a short pulse of TNF improved the predictability of the NF-κB response.

In Chapter 5, to complement our statistical analysis of noisy TNF-NF-κB signaling data, we followed a mechanistic approach to understand heterogeneity in TNF-NF-κB signaling. Building on previous approaches [59], we calibrated a mechanistic ODE model of NF-κB signaling to heterogeneous data from live-cell imaging, by considering variation in the abundances of NF-κB and IKK. Using the calibrated model, we showed examples of simulations probing dose-response capabilities at a single-cell level. It is now widely accepted, and accounted for while designing experiments and drug studies, that dose-response properties, such as the range of responsiveness, can vary widely across cell-lines [46]. However, given the heterogeneity that we observe within a genetically identical population of cells, is it possible that these properties also vary considerably within a cell line? This is a key question in shifting from designing population-level experiments to single-cell experiments that effectively probe single-cell signaling mechanisms and manipulate single-cell behavior. However, experimental studies of single-cell dose response behavior, where the same cell is in turn exposed to a range of increasing stimulus doses [184], are currently not possible in signaling systems with long time-scale responses like the TNF-NF-κB signaling pathway, where the response is on the order of hours. The presence of several long time-scale feedback elements

means that once stimulated, it is difficult to return the cell to its original biochemical state for a repeat stimulation. Additional technical limitations include phototoxicity effects with longer imaging durations, and biological limitations include the increased probability that the observed cell will divide in a time scale shorter than what is needed to repeatedly observe the response to a range of stimuli. A well-calibrated computational model of NF-κB signaling heterogeneity gives us unique opportunities to probe single-cell dose response properties in silico. Using our mechanistic model, we presented an example of this kind of analysis, and demonstrated how variation in protein abundances, in a model of NF-κB signaling, can lead to considerable differences in saturating NF-κB responses to TNF, and differences in the capacities of simulated single cells to transmit information about TNF dose.

Biological systems are complex, and often we need a plethora of complementary approaches, both experimental and computational, to investigate different aspects of the same system, ultimately increasing our understanding of the system as a whole. Overall, in this thesis we have employed a variety of complementary computational approaches and have addressed several challenges related to modeling cell signaling systems. Using the NF-κB signaling pathway as a model system, we demonstrated how Bayesian methods for problems like parameter estimation and regularization can be useful for improving our understanding of key regulatory mechanisms in single cells. We showed how information theoretic and statistical modeling tools, combined with data from cutting edge experimental techniques, can address questions regarding the capabilities of noisy signaling pathways in detecting environmental cues. Finally, we complemented our statistical analysis of noisy signaling data, with mechanistic approaches to model heterogeneity in NF-κB signaling, and demonstrated the use of simulation-based analyses to investigate information transmission capacities of single cells.

Appendix A

Supplementary material for Chapter 3: Parallel Tempering with Lasso for model reduction in systems biology

Figure A1: 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 true parameter values. Each colored line represents a noisy trajectory obtained by adding Gaussian noise (mean = 0, standard deviation = 30% of the true data value) 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 (from 4,000 parameter samples) with μ and b, (red points show the mean, and black lines show mean $\pm$ standard deviation). The hyperparameters selected ($\mu = -10$, $b = 1$) provide the most regularization while not substantially increasing the negative log likelihood. C) Comparison of the log likelihood distributions (from 4,000 parameter 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. Boxes show data in the 25th–75th percentile and the circles show the mean. D) Example of PTLasso fits (from 4,000 parameter 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 (from 400,000 parameter samples). Since the regularization strength was too high, none of the parameters deviated from the prior. *`http://alex.bikfalvi.com/research/advanced_matlab_boxplot/`
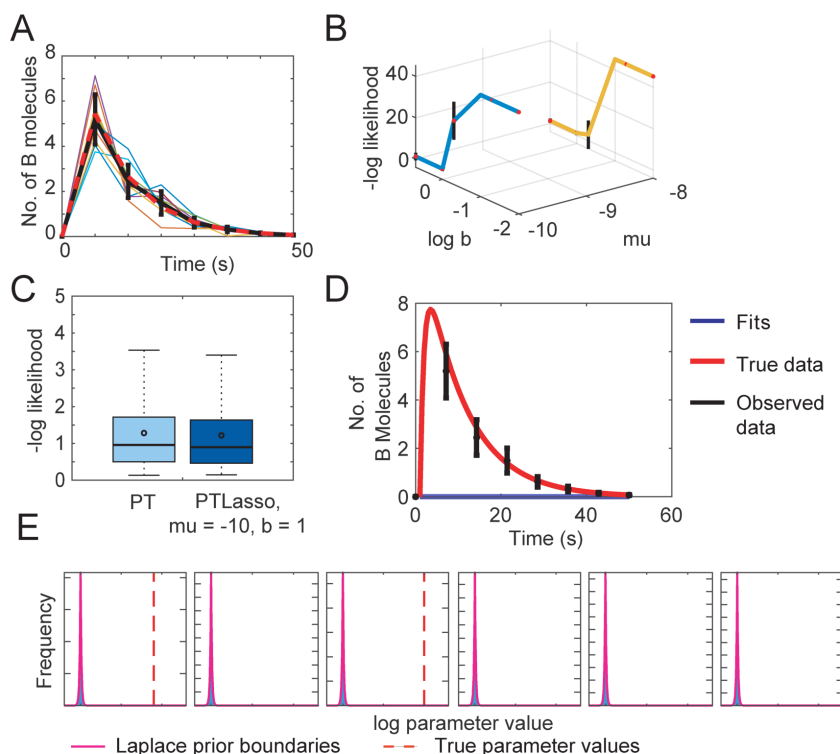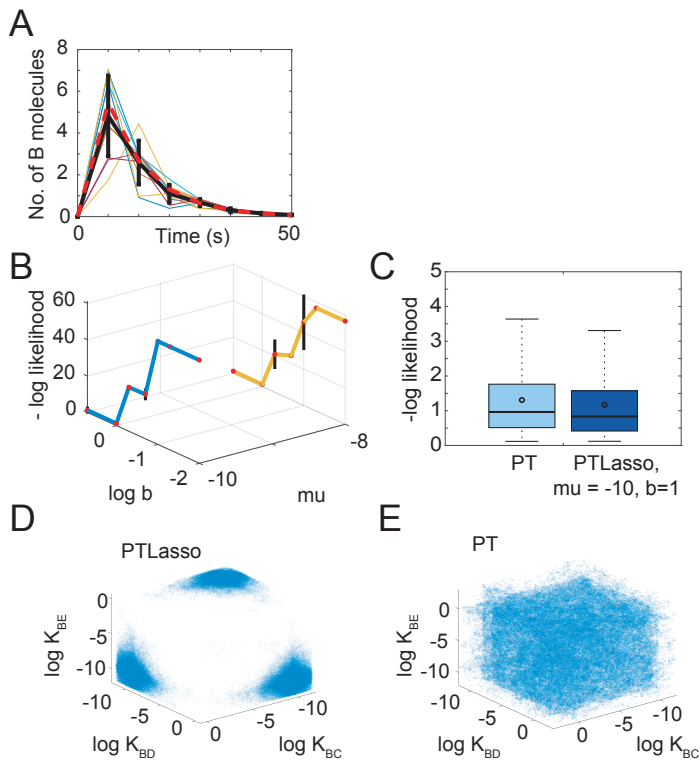
Figure A2: Hyperparameter tuning for PTLasso with a fully connected 5-node graph. A) Data generated for fitting. Red dashed lines show the model simulation at 8 time points with the true parameter values. Each colored line represents a noisy trajectory obtained by adding Gaussian noise (mean = 0, standard deviation = 30% of the true data value) 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 (from 7,000 parameter samples, red points show the mean, and black lines show mean $\pm$ standard deviation). 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 (from 7,000 parameter samples) obtained with PT and PTLasso for the chosen values of hyperparameters. Box plots are obtained using a third party MATLAB library, aboxplot*, with outliers not shown. Boxes show data in the 25th–75th percentile and the circles show the mean. D). Parameter covariation of the three selected parameters with PTLasso and E) with PT shown as a 3D scatter plot with transparent points (from 700,000 parameter samples). *http://alex.bikfalvi.com/research/advanced_matlab_boxplot/
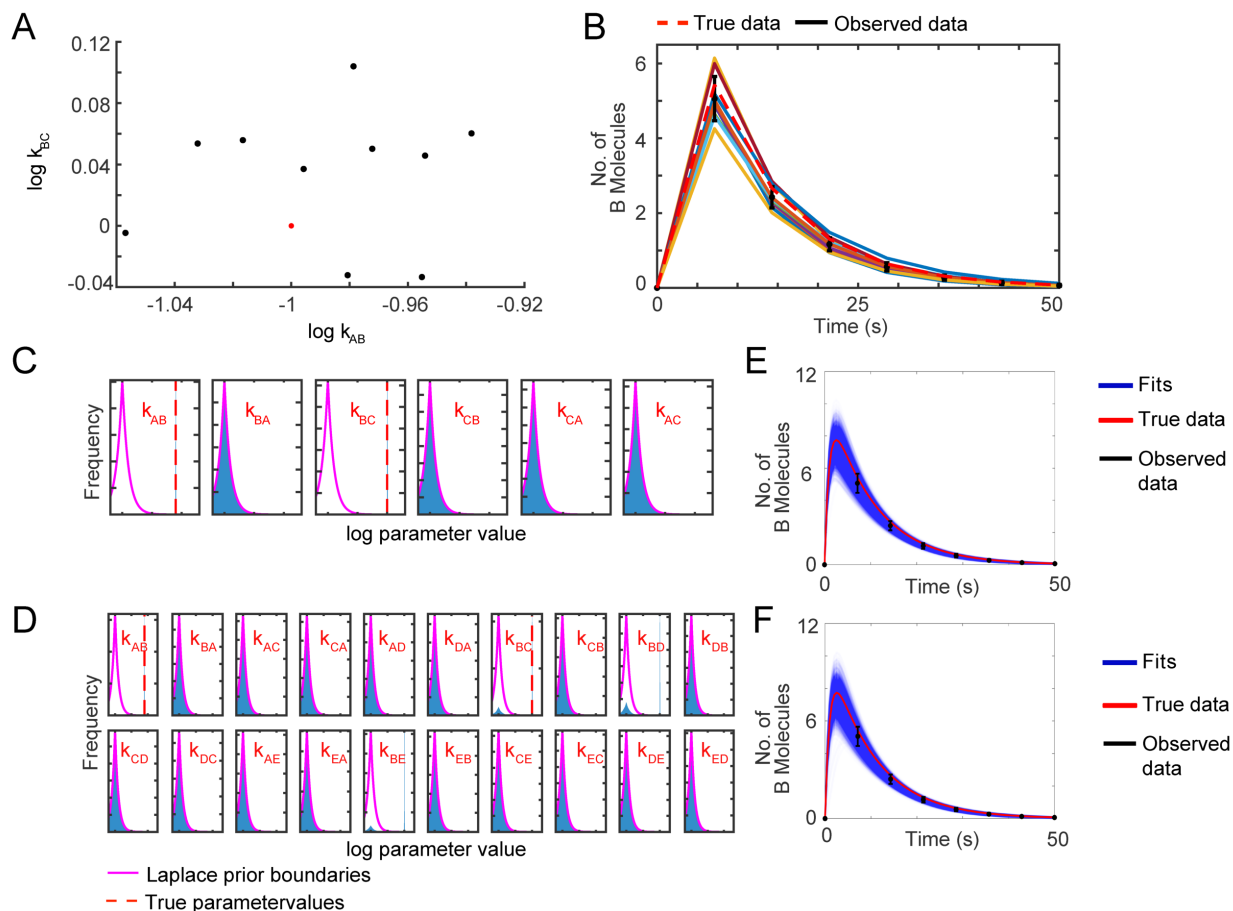
Figure A3: Model reduction using PTLasso with fully connected 3-node and 5-node graphs when the observed data is generated from noisy parameters. A) Noisy parameter values (black) used to generate the observed data. The log true parameters (red) of the known model were perturbed 10 times with Gaussian noise (mean = 0, standard deviation = 0.05). B) Colored lines show model outputs for each of the 10 noisy parameter sets. The black error bars shows the mean and standard deviation of the colored lines and is the observed data for fitting. Red dashed line shows the model simulation at 8 time points with the true parameter values. C) Frequency histograms showing probability distributions of the parameters (from 800,000 parameter samples) for PTLasso fits of a fully connected three node graph and D) fully connected five node graph. The range of log parameter values on each x-axis is −12 to 3, which covers the full range over which parameters were allowed to vary. The y-axis of each panel is scaled to the maximum value of the corresponding distribution to emphasize differences in shape. The pink lines show the boundaries of the Laplace prior with $\mu = -10$, $b = 1$, and the dashed red lines in panels for k_{AB} and k_{BC} show the true parameter values. A parameter distribution confined within the Laplace prior boundaries indicates that the parameter is extraneous. E) PTLasso fits to the data for a fully connected three node graph and F) five node graph. Transparent blue lines show ensemble fits (from 8,000 parameter samples, 100 time points per trajectory), red line shows the true data (100 time points), and the black error bars show the mean ± standard deviation of the observed data (8 time points).
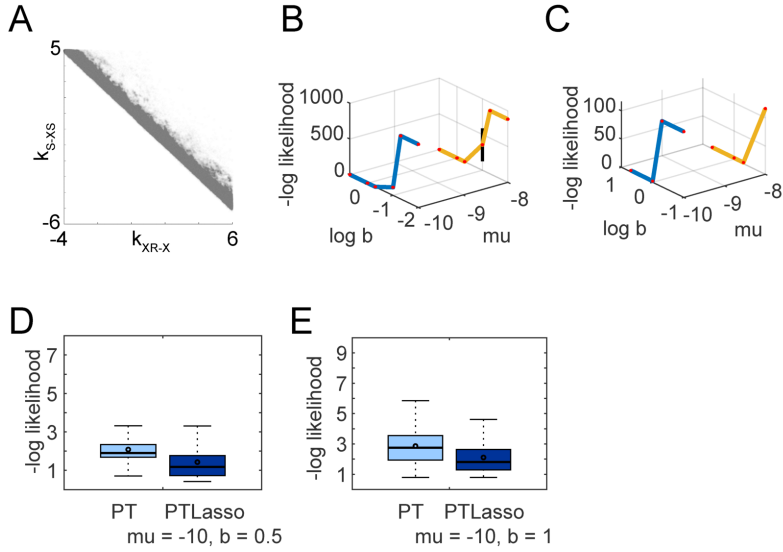
Figure A4: Hyperparameter tuning for PTLasso with dose-response motifs inferred from a prior network. A). Linear correlation of non identifiable parameters in the reduced perfectly adapting model shown as a scatter plot (axes show log parameter values). B) Hyperparameter tuning plot for the linear dose 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 (from 400 parameter samples for the linear dose response model and from 800 parameter samples for the perfectly adapting dose response model. Red points show the mean, and black lines show mean $\pm$ standard deviation). 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 (from 400 parameter samples) and E) the perfectly adapting dose response model (from 800 parameter samples). Box plots are obtained using a third party MATLAB library, aboxplot*, with outliers not shown. Boxes show data in the 25th–75th percentile and the circles show the mean. *http://alex.bikfalvi.com/research/advanced_matlab_boxplot/
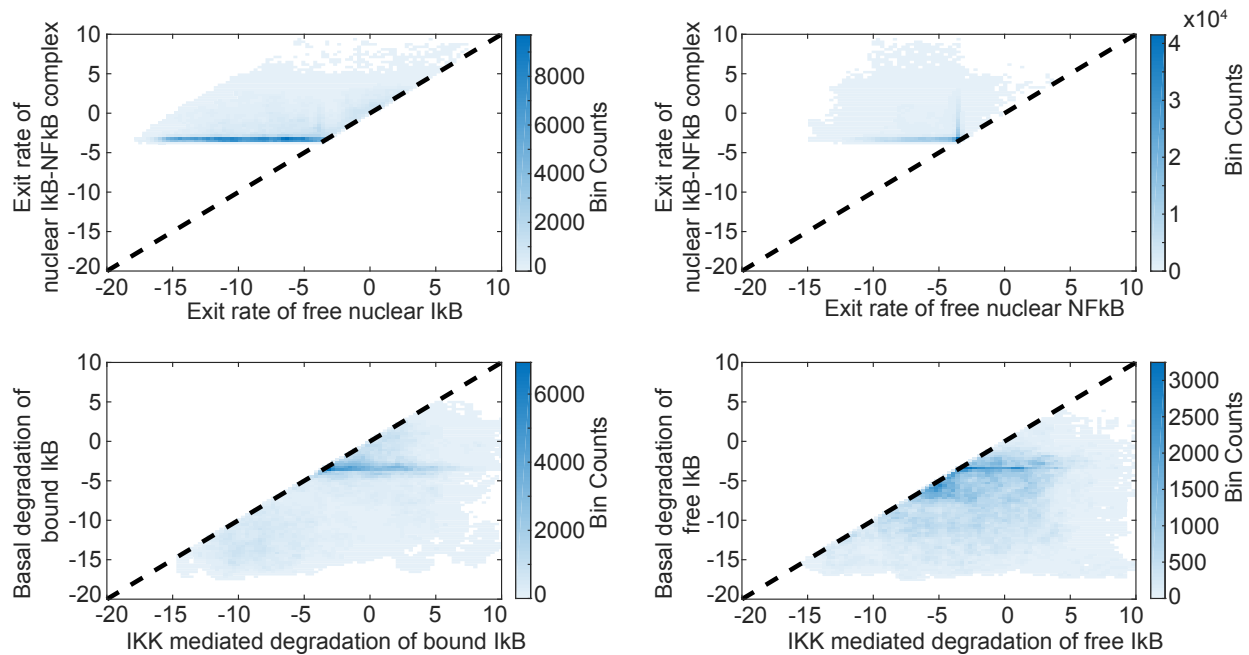
Figure A5: Hard constraints on parameter covariation in NF-κB signaling. Binned scatter plots (MATLAB function binscatter with 940,000 parameter samples from a PTLasso fit for an NF-κB response to pulsatile TNF stimulation) show the joint distributions for the pairs of parameters for which covariance was constrained during fitting.
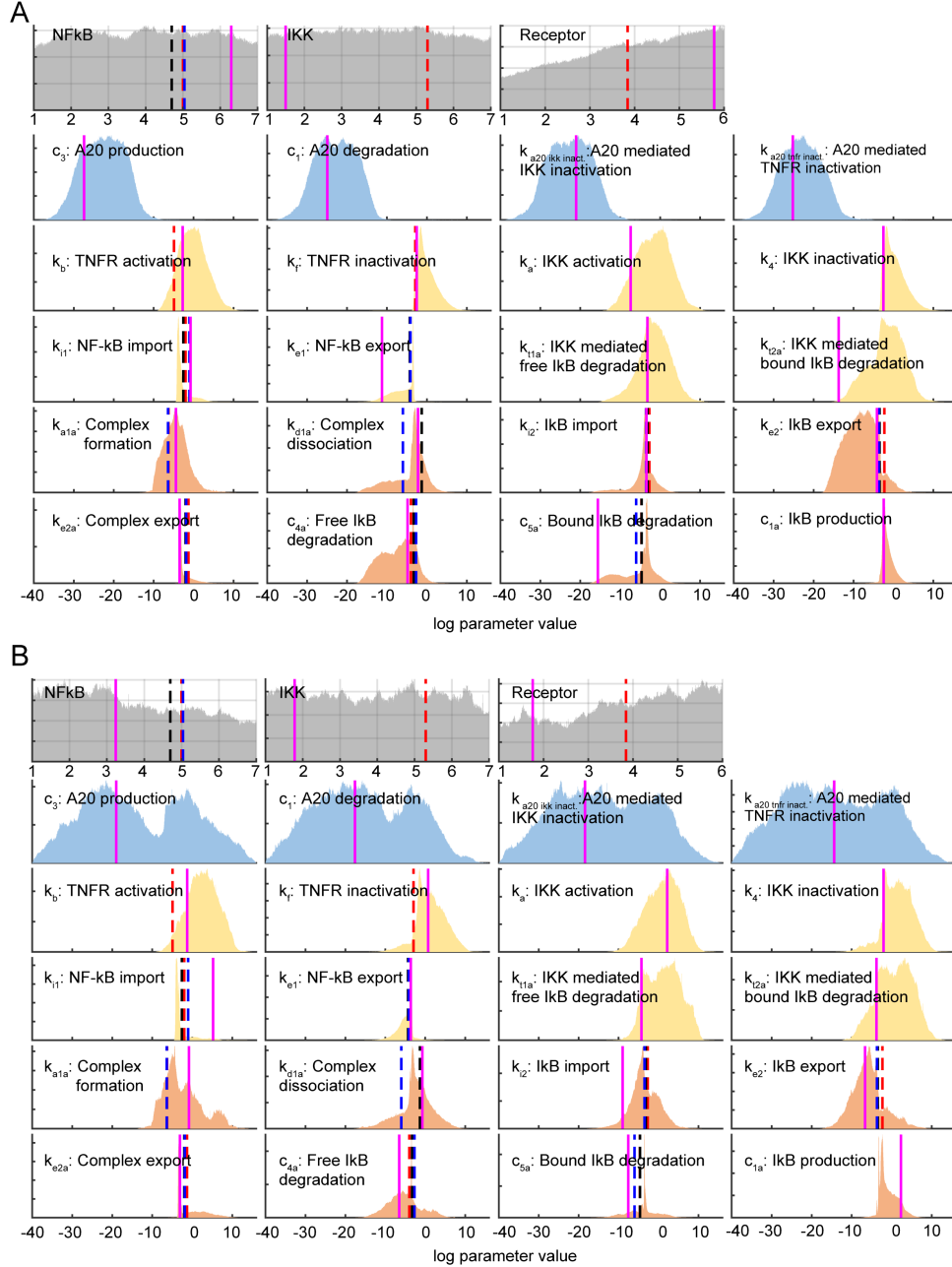
Figure A6: Posterior probability distributions of model parameters shown with the corresponding published values for a representative NF-κB response to pulsatile TNF stimulation A) with PTLasso and B) with PT. Distributions of total protein abundance parameters and rate constant parameters (from 5,640,000 parameter samples) from the A20 module (blue), Activation module (yellow) and IκB module (orange). All parameters are in logscale. Total 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. The dashed lines correspond to published values of parameters – Pekalski et al. [15] (red), Lee et al. [56] (black), and Kearns et al. [121] (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. [56], $1\mu M$ of NF-κB = 50,000 molecules/cell and applied this to other species in the models. In the parameter labels "import" refers to translocation from the cytoplasm into the nucleus and "export" is the reverse. "Complex" refers to the NF-κB-IκB complex. The y-axis for each panel is scaled from 0 to the maximum value of the distribution to emphasize differences in the shapes of the distributions.
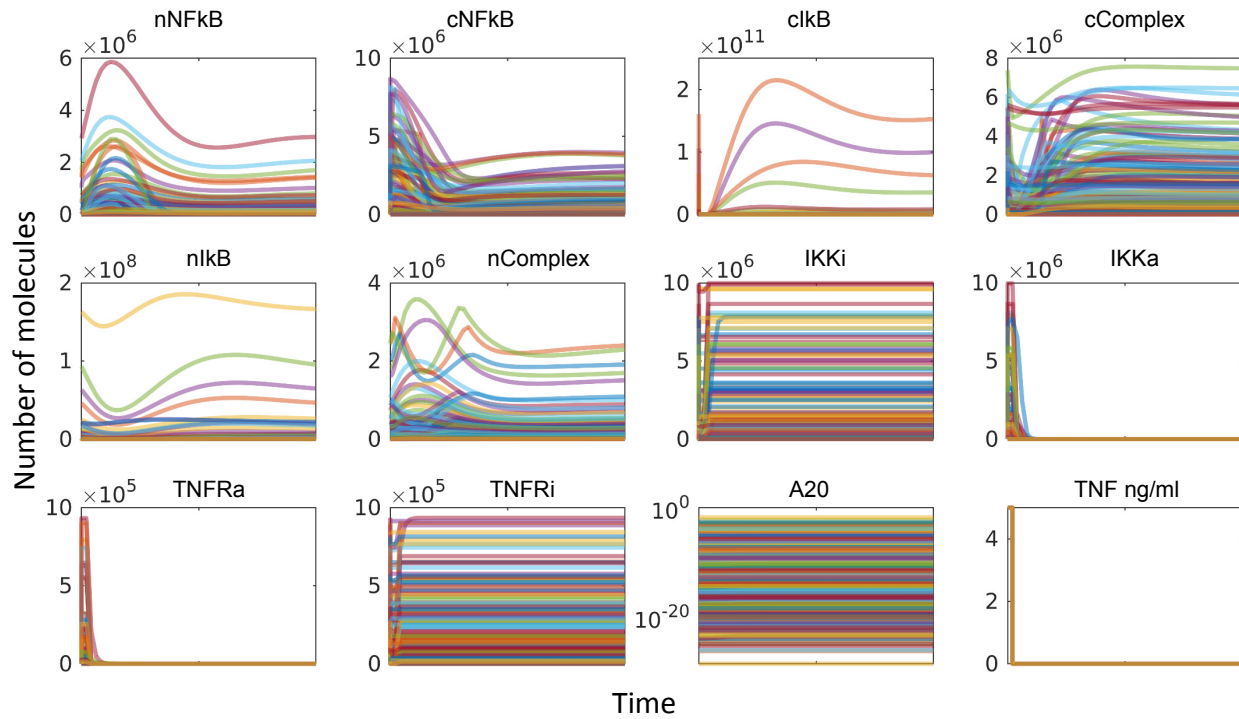
Figure A7: NF-κB signaling model predictions. Model predictions for non-fitted variables for a representative NF-κB response to pulsatile TNF stimulation (from 500 parameter samples from one of the PTLasso repeats). Suffix "a" and "i" refer to active and inactive versions of a species respectively. Prefix "c" and "n" refer to cytoplasmic and nuclear versions of a species respectively. "Complex" refers to the NF-κB-IκB complex. Time courses are shown for 4 hours after the initial 5ng/ml TNF stimulation. The TNF concentration is set to 0 at the 5 minute time point.

Table A1: Hard constraints in the NF-κB signaling fits.

Index	Constraint description	Constraint
1	Exit rate of NF-κB-IκB from the nucleus is greater than free NF-κB	ke2a > ke1
2	Exit rate of NF-κB-IκB from the nucleus is greater than free IκB	ke2a > ke2
3	IKK mediated degradation of free cytosolic IκB is greater than the basal rate	kt1a > c4a
4	IKK mediated degradation of NF-κB bound cytosolic IκB is greater than the basal rate	kt2a > c5a
5	At equilibrium there is some NF-κB in the nucleus	$\dfrac{\text{nNF}\kappa\text{B}_{eq} + \text{nNF}\kappa\text{B} - \text{I}\kappa\text{B}_{eq}}{\text{TotalNF}\kappa\text{B}} \in [0.05,0.5]$
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 A2: Maximum PSRF across all model parameters for each example shown up to 4 significant digits. Parameter distributions are constructed from the lowest temperature chain.

Model name, No. of swaps	PTLasso	PT
3-node graph, 400,000	1.000	1.002
5-node graph, 700,000	1.025	1.003
Linear Dose-Response, 400,000	1.001	1.000
Perfectly adapting Dose-Response, 800,000	1.015	1.004
NF-κB signaling (pulse stimulus):		
Trajectory 1, 5,640,000	1.025	1.024
Trajectory 2, 5,640,000	1.027	1.057
Trajectory 3, 5,640,000	1.027	1.006
NF-κB signaling (continuous stimulus) 3,200,000	1.05	N/A

Table A3: MPSRF values for parameter distributions from each example shown up to 4 significant digits. Parameter distributions are constructed from the lowest temperature chain.

Model name, No. of swaps	PTLasso	PT
3-node graph, 400,000	1.001	1.002
5-node graph, 700,000	1.011	1.007
Linear Dose-Response, 400,000	1.000	1.001
Perfectly adapting Dose-Response, 800,000	1.012	1.006
NF-κB signaling (pulse stimulus)		
Trajectory 1, 5,640,000	1.074	1.094
Trajectory 2, 5,640,000	1.118	1.200
Trajectory 3, 5,640,000	1.129	1.014
NF-κB signaling (continuous stimulus), 3,200,000	1.069	N/A

Table A4: PSRF to show convergence of energy distributions when combining PT or PTLasso chains for the NF-κB signaling fit with pulsatile TNF stimulation. M is the number of independent chains that are combined for each group and N is the total number of swaps. The length of each chain is N/M. Energy distributions are constructed from the lowest temperature chain.

Trajectory index	PTLasso group 1, M=6, N=5,640,000	PTLasso group 2, M=6, N=5,640,000	PT group 1, M=6, N=5,640,000	PT group 2, M=6, N=5,640,000
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 A5: PSRF to show convergence of energy distributions when combining PTLasso chains for the NF-κB signaling fit with continuous TNF stimulation. M is the number of independent chains that are combined for each group and N is the total number of swaps. The length of each chain is N/M. Energy distributions are constructed from the lowest temperature chain.

PTLasso group 1, M=8, N=3,200,000	PTLasso group 2, M=8, N=3,200,000
1.113	1.063

Table A6: Step acceptance rates for the lowest temperature chain for each example.

Model name, No. of swaps	PTLasso (repeat 1, 2)	PT (repeat 1, 2)
3-node graph, 400,000	0.24, 0.25	0.24, 0.25
5-node graph, 700,000	0.24, 0.23	0.24, 0.26
Linear Dose-Response, 400,000	0.24, 0.22	0.24, 0.24
Perfectly adapting Dose-Response, 800,000	0.24, 0.24	0.25, 0.25
NF-κB signaling (pulse stimulus), 5,640,000		
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
NF-κB signaling (continuous stimulus), 3,200,000		
	0.28, 0.27	N/A

Table A7: Swap acceptance rates for the two lowest temperature chains for each example.

Model name, No. of swaps	PTLasso (repeat 1, 2)	PT (repeat 1, 2)
3-node graph, 400,000	0.24, 0.33	0.25, 0.35
5-node graph, 700,000	0.33, 0.47	0.32, 0.26
Linear Dose-Response, 400,000	0.26, 0.97	0.24, 0.24
Perfectly adapting Dose-Response, 800,000	0.17, 0.37	0.25, 0.26
NF-κB signaling (pulse), 5,640,000		
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
NF-κB signaling (continuous), 3,200,000	0.22, 0.22	N/A

Table A8: Reactions in fully connected three node network. The "parameters" column specifies the forward and reverse rate constant pair. True parameters are shown in red. All the reactions follow mass action kinetics. First order reaction rate constants are in units of s^{-1}. Second order reaction rate constants are in units of molecule$^{-1}s^{-1}$.

Reaction	Parameters
A $\longleftrightarrow$ B	k_{AB}, k_{BA}
B $\longleftrightarrow$ C	k_{BC}, k_{CB}
C $\longleftrightarrow$ A	k_{CA}, k_{AC}

Table A9: Reactions in fully connected five node network. The "parameters" column specifies the forward and reverse rate constant pair. True parameters are shown in red, and parameters of the inferred alternate reduced models are shown in blue. All the reactions follow mass action kinetics. First order reaction rate constants are in units of s^{-1}. Second order reaction rate constants are in units of molecule$^{-1}s^{-1}$.

Reaction	Parameters	Reaction	Parameters
A $\longleftrightarrow$ B	k_{AB}, k_{BA}	C $\longleftrightarrow$ D	k_{CD}, k_{DC}
A $\longleftrightarrow$ C	k_{AC}, k_{CA}	A $\longleftrightarrow$ E	k_{AE}, k_{EA}
A $\longleftrightarrow$ D	k_{AD}, k_{DA}	B $\longleftrightarrow$ E	k_{BE}, k_{EB}
B $\longleftrightarrow$ C	k_{BC}, k_{CB}	C $\longleftrightarrow$ E	k_{CE}, k_{EC}
B $\longleftrightarrow$ D	k_{BD}, k_{DB}	D $\longleftrightarrow$ E	k_{DE}, k_{ED}

Table A10: Prior network comprising linear dose-response model reactions and adaptive dose-response model reactions. The "parameters" column specifies the forward and reverse rate constant pair. All the reactions follow mass action kinetics. First order reaction rate constants are in units of s^{-1}. Second order reaction rate constants are in units of $\text{molecule}^{-1}\text{s}^{-1}$.

Reaction	Parameter(s)
S $\longrightarrow$ S + R	k_{S-RS}
X + R $\longrightarrow$ X	k_{XR-X}
S $\longrightarrow$ S + X	k_{S-XS}
X $\longrightarrow$ 0	k_{X-0}
R $\longleftrightarrow$ 0	k_{R-0}, k_{0-R}

Table A11: Reactions in NF-κB signaling model. The "parameters" column specifies the forward and reverse rate constant pair. All the reactions follow mass action kinetics. First order reaction rate constants are in units of s^{-1}. Second order reaction rate constants are in units of molecule^{-1}s^{-1} except k_b which is in units of (ng/ml)$^{-1}$s^{-1}. * 3.3 is a volume correction factor [56, 145].

Reaction	Parameter(s)	Module
TNFR activation		
TNFR(inactive) + TNF $\longrightarrow$ TNFR(active) + TNF	k_b	Activation
TNFR inactivation		
TNFR(active) $\longrightarrow$ TNFR(inactive)	k_f	Activation
IKK activation		
IKK(inactive) + TNFR(active) $\longrightarrow$ IKK(active) + TNFR(active)	k_a	Activation
IKK inactivation		
IKK(active) $\longrightarrow$ IKK(inactive)	k_4	Activation
Reversible cytoplasmic NF-κB-IκB complex formation		
NF-κB (cytoplamic) + IκB (cytoplasmic) $\longleftrightarrow$ NF-κB-IκB (cytoplasmic)	k_{a1a}, k_{d1a}	IκB
Reversible nuclear NF-κB-IκB complex formation		
NF-κB (nuclear) + IκB (nuclear) $\longleftrightarrow$ NF-κB-IκB (nuclear)	3.3* $\times k_{a1a}$, k_{d1a}	IκB
IKK mediated cytoplasmic free IκB degradation		
IKK(active) + IκB (cytoplasmic) $\longrightarrow$ IKK(active)	k_{t1a}	Activation
IKK mediated cytoplasmic bound IκB degradation		
IKK(active) + NF-κB-IκB (cytoplasmic) $\longrightarrow$ IKK(active) + NF-κB (cytoplasmic)	k_{t2a}	Activation
Basal cytoplasmic free IκB degradation		
IκB (cytoplasmic) $\longrightarrow$ 0	c_{4a}	IκB
Basal cytoplasmic bound IκB degradation		
NF-κB-IκB (cytoplasmic) $\longrightarrow$ NF-κB (cytoplasmic)	c_{5a}	IκB
NF-κB shuttling		
NF-κB (cytoplasmic) $\longleftrightarrow$ NF-κB (nuclear)	k_{i1}, k_{e1}	Activation
IκB shuttling		
IκB (cytoplasmic) $\longleftrightarrow$ IκB (nuclear)	k_{i2}, k_{e2}	IκB
Linear production of IκB		
NF-κB (nuclear) $\longrightarrow$ NF-κB (nuclear) + IκB (cytoplasmic)	c_{1a}	IκB
Export of NF-κB-IκB complex from the nucleus		
NF-κB-IκB (nucleus) $\longrightarrow$ NF-κB-IκB (cytoplasmic)	k_{e2a}	IκB
Linear production of A20		
NF-κB (nuclear) $\longrightarrow$ NF-κB (nuclear) + A20	c_1	A20
A20 mediated inactivation of IKK		
A20 + IKK(active) $\longrightarrow$ A20 + IKK(inactive)	$k_{\text{a20-ikk-inact}}$	A20
A20 mediated inactivation of TNFR		
A20 + TNFR(active) $\longrightarrow$ A20 + TNFR(inactive)	$k_{\text{a20-tnfr-inact}}$	A20
A20 degradation		
A20 $\longrightarrow$ 0	c_3	A20

Appendix B

Supplementary material for Chapter 4: NF-κB Dynamics Discriminate between TNF Doses in Single Cells

Table B1: Related to main Figure 4.2. Summary of cell numbers for fixed cell experiments. For each condition cells were exposed continuously or to a pulse of cytokine, and fixed after 0, 10, 30, 60, 90, 120, 180, 240, or 360 minutes. 'Total' is the aggregate number of cells measured across all time points, 'Avg #' quantifies the average number of cells measured in each time point, and 'StDev' quantifies the standard deviation across time points for indicated condition.

	1 ng/ml	10 ng/ml	100 ng/ml	1 ng/ml	10 ng/ml	100 ng/ml
	Pulse	Pulse	Pulse	Continuous	Continuous	Continuous
Total #						
KYM1 TNF	8848	5390	8894	8981	8034	7376
HeLa TNF	4356	4312	3818	5371	4865	5077
U2OS TNF	8550	8280	8864	10706	10031	8856
HCT116 TNF	29243	23119	27174	33410	32012	29683
A549 TNF	3512	2882	9066	6354	5689	8490
Avg #						
KYM1 TNF	983.11	598.89	988.22	997.89	892.67	819.56
HeLa TNF	484	479.11	424.22	596.78	540.56	564.11
U2OS TNF	950	920	984.89	1189.56	1114.56	984
HCT116 TNF	3249.2	2568.78	3019.33	3712.22	3556.89	3298.11
A549 TNF	390.22	320.22	1007.33	706	632.11	943.33
StDev						
KYM1 TNF	120.48	103.66	167.87	230.21	321.18	370.25
HeLa TNF	22.51	19.73	52.29	30.35	39.4	46.64
U2OS TNF	295.58	152	241.88	333.66	343.66	172.45
HCT116 TNF	415.67	305.31	404.14	402.29	290.18	456.39
A549 TNF	67.44	42.29	129.58	133.76	84.9	177.26

Table B2: Related to main Figure 4.4. Summary of pairwise p-values for AUC2/AUC1 compared between test pulse conditions. For each pair of test pulse conditions in Figure 4.4D, p-values are lower than listed. Concentrations listed are relative to a test pulse concentration of 0.2 ng/ml TNF.

Test pulse	0x	1x	5x	50x	500x
0x	-	10^{-17}	10^{-36}	10^{-27}	10^{-22}
1x	10^{-17}	-	10^{-27}	10^{-24}	10^{-21}
5x	10^{-36}	10^{-27}	-	10^{-11}	10^{-16}
50x	10^{-27}	10^{-24}	10^{-11}	-	10^{-6}
500x	10^{-22}	10^{-21}	10^{-16}	10^{-6}	-

A

Model 1: Switch-like response

$$Response = \eta_{basal} + H(dose) \times SF_1 \times CellState \times \eta_{induced}$$

Model 2: Graded response

$$Response = \eta_{basal} + dose \times SF_2 \times CellState \times \eta_{induced}$$

Model 3: Graded response with threshold

$$Response = \eta_{basal} + H(dose) \times dose \times SF_3 \times CellState \times \eta_{induced}$$

$$H(dose) \equiv \begin{cases} 0, dose \leq thresh \\ 1, dose > thresh \end{cases}$$

B

Parameter	Value	Description
η_{basal}	0-5	Adds a random value to cellular responses to simulate basal noise and cell-to-cell heterogeneity
$\eta_{induced}$	±0-15%	Random fluctuations were incorporated into the dose-dependent response of each cell to simulate biological noise
dose	[0, 1, 2, 3]	Each model was simulated with 4 integer doses, with '0' representing the untreated condition
SF_1	6	Scaling factor for Model 1 to normalize responses between models
SF_2	1.5	Scaling factor for Model 2
SF_3	2	Scaling factor for Model 3
CellState	[1,2,…,100]	Each cell was initialized with a unique integer value to describe its state
thresh	0-4	Each cell was initialized with a random value to describe its threshold for activation (used with *H(dose)* in models 1 and 3)

Figure B1: Description for 3 models of single cell dose responses. (A) Formulae for (M1) Switch-like, (M2) Graded, and (M3) Graded with threshold (top, middle and bottom respectively) models of single cell responses. M1 and M3 differ from M2 because they incorporate a heavyside function to model a switch-like behavior. For all models 100 cells were simulated, each initialized with a unique integer value between 1 and 100 to approximate heterogeneity in the cellular state. During initialization each cell was also assigned a threshold for activation (thresh) used in models M1 and M3 to determine whether a cell responds to a given dose. In all models, the strength of response to a dose of cytokine scales linearly with the cell's state, whereby a cell with state 100 and a cell with state 0 will have the strongest and weakest responses respectively. (B) Parameter values used for simulations described in (A) as well as Figure 4.1 of the main text. Basal noise (η_{basal}) was used as a cut-off to quantify the non-responsive fraction in the right column of Figure 4.1B. That is, non-responders are defined as: $Response \leq (\eta_{basal})_{\mathrm{MAX}}$
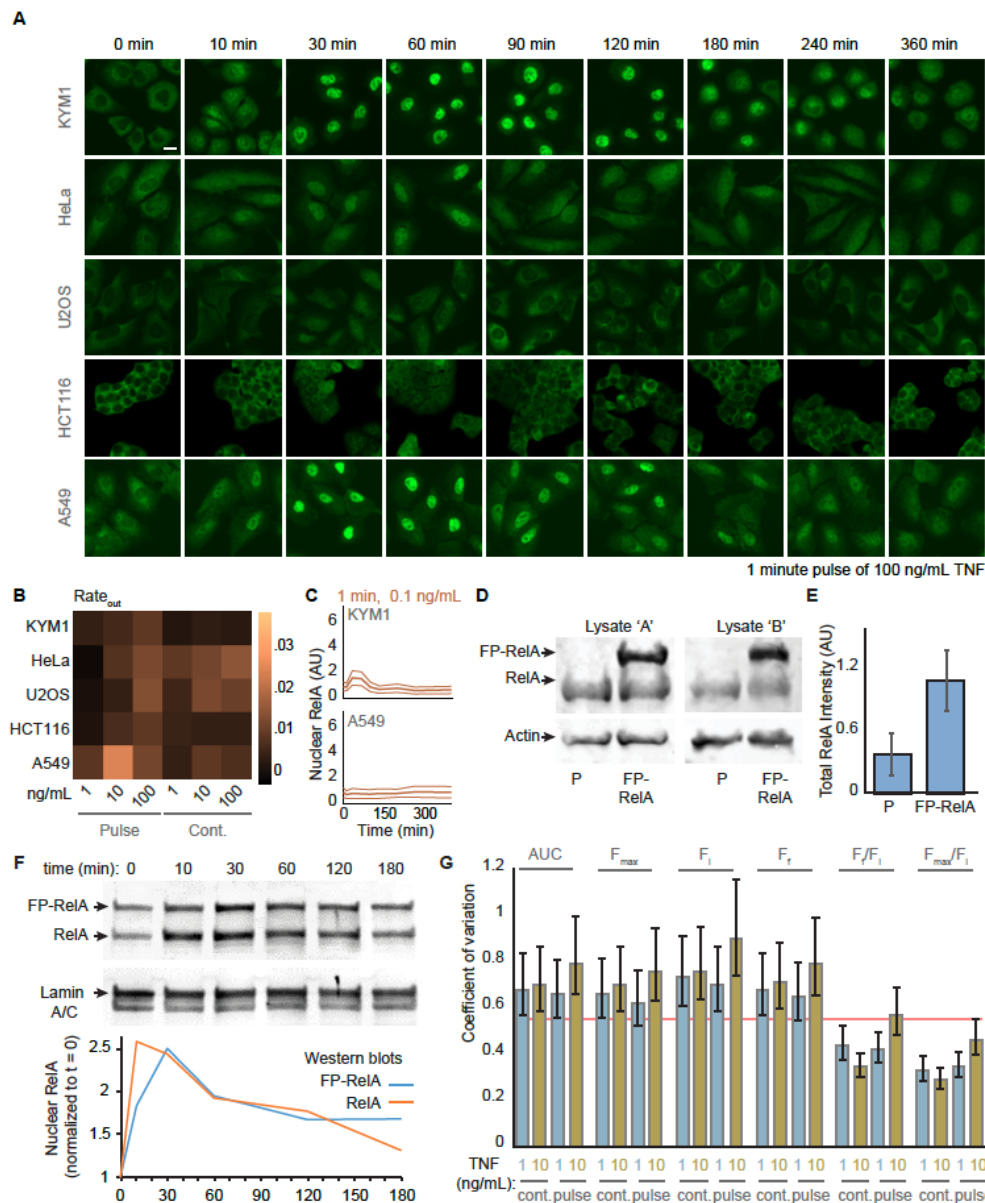
Figure B2: Subcellular localization of endogenous RelA in human cancer cells. (A) Fixed-cell RelA immunofluorescence images of KYM1, HeLa, U2OS, HCT116, and A549 cells. Cells were treated with a 1 minute pulse of 100 ng/mL of TNF and fixed at the indicated time after cytokine exposure; scale bar 10 μm. (B) Heatmap for Rate$_{out}$ descriptor does not show a significant trend. (C) KYM1 cells show greater sensitivity to low concentrations of TNF when compared with A549 cells, the second most responsive cell line to TNF. (D) Western blot of RelA in lysates from parental KYM1 cells (P) and KYM1 cells that express mVenus-RelA (FP-RelA). Biological replicates are shown. (E) Quantification of bands from (D) show that FP-RelA is overexpressed 2.9-fold when compared with P KYM1 cells. For FP-RelA, endogenous and FP-RelA bands were summed; $\pm$SEM. (F) Western blot of nuclear lysates (top) from FP-RelA cells collected at indicated times after exposure to a 1 minute pulse of 10 ng/mL TNF. Quantification of endogenous and FP-RelA bands (bottom) show qualitatively similar dynamics to each other. (G) Bar graphs of the coefficients for variation for descriptors (described in Figure 4.2B) of nuclear FP-RelA density in cells exposed to TNF conditions as indicated. Error bars represent 95% confidence interval of the coefficient of variation.
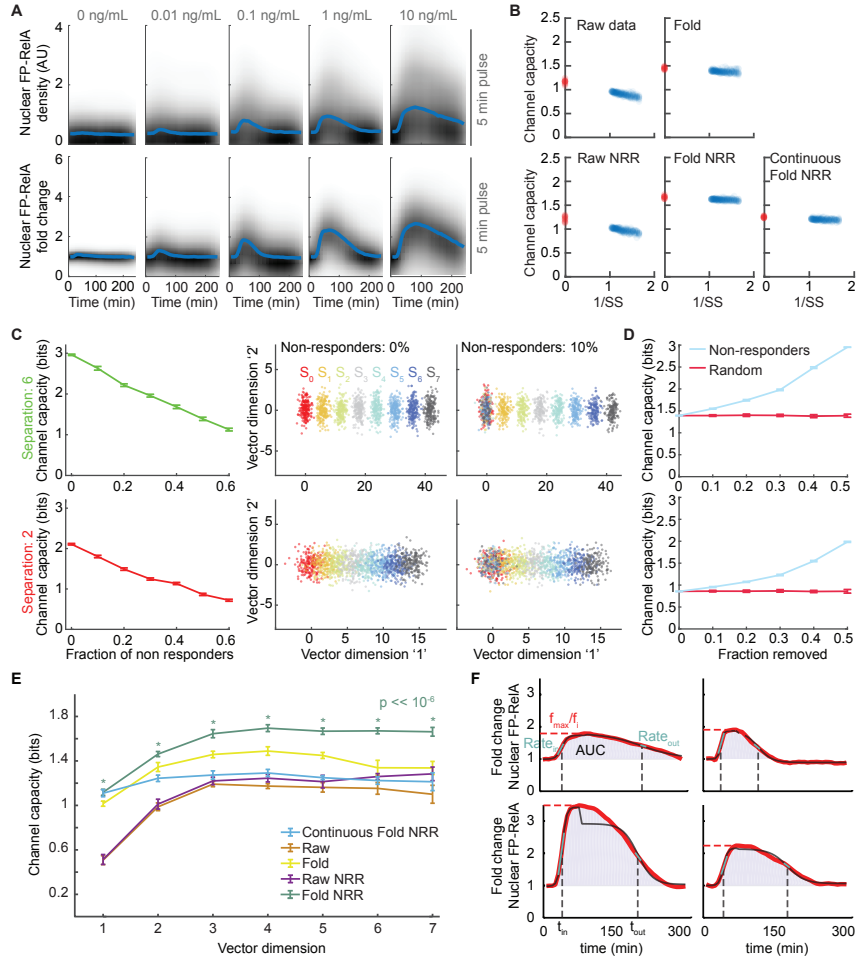
Figure B3: Details of control analysis, models, and features used in channel capacity calculations. (A) Density plots for nuclear FP-RelA in KYM1 cells exposed to a 5 minute TNF pulse of indicated concentration. Raw trajectories (top) and fold-transformed trajectories (bottom) are shown with the untreated control sample in the left-most column. (B) Plots showing jackknife analysis for channel capacity calculations. To correct for sample size bias and determine the variance of channel capacity values, the data set was jackknifed at sequentially increasing fractions, SS (points shown in blue) followed by linear regression to extrapolate the channel capacity at a sample size of infinity (i.e. $1/SS = 0$; points shown in red). The lines obtained via linear regression are relatively flat, indicating that the size of our dataset is large enough to accurately estimate the channel capacity. (C) Plots showing the degradation in channel capacity of a system of eight 2-D Gaussians, when subpopulations of non-responders are considered. Fully separated Gaussians show the theoretical maximum channel capacity of 3 bits for this system (middle top). Overlap in the Gaussians (middle bottom), or the presence of bimodal distributions with an overlapping basal population (right) significantly reduces the values of channel capacity calculations (left); $k = 10$; $\pm$ standard deviation. (D) The effects of subsampling were simulated for the Gaussian system described in (C), except starting with 50% of the population sampled from the basal distribution for fully separated Gaussians (top) or overlapping Gaussians (bottom). In contrast with randomly targeted removal (red), channel capacity values after selectively removing a defined fraction of non-responders (blue) increases the channel capacity up to the defined limit revealing the true channel capacity of the responsive subpopulation; $\pm$ standard deviation. (E) Channel capacity values calculated for each data set over a range of vector dimensions ($d = 5$ is reported in the manuscript): Single-cell trajectories for 'Raw' data sets are in arbitrary fluorescence units; for 'Fold' data sets, each single-cell time course is represented as fold change (A); for 'NRR' data sets, single-cell trajectories with < 1.2-fold change in nuclear FP-RelA are removed; the continuous data set only includes conditions with continuous exposure to TNF (bottom row of Figure 4.3A). Green stars compare Fold-NRR and Raw-NRR ($p < 10^{-6}$, t test). (F) Representative sigmoid fits of five single cell time courses. The red lines show representative experimental trajectories for a single cell's nuclear FP-RelA time course. Black lines show two sigmoid fits used to approximate phases of nuclear entry and exit for FP-RelA and to calculate t_{in} and t_{out}. $Rate_{in}$ and $Rate_{out}$ were calculated from the slopes of adjacent time points from the experimental trajectory (teal line segments). The region used to calculate the AUC is indicated by the shaded area.
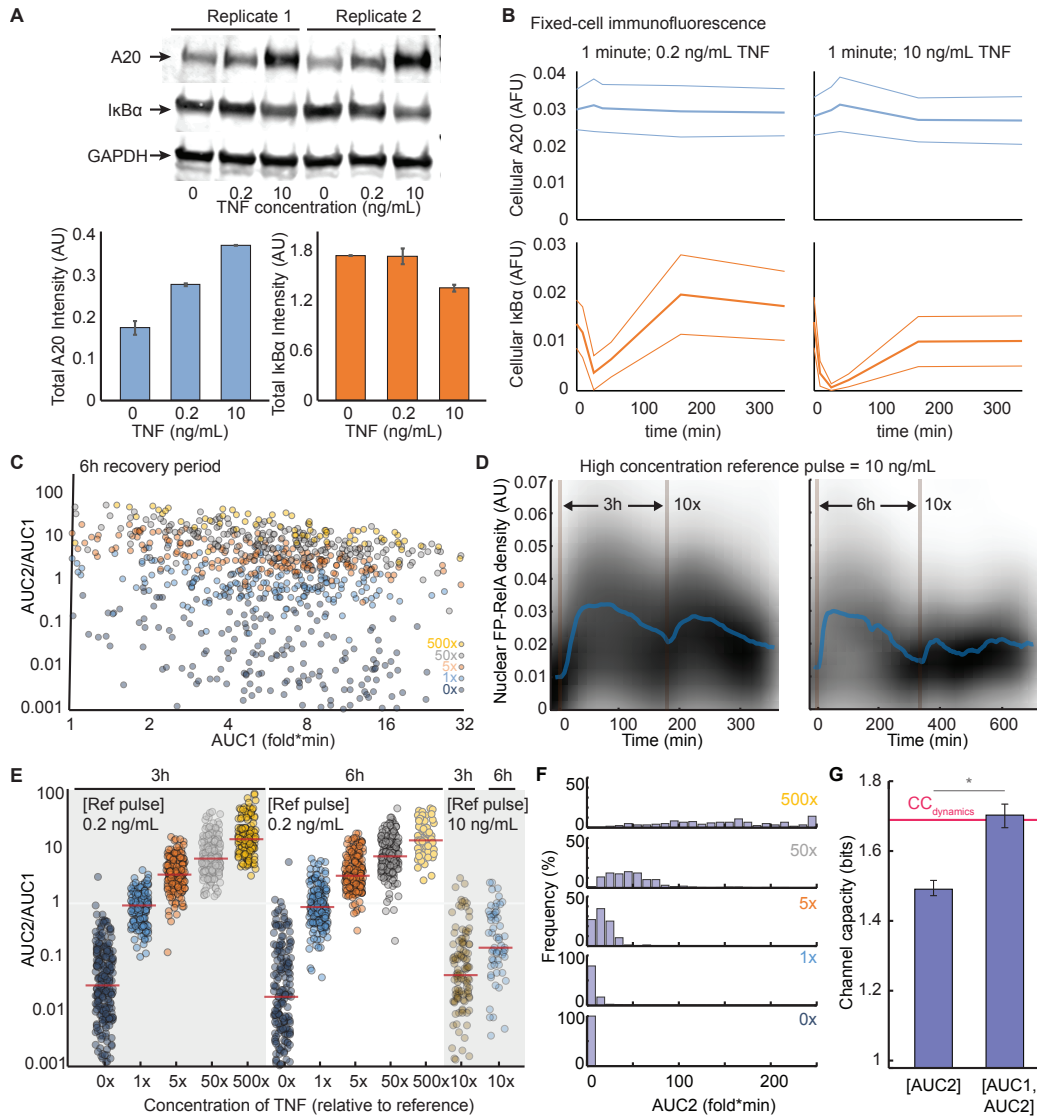
Figure B4: Negative feedback and dampened responses in cells exposed to a high concentration reference pulse of TNF. (A) Western blots (top) for regulators of negative feedback A20 and IκBα in KYM1 cells 3 hours after exposure to a 1 minute reference pulse of TNF at indicated concentration. GAPDH-normalized band intensities quantified in bar graphs (bottom) show that negative regulators are perturbed greater in response to a high concentration reference pulse (10 ng/mL) in comparison with a low concentration reference (0.2 ng/mL); ±SEM of biological duplicate. (B) Results from (A) are confirmed by time courses for average cellular A20 and IκBα from fixed-cell immunofluorescence KYM1 exposed to indicated TNF conditions. Light colored lines indicate the standard deviation. (C) Scatter plots showing AUC2/AUC1 stratified along AUC1 across the range of test conditions in cells after a 6 hour recovery period. (D) Density plots of single-cell FP-RelA time courses for cells exposed to a high concentration (10 ng/mL) reference pulse of TNF followed by a TNF pulse with a 10× increase in concentration 3 hours (left) or 6 hours (right) later. (E) Jitter plots comparing AUC2/AUC1 responses in single cells exposed to a low concentration reference pulse (0.2 ng/mL) or a high concentration reference pulse (10 ng/mL) of TNF with either a 3 hour or 6 hour intervening recovery window. Responses to test pulses are significantly dampened when cells are exposed to a high concentration reference pulse regardless of the duration of the recovery window. (F) Frequency distributions of AUC2 across the range of test conditions. (G) Channel capacity values calculated using AUC2 are compared with a vector that includes reference information about each cell ([AUC1, AUC2]). Red line indicates the channel capacity calculated from the full matrix of conditions (Figure 4.3); ± standard deviation; $p < 10^{-11}$.

117

Appendix C

Supplementary material for Chapter 5: Modeling heterogeneity in TNF-NF-κB signaling
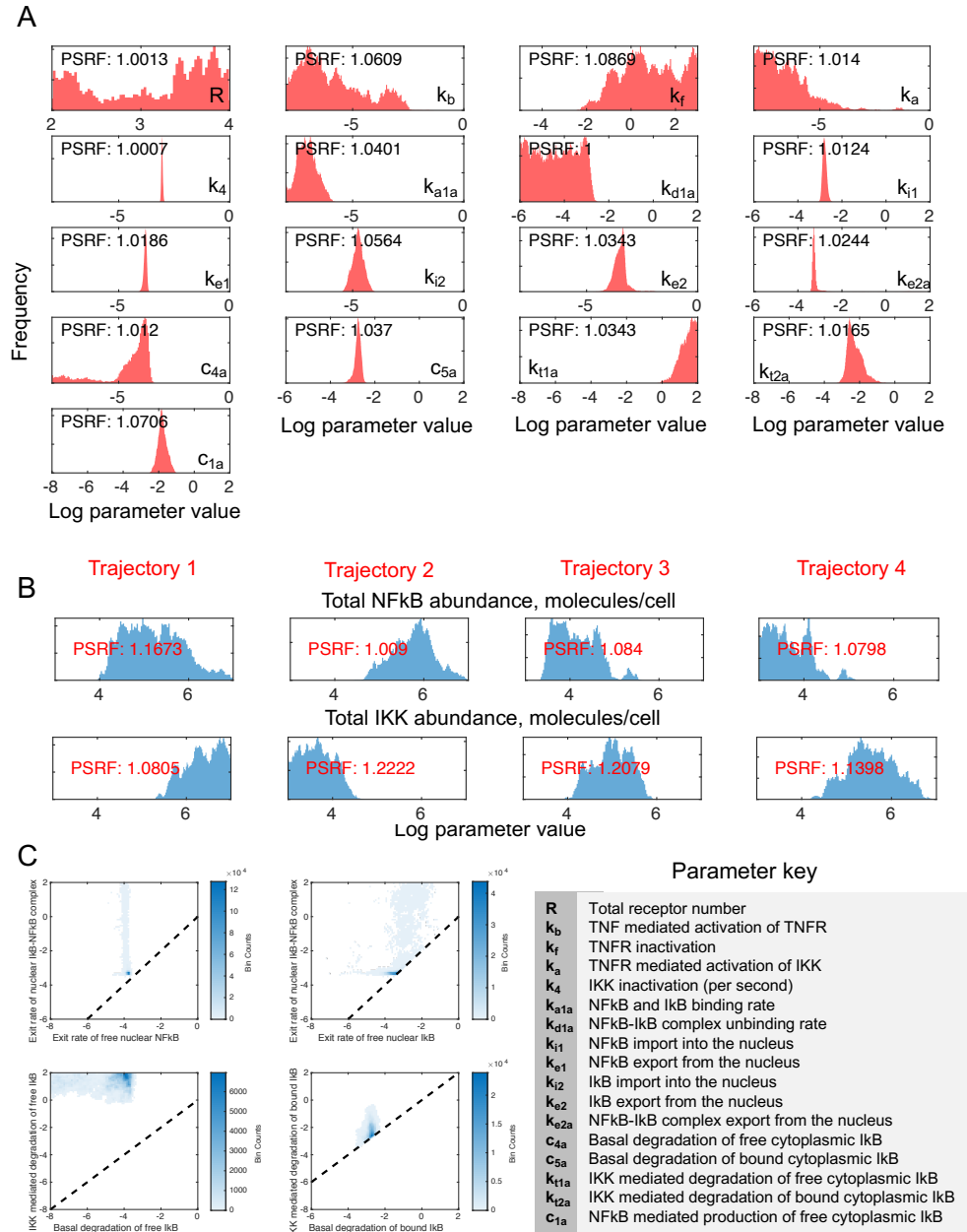
Figure C1: Parameter distributions obtained from PT by fitting the model to multiple single-cell NF-κB trajectories A) Posterior distributions of the shared parameters. B) Posterior distributions of the NF-κB (top row) and IKK (bottom row) abundance parameters that vary between cells. Each column corresponds to the parameters for a particular single-cell NF-κB response. The PSRF for model parameters is calculated from a single PT chain (the code splits the chain into two halves), using third party code available here: https://research.cs.aalto.fi/pml/software/mcmcdiag/ C) Parameter distributions correctly follow hard constraints imposed on key parameter covariation during fitting.
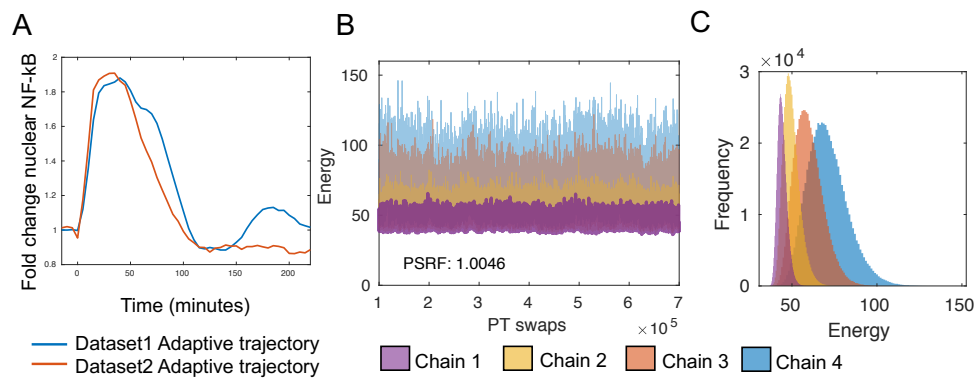
Figure C2: Fitting to $dataset_2$ A) The trajectory that is different between $dataset_1$ and $dataset_2$. B) Energy chains when fitting to $dataset_2$. Chains 1–4 are in increasing order of temperature. C) Histograms of the energy values, from panel B, corresponding to the different chains.
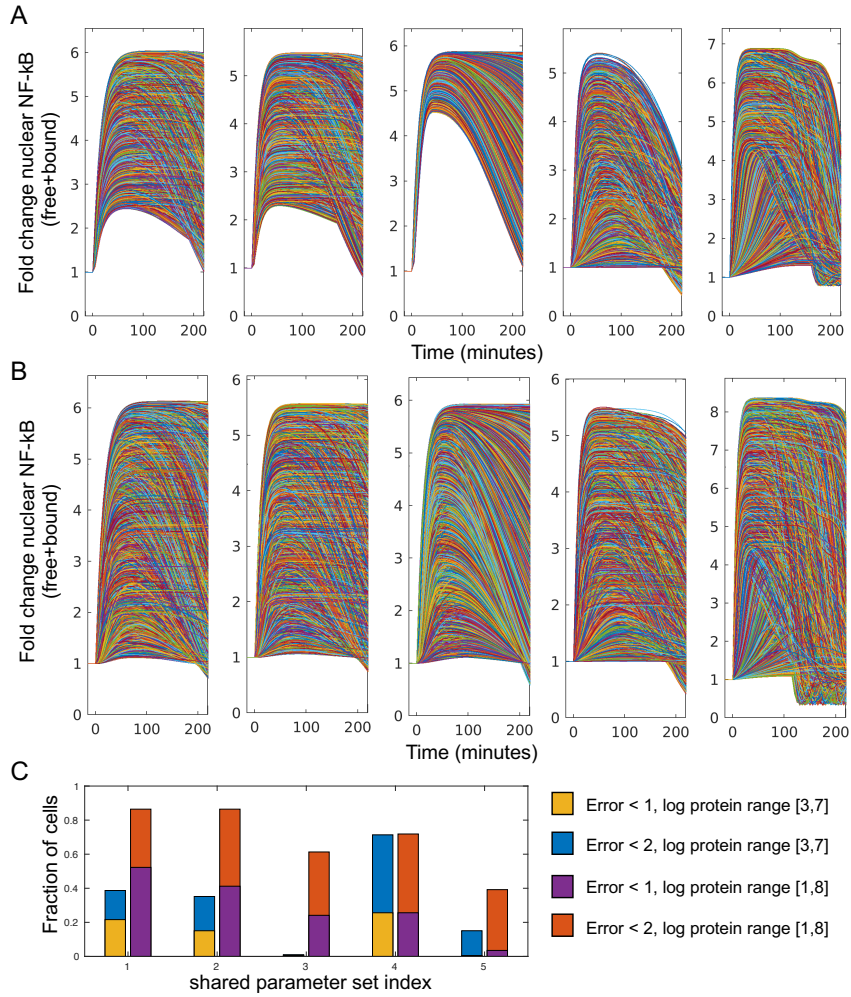
Figure C3: Evaluating the effect of increasing the range of sampled protein abundances for the 1-trajectory fit. A) Simulated NF-κB time courses obtained by fixing a shared parameter set and sampling NF-κB and IKK abundances in the $\log_{10}$ molecules/cell range of 3–7. Each panel shows results for a distinct representative shared parameter set. B) Simulated NF-κB time courses obtained by fixing a shared parameter set and sampling NF-κB and IKK abundances in the $\log_{10}$ molecules/cell range of 1–8. Each panel shows results for the shared parameter set used in corresponding panel (directly above) from (A). Errors corresponding to the simulated distributions in panels A and B.

121
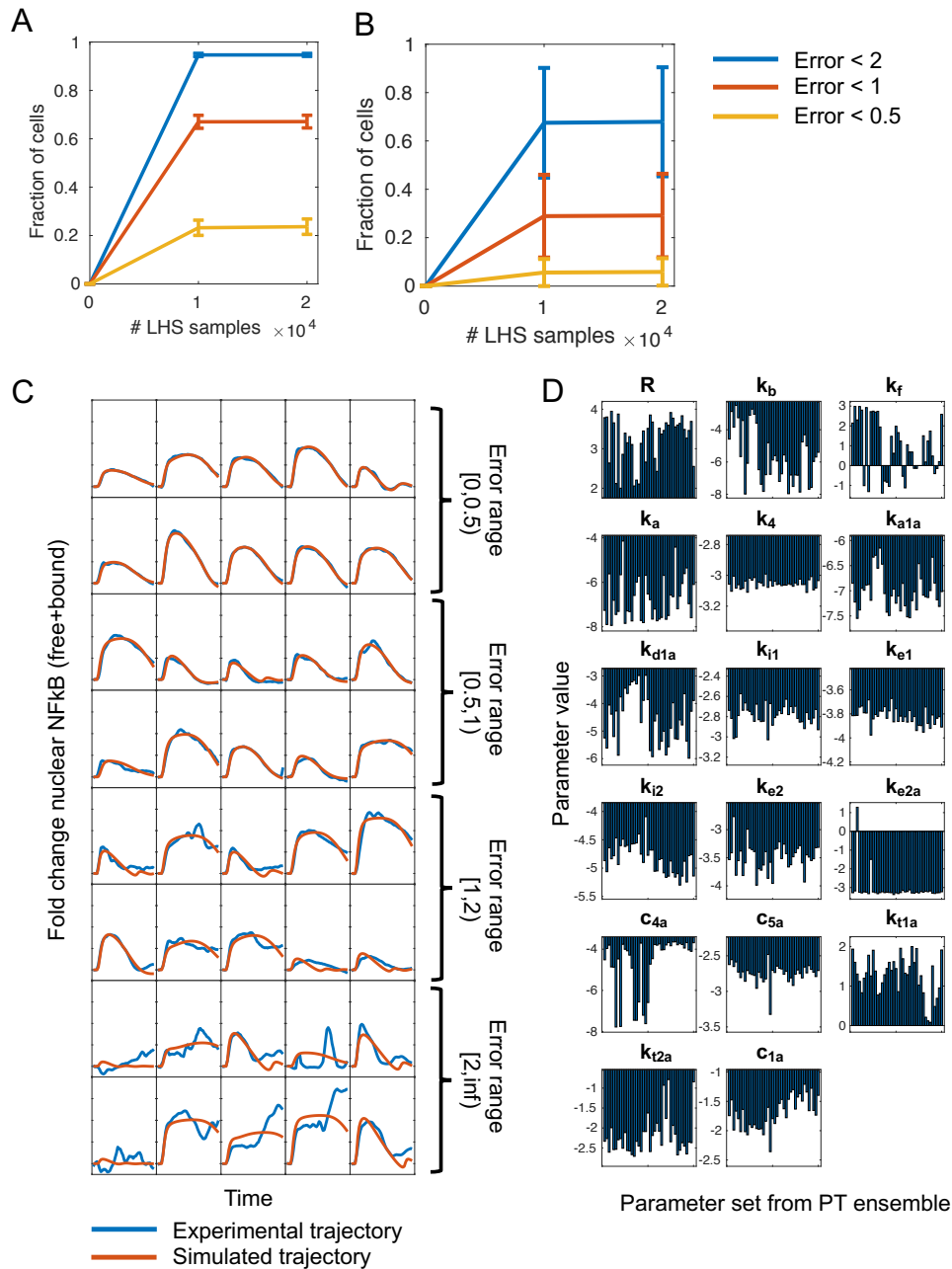
Figure C4: Evaluating the hyperparameters of the sampling procedure. A) Checking for convergence in the number of LHS samples drawn when fitting to $dataset_1$ and B) a single trajectory. Error bars show the mean and standard deviation of the fraction of experimental data captured within an error threshold of 0.5 (yellow), 1 (orange), or 2 (blue), calculated across 40 shared parameter sets, by either drawing 10^4 of 2×10^4 protein samples. The [0,0] point is included by default. C) Representative examples of simulated nearest neighbors for the experimental data that show different error brackets. D) Shared parameter sets (40 sets) corresponding to Figure 5.3D. Each panel corresponds to a different model parameter, and the ith bar in the panel, shows the value of that model parameter in the ith parameter set. The randomly selected sets are considerably dispersed and reflect the non-identifiability of the model.
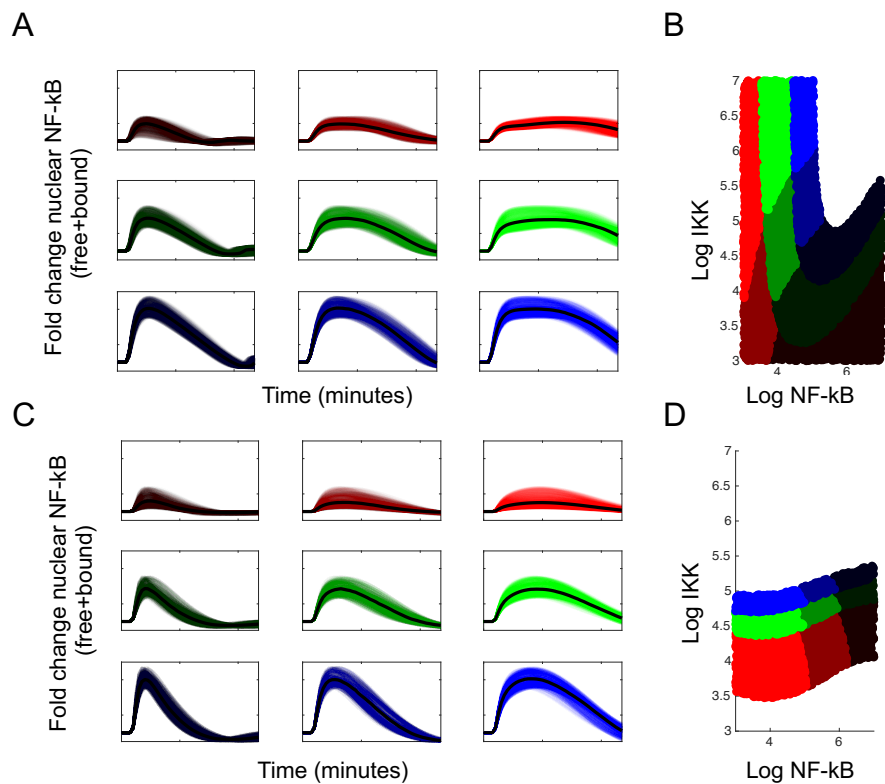
Figure C5: Clustering analysis. A) Binned and clustered trajectories from an LHS sampled dataset obtained using a shared parameter set fit to $dataset_1$, and B) the corresponding clusters (same color coding) in protein expression space. C) Binned and clustered trajectories from an LHS sampled dataset obtained using a shared parameter set fit to $dataset_2$, and D) the corresponding clusters (same color coding) in protein expression space.
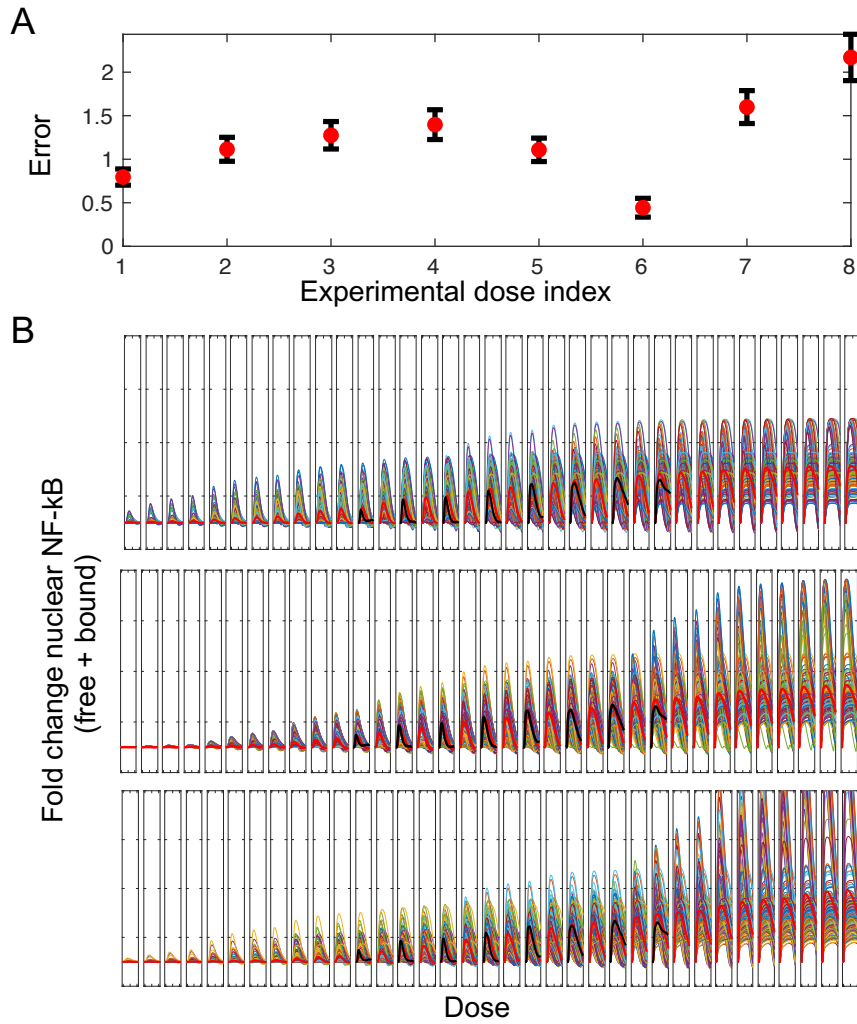
Figure C6: Comparing dose response predictions to experimental data. A) Error bars showing mean and standard deviation of the dose response prediction error across 20 shared parameter sets, for 8 experimentally tested doses. The dose response prediction error is calculated as the Euclidean distance between the simulated mean and experimental mean. B) Examples of dose response for 3 different shared parameter sets. Colored lines show simulations of single-cell responses, the red line shows the mean of the simulated data, and the black line shows the mean of the experimental data for available conditions. y-axis limits are [0,8]. The displayed dose range varies from 2.5×10^{-5} to 7500 ng/ml.
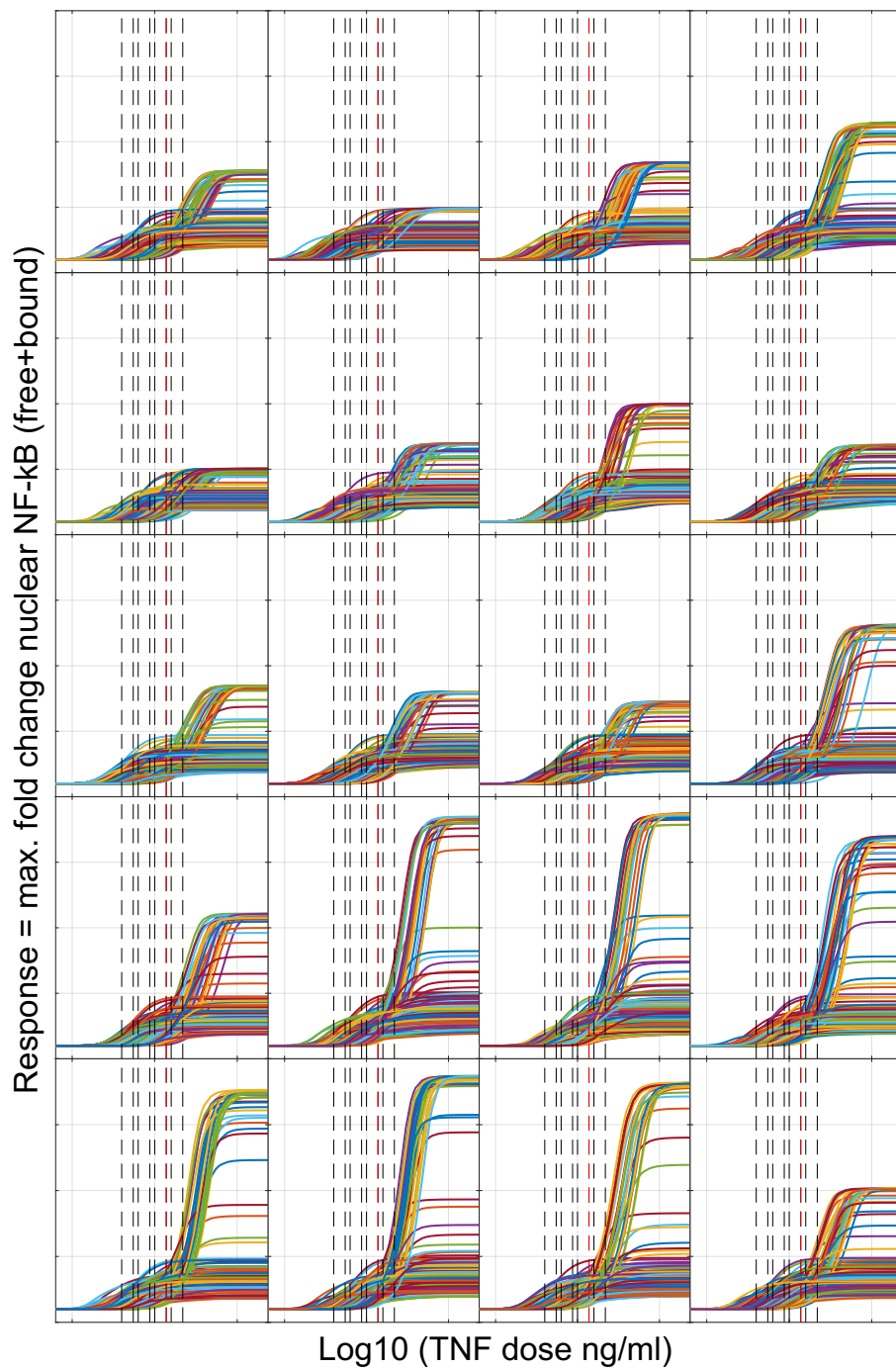
Figure C7: Evaluating the sensitivity of dose-response predictions with respect to shared parameter sets. Each panel shows dose-response curves for a 1-NN dataset derived from NF-κB and IKK variation using a particular shared parameter set. Overall the figure shows results for 20 shared parameter sets. The dashed red line shows the calibration dose of 5 ng/ml. The dashed black lines show the doses [0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50] where we have experimental data, and that was treated as the range of NFkB responsiveness in the experiments. For all panels, the y-axis limits are [0 20], and the x-axis limits are $[10^{-6}, 7.5 \times 10^{6}]$

Table C1: Hard constraints when jointly fitting the NF-κB signaling model to single-cell responses.

Index	Constraint description	Constraint
1	Exit rate of NF-κB-IκB from the nucleus is greater than free NF-κB	ke2a > ke1
2	Exit rate of NF-κB-IκB from the nucleus is greater than free IκB	ke2a > ke2
3	IKK mediated degradation of free cytosolic IκB is greater than the basal rate	kt1a > c4a
4	IKK mediated degradation of NF-κB bound cytosolic IκB is greater than the basal rate	kt2a > c5a
5	At equilibrium there is some NF-κB in the nucleus	$\dfrac{\mathrm{nNF\kappa B}_{eq} + \mathrm{nNF\kappa B} - \mathrm{I\kappa B}_{eq}}{\mathrm{TotalNF\kappa B}} \in$ [0.05,0.5]
6	Equilibrium check before TNF stimulation	Rate of change for each model species < 1e-3 molecules/5e5 seconds

Bibliography

[1] Paolo Mazzarello. A unifying concept: The history of cell theory. *Nature Cell Biology*, 1:E13–E15, 1999.

[2] Norbert Perrimon, Chrysoula Pitsouli, and Ben-Zion Shilo. Signaling mechanisms controlling cell fate and embryonic patterning. *Cold Spring Harbor Perspectives in Biology*, 4, 2012.

[3] Julian Downward. The ins and outs of signalling. *Nature*, 411:759–762, 2001.

[4] Richard Sever and Joan S. Brugge. Signal transduction in cancer. *Cold Spring Harbor Perspectives in Medicine*, 5, 2015.

[5] João Antonio Chaves de Souza, Carlos Rossa Junior, Gustavo Pompermaier Garlet, Andressa Vilas Boas Nogueira, and Joni Augusto Cirelli. Modulation of host cell signaling pathways as a therapeutic approach in periodontal disease. *Journal of Applied Oral Science*, 20(2):128–138, 2012.

[6] Seyyed Shamsadin Athari. Targeting cell signaling in allergic asthma. *Signal Transduction and Targeted Therapy*, 4, 2019.

[7] Tuomas Sandholm. Steering Evolution Strategically: Computational Game Theory and Opponent Exploitation for Treatment Planning, Drug Design, and Synthetic Biology. *Proceedings of the Twenty-Ninth AAAI conference on Artificial Intelligence, Senior Member Track*, 2015.

[8] Robin Schmucker, Gabriele Farina, James Faeder, Fabian Fröhlich, Ali Sinan Saglam, and Tuomas Sandholm. Combination Treatment Optimization Using a Pan-Cancer Pathway Model. *biorxiv*, 2020.

[9] Christian Kroer and Tuomas Sandholm. Sequential Planning for Steering Immune System Adaptation. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016.

[10] Jeremy Gunawardena. Signals and Systems: Towards a Systems Biology of Signal Transduction. *Proceedings of the IEEE*, 96(8):1386–1397, 2008.

[11] Arathi Nair, Prashant Chauhan, Bhaskar Saha, and Katharina F. Kubatzky. Conceptual evolution of cell signaling. *International Journal of Molecular Sciences*, 20(13):3292, 2019.

[12] Sanjana Gupta, Robin E.C. Lee, and James R. Faeder. Parallel Tempering with Lasso for model reduction in systems biology. *PLoS Computational Biology*, 16(3):e1007669, 2020.

[13] Narat J. Eungdamrong and Ravi Iyengar. Modeling cell signaling networks. *Biology of the Cell*, 96(5):355–362, 2004.

[14] Jeremy Gunawardena. Some lessons about models from Michaelis and Menten. *Molecular Biology of the Cell*, 23(4):517–519, 2012.

[15] Jakub Pekalski, Pawel J. Zuk, Marek Kochańczyk, Michael Junkin, Ryan Kellogg, Savaş Tay, and Tomasz Lipniacki. Spontaneous NF-κB activation by autocrine TNFα signaling: A computational analysis. *PLoS ONE*, 8(11), 2013.

[16] J. R. Faeder, W. S. Hlavacek, I. Reischl, M. L. Blinov, H. Metzger, A. Redondo, C. Wofsy, and B. Goldstein. Investigation of Early Events in FceRI-Mediated Signaling Using a Detailed Mathematical Model. *The Journal of Immunology*, 170(7):3769–3781, 2003.

[17] Michael L. Blinov, James R. Faeder, Byron Goldstein, and William S. Hlavacek. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *BioSystems*, 83:136–151, 2006.

[18] Robert D. Phair. Mechanistic modeling confronts the complexity of molecular cell biology. *Molecular Biology of the Cell*, 25(22):3494–3496, 2014.

[19] Ruth E Baker, Jose-Maria Peña, Jayaratnam Jayamohan, and Antoine Jérusalem. Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5), 2018.

[20] Juliane Liepe, Paul Kirk, Sarah Filippi, Tina Toni, Chris P Barnes, and Michael P H Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9(2):439–456, 2014.

[21] Sanjana Gupta, Jacob Czech, Robert Kuczewski, Thomas M Bartol, Terrence J Sejnowski, Robin E C Lee, and James R Faeder. Spatial Stochastic Modeling with MCell and CellBlender. In *Quantitative Biology: Theory, Computational Methods, and Models, Brian Munsky, William S. Hlavacek, and Lev S. Tsimring, Eds.*, pages 485–511. MIT Press, 2018.

[22] Alexander D. Malkin, Robert P. Sheehan, Shibin Mathew, William J. Federspiel, Heinz Redl, and Gilles Clermont. A Neutrophil Phenotype Model for Extracorporeal Treatment of Sepsis. *PLoS Computational Biology*, 11(10), 2015.

[23] William W. Chen, Mario Niepel, and Peter K. Sorger. Classic and contemporary approaches to modeling biochemical reactions. *Genes and Development*, 24:1861–1875, 2010.

[24] T S Ahearn, R T Staff, T W Redpath, and S I K Semple. The use of the Levenberg–Marquardt curve-fitting algorithm in pharmacokinetic modelling of DCE-MRI data. *Physics in Medicine and Biology*, 50:N85–N92, 2005.

[25] Fuchang Gao and Lixing Han. Implementing the Nelder-Mead simplex algorithm with adaptive parameters. *Computational Optimization and Applications*, 51(1), 2012.

[26] Attila Gábor and Julio R Banga. Robust and efficient parameter estimation in dynamic models of biological systems. *BMC systems biology*, 9, 2015.

[27] Pu Li and Quoc Dong Vu. Identification of parameter correlations for parameter estimation in dynamic biological models. *BMC Systems Biology*, 7, 2013.

[28] Andreas Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 2009.

[29] David J. Klinke. An empirical Bayesian approach for model-based inference of cellular signaling networks. *BMC Bioinformatics*, 10:371, 2009.

[30] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910, 2005.

[31] H. Eydgahi, W. W. Chen, J. L. Muhlich, D. Vitkup, J. N. Tsitsiklis, and P. K. Sorger. Properties of cell death models calibrated and compared using Bayesian approaches. *Molecular Systems Biology*, 9(1):644–644, 2014.

[32] Sanjana Gupta, Liam Hainsworth, Justin Hogg, Robin Lee, and James Faeder. 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*, pages 690–697, 2018.

[33] Lea Goentoro, Oren Shoval, Marc W. Kirschner, and Uri Alon. The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. *Molecular Cell*, 36(5):894–899, 2009.

[34] Uri Alon. Network motifs: theory and experimental approaches. *Nature reviews. Genetics*, 8(6):450–61, 2007.

[35] Félix Proulx-Giraldeau, Thomas J. Rademaker, and Paul François. Untangling the Hairball: Fitness-Based Asymptotic Reduction of Biological Networks. *Biophysical Journal*, 113:1893–1906, 2017.

[36] Qiang Zhang, Sudin Bhattacharya, Melvin E Andersen, and Rory B Conolly. Computational Systems Biology and Dose-Response Modeling in Relation to New Directions in Toxicity Testing. *Journal of Toxicology and Environmental Health, Part B*, 13:253–276, 2010.

[37] Arthur D. Lander. The edges of understanding. *BMC Biology*, 8:40, 2010.

[38] Lily A. Chylek, David A. Holowka, Barbara A. Baird, and William S. Hlavacek. An interaction library for the FcϵRI signaling network. *Frontiers in Immunology*, 5:172, 2014.

[39] R Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B*, 58(1):267–288, 1996.

[40] Antero Salminen, Jari Huuskonen, Johanna Ojala, Anu Kauppinen, Kai Kaarniranta, and Tiina Suuronen. Activation of innate immunity system during aging: NF-kB signaling is the molecular culprit of inflamm-aging. *Ageing Research Reviews*, 7:83–105, 2008.

[41] Nicolas A. Pabon, Qiuhong Zhang, J. Agustin Cruz, David L. Schipper, Carlos J. Camacho, and Robin E.C. Lee. A network-centric approach to drugging TNF-induced NF-κB signaling. *Nature Communications*, 10(1), 2019.

[42] Andrea Oeckinghaus and Sankar Ghosh. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harbor perspectives in biology*, 1, 2009.

[43] Chaitanya S. Mokashi, David L. Schipper, Mohammad A. Qasaimeh, and Robin E.C. Lee. A System for Analog Control of Cell Culture Dynamics to Reveal Capabilities of Signaling Networks. *iScience*, 19:586–596, 2019.

[44] Robin E.C. Lee, Mohammad A. Qasaimeh, Xianfang Xia, David Juncker, and Suzanne Gaudet. NF-κB signalling and cell fate decisions in response to a short pulse of tumour necrosis factor. *Scientific Reports*, 6, 2016.

[45] Savas Tay, Jacob J. Hughey, Timothy K. Lee, Tomasz Lipniacki, Stephen R. Quake, and Markus W. Covert. Single-cell NF-kB dynamics reveal digital activation and analogue information processing. *Nature*, 466(7303):267–271, 2010.

[46] Qiuhong Zhang, Sanjana Gupta, David L. Schipper, Gabriel J. Kowalczyk, Allison E. Mancini, James R. Faeder, and Robin E.C. Lee. NF-κB Dynamics Discriminate between TNF Doses in Single Cells. *Cell Systems*, 5(6):638–645.e5, 2017.

[47] Raymond Cheong, Alex Rhee, Chiaochun Joanne Wang, Ilya Nemenman, and Andre Levchenko. Information transduction capacity of noisy biochemical signaling networks. *Science*, 334(6054):354–358, 2011.

[48] Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, and Roy Wollman. Accurate information transmission through dynamic biochemical signaling networks. *Science*, 346(6215):1370–1373, 2014.

[49] Steven J. Altschuler and Lani F. Wu. Cellular Heterogeneity: Do Differences Make a Difference? *Cell*, 141:559–563, 2010.

[50] Sabrina L Spencer, Suzanne Gaudet, John G Albeck, John M Burke, and Peter K Sorger. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, 459(7245):428–432, 2009.

[51] Alejandro Colman-Lerner, Andrew Gordon, Eduard Serra, Tina Chin, Orna Resnekov, Drew Endy, C. Gustavo Pesce, and Roger Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437:699–706, 2005.

[52] Ting Liu, Lingyun Zhang, Donghyun Joo, and Shao-Cong Sun. NF-κB signaling in inflammation. *Signal Transduction and Targeted Therapy*, 2, 2017.

[53] Andre Levchenko and Ilya Nemenman. Cellular noise and information transmission. *Current Opinion in Biotechnology*, 28:156–164, 2014.

[54] Raymond Cheong, Alex Rhee, Chiaochun Joanne Wang, Ilya Nemenman, and Andre Levchenko. Information transduction capacity of noisy biochemical signaling networks. *Science*, 334(6054):354–358, 2011.

[55] Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, and Roy Wollman. Accurate information transmission through dynamic biochemical signaling networks. *Science*, 346(6215):1370–1373, 2014.

[56] Robin E.C. Lee, Sarah R. Walker, Kate Savery, David A. Frank, and Suzanne Gaudet. Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Molecular Cell*, 53(6):867–879, 2014.

[57] Jason Yao, Anna Pilko, and Roy Wollman. Distinct cellular states determine calcium signaling response. *Molecular Systems Biology*, 12(12):894, 2016.

[58] Lekshmi Dharmarajan, Hans Michael Kaltenbach, Fabian Rudolf, and Joerg Stelling. A Simple and Flexible Computational Framework for Inferring Sources of Heterogeneity from Single-Cell Dynamics. *Cell Systems*, 8(1):15–26, 2019.

[59] Mridul K. Kalita, Khachik Sargsyan, Bing Tian, Adriana Paulucci-Holthauzen, Habib N. Najm, Bert J. Debusschere, and Allan R. Brasier. Sources of cell-to-cell variability in canonical nuclear factor-κB (NF-κB) signaling pathway inferred from single cell dynamic images. *Journal of Biological Chemistry*, 286(43):37741–37757, 2011.

[60] Marc R. Birtwistle, Jens Rauch, Anatoly Kiyatkin, Edita Aksamitiene, Maciej Dobrzyński, Jan B. Hoek, Walter Kolch, Babatunde A. Ogunnaike, and Boris N. Kholodenko. Emergence of bimodal cell population responses from the interplay between analog single-cell signaling and protein expression noise. *BMC Systems Biology*, 6, 2012.

[61] Ofer Feinerman, Joël Veiga, Jeffrey R Dorfman, Ronald N Germain, and Grégoire Altan-Bonnet. Variability and Robustness in T Cell Activation from Regulated Heterogeneity in Protein Levels. *Science*, 321, 2008.

[62] Jette Strasen, Uddipan Sarma, Marcel Jentsch, Stefan Bohn, Caibin Sheng, Daniel Horbelt, Petra Knaus, Stefan Legewie, and Alexander Loewer. Cell-specific responses to the cytokine TGF β are determined by variability in protein levels. *Molecular Systems Biology*, 14:e7733, 2018.

[63] Byron Goldstein, James R Faeder, and William S Hlavacek. Mathematical and computational models of immune-receptor signalling. *Nature Reviews Immunology*, 4:445–456, 2004.

[64] Robin E C Lee, Sarah R Walker, Kate Savery, David A Frank, and Suzanne Gaudet. Fold Change of Nuclear NF-kB Determines TNF-Induced Transcription in Single Cells. *Molecular Cell*, 53(6):867–879, 2014.

[65] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A. Kaandorp, and Joke G. Blom. Systems biology: Parameter estimation for biochemical models. *FEBS Journal*, 276:886–902, 2009.

[66] Alexander D. Malkin, Robert P. Sheehan, Shibin Mathew, William J. Federspiel, Heinz Redl, and Gilles Clermont. A Neutrophil Phenotype Model for Extracorporeal Treatment of Sepsis. *PLoS Computational Biology*, 11(10):1–30, 2015.

[67] Sergio Iadevaia, Yiling Lu, Fabiana C. Morales, Gordon B. Mills, and Prahlad T. Ram. Identification of optimal drug combinations targeting cellular networks: Integrating phospho-proteomics and computational network analysis. *Cancer Research*, 70(17):6704–6714, 2010.

[68] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally Sloppy Parameter Sensitivities in Systems Biology Models. *PLOS Computational Biology*, 3(10):1871–1878, 2007.

[69] Carlos F. Lopez, Jeremy L. Muhlich, John A. Bachman, and Peter K. Sorger. Programming biological models in Python using PySB. *Molecular Systems Biology*, 9(1):1–19, 2013.

[70] Fengkai Zhang, Bastian R Angermann, and Martin Meier-Schellersheim. The Simmune Modeler visual interface for creating signaling networks based on bi-molecular interactions. *Bioinformatics*, 29(9):1229–30, 2013.

[71] Leonard A. Harris, Justin S. Hogg, José Juan Tapia, John A.P. Sekar, Sanjana Gupta, Ilya Korsunsky, Arshi Arora, Dipak Barua, Robert P. Sheehan, and James R. Faeder. BioNetGen 2.2: Advances in rule-based modeling. *Bioinformatics*, 32(21):3366–3368, 2016.

[72] Lily A. Chylek, Vyacheslav Akimov, Jörn Dengjel, Kristoffer T. G. Rigbolt, Bin Hu, William S. Hlavacek, and Blagoy Blagoev. Phosphorylation site dynamics of early T-cell receptor signaling. *PLoS ONE*, 9(8):e104240, 2014.

[73] Ulrich H.E. Hansmann. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters*, 281:140–150, 1997.

[74] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141–151, 1999.

[75] Sarah Lukens, Jay DePasse, Roni Rosenfeld, Elodie Ghedin, Ericka Mochan, Shawn T Brown, John Grefenstette, Donald S Burke, David Swigon, and Gilles Clermont. A large-scale immuno-epidemiological simulation of influenza A epidemics. *BMC Public Health*, 14(1):1019, 2014.

[76] Diego Fernández Slezak, Cecilia Suárez, Guillermo A. Cecchi, Guillermo Marshall, and Gustavo Stolovitzky. When the optimal is not the best: Parameter estimation in complex biological models. *PLoS ONE*, 5(10), 2010.

[77] Larry Wasserman. *All of Statistics. A Concise Course in Statistical Inference.* Springer Publishing Company, Incorporated, 2005.

[78] Trevor Park and George Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.

[79] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67:301–320, 2005.

[80] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[81] S. Chib and E. Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

[82] Alan C. Hindmarsh, Peter N. Brown, Keith E. Grant, Steven L. Lee, Radu Serban, Dan E. Shumaker, and Carol S. Woodward. Sundials: Suite of nonlinear and differential/algebraic equation solvers. *ACM Transactions on Mathematical Software*, 31(3):363–396, 2005.

[83] Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, Adam P Arkin, Benjamin J Bornstein, Dennis Bray, Athel Cornish-Bowden, et al. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[84] Brandon R Thomas, Lily A Chylek, Joshua Colvin, Suman Sirimulla, Andrew HA Clayton, William S Hlavacek, and Richard G Posner. BioNetFit: a fitting tool compatible with BioNetGen, NFsim and distributed computing environments. *Bioinformatics*, 32(5):798–800, 2015.

[85] Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael P H Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6:187–202, 2009.

[86] Nayantara Bhatnagar, Andrej Bogdanov, and Elchanan Mossel. The Computational Complexity of Estimating MCMC COnvergence Time. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques. Lecture Notes in Computer Science, vol 6845. Springer, Berlin, Heidelberg*, 6845, 2011.

[87] U. Kummer, L. F. Olsen, C. J. Dixon, A. K. Green, E. Bornberg-Bauer, and G. Baier. Switching from simple to complex oscillations in calcium signaling. *Biophysical Journal*, 79(3):1188–1195, 2000.

[88] John J. Tyson, Katherine C. Chen, and Bela Novak. Sniffers , buzzers , toggles and blinkers : dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, 2003.

[89] Dipak Barua, James R Faeder, and Jason M Haugh. Structure-based kinetic models of modular signaling protein function: focus on Shp2. *Biophysical journal*, 92:2290–300, 2007.

[90] Kevin S. Brown and James P. Sethna. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E*, 68:021904, Aug 2003.

[91] Erin M Shockley, Jasper A Vrugt, and Carlos F Lopez. PyDREAM: High-dimensional parameter inference for biological models in Python. *Bioinformatics*, doi:10.1093/bioinformatics/btx626, 2017.

[92] Heikki Haario, Marko Laine, Antonietta Mira, and Eero Saksman. DRAM: Efficient adaptive MCMC. *Statistics and Computing*, 16:339–354, 2006.

[93] Beata Hat, Marek Kochańczyk, Marta N. Bogdał, and Tomasz Lipniacki. Feedbacks, Bifurcations, and Cell Fate Decision-Making in the p53 System. *PLoS Computational Biology*, 12(2), 2016.

[94] Thomas J. Snowden, Piet H. van der Graaf, and Marcus J. Tindall. Methods of Model Reduction for Large-Scale Biological Systems: A Survey of Current Methods and Trends. *Bulletin of Mathematical Biology*, 79(7):1449–1486, 2017.

[95] Tom Quaiser, Anna Dittrich, Fred Schaper, and Martin Mönnigmann. A simple work flow for biologically inspired model reduction - application to early JAK-STAT signaling. *BMC Systems Biology*, 5, 2011.

[96] Tim Maiwald, Helge Hass, Bernhard Steiert, Joep Vanlier, Raphael Engesser, Andreas Raue, Friederike Kipkeew, Hans H. Bock, Daniel Kaschek, Clemens Kreutz, and Jens Timmer. Driving the model to its limit: Profile likelihood based model reduction. *PLoS ONE*, 11(9), 2016.

[97] Thomas P. Prescott and Antonis Papachristodoulou. Layered decomposition for the model order reduction of timescale separated biochemical reaction networks. *Journal of Theoretical Biology*, 356:113–122, 2014.

[98] Andrea Ciliberto, Fabrizio Capuani, and John J. Tyson. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Computational Biology*, 3(3):0463–0472, 2007.

[99] David J. Klinke and Stacey D. Finley. Timescale analysis of rule-based biochemical reaction networks. *Biotechnol Prog*, 28(1):33–44, 2012.

[100] Michael Gabel, Tobias Hohl, Andrea Imle, Oliver T. Fackler, and Frederik Graw. FAMoS: A Flexible and dynamic Algorithm for Model Selection to analyse complex systems dynamics. *PLOS Computational Biology*, 15(8):e1007230, 2019.

[101] M R Maurya, S J Bornheimer, V Venkatasubramanian, and S Subramaniam. Mixed-integer nonlinear optimisation approach to coarse-graining biochemical networks. *IET systems biology*, 3(1):24–39, 2009.

[102] Binita Bhattacharjee, Douglas A. Schwer, Paul I. Barton, and William H. Green. Optimally-reduced kinetic models: Reaction elimination in large-scale kinetic mechanisms. *Combustion and Flame*, 135(3):191–208, 2003.

[103] Linda Petzold and Wenjie Zhu. Model reduction for chemical kinetics: An optimization approach. *AIChE Journal*, 45(4):869–886, 1999.

[104] Richard Bonneau, David J. Reiss, Paul Shannon, Marc Facciotti, Leroy Hood, Nitin S. Baliga, and Vesteinn Thorsson. The inferelator: An algorithn for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biology*, 7(5), 2006.

[105] Tong Tong Wu, Yi Fang Chen, Trevor Hastie, Eric Sobel, and Kenneth Lange. Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics*, 25(6):714–721, 2009.

[106] H. Robert Frost and Christopher I. Amos. Gene set selection via LASSO penalized regression (SLPR). *Nucleic Acids Research*, 45(12), 2017.

[107] Yiming Lu, Yang Zhou, Wubin Qu, Minghua Deng, and Chenggang Zhang. A Lasso regression model for the construction of microRNA-target regulatory networks. *Bioinformatics*, 27(17):2406–2413, 2011.

[108] Matthieu Vignes, Jimmy Vandel, David Allouche, Nidal Ramadan-Alban, Christine Cierco-Ayrolles, Thomas Schiex, Brigitte Mangin, and Simon de Givry. Gene regulatory network reconstruction using bayesian networks, the dantzig selector, the lasso and their meta-analysis. *PLoS ONE*, 6(12), 2011.

[109] Jiahan Li, Kiranmoy Das, Guifang Fu, Runze Li, and Rongling Wu. The Bayesian lasso for genome-wide association studies. *Bioinformatics*, 27(4):516–523, 2011.

[110] Swati Biswas and Shili Lin. Logistic Bayesian LASSO for Identifying Association with Rare Haplotypes and Application to Age-Related Macular Degeneration. *Biometrics*, 68(2):587–597, 2012.

[111] Bernhard Steiert, Jens Timmer, and Clemens Kreutz. L1 regularization facilitates detection of cell type-specific parameters in dynamical systems. *Bioinformatics*, 32(17):i718–i726, 2016.

[112] O. Atay and J. M. Skotheim. Modularity and predictability in cell signaling and decision making. *Molecular Biology of the Cell*, 25(22):3445–3450, 2014.

[113] Lukas Meier, Sara Van De Geer, and Peter Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 70(1):53–71, 2008.

[114] Henry Kautz, Bart Selman, and Yueyen Jiang. A general stochastic approach to solving problems with hard and soft constraints. In *The Satisfiability Problem: Theory and Applications*, pages 573–586. American Mathematical Society, 1996.

[115] Stephen P. Brooks and Andrew Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

[116] G. O. Roberts, A. Gelman, and W. R. Gilks. Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability*, 7(1):110–120, 1997.

[117] Mathieu Rosenbaum and Alexandre B. Tsybakov. Sparse recovery under matrix uncertainty. *Annals of Statistics*, 38(5):2620–2651, 2010.

[118] A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochemical Society Transactions*, 31(6):1472–1473, 2003.

[119] D. A. Lauffenburger. Cell signaling pathways as control modules: Complexity for simplicity? *Proceedings of the National Academy of Sciences*, 97(10):5031–5033, 2002.

[120] Leland H. Hartwell, John J. Hopfield, Stanislas Leibler, and Andrew W. Murray. From molecular to modular cell biology. *Nature*, 402:C47–C52, 1999.

[121] Jeffrey D. Kearns, Soumen Basak, Shannon L. Werner, Christine S. Huang, and Alexander Hoffmann. IκBϵ provides negative feedback to control NF-κB oscillations, signaling dynamics, and inflammatory gene expression. *Journal of Cell Biology*, 173(5):659–664, 2006.

[122] Nadine Tarantino, Jean Yves Tinevez, Elizabeth Faris Crowell, Bertrand Boisson, Ricardo Henriques, Musa Mhlanga, Fabrice Agou, Alain Israël, and Emmanuel Laplantine. Tnf and il-1 exhibit distinct ubiquitin requirements for inducing NEMO-IKK supramolecular structures. *Journal of Cell Biology*, 204(2):231–245, 2014.

[123] Eva Balsa-canto and Julio R Banga. AMIGO , a toolbox for advanced model identification in systems biology using global optimization. *Bioinformatics*, 27(16):2311–2313, 2011.

[124] Jonathan R. Karr, Jayodita C. Sanghvi, Derek N. MacKlin, Miriam V. Gutschow, Jared M. Jacobs, Benjamin Bolival, Nacyra Assad-Garcia, John I. Glass, and Markus W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[125] Xiaoqiang Sun, Jiguang Bao, Zhuhong You, Xing Chen, and Jun Cui. Modeling of signaling crosstalk-mediated drug resistance and its implications on drug combination. *Oncotarget*, 7(39):63995–64006, 2016.

[126] Li Ang Zhang, Alisa Urbano, Gilles Clermont, David Swigon, Ipsita Banerjee, and Robert S. Parker. APT-MCMC, a C++/Python implementation of Markov Chain Monte Carlo for parameter identification. *Computers and Chemical Engineering*, 110:1–12, 2018.

[127] Bharat B. Aggarwal. Signalling pathways of the TNF superfamily: A double-edged sword. *Nature Reviews Immunology*, 3(9):745–756, 2003.

[128] Taro Kawai and Shizuo Akira. Signaling to NF-κB by Toll-like receptors. *Trends in Molecular Medicine*, 13(11):460–469, 2007.

[129] Lauren M. Workman and Hasem Habelhah. TNFR1 signaling kinetics: Spatiotemporal control of three phases of IKK activation by posttranslational modification. *Cellular Signalling*, 25(8):1654–1664, 2013.

[130] Matthew Hayden and Sankar Ghosh. Signaling to NF-kappaB. *Genes & development*, 18(18):2195–2224, 2004.

[131] Heike L. Pahl. Activators and target genes of Rel/NF-κB transcription factors. *Oncogene*, 18(49):6853–6866, 1999.

[132] Harald Wajant and Peter Scheurich. TNFR1-induced activation of the classical NF-κB pathway. *FEBS Journal*, 278(6):862–876, 2011.

[133] Shailaja Kasibhatla, Thomas Brunner, Laurent Genestier, Fernando Echeverri, Artin Mahboubi, and Douglas R. Green. DNA damaging agents induce expression of Fas ligand and subsequent apoptosis in T lymphocytes via the activation of NF-κB and AP-1. *Molecular Cell*, 1(4):543–551, 1998.

[134] Toby Lawrence. The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harbor perspectives in biology*, 1(6), 2009.

[135] Paul P. Tak and Gary S. Firestein. NF-κB: A key role in inflammatory diseases. *Journal of Clinical Investigation*, 107(1):7–11, 2001.

[136] Claire E. Lewis and Jeffrey W. Pollard. Distinct role of macrophages in different tumor microenvironments. *Cancer Research*, 66(2):605–612, 2006.

[137] Jean Marx. Inflammation and cancer: The link grows stronger. *Science*, 306(5698):966–968, 2004.

[138] D. Schottenfeld and J. Beebe-Dimmer. Chronic Inflammation: A Common and Important Factor in the Pathogenesis of Neoplasia. *CA: A Cancer Journal for Clinicians*, 56(2):69–83, 2006.

[139] Arjun Raj and Alexander van Oudenaarden. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(2):216–226, 2008.

[140] Mads Kærn, Timothy C. Elston, William J. Blake, and James J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[141] Alexander Hoffmann, Andre Levchenko, Martin L. Scott, and David Baltimore. The IκB-NF-κB Signaling Module: Temporal Control and Selective Gene Activation. *Science*, 298(5596):1241—-1245, 2002.

[142] Bing Tian, David E. Nowak, Mohammad Jamaluddin, Shaofei Wang, and Allan R. Brasier. Identification of direct genomic targets downstream of the nuclear factor-κB transcription factor mediating tumor necrosis factor signaling. *Journal of Biological Chemistry*, 280(17):17435–17448, 2005.

[143] Bing Tian, David E. Nowak, and Allan R. Brasier. A TNF-induced gene expression program under oscillatory NF-κB control. *BMC Genomics*, 6, 2005.

[144] Shannon L. Werner, Derren Barken, and Alexander Hoffmann. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science*, 309(5742):1857–1861, 2005.

[145] Louise Ashall, Caroline A. Horton, David E. Nelson, Pawel Paszek, Claire V. Harper, Kate Sillitoe, Sheila Ryan, David G. Spiller, John F. Unitt, David S. Broomhead, Douglas B. Kell, David A. Rand, Violaine Sée, and Michael R.H. White. Pulsatile stimulation determines timing and specificity of NF-κB-dependent transcription. *Science*, 324(5924):242–246, 2009.

[146] Ryan A. Kellogg, Chengzhe Tian, Tomasz Lipniacki, Stephen R. Quake, and Savaş Tay. Digital signaling decouples activation probability and population heterogeneity. *eLife*, 4, 2015.

[147] Samuel Zambrano, Ilario de Toma, Arianna Piffer, Marco E. Bianchi, and Alessandra Agresti. NF-κB oscillations translate into functionally related patterns of gene expression. *eLife*, 5, 2016.

[148] Tomasz Lipniacki, Krzysztof Puszynski, Pawel Paszek, Allan R Brasier, and Marek Kimmel. Single TNFalpha trimers mediating NF-kappaB activation: stochastic robustness of NF-kappaB signaling. *BMC Bioinformatics*, 8:376, 2007.

[149] David A. Turner, Pawel Paszek, Dan J. Woodcock, David E. Nelson, Caroline A. Horton, Yunjiao Wang, David G. Spiller, David A. Rand, Michael R.H. White, and Claire V. Harper. Physiological levels of TNFα stimulation induce stochastic dynamics of NF-κB responses in single living cells. *Journal of Cell Science*, 123(16):2834–2843, 2010.

[150] Marcelo Behar, Derren Barken, Shannon L. Werner, and Alexander Hoffmann. The dynamics of signaling as a pharmacological target. *Cell*, 155(2):448, 2013.

[151] Cellina Cohen-Saidon, Ariel A. Cohen, Alex Sigal, Yuvalal Liron, and Uri Alon. Dynamics and Variability of ERK2 Response to EGF in Individual Living Cells. *Molecular Cell*, 36(5):885–893, 2009.

[152] Jeremy E. Purvis, Kyle W. Karhohs, Caroline Mock, Eric Batchelor, Alexander Loewer, and Galit Lahav. p53 dynamics control cell fate. *Science*, 336(6087):1440–1444, 2012.

[153] Jeremy E. Purvis and Galit Lahav. Encoding and decoding cellular information through signaling dynamics. *Cell*, 152(5):945–956, 2013.

[154] Gal Chechik and Daphne Koller. Timing of gene expression responses to environmental changes. *Journal of Computational Biology*, 16(2):279–290, 2009.

[155] Garrett D. Potter, Tommy A. Byrd, Andrew Mugler, and Bo Sun. Dynamic Sampling and Information Encoding in Biochemical Networks. *Biophysical Journal*, 112(4):795–804, 2017.

[156] James E. Ferrell and Sang Hoon Ha. Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity. *Trends in Biochemical Sciences*, 39(10):496–503, 2014.

[157] Lea Goentoro and Marc W. Kirschner. Evidence that Fold-Change, and Not Absolute Level, of β-Catenin Dictates Wnt Signaling. *Molecular Cell*, 36(5):872–884, 2009.

[158] Ryan Suderman, John A. Bachman, Adam Smith, Peter K. Sorger, and Eric J. Deeds. Fundamental trade-offs between information flow in single cells and cellular populations. *Proceedings of the National Academy of Sciences of the United States of America*, 114(22):5755–5760, 2017.

[159] Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Joo Han Chang, Robert A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, 7(10), 2006.

[160] Steve Simon. SCATTERCLOUD. `https://www.mathworks.com/matlabcentral/fileexchange/6037-scattercloud,MATLABCentralFileExchange`.

[161] Paul H.C. Eilers and Jelle J. Goeman. Enhancing scatterplots with smoothed densities. *Bioinformatics*, 20(5):623–628, 2004.

[162] Sarah Filippi, Chris P. Barnes, Paul D.W. Kirk, Takamasa Kudo, Katsuyuki Kunida, Siobhan S. McMahon, Takaho Tsuchiya, Takumi Wada, Shinya Kuroda, and Michael P.H. Stumpf. Robustness of MEK-ERK Dynamics and Origins of Cell-to-Cell Variability in MAPK Signaling. *Cell Reports*, 15:2524–2535, jun 2016.

[163] Mario Niepel, Sabrina L Spencer, and Peter K Sorger. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Current Opinion in Chemical Biology*, 13:556–561, 2009.

[164] Jérémie Roux, Marc Hafner, Samuel Bandara, Joshua J Sims, Hannah Hudson, Diana Chai, and Peter K Sorger. Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold. *Molecular Systems Biology*, 11:803, may 2015.

[165] Deborah A. Flusberg, Jérémie Roux, Sabrina L. Spencer, and Peter K. Sorger. Cells surviving fractional killing by TRAIL exhibit transient but sustainable resistance and inflammatory phenotypes. *Molecular Biology of the Cell*, 24:2186–2200, jul 2013.

[166] Deborah A. Flusberg and Peter K. Sorger. Surviving apoptosis: Life-death signaling in single cells. *Trends in Cell Biology*, 25(8):446–458, aug 2015.

[167] Zintis Inde and Scott J Dixon. The Impact of Non-Genetic Heterogeneity on Cancer Cell Death. *Critical Reviews in Biochemistry and Molecular Biology*, 53(1):99–114, 2018.

[168] Kyoung Ae Kim, Sabrina L. Spencer, John G. Albeck, John M. Burke, Peter K. Sorger, Suzanne Gaudet, and Do Hyun Kim. Systematic calibration of a cell signaling network model. *BMC Bioinformatics*, 11:202, 2010.

[169] Jeremy E. Purvis and Galit Lahav. Encoding and decoding cellular information through signaling dynamics. *Cell*, 152:945–956, 2013.

[170] X. Xia, M. S. Owen, R. E.C. Lee, and S. Gaudet. Cell-to-cell variability in cell death: Can systems biology help us make sense of it all? *Cell Death and Disease*, 5:e1261, 2014.

[171] Zhang Cheng, Brooks Taylor, Diana R. Ourthiague, and Alexander Hoffmann. Distinct single-cell signaling characteristics are conferred by the MyD88 and TRIF pathways during TLR4 activation. *Science Signaling*, 8(385):ra69, 2015.

[172] L Naomi Handly, Jason Yao, and Roy Wollman. Signal transduction at the single-cell level: Approaches to study the dynamic nature of signaling networks. *Journal of Molecular Biology*, 428(19):3669–3682, 2016.

[173] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, and Garry P. Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

[174] Steffen Waldherr. Estimation methods for heterogeneous cell population models in systems biology. *J. R. Soc. Interface*, 15:20180530, 2018.

[175] Jan Hasenauer, Steffen Waldherr, Malgorzata Doszczak, Nicole Radde, Peter Scheurich, and Frank Allgöwer. Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics*, 12:125, 2011.

[176] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[177] Stefan M. Kallenberger, Joël Beaudouin, Juliane Claus, Carmen Fischer, Peter K. Sorger, Stefan Legewie, and Roland Eils. Intra- and interdimeric caspase-8 self-cleavage controls strength and timing of CD95-induced apoptosis. *Science Signaling*, 7(316):ra23, 2014.

[178] Carolin Loos and Jan Hasenauer. Mathematical modeling of variability in intracellular signaling. *Current Opinion in Systems Biology*, 16:17–24, 2019.

[179] Fabian Konrath, Anna Mittermeier, Elena Cristiano, Jana Wolf, and Alexander Loewer. A systematic approach to decipher crosstalk in the p53 signaling pathway using single cell dynamics. *PLOS Computational Biology*, 16(6):e1007901, 2020.

[180] Karolina Tudelska, Joanna Markiewicz, Marek Kochańczyk, Maciej Czerkies, Wiktor Prus, Zbigniew Korwek, Ali Abdi, Sławomir Błoński, Bogdan Kaźmierczak, and Tomasz Lipniacki. Information processing in the NF-κB pathway. *Scientific Reports*, 7:15926, 2017.

[181] Budiman Minasny. Latin Hypercube Sampling. `https://www.mathworks.com/matlabcentral/fileexchange/4352-latin-hypercube-sampling`, MATLABCentralFileExchange.

[182] Richard Crozier. tightfig(hfig). `https://www.mathworks.com/matlabcentral/fileexchange/34055-tightfig-hfig),MATLABCentralFileExchange`.

[183] Pekka Kumpulainen. tight subplot. `http://www.mathworks.com/matlabcentral/fileexchange/27991`, MATLAB Central File Exchange.

[184] Amiran Keshelava, Gonzalo P. Solis, Micha Hersch, Alexey Koval, Mikhail Kryuchkov, Sven Bergmann, and Vladimir L. Katanaev. High capacity in G protein-coupled receptor signaling. *Nature Communications*, 9, 2018.

[185] Eshan D. Mitra, Raquel Dias, Richard G. Posner, and William S. Hlavacek. Using both qualitative and quantitative data in parameter identification for systems biology models. *Nature Communications*, 9:3901, 2018.

[186] Hans Peter Fischer. Mathematical modeling of complex biological systems: From parts lists to understanding systems behavior. *Alcohol Research and Health*, 31(1):49–59, 2008.

[187] Marc H.V. Van Regenmortel. Reductionism and complexity in molecular biology. *EMBO Reports*, 5:1016–1020, 2004.

[188] Srdjan Kesić. Systems biology, emergence and antireductionism. *Saudi Journal of Biological Sciences*, 23:584–591, 2016.

[189] Fulvio Mazzocchi. Complexity and the reductionism-holism debate in systems biology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 4:413–427, 2012.

[190] Hiroaki Kitano. Systems Biology: A Brief Overview. *Science*, 295, 2002.

[191] Mariana Gómez-Schiavon and Hana El-Samad. Complexity-aware simple modeling. *Current Opinion in Microbiology*, 45:47–52, 2018.

[192] Fengkai Zhang, Bastian R Angermann, and Martin Meier-Schellersheim. The Simmune Modeler visual interface for creating signaling networks based on bi-molecular interactions. *Bioinformatics*, 29:1229–1230, 2013.

[193] Michael W. Sneddon, James R. Faeder, and Thierry Emonet. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nature Methods*, 8:177–183, 2011.

[194] Pierre Boutillier, Mutaamba Maasha, Xing Li, Hector F Medina-Abarca, Jean Krivine, Jérôme Feret, Ioana Cristescu, Angus G Forbes, and Walter Fontana. The Kappa platform for rule-based modeling. *Bioinformatics*, 34, 2018.

[195] Martin Meier-Schellersheim, Rajat Varma, and Bastian R. Angermann. Mechanistic Models of Cellular Signaling, Cytokine Crosstalk, and Cell-Cell Communication in Immunology. *Frontiers in Immunology*, 10, 2019.

[196] Tina Toni and Michael P H Stumpf. Parameter inference and model selection in signaling pathway models. *Methods in molecular biology (Clifton, N.J.)*, 673:283–295, 2010.

[197] Hans V. Westerhoff, Catherine Winder, Hanan Messiha, Evangelos Simeonidis, Malgorzata Adamczyk, Malkhey Verma, Frank J. Bruggeman, and Warwick Dunn. Systems Biology: The elements and principles of Life. *FEBS Letters*, 583:3882–3890, 2009.

[198] Fridolin Gross. Occam's Razor in Molecular and Systems Biology. *Philosophy of Science*, 86, 2019.